Science & Technology Studies

4/2019

Science & Technology Studies

ISSN 2243-4690

Subscriptions

Subscriptions and enquiries about back issues should be addressed to:

Email: johanna.hokka@uta.fi

The subscription rates (2019) for access to the electronic journal is 40 euros for individual subscribers and 100 euros for institutional subscribers.

Science & Technology Studies

Volume 32, Issue 4, 2019

Editorial

Articles

Book reviews

Visit our web-site at

www.sciencetechnologystudies.org

The Politics of Anticipatory Expertise: Plurality and Contestation of Futures Knowledge in Governance — Introduction to the Special Issue[1]

Stefan C. Aykut

Universität Hamburg, Faculty WISO, Germany/ stefan.aykut@uni-hamburg.de

David Demortain

INRA, Laboratoire Interdisciplinaire Sciences Innovations Sociétés (LISIS), France

Bilel Benbouzid

UPEM, Laboratoire Interdisciplinaire Sciences Innovations Sociétés (LISIS), France

The future is, as Arjun Appadurai (2013) puts it, a "cultural fact". Anticipatory practices – prophecies, utopias, predictions, forecasts, or scenarios – fulfil important societal functions in pre-modern and modern societies alike (Mannheim, 1991; Koselleck, 1985). They provide orientation in social interactions (Luhmann, 1997), shape expectations in market transactions (Merton, 1948; Beckert, 2013) and technology development (Borup et al., 2006), and constitute key resources for social organisation (Krämer and Wenzel, 2018) and political power (Scott, 1998; Mitchell, 2006; Ezrahi, 2012). Since World War II, such practices have undergone a process of *scientisation* and *professionalisation*. Formalised approaches draw on cybernetics, cognitive and behavioural sciences, gaming and econometrics, as well as techniques of computational modelling and simulation. They were forged in Cold War think tanks such as the US RAND Corporation (Andersson, 2012) and the International Institute for Applied Systems Analysis (IIASA) in Laxenburg, Austria (Rindzevičiūtė, 2016), as well as in Soviet, Dutch and French planning circles (Desrosières, 1999; Andersson and Rindzevičiūtė, 2015). New impetus for the development of such techniques came in the 1970s, as growing concerns about industrial, environmental, and health risks prompted policymakers and governance bodies to evaluate the long-term consequences of their decisions (Buttel et al., 1990; Dahan, 2007; Seefried, 2015). Anticipatory knowledge practices turn 'the future' into an object of scientific enquiry and political intervention, hence giving shape to novel forms of "anticipatory governance" (Guston, 2014; Anderson, 2010). By producing information about what has not yet happened (and may indeed never happen), they reduce social complexity and constitute problems for acting in the present (Mallard and Lakoff, 2011). In doing so, however, they also reduce the inherent openness of the future, and risk closing down the "horizon of the possible for social and political creation" (Schulz, 2015: 132).

In sum, anticipatory expertise has become an indispensable core ingredient of contemporary attempts to govern complex problems and exhibits features of an institutionalised *regulatory science* (Demortain, 2017). These include the standardisation of dominant knowledge practices and the emergence of professional cadres of experts forming a part of, or entertaining close ties to, the state apparatus. At the same time, it has also become a *field of contention*, wherein a variety of technologies of knowledge production, social groups, and visions of the future coexist and compete. Indeed, the production of predictions and forecasts for public policy is not only an epistemic endeavour, but also an intrinsically contested and political activity. Its internal diversity – most policy environments count a variety of available technologies for producing anticipatory expertise – tends to reflect the politics and power contests of the corresponding worlds of policy formulation and implementation. As a result, the landscape of anticipation-for-policy is rapidly evolving. Profound changes affect the production and validation of knowledge, as well as its use in policy environments. The diversity of actors, sites and knowledge practices involved in anticipatory expertise is increasing.

Hence, while the study of anticipatory knowledge practices is hardly a new theme in science and technology studies, we see the need to take stock of these new developments. We also believe that such an endeavour necessitates to extend the scope of existing STS research. Previous studies have mostly concentrated on isolated and/or dominant knowledge practices, without paying much attention to the dynamic interplay or competition between different forms of anticipatory knowledge and their relation to the policy process. The special issue therefore builds on ongoing scholarly discussions about anticipatory knowledge and its 'performativity', to take a fresh look at the *politics* of anticipatory expertise. It also goes beyond the political science literature, which has investigated the role of forecasts and predictions in the political struggles and framing contests that accompany policy formulation (e.g. Baumgartner and Midttun, 1987; Grunwald, 2009), while showing only limited interest in unpacking the social dynamics underlying model construction, or the validation practices that found the credibility claims of anticipatory knowledge. Various new developments in the interstitial space between these two, scarcely intersecting strands of literature thus remain unattended, which this introduction and the contributions comprising the special issue offer to identify and explore.

The performativity agenda

Scholars have thoroughly investigated how the growing reliance on model-based simulations and predictions has reshaped entire scientific disciplines (Morgan and Morrison, 1999) and fields of technological innovation (Van Lente and Rip, 1998). A highly influential way to capture this rise of anticipatory practices has been through the notion of *performativity* in economic sociology (Callon, 1998; Mackenzie, 2004). Following Mackenzie (2006: 250), a broad, generic sense of the notion simply points to the fact that numerical models and other artefacts created by economists are taken up in economic practice. In a stronger sense, they materially 'equip' economic transactions and 'format' the ways in which such markets function. To be performative, future knowledge travels through socio-technical "arrangements of prediction" (Schubert, 2015) – instruments, infrastructure, and shared practices – which bring together social actors and redefine their preferences. The performativity argument has be extended to the making of economic policy. Braun (2014: 51) argues that models are key ingredients of "governability paradigms", which stabilise "when a sufficiently large part of the macroeconomic discipline is in agreement over the causal relationships between instrument and target variables, as well as over the way in which the former should be used by policymakers." Macroeconomic models have also been found to play key roles in policy change (Henriksen, 2013; Angeletti, 2011): as new models come to be embedded in policy circles and administrative practice, they contribute to wider transformations in the ways in which problems are identified and discussed, and solutions designed and implemented.

Other policy domains and model-based sciences are similarly concerned by these developments. In an early study on IIASA's first global energy forecast, (Wynne, 1984: 277) argues that this predictive exercise was highly consequential for policy formulation, as it delimited "what policies are even conceivable". The complex formalism of the underlying energy model tacitly embeds, and thereby reproduces, the worldviews and normative assumptions of its architects, imbuing them with an aura of scientific objectivity. Similarly, climate change forecasts simultaneously define through which mechanisms climate risks occur, what and who is at risk, and which adaptation policies might be appropriate (Jasanoff, 2010). Kieken (2004) goes one step further in a study on integrated assessment models (IAMs) – a type of model widely used in climate policy assessments. He holds that it would be wrong to depict IAMs as primarily *scientific* objects. Much to the contrary, exerting political influence is a constitutive goal of the IAM community and a fundamental ambition of IAM modellers (see also Cointe et al., in this issue). Following on this, Beck and Mahony (2017) show that the projections produced by such models even bring into being new political objects – such as 'negative emission technologies' – and argue that the resulting "politics of anticipation" pose challenges to common conceptions of scientific neutrality.

Another way in which models 'perform' is by coordinating the various social worlds implicated in policymaking. Models and their productions function as *boundary objects* that structure collaboration between academic communities (Edwards, 2010), and between experts and politicians (van der Sluijs et al., 1998; van Egmond and Zeiss, 2010). The MARKAL energy model, for instance, has been found "to perform different roles for different groups", a capacity which "has served to embed and institutionalise the model in the energy policy community" (Taylor et al., 2014: 32). While this has assured the model a rare longevity in administrative practice, over the years its technological focus has also reinforced an existing bias toward technical solutions in energy and climate policy. Modelling of environmental hazards is another case in point. It is practiced by specialised consultants who have established privileged relationships with policy managers. In a study on flooding risks, Catharina Landström and colleagues show that even though modelling practices are rarely subject to academic peer review, modellers tacitly "define what society needs to know about flooding in order to undertake risk management that is considered satisfactory" (Landström et al., 2011: 18,19).

Pluralizing performativity

The emergence of complex, socio-material "machineries" of anticipation (Nelson et al., 2008: 549) has wide-ranging consequences for policymaking. As it confers political influence to the expert communities that control the 'means of anticipation' – so to speak – the literature has also embraced normative considerations. Model-based forecasts have been criticised for depoliticising policymaking and public debate (Voß, 2013), silencing the voices of lay publics and local populations (Miller, 2004; Mahajan, 2008) and restricting the expression of alternative imaginaries of the future (Jasanoff, 2010). When techno-scientific or catastrophic future visions "colonise" or "abduct" the present (Adams et al., 2009: 255; Kaiser, 2015), they contribute to processes of "de-futurisation" (Luhmann, 1990). This raises important questions pertaining to the accountability of modellers, and to possible ways of 'democratising' anticipatory expertise by associating wider publics in modelling or scenario building. It also points to the need to gain a better understanding of the diversity of ways of forging futures in policy the variety of actors involved.

No longer the monopoly of a few academic or state institutions, quantified future visions are produced within broader networks spanning public agencies and global governance bodies, scientific institutes and think tanks, as well as firms and civil society organisations (Voß et al., 2006; Guston, 2014). Policy intermediaries and knowledge brokers invest in anticipatory practices to sustain their role in changing policy environments, while transnational organisations and governance bodies play key roles in processes of model evaluation, validation and standardisation. Anticipatory knowledge production hence spans a variety of institutional loci at local, national and transnational scales. This geographic and spatial diversity in turn imposes differentiated requirements on the design and scope of model architectures. In addition, sharp increases in computing power and data availability have renewed existing knowledge practices, and led to the emergence of new ones. Classical quantitative models based on the law of large numbers and the associated notions of norms and means (Desrosières, 2000) now compete with machine-learning based techniques in which the model is no longer an input into the calculation, but an output (Cardon, 2015), as they proceed by testing all possible correlations between an ever-increasing number of features. As a result, predictions, forecasts and scenarios in many policy domains now form "ecologies" or "assortments of futures" (Michael, 2017) among which policy actors, stakeholders and activist groups can choose and within which they must navigate. Moreover, as the diversity of actors, instruments and governance scales in policymaking increases, modelling techniques tend to vary in form and content depending on the political context of knowledge production, and on the demonstrations that those who use models and their outputs are interested in making.

By contrast, the theme of performativity has typically been applied to situations characterised either by the existence of one dominant knowledge practice, or by a privileged relationship between a producer of anticipatory expertise and a (political) centre of decision-making – often the state (e.g. Nelson et al., 2008). We therefore feel the need to enrich the focus on performativity through a political sociology lens, which pays closer attention to conflicts and power asymmetries in knowledge production, as well as normalisation and regulation activities by public agencies and private actors (Frickel and Moore, 2006; Bonneuil and Joly, 2013). The *political sociology of anticipatory expertise* that we defend approaches predictions, forecasts and scenarios as one of the many 'currencies' mobilised by competing actor groups seeking to bear on the governance of public problems. As with other forms of policy-relevant knowledge, the social dynamics of modelling fields therefore reflect the politics of policymaking, the variety of intentions and actors involved, the power struggles among them, and wider shifts in policy frames. How do *performance contests*, so to speak, play out? How do various forms of anticipatory expertise co-construct or exclude each other in policymaking and governance? Amongst a diversity of anticipatory instruments and knowledge practices, what explains the success of some techniques and the failure of others? And how do such epistemic conflicts shape knowledge production in the first place?

The contributions to the special issue

To develop a differentiated take on contemporary transformations affecting anticipatory expertise, the articles assembled here span a variety of governance scales (local, national, global), countries (the US and the European Union, France, Germany and Belize) and policy domains (energy, climate, agriculture and food policy, risk regulation, forest and water management, policing). They cover different forms of knowledge production, ranging from formalised and computational to non-formalised, lay practices of anticipation. The authors also approach their objects from different theoretical and methodological perspectives, combining STS approaches with ethnography, history, political science, and sociology, while using analytical frameworks as diverse as coproduction, performativity, and field theory. All papers study the dynamics of anticipatory expertise against the backdrop of evolution in the corresponding policy field, to better understand how such expertise becomes embedded in, and co-constitutes the governance of, complex and contested policy issues. As the following summaries cannot do justice to the empirical richness and analytical diversity of these studies, we invite the interested reader to take a closer look at the articles themselves.

Three contributions focus on forecasts and scenarios in energy, climate and agriculture governance. Stefan Aykut provides a historical study on German and French energy policy, which links evolutions in energy modelling to key moments of policy change. Energy policy is envisioned as a field populated by competing "predictive policy assemblages" made up of discourses, human agents, knowledge practices and material artefacts. Dynamics of model development therefore tend to reflect wider political struggles: in the post-war decades, energy models helped constitute 'energy policy' as an autonomous policy domain structured around a specific representation of the energy system and of available policy options. This changed in the 1970s, when activist groups 'equipped' with new modelling techniques proposed alternative future-visions and political ontologies, which enabled new forms of political intervention. The article concludes by specifying a series of conditions under which changes in knowledge practices can be expected to contribute to wider policy change. Béatrice Cointe, Christophe Cassen and Alain Nadai examine how IAMs became over the past decades the main tool for producing emissions scenarios for the Intergovernmental Panel on Climate Change (IPCC). IAMs represent interactions among environmental, technological and human systems in a single integrated framework. The authors retrace the structuration of the modelling community around both a common normative commitment to produce 'policy-relevant' knowledge, and of a shared 'repertoire' of organisational techniques and knowledge infrastructures, such as the creation of a research consortium, common scenario databases, and model intercomparison projects. This, they argue, created a wider convergence of research practices and agendas among scientists involved in climate expertise. It also anchored this nascent modelling technique and the corresponding epistemic community in global climate governance. Lise Cornilleau's contribution shows that similar dynamics are at work in global agriculture and food security governance. Drawing on neo-institutionalist and Bourdieusian field theory, she examines the processes through which certain knowledge practices come to be considered as more legitimate than others in policy contexts. Empirically, the paper centres on the competition between two distinct modelling communities. It narrates how the architects of a new model-type attempted to gain a better position in global expertise by emulating central features of the dominant modelling approach. The study demonstrates that modellers act as 'institutional entrepreneurs', actively creating and maintaining structural homologies between fields of expertise and governance. Echoing Cointe et al.'s central argument, Cornilleau contends that dominant modellers strategically used intercomparison exercises to maintain the high ground in policy advice, by setting implicit standards for modelling techniques that suit the incumbent expert community.

Two articles look into the prediction and management of environmental resources. Antoine Dolez, Céline Granjou and Séverine Louvel investigate forest science and management in France. Anticipatory expertise in this domain does not form a monolithic whole, as well-established practices of knowing and governing forest development coexist alongside newer ones that emerged in the context of climate change debates. The authors identify three "micro-regimes of anticipation": the adaptation of forestry to future climates; the prediction of tree biology; and the monitoring of forests as indicators of climate change. This diversity, they contend, both reflects the impact of the "big future" of climate change on knowledge production, while also pointing to wider changes in forest management, as different expert communities tend to maintain privileged relations with policy actors. Mapping such regimes helps understand the evolution, interaction and hybridisation of knowledge practices, as well as the conflicting politics of environmental anticipation. Sophie Haines offers an ethnographic study of anticipatory water management in Belize. Alongside formalised predictions based on statistics and modelling, she foregrounds the ways in which scientists, practitioners and policymakers navigate water futures through relational "reckoning" work. Necessary to "mak[e] measures and measurement meaningful", such work rests on a scientific ambition to know and predict, and on artefacts like datasets, models, and maps. However, it is also embedded in a complex web of social relations, which are shaped by political and economic struggles, and are thus affective, situated, and experiential. Haines argues that the indeterminacies surrounding data and its interpretation, as well as the political use of predictions, frequently lead to frictions, disorientation, and discontent. This material, social, and emotional context shapes what she calls the "political lives of anticipation". Anticipatory governance, the article demonstrates, has as much to do with scientific knowledge, as with situated practices of coping with non-knowledge and uncertainty.

The three final papers consider predictive modelling in risk regulation and security interventions. Bilel Benbouzid focuses on a commercially successful software package for predicting crime, PredPol. He analyses the construction and dissemination of the PredPol algorithm against the backdrop of other predictive technologies that motivated and inspired its development: statistical systems for identifying crime 'hotspots', and algorithms used by geologists to predict earthquakes and seismic aftershocks. By opening the black box of PredPol's mathematical composition – partially, as the algorithm is not public – Benbouzid shows how it embeds specific normative assumptions and policy frames. PredPol aims to 'optimise' police work, which makes it compatible with neo-managerial cost reduction efforts. In addition, PredPol's way of predicting crime independently of any consideration of the underlying social dynamics obfuscates the social causes that drive criminal behaviour. Henri Boullier, David Demortain, and Maurice Zeeman zoom in on modelling practices at the U.S. Environmental Protection Agency (EPA). The EPA has in recent decades become the site of the formalisation of a new practice of chemical hazards' prediction, derived from the modelling of statistical relationships between chemical structures and their biological activity. This might seem surprising, as the agency operates in a highly constrained epistemic environment: legal frameworks combine with organisational cultures to define the kind of knowledge that is routinely used, and limit the ability to embrace other types of information. The study examines against this backdrop how the EPA succeeded in turning highly uncertain and experimental modelling techniques into credible regulatory knowledge. It also shows that while this so-called SAR technique has come to occupy a central position in chemical's regulation, it does not predetermine risk regulation. Instead, modelling articulates with other ways of establishing chemical risks, such as empirical experiments. Brice Laurent and François Thoreau further explore chemical hazards modelling, this time in the European context. They also find that modelling of structure-agency relations does not replace expert judgment. It requires large doses of human intervention, for instance to determine the chemical substance to be modelled in each new case, or to correct problems of over- or underfitting with external data through parameterisations. It therefore constitutes more than a simple mechanical tool, and tends to resists standardisation. Laurent and Thoreau argue that this feature of chemical hazards' modelling does not easily fit with the ideal of "mechanical objectivity" (Porter, 1995) that structures the European Union's risk governance. They also highlight that model-based regulation of chemicals poses critical questions of transparency, as it institutes power asymmetries between model developers in private companies interested in avoiding public scrutiny of their knowledge practices, and experts in the regulatory agencies in charge of assessing the models.

Unpacking the politics of anticipatory expertise

Despite the plurality of objects and approaches, a series of common themes and insights emerges from the articles. First, *model-based predictions form part of, and are embedded in, a larger set of anticipatory practices* that inform contemporary policymaking. The papers show that anticipatory expertise cannot be reduced to quantified predictions, nor to a single epistemic culture. It comprises various knowledge practices, tools, and organisations. New modelling techniques are commonly related to, contested by, and constructed against other models and claims about the future. Instead of replacing established (lay or expert) practices of anticipation, they complement them by addressing new questions, or produce 'frictions' when they challenge socially entrenched practices of anticipation. They also require new forms of judgment and human intervention, to calibrate them, contextualise their knowledge claims, and interpret their results.

Second, *anticipatory expertise intervenes in all phases of the policy process*. Modellers are consistent actors in policymaking, not occasional advisers brought into the process upon request by policymakers. This is not to say that anticipatory knowledge determines policymaking. Neither on top nor "on tap" – to paraphrase Winston Churchill – for policymakers, anticipation experts actively participate in policy formulation. Their tools, judgements and simulations are among the information sources and framing devices that shape public debates and agenda setting. They are also routinely used in policy implementation, for instance to calibrate policy instruments or monitor the impact of political measures.

Third, *anticipatory knowledge is produced within sociotechnical 'collectives' or 'assemblages'*. Knowledge about the future is inherently uncertain and thus particularly exposed to critique. To become a relevant and lasting feature of policymaking, it needs to be validated according to collectively held norms of credibility, through processes involving governmental assessment bodies, global standardisation organisations, and intercomparison projects. It also relies on 'infrastructures of anticipation' – in Edward's (2010) sense of vast machines of technical artefacts, epistemic infrastructures and social institutions – within which models and their results are controlled, compared and interpreted. This in turn creates path dependences in terms of knowledge practices and expert communities. In most of the cases analysed in the special issue, the tools, practices, and organisations involved in collecting data, constructing models and validating them, and producing anticipatory expertise, all have long histories.

Fourth, *fields of anticipatory expertise are driven by struggles for hegemony* between different knowledge practices and their proponents. The dynamics of such competition shape the relations between expert collectives and produce hierarchies among them. On one extreme, we find 'open markets' for expertise, in which a wide variety of knowledge producers compete for public attention and contracts. Here, entry barriers tend to be low, and validation and standardisation processes collectively negotiated. In cases where the capacity to produce formalised, authoritative knowledge about the future is more unevenly distributed, the field of expertise can take an 'oligopolistic' structure, or, on the other extreme, be dominated by a (near-)hegemonic knowledge practice.[2] In these cases, the diversity of knowledge about the future is channelled by a general drive toward the standardisation of knowledge production, as well as by the fact that anticipation is sustained by and embedded in material artefacts and technical infrastructures, which constitute potent entry barriers for potential newcomers. Existing sociotechnical infrastructures of anticipation restrict competition in futures and policy knowledge.

Fifth, the *dynamics of knowledge production reflect central features of policymaking* in a given domain. The papers in this special issue show that anticipatory expertise in risk regulation (toxicology), strategic planning (energy, climate), administrative management (forest, water policies) and security interventions (policing) takes very different forms, depending on the actor configurations and prevalent modes of policy intervention at play in public policy. This raises interesting questions as to the ways in which expert communities and anticipatory practices articulate with policymaking and governance, and in which predictions, forecasts, and scenarios come to be translated into policy-relevant knowledge that circulates outside expert communities to become an integral element of policymaking.

Lastly, *field-specific relations between experts and policy actors structure knowledge production and uptake*. Such "patterns of interaction" (Miller, 2001) enclose distinct ways of envisioning and organising the translation of scientific knowledge into policy-relevant expertise. In climate governance for instance, modelling communities are formally incorporated into a larger "governance apparatus" (Feldman, 2011) through the mediation of technical assessment bodies. In energy policy, modellers tend to merge with competing policy networks by establishing privileged relationships with specific policy actors. This contrasts with fields like predictive policing, where relations between knowledge producers and users are only weakly formalised and more fluid. Such cases tend to accommodate a greater variety of knowledge practices. As these examples show, intermediary organisations of assessment, comparison, and standardisation play a critical role in structuring and strengthening the science-policy link. Accordingly, such organisations constitute key sites in struggles over economic resources and political influence. Where they do not exist, knowledge users typically struggle to evaluate competing predictive techniques, as knowledge producers have little interest in disclosing their epistemic practices.

Taken together, the papers in this special issue illustrate the productivity of a perspective that combines STS and political sociology to gain a finer-grained understanding of anticipatory expertise in public policy. Such an analytical angle, we contend, leads beyond visions of anticipatory governance either as a process in which 'enlightened' policymakers base their decisions on rational assessments of the long-term consequences of different policy options, or as a process in which modelling experts and their predictions indistinctly 'depoliticise' public debate and predetermine policymaking. As models, forecasts, and algorithms have become a common – and in many ways indispensable – feature of contemporary policy debates, we hope that this special issue will stimulate further research that jointly analyses the politics of anticipatory knowledge production and the multiple ways in which such knowledge informs, intervenes in, and contributes to shaping the governance of public problems.

References

Adams V, Murphy M and Clarke AE (2009) Anticipation: Technoscience, life, affect, temporality. *Subjectivity* 28(1)**:** 246-265.

Anderson B (2010) Security and the future: Anticipating the event of terror. *Geoforum* 41(2)**:** 227-235.

Andersson J (2012) The Great Future Debate and the Struggle for the World. *American Historical Review* 117(5)**:** 1411-1431.

Andersson J and Rindzevičiūtė E (2015) *The Struggle for the Long Term in Transnational Science and Politics. Forging the Future.* New York: Routledge.

Angeletti T (2011) Faire la réalité ou s'y faire ? La modélisation et les déplacements de la politique économique au tournant des années 1970. *Politix* 2011/3(95)**:** 47-72.

Appadurai A (2013) *The Future as Cultural Fact. Essays on the Global Condition*. London, UK: Verso.

Baumgartner T and Midttun A (1987) *The politics of energy forecasting. A comparative Study of Energy Forecasting in Western Europe and North America.* Oxford: Clarendon Press.

Beck S and Mahony M (2017) The IPCC and the politics of anticipation. *Nature Climate Change* 7(5)**:** 311-313.

Beckert J (2013) Imagined Futures. Fictional Expectations in the Economy. *Theory and Society* 42(3)**:** 219-240.

Bonneuil C and Joly P-B (2013) *Sciences, techniques et société.* Paris: La Découverte.

Borup M, Brown N, Konrad K and Van Lente H (2006) The sociology of expectations in science and technology. *Technology Analysis and Strategic Management* 18(3/4)**:** 285-298.

Braun B (2014) Why models matter: the making and unmaking of governability in macroeconomic discourse. *Journal of critical globalisation studies* 7**:** 48-79.

Buttel FH, Hawkins AP and Power AG (1990) From Limits to Growth to Global Change: Constraints and contradictions in the evolution of environmental science and ideology. *Global Environmental Change* 1**:** 57-66.

Callon M (1998) *The Laws of the Markets.* Oxford, UK: Oxford University Press.

Cardon D (2015) *A quoi rêvent les algorithmes. Nos vies à l'heure des big data.* Paris: Le Seuil.

Dahan A (2007) *Les modèles du futur.* Paris: La Découverte.

Demortain D (2017) Expertise, regulatory science and the evaluation of technology and risk: Introduction to the Special Issue. *Minerva* 55(2)**:** 139-159.

Desrosières A (1999) La commission et l'équation: une comparaison des Plans français et néerlandais entre 1945 et 1980. *Genèses* 34**:** 28-52.

Desrosières A (2000) *The Politics of Large Numbers: A History of Statistical Reasoning.* Cambridge, MA: Harvard University Press.

Edwards PN (2010) *A vast Machine. Computer Models, Climate Data, and the Politics of Global Warming.* Cambridge, MA: The MIT Press.

Ezrahi Y (2012) *Imagined Democracies: Necessary Political Fictions.* Cambridge and New York: Cambridge University Press.

Feldman G (2011) Illuminating the apparatus: Steps toward a nonlocal ethnography of global governance. In: Shore C, Wright S and Però D (eds) *Policy worlds: anthropology and the analysis of contemporary power.* New York, Oxford: Berghahn Books, pp. 32-49.

Frickel S and Moore K (2006) *The New Political Sociology of Science- Institutions, Networks and Power.* Madison: The University of Wisconsin Press.

Grunwald A (2009) Energiezukünfte vergleichen – aber wie? In: Möst D, Fichtner W and Grunwald A (eds) *Energiesystemanalyse*. Karlsruhe: Universitätsverlag Karlsruhe, pp. 33-47.

Guston DH (2014) Understanding 'anticipatory governance'. *Social Studies of Science* 44(2)**:** 218-242.

Henriksen LF (2013) Economic models as devices of policy change: Policy paradigms, paradigm shift, and performativity. *Regulation & Governance* 7(4)**:** 481-495.

Jasanoff S (2010) A New Climate for Society. *Theory, Culture & Society* 27**:** 233-253.

Kaiser M (2015) Über Folgen. Technische Zukunft und politische Gegenwart. Weilerswist-Metternich Velbrück Wissenschaft.

Kieken H (2004) RAINS: Modéliser les pollutions atmosphériques pour la négociation internationale. *Revue d'histoire des sciences* 57(2)**:** 377-406.

Koselleck R (1985) *Futures Past: On the Semantics of Historical Time.* Cambridge, MA: MIT Press.

Krämer H and Wenzel M (eds) (2018) *How Organizations Manage the Future: Theoretical Perspectives and Empirical Insights*. Basel: Springer.

Landström C, Whatmore SJ and Lane SN (2011) Virtual engineering: computer simulation modelling for flood risk management in England. *Science & Technology Studies* 24(2)**:** 3-22.

Luhmann H-J (1997) *Die Gesellschaft der Gesellschaft*. Frankfurt am Main: Suhrkamp.

Luhmann N (1990) Die Zukunft kann nicht beginnen. Temporalstrukturen der modernen Gesellschaft. In: Sloterdijk P (ed) *Vor der Jahrtausendwende: Berichte zur Lage der Zukunft.* Frankfurt am Main: Suhrkamp, pp. 119-150.

Mackenzie D (2004) The Big, Bad Wolf and the Rational Market : Portfolio Insurance, the 1987 Crash and the Performativity of Economics. *Economy and Society* 33**:** 303-334.

Mackenzie D (2006) *An Engine, Not a Camera. How Financial Models Shape Markets.* Cambridge, MA: MIT Press.

Mahajan M (2008) Designing epidemics: models, policy-making, and global foreknowledge in India's AIDS epidemic. *Science and Public Policy* 35(8)**:** 585-596.

Mallard G and Lakoff A (2011) How claims to know the future are used to understand the present. In: Camic C, Gross N and Lamont M (eds) *Social Knowledge in the Making.* Chicago: University of Chicago Press, pp. 339-378.

Mannheim K (1991) *Ideology and utopia*. London: Routledge.

Merton RK (1948) The Self-fulfilling Prophecy. *The Antioch Review* 8(2)**:** 198-210.

Michael M (2017) Enacting Big Futures, Little Futures: Toward an ecology of futures. *The Sociological Review* 65(3)**:** 509-524.

Miller CA (2001) Scientific Internationalism in American Foreign Policy: The Case of Meteorology (1947-1958). In: Miller CA and Edwards PN (eds) *Changing the Atmosphere. Expert Knowledge and Environmental Governance*. Cambridge, MA: MIT Press.

Miller CA (2004) Resisting Empire: Globalism, Relocalization, and the Politics of Knowledge. In: Long-Martello M and Jasanoff S (eds) *The Patchwork Planet: Local and Global in Environmental Politics.* Cambridge, MA: MIT Press.

Mitchell T (2006) Society, economy and the state effect. In: Sharma A and Gupta A (eds) *The anthropology of the state: a reader.* Malden, MA: Blackwell Publishing, pp. 169-186.

Morgan M and Morrison M (eds) (1999) *Models as Mediators. Perspectives on Natural and Social Science.* Cambridge, UK: Cambridge University Press.

Nelson N, Geltzer A and Hilgartner S (2008) Introduction: the anticipatory state: making policy-relevant knowledge about the future. *Science and Public Policy* 35(8)**:** 546-550.

Porter T (1995) *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, N.J: Princeton University Press.

Rindzevičiūtė E (2016) *The power of systems. How policy sciences opened up the cold war world.* Ithaca and London: Cornell University Press.

Schubert C (2015) Situating technological and societal futures. Pragmatist engagements with computer simulations and social dynamics. *Technology in Society* 40**:** 4-13.

Schulz MS (2015) Future moves: Forward-oriented studies of culture, society, and technology. *Current Sociology Monograph* 63(2)**:** 129-139.

Scott JC (1998) *Seeing like a state: How certain schemes to improve the human condition have failed.* New Haven, CN: Yale University Press.

Seefried E (2015) *Zukünfte. Aufstieg und Krise der Zukunftsforschung 1945-1980.* Berlin: De Gruyter.

Taylor PG, Upham P, Mcdowall W and Christopherson D (2014) Energy model, boundary object and societal lens: 35 years of the MARKAL model in the UK. *Energy Research & Social Science* 4**:** 32-41.

Van Der Sluijs J, Van Eijndhoven J, Shackley S and Wynne B (1998) Anchoring Devices in Science for Policy. The Case of Consensus around Climate Sensitivity. *Social Studies of Science* 28(2)**:** 291-323.

Van Egmond S and Zeiss R (2010) Modeling for policy science-based models as performative boundary objects for Dutch policy making. *Science Studies* 23(1)**:** 58-78.

Van Lente H and Rip A (1998) Expectations in Technological Developments: An Example of Prospective Structures to be Filled in by Agency. In: Disco C and Van Der Meulen B (eds) *Getting New Technologies Together: Studies in Making Sociotechnical Order.* Berlin: Walter de Gruyter, pp. 202-230.

Voß J-P (2013) Technikwissenschaftliches Wettrüsten für nachhaltige Entwicklung? Wo transformative Wissenschaft hinführen könnte. Ökologisches Wirtschaften 2013(2)**:** 28-29.

Voß J-P, Bauknecht D and Kemp R (2006) *Reflexive Governance for Sustainable Development*. Cheltenham: Edward Elgar.

Wynne B (1984) The institutional context of science, models, and policy: the IIASA energy study. *Policy Sciences* 17(3)**:** 277-320.

Notes

1 This special issue derives from the research project INNOX - Innovation in Expertise. Modeling and simulation as Tools of Governance, supported by the French Agence Nationale de la Recherche (Grant n°ANR-13-SOIN-0005), between 2014 and 2017. The majority of papers included in this special issue were presented at various workshops held during this project or at its concluding event "Governing by prediction? Models, data and algorithms in and for governance" held in Paris in September 2017 (innox.fr). The three guest editors would like to thank the editors of the journal for supporting the publication of this special issue and the many colleagues who participated in the conversations that took place throughout the project and during the preparation of this special issue.

2 Hence, neoclassical economics has come to claim a quasi-"monopoly over the future" (Appadurai, 2013), partly through early investment in predictive techniques of computational modelling.

Reassembling Energy Policy: Models, Forecasts, and Policy Change in Germany and France[1]

Stefan C. Aykut

Faculty of business, economics and social sciences, Universität Hamburg, Germany & Laboratoire Interdisciplinaire Sciences Innovations Sociétés (UMR 1326), France / stefan.aykut@uni-hamburg.de

Abstract

Forecasts and scenarios calculated by energy system models are ubiquitous in energy debates. They are used by a wide range of public and private actors to make investment decisions, identify problems, and support or criticise specific forms of political intervention. The article presents an analytical framework for studying such entanglements between predictive practices and policy-making. Drawing on work in STS and the anthropology of politics, energy policy is conceptualised as a field of contention, populated by competing *predictive policy assemblages*. This concept is applied to a historical study on German and French energy policy-making, focusing on two periods. In the post-WWII decades, energy forecasts contributed to the structuring of 'energy policy' as an autonomous policy domain concerned with choosing between different energy *supply* options. This dominant paradigm was challenged in the 1970s and 1980s, when new modelling techniques forged by civil society groups brought energy demand and renewable energies to the fore politically and helped structure new political alliances. The article concludes by arguing that new ways of 'assembling' energy systems in models and forecasts can contribute to policy change, if they successfully 'perform' energy policy along three dimensions: by instituting alternative future-visions; by enabling new forms of political intervention; and by contributing to the formation of new 'predictive policy assemblages'.

Keywords: modelling, foreknowledge, policy change, performativity, policy assemblage.

Introduction

In a section entitled "The War of the Models" in *Adults in the Room*, a book on his time in office as the Greek finance minister during the European debt crisis, Yanis Varoufakis (2017: 603) recalls the following situation:

> … whenever I argued that in a struggling economy marred by poverty and tax evasion the best way to increase the state's revenues from VAT or from corporate tax was to *reduce* VAT and corporate tax rates, the troika would retort that their models showed the opposite: only by *increasing* the rate of VAT and corporate tax would tax revenue rise. And my country's Council of Economic Advisers, under Georg Chouliarakis, was using the same models to produce the same argument in favour of austerity. One day, incensed and incredulous, I asked to be allowed a glimpse inside the models. I was told that such models were complex, the implication being I would not understand, but I insisted: in a previous life I had been an econometrician, I replied.

This is a telling tale of the role of economic models in public policy. In a context requiring decisions about complex economic phenomena, such models provide political actors with a tool to evaluate possible directions of economic change, weigh the effects of alternative policy options, and legitimise potentially controversial decisions. The authority of such models stems from the professional prestige of economists and their seemingly objective quantifications, but also from their opacity to non-experts, which protects the underlying assumptions and worldviews from critique.

Like econometric models in economic policy, models and their outputs – forecasts, simulations or scenarios – play an increasingly important role in a wide range of policy fields (Nelson et al., 2008; Guston, 2014). This applies particularly to energy policy, where models have been used since the post-WW II decades to inform governments, energy utilities and the public about future trends in energy demand and supply, identify potentially problematic evolutions, and choose between different policy options (Baumgartner and Midttun, 1987). And yet, claims to objectivity in foreknowledge are circumscribed by the complexity of social processes and a range of 'if-then' assumptions that characterise the model-world. Consequently, scholarly debates soon questioned the 'knowability' of the future (Polak, 1973) and focused instead on the 'construction' of futures and the wider social and cultural settings in which these are embedded (e.g. Andersson and Rindzevičiūtė, 2015).

Some of the more productive current lines of research in this direction currently cluster around what could be called the 'performativity paradigm' in economic sociology (Callon, 2007; MacKenzie et al., 2007), which holds that the discipline of economics, rather than simply *describing* or *representing* economic activity, actively contributes to *shaping* it. This argument echoes longstanding sociological debates, on the constitutive nature of speech acts (Austin, 1962), the ways in which actors' definitions of a situation alter these very situations (Merton, 1948), and the self-validating and self-referential nature of social institutions (Barnes, 1983).

The paper attempts to widen the analytical scope of the "performativity idiom" (Pickering, 1995: 5) by connecting it to recent work in political anthropology on 'policy assemblages' (McCann and Ward, 2012). It envisions energy policy-making as a field populated by different *predictive policy assemblages*. This new concept points to material-semiotic constellations of actors, practices, discourses and material artefacts, which *compete* in the enactment of different energy futures. Such a perspective 're-embeds' models and forecasting practices in their socio-political environment, so as to better capture their circulation across social spaces and their involvement in administrative practice and policy debates. It also displaces the common focus in the performativity literature on dominant models and theories, to take into account alternative approaches. The concept hence allows obtaining a more fine-grained understanding of the conditions under which predictive practices indeed 'perform' successfully.

The first section of the paper develops this conceptual framework against the backdrop of the existing literature. The second and third sections apply the framework on a historical comparative study on the role of models and predictive practices in German and French energy policy. The study covers the period from the post-WWII decades to the late 1980s, and draws on an extensive document review, archival research, and 30 semi-structured interviews with energy modellers as well as public and private end-users of modelling results.[2] The article closes with general reflections on the role of predictive practices in energy policy-making, and on the interplay between evolutions in such practices and broader dynamics of policy change.

Performing in a contested environment: predictive policy assemblages

How does a focus on *predictive policy assemblages* change established ways of studying the role of predictive practices in energy policy-making? In what follows, I first review the existing literature to examine how different authors understand and operationalise performativity. I then introduce the notion of 'predictive policy assemblages' against the backdrop of recent critiques of the performativity paradigm. Lastly, I distinguish different components of predictive practices in the energy field and assess their role in policy-making.

Unpacking performativity: discourses, practices and social organisation

As to the question of how exactly predictive practices affect or 'perform' social reality, (at least) three broad approaches can be distinguished in the literature.

A first line of reasoning points to the ways in which foreknowledge influences the beliefs and expectations of political and economic actors (Beckert, 2013). Such a view informed one of the earliest social science analyses on energy modelling, which examined the making of the first global energy forecast in the late 1970s (Wynne, 1984; Thompson, 1984). The authors contend that normative assumptions built into model-design deeply biased the forecast, which in turn underpinned a policy paradigm centred exclusively on large-scale energy supply technologies. The argument here is that *model outputs* (forecasts and scenarios) influence the *discursive context* of policy-making by reducing the undetermined, 'open' future into an actionable set of 'plausible' development trajectories.

A second approach foregrounds that the production of forecasts and scenarios unfolds within organisational networks of state administrations, energy experts, firms and activists. In an early comparative study on energy forecasting in different Western countries, Baumgartner and Midttun (1987) show that ministerial forecasting committees, scenario-building exercises and participatory foresight groups constitute specific ways of assembling energy policy worlds and may either stabilise or, to the contrary, unsettle and recompose these networks. Such "social performativity" (Schubert, 2015) points not to discourses and beliefs, but to the ways in which *anticipatory exercises* affect *social organisation*.

A third school of thought – often labelled the 'new economic sociology' – has shifted the analytical focus from discourses, expectations and social organisation, to 'sociotechnical arrangements' and the use of material devices in economic practice. MacKenzie (2008) for instance shows how the Black-Scholes model of option pricing altered the functioning of stock markets, as traders equipped with such models tend to act very differently than 'naked' agents. The circulation of such models in turn draws the contours of peculiar social spaces, in which agents are linked not by shared causal beliefs (although this may be the case), but by the common use of a material-semiotic artefact, the numerical model. Here, it is claimed that *models themselves* shape social reality because they form part of *shared practices* of prediction and planning.

This last approach has proved particularly fertile in scholarship on the construction of markets, including energy markets (Silvast, 2017). However, a series of recent papers has taken a more critical stance towards the new paradigm, pointing to conceptual flaws (Mäki, 2013), theoretical shortcuts (Miller, 2002) and empirical problems (Brisset, 2016). While some of the critiques arguably only address a "stripped-down version" of the performativity thesis, which holds that markets would materialize more or less directly from economic theory (Silvast, 2017), others point to more substantial deficiencies. Hence, markets are also shaped by wider social institutions and political struggles (Cochoy et al., 2010), which may drive market design in quite different directions than economic theories. Furthermore, by (rightly) stressing the constitutive role of economics, the 'new economic sociology' may at times have underestimated the internal diversity of economic theory, which offers not a single reality to be enacted, but many (Henriksen, 2013). The question of which theory or model finally prevails and 'performs' reality hence constitutes a research puzzle in its own right.

Predictive policy assemblages

In other words, the performativity paradigm appears ill-equipped to properly account for the interplay between predictive practices and wider social and political dynamics, such as, for instance, the contentious *politics* of policy-making in a field like energy policy. Indeed, the bulk of social science research on foreknowledge has been dedicated to showing how dominant models and predictive practices reproduce dominant worldviews and stabilise social order (e.g. Callon, 1998; Mackenzie, 2006), while non-hegemonic models and practices have received far less attention. Conversely however, historical accounts of major shifts in energy policy mostly focus on political struggles between powerful actor coalitions (e.g. Unruh, 2000; Meadowcroft, 2009) and tend to overlook that actors involved in policy-making have to justify their decisions and policy preferences in the light of appropriate foreknowledge. Building capacities to produce foreknowledge therefore constitutes an important strategy that actors employ to advance their respective agendas (Chateauraynaud, 2013), and established practices of energy forecasting have historically been challenged by civil society groups, leading to a diversification of modelling sites and techniques (Aykut, 2015).

This is precisely where a perspective in terms of *predictive policy assemblages* complements existing approaches. "Policy assemblages" designate constellations of discourses, human bodies, social practices and material artefacts, in which specific forms of governing are "enacted" (McCann and Ward, 2012). This allows models and their material supports (computers, programs, databases), as well as resources and energy technologies, to be considered as central parts of wider political formations. Additionally, the assemblage metaphor draws attention not only to stable and formalised 'coalitions', but also to more loosely coupled ensembles. Their constitution, evolution and disappearance can be retraced by looking at processes of "translation" (production of equivalence, comparison, representation) and "linking" (associations, networks, compositions) between heterogeneous elements (DeLanda, 2006).

Placing predictive policy assemblages at the centre of the study of energy policy-making hence allows capturing both the central role and 'performative' effects of predictive practices, and the ways in which different actor-coalitions use models and forecasts in their quest for public attention and political influence. Instead of focussing on 'naked' actors and discourses, or on dominant models and their performative effects, it foregrounds the emergence and expansion of different actor-coalitions 'equipped' with their respective models and forecasts. By focusing on the *competition* of such assemblages for the enactment of distinct energy futures, such a view also opens up new perspectives on policy change. It allows an examination of the role of predictive practices in the formation, stabilisation, and transformation of dominant policy networks and paradigms, while also shedding light on the ways in which the emergence of new actor coalitions and problem framings may trigger innovations in model-design. Such innovations can in turn enable new forms of political intervention.

Predictive practices in the energy field

To understand how predictive practices intervene in the formation and competition of different predictive policy assemblages, we must further distinguish between the different components of such practices: energy models, databases, scenarios, and anticipatory exercises.

Energy (system) models emerged in the 1950s in the industrialised world, as the need for heavy investments in energy infrastructure drove the development of new planning and forecasting techniques. Part of a wider trend toward quantification and scientisation in public policy (Porter, 1995), such models isolate and 'represent' specific features of energy systems in stylised fashion, thereby constructing a 'mini-world' populated by a set of 'components' which are either endogenous (calculated by the model) or exogenous (external inputs), and which are related to each other in specific ways (e.g., linear or other forms of coupling). Since the 1970s, model development draws on methods from a wide array of disciplines, "including engineering, economics, operations research and management science" and uses different techniques, such as "mathematical programming (especially linear programming), econometrics and related methods of statistical analysis and network analysis" (Hoffman and Wood, 1976: 423). Despite this diversity, energy models broadly fall into two categories: top-down or economic models take an aggregate view and highlight the role of prices and markets in driving energy demand and supply. Bottom-up, process or engineering models stress the specificities of energy technologies and the technical determinants of energy demand. Such differences in model design have important political implications, as different model types tend to foreground different processes, and enable particular forms of political intervention.

Databases aggregate statistical time-series of real-world evolutions of key variables that drive energy demand and supply. They serve to establish basic relationships between different model components – usually a specific coefficient or 'elasticity' – and to 'calibrate' the model-world on observations (Edwards, 2010). Data production is a highly time-consuming process of assembling (sometimes purchasing), homogenising, and standardising heterogeneous information from different public and private sources. Databases thus form the backbone of energy models, which in turn inherit the value-laden categorical definitions of each of the initial datasets (Bowker, 2000).

Scenarios consist in coherent narratives about possible evolutions of the world that are used to simulate specific developments. Operationalised as quantified hypotheses for the evolution of key variables (e.g. GDP or population growth), such narratives are applied to the modelled mini-world. Scenarios broadly fall into three categories: *forecasts* extrapolate the most likely developments from existing trends; *exploratory scenarios* simulate specific changes or policy interventions; and *normative scenarios* aim to attain a specific policy objective. The term is ambiguous, as it also applies to the output of such model-simulations. In both cases, scenarios play a crucial role in mediating between models and their users. They make the abstract formalisations and quantifications of models intelligible, and also contribute to model-development, as a new scenario may demand the representation of new model components.

Lastly, *anticipatory exercises* designate a series of techniques through which scenarios are built and forecasts produced (Baumgartner and Midttun, 1987). As such exercises oftentimes associates major political, industrial, and civil society actors, they also contribute to the circulation and public uptake of forecasts. Conversely, discussion in such contexts may in turn stimulate new research and even alter model design (Angeletti, 2011). Practices of foreknowledge production thus involve not only *epistemic* representation (reconstructing energy systems in models), but also forms of *political* representation (reconstructing policy communities in forecasting committees).

To sum up, predictive practices influence energy policy-making in several distinct ways: scenarios and forecasts shape actors' expectations and provide them with 'actionable' future visions; energy models circulate in public and private planning practices and enable specific forms of political intervention; and anticipatory exercises assemble stakeholders in ways that may strengthen, sideline or rearrange existing policy communities. As I will show in the following sections, competing predictive policy assemblages can usefully be differentiated along one or several of these dimensions, as they often rely on different future-visions, use different types of models, and engage in distinct types of anticipatory exercises.

Making national energy (supply) policies (1950-1975)

In most Western countries, energy policy was characterized until the 1970s by the "energy syndrome" described by Leon N. Lindberg (1977), that is, increasing energy demand combined with weak national energy policies and the dominant role of energy utilities. This particularly applies to Germany, where the evolution of the energy sector was largely driven by industrial actors until the federal State claimed a more central role in the wake of the 1973 oil price shock. Such claims were underpinned by energy demand forecasts warning of a looming "energy gap" if no action was taken (e.g. Bundestag, 1979: 14809-14814). New modelling techniques also represented the national economy in greater detail, and examined processes of substitution between different energy carriers, especially domestic coal and imported oil. The result of these discussions was a progressive redefinition of the respective roles of the state and the market in energy policy. The French case is singularised by the existence, well before the 1970s, of nationalised energy companies and an institutionalised practice of forecasting and planning carried out by the French planning bureau (the *Commissariat général au Plan*, hereafter: CGP), which associated major stakeholders and administrations. Here too, energy demand forecasts played an important role in stabilising a new policy assemblage: produced by the monopoly of Electricité de France (EDF), they contributed to aligning political and economic actors on an acceleration of the nuclear program.

'Elasticity': calibrating France on a nuclear future

The structure of the French energy sector is the result of a historical process that led to the progressive institutionalisation, after World War Two, of a productivist, centralized energy policy paradigm (Lucas, 1985). This placed the state and a small number of nationalised energy utilities – especially the electricity monopolist EDF – at the centre of policy formulation and implementation. The nationalisation endowed these companies with a public function, and convinced their directors that the optimum for all and for their company were one and the same (Wieviorka and Trinh, 1991: 40). The intellectual coherence of this configuration of actors was ensured by the omnipresence of state engineers from the prestigious *Corps des Mines* in key positions in public companies and ministries, such as the all-mighty *Direction générale de l'énergie et des matières premières* (DGEMP). This cemented a relatively closed network that monopolized the decision-making process on energy and ensured coherence of discourse and values, centred on notions of technological *grandeur* and national independence (Hecht, 1998).

This institutional and intellectual context was paramount to the constitution of a French energy mix that is particularly atypical in its heavy reliance on nuclear energy. The foundations of this policy were laid in 1974, when the conservative government of Pierre Messmer (1972-1974) decided to accelerate the deployment of nuclear energy projects in the aftermath of the first oil price shock (Radanne, 2006). While it is generally argued that this new orientation resulted from considerations of energy dependence (Puiseux, 1982) this explanation overlooks an essential factor that made the energy dependence argument plausible in the first place: the discursive and political construction of future electricity demand.

The locus of French 'future-making' – in the double sense of knowledge production and political intervention – and the centre of a near-hegemonic 'predictive policy assemblage' at that time was the French planning bureau. Created in 1946, the CGP was unique among Western countries in associating major stakeholders – ministerial bureaucrats, industry representatives, and union leaders – and experts in a given policy domain to prepare five-year plans that should, in de Gaulle's words, serve as "orientation" not "coercion" for policy and investment decisions (Massé, 1965). Although the mobilisation of foreknowledge – from quantified forecasts to qualitative assessments – was commonplace in the CGP's various commissions, in-house models were exceptional until the 1980s (Angeletti, 2011). The practice of future-making institutionalised by CGP is described by Puiseux (1987), former head of the forecasting division of EDF and member of the CGP energy commission, as "technocratic elitism", in that it resembled more a cordial and expert-led "gentlemen's discussion" between high-ranking officials than a rigorous science-based assessment. In other words, the aim was not to 'discover' the most plausible future, but to collectively 'construct' a future that would at the same time prove reasonably plausible and acceptable enough to all that it could then be implemented collectively (Desrosières, 1999).

In line with political action horizons, the CGP produced 5-10 year energy demand forecasts, on which the state was to base its investment decisions (Château, 1985: 2). Estimates of future electricity demand were quite naturally provided by EDF, the only actor with the technical expertise, data, and modelling tools required for this task. With its status as a state-owned company, it could also claim to produce objective, non-biased results. The reluctance of the energy commission and relevant public administrations (especially DGEMP) to produce their own energy models or rely on independent expertise hence institutionalised an asymmetry in the production of authoritative knowledge claims about energy futures, which limited the discursive space of the deliberations. Bernard Laponche[3], a nuclear physicist working for the leading public nuclear research facility CEA, and who participated in the commission as a representative of the trade union CDFT, recalls:

> It was EDF who showed up saying "All right, I've made my forecasts, we need a trillion kilowatt hours in the year 2000", oh really. So Syrota[4] says "but we can find some savings," etc. […] and the Chairman says "come on, maybe you can at least explain this to us," and Boiteux says "well Mr. President sir, if you would like to look at the code I can have it delivered by truck"… (Interview 9).

As in other countries, the modelling techniques used by EDF at that time consisted mainly in more or less sophisticated extrapolations from the past, and reproduced the prevalent dogma in expert circles of a doubling of energy consumption every 10 years.[5] However, there also was a specific 'French touch' to EDF's estimates, which resulted from the ambiguous institutional status of the company. Marcel Boiteux, who directed EDF from 1967 to 1987 and was its first CEO without an engineering background,[6] had gradually modernized the company and provided it with a commercial and industrial strategy. Formerly director of the company's Department of General Economic Studies, he had championed a new approach to the calculation of electricity tariffs, which aimed to 'optimise' pricing and investment decisions by linking the investment-reimbursement cycle of plant construction to the evolution of electricity demand and the load profiles[7] of power plants (Romeiro, 1994: 27). The approach stressed the importance, from an industrialist's point of view, of the foreseeability of future electricity demand. This in turn transformed the nature of EDF's projections, as described vividly by Puiseux (1987: 190):

> On that day the chairman of the Energy Commission of the Planning Bureau suggested privately to me that if only EDF would decide to engage in somewhat more vigorous commercial activities, it would be possible substantially to increase the value of the GNP elasticity of electricity consumption. In this way the numbers which resulted from my regression calculations stopped being natural constants and became instead political action variables. This was quite a shattering discovery for a naïve soul.

To understand the distress of EDF's chief forecasting expert, recall that by the mid-1970s, a controversy opposed the electricity monopoly, which favoured an acceleration of the French nuclear program, and the Ministry of Finance, which was concerned about the associated investment risks. EDF backed its arguments with demand forecasts using consistently overestimated values for the 'elasticity' of electricity demand, i.e., the relationship between GNP growth and growth in electricity consumption (Château, 1985). The discrepancy between the modelled and observed relationship between these two variables became plainly visible in the 1970s, when electricity demand grew less than expected, and even stagnated briefly in 1974/1975. While this resulted partly from the oil price shocks and ensuing economic downturn, it also reflected a long-term evolution: economic growth in the after-war period, on which the models were calibrated, had been particularly electricity-intensive because of the imperatives of reconstruction, industrial development, and rural electrification. In the 1970s France entered a new era, in which the basic relationships between key variables changed.

The company did not respond to this discrepancy by adjusting its models to observed changes. Instead, a public campaign for household electrification, summed up by Boiteux's famous slogan *tout* électrique*, tout nucléaire* – "all electric, all nuclear" – was designed to ensure that the electricity intensity of economic development would be in phase with EDF's industrial strategy (Romeiro, 1994). Backed by the state, its implementation temporarily restored the relationship between electricity demand and economic growth observed in the 1960s (Puiseux, 1987: 193). In other words, the French economy had successfully been 'calibrated'[8] by the dominant predictive policy assemblage to fit EDF's models and official forecasts based on the company's calculations.

'Substitution': the making of German energy policy

The German energy sector has historically been structured around private or semi-public energy utilities with regional monopolies. This mode of organisation was codified in 1935, in an energy bill (*Gesetz zur Förderung der Energiewirtschaft*, EnWG) that excluded economic competition and aimed instead to ensure a stable energy supply and the construction of power grids in a context of ongoing war preparations. As the regulation of the sector was in the hands of the federated *Länder*, Germany did not have a genuine federal energy policy before the 1970s (Stier, 1999). The limited role of the federal level did not mean, however, that the state was not engaged in multiple ways in energy governance – through subsidies and funding for energy-related research; legislative or administrative rule-making affecting the building of transmission lines, power plants, and resource-extraction; and even as a market actor, through utilities that were partly or wholly owned by municipalities or *Länder*. This entanglement of regulated energy companies and the state administrations that regulated them was a characteristic feature of German corporatism (Beyer, 2002). It created a complex terrain for energy policy, whose main actors were the federal state, the *Länder*, municipalities, energy utilities, and large industrial consumers (Kleinwächter, 2007).

The dominant fuel during the reconstruction period was coal, which provided over 90% of primary energy in 1950. However, in accordance with Germany's post-war ideology of market liberalism, market forces were to drive the choice of energy fuels and the construction of new power plants. Forecasts and the first energy system models emerged in this context as planning tools for energy companies that had to make decisions about how to meet steeply rising energy demand, and convince public and private investors to fund the construction of ever-larger coal, gas, and later nuclear power plants (Kraus, 1988; Herbst et al., 2012: 112).

Rather loosely structured and lacking a central anchoring point like the CGP in France, this dominant 'predictive policy assemblage' was challenged in the 1950s and 60s by quickly rising consumption of imported oil and gas, and the opening of the German market to imported coal in 1956. These developments heavily impacted the domestic coal industry, which entered a phase of decline, and provoked a rise in energy dependence from 8% in 1960 to 60% in 1977 (Meyer-Abich and Dickler, 1982). The crisis in the coal industry spurred heated debate within government: while social conservatives led by Chancellor Adenauer defended government support for the mining industry, market liberalists around the Minister of the Economy Ludwig Erhard refused government intervention.

In this context, Adenauer encouraged the creation of a parliamentary commission on energy policy (*Energie-Enquete*) in 1959. The commission was to evaluate the future prospects of German coal against the backdrop of the evolution of global energy markets and domestic energy demand. Its final report presented an analysis based on a 10-year forecast established by a consortium of major German economic institutes.[9] The study championed a new modelling methodology that was considered highly innovative at the time (Wessels, 1962): while earlier energy demand forecasts had represented the national economy as an aggregate whole, the new technique disaggregated the economy into three major sectors – industry, transportation, and households – and went into further detail in the industry and transportation sectors (three subsectors each). Designed to provide finer-grained descriptions of substitution processes between different energy technologies and sources, the method was thought to allow for more robust estimations of the future energy mix and its implications for the coal sector. The report also initiated an extensive data collection program supported by the federal government and major energy utilities, which made it possible to represent the German economy in unprecedented detail. Finally, it included a discussion of plausible alternative evolutions to the main, 'business-as-usual' forecast. While the sectorial approach highlighted ongoing substitution processes between German coal and imported fuels, the discussion of alternative evolutions gave these substitution processes a political dimension: instead of 'natural' evolutions in a market-driven economy, they now appeared as the result of a voluntary choice between political intervention and non-action.

In other words, through its method and the way it presented its results, the report backed calls for a genuine federal energy policy. On the basis of its conclusions, successive federal governments forged a "coal-priority-policy" (*Kohlevorrang-politik*), an unprecedented and massive infringement of the dominant free-market ideology (Krisp, 2007: 26, 27). However, the report's impact was not merely due to its methodological sophistication or empirical detail. The workings of the *Energie-Enquete* also contributed to further structuring and stabilising the dominant predictive policy assemblage, which in turn ensured the reception and uptake of the report by relevant actors: in the preceding years, economic institutes with close ties to industry and government[10] had begun to establish energy forecasts based on econometric models that became increasingly complex over time, and could therefore only be understood and challenged by a handful of actors (Seefried, 2010a). Designed with help and crucial input from main actors in the energy establishment, including energy producers, large industrial consumers and state bureacracies, they tended to reproduce the views of these actors (Kraus, 1988: 25). As in France, a characteristic feature of such models was their tight linear coupling of economic growth and growth in energy demand. Combined with the post-war ideology of economic development as a foundation of the West German social contract, this left no room for demand-oriented interventions in the energy system. However, following the methodology introduced by the *Energie-Enquete*, models progressively went from assembling the economy in a highly aggregated fashion to more detailed representation of some sectors that were subject to 'structural changes', and were therefore of particular political and economic interest (like the coal and steel industries). In line with the framing provided by energy models, energy policy hence emerged as 'energy supply policy': demand was considered outside the realm of politics, and policy-making limited to a choice between different fuel and technology options.

Accordingly, the first federal energy programme, launched on 3 October 1973, complemented the coal-priority-policy with a series of measures designed to kick-start an ambitious German nuclear program. Once again, this was justified on the basis of modelling results, which suggested that the macroeconomic costs of coal subsidies could be counterbalanced by the development of an alternative, supposedly cheap energy source (Bundestag, 1979: 14812). Rising oil prices at the end of 1973 accelerated the move from fragmented measures to a coherent and encompassing federal energy policy. Resumed by the formula "CoCoNuke" – for the triptych of conservation, coal and nuclear – the emerging paradigm for the first time included a focus on energy demand reduction, so as to diminish energy dependency (Düngen, 1993). Lacking significant political support, demand reduction measures were, however, not forcefully implemented at the time. This contrasted with the other two objectives: the proportion of primary energy consumption supplied by coal was stabilised at around 30% in the 1970s, and atomic energy's contribution to electricity production rose from 3.7% in 1970 to 40% in 1985 (Herzig, 1992: 153). The capital-intensity of the nuclear programme also accelerated concentration tendencies in the energy sector, where, already in 1974, two companies alone (RWE and Veba AG) controlled over 50% of the market (Nelkin and Pollak, 1981: 18).

Nonetheless, German energy policy did not form a monolithic whole. Beneath the dominant focus on energy supply, two policy assemblages struggled over the definition of energy policy. The first assemblage included abundant black and brown coal reserves in the Ruhr basin of North-Rhine-Westphalia, the Social Democratic Party that ruled the most populated federated State continuously from 1966 to 2005, the trade unions and the largest coal producer RWE. Together they enacted a policy that articulated social concerns for coal workers with a strategic focus on energy independence, understood as the capacity to fuel economic development using domestic resources. The second assemblage brought together the less densely populated areas in both northern and southern Germany whose rural geography allowed for the construction of atomic power plants far from urban centres, as well as banks and industrial conglomerates in southern economic centres and the two Christian democratic parties that governed Bavaria (CSU, since 1953) and Baden-Württemberg (CDU, 1957). This assemblage enacted an energy policy framed as industrial policy, and aimed at ensuring economic competitiveness through low energy prices. Energy independence was defined not in resource terms, but in technological terms, as the need to acquire nuclear know-how, so as to stay competitive in a globalised economy.

The general orientation of German energy policy in the 1970s and 80s thus involved a fragile compromise. Institutionalised in the "coal round tables" (*Kohlerunden*)[11] and the "atomic forum" (*Atomforum*),[12] the bipartition of energy policy-making hindered the emergence of a unified pro-nuclear front and favoured the emergence, in the 1980s, of a new political constellation.

Toward Change? The politicisation of Energy Futures (1975-1990)

Towards the end of the 1970s, the energy discourse changed yet again. The failure of dominant econometric models based on linear coupling of economic growth and energy demand to explain the impact of the oil-price shocks had cast macroeconomic forecasts into doubt and triggered the development of new modelling techniques (Château, 1985; Seefried, 2010b). Bottom-up (or engineering) models improved the representation of energy efficiency and alternative energy production techniques such as distributed renewables. These modelling approaches, which emerged first in the US (e.g. Ford Foundation, 1974) and France, and only some years later in Germany, suggested possible ways to decouple growth from energy demand through efficient resource-use. In the context of growing opposition to atomic energy, forecasts also became politically contested, and a multiplicity of contrasting energy futures came to populate public debate. Forecasts would no longer be created only by State administrations, research institutes, and energy companies, but also by experts close to the anti-nuclear movement (Kraus, 1988: 18). This also changed their status: from a public policy instrument, forecasts evolved into a weapon in energy controversies, used by civil society groups to repoliticise energy futures. In other words, the landscape of energy modelling and anticipatory exercises diversified, resulting in the emergence of competing 'predictive policy assemblages'.

'Useful Energy': establishing demand-side policies in France

In France as in other Western democracies, nuclear energy provoked widespread opposition. In contrast to other countries, however, the anti-nuclear movement struggled to institutionalise into a lasting political force (Nelkin and Pollak, 1981; Szarka, 2002). Two main explanatory factors are invoked to explain this specific French trajectory: a particularly powerful policy community around nuclear energy (Simmonot, 1978; Kitschelt, 1986), and the fragility of counter-expertise in a country where state engineers in ministries, the public research body CEA, and EDF enjoyed a near-monopoly in energy expertise (Restier-Melleray, 1990; Topçu, 2013). But the applicability of the second, at least, appears less straightforward than is frequently assumed. France has been at the forefront of the development of sophisticated modelling tools for energy demand, and the elaboration of alternative energy futures. Accordingly, what has to be explained is less the lack of alternative expertise than its failure to 'perform', i.e., by federating a new policy assemblage that would enact an alternative vision of the French energy future.

Throughout the 1970s, the anti-nuclear movement was supported by scientists engaged in fundamental research outside the nuclear establishment, as well as unionists from EDF and CEA (Topçu, 2006: 253). Discursively, it could draw on alternative forecasts produced by a new type of models, in which France soon became a front-runner. The most prominent example is the MEDEE model family developed at the *Institut Economique et Juridique de l'Energie* in Grenoble. First set out in a doctoral thesis co-authored by two engineers, Bertrand Château and Bruno Lapillonne, MEDEE pioneered a bottom-up approach to energy demand. The basic structure of the model (figure 1) couples a macroeconomic module with sectoral modules (households, industry, transportation, etc.) to determine "useful energy demand", which is distinguished from "final energy", i.e., the energy delivered to end-users in the form of electricity, natural gas or fuel. In replacing aggregate demand with a focus on the satisfaction of particular social needs like transportation, heating and production, MEDEE departed radically from existing approaches and helped establish demand as a politically influenceable variable (Interview 5).

Throughout the 1970s, MEDEE gradually evolved from a set of equations into a numerical model (Interview 12). This heavy "investment in form" (Thévenot, 1984) proved to be a crucial

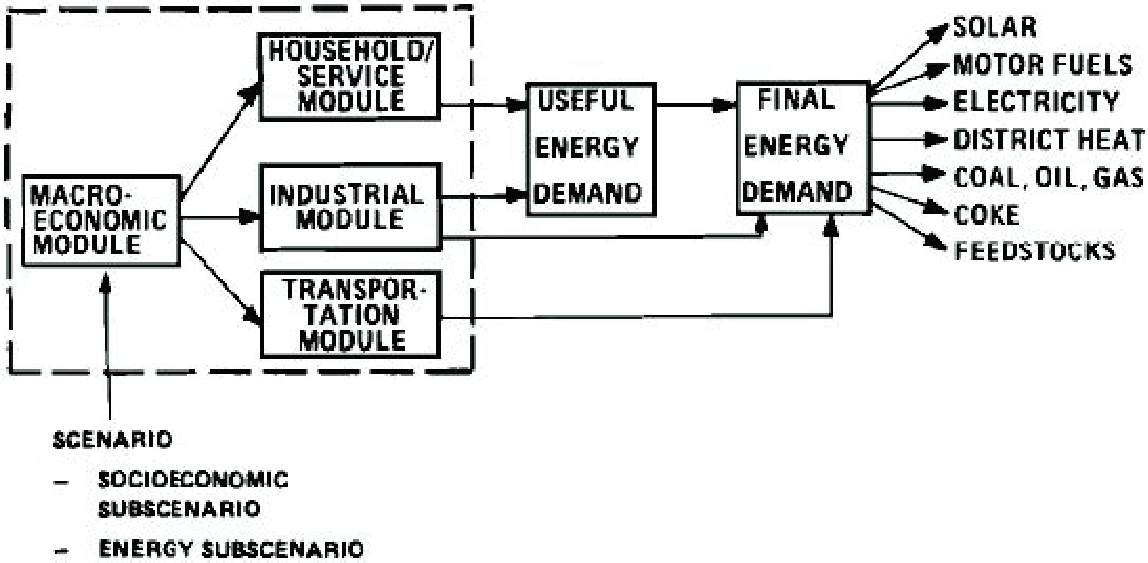

Figure 1. The MEDEE model (Lapillonne, 1978: 8).

and lasting asset: the model could now be adapted and used indifferently by a wide array of end-users, and its formalisation increased its legitimacy in public discourse. Although its bottom-up approach at first aroused resistance, both empirical observations and international evolutions in energy modelling soon seemed to validate its basic hypotheses. Both model-development and the extensive data collection programme it necessitated were supported by French and European research funding, and major industrial actors in the energy field contributed by providing data. MEDEE's role in the energy debate gradually evolved as a result: first used to provide alternative forecasts to official estimates, it progressively came to be included in official forecasts in the 1980s.

A second alternative approach, the ALTER project, emerged when a small group of researchers close to anti-nuclear circles – mathematician Philippe Courrège, agronomist Philippe Chartier, and economist and engineer Benjamin Dessus – carried out a normative scenario-building exercise to demonstrate that France could in principle satisfy all its energy needs from renewable sources (Collectif de Bellevue, 1976). The heart of the project consisted in a modelling effort that reassembled the French economy through its basic energy flows and provided a static physical representation[13] of a future energy system in which industrial and social activities were fuelled exclusively by solar energy and biomass. As recalled by Benjamin Dessus, this made it possible to reconceive the relationship between demand- and supply-side policies:

> So I was like a lot of people: do we do solar, wind turbines, or nuclear? But that wasn't enough of a response at all. This exercise taught us an idea, that you have to bring the whole system into play, on energy demand as well as supply, which was not at all… in the culture we came from. Engineers are used to making things. (Interview 1)

While the insistence on small production units and demand reduction measures facilitated the uptake of the scenario in local ALTER plans elaborated by civil society groups, its disruptive vision of the future encountered strong resistance in national policy circles. An especially controversial point was that the authors had based their estimations on precise calculations of future 'energy needs', defining the average size of apartments, heating temperature, electrical equipment, transportation kilometres, and so on. Many policy-makers considered precise determination of such variables to be beyond the state's mandate (Interview 1). Moreover, the absence of market mechanisms in the model that formed the basis of ALTER seemed to suggest an all-encompassing planning approach to energy policy. In retrospect, Philippe

Courrège argues that this exposed the scenario to critiques of its underlying political philosophy, as opponents conjured the Orwellian vision of a centralised, almighty State controlling the national economy and intruding even into the personal life of its citizens (Interview 3).

The trajectories of the two projects converged in the early 1980s, as actors from both ALTER and MEDEE came to occupy influential positions in the newly founded Agency for the reduction of energy demand (AFME).[14] Its creation was a sign of growing political interest in energy efficiency, fuelled by rising energy prices, increasing overcapacity in electricity production (Puiseux, 1987: 185), and the brief political opening created by the arrival into power of Mitterrand's socialists. The agency, which was soon at the centre of a major political battle over energy demand, became the home of critical researcher-activists: Bernard Laponche was appointed director general, while Dessus became director of technical services, Chartier scientific director and Château director of economic studies. Château brought with him the MEDEE model, which enabled the agency to challenge official forecasts not only from a normative perspective, but also on technical, quantified grounds. As a consequence, the MEDEE model became the agency's official modelling tool and contributed to its international strategy,[15] while the agency gradually came to constitute the central node of an emerging, alternative 'predictive policy assemblage'.

The model informed the French planning bureau's first long-term energy forecast in 1983, carried out in preparation of the Ninth Plan. Designed to test the viability of the nuclear program, the exercise was the first official forecast to project decreasing energy demand as a result of the economic crisis and changing policy orientations. The final report estimated that the number of planned reactors in 1990 was too high by 25 to 30%, and suggested that no new reactors would be needed before the end of the decade (CGP, 1983: 21, 51-55). These conclusions were a shock to the nuclear establishment and spurred heated debate. Yet, once again, considerations of industrial policy prevailed.

EDF reacted to the problem of overcapacity – which now took the form of an impending industrial catastrophe rather than a distant and abstract economic risk – with a twofold strategy: an ambitious program to provide electricity to neighbouring countries (especially Switzerland), and an intensification of household electrification. This was fundamentally at odds with a reorientation of energy policy towards demand reduction. The electricity monopoly's position was strengthened in the mid-1980s, when falling oil prices not only decreased political interest in energy efficiency, but also seemed to contradict the gloomy predictions of peak oil and rising energy costs that had come to populate public debate in the 1980s. In 1986, the election of a conservative government put an end to the controversy: the budget of the agency was cut by almost 80% and its personnel diminished by one third (Evrard, 2013). EDF was allowed free reign to define French energy policy. Alongside this gradual disengagement of the state, the central locus of future-making, CGP, progressively lost its importance in the 1980s. Until well into the 2000s, no other institution emerged which could perform somewhat authoritative collective forecasting or scenario-building exercises.[16]

'Energiewende': reassembling German energy policy

In Germany as well, alternative models and scenarios emerged in a context of growing anti-nuclear protest, which reached a peak at the end of the 1970s with massive demonstrations against a fast-breeder reactor under construction in Kalkar (North Rhine-Westphalia) and a planned atomic waste storage facility in Gorleben (Lower Saxony). Contrary to France, different safety standards in the different *Länder* facilitated legal challenges, and courts progressively evolved into a public forum for anti-nuclear experts and a range of grassroots, popular education, or research institutions like *VHS Wyhler Wald* (created in 1975) and the Öko-Institut[17] (1977), which provided expertise on nuclear risks and informed about alternative, renewable energy sources.

Öko-Institut proved to be particularly influential. The 1980 "energy turnaround" report (Krause et al., 1980) by three of its experts – Florentin Krause, a chemist, Hartmut Bossel, an engineer and philosopher, and Karl-Friedrich Müller-Reissmann, a theologian and computer scientist – outlined a far-reaching transformation pathway that durably influenced the German energy debate. As in the French case, it rested on a bottom-up approach with detailed representation of potential energy savings and possible contributions of decentralized renewables. The subtitle of the study, "growth and prosperity without oil and uranium", indicated that the proposed energy transition would neither entail material sacrifices nor imply a departure from Germany's post-war ideology combining market liberalism and a social contract based on economic growth.

Alongside this concession to the dominant discourse, the authors operated a series of *strategic displacements* in the report, by redefining basic notions and concepts from mainstream forecasts. In line with recent modelling trends in other countries, they proposed to disaggregate energy demand in 'energy services' – heat, light, kinetic force, transportation kilometres – and criticised the domination of energy debates by neoclassical economic theory. The study then concentrated its attacks on three elements of mainstream forecasts: the tight coupling between economic growth and energy demand; the concentration of energy policy on the production side; and the reliance on oil and nuclear as basic pillars of the energy system.

The scenario-technique played a central role in establishing these arguments. Unlike the ALTER project, the report was based on a pragmatic and dynamic (as opposed to static) approach that used official economic forecasts and excluded deep changes in the economy (like a departure from industrialism) or energy consumption patterns (such as lifestyle changes). Accordingly, the authors qualified their method as a "technical fix" approach, aimed at satisfying projected energy needs even of "overtly growth-euphoric forecasts" (Krause et al., 1980: 10). On the basis of a detailed analysis of the evolution of energy needs and services, the report proposed three scenarios: a "business-as-usual" pathway, assuming unchanged production and consumption patterns, which the authors labelled the "suicide scenario" and dismissed as "unrealistic"; a "coal and gas" scenario that attempted to convince moderate critics of atomic energy that it was possible to phase out nuclear energy by implementing ambitious policies to favour energy efficiency, coal and gas; and a "sun and coal" or *Energiewende* ('energy turnaround') scenario, which the authors clearly preferred, and which presented the advantage of relying exclusively on domestic resources. This, they contended, would not only minimize risks, but also make the German economy virtually self-sufficient in terms of energy supply.

Most importantly, the latter scenario was carefully designed to construct alliances with major actors in West German energy politics. Not only did the 'coupling' of coal and renewables contribute to building bridges between the ecological movement, the trade unions, and parts of the coal industry; its technology-oriented bottom-up approach also provided a discursive underpinning for advocates of an "ecological modernization" of the German economy (Mol and Jänicke, 2009). The calculations laid out in the report were thus in line with a larger reconceptualization of environmental policies as *industrial* policies. Its pragmatic outline hence proved instrumental in ensuring the report would find an audience well beyond environmental activist circles.

The long-term performative effects of the *Energiewende* report, however, cannot be understood by analysing its content alone. Equally important are concomitant political changes, through which parliamentary Enquete Commissions gradually emerged as a central forum for official forecasting practices, especially in the energy field. Such commissions had been institutionalised and endowed with considerable autonomy and resources after a parliamentary reform in 1969, aimed at strengthening parliament's independence from ministerial expertise (Knelangen, 2000). In line with this objective, these commissions are composed of equal numbers of parliamentarians and experts. A corollary of this practice is that experts, who are full, voting members of the commission, are chosen not only to provide specialist knowledge, but also to represent a particular social force or political position. This way of organising the production of policy-relevant knowledge through a dialogic process that associates relevant viewpoints has been described as a specifically German "civic epistemology" which holds the potential (but not the guarantee) of opening such processes up for contesting voices (Jasanoff, 2005; Beck, 2004).

In line with these developments, the wider uptake of both Öko-Institut's future vision and its modelling approach are intimately linked to the workings of a parliamentary Enquete Commission established in 1979 on "future nuclear energy policy" (Altenburg, 2010; Aykut, 2015). The political context was explosive. Public opposition to atomic energy, and in particular to the fast-breeder in Kalkar, had been growing, fuelled by external events like the nuclear accident in Harrisburg (USA) in 1979, and events in neighbouring Austria, where atomic energy had just been rejected in a national referendum. As a consequence, all parties, but especially the Social Democrats, had to cope with internal division over the nuclear issue. Accordingly, key criteria for the selection of experts in the Enquete Commission were their ties either to atomic research or the environmental movement, and more generally their stance towards nuclear energy. "The commission was composed politically", as Klaus-Michael Meyer-Abich, a 'natural philosopher' and moderate critic of atomic energy who participated in the commission, recalls (Interview 14); its members further included Günter Altner, one of the founders of Öko-Institut, but also Wolf Häfele, the former head of fast-breeder development at the nuclear research centre KfK and one of Germany's most vocal nuclear advocates, and Klaus Knizia, CEO of VEW, a local electricity producer with interests in both coal and nuclear.

In this heated atmosphere, the stated aim of the commission was to channel open confrontation into a "rational" debate (PEK, 1980: 2). Interestingly, its members believed that such a debate could be furthered through a systematic clarification of different future visions. Based on a comprehensive research programme that included major German energy research institutes and a long series of hearings with energy experts, they elaborated four scenarios, two with and two without nuclear energy, and set out to analyse their implications in political, social and economic terms:

> To further mutual understanding, the commission has attempted to make the visions of the energy future that result from different convictions amenable to reasoned discussion. It therefore agreed to represent these in four internally coherent energy policy paths. This required the willingness of all to outline the limitations and consequences of the respective energy paths. The commission thereby sought to create the conditions to sound out the prospects for a broad consensus on energy policy in a manageable time frame. (PEK, 1980: 23)

The four energy paths were designed to represent important standpoints in the German energy debate. The first reflected the vision of the nuclear industry and mainstream energy economists. It projected a doubling of energy demand by 2030, almost all of which was to be satisfied through atomic energy. The second path expressed a view shared by industrialists and parts of the governing coalition, and combined moderate demand reduction with diversification of (conventional) energy technologies. It projected 50% demand growth, to be met by increasing nuclear and coal. In the third path, demand was stabilized and nuclear energy progressively phased out. This roughly corresponded to the position of nuclear critics within the political establishment. Finally, the fourth path involved a rapid nuclear phase-out, associated to heavy energy savings and deployment of renewables. This adapted version of the *Energiewende* scenario mirrored the views of anti-nuclear activists.

The fifteen commission members also established four common criteria – "economic viability", "international compatibility", "environmental compatibility", and "social compatibility" – to evaluate the scenarios and create common ground for policy recommendations. Although a consensual assessment proved difficult (Interview 14), the commission succeeded in forging a common position on measures for the medium term. Based on Lovins' (1976) famous distinction, it advocated a temporary "parallel approach" aimed at giving both the "hard path" (combining fossil and fissile technologies) and the "soft path" (efficiency and renewables) an equal chance. This was justified on the grounds that the evolution of key variables, such as structural changes in the economy, public acceptance of nuclear energy, effects of energy-saving policies and the feasibility of fast-breeder technology, was too uncertain to be forecasted properly. The commission therefore suggested pursuing both the construction of the fast-breeder *and* stringent energy savings until the end of the decade. By postponing the choice between the two paths, it delegated the final decision on the energy future to the political system. For anti-nuclear activists, an important result was that the commission considered a nuclear phase-out to be a viable option at all:

> The whole thing was decided unanimously. And at the beginning, people always said: it's not possible without […] This commission was the first one where everyone decided collectively: yeah, it's possible without. It's possible with, but it's also possible without. Politically, this was already quite a success at that time (Interview 14).

The commission participated in a redefinition of the front lines in the energy controversy. Widely discussed and publicised, the four energy paths made it clear that an energy transition was not only in the interest of radical ecologists, but could benefit wider parts of the industrial and political establishment (Interviews 14, 28). Not only did the struggling coal industry, trade unions, and their social democratic allies find – at least temporarily – that they had common strategic interests with anti-nuclear activists; other industrial branches and the emerging 'green sector', as well as local actors and municipalities[18] progressively discovered that they might well profit from an alternative path that, by not relying on capital-intensive energy supply technologies, could allow them to develop and commercialize energy saving and efficiency technologies (Weidner and Mez, 2008). This was backed by environmental economists who began to collect evidence that such a transformation could be accomplished within a reformed social market economy (Binswanger et al., 1981). The resulting redefinition of roles and interests contributed to structuring and consolidating an alternative 'predictive policy assemblage', which was equiped with its own modelling tools and a future vision that broadly corresponded to the *Energiewende* scenario. Parts of this vision were enacted almost 20 years later by a coalition government of Social Democrats and Greens (1998-2005).

Conclusion

Models and forecasts occupy a central position in energy debates. They propose the future-visions that populate public discourse, provide market actors and policy-makers with ontologies to understand energy systems, and shape wider policy networks in scenario-building exercises and through the circulation of models across social spaces. In doing so, they can stabilise dominant framings, practices, and policy assemblages, or rearrange and reorder policy worlds, thereby contributing to the formation of new assemblages that enact alternative conceptions of energy policy. Energy controversies therefore unfold not only as political or ideological struggles about the problems of energy production and suitable ways of dealing with such problems; I have argued here that they can be understood as struggles between competing 'predictive policy assemblages', in which new actors, their problem-framings and predictive practices challenge both how established models compose energy systems and how major anticipatory exercises include relevant actors in the production of energy futures.

Unsurprisingly, then, there were close parallels between model-development and evolutions in policy-making in Germany and France. Social and political events in different periods triggered innovations in modelling, which required the production of new data. This in in turn contributed to transforming problem definitions, induced or accompanied changes in energy policy, and helped to sustain novel institutions and organisations. The way energy forecasts relate to energy policy, however, has differed in the two countries, and this relationship has changed over time.

Anticipatory exercises in France were traditionally carried out by the national planning bureau CGP, and their status was not only epistemic, but also explicitly political. Although CGP associated major actors in the energy field, it was characterised by an inherent asymmetry: estimations of electricity demand were almost exclusively calculated by EDF's models, and matched the company's industrial strategy. When model predictions and real-world developments diverged in the 1970s, the dominant 'predictive policy assemblage' enacted a policy of household electrification that re-calibrated the electricity-intensity of economic development to a level compatible with the country's ambitious nuclear program.

In Germany, national energy forecasts emerged in a context of crisis in the coal sector, which challenged the role of the federal state in energy policy. The 1959 *Energie-Enquete* commission was created in response. It introduced innovations in modelling techniques that made visible substitution processes in economic sectors, and called for government to take a more proactive role. Progressively institutionalised in the preparation and evaluation of federal policies, the modelling approach envisioned energy policy as a choice between different energy carriers, and furthered the emergence of new dominant framings such as the 'coal priority' and 'CoCoNuke' policies.

In both countries, established predictive practices and dominant policy paradigms were challenged by new actors and modelling techniques in the 1970s. A situation where forecasts were more or less directly embedded in policy-making and models established by experts close to the energy policy establishment gave way to a new configuration, characterised by a multiplication of model-types and a politicisation of forecasts, which were produced and taken up by a wide range of actors in an increasingly controversial debate. But while alternative scenarios succeeded in reassembling German energy policy along lines that proved to be conductive to policy change, this did not occur in France, where demand-side modelling was institutionalised in the energy savings agency *Ademe*, but failed to enrol potential agents of a new political constellation.

Common attempts to explain this rigidity of French energy policy point to the homogeneity of the dominant actor-coalition and the heavy investments made by EDF (e.g. Puiseux, 1987: 195). A focus on predictive practices adds two important elements to the puzzle: first, while bottom-up models did provide a powerful tool to counter dominant discourse, and formed the quantitative backbone for demand-reduction policies in public discourse and inter-ministerial negotiations, they did not deliver a coherent future vision in which major actors of French energy and industrial policy could recognize or project themselves. As for the ALTER scenario, which could have provided such a vision, its architecture was too uncompromising to offer such actors – progressive industrialists, entrepreneurs or municipalities – a plausible and desirable future in which their expertise and activities would be valued.

In Germany, by contrast, the *Energiewende* scenario not only 'equipped' the ecological movement with new arguments in its battle for a non-nuclear future; it also 'reassembled' energy policy in a way that opened up energy debates. Öko-Institut's vision functioned as a "prospective structure to be filled in by agency" (Van Lente and Rip, 1998), proposing both a new narrative and a new arrangement of energy policy that could subsequently be enacted. This was accompanied by a formalisation of bottom-up models at Öko-Institut and other modelling centres. Increasingly used in policy-making and administrative practice, these models contributed to durably anchoring efficiency and renewables policies in policy circles (Interviews 15, 22).

Finally, a symmetrical, yet opposite evolution in the 1980s increased the discrepancies between the energy trajectories of the two countries. The established French locus of future-making, the CGP, progressively lost its central position, making it more difficult for alternative modelling approaches and future-visions to enter policy circles and gain public acceptance. In Germany, in contrast, the parliamentary Enquete Commissions, which provided a forum for contesting actors and alternative futures, became a central node in energy debates in the following decade. The scenario technique introduced by the 1979 commission also durably changed official anticipatory exercises: used in subsequent commissions and committees on climate and energy policy, it was instrumental in organising energy discourse around a set of distinct, mutually exclusive future-visions, which not only reflected divergent policy preferences, but also corresponded to different ways of 'assembling' energy system and envisioning political interventions in such systems (Interviews 27, 28).

The long-term 'success' of the *Energiewende* report can therefore only be fully appreciated by jointly analysing how it 'performed' along three dimensions, through: an alternative future-vision that contributed to re-structuring German energy debates; a bottom-up modelling approach that circulated in administrative and civil society practice and enabled new forms of political intervention; and the formation of a new 'predictive policy assemblage' capable of enacting an alternative energy future.

References

Altenburg C (2010) *Kernenergie und Politikberatung. Die Vermessung einer Kontroverse*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Andersson J and Rindzevičiūtė E (2015) *The Struggle for the Long Term in Transnational Science and Politics. Forging the Future*. New York, NY: Routledge.

Angeletti T (2011) Faire la réalité ou s'y faire ? La modélisation et les déplacements de la politique économique au tournant des années 1970. *Politix* 2011/3(95)**:** 47-72.

Austin JL (1962) *How to Do Things with Words*. Oxford: Clarendon Press.

Aykut SC (2015) Energy futures from the social market economy to the Energiewende: The politicization of West German energy debates, 1950-1990. In: Andersson J and Rindzevičiūtė E (eds) *The Struggle for the Long Term in Transnational Science and Politics: Forging the Future*. London: Routledge, pp. 63-91.

Barnes B (1983) Social life as bootstrapped induction. *Sociology* 17(4)**:** 524-545.

Baumgartner T and Midttun A (1987) *The politics of energy forecasting. A comparative Study of Energy Forecasting in Western Europe and North America*. Oxford: Clarendon Press.

Beck S (2004) Localizing Global Change in Germany. In: Jasanoff S and Long-Martello M (eds) *Earthly Politics. Local and Global in Environmental Governance*. Cambridge, MA and London: The MIT Press, pp. 173-194.

Beckert J (2013) Imagined Futures. Fictional Expectations in the Economy. *Theory and Society* 42(3)**:** 219-240.

Beyer J (2002) Deutschland AG a.D.: Deutsche Bank, Allianz und das Verflechtungszentrum großer deutscher Unternehmen. *MPIfG Working Paper* 02/4.

Binswanger HC, Bonus H and Timmermann M (1981) *Wirtschaft und Umwelt. Möglichkeiten einer ökologieverträglichen Wirtschaftspolitik*. Stuttgart: Kohlhammer.

Bowker GC (2000) Biodiversity datadiversity. *Social Studies of Science* 30(5)**:** 643-683.

Brisset N (2016) Economics is not always performative: some limits for performativity. *Journal of Economic Methodology* 23(2)**:** 160-184.

Bundestag (1979) Plenarprotokoll 8/188. Bonn, 28. November 1979: Deutscher Bundestag.

Callon M (1998) *The Laws of the Markets*. Oxford, UK: Oxford University Press.

Callon M (2007) What Does It Mean to Say That Economics Is Performative? In: Mackenzie D, Muniesa F and Siu L (eds) *Do Economists Make Markets? On the Performativity of Economics*. Princeton: Princeton University Press, pp. 311-357.

Cgp (1983) Énergie. Rapport du Groupe Long-Terme du Commissariat général au Plan. Paris: La Documentation Française.

Château B (1985) La prévision énergétique en mutation? . *Revue de l'énergie* 36(janvier)**:** 1-11.

Chateauraynaud F (2013) Des prises sur le futur. Regard analytique sur l'activité visionnaire. In: Bourg D, Joly P-B and Kaufmann A (eds) *Du risque à la menace. Penser la catastrophe*. Paris: PUF, pp. 287-309.

Cochoy F, Giraudeau M and Mcfall L (2010) Performativity, economics and politics: An overview. *Journal of Cultural Economy* 3(2)**:** 139-146.

Collectif De Bellevue (1976) *Projet Alter. Etude d'un avenir énergétique pour la France axé sur le potentiel renouvelable. Esquisse d'un régime à long terme tout solaire*. Paris: Syros.

Delanda M (2006) *A new philosophy of society: Assemblage theory and social complexity*. Bloomsbury Publishing.

Desrosières A (1999) La commission et l'équation: une comparaison des Plans français et néerlandais entre 1945 et 1980. *Genèses* 34**:** 28-52.

Düngen H (1993) Zwei Dekaden deutscher Energie- und Umweltpolitik: Leitbilder, Prinzipien und Konzepte. In: Hohensee J and Salewski M (eds) *Energie-Politik-Geschichte : nationale und internationale Energiepolitik seit 1945*. Stuttgart: Franz Steiner Verlag, pp. 35-50.

Edwards PN (2010) *A vast Machine. Computer Models, Climate Data, and the Politics of Global Warming*. Cambridge, MA: The MIT Press.

Evrard A (2013) *Contre vents et marées. Politiques des énergies renouvelables en Europe*. Paris: Presses de Sciences Po.

Ford Foundation (1974) *A Time to Choose: America's Energy Future*. Pensacola, FL: Ballinger Publishing Co.

Guston DH (2014) Understanding 'anticipatory governance'. *Social Studies of Science* 44(2)**:** 218-242.

Hecht G (1998) *The radiance of France: Nuclear Power and National Identity after World War II*. Cambridge, MA: MIT Press.

Henriksen LF (2013) Economic models as devices of policy change: Policy paradigms, paradigm shift, and performativity. *Regulation & Governance* 7(4)**:** 481-495.

Herbst A, Toro F, Reitze F and Jochem E (2012) Introduction to Energy System Modelling. *Swiss Journal of Economics and Statistics* 148(2)**:** 111-135.

Herzig T (1992) Wirtschaftsgeschichtliche Aspekte der deutschen Elektrizitätsversorgung 1880 bis 1990. In: Fischer W (ed) *Die Geschichte der Stromversorgung*. Frankfurt am Main: Verlags- und Wirtschaftsgesellschaft der Elektrizitätswerke m.b.H, pp. 122-166.

Hoffman K and Wood DO (1976) Energy system modeling and forecasting. *Annual Review of Energy* 1**:** 423-453.

Jasanoff S (2005) *Designs on Nature. Science and Democracy in Europe and the United States*. Princeton: Princeton University Press.

Kitschelt HP (1986) Political Opportunity Structures and Political Protest: Anti-Nuclear Movements in Four Democracies. *British Journal of Political Science* 16(01)**:** 57-85.

Kleinwächter K (2007) Das „Eiserne Pentagramm" - Strommarktregulierung in Deutschland. In: Kleinwächter L (ed) *Deutsche Energiepolitik*. Potsdam: Brandenburgische Landeszentrale für politische Bildung, pp. 65-79.

Knelangen W (2000) Enquete-Kommissionen. In: Andersen U and Woyke W (eds) *Handwörterbuch des politischen Systems der Bundesrepublik Deutschland. (4te Auflage)*. Opladen: Leske+Budrich, pp. 157-159.

Kraus M (1988) Energieprognosen in der Retrospektive. Analyse von Fehlerursachen der Prognose/Ist-Abweichungen von Energiebedarfsschätzungen in der Bundesrepublik Deutschland von 1950 bis 1980. *Dissertation*. Karlsruhe: Universität Karlsruhe.

Krause F, Bossel H and Müller-Reissmann K-F (1980) *Energie-Wende: Wachstum und Wohlstand ohne Erdöl und Uran. Ein Alternativ-Bericht des Öko-Instituts*. Frankfurt am Main: Fischer.

Krisp A (2007) *Der deutsche Strommarkt in Europa - zwischen Wettbewerb und Klimaschutz*. Dissertation: Fachbereich Sozial- und Kulturwissenschaften der Justus-Liebig-Universität Gießen.

Lapillonne B (1978) *MEDEE 2: A Model for Long-Term Energy Demand Evaluation. IIASA Research Report*. Laxenburg, Austria: IIASA.

Lindberg LN (ed) (1977) *The energy syndrome: Comparing national responses to the energy crisis*. Lexington: Lexington Books.

Lovins AB (1976) Energy Strategy: The Road Not Taken. *Foreign Affairs* 55(Fall)**:** 65-96.

Lucas N (1985) *Western European energy policies : a comparative study of the influence of institutional structure on technical change*. Oxford: Clarendon Press.

Mackenzie D (2006) *An Engine, Not a Camera. How Financial Models Shape Markets.* Cambridge, MA: MIT Press.

Mackenzie D (2008) *Material markets: How economic agents are constructed.* Oxford, UK: Oxford University Press.

Mackenzie D, Muniesa F and Siu L (2007) *Do Economists Make Markets? On the Performativity of Economics.* Princeton: Princeton University Press.

Mäki U (2013) Performativity: Saving Austin from MacKenzie. In: Karakostas V and Dieks D (eds) *EPSA11 Perspectives and Foundational Problems in Philosophy of Science.* Heidelberg: Springer, pp. 443-453.

Massé P (1965) *Le Plan ou l'anti-hasard.* Paris.

Mccann E and Ward K (2012) Policy assemblages, mobilities and mutations: toward a multidisciplinary conversation. *Political studies review* 10(3)**:** 325-332.

Meadowcroft J (2009) What about the politics? Sustainable development, transition management, and long term energy transitions. *Policy Sciences* 42(4)**:** 323-340.

Merton RK (1948) The Self-fulfilling Prophecy. *The Antioch Review* 8(2)**:** 198-210.

Meyer-Abich KM and Dickler RA (1982) Energy Issues and Policies in the Federal Republic of Germany. *Annual Review of Energy* 7**:** 221-259.

Miller D (2002) Turning Callon the right way up. *Economy and society* 31(2)**:** 218-233.

Mol APJ and Jänicke M (2009) The origins and theoretical foundations of ecological modernisation theory. In: Mol APJ, Sonnenfeld DA and Spaargaren G (eds) *The Ecological Modernisation Reader: Environmental Reform in Theory and Practice.* London and New York: Routledge, pp. 17-27.

Nelkin D and Pollak M (1981) *The Atom Besieged. Antinuclear protest in France and Germany.* Cambridge, MA: MIT Press.

Nelson N, Geltzer A and Hilgartner S (2008) Introduction: the anticipatory state: making policy-relevant knowledge about the future. *Science and Public Policy* 35(8)**:** 546-550.

Pek (1980) Bericht der Enquete-Kommission 'Zukünftige Kernenergie-Politik'. *Drucksache 8/2628.* Bonn: Deutscher Bundestag.

Pickering A (1995) *The Mangle of Practice: Time, Agency, and Science.* Chicago: University of Chicago Press.

Polak LF (1973) *The Image of the Future.* Amsterdam, London, New York: Elsevier Scientific Publishing Company.

Porter TM (1995) *Trust in numbers: The pursuit of objectivity in science and public life.* Princeton: Princeton University Press.

Puiseux L (1982) Les bifurcations de la politique énergétique française depuis la guerre. *Annales : Economies, Sociétés et Civilisations* 37(4)**:** 609-620.

Puiseux L (1987) The Ups and Downs of Electricity Forecasting in France: Technocratic Elitism. In: Baumgartner T and Midttun A (eds) *The politics of energy forecasting. A comparative Study of Energy Forecasting in Western Europe and North America.* Oxford: Clarendon Press, pp. 180-210.

Radanne P (2006) *Energies de ton siècle. Des crises à la mutation.* Paris: Lignes de repères.

Restier-Melleray C (1990) Experts et expertise scientifique. Le cas de la France. *Revue française de science politique* 4**:** 546-585.

Romeiro F (1994) *Energie, économie, environnement. Le cas du secteur de l'électricité en Europe entre passé, présent et futur.* Geneva, Paris: Librairie Droz.

Schubert C (2015) Situating technological and societal futures. Pragmatist engagements with computer simulations and social dynamics. *Technology in Society* 40**:** 4-13.

Seefried E (2010a) Experten für die Planung? „Zukunftsforscher" als Berater der Bundesregierung 1966-1972/73. *Archiv für Sozialgeschichte* 50**:** 109-152.

Seefried E (2010b) Prognostik zwischen Boom und Krise. Die Prognos AG und ihre Zukunftsprognosen für die Entwicklung der Bundesrepublik in den sechziger und siebziger Jahren. In: Hartmann H and Vogel J (eds) *Zukunftswissen. Prognosen in Wirtschaft, Politik und Gesellschaft seit 1900.* Frankfurt/New York Campus, pp. 76-106.

Silvast A (2017) Energy, economics, and performativity: Reviewing theoretical advances in social studies of markets and energy. *Energy Research & Social Science* 34**:** 4-12.

Simmonot P (1978) *Les nucléocrates.* Grenoble: Presses Universitaires de Grenoble.

Stier B (1999) *Staat und Strom - Die politische Steuerung des Elektrizitätssystems in Deutschland 1890 - 1950.* Mannheim: Verlag Regionalkultur.

Szarka J (2002) *The shaping of environmental policy in France.* Oxford: Berghahn Books.

Thévenot L (1984) Rules and implement: investment in forms. *Social Science Information* 23(1)**:** 1-45.

Thompson M (1984) Among the Energy Tribes: A Cultural Framework for the Analysis and Design of Energy Policy. *Policy Sciences* 17(3)**:** 321-339.

Topçu S (2006) Nucléaire: de l'engagement "savant" aux contre-expertises associatives. *Natures, Sciences, Sociétés* 14**:** 249-256.

Topçu S (2013) *La France nucléaire. L'art de gouverner une technologie contestée.* Paris: Seuil.

Unruh GC (2000) Understanding carbon lock-in. *Energy Policy* 28(12)**:** 817-830.

Van Lente H and Rip A (1998) Expectations in Technological Developments: An Example of Prospective Structures to be Filled in by Agency. In: Disco C and Van Der Meulen B (eds) *Getting New Technologies Together: Studies in Making Sociotechnical Order.* Berlin: Walter de Gruyter, pp. 202-230.

Varoufakis Y (2017) *Adults in the Room. My Battle with Europe's Deep Establishment*. London: The Bodley Head.

Weidner H and Mez L (2008) German Climate Change Policy: A Success Story With Some Flaws. *The Journal of Environment & Development* 17(4)**:** 356-378.

Wessels T (1962) Die Energie-Enquete : Ergebnisse und wirtschaftspolitische Konsequenzen. *Tagungsberichte des Energiewirtschaftlichen Instituts an der Universität Köln, Heft 12.* München: Verlag R. Oldenbourg.

Wieviorka M and Trinh S (1991) *Le modèle EDF: essai de sociologie des organisations.* Paris: La Découverte.

Wynne B (1984) The institutional context of science, models, and policy: the IIASA energy study. *Policy Sciences* 17(3)**:** 277-320.

Yon G (2014) L'économicité d'EDF. La politique tarifaire d'Électricité de France et la reconstruction de l'économie nationale, de la nationalisation au milieu des années 1960. *Politix* 2014/1(105)**:** 91-115.

Notes

1 This research was supported by a grant of the French Agence Nationale de la Recherche to the project "Innovation in Expertise. Modeling and simulation as tools of governance" (ANR-13-SOIN-0005), coordinated by David Demortain. Comments from participants to two project workshops, especially David Demortain, Bilel Benbouzid, Pierre-Benoît Joly and Bernard Laponche, as well as from Amy Dahan, Michel Armatte, Antti Silvast and two anonymous reviewers greatly improved the paper. I also thank Paul Reeve for his excellent language editing and insightful comments.

2 Interviews were conducted between 2010 and 2017. Documents were collected from parliamentarian, ministerial and personal archives. All quotes have been translated to English by the author.

3 Laponche, who had critically examined energy forecasting techniques in his PhD thesis, later became a leading figure among anti-nuclear activists.

4 A senior civil servant and industrialist, Jean Syrota directed Cogema (later Areva) (1988-1999) and chaired the *Corps des Mines* (1993-1997). He distinguished himself by opposing EDF's strategy, advocating for energy savings policies instead.

5 Modelling techniques used at that time all entailed the implicit assumption of a stable relationship between energy demand and economic growth (Puiseux, 1987: 188,189).

6 Marcel Boiteux is a mathematician and graduate of the elite institutions *Ecole Normale Supérieure* and *Institut d'études politiques de Paris* (Sciences Po).

7 In electrical engineering, load profiles are graphs that represent variation in electrical load over time.

8 Yon (2014) shows how French state engineers invented the marginal cost curve to "calibrate France".

9 The consortium was led by the University of Cologne's *Energiewirtschaftliche Institut* (founded in 1943).

10 EWI is financed by a consortium including the energy utility RWE and the federated State of North-Rhine Westphalia. Two other economic institutes with important energy divisions, Ifo (Munich, 1949) and RWI (Essen, 1943), have close ties to industry. Other actors in the field included the State-financed DIW (Berlin, 1925), *Prognos AG*, a Suisse institute (1959), as well as technical universities and atomic research institutes (e.g., *TU Karlsruhe*, *Kernforschungszentrum Jülich*).

11 Initiated in 1983, these negotiation cycles associate firms, trade unions, the *Länder*, and the federal State.

12 Founded in 1959, the lobbying association is composed of major industrial actors and research institutes.

13 Such 'physical economics' were opposed by the authors to the dominant econometric models (Interview 3).

14 The *Agence française pour la maîtrise de l'énergie* (transformed to Ademe in 1991) resulted from the merger of two pre-existing public bodies in 1982.

15 To encourage energy demand policies in the global South, the model – together with its architect Château – were 'exported' to developing countries (Interviews 5, 12, 9).

16 While CGP continued to produce energy forecasts until the late 1990s, its last five-year plan was adopted in 1989. The institution was abolished in 2006.

17 Founded by activists in Wyhl (*Baden-Württemberg*), the institute is financed by a philanthropic association.

18 Local *Energiewende* committees took up the transition scenario and used it for community level activism, thereby giving it a wider audience.

Appendix 1. List of Interviews

French energy experts and modellers (interviews 1-13)

Benjamin Dessus (energy expert and activist, formerly director of research at Ademe; *interview conducted on* 1.6.2011), Pierre Radanne (energy expert, formerly president of Ademe; 7.12.2008), Philippe Courrège (energy modeller and activist; 6.5.2011), Pierre Matarasso (energy expert and activist; 6.5.2011), Patrick Criqui (energy modeller, EDDEN; *first interview*: 18.03.2015), Patrick Criqui (*second interview*, *conducted with Alain Nadaï*, 25.1.2017), Kimon Keramidas (energy modeller, JRC of the EU; 28.5.2015), Silvain Cail (energy modeller, head of global forecasting at Enerdata; 29.05.2015), Nadia Maïzi (energy modeller, CMA; 21.1.2016), Bernard Laponche (energy expert and activist, formerly CEA and director of Ademe; *conducted with Alain Nadaï,* 19.01.2017), Jean-Charles Hourcade (energy modeller, Cired; *conducted with Alain Nadaï* 19.4.2017), Michel Colombier (energy expert, Iddri; *conducted with Alain Nadaï* 15.3.2017), Bertrand Château (energy modeller, Enerdata, formerly IEPE and Ademe; 31.5.2017), Bruno Lapillonne (energy modeller, Enerdata, formerly IEPE, IIASA and Ademe; 15.6.2017)

German energy experts and modellers (interviews 14-22)

Klaus Michael Meyer-Abich (philosopher, energy expert and member of PEK atomic energy; 20.4.2010), Wolf-Peter Schill (energy modeller, DIW; 15.7.2016), Nico Bauer (energy modeller, PIK; 22.1.2015), Alexander Popp (energy modeller, PIK; 25.9.2015), Elmar Kriegler (energy modeller at PIK; 25.9.2015), Jan C. Minx (energy & environmental policy expert, Mercator Institute, formerly IPCC; 25.9.2015), Julia Repenning and Ralph O. Harthan (energy modellers, Öko-Institut; 31.10.2016), Sabine Gores (energy modeller, Öko-Institut; 26.10.2016), Felix Matthes (energy modeller, head of Öko-Institut; 31.1.2016)

French civil servants and politicians (interviews 23-26)

Dominique Chauvin (head of sustainability at Total, member of several public energy forecasting exercises; 16.7.2016), anonymous interviewee (civil servant, ministry of economy, formerly ministry of ecology; *interview by Alain Nadaï*, 24.11.2016), anonymous interviewee (energy modeller, Ademe; 2.2.2017), anonymous interviewee (civil servant at ministry of economy, formerly ministry of ecology; *conducted with Alain Nadaï*, 26.4.2017)

German civil servants and politicians (interviews 27-30)

Klaus Töpfer (former minister of environment; 29.4.2010), Reinhard Loske (former member of parliament, member of PEK climate; 27.4.2010), Martin Weiss (civil servant and energy expert, ministry of environment; 18.7.2016), Kai Kuhnenn (energy expert, formerly at UBA; 15.7.2016)

Organising Policy-Relevant Knowledge for Climate Action: Integrated Assessment Modelling, the IPCC, and the Emergence of a Collective Expertise on Socioeconomic Emission Scenarios

Béatrice Cointe

IK Centre for Technology, Innovation and Culture, University of Oslo, Norway/beatrice.cointe@tik.uio.no

Christophe Cassen

Centre International de Recherche sur l'Environnement et le Développement (CIRED), CNRS, France

Alain Nadaï

Centre International de Recherche sur l'Environnement et le Développement (CIRED), CNRS, France

Abstract

Greenhouse gas emission scenarios are key to analyse of human interference with the climate system. They are mainly produced by one category of computer models: Integrated Assessment Models (IAMs). We analyse how IAM research organised into a community around the production of socio-economic scenarios during the preparation of the IPCC AR5 (2005-2014). We seek to describe the co-emergence of a research community, its instruments, and its domain of applicability. We highlight the role of the IPCC process in the making of the IAM community, showing how IAMs worked their way to an influent position. We then survey three elements of the repertoire that served to organise collective work on scenarios in interaction with the IPCC and the European Union, and which now frames the community and its epistemic practices. This repertoire needs to articulate epistemic practices with the pursuit of policy relevance, which shows how epistemic communities and patterns of co-production materialise in practical arrangements.

Keywords: climate change, models, scenarios, repertoires, epistemic community

Introduction

The apprehension of possible futures is a crucial part of how we understand and tackle climate change. Long-term quantitative scenarios are one of the main devices used for this apprehension. They are mobilised as descriptions of "plausible trajectories of different aspects of the future that are constructed to investigate the potential consequences of anthropogenic climate change" (IPCC, undated). Among such scenarios, greenhouse gas emission scenarios, usually associated with socioeconomic narratives,[1] play a pivotal role – here we refer to them as 'socioeconomic emission scenarios'. They map possible evolutions of the drivers of climate change, and are used as input for climate models and as a basis for assessing climate policy options. Successive sets of socioeconomic emission scenarios have informed IPCC Assessment Reports since 1990, serving as references across climate research communities (Moss et al., 2010; O'Neill and Nakicenovic, 2008). They provided the backbone of the assessment of mitigation in the IPCC Fifth Assessment Report (AR5) (IPCC, 2014a). But how are such scenarios made?

Socioeconomic emission scenarios are now almost exclusively produced using one type of models labelled 'Integrated Assessment Models' (IAMs), to the extent that it is difficult to consider the scenarios and the models independently. IAMs are large-scale, complex numerical models that represent interactions among environmental, technological and human systems in a single integrated framework. They are used to generate quantified scenarios about the long-term evolutions of these interactions, usually on a global scale. To do so, they integrate contributions from various disciplines, among which environmental sciences, economics and engineering, with the express intention to inform policy-making (Weyant et al., 1996). In the AR5, they are opposed to other approaches to scenario generation, such as qualitative scenarios or aggregated models for cost-benefit analysis (Clarke et al., 2014: 422).

Most IAMs emerged in the 1990s. They have developed and expanded over the past 15 years, and a dedicated research community has gradually formed around them. IAM research is now a rather prominent source of expertise on climate change policy, as suggested by its central position within the latest report of the IPCC Working Group III (WG III), dedicated to mitigation. In this paper, we analyse how IAM research has organised around the production of socioeconomic scenarios during the period leading up to the publication of the IPCC AR5, from 2005 to 2014. This episode was about the production of specific kind of knowledge on the future as much as about the definition of IAMs as a category of models and as a field of research. By retracing it, we seek to understand the establishment of IAMs as central devices in the production of climate projections, and, it follows, that of IAM research as legitimate expertise on climate action.

IAMs – and expertise on climate change mitigation more broadly – have received limited attention from STS scholars. So far, only a handful of STS publications have scrutinised them (Wynne, 1984; Shackley and Wynne, 1995b; Edwards, 1996; Lövbrand, 2011; Weszkalnys and Barry, 2013; Beck and Krueger, 2016). Yet, their central position within the IPCC, their intrinsic interdisciplinarity and their overt ambition for policy-relevance make them particularly intriguing objects from an STS standpoint.

As a first step into the exploration of this object, our objective is to characterise IAM research in terms of its social, material, theoretical and institutional organisation: what is it, and how did it take shape? How does its vocation for policy-relevance translate in this organisation, and in its relations with institutions such as the IPCC or the EU? Like Knorr-Cetina (1999) in her analysis of epistemic cultures, our focus is on the organisation of knowledge production rather than on the content and circulation of knowledge. Our paper is thus meant as a contribution to studies of the social as an instrument in the production of knowledge and to practice-oriented accounts of the constitution of scientific communities and fields. It investigates the emergence, stabilisation and dynamics of research communities as related to the applications of their research, especially when such applications have to do with policy. In that, it joins up with recent attempts to unpack the making of epistemic communities (Akrich, 2010; Lorenz-Meyer, 2010; Meyer and Molyneux-Hodgson, 2010; Demortain, 2017).

We rely on a qualitative study combining document analysis, interviews and ethnographic observation. We identified the research teams involved, retraced the chronology of the production of socioeconomic emission scenarios for the AR5, and collected materials related to the events and projects that contributed to it (reports, presentations, articles). One of us had first-hand knowledge of some of these projects as a member of CIRED's IAM team since 2008. We also participated in IAM-related conferences, and interviewed 15 modellers and experts from seven institutions to understand the individual and collective aspects of their work and get insights into the evolutions and challenges of IAM research.[2]

The first section clarifies our take on the IAM community and explains how we relate to work on epistemic communities, on the interface between climate science and policy, and on the collective dynamics of scientific research. We then retrace the history of the interactions between IAM research and the IPCC and relate how IAMs came to play a central role in the preparation of the latest IPCC report. Following this historical account, we look closer into the details of IAM research, first pointing out the heterogeneity across IAM models, and then analysing elements in the repertoire that enabled the constitution of a coherent community out of this heterogeneity.

Investigating the co-emergence of a policy-relevant science and of its applications

This paper is interested in how integrated assessment modelling holds together as a research community sharing a reliance on a type of large, complex numerical models labelled as IAMs and an ambition for climate policy-relevance. Our focus is on the collective organisation and establishment of an emerging research community as a policy-relevant field.

In studies of the science-policy interface, the concept of epistemic communities is prominent. The most influential definition of the term is that proposed by Peter Haas: epistemic communities as "network[s] of professionals with recognized expertise and competence in a particular domain and an authoritative claim to policy-relevant knowledge within that domain or issue area" that share both a set of epistemic values and an orientation towards specific policy action (Haas, 1992: 9). While Haas's conceptualisation of epistemic communities explicitly draws on sociological approaches to the collective dimensions of scientific activities (such as Fleck's notion of 'thought collective' and Kuhn's paradigms), he used it primarily to introduce knowledge and expertise as relevant factors in the analysis of international politics, especially on environmental issues. His focus was on what epistemic communities do, rather than on how they are made. Since then, the notion has been taken up widely, with various reinterpretations (Lorenz-Meyer, 2010). In particular, it has been applied to the analysis of climate change as a global political issue largely framed by scientific expertise (Godard, 2001). Edwards (1996) argued that global comprehensive modelling (including both IAMs and Earth System Models) contributed to the emergence of an epistemic community by acting as a vehicle for shared knowledge, values, tools and data. Some studies of the IPCC have also relied on the epistemic community model to analyse the production of usable scientific knowledge on climate change and its institutionalization (Hughes and Paterson, 2017).[3]

IAM research is based upon trust in scientific knowledge and upon academic standards of validation; it has an ambition for policy-relevance; and it is a significant contributor to the IPCC reports, hence part of recognized expertise on climate issue. It can then be considered as part of a climate science epistemic community. It may even constitute an epistemic community in its own right, with distinct standards and norms of validation and, possibly, its own channels of policy influence, but so far, little is known of these standards, norms and validation. Our objective is to explore and understand the specificities of IAM research: what distinguishes it from other types of climate-related science? How is it organised? What does its authority within climate change expertise rely on? In other word, we seek to investigate how IAMs and socio-economic emission scenarios are made, which we consider a prerequisite to analysing how they are used and to inquire into the precise nature of their interactions with policy-making.

Our approach raises two questions related to different topics of research in STS. The first relates to diversity within climate science. While climate science has expanded dramatically over the past few decades, few studies have tried to unpack the plurality of scientific perspectives that make it up, and the relationships among them. Detailed analyses of the making of climate knowledge have tended to focus on the natural science side of climate research, and especially on Global Circulation Models (GCMs). Shackley and Wynne (1995a, 1995b), Shackley et al. (1999) and Demeritt (2001) have analysed how GCMs came to dominate climate science in a context of "mutual construction of climate science and policy" (Shackley and Wynne, 1995a) and of their expectations towards one another. However, when it comes to expertise on climate change adaptation and mitigation (that is, the domains of WG II and WG III of the IPCC), we know much less (Hulme and Mahony, 2010). Recent studies have mapped the research networks and disciplines involved in WG III of the IPCC (Corbera et al., 2015, Hughes and Paterson, 2017), but with a focus on personal trajectories and institutional affiliations, rather than on the scientific perspectives, instruments and practices used. Investigating IAM research as one of the scientific approaches represented within WG III – and a particularly influent one – is a step towards a more refined understanding of climate expertise.

Our practice-oriented approach relates to a second STS question, which has to do with the emergence of scientific communities and in particular epistemic communities. Early laboratory studies tended to shun approaches in terms of scientific community: instead of imposing abstract, predefined social units to the description of science, they argued, one should focus on empirical accounts of scientific practices and emphasise the multiple relationships in which such practices are entangled (Knorr-Cetina, 1982). Since then, STS have re-appropriated the notion of scientific community. While rejecting the characterisation of such communities as "focused largely on shared theories and constitutive of a discipline or field" (Leonelli and Ankeny, 2015: 702), they approach them from a practice-oriented perspective, providing material and situated accounts of the collective organisation of research (Molyneux-Hodgson and Meyer, 2009; Meyer and Molyneux Hodgson, 2010; Leonelli and Ankeny, 2015; Merz and Sormani, 2016). Similarly, STS have seized the notion of epistemic community with a view to enriching it. Meyer and Molyneux-Hodgson (2010), Akrich (2010) and Demortain (2017) have all pointed out a lack of research on how epistemic communities emerge and how they produce knowledge, share it, and make it relevant for policy: epistemic communities often appear as finished products, that is as homogeneous and readily available when policy problems emerge. This leaves behind the question of how policy-relevant research is produced and stabilised. Meyer and Molyneux-Hodgson (2010) suggest that we view epistemic communities as dynamic entities and call for studies on how they come into being, how they are made and materialised. This amounts to investigating how scientific knowledge is made so as to be policy-relevant and how scientists organise to act with knowledge. More broadly, this is an invitation to study the joint emergence and structuration of research and of its domains of applicability, especially when these domains of applicability have to do with policy-making.

IAM research appears particularly suited to take such analyses further: it predominantly takes place in academic settings, but its vocation for policy usefulness is explicit (Shackley and Wynne, 1995b: 122; Edwards, 1996; Weyant et al., 1996). It is also in large part organised in project-based collaborations of various types, such as large model intercomparison projects, EU-funded consortia, or contributions to the work of the IPCC. This entails specific conditions for knowledge production. For instance, in an analysis of the European Commission-funded project ADAM, Eva Lövbrand showed how integrated assessment modellers involved in the project aligned their scientific objectives with the European Commission's expectations, providing assessments that supported (rather than challenged) the EU's policy goals; but, in so doing, they opened new research questions and expanded their scientific horizons (Lövbrand, 2011: 232-233).

To account for the peculiar dynamics and organisation of IAM research, we borrow Leonelli and Ankeny's notion of 'repertoires' (Leonelli and Ankeny, 2015). Repertoires are shared sets of norms, infrastructures, procedures and resources that successfully adapt to the broader research and funding context, and that come to structure the development of communities committed to using them. Leonelli and Ankeny are interested in how some temporary projects perpetuate into lasting communities. For them, the notion of repertoire "captures what happens when specific projects become blueprints for the way in which whole communities should do science" (Leonelli and Ankeny, 2015: 705). By communities, they refer to "group[s] of individuals brought together by repeated interaction around one or more goals, which can range from the pursuit of a given interest to the production of a tool, the development of a procedure, or the use of a common space (whether physical or intellectual)" (Leonelli and Ankeny, 2015: 702). We use the term in a similar way when we refer to 'the IAM community'. Our objective is to map out how such a community emerged out of punctual projects, and what shared repertoire holds it together. However, contrary to the biological research communities that Leonelli and Ankeny study, IAM research is almost constantly concerned with its relevance and applicability: constructing policy-relevance is part of the work of establishing a repertoire.

The centrality of policy-relevance in the organisation of the IAM community shows in the type of projects in which the repertoire emerged. Indeed, these were largely driven by the agenda of the IPCC, especially in the preparation of the AR5, and, to a lesser extent, by requests from the European Commission to assess options for climate policy. The involvement of IAMs within the IPCC process thus appears as a key driver in the constitution of the repertoire of IAM research. The following section retraces the history of this involvement with a focus on how IAMs became central to the production of socioeconomic scenarios within the IPCC process.

The production of socioeconomic emissions scenarios for the AR5: a catalyst for structuring the IAM community

Socioeconomic emissions scenarios and IAMs in the IPCC AR5

IAMs have played a significant role in the AR5 (Edenhofer et al., 2014: 48), where they are described as "invaluable to help understand how possible actions or choices might lead to different future outcomes" (Edenhofer et al., 2014: 51), that is as guides for political decision. Two types of IAM-generated scenarios appear in the AR5. First, four 'Representative Concentration Pathways' (RCPs) representing contrasted possible emission trajectories to 2100 served as input for the elaboration of new climate change projections by climate models (IPCC, 2013: 164, 1060). They were produced by four IAMs. Second, IPCC WG III collected a database of 1184 peer-reviewed socioeconomic scenarios (IPCC, 2014b). Thirty models contributed to the scenario database, with eleven providing 966 out of 1184 scenarios (IPCC, 2014a: 1309-1310).

Box 1. Climate science acronyms

GCM: General Circulation Models are physics-based models of the atmosphere and ocean, used for weather forecasting, to study the climate, and to generate long-term projections of climate change.

ESM: Earth System Models are natural sciences-based models that represent biogeochemical cycles (especially the carbon cycle) in addition to the climate system.

IAM: Integrated Assessment Models draw on engineering, economics and natural sciences to represent interactions between human, technological and environmental systems. They are used to produce socioeconomic emission scenarios and to assess global climate policy options.

IAV: Impact, Adaptation and Vulnerability is a heterogeneous field that studies the vulnerability and adaption of socioeconomic and natural systems to the consequences of climate change, often at a regional rather than global scale.

IPCC: The International Panel on Climate Change, created in 1988, regularly produces overview of peer-reviewed climate science. It does not produce research and it claims to be policy-relevant, but not policy-prescriptive. It comprises three Working Groups: WG I on the physical basis of climate change; WGII on impacts, adaption and vulnerability; WG III on mitigation.

The use of IAMs to produce such scenarios specifically in view of IPCC assessment reports was not a novelty, though it gained importance in the AR5. The website of the Integrated Assessment Modeling Consortium (IAMC), a key forum for IAM research, mentions the involvement of IAMs as early as the First Assessment Report, and portrays the histories of IAMs and of the IPCC as intertwined:

> Development and analysis of global to regional and country scenarios have been at the heart of integrated assessment modelling from its earliest days: scenarios to underpin the 1st Assessment Report of the IPCC were elaborated with 1st generation IAMs. (IAMC Website, undated)

Looking back to the early days of the IPCC, this section shows how the production of socioeconomic emissions scenarios drove the development of IAMs and their involvement in the IPCC process. This intensified in the late 2000s, when the IPCC delegated the production of scenarios to the ‘scientific community’ (IPCC, 2006), and through the intermediary of a few IAM teams.

The historic role of IAMs in the production of IPCC socioeconomic emissions scenarios

In 1988, the IPCC received a mandate from the UN General Assembly to produce regular assessments of the physical impacts and climate policy aspects of climate change in order to inform policy-makers. Over its first four Assessment Reports, the IPCC has orchestrated the elaboration of three generations of socioeconomic emissions scenarios used as references for the evaluation of future climate change, its impacts, and its techno-economic implications.

The first two generations of IPCC socioeconomic emissions scenarios, the SA90 and IS92 (Leggett et al., 1992) were produced respectively in 1990 and 1995, as part of the IPCC’s First and Second Assessment Reports. They were produced using the two main IAMs operational at the time: ASF, developed by the EPA in the US, and IMAGE, developed by the Dutch National Institute for Public Health and the Environment (RIVM).

In 2000-2001, the preparation of the Third Assessment Report (TAR) was a turning point regarding the substance of scenarios. Their construction was recognized as a means for organising and communicating the uncertainties associated with climate policy. Four storylines were developed using a forward-looking approach: first, describing socioeconomic driving forces, then modelling resulting emissions and atmospheric concentrations of greenhouse gas and aerosols. The reference scenarios were published in 2000 in the *Special Report on Emissions Scenarios* (SRES) (Nakicenovic et al., 2000).[4] They served as references in the Third and Fourth IPCC assessment reports (IPCC, 2001; 2007). In a sequential approach, the SRES scenarios provided emissions trajectories both for climate models (using them as input to project the magnitude and pattern of climate change) and for impact models (using them as input to evaluate climate change impacts).

The productions of SRES scenarios mobilised more modelling teams than the two previous generations of scenarios. Six IAMs were selected to develop reference scenarios: MESSAGE, IMAGE, MARIA, AIM, MiniCam, ASF. These models were developed in research institutes focusing on modelling environment and climate issues that were based in Europe (IIASA, RIVM), in Japan (RITE, NIES) and in the US (PNNL, EPA).

The SRES provided a first opportunity for IAM teams to work together. It also brought crucial upstream inputs for the two other IPCC Working Groups (Interview 6). The International Institute for Applied Systems Analysis (IIASA) played a central role. Founded during the Cold War to foster collaborations between scientists from the East and the West, IIASA builds upon a long tradition in the modelling of energy and environmental systems initiated in the 70s. Researchers in these fields have participated in IIASA’s Energy Systems Program over the years. In the 90s, energy research and climate research began to merge and grew increasingly involved in the IPCC. Nebojsa Nakicenovic, head of the “Transitions to New Technologies Project”, gradually emerged as a leading figure. Together with Bert de Vries (head of the IMAGE team at RIVM), he took on a major role in the coordination of the SRES as convening lead author.

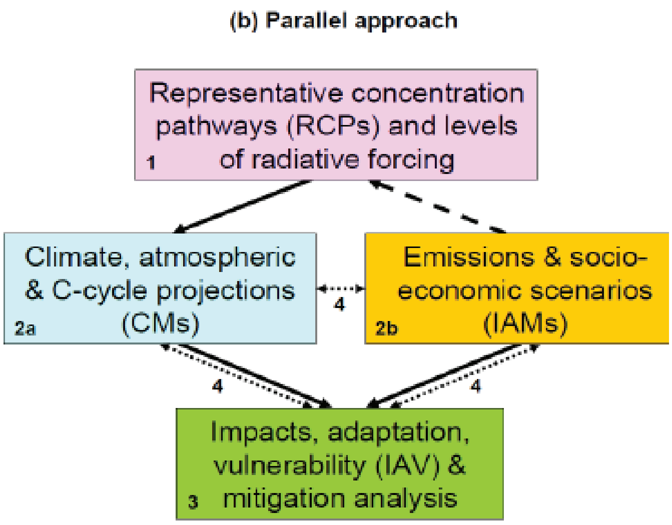

Figure 1. The AR5 new scenario approach (Source: IPCC, 2007)

Socioeconomic scenarios for the AR5: a new approach

The preparation for the Fifth Assessment Report (AR5) started in 2004, even ahead of the publication of the Fourth. The IPCC wanted to develop a new approach to socioeconomic emissions scenarios and, subsequently, to the interactions between IPCC and IAMs. Indeed, faced with criticism of the limitations of the SRES scenarios (Webster et al., 2003; van Vuuren and O'Neill, 2006), the IPCC acknowledged the need for new 'emissions scenarios' following the AR4. In 2005, it created a *Task Group on New Emission* Scenarios (TGNES) to study the matter. In 2006, the TGNES suggested a parallel approach to scenarios production associating the communities of climate modelling, Impact Adaptation and Vulnerability (IAV) and IAMs for the AR5 (Fig.1, Box 2). The IPCC endorsed this strategy at the Noordwijkerhout expert meeting in 2007 (Moss et al., 2008).

IAM teams, especially those that played a leading role in the SRES process, were significantly involved in both the TGNES (9 out of 31 members) and in the and in the preparation of the Noordwijkerhout meeting.

The delegation to IAMs: a catalyst for community-making

Alongside discussions about the production of new emissions scenarios, debates focused on the role of the IPCC in their development.[6] The IPCC had directly organised the development of the previous sets of socioeconomic scenarios. This time, the options favoured following the expert

Box 2. The "parallel approach" to scenarios

The parallel approach adopted for the AR5 process was intended to address the limitations of the SRES scenarios, especially the delay required to use the scenarios in studies of impacts, adaptation, and vulnerability (Parson et al., 2007; Moss et al., 2010) and the difficulty for models to completely reflect storylines decided separately from model construction. This new approach started with the selection of four RCPs, from available scenarios in the IAM literature. The RCPs were finalised after a huge work of harmonization with climate scientists. They were supposed to help develop new set of climate model simulations "at the same time that new work [was] carried out in the Integrated Assessment Model (IAM) and Impact, Adaptation and Vulnerability (IAV) communities" (van Vuuren et al., 2008). They served as references for the evaluation of socioeconomic implications and climate policy options by IAMs for the AR5 WG III[5]. IAM played a central role throughout this process, as they produced both RCPs (4 out of the 6 used in the SRES) and socioeconomic emissions scenarios for WG III.

meetings were for the IPCC to simply facilitate or coordinate the development of new socio-economic emission scenarios by independent researchers, rather than carry it out itself. This was seen as a way to avoid a technocratic process while guaranteeing the independence of the IPCC and the work across Working Groups.

At its 25th Session in 2006, the IPCC delegated the preparation of those scenarios to the 2research community" describing itself as a "catalyst" for scenario production (IPCC, 2006). Lobbying from the research community weighed in the decision. Our interviews suggest that IAM teams informally convinced the IPCC bureau, in particular the co-chair of the IPCC, Rajendra Pachauri, that they were able to convene the process. A group of modelling teams from IIASA and NIES, headed respectively by Nebojsa Nakicenovic and Mikiko Kainuma (who had both played a key role in the SRES), together with John Weyant (director of the Energy Modeling Forum, in Stanford), decided to establish a specific consortium to that end, the IAMC.

> So, we had convinced Pachauri and the bureau during the four workshops on scenarios during the AR4, this is in preparation for the next round essentially, we could do it. Toward the end of the period Pachauri told us: I know you guys want to do this, but you do not have any funding and support. He wanted to be sure we could make it. So, a few of us, N. Nakicenovic, M. Kainuma, got together overnight and said: we are starting a new institution, and we are going to have the Integrated Assessment Modelling teams and we are going to call it the Integrated Assessment Modeling Consortium with annual meetings, scientific assessment committee and so on. (Interview 11)

The IAMC was formally created in 2007 and a consortium agreement was elaborated with the IPCC. It is overseen by a "Scientific Steering Committee" comprising the three founding teams, PNNL, as well as several European (PBL/the Netherlands, PIK/Germany, FEEM/Italy) and Southern teams (IMA/India, ERI/China, UFRJ-COPPE/Brazil). The list of participants in the IAMC overlaps with that of lead authors in the contribution of WG III to the AR5, particularly those of chapter 6 on "transformation pathways", which was based on IAM outputs. So-called "transformation pathways" – IAM-generated scenarios – served as a red thread to ensure the overall coherence of the report, and researchers working with scenarios were spread out as authors across the chapters (Interview 3).

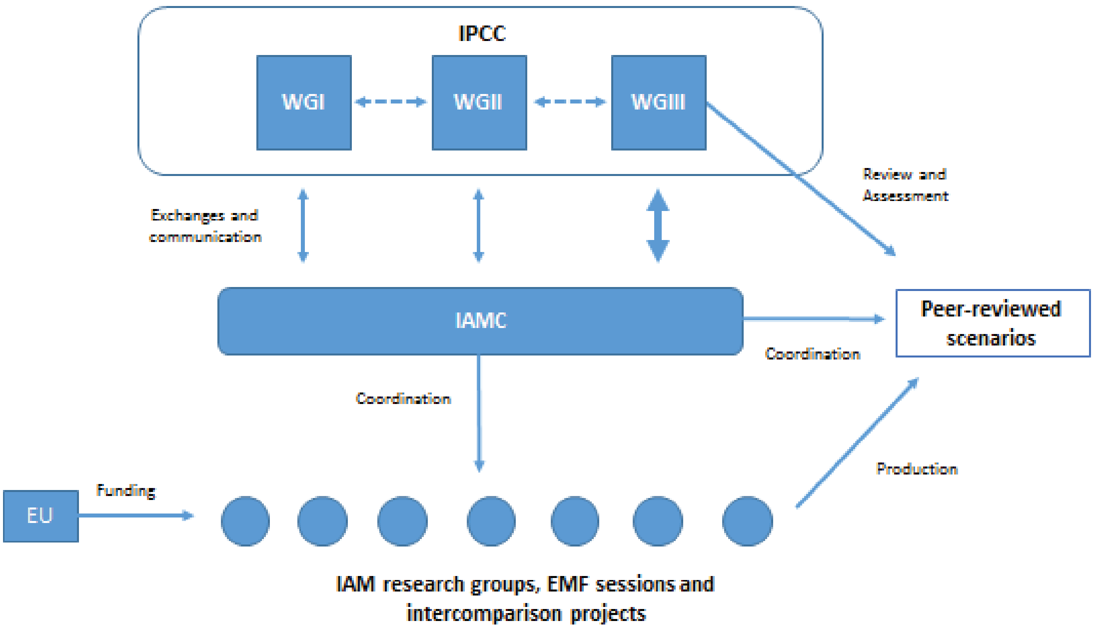

Figure2. Formal organisation of interactions between IAMC, IPCC and the IAM community during the AR5 process. Many people circulate across institutions and are involved in several of these groupings.

The main role of the IAMC is the coordination of research activities within the IAM community, initially focusing on the preparation of RCPs. The IAMC itself does not produce scenarios or conduct research. The scenarios referenced in the AR5 were produced either by individual teams (as in the case of RCPs) or in a series of Model intercomparison projects (MIPs). MIPs have been a regular feature of IAM research since the early 90s, in particular those organised by the Energy Modeling Forum (EMF). The EMF is a point of reference with an "enormous convening power on [the IAM] community" (Interview 4), and two sessions related to climate change took place within the AR5 timeframe (i.e. between 2004 and 2014). A new feature in this period was the increasing role of European projects funded by the Commission within the 7th research Framework Programme, which pushed for evaluation of the 2°C target as it was the backbone of EU climate policies (Lövbrand, 2011). EU-funded projects largely focused on estimating the possibility to comply with the 2°C objective and the related mitigation costs according to the ambition of countries' commitments or the availability of low carbon technologies among major emitters from North and South countries (Table 1). The outcomes of these projects were collected into dedicated databases, published in peer-reviewed publications, and assessed by the IPCC WG III (Fig. 2). These scenarios provided the backbone for the WG III report on mitigation (IPCC, 2014a; interview 5).

Since the early 90s, the production of socioeconomic emission scenarios has fostered collaborations among IAM teams. This was in large part a result of the demand from the IPCC for reference socioeconomic emission scenarios and, in the preparation of AR5, for a wide range of emission scenarios testing diverse options for climate policy. The preparation of the AR5 was a milestone in the structuration of the IAM community: it framed the production of a new generation of socioeconomic emission scenarios as a community-wide effort and gave more attention to IAMs than previous report. What are the specificities of this community, and what holds it together, aside from the participation in the production of emission scenarios?

Variety and convergence among IAMs

From the account of the interactions between the IPCC and the IAM community, IAM research can appear as a relatively small and close-knit field, comprising of about 30 teams that regularly collaborate in projects. However, taking a closer look, the IAMs referred to in the AR5 turn out to be quite diverse: they do not constitute a homogeneous category. In this section, we review the specif-

Table 1. Model Intercomparison Projects contributing to the AR5 database (adapted from IPCC, 2014a: 1311)

Modelling Intercomparison Projects	Date	Type	Leader	Teams	Models	Scenarios in the AR5 database
ADAM	2006-2009	EUFP6	Tyndall Center	7	5	15
EMF22	2008-2009	EMF session	Stanford University	23	17	70
AME	2009-2011	International	PNNL	17	27	83
POeM	2008-2011	EUFP7	Chalmers Univ of Technology	8	3	4
RECIPE	2008-2010	European	PIK	4	3	18
EMF27	2010-2013	EMF session	Stanford Univ	24	17	362
ROSE	2010-2012	Stiftung-Mercator funded	PIK	5	4	105
AMPERE	2011-2014	EUFP7	PIK	23	14	378
LIMITS	2011-2014	EUFP7	FEEM	13	7	84
GEA	2008-2012	IIASA	IIASA	2	2	1

icities of IAMs and outline what such specificities entail in terms of talking of an IAM community.

The heterogeneity of IAMs

IAMs share a few characteristics: their complexity, their global scale, the fact that they represent both physical and social phenomena, their vocation to "help understand how possible actions or choices might lead to different future outcomes" (Edenhofer et al., 2014: 51), and their use of economics as a basis for decision-making. Beyond these similarities, IAMs encompass a wide diversity of approaches to modelling, which is reflected across the thirty models referenced in the AR5 (Clarke et al., 2014: 422). Our point here is not to map this diversity or discuss classifications of IAMs – a task best undertaken by modellers themselves (Dowlatabadi, 1995; Weyant et al., 1996; Hourcade et al., 2006; Crassous, 2008; Clarke et al., 2014; Lefevre, 2016) – but to point to some manifestations of the disparate character of the IAM category besides the technical features of models.

First, integrated assessment modelling is interdisciplinary by definition. It shows in the backgrounds of modellers: among those we interviewed, aircraft engineering, physics, economics, applied mathematics, biology, operations research, environmental sciences, chemistry and government studies are represented (and sometimes combined). All IAM teams and models integrate elements from several disciplines, but they do so in different ways, and it tends to influence modelling choices and methods.

Second, whereas GCMs share a common physical basis, there is no single unifying theory of integrated assessment. Integrated assessment modelling borrows from various intellectual traditions, including energy systems modelling, macroeconomic forecasting and systems dynamics. Models vary in their architecture, in their philosophy and in the type of questions they are designed to address, with clear differences between natural science or engineering-based models and economics-oriented ones (Risbey et al., 1996: 372).

For instance, IMAGE is a geographically explicit simulation model that started from a systems dynamics approach. It has a distinctly environmental science orientation and a rather detailed energy system module:

> With IMAGE, our vision is to represent the world as much as possible in terms of the physical reality, so I prefer to describe the agricultural system in number of cows and the tons of cereals that are produced, and I'm not very interested in how many euros are produced, because the connections to the environmental change parameters are the cows in the area. (Interview 7)

By contrast, ReMIND, developed at PIK in Potsdam, started as an economic model to study endogenous technological innovation dynamics, and evolved to incorporate details on different energy technologies; it is now coupled to a land-use model developed in the same institute. In Italy, the FEEM's WITCH model is "a spin-off, essentially" in the tradition of Nordhaus' RICE model,[7] to which it added elements on technological progress as well as game-theoretical structure (Interview 12b).

Each IAM carves its own niche along the way. Common theoretical filiations, inclusions of modules from other models, shared assumptions, and other forms of kinships and coupling make the precise classification of IAMs a difficult business. IAMs tend to incorporate bits and pieces from different disciplines and intellectual traditions from which they build their own internal logic and relevance. These evolution patterns deserve further investigation, but IAMs appear to evolve through a complex interplay of available skills and interests, orientations given by funding, and research questions and insights that emerge along the way (Interview 10).

Differences across models affect the type of questions each model can address and the messages that can be derived from its outputs. Different IAMs have different strengths and limitations, and they are not all tailored to say the same things – though some limitations are shared by all, for instance the focus on technology change instead of lifestyle change or the reliance on economics as a basis for decision-making.

Increased cooperation across IAM teams

In such a disparate landscape, common ground cannot be taken for granted. The need for common standards and procedures, especially for model evaluation, was already pointed out in the beginnings of IAM research in the 1990s (Risbey et al., 1996). It remains a matter of reflection to this day (Schwanitz, 2013; Wilson et al., 2017).

At the same time, attempts at defining common standards indicate a will to orchestrate efforts so as to be able to speak as a community. From its early days, integrated assessment was conceived as a collective enterprise benefitting from a "healthy diversity of modelling approaches" that it was a good strategy to "maintain and extend" (Toth, 1995: 266). The challenge was – and still is – to articulate coherent messages out of this plurality, because the vocation of IAM research is to feed into the policy process (Weyant et al., 1996: 366). Though the IAM landscape has evolved, Risbey et al.'s diagnosis (1996) seems characteristic of the ambitions of IAMs:

> *As a community* we need to establish norms and procedures that distinguish good analyses from bad ones, to be more reflexive about our own analyses, and to make all efforts to guard against *our* analyses falling prey to political expediency. Otherwise, *we* run the risk of not being heard at all, or of speaking with the wrong voice in the political clamor over climate change (Risbey et al., 1996: 370, emphasis added)

In fact, the unifying principle behind IAM research does not lie in a core theoretical basis, but in the dual ambition to represent complex systems through a combination of disciplinary insights and to provide policy-relevant assessments – but its legitimacy to do so rests on epistemic grounds whose soundness needs to be collectively guaranteed. This distinctive feature of IAM research has largely shaped its collective organisation. It can account, at least partly, for the prominence of collective projects and institutional hubs in the IAM community. Institutions such as the EMF in Stanford, which has coordinated model intercomparisons since the 1970s, or IIASA, where many IAM researchers have spent time (Corbera et al., 2015), have served as nodes for sharing and comparing modelling perspectives and results. The IPCC seems to constitute a similar nodal point for IAM research. Corbera et al. (2015: 96) have analysed patterns of authorships in WG III, showing that a small number of researchers co-author regularly with each other; most have contributed to several IPCC reports, suggesting they may have organised their career around the IPCC process. Out of the top 20 authors in this group, we identified a dozen as directly involved in IAM research networks.[8]

Since the mid-2000s, contacts and common projects have multiplied, driven by the momentum provided by the preparation of the AR5 and by a series of EU-funded projects on climate mitigation options. Networks that used to be separate have merged (Interviews 4, 12a, 12b), and interactions across teams have intensified and stabilised. The following section analyses the repertoire that emerged and stabilised as IAMs were mobilised to produce scenarios in view of the AR5, and how it contributed to the organisation of the IAM community.

The IAM community's repertoire

Several large projects took place between 2005 and 2014 (Table 1), which was a period of intense collective activity for IAM researchers who worked on the RCPs and were expected to assess a range of mitigation scenarios. These projects were framed both by demands from the IPCC (a new generation of reference socioeconomic emission scenarios; an evaluation of the implications of a 2°C target), and by the EU's request for science-based support for its own climate policy objectives (Lövbrand, 2011). They stimulated cooperation, intensified interactions across IAM teams, and led to the setting up of devices and institutions to work with the heterogeneity of IAMs. These were not only crucial in the preparation of the AR5, they have remained in place to this day. They now shape the way IAM research is carried out, and delineate the IAM community: they constitute a 'repertoire' enabling the continuation of collective work, framing common goals and standards, and ensuring the transferability and legitimacy of IAM results (Leonelli and Ankeny, 2015). In this last part, we focus on the three main features of this repertoire: the IAMC, the organisation of work in Modelling Intercomparison Projects, and scenario databases. All address the same core challenge of IAM research: extracting a policy-relevant message out of diverse modelling approaches and philosophies.

The Integrated Assessment Modeling Consortium (IAMC)

The IAMC is a *sui generis* institution that has become a central node in the collective organisation of IAM research. It was created in 2007 to coordinate the production of RCPs for the IPCC but outlived this purpose to become a community organisation. As one of its founders summarised, "we are trying to do our own community, this is the role of IAMC now" (Interview 11).

From the start, the IAMC served as a forum to discuss the evaluation of modelling outputs in reaction to external requests and to organise relationships with end-users of scenarios, chiefly climate scientists. The coordination of the RCP process required the harmonisation of scenarios produced by different IAMs and their adaptation to the needs of the climate scientists who would use them. Once the RCPs were ready, the IAMC broadened its scope to become a proper discipline organisation. As one modeller recalls:

> All these people that were coming to the IAMC and were not part of the RCP development – I didn't know why they were coming, because on our annual meetings we were all the time discussing the RCPs. And so, a couple of years ago, we decided to completely reform the IAMC into a much more useful organisation, which is now this discipline organisation, similar to the AGU [American Geophysical Union] for geoscience. So, we want to become this discipline organisation which organises this annual conference to look into interesting topics and to share knowledge. Also at other moments of time, we have our working groups to help the community. (Interview 7)

The IAMC convenes annual meetings since 2008, and the number of attendees is slowly, but regularly, increasing. These meetings consist in an open conference, after which 'Scientific Working Groups' meet to discuss issues at the core of the practice of integrated assessment, such as data protocols, shared model documentations, or model evaluation and diagnostics. While actual work on these issues mostly takes place within specific projects (Interview 12a), the IAMC serves to bring it together in front of the whole community.[9]

The IAMC gradually established itself as a focal point for IAM research, "the central point where everything should go" and "the organisation that should coordinate activities" (Interview 12b). It provides an arena for negotiating and stabilising the epistemic culture of integrated assessment modelling as well as an institutional embodiment of the IAM community. All the same, it is a young organisation without permanent funding, which depends on the financial resources made available by member organisations. This limits its capacity to undertake much work beyond communication and meetings. The community remains largely dependent on government-funded projects for its activities. The IAMC also lives in the shade of better-known institutions such as the IPCC or IIASA: "it still needs to be credited" because "no one knows about it" (Interview 12b).

Model Intercomparison Projects (MIPs)

A significant part of IAM research occurs in MIPs. They aim to compare the outputs and behaviours of several models and to test how they react to specific sets of assumptions. Though inspired by climate science practices, this type of project is quite specific to the IAM community. The EMF has regularly organised such model intercomparisons since the 1990s, but the practice intensified and institutionalised during the preparation of the AR5. 95% of the scenarios considered in the report of WG III in the AR5 were generated in nine MIPs (Table 1). Each of these projects brought together more than a dozen of modelling teams from all over the world, strengthening interactions among them. Two were organised by the EMF, which coordinates but does not directly fund research; five were funded under the EU Framework Programmes. The EU-funded projects constituted a change in the scale and scope of MIPs, bringing European teams closer together and enabling their growth.

Participation in model intercomparisons, especially those convened by the EMF, is, as the leader of one team told us, "a matter of pedigree" (Interview 4): it is a sign that you belong to the community and that your model is recognised by this community. Besides, with the multiplication of EU-funded MIPs in the late 2000s, these projects have become one of the main sources of funding for IAM teams. They are also one of the main venues for collaboration across teams, influencing the organisation of IAM research. Because MIPs seek to pool together and make comparable scenarios produced by very different models, they entail both practical and theoretical reflections about the organisation and verification of modelling.

MIPs as they developed since the late 2000s tend to follow a similar pattern. Work is divided into several work packages, and it is shared between a project coordinator, leaders of work packages, and the rest of the participants. Starting from a central question, protocols are established to analyse sub-questions in separate work packages that define which types of scenarios need to be generated.

> [Protocols] change, of course, because they answer to different questions, but the structure is the same to isolate factors. Questions change but this is the same matrix with two axes: one axis with climate policy, typically 2°C, and on the other axis you have what you want to understand: technology, policy, structure, anticipation… and then you have scenarios – from 10 to 20 – in the matrix. The matrix has scenarios which are compulsory, optional. (Interview 12b)

The production of scenarios follows a standard protocol based on the comparison of policy scenarios against baselines without climate policies.

In a first step, a diagnostic describes how models differ in their response to climate policy, looking for instance at the rate of emission reductions for a given carbon price trajectory. Diagnostics aim at identifying patterns of model behaviour and contributing to their validation. This was inspired by the Coupled Model Intercomparison Project (CMIP) in climate science[10], as a consequence of the parallel process of scenarios production for the AR5 that encouraged exchanges across climate research communities. Scenarios assumptions are then implemented in a chosen set of models. This requires some harmonization across models in order to manage global uncertainty. Scenarios are generated by models and results analysed and compared in each work package, and each work package leader has scrutiny over the protocol and the processing of data.

MIPs entailed intense collaborations, fostering mutual learning about the specificities of each IAM. They drove improvements in model documentation and evaluation that helped to map and characterise differences between models. They contributed to the development of common modelling practices across teams and fostered innovation.

MIPs result in the production of sheer numbers of scenarios, and the need to manage them contributed to the development and stabilisation of another key element of the repertoire of the IAM Community: scenario database.

Scenario Databases

The first initiatives to constitute databases of socio-economic and emission scenarios date back to the 90s, but the practice gained prominence in the late 2000s. Here again, the development of standardised, publicly available scenario databases can be traced to the AR5 process. A first database was elaborated to gather the RCPs and make them available to their users. A second IPCC-related database followed, to collect scenarios as part of the preparation of the report of WG III, which planned to use them to map "the solution space" (Edenhofer, 2014).

In 2012, WG III issued an open call for scenarios: IAM teams were invited to submit socioeconomic emission scenarios for consideration by WG III. To be included in the database, scenarios had to meet a series of criteria: being peer reviewed, providing a minimum set of mandatory variables, scenario documentation, or coming from "formal energy-economic or integrated assessment models" with a large coverage of energy sectors (IAMC, 2012). The data template (an excel file) gathered general instructions, the description of the scenarios, a model classification and data breakdown by models/regions/variables every ten years until 2100.

These two databases are hosted on IIASA servers. IIASA has devoted human resources to the maintenance and operation of the databases in coordination with the IAMC (Interview 12a; Guivarch 2016, personal communication). A web-based infrastructure was built to enable modellers to upload their results and to allow outside users to access the data. It includes detailed information about the purpose of the database, the regions and sectors covered, the list of MIPs that contribute to the database, etc. (e.g. IPCC, 2014b)

With such infrastructure available, most MIPs now gather scenarios in similar databases. These are built according to similar templates and usually hosted on the same IIASA server, though not all of them are public. The evolution towards more standardised data management indicates an increasing professionalization of IAM-based scenario production, with a move from spreadsheets to big databases.

> While we previously, fifteen years ago, would run 3-4 scenarios and submit them via an Excel spreadsheet, and somebody would make – at hand – a PowerPoint presentation out of it, now we have these tools where we submit to a database, maybe 20 scenarios from each team, people have R scripts where they are able to pick up directly all kinds of analyses from these databases… So, we have become much more professional. (Interview 7)

Databases shape IAM research in several ways. First, they are tools for scientific research: they serve to organise collaboration, allow for rapid checking of errors in reported data, and have become "standards to read data" that modellers "use for themselves" as "a way to learn the model" (Interview 13). Second, they pool and order scenario data, making it available in usable forms. Some databases are public, and anyone can access the data and work with it: modellers consider they "make a huge service to other scientific communities" by creating and maintaining them (Interview 10). Third, the increased reliance on standardised database encourages a degree of convergence in modelling approaches.

Last, these databases pool together scenarios from a set of diverse models and organise them according to standardised templates, making them easily available. However, they do not stand alone: to work with them meaningfully, one must have a sense of how scenarios were produced, to answer which questions, and by which models. This is mostly transmitted informally, via discussions and mutual understanding fostered by regular interactions (Interview 5). The databases lose part of their meaning when separated from the collective that contributed to them, and in that respect subtly demarcate the IAM community. Integrated assessment modellers thus sometimes blame those who use the database without having access to this informal knowledge for treating these datasets "as numbers that are all the same", or as "statistical samples", whereas "to do justice to the database you would need to go through all the study protocols" (Fieldnotes, 2015b).

Effects of the repertoire: professionalization and convergence

The IAMC, MIPs and scenario database all emerged during the period of preparation of the IPCC AR5, largely as tools to organise the contribution of IAM research to it They have perpetuated and they play a crucial role in the current configuration of IAM research, shaping day-to-day work within research groups, collective organisation, and communication with external audiences. Modellers argue that the field has become "more professional" owing to the consolidation of this repertoire (Interview 7). Interactions among teams are more sustained and institutionalised, leading to better cross-knowledge of the strengths and weaknesses of models and to less heated debates about basic modelling approaches (Interviews 4, 7). The repertoire seems to allow for the articulation of common purposes. In particular, it has enabled a coordinated reflexion on model evaluation to eventually take place (Schwanitz, 2013; Wilson et al., 2017), mostly within MIPs but also via the IAMC and its scientific working groups. This comes with added visibility – and scrutiny – for IAMs and their results, especially since some scenario databases are publically available.

All the same, the consolidation of a repertoire for IAM research generates its own challenges and constraints. As Leonelli and Ankeny (2015: 706) noted:

> The adoption of a repertoire unavoidably creates strong commitments to particular techniques, assumptions, values, institutions, funding sources, and methods, which although initially productive, can sometimes act as constraints to future integration and innovation.

MIPs contributed to a proliferation of scenarios, to the extent that "[we might] reach a point where no two papers will use the same reference scenarios" (Fieldnotes 2015b). More crucially, the shared protocols of intercomparison and the standardisation required for inclusion in databases lead to a form of convergence and "group thinking" (Interview 7) that some critics consider as insularity (Fieldnotes 2015a; Interview 6). One widely acknowledged issue is that this organisation of research favours the investigation of common questions, the reproduction of scenarios, hence the development of similar features across models, to the expense of the improvement of models and the exploration of their core specificities (Interviews 5, 13). The convergence of research agendas is reinforced by the IAM community's current dependence on EU funding, which directs research towards the assessment of specific climate policy objectives, such as the 2°C or 1.5°C targets (Interviews 12b, 13).

Conclusions

As part of an investigation of how the socioeconomic emission scenarios used to study human interference with the climate are produced, we have studied the research community that works on them. This community is unified not only by the scenarios that it produces, but also by the fact that it uses models labelled as IAMs to do so. Despite constituting a heterogeneous set of models, these IAMs are all interdisciplinary models with a vocation for policy-relevance.

We have shown how interactions within the IAM community have intensified and organised since 2005, spurred by the IPCC and the EU. The elaboration of the IPCC AR5 was instrumental in this process. When the IPCC delegated the preparation of new scenarios to the scientific community, IAM researchers were on the front line: RCPs used by climate models were generated by IAMs, and so were the socioeconomic emission scenarios assessed by WG III. As for the EU, it funded several large IAM projects to inform and support its climate policy and to feed into the IPCC process; these enabled IAM research groups to capitalise on existing collective arrangements like the EMF, and to develop new methods and tools for cooperation, thereby equipping the definition of IAMs as a category of models.

Last, we analysed three elements of the repertoire that emerged during this period of intense collaboration: the IAMC, Model Intercomparison Projects, and scenario databases. These are now central to the way integrated assessment modellers do research. They frame epistemic practices and demarcate the IAM community. This repertoire organises harmonisation and professionalization as well as increased interpersonal and informal exchanges. We showed how it seeks to articulate the technicity and diversity of IAMs, the ambition to combine the variety of perspectives they offer, and the need for transparency heightened by their ambition to inform policy. This repertoire was also shaped by an ambition for policy-relevance that is constitutive of IAM research and that translated in close ties to the IPCC agenda. Indeed, it was developed during projects that were largely driven and framed by the preparation of the IPCC AR5. As a result, on top of framing common epistemic practices, the repertoire ensures that IAM research works towards specific applications – namely, informing climate policy choices. We have thus analysed the emergence and organisation of one applied science. This leads us to a set of empirical observations that raise general issues about applied sciences and their relations to their expected users.

The applied character of IAM research and its ambition for policy-relevance materialise in its symbiotic relationship with the IPCC. The IPCC acts as a communication channel between climate negotiations and climate science, as well as among climate research communities; it has become a central feature in the organisation of climate change research itself. This is particularly striking in the case of IAM research, whose origins story ties the evolution of IAMs to that of IPCC reference scenarios. We showed how the IPCC plays in the orientation, rhythm and domain of applicability of IAM research, but also that IAM researchers were heavily involved in the IPCC process, thereby influencing it, particularly during the AR5 process. This interrogates the demarcation between research and assessment at the core of the IPCC, at least in the case of WG III. To an extent, the separation between the two activities seems artificial and mostly institutional, especially when many researchers are involved, and influential, in both.

All the same, and however difficult it is to maintain, the demarcation comes with constraints that partly shaped the repertoire of the IAM community. It led to the creation of an intermediary institution, the IAMC, to coordinate the preparation of RCPs. The peer-review criteria imposed a deadline for the publication of project results. Last, the choice by WG III to assess "the full breadth of baseline and mitigation scenarios in the literature" (Edenhofer et al., 2014: 51) spurred the creation of a scenario database, a practice which has now become a standard of multi-teams IAM projects.

The joint construction of the IAM community and of the applications for IAM research also appears through the influence of the EU. Since the mid-2000s, most of the funding for European IAM teams came from MIPs funded by the DG Research Framework Programmes (FP6 and FP7). These projects stimulated interactions among teams and heightened the need for common databases and protocols. The need for comparability and the expectations from the EU inevitably influenced IAM research priorities.

By describing this process, we have shown how the IAM community worked its way to its current position in the academic landscape and with respect to climate change discussions. This position, we suggest, rests upon the articulation of epistemic practices with the pursuit of policy relevance. Emphasising this articulation as constitutive of certain scientific communities can inform a dynamic conception of epistemic communities as scientific communities that manage the balance and tensions between epistemic practices and policy relevance.

In the case of IAMs, the vocation for policy-relevance does not necessarily curtail scientific dynamism. Lövbrand (2011) found that the ability of modellers to align to the European Commission's expectations actually opened new scientific perspectives. Similarly, the repertoire that was constituted to enable IAM research to meet demands from the IPCC and the EU seems to stimulate research and to give IAM teams "an innovation boost" (Interview 4). However, the positioning of IAM research as policy-relevant also generates constraints and tensions, especially since policy framing and priorities evolve. For the IAM community, maintaining and reinforcing its current position implies adjusting to the dynamics of assessment and policies, which could come at a cost for research in the long-run. For instance, since the 2015 Paris Agreement, international climate negotiations focus on bottom-up initiatives and national mitigation policies, rather than on global action. IAMs being less suited for studies at the national scale, the IAM community seeks to assert its relevance in the face of competing expertise. In the context of EU funding, it is also expected to assess increasingly stringent climate objectives, such as the 1.5°C, and *de facto* contributes to their institutionalisation even when models have to be pushed to their limits to achieve them (Beck and Mahony, 2017; Interview 13).

Our analysis of the emergence, workings and dynamics of 'epistemic communities' complement studies on the co-production of climate futures, because they account for the way patterns of co-production take shape and evolve. It provides a necessary basis to analyse the uptake of IAM results in the policy process and the implications of the IAM community's presumably dominant position within climate expertise.

Acknowledgements

This research was supported by two CNRS grants for exploratory projects (SOCIOMODEL in 2015 and PROSPER in 2016) and by the Chair 'Modeling for sustainable development' of MINES ParisTech, Ecole des Ponts ParisTech, supported by ADEME, EDF, GRTgaz, RTE, and SCHNEIDER ELECTRIC. This paper has benefitted from numerous feedback and discussions during its preparation, in particular from the researchers at CIRED, from participants in workshops of the INNOX project, and from three anonymous reviewers. Stefan Aykut and David Demortain have greatly helped this paper to take shape. Last, the authors extend their thanks to all the modellers interviewed.

References

Akrich M (2010) From communities of practice to epistemic communities: health mobilizations on the Internet. *Sociological Research Online* 15(2):10.

Beck S and Mahony M (2017) The IPCC and the politics of anticipation. *Nature Climate Change* 7(5): 311-313.

Beck M and Krueger T (2016) The epistemic, ethical, and political dimensions of uncertainty in integrated assessment modelling. *WIREs Climate Change* 7:627-645.

Clarke L, Jiang K, Akimoto M et al. (2014) Assessing Transformation Pathways. In: Edenhofer O, Pichs-Madruga R, Sokona Y et al. *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge: Cambridge University Press, pp. 413-510.

Corbera E, Calvet-Mir L, Highes H and Paterson M (2015) Patterns of authorship in the IPCC Working Group III report. *Nature Climate Change* 6:94-100.

Crassous R (2008) Modéliser le long terme dans un monde de second rang : application aux politiques climatiques, Thesis, AgroParistech, France.

Demeritt D (2001) The construction of global warming and the politics of science. *Annals of the Association of American Geographers* 91(2): 307-337.

Demortain D (2017) Experts and the production of international policy knowledge: do epistemic communities do the job? In: Littoz-Monnet A (ed) *The Politics of Expertise in International Organizations: How International Bureaucracies Produce and Mobilize Knowledge*. London and New York: Routledge, pp. 76-92.

Dowlatabadi H (1995) Integrated assessment models of climate change: an incomplete overview. *Energy Policy* 23(4/5): 289-296.

Edenhofer O (2014) A Map of the Solutions Space: Mitigation. Our Common Future Under Climate Change, 7-10 July 2015, Paris, France.

Edenhofer O, Pichs-Madruga R, Sokona Y et al. (2014) Technical Summary. In: Edenhofer O, Pichs-Madruga R, Sokona Y et al. *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge: Cambridge University Press, pp. 33-107.

Edwards P (1996) Global comprehensive models in politics and policymaking. *Climatic Change* 32: 149-161.

Edwards P, Mayernik MS, Batcheller AL, Bowker GC and Borgman CL (2011) Science friction: data, metadata, and collaboration. *Social Studies of Science* 41(5): 667-690.

Fieldnotes (2015a) Our Common Future Under Climate Change Conference, Paris, July 2015.

Fieldnotes (2015b) Eighth annual meeting of the IAMC, Potsdam, November 2015.

Godard O (2001) L'expertise économique du changement climatique planétaire – 1. Modèles d'organisation de l'expertise. - 2. Sur la scène internationale, le GIEC. - 3. Sur la scène française, autour du Commissariat général du Plan, Paris: Annales des Mines, Série Responsabilité et environnement (21), pp. 23-65.

Guillemot H (2007) Les modèles numériques du climat. In: Dahan-Dalmedico A (ed) *Les modèles du futur. Changement climatique et scénarios économiques: enjeux scientifiques et politiques*, Paris: La Découverte, pp. 91-112.

Haas P (1992) Introduction: Epistemic communities and international policy coordination. *International Organization* 46(1): 1-35.

Hourcade JC, Jaccard M, Bataille C and Ghersi F (2006) Hybrid modeling: New answers to old challenges. *The Energy Journal* 2:1–12.

Hughes H R and Paterson M (2017) Narrowing the climate field: The symbolic power of Authors in the IPCC's Assessment of Mitigation. *Review of Policy Research* 34(6):744-766.

Hulme M and Mahony, R. (2010) Climate change: What do we know about the IPCC?, Progress in Physical Geography: Earth and Environment, 34(5): 705-718.

IAMC (2012) Request for Long-Term, Integrated Scenarios in Support of the IPCC's 5th Assessment Report, Working Group III.

IAMC (undated) Integrated Assessment Modeling Consortium homepage. Available at: http://www.global-change.umd.edu/iamc/ (accessed 14.06.2017)

IPCC (2001) Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change [Houghton JT, Ding Y, Griggs DJ et al. (eds)]. Cambridge: Cambridge University Press..

IPCC (2006) *Recommendations on New Emission Scenarios*, final report, IPCC-XXV/Doc. 11

IPCC (2007) Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon S, Qin D, Manning M et al. (eds)]. Cambridge: Cambridge University Press.

IPCC (2013) *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Stocker TF, Qin D, Plattner GK et al. (eds)]. Cambridge: Cambridge University Press..

IPCC (2014a). *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Edenhofer O, Pichs-Madruga R, Sokona Y et al. (eds)]. Cambridge: Cambridge University Press.

IPCC (2014b) WG III AR5 Scenario Database. Available at: https://tntcat.iiasa.ac.at/AR5DB (accessed 14.06.2017).

IPCC (undated) Data Distribution Center. Available at: http://sedac.ciesin.columbia.edu/ddc/ar5_scenario_process/scenario_background.html (accessed 14.06.2017).

Knorr-Cetina K (1982) Scientific communities or transepistemic arenas of research? A critique of quasi-economic models of science. *Social Studies of Science* 12(1): 101-130.

Knorr-Cetina K (1999) *Epistemic Cultures: How the sciences make knowledge*. Cambridge, MA: Harvard University Press.

Kriegler E, O'Neill BC, Hallegatte S et al. (2012) The need for and use of socio-economic scenarios for climate change analysis: a new approach based on shared socio-economic pathways. Global Environmental Change 22: 807–822.

Lefevre J (2016) Hybridization challenges in energy-economy integrated models and representation of the low carbon transition: an application to the Brazilian case, Thesis, AgroParisTech, Paris, France.

Leggett J, Peppe WJ, Swart RJ et al. (1992) Emissions Scenarios for IPCC: An Update. In: Houghton JT, Callander BA and Varney SK (eds) *Climate Change 1992. The Supplementary Report to the IPCC Scientific Assessment*. Cambridge: Cambridge University Press, pp. 69-96.

Leonelli S and Ankeny RA (2015) Repertoires: How to transform a project into a research community. *BioScience* 65(7): 701-708.

Lorenz-Meyer D (2010) Possibilities of enacting and researching epistemic communities. *Sociological Research Online* 15(2):13.

Lövbrand E (2011) Co-producing European climate science and policy: a cautionary note on the making of useful knowledge. *Science and Public Policy* 38(3): 225-236.

Meyer M and Molyneux-Hodgson S (2010) Introduction: The dynamics of epistemic communities. *Sociological Research Online* 15(2).

Merz M and Sormani P (2016) *The local configuration of new research fields: on regional and national diversity. Sociology of the Sciences Yearbook* Volume 29. Heidelberg: Springer.

Molyneux-Hodgson S and Meyer M (2009) Tales of emergence: Synthetic biology as a scientific community in the making. *Biosocieties* 4(2-3): 129-145.

Moss R, Babiker M, Brinkman S et al. (2008) *Towards New Scenarios for Analysis of Emissions, Climate Change, Impacts, and Response Strategies.* Intergovernmental Panel on Climate Change, Geneva

Moss RH, Edmonds JA, Hibbard KA et al. (2010) The next generation of scenarios for climate change research and assessment. *Nature* 463(7282):747–756.

Nakicenovic N, Alcamo J, Davis G et al. (2000) *Special Report on Emissions Scenarios*, Cambridge: Cambridge University Press.

O'Neill BC and Nakicenovic N (2008) Learning from global emissions scenarios. *Environmental Research Letters* 3(4): 045014.

Parson E, Burkett V, Fisher-Vanden K et al. (2007) Global-change scenarios: their development and use. US Department of Energy Publications.

Risbey JM, Kandlikar M and Patwardhan A (1996) Assessing Integrated Assessments. *Climatic Change* 34: 369-395.

Schwanitz VJ (2013) Evaluating integrated assessment models of global climate change. *Environmental Modelling & Software* 50:120–131.

Shackley S and Wynne B (1995a) Global climate change: the mutual construction of an emergent science-policy domain. *Science and Public Policy* 22(4): 218-230.

Shackley S and Wynne B (1995b) Integrating knowledges for climate change: Pyramids, nets and uncertainties. *Global Environmental Change* 5(2): 113-126.

Shackley S, Risbey J, Stone P and Wynne B (1999) Adjusting to policy expectations in climate change modeling. An interdisciplinary study of flux adjustments in coupled atmosphere-ocean general circulation models. *Climate Change* 43: 413-454.

Sundberg M (2010) Organizing simulation code collectives. *Science Studies* 23(1):37-57

Toth FL (1995) Practice and progress in integrated assessments of climate change: A workshop overview. *Energy Policy* 23(4-5): 253-267.

Vardy M, Oppenheimer M, Dubash NK, O'Reilly J and Jamieson D (2017) The Intergovernmental Panel on Climate Change: challenges and opportunities. *Annual Review of Environment and Resources* 42:55-75.

van Vuuren DP and O'Neill BC (2006) The Consistency of IPCC's SRES Scenarios to 1990–2000 Trends and Recent Projections. *Climatic Change* 75(1-2): 9-46.

van Vuuren DP, Feddema J, Lamarque J-Fr et al. (2008) *Work plan for data exchange between the Integrated Assessment and Climate 1 Modeling community in support of Phase-0 of scenario analysis for climate change 2 assessment (Representative Community Pathways)*. International Panel on Climate Change (IPCC). Available at https://cmip.llnl.gov/cmip5/docs/RCP_handshake.pdf (accessed 10.02.2018).

van Vuuren DP, Riahi K, Moss RH et al. (2012) A proposal for a new scenario framework to support research and assessment in different climate research communities. *Global Environmental Change* 22(1): 21-35.

Webster M, Forest C, Reilly J et al. (2003) Uncertainty Analysis of Climate Change and Policy Response. *Climatic Change* 61: 295–320

Weyant J, Davidson, Dowlatabadi H et al. (1996) Integrated Assessment of Climate Change: An Overview and Comparison of Approaches and Results, In: Bruce JP, Lee H and Haites E (eds) *Climate Change 1995: Economic and Social Dimensions of Climate Change*. Cambridge: Cambridge University Press, pp. 374-380.

Weszkalnys G and Barry A (2013) Multiple Environments: accountability, integration and ontology. In: Barry A, Born G (eds) *Interdisciplinarity: reconfiguration of the social and natural sciences*, London: Routledge, pp. please add page range.

Wilson C, Kriegler E, van Vuuren DP et al. (2017) Evaluating Process-based integrated assessment models of climate change mitigation. IIASA Working Paper. IIASA, Laxenburg, Austria, WP-17-007.

Wynne B (1984) The institutional context of science, models, and policy: the IIASA energy study. *Policy Sciences* 17: 277-320.

Notes

1 Emission scenarios project the evolution of greenhouse gases emissions and the resulting atmospheric concentrations on the basis of socioeconomic hypotheses. Socioeconomic narratives are coherent sets of assumptions about the evolution of key socioeconomic variables (such as world population or GDP). The distinction is not always clear-cut: for instance, the SRES scenarios perform both functions (Nakicenovic et al., 2000). As the emission scenarios we consider all rely on socioeconomic assumptions, we refer to them as "socioeconomic emission scenarios", while acknowledging that the status and origins of the socioeconomic assumptions underpinning them may vary.

2 We interviewed researchers from COPPE/UFRJ, NIES, PBL, PIK, EMF, FEEM, and the IPCC WG III Technical Support Unit between 2015 and 2017 (Appendix 2). Observation took place during the Conference "Our Common Future Under Climate Change", Paris, July 2015 and at the Eighth meeting of the IAMC, Potsdam, November 2015.

3 According to Hughes and Paterson (2017), analyses of the IPCC in terms of epistemic community tend to emphasise the need for a separation between scientific production and political action, whereas those viewing the IPCC as a boundary organisation stress the interrelations between science and politics.

4 Special reports provide assessments of a specific issue related to climate science or policy. They generally follow the same structure as a volume of an Assessment Report.

5 Further integration between RCPs, climate model results and IAMs failed, as the new socioeconomic narratives, the so-called Shared socio-economic pathways (SSPs), were not ready on time (Kriegler et al., 2012; Moss et al., 2010; van Vuuren et al., 2012).

6 New emissions scenarios for the IPCC process were discussed in workshops in Washington (January 2005), Laxenburg (July 2005) and Seville (March 2006).

7 RICE (Regionally Integrated Climate-Economics) is an economics-based model initially developed by Nordhaus in the 1990s.

8 As is the case for the IPCC more broadly, IAM research mostly takes place in developed countries. While there are IAM teams and WG III authors from developing countries, they usually have strong links with institutions based in developed countries (e.g. having spent time there or using models based on those of developed countries teams) (Corbera et al., 2015; Vardy et al., 2017)

9 However, those in charge of a specific Scientific Working Groups are often in charge of the same issue within projects. For instance, one of the co-chairs of the "data protocol and management working group" oversees the database infrastructure at IIASA; and two of the co-chairs of the "scenario working group" are representatives of the IAM Community in the "Scenario-MIP project".

10 https://cmip.llnl.gov/index.html [accessed 31/01/2018]; on collective practices in climate science, see Guillemot (2007), Sundberg (2010) and Edwards et al (2011).

Appendix 1 - Acronyms

AR: Assessment Report
AR4: Fourth Assessment Report
AR5: Fifth Assessment Report
EMF: Energy Modeling Forum
EPA: Environmental Protection Agency
FEEM: Fondazione Eni Enrico Mattei
GCM: General Circulation Models
IAM: Integrated Assessment Models
IAMC: Integrated Assessment Modeling Consortium
IIASA: International institute for Applied System Analysis
NIES: National Institute for Environmental Studies
PIK: Potsdam Institute for Climate
PNNL: Pacific Northwest National Laboratory
RCP: Representative Concentration Pathway
RITE: Research Institute of Innovative Technology of the Earth
RIVM: National Institute of Public Health and the Environment
SRES: Special Report on Emissions Scenarios
TAR: Third Assessment Report

Appendix 2 – List of interviews

1	IPCC author, background in engineering and economics
2	Modeller, IPCC author, member of IAMC Scientific Committee, background in applied mathematics and physics
3	Modeller (land-use), IPCC contributor, background in biology
4	Modeller, IPCC author, member of IAMC Scientific Steering Committee, background in physics
5	Former member of the IPCC Working Group III Technical Support Unit
6	Emeritus professor, former modeller, IPCC author, background in chemistry
7	Modeller, leader of an IAM team, background in chemistry and environmental science
8	Modeller (climate policy), background in economics and earth and life sciences
9	Junior modeller (climate and energy policies), background in climate studies
10	Modeller, background in economics
11	Coordinator of EMF, IPCC author, member of IAMC Scientific Steering Committee, background in engineering
12a	Modeller, IPCC author, background in environmental economics and operations research
12b	Modeller, IPCC author, member of IAMC Scientific Steering Committee, background in engineering and economics
13	Modeller, background in applied mathematics and economics
14	Modeller (economics), background in economics
15	Climate scientist

Magicians at Work: Modelers as Institutional Entrepreneurs in the Global Governance of Agriculture and Food Security

Lise Cornilleau

Laboratoire Interdisciplinaire Sciences Innovations Sociétés, Paris-Est University & Centre de Sociologie des Organisations, Sciences Po, France/ lise.cornilleau@sciencespo.fr

Abstract

Global models of agriculture act as the epistemic basis for quantitative foresight, which guides international policymaking and research on agriculture. With the new political sociology of science as a backdrop, this article studies the actors who develop and use these models through the lens of field theory. Contributing to the dialogue between the neo-institutionalist field theory and its Bourdieusian version, it describes the structure and the dynamics of the strategic action field of modelling organizations, using the Bourdieusian notions of 'succession' and 'subversion' to refine the characterization of challengers. It also discusses the insights of the Bourdieusian concept of 'homology' to analyse the relations between the field of model producers and the field of model users. Whereas Bourdieu provides a primarily descriptive account of homologies, which are close to a "social magic without magicians" for Roueff, the present text describes magicians doing the work of producing homologies. Some modelers use intercomparison to reduce competition and to have their models used in the field of global governance, thus strategically producing homologies, while resolving the main modelling conflict of the field. These actors benefit from the recent evolution of the modelling field under the influence of climate change, to behave as what Fligstein and McAdam have called 'institutional entrepreneurs'. The article concludes that field theory makes it possible to describe the co-construction of a range of models developed by competing organizations and the controversial making of global agricultural governance. Doing so, it complements the co-production framework, which often focuses on a given site of expertise production and a site of global governance.

Keywords: field theory, global modelling, intercomparison, agriculture, climate change

Introduction

The 2008 economic and financial crisis shook the agricultural sector. A sudden rise in global agricultural prices (mainly grains) and massive acquisition of farmland by foreign investors hit several developing countries, and fanned the flames of political disputes in the so-called Arab Spring. This crisis reminded us of the vulnerability of agro-food systems that have become globalized.

Climate change and the financialization of agricultural markets creates concerns that resource pressures, and resulting episodes of agricultural and food price volatility, would become more frequent. In this troubled context, economic models of world agriculture are the primary knowledge tools that decision-makers use to reflect on the future of agriculture and to arbitrate between policy options at an international scale. The accuracy of these models is under scrutiny by experts from the Food and Agriculture Organizations of the United Nations (FAO), the World Bank, the International Food Policy Research Institute (IFPRI) – a research institute in economics part of the Consultative Group in International Agricultural Research (CGIAR) – and more broadly within the academic field of agricultural economics, dominated by American, European, and Chinese universities. The 2008 crisis questions the way these models are elaborated and evaluated, and by whom. Yet, if the role of models in the global governance of agriculture is crucial, they have received less focus than in the cases of finance (McKenzie, 2006) or climate change (Dahan, 2007).

The co-production framework (Jasanoff, 2004) helps to understand the politics of models. Models are 'mediators' between theory and the real system (Morgan and Morrison, 1999). They stabilize a state of the art on a given topic in order to build prospective scenarios, which are increasingly used by decision-makers in a normative fashion. Based on a specific set of knowledge and assumptions, they contribute to 'perform' public policies in a specific way (Armatte, 2010; Henriksen, 2013). Previous work has shown how models of world agriculture based in neoclassical theory lay the groundwork for scenarios which offer a limited set of options for the future. As small-scale agriculture is 'invisibilized' in such models (Leblond and Trottier, 2016), the scenarios they enable to develop tend to promote technology-intensive agriculture over agroecology, and free-trade over food sovereignty (Fouilleux, 2010; Cornilleau, 2016). Hence, economic models represent certain 'worlds' that they help perform through scenario building. But why are certain models deemed more legitimate than others? What are the mechanisms of the competition between different models of a given problem, such as world agriculture?

In line with the new political sociology of science, I seek to complement the co-production framework by stressing the "unequal distributions of power and resources" (Frickel and Moore, 2006: 10) in modelling activity. The co-production perspective, by paying symmetrical attention to the "dominant point of view" and the "marginalized alternatives" (Jasanoff, 2004: 280), has already worked in this direction. Research inspired by Desrosières (2008) has analysed the production and use of alternative statistics, using the concept of 'statactivism' (Bruno et al., 2014; Kurunmäki et al, 2016). An actor-network theory approach (Callon, 1986) has highlighted the drivers of the success of a model, defined as its ability to enrol users at an early stage of scenario development (Kieken, 2004). These studies focus respectively on the attempts to create alternative quantifications (i.e. the "marginalized alternatives"), or on the modalities of a success (i.e. the "dominant point of view"). In this article, I use a complementary perspective to document inequalities in modelling activity. My goal is to map the field of modelling organizations (research institutes, international organizations, ministries of agriculture), and to enlighten how world agricultural models are being hierarchized. Inspired by previous research on metrics (Paradeise and Filliatreau, 2016), I use neo-institutionalist field theory (Fligstein and McAdam, 2012) while borrowing some concepts to Bourdieu (2015). If these perspectives rely on different hypotheses (Martin, 2003), the former was inspired by the latter. In both cases, a field is defined as a socially structured space of positions, organized through a struggle over what is specifically at stake in the field.

As the results of the modelled foresight are intended to circulate among experts and decision-makers, competition between models cannot be analysed only at the level of modelling organizations themselves. Models are part of the toolbox that States and stakeholders (companies, NGOs, modelling organizations themselves) use to influence the making of the global governance of agriculture. The introduction of 'good practices' (Bernstein and Van der Ven, 2017) encourages stakeholders to use quantified indicators in advocacy and lobbying strategies, which explains the increased competition between indicators of world problems (Rottenburg et al., 2015), as in the case of women's rights (Merry, 2016). Field theory accounts both for the dynamics in a given social space, and for the interactions with other social spheres: the modelling field can be analysed *per se*, and in interaction with the academic field or the field of global governance. This is different from the ecology of knowledge (Akera, 2007), which describes how forms of knowledge co-produce institutional hierarchies, yet does not address the dynamics *within* science and *within* society. Field theory is also more appropriate to analyse the competition between models than a market-based approach, which has been used on standards (Reinecke et al., 2012; Fouilleux, Loconto, 2017). World agricultural models interact with economic stakes[1], but they are mostly evaluated in scientific (they are deemed credible) or political (they are deemed legitimate) terms. A last advantage of field theory is its potential to illuminate the international dimension of the modelling field. Sociologists (Bigo, 2011, Go and Krause, 2016) stress that the internationalization of a field is non-linear, and that interactions with national stakes are decisive.

A field exists only if the space of interactions structures actors' behaviour to such an extent that it becomes a relevant level for analysis. Semi-structured interviews with modelers showed that they situated their models in relation to others, and their representation of this space shaped the way they evaluated their own tools. Interviews gave insight into the topologies of the field: actors define the boundaries and the structure of the social space they moved within, and they describe their efforts to improve their position within it. In order to get a sense of what is at stake at the international level of this field, I conducted interviews both with representatives of national (French) and international modelling organizations. To map the positions of each modelling organization, and to track the circulation of models in governance and expertise, I analysed reports and minutes from an international research group – the Global Economics Team of the Agricultural Models Intercomparison and Improvement Project (AgMIP). This gave me access to debates amongst modelers about the evaluation and expected uses of their tools. The Global Economics Team of the AgMIP initially (2011-2013) included modelers from international organizations and US, Australian, Japanese, German, and Dutch research organizations, with ten models represented[2], while the French modelling team joined the exercise in its second phase.

My argument has a three-part structure. In the first section, I show that the field of global agricultural models is not autonomous: its evolution is in sync with the 'climatisation' (Aykut et al.., 2017) of agricultural policies and research, i.e. the political demand that they take climate change into account. I describe how this 'climatisation' incites newcomers to enter the modelling field and how they tried to challenge the incumbent models. In the second section, the homologies between the field of model producers and the field of model users (including modelling organizations themselves) are highlighted: using a credible and legitimate model has direct effects on the position that institutions hold in the field of global governance. Last, I seek to explain the mechanisms standing behind these homologies, which are close to a "social magic without magicians" in Bourdieu's theory (Roueff, 2013). However, there are magicians here, and they play a central role. This last section shows these magicians at work: modelers who benefit from the recent evolution of the field and behave as 'institutional entrepreneurs' (Fligstein and McAdam, 2012). They use intercomparison within the AgMIP to reduce competition between models and to facilitate the circulation of their own models in the political field, thus producing homologies between the two fields, while solving the main conflict in the modelling field. I conclude that this amended field theory is a complement to the co-production framework, which has analysed jointly a given site of expertise and a site of global governance.

The 'climatisation' of the field of global agricultural models

This section presents recent shifts in the field of global agricultural models. It has changed under the imperative to integrate climate change. This 'climatisation' of the modelling field went hand in hand with a change in its structure. A conflict emerged between two different traditions: the equilibrium models' tradition focused on the economy was challenged by the integrated models' tradition, which better represented the links between the economy and climate change.

From trade to climate: Models and the international agenda of agriculture

For Bourdieu, as well as for Fligstein and McAdam, a field is only partially autonomous from other fields, and inter-field relationships influence the structure and dynamics of a given field. Fligstein and McAdam (2012) think of inter-field relationships in terms of their embeddedness within each other[3], from the macro level (e. g. the global governance field) to the micro level (e. g. an office in the department of a State), and compare fields to 'Russian dolls'[4]. The modelling field is not autonomous from the global governance field, and can be considered as one of its subfields. Global agricultural models have changed in accordance with shifts in international agricultural policies; they initially focused on the trade agenda and on the preparation of a 'Doubly Green Revolution' (Cornilleau and Joly, 2014), only to change their orientation when climate change became a priority.

The trade agenda determined the first shift in economic agricultural models. In the 1970s, two categories of economic models coexisted: (i) time series models, which describe the dynamics of physical aggregates over time, whereas the (ii) first *equilibrium models*, which compute prices by balancing global supply and demand, appeared in the context of the 1974 world food crisis (Cornilleau, 2016). At the end of the 1970s, economists from agro-exporter countries (mostly from the United States of America (USA)), from the Organization for Economic Cooperation and Development (OECD) and from the IFPRI questioned the ability of time series models to simulate the impacts of trade policies on domestic agricultural prices. The reason for this was that agricultural trade liberalization was under examination in these countries and organizations, before the first steps of its implementation were negotiated at the Uruguay Round (1986-1994) of the General Agreement on Tariffs and Trade (GATT). Equilibrium models[5] ended up mediating controversies on trade, notably on the level of agricultural subsidies in the USA versus in the European Economic Community in the 1980s (Fouilleux, 2000). American universities and research institutes played a leading role in the development of these equilibrium models. The US Department of Agriculture and the Ford Foundation funded the International Agricultural Trade Research Center (IATRC), a think tank whose goal was to equip the OECD and the GATT negotiations with the economic toolkit that delegitimized subsidies and led to trade liberalization in agriculture (Joly and Lacombe, 2017). The US also supported networks of modelers working on the development of an international agricultural database, such as the Global Trade Analysis Project (GTAP) created in 1991 at the Purdue University in interaction with the FAO and the OECD (Leblond and Trottier, 2016; Dorin and Joly, 2019), which is still widely used today. In the 1990s, equilibrium models were developed by other institutions for different purposes, such as the IFPRI with the International Model for Policy Analysis of Agricultural Commodities and Trade (IMPACT) in 1995. This model became a flagship for the IFPRI, which supported foresight aiming at funding international agricultural research around the project of a "doubly Green Revolution", thanks to its format emphasizing the promises of technologies (Cornilleau, 2016). This brief genealogy[6] shows how economic models are both tools of proof and tools of power to defend national, or institutional, interests in the making of international agricultural policies.

The field of global agricultural models underwent another major change, which is at the heart of this article. Agricultural models have received renewed attention since the 4th report of the Intergovernmental Panel on Climate Change (IPCC) concluded in 2007 that agriculture is both a major driver of climate change and an area particularly affected by it, with worrying impacts on food security. Decision-makers now expect models to allow them to reflect on the interactions between agriculture, food, energy, and the environment. The 'climatisation' of the agricultural agenda had a counterpart in the 'climatisation' of the modelling field, which is twofold. First, equilibrium models seek to better represent the environment and climate change by relying on agroecological zones instead of political divides (such as countries or regions) and they tend to adopt a modular structure (what enables them to add modules representing the environment). Second, *integrated models*, another modelling tradition[7] grounded in the Club of Rome perspective focused on the interactions between human activities and the environment, were increasingly seen as more credible to represent the environment than equilibrium models, which conceive of agriculture as an industrial sector. This was all the more the case that in the 2000s, integrated models sought to provide more details on the equilibrium of the economy, hence undermining equilibrium models' added value. These evolutions[8] show that the boundaries and the structure of the field evolved after the 'climatisation' of agriculture.

Incumbents and challengers in the 'climatised' modelling field

The structure of the modelling field is defined by the relative positions of modelling organizations according to the 'capital' (Bourdieu, 2015) of their model. Each model positioned in the field draws its properties from other models, and these relations are conveyed in the formalism of models. There is a consensus on the capital specific of the field, on "what is at stake", and on the 'rules' of the field (Fligstein and McAdam, 2012: 10-11), i.e. modelers agree on a broad definition of what a good model is, and on the ways for a model to improve its value and its position in the field.

Part of this consensus is a recognition of the shortcomings inherent to these models. Modelers recognize that the existing theories are inadequate to describe the complexity of the agricultural and food system at a global scale. They nonetheless have to translate them into models *via* a set of equations linking the variables that determine global agricultural production (population, eating habits, urbanization, etc.) and global food consumption (yield or production estimates, price elasticities[9], etc.). Once a representation of agriculture established, data are not always available[10], and are often of poor quality[11]. Consequently, modelers evaluate models through two channels: (i) the reputation of the model in the academic field, assessed through peer review, (ii) the model's ability to help decision making through quantified scenarios. The modelling field interacts closely with the academic field and the political field, and the capital specific to the modelling field depends on the amount of capital obtained in these two fields: a good model is a model which is deemed both credible and legitimate. Scenario building is the main source of revenue: dominant modelling teams reinvest the profits they make with scenarios in improving their model (by adding a new module for instance), which then helps them find new clients, and thus improve their position in the field. This creates a hierarchy, with models "which received a seal of approval" and "dominate whatever the subject is" (in a modeler's terms) above the others.

The following subsection describes the dynamics of the field since its 'climatisation'. First, integrated models tried to benefit from the new framework to improve their position and to challenge the 'incumbents' of the field, i.e. equilibrium models (Fligstein and McAdam, 2012). Then, other challengers use 'subversion' or 'succession' strategies (Bourdieu, 2015).

Integrated models challenge equilibrium models

Two models were characterized by interviewees as dominant, stemming from different modelling traditions (equilibrium models vs. integrated models): the IFPRI's IMPACT model, and the Global Biosphere Management Model (GLOBIOM) developed since 2008 by the International Institute for Applied System Analysis (IIASA). The IIASA is a research institute created in 1971 to conduct research on 'world problems' in the context of the Cold War, which became a major research institute in the modelling of climate change (Dahan, 2007). Although they come from different traditions, they both changed to be more relevant to the 'climatized' agenda and now have a similar modular form and both propose a spatialization of their results. IMPACT transformed its partial equilibrium format into a modular structure, and uses agroecological zones. As for GLOBIOM, it relates human activities (food, fiber, energy, industry) with both environmental and economic equilibriums. IMPACT has long dominated the field and still benefits from its "return on investment", in a modeler's terms. It has become a reference tool for providing foresight on food security and malnutrition. However, it has been deemed less relevant than GLOBIOM to provide scenarios on the interactions between agriculture, environment, and energy, as it uses a database representing both forests and farming. According to interviewees, IMPACT would be threatened if GLOBIOM were able to improve the way it represents food security:

> **– You were saying that GLOBIOM is the most dominant model?**
> […] It is a dominant model in terms of publications, and in terms of the issues dealt with: […] they've got farming, they've got a very good forestry model, etc. […] GLOBIOM is heavily requested by the European Commission and has been chosen to evaluate everything that has to do with biofuels issues. […] [But] IMPACT is always the model used to evaluate food security. So GLOBIOM on the issues of food security has not reached that supremacy yet, but given what and how much they publish … The farming and climate variability issues are blind spots for food security issues and IMPACT does not have that at all. […] So in a couple of years, GLOBIOM will certainly strongly compete with IMPACT on those issues.

The 'climatisation' of the field has helped challengers compete with incumbents, and GLOBIOM has improved its position in the field, accumulating both scientific and political capital quickly.

Subversion and succession attempts

Not all models are as successful as GLOBIOM when entering the 'climatised' field. I use Bourdieu's (2015) concepts of 'subversion' and 'succession' to describe their strategies. Some new entrants try to challenge the dominant actors by offering a slightly different definition of what is at stake in the field, i.e. a subversion strategy. For example, French modelling organizations put more emphasis on the analytical power of models and their heuristic use. They see foresight as a way "to construct desired futures and test their consistency and viability" with the help of models, which is typical of the French foresight culture (Dorin and Joly, 2019). Expert knowledge is more important than the model itself, what contrasts with IFPRI's foresight practices, as a French modeler explains:

> The idea [with models] is to carry out a "plausibility test" of the scenarios. Basically we do scenarios, then we look at whether they are coherent, whether it works or not. This is something that is done differently from what IFPRI does, where the IMPACT model is really at the centre of the work. Here, let's say, the centre of the work is really all that is done with the experts [who build scenarios] and then we look, we test if the coherence works [with the model] and that's it. IFPRI's scenarios are based more on the results of these models: [...] price evolution, need for investment, etc. is calculated by the IMPACT model, so it is really an approach that is completely different.

This approach is grounded in a definition of economics as a social science, whose predictions could therefore only be considered as insights. Even if decision-makers tend to see modelled scenarios as a 'crystal ball' in the terms of a modeler, French modelers consider that this should be avoided, for instance though a participatory and transparent approach to foresight. This vision is supported by a different model form than dominant models, as in the case of the French model Globagri, which was used in integrated modelling (the Nexus Land-Use (NLU) model) and in scenario building (the Agrimonde foresight). Globagri focuses on world physical food biomass balances (resources, trade, uses), without considering prices. NLU, developed by a French research centre, provides scenarios such as reduction in meat consumption or a balance between undernutrition in the South and over-nutrition in the North, which are not possible with other models. Its simplicity renders NLU transparent for users, and it has easily modifiable parameters: its developers argue that it enables debate with non-specialists. French research centres also used Agribiom (Joly and Dorin, 2019), then Globagri, to develop the Agrimonde foresight, which explores a certain definition of agroecology: (i) change in diet, (ii) reducing food and agricultural waste, (iii) favouring biodiversity and ecological intensification.

As a model's legitimacy depends on its links with other legitimate models (Cornilleau, 2016; Leblond and Trottier, 2016), other new entrants adopt a 'succession' strategy: they try to copy the form used by incumbents or to use some of their data. They also agree with incumbents' definition of the capital of the field: they consider modelling as a tool for evidence-based policy-making. These strategies face two difficulties. First, accessing the data, the code, or the modules developed by a dominant model is not easy: they are kept secret as they provide the dominant model with a temporary monopoly on certain topics, and the associated revenues in terms of scenario building. This explains for instance the difficulties faced by a team in their collaboration with IIASA, the goal being to access some of GLOBIOM's modules, as a modeler relates:

> They have a strategy, aimed at GLOBIOM. […] So, well, they try to collaborate. [Sarcastic] They send people who never come back. They try to retrieve bits. Now, they have succeeded, but completely indirectly thanks to [an international expert in modelling]. He succeeded in obtaining data from […] GLOBIOM livestock modules. So in the end, it ended up at [the team] via a convoluted path.

Another challenge of 'succession' strategies is to identify the added value of a new model as compared to incumbents, in order to generate orders for scenarios that would make the model profitable. The FAO has recruited a well-known modeler in the sub-field of equilibrium models to develop a model, with the hope that it would challenge the supremacy of the IFPRI's model. This modeler explained that the purpose of this project was for FAO to obtain a tool that reflected its "own view":

> **– So the FAO wanted a model to compete with IFPRI's model?**
> That's a very good question; there is a lot of discussion about that. I'm not sure we've resolved it yet, but it's possible that these efforts will eventually merge with the IFPRI's [as they] […] are very similar. […] But we've had problems in the past. There are concerns in the FAO that it needs to represent our own view, that it's our scenarios, and that we are not just taking IFPRI scenarios.

Yet the FAO foresight team had little human and financial resources, so the profitability of its new model was all the more pressing. This hostile context made a succession strategy difficult, and the FAO finally decided to merge its project with the IFPRI's model. The FAO is now compelled to commission IFPRI (as the previous team using 'succession' does with IIASA) to get scenarios: they are not allowed to develop their "own view" through models, but have outsourced this research.

In this first section, I developed an overview of the dynamics of the field of global agricultural models, showing that all models are not deemed equally credible and legitimate to represent agriculture under climate change. The 'climatisation' of the field generates a competition between equilibrium models (as incumbents) and integrated models (as challengers). New entrants also attempt to replace the incumbent models *via* subversion or succession. If succession fails, subversion lays the groundwork for a definition of the stake of the field which is different from that of the incumbents, in which models are as much analytical tools for imagining the future as they are evidence-based policy instruments.

Homologies between the modelling field and the field of global governance

In what follows, I describe how the field of model producers is related to the field of model users. These users are the modelling organizations themselves, but also the stakeholders of the global governance of agriculture (States, NGOs, companies, etc.) which is fragmented between international organizations (United Nations, FAO, World Bank, World Trade Organization (WTO), etc.). I define global governance as "complex, dense, and multidirectional networks" in which governments are influenced by international organizations, research institutes, NGOs, and companies (Djelic and Sahlin-Andersson, 2006). This governance increasingly relies on the use of models: through new institutions, such as science-policy interfaces inspired by the IPCC (Miller, 2007; Haas, 2017), and through new practices, such as the quantification of the performance of international policies (Bezes et al., 2016). In this section, I introduce the concept of 'homology' (Bourdieu, 2015) to account for certain inter-field relations which cannot be captured by Fligstein & McAdam's 'russian dolls' metaphor. This concept refers to structural parallels between fields because of their relative autonomy, for example between the field of art producers and the field of art consumers, or between the field of higher education and the field of power (Bourdieu, 2015). In this line, I show that there is a homology between the field of model producers and the field of model users: users of legitimate models are more likely to have powerful positions in the field of global governance than the users of dominated models, what I show through two examples.

First, I investigate the competition between models as homologous with the competition between participants in a science-policy interface launched in 2002 by the World Bank and the FAO: the International Assessment of Agricultural Science and Technology (IAASTD), which has an intergovernmental and multi-stakeholder format. Then, I consider these homologies through the use of models as indicators for international policy-making, through the cases of the United Nations Sustainable Development Goals (SDGs) and the design and evaluation of agricultural policies for Africa by international organizations. Previous research on science-policy interfaces has emphasized the 'mutual construction' of a modelling type and the associated political field (Shackley and Wynne, 1995). What I show here, however, is the interest of describing the struggles between different models *within* science-policy interfaces or evidence-based policy making. The concept of homology allows to look at the co-construction of a plurality of competing models with a global governance, which is both complex and multi-sited and riddled with tensions between diplomacies and other stakeholders.

The politics of modelling in science-policy interfaces: No model, no voice

To give a sense of how these homologies are revealed in science-policy interfaces, I consider the role of the IFPRI's IMPACT model in the IAASTD process, and how it laid the groundwork for the French Agribiom model described in the previous section. The IAASTD was initially asked to reflect on the possible futures for agriculture on the basis of scenarios created with IMPACT. Doing so, the IASSTD Advisory Committee capitalised on the excellent academic reputation of the IMPACT model and also followed science-policy interfaces, such as the Millenium Ecological Assessment (MEA) and the IPCC, which used IMPACT to represent world agriculture. Yet the multi-stakeholder format of the IAASTD led to a debate on scenario making, and several authors disagreed with the weight given to economic models and/or with the choice of this specific model. Resistance was encountered from certain NGOs represented in the IAASTD, who accused models of being an elitist tool. They argued that models are hardly transparent for those without economics training and that they therefore do not facilitate inclusive deliberations on the future of agriculture (Scoones, 2009). Being based on neo-classical economics, the IMPACT model was also rejected by economists from other traditions. A participant recounts the criticism that a modeler from the IFPRI received during his presentation of the first version of IMPACT-based scenarios:

> So they began with a General Assembly […] and there were reactions [from authors] in the room, they said: 'no, but wait, the parameters you choose for your models'. Models have been criticized as being econometric models by that kind of heterogeneous group, with native peoples, farmers from Zimbabwe, neo-institutionalist economic researchers. Everyone had something to say, because everyone could speak, I mean, you've just got to raise your hand and say what you want to say. So they took a beating!

In the end, the IMPACT model was not used to prepare the scenarios of the IAASTD, which became mostly qualitative. The IFPRI decried this defeat, as a failure to demonstrate the relevance of the IMPACT model. The two sponsors of the IAASTD also joined in this disappointment: the World Bank and the FAO, who develop and promote similar equilibrium models. Company representatives likewise attacked the scientific validity of this assessment, because of the absence of modelled scenarios and of their perception of an overrepresentation of the social sciences, at the expense of mainstream economics and agricultural sciences. This was the case of Syngenta; whose representative wrote an article on IAASTD in NewScientist.com. The IAASTD has been weakened by these attacks: despite its "business as usual is not an option" motto, and its recommendations to consider agroecology and food sovereignty, it had less political impact than previous assessments using incumbent models (the MEA or the IPCC).

After their participation in the IAASTD, some actors were motivated to enter the modelling field. This was the case of French research institutes, whose experts have had the impression of being left out because their institution had not developed a model. Models structured the debate to such an extent that it seemed impossible for experts to "have a say" without referring to a model, hence the French disarray:

> We had French participants in the chapters on scenarios. […] [A French author] came back completely miserable saying: "well, we can't say anything; the IFPRI has a say in it, the Indians, the Dutch, their model this, their model that, but we didn't have a say in it, I won't go back to that group, […] I won't be the one who hasn't got anything to say". It was after that that we launched Agrimonde, even though we knew we wouldn't have Agrimonde ready in time for the IAASTD, but, at least, the next time we would be in the game, we would have our say.

On this basis, French institutes decided to develop their own model (Agribiom), which would be the basis of the first Agrimonde foresight. The second version of Agrimonde, based on the Globagri model, associated NGOs (OXFAM) and social movements (The International Planning Committee for Food Sovereignty), who believed it could be an advocacy tool for agroecology. This example shows that scientific controversies on agriculture, which are debated within science-policy interfaces such as the IAASTD or the IPCC, are both reflected in and reinforced by the struggles in the field of agricultural models.

Models as indicators: Unequal access to evidence-based policy making

Models are instrumental in science-policy interfaces, as foresight is the basis of their policy recommendations. But economic models are also increasingly being used as indicators to develop evidence-based policies and to quantify the performance of policy-making, as it has been shown for example in healthcare (Sjögren and Helgesson, 2007). Agriculture is no exception, especially at the international level. International organizations involved in the global governance of agriculture claim to develop evidence-based public policies, i.e. using data to target measurable objectives, define policy options, and evaluate public policies. Although United Nations (UN) agencies have always used data to demonstrate the effectiveness of their programmes to convince donors, this trend has increased since the launch of the Millennium Development Goals (MDGs) in 2000. These 'goals' are used by international organizations to quantify the performance of development policies, to harmonize their actions through a common framework, and to make funding allocation decisions. The MDGs have been seen as evidence of a shift toward a "new public management" of the UN (Bezes et al., 2016). Previously, since the late 1980s, the World Bank, which has always invested heavily in evaluation (Goldman, 2005), has recommended using economic equilibrium models as instruments for *ex ante* or *ex post* evaluation in various areas, including agriculture and food security (Dervis et al., 1989).

The IMPACT model was used to assess progress toward the Millennium Development Goal of reducing hunger. The dynamics of the modelling field, i.e. the fact that GLOBIOM challenged IMPACT had an impact on the respective roles of their modelling organizations in making these performance indicators. IIASA invested in the new UN agenda with GLOBIOM, through the preparation of the Sustainable Development Goals (SDGs), a set of 169 indicators meant as a successor to the MDGs for the post-2015 agenda. Unlike the MDGs, the SDGs highlight the links between overconsumption in the North and poverty in the South and examine the interactions between the environment, energy, climate change, and food security (Figure 1). On these two issues, GLOBIOM was deemed better able than IMPACT to assess possible interactions and trade-offs, due to its integrated modelling tradition.

Equilibrium models close to IMPACT are also used to define goals and to evaluate the renewed framework of African agricultural policies, in the context of the Comprehensive Africa Agriculture Development Programme (CAADP), prepared by the African Union with technical support from the United Nations and the World Bank. The CAADP encourages foreign investment, through the preparation of investment projects at the country level and legislative changes (e.g. biotechnology regulations or free trade). Models intervene in the three steps of policy design within the CAADP: 1) analysis of investment projects, 2) evaluation of agricultural legislation and approval of an investment plan, and 3) evaluation of programme implementation. The instrumental role given to models in the evaluation made it so that the IFPRI, which benefited from the legitimacy of the IMPACT model, took the role that might have been expected the FAO in its work with the CAADP.

The competition between the IFPRI/IIASA in the MDGs/SDGs and the FAO/IFPRI in the CAADP show that using a dominant model is a better guarantee for an institution to shape evidence-based policy-making. This result reinforces what happens in science-policy interfaces: institutions relying on a legitimate model are more likely to be well positioned in the field of global governance. The modelling field and the field of global governance are homologous.

Magicians at work: Shaping the field through standardization

Is the alignment between the modelling field and the field of global governance simply 'magic' (Roueff, 2013)? This would be the Bourdieusian reading of the homologies that I described in the previous section, which denies any underlying intentionality. Yet fieldwork shows that some modelers seek to strategically produce homologies. This section describes these 'magicians' at work, i.e. the efforts of modelers to have their models used in the field of global governance. In this respect, I show the decisive role of an intercomparison research project, the AgMIP. This project was founded in 2008 on the impetus of crop modelers from the NASA-Goddard Institute for Space Studies at Columbia University (which still leads the project), with the support of incumbents in the field of economic models. This alliance of prominent institutions convinced prestigious American universities such as Washington University, national and international agronomic research organizations (CGIAR etc.), governments (United Kingdom, USA, European Commission), and companies (Monsanto) to provide funding and in-kind support. The AgMIP aims to connect three types of models: economic models, climate models, and newly developed global crop models, through the collaboration of three dedicated modelling teams (Global Economics, Crop, and Climate teams) (Figure 2). The economic models presented in the previous sections are part of the Global Economics team. What is at stake within the AgMIP is to propose scenarios on global food security which can localize the impacts of climate change on crops and calculate the related economic risks. Through these scenarios, investors and policymakers can prepare themselves for such futures[12], for example by buying "climate-smart" biological or financial technologies. Another related goal of the AgMIP is to

Figure 1: The United Nations Sustainable Development Goals. Source: http://www.un.org/sustainabledevelopment/news/communications-material/

compare the models of each type amongst themselves – including the economic models, on which I focus here – to explore the impacts of their differences on foresights, and to try to reduce their heterogeneity in order to build foresights that are as consensual as possible.

In this section, I explain first how the AgMIP was conceived and created by modelers as a decisive tool for producing homologies. Then, I show how the AgMIP, by promoting a standardization of models, shapes the modelling field according to the interests of the 'institutional entrepreneurs' (Fligstein, McAdam, 2012) who have accompanied, and benefited from, the 'climatisation' of the field: it solves the conflict IMPACT vs. GLOBIOM, and it discourages models using subversion strategies.

Producing homologies: From social skills to institutional entrepreneurship

In the neo-institutionalist perspective, all actors are skilled, yet "resources […] matter a lot, [such as] the ability to deploy money, connections to the government" (Fligstein and McAdam, 2012: 181). Interviews with modelers show that they make use of such resources such as the head of the Foresight Team of an international organization who developed equilibrium models for many international organizations (OECD, World Bank, FAO), i.e. a well-endowed team directly connected with governance. These resources helped this actor to have his models mobilized as mediators ("we were in the middle", he explains) of policy-decisions. Most revealingly, he presents the milestones of his career through a list of the international policies that he quantified and evaluated with his models, from trade negotiations to climate policies and the Millennium Development Goals:

> I was in the OECD for ten years at the beginning of my career. And one of the big studies I worked on initially […] [was] a study of OECD agricultural policies, so the different subsidies and protection measures that were in place in the late 1980s, which was at the time a very contentious issue among the high-income countries, […] the United States and the European Commission. So, we were in the middle, trying to assess what were the economic impacts of this. […] I did a lot of work on the [WTO] Uruguay Round and an assessment of the Uruguay Round. […] And then I switched to working on climate change, in the mid-1990s. […] When I was at the World Bank I did a lot of projections for the MDGs.

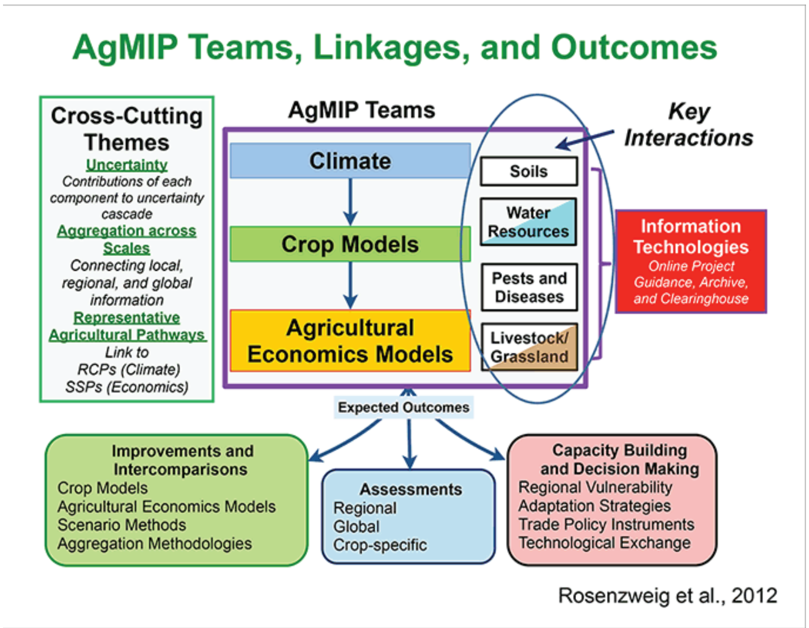

Figure 2: The collaboration between AgMIP's three teams (Agricultural Economics, Crop, Climate) and their expected outputs: improvements and intercomparisons of each type of models; assessments and decision making/capacity building. Source: www.agmip.org

Later in the interview, this actor explained that he had long wanted to create an intercomparison project of agricultural models, as he believed it would facilitate the circulation of model results in international policies and expertise. He hoped to replicate what he presented as the "success" of the intercomparison of climate models within the Energy Modelling Forum (EMF)[13] created in the 1970s during the energy crisis. Indeed, most of the research conducted within the EMF, and the resulting changes in the climate models, have informed the IPCC. Even before the AgMIP was created, this modeler imagined the benefits of a hypothetical "Agricultural Modelling Forum" based on the EMF example:

> **– Why is your Team interested in the AgMIP?**
> […] EMF started in the 1970s because of the energy crisis. So, it was mainly a gathering of energy modelers […]. When the energy crisis kind of melted away, well climate change appeared! So, EMF became basically the place where a lot of modelling of climate change occurs; so if you look at the IPCC, the 3rd volume on mitigation, almost all that work comes from the EMF modelling group. It has been incredibly successful, EMF! […] So, my hope was that we could recreate something like EMF but for agriculture, that I actually called "AMF" for "Agriculture Modelling Forum" but right now we are in AgMIP, which is fine. […]
>
> **– As the EMF was largely used in the IPCC, your objective [with AgMIP] is also to be more influential in global assessments?**
> Sure. We'll also influence the IPCC, especially volume 2, you know, on impacts, adaptation, etc., and vulnerability and… [...] we'll be cited in the 5th Assessment Report, but I think as this work progresses, you know, I think we'll have more to feed into the volume 2 report.
>
> **– Are there other reports that you will influence?**
> Yes, there are always things coming out, there will be a post-Rio agenda, you know the Beyond the MDG's [i.e. the Sustainable Development Goals]. […] I'm sure we'll be asked to provide some assessment of the goals.

This modeler is confident in his ability to produce some of the homologies that we described in the previous section: he hopes to assess the Sustainable Development Goals, as well as to feed into one of the most influential science-policy interfaces, the IPCC. He was willing to create an intercomparison device such as the AgMIP, as he was convinced that it would be a decisive instrument for this purpose.

Hence, putting intercomparison at the service of homology production has been a conscious strategy for some modelers. However, the AgMIP only appeared in 2008. The 'climatisation' of agriculture made economic and crop modelers focus on the necessity of creating the AgMIP and constituted an opportunity for them to behave as institutional entrepreneurs, i.e. to try and shape the field according to their interests through "new identities, coalitions and hierarchies" (Fligstein and McAdam, 2012: 84). This is consistent with Fligstein and McAdam's (2012: 181) hypothesis: "entrepreneurs appear not in settled social fields but in those that are emerging or those that are on the verge of transformation". The new framework of climatised agriculture was used by crop modelers to tout the merits of their global crop models, in alliance with economic modelers who had an interest in producing spatialized scenarios. They planned to feed into the 5th Assessment Report of the IPCC and the making of development indicators. More generally, the AgMIP was driven by the preparation of scenarios for the conferences, programmes, etc. of the 'climatised' agricultural agenda. As an interstitial organization at the crossroads between the modelling field and the field of global governance, the AgMIP has enabled these actors to produce homologies between these two fields.

Shaping the field through standardization

Not only was the AgMIP intended to help modelers circulate their results in the field of expertise and governance, but it also favoured certain actors in the modelling field. This was achieved through the definition of a given objective and certain rules for AgMIP, from which dominant modelers would benefit more than other modelling teams. The AgMIP has organized the preference for certain types of models with the alleged objective of harmonizing them, as other interstitial organizations do, for example in the case of analysis techniques used in the regulation of risks (Demortain, 2011). The AgMIP aimed at reducing the diversity of models' outputs *before* these instruments are used by experts and decisionmakers, and this objective has resulted in a certain standardization of the models themselves. However, the AgMIP was not intrinsically intended to shape the field through standardization. There was initially an internal controversy within the group, at the end of which this vision of the AgMIP prevailed.

Harmonizing is a founding principle of the AgMIP (Rosenzweig et al., 2013), which seeks to answer the question: "why do global long-term scenarios for agriculture differ?" (Lampe et al., 2014). Participants in the Global Economics Team of the AgMIP, whatever their position in the modelling field, agree that such an objective is needed in order for models to seem credible. This harmonization is justified in their eyes by the need to protect modelers from criticism that would emerge from greater heterogeneity between models, in the aftermath of the controversy on climate model reliability, following the publication of the 4th Report of the IPCC. As a modeler explains, what is at stake is to build a "robust decision" on science:

> When you have results with very strong heterogeneities [between models], in the end you can't do anything with them. [...] In the previous IPCC results, we finally said: "well, you can have a warming between 2 degrees, or even 0.5 degree and 8 degrees"! When we come to the decision-maker with that kind of conclusion, well, we didn't say anything! So there is a need to reduce the heterogeneity of the models, so all these intercomparison exercises have the ultimate will to reduce the heterogeneity of the models to finally reach a robust decision.

To achieve this harmonization objective, the AgMIP tests the predictions of the different models of the Global Economics Team using the baselines of the IPCC – both its climate scenarios called Representative Concentration Pathways (RCPs) and its socio-economic scenarios, called Shared Socio-economic Pathways (SSPs). Through this lens, various economic global agricultural models propose very different results, for instance when one considers the impact of a given RCP scenario on food prices (Lampe et al., 2014). This is due to their specific form, as shown in the first section, but also to the calibration of elasticities, and the choice of hypotheses (e.g. is land supply rigid? How does food consumption evolve when prices rise?). The AgMIP's objective is then to identify as precisely as possible the origin of these differences, to question the choices that led to them, and *to suggest changes in the models* to limit divergent outcomes. The overall objective of this intercomparison is for the AgMIP to establish agricultural scenarios of reference, called Representative Agricultural Pathways (RAPs) which are intended to be as influential as the IPCC's SSPs and RCPs, and to complement them.

Furthermore, this organization of work, focused on harmonisation and on the making of reference agricultural scenarios, reinforces the structure of the field (Table 1). As we saw in the first section, challengers using subversion strategies value more the heuristic dimension of models over the use of models as evidence-based tools for decision making, which is where incumbents and challengers using 'succession' strategies put more emphasis. For incumbents, the main objective of AgMIP is to have the results of models converge. For example, the IFPRI transformed IMPACT's parameters after participating in the AgMIP, as one participant explains:

> IFPRI came out with a report in 2010 [Nelson et al., 2010, which is based on the IMPACT model] […]: they showed a doubling of food prices by 2050, and then another doubling with climate change. We were very surprised by these results. […] And actually IFPRI has changed its scenarios very significantly, not based on what we said, but because of the AgMIP process. They were confronted and asked "why do you plan such high prices?" and they answered. Largely the problem was that they used models for kinds of medium-term analysis and they had pretty low elasticities, while when you think of the long-term, there is much more flexibility. And when they increased their elasticities, food prices came down.

A modeler from a less dominant institution in the field considers that participating in AgMIP has challenged the structure of his model, e.g. its number of crops. Upon reflection, however, he did not transform his model, but instead wished to preserve what he calls the "good" heterogeneity between models, i.e. the differences in results due to divergent hypotheses that are all equally valid:

> **– For your model, do you feel that your participation in the AgMIP has made a difference?**
> [...] when you are in this community and I come with a very simple representative crop and next to me there is a guy who says "there I have 21 crops", well, I feel a little bad. But you have to resist: everyone has his own questions, his own way of doing things, and it's true that when I came back from an AgMIP meeting, I asked myself, 'you have to break down the model a little more' and then, in retrospect, I said to myself: 'well, no, it doesn't mean anything, considering what we want to achieve, it doesn't mean anything'.

These challengers using subversion strategies have an interest in intercomparison projects, as they could help make their models known through publications and establish contacts with dominant teams. They also seek to situate their model in relation with others and to strengthen their choice of hypotheses and functional forms. They would have preferred that the AgMIP follow the objective of deepening this "technical" comparison of models, precisely in order to avoid standardization. However, this option was not popular among modelers taking part in the project, and it was rejected by the AgMIP's Steering Committee, who focused on harmonizing models and developing scenarios to feed into the IPCC, as a participant recalled:

> I arrived after the beginning of a second phase where there was some hesitation regarding which objective to pursue. […] Meaning when do we really dive into what [one of the coordinators of the Global Economics Team] calls 'deep diving', meaning when do we dive into the model's mechanics, or do we finally ignore all that, and go for politically orientated outcomes, by making scenarios, etc. […]. There was the influence of the Steering Committee […] which was going for the Representative Agricultural Pathways, so we eventually went for that.

The overall effect of the AgMIP on the modelling field has been to have helped GLOBIOM challenge IMPACT, whereas it has kept models using subversion strategies out of the game. Some AgMIP participants consider that the IIASA used the AgMIP as a "launching pad", as GLOBIOM's superiority has been stressed in the publications which came out of the first phase of the AgMIP (Lampe et al., 2014). These publications then became a cornerstone of the IPCC's 5th Assessment Report. IIASA used intercomparison strategically with the GLOBIOM model, producing a new homology between the modelling field and the field of global governance. GLOBIOM, formerly a challenger in the field, has become a new incumbent, partly thanks to AgMIP. This does not mean that all the homologies I documented before arise from the participation of modelling institutions in the AgMIP: not only modelers use more generally their skills for this purpose, but other arenas – such as academic conferences – may have also created standardization between models. Yet, the AgMIP helps us understand recent changes in the field of agricultural models since it was 'climatised', as it reveals certain mechanisms through which the competition between models has been channelled. The construction of an intercomparision between different models resolves the modelling conflict in favour of a new dominant model, and at the same time allows for homology between modelling and politics.

Conclusion

In this article, I used field theory to analyse the competition between organizations which develop world agricultural models since the 'climatisation' of the global agricultural agenda. I showed a change in the field, with the integrated modelling tradition taking the lead over equilibrium models. Despite this change in its structure, the capital, the stake, and the rules of the field do not change. The dominant models are still those that manage to play to the best of their ability as intermediaries between modelling and politics, i.e. to accumulate both scientific and political capital. Even though GLOBIOM (IIASA) challenged IMPACT (IFPRI), both models are similar in this respect. This specific structure of the modelling field explains the homologies that I documented between the modelling field and the field of

Table 1: The participation in the AgMIP from two contrasted positions in the modelling field: a challenger using a subversion strategy vs. an incumbent. Source: interviews

	AgMIP experience / Capacity to structure the field	**Goals for AgMIP /** Definition of what is at stake in the field	**Effects on his own model /** Trajectory in the field	**Expected relationship with other models/** Position in the field
Challenger / using a subversion strategy	"I don't understand a thing"; "There is a core group we are not part of"	Having "substantial" debates through "deep diving" i.e. comparing the technical properties of models	No change, but clarifying their choices of hypotheses *vis-à-vis* other models	Keep the "good" scientific heterogeneity between models, while reducing the "detrimental" political one
Incumbent	"I have wanted to create the AgMIP for a long time"	Influence expertise (e.g. IPCC) and indicators for evidence-based public policies (e.g. Sustainable Development Goals)	New specifications of models to obtain more similar outcomes	Distribution of the topical issues on the political agenda according to the comparative advantage of each model

global governance. I also demonstrated that these homologies are reinforced by the strategies of certain modelers, who behave as institutional entrepreneurs, feeding into the 'climatised' agricultural framework and using the change in the field to improve their positions. Intercomparison projects, such as the AgMIP, are instrumental in this regard: they were created by modelers who had long wanted to have the results of their models more easily circulated in the field of global governance. These actors had an interest in the objectives and rules of the AgMIP, which generated standardization. I explained how these rules helped to solve the main conflict of the field between the two modelling traditions, and at the same time facilitated homologies between modelling and politics. Challengers using subversion, who would have assigned other goals to this intercomparative work, have been marginalized in both the modelling field and in the field of global governance. As suggested by Roueff (2013), homologies are not magic without magicians, i.e. an automatic alignment of positions between the field of model producers and the field of model users. Homologies are rather strategically produced.

This article contributes to the interest of STS in field theory (Berman, 2014; Hess and al., 2017), but it specifically aims to analyse knowledge-power relations in global governance. The co-productionist research agenda on global governance has often looked at individual given expert committees and how they co-construct their own scientific legitimacy with a political body – such as the IPCC with climate governance (Miller, 2007) or the Codex Alimentarius with the WTO (Winickoff and Bushey, 2010). As global governance is complex, and often multi-level and overlapping, a recent research invites to go beyond a unified vision of 'science' coproducing 'policy' at a global scale, by describing several standards-setting bodies' attempt to earn a form of 'epistemic jurisdiction'[14] (Winickoff and Mondou, 2017). In this line, field theory allows to account for the strategies of a plurality of scientific modelling organizations in the field of global agricultural governance, riddled with conflicting visions and interests. Even if the mathematical form of models suggests that they would be universally valid (Fourcade, 2006), global models are 'situated' and their respective preferences for a given agriculture are not representative of all countries' nor all stakeholders' interests. Despite attempts to standardize the field of models, research organizations using subversion keep developing their models, claiming a right to define agricultural policies, and hoping to become influential in the powerful organizations of the field of governance. Based on this result, further research could consider through the lens of field theory how the stakeholders of global governance struggle through the production and/or use of scientific devices (such as models) for the right to define international policies.

Acknowledgements

This research was supported by a grant from the French National Research Agency for the INNOX project, "Innovation in Expertise. Modelling and simulation as tools of governance" (ANR-13-SOIN-0005), coordinated by David Demortain. I am grateful for comments on earlier versions of this article from participants in INNOX workshops, especially from Stefan Aykut, Bilel Benbouzid, David Demortain, and Bertrand Schmitt. Additionally, advice from Jérôme Aust, Olivier Borraz, Patrick Castel, Pierre-Benoît Joly, Jean-Noël Jouzel, Scott Viallet-Thévenin, and two anonymous reviewers of *Science & Technology Studies* greatly improved the paper. I also thank Annelies Fryberger for her excellent language editing.

References

Akera A (2007) Constructing a Representation for an Ecology of Knowledge: Methodological Advances in the Integration of Knowledge and its Various Contexts. *Social Studies of Science* 37(3): 413-441.

Armatte M (2010) *La science économique comme ingénierie. Quantification et modélisation*. Paris: Presses de l'École des Mines.

Cornilleau L and Joly PB (2014) La Révolution Verte, un instrument de gouvernement de la "faim dans le monde". Une histoire de la recherche agronomique internationale. In: Pestre D (ed) *Le Gouvernement des technosciences. Gouverner le progrès et ses dégâts*. Paris: La Découverte, pp. 171-201.

Cornilleau L (2016) La modélisation économique globale, une technologie de gouvernement à distance ? Généalogie, circulations et traductions d'un modèle de la sécurité alimentaire globale de l'IFPRI. *Revue d'Anthropologie des Connaissances* 10(2): 171-196.

Aykut SC, Foyer J and Morena E (2017) *Globalising the Climate: Cop21 and the Climatisation of Global Debates*. London and New York: Routledge.

Berman EP (2014) Field Theories and the Move toward the Market in US Academic Science. *Political Power and Social Theory* 27: 193-221.

Bernstein S and Van der Ven H (2017) Best Practices in Global Governance. *Review of International Studies* 43(3): 534-556.

Bezes P, Chiapello È and Desmarez P (2016) Introduction : la tension savoirs-pouvoirs à l'épreuve du gouvernement par les indicateurs de performance. *Sociologie du travail* 58(4): 347-369.

Bigo D (2011) Pierre Bourdieu and International Relations: Power of Practices, Practices of Power. *International Political Sociology* 5 (3): 225-258.

Bourdieu P (2015) *Sociologie générale: Cours au Collège de France 1981-1983* (Vol. 1). Paris: Le Seuil.

Bruno I, Didier E and Prévieux J (2014) *Statactivisme. Comment lutter avec des nombres*. Paris: La Découverte.

Callon M (1986) Éléments pour une sociologie de la traduction : la domestication des coquilles Saint-Jacques et des marins-pêcheurs dans la baie de Saint-Brieuc. *L'Année sociologique* 36: 169-208.

Dahan A (2007) *Les modèles du futur*. Paris: La Découverte.

Demortain D (2011) *Scientists and the Regulation of Risks: Standardising Control*. Cheltenham: Edward Elgar.

Dervis K, De Melo J and Robinson S (1989) *General equilibrium models for development policies*, World Bank: Washington DC.

Desrosières A (2008) *L'argument statistique. Tomes I et II*. Paris: Presses des Mines.

Djelic ML and Sahlin-Andersson K (2006) *Transnational Governance. Institutional Dynamics of Regulation*. Cambridge: Cambridge University Press.

Fligstein N and McAdam D (2012) *A theory of fields*. Oxford: Oxford University Press.

Fouilleux E (2000) Entre production et institutionnalisation des idées : la réforme de la politique agricole commune. *Revue française de science politique* 50(2): 277-305.

Fouilleux E and Loconto A (2017) Voluntary standards, certification, and accreditation in the global organic agriculture field: a tripartite model of techno-politics, *Agriculture and Human Values* 34: 1-14.

Fourcade M (2006) The Construction of a Global Profession: The Transnationalization of Economics. *American Journal of Sociology*, *112*(1), pp. 145-194.

Frickel S and Moore K (2006) *The New Political Sociology of Science. Institutions, Networks and Power*. Madison: The University of Wisconsin Press.

Go J and Krause M (2016) Fielding Transnationalism: An Introduction. *The Sociological Review Monographs* 64: 6-30.

Goldman M (2005) *Imperial Nature. The World Bank and Struggles for Social Justice in the Age of Globalization*. New Heaven: Yale University Press.

Haas PM (2017) Coupling science to governance: straddling the science-policy interface. In : Littoz-Monnet A (ed) *The politics of expertise in international organizations: how international bureaucracies produce and mobilize knowledge*, Abingdom & New York: Routledge, pp.54-75.

Henriksen LF (2013) Economic models as devices of policy change: Policy paradigms, paradigm shift, and performativity. *Regulation & Governance* 7(4): 481-495.

Hess DJ, Amir S, Frickel S, Kleinman DL, Moore K and Williams LDA (2017) Structural Inequality and the Politics of Science and Technology. In: Felt U (ed) *The Handbook of Science and Technology Studies.* Fourth Edition. Cambridge, MA: The MIT Press, pp. 319-346.

Jasanoff S (2004) *States of Knowledge: the co-production of science and the social order*. London and New York: Routledge.

Joly PB and Dorin B (2019) Modelling World Agriculture as a Learning Machine? From mainstream models to Agribiom 1.0, *Land Use Policy*, March.

Joly PB and Lacombe P (2017) Sciences sociales et politiques agricoles. Une analyse socio-historique des rapports savoir/pouvoir. *Histoire de la recherche contemporaine* VI (2) : 163-179.

Kieken H (2004) RAINS : Modéliser les pollutions atmosphériques pour la négociation internationale. *Revue d'histoire des sciences* 57(2): 377-406.

Kurunmäki L, Mennicken A and Miller P (2016) Quantifying, Economising, and Marketising : Democratising the social sphere ? *Sociologie du travail* 58 : 390-402.

Lampe M, Willenbockel D, Ahammad H et al. (2014) Why do global long-term scenarios for agriculture differ? An overview of the AgMIP global economic model intercomparison. *Agricultural Economics* 45(1): 3-20.

Leblond N and Trottier J (2016) Performing an Invisibility Spell: Global Models, Food Regimes and Smallholders. *International Journal of Sociology of Agriculture & Food* 23(1): 21-40.

Martin JL (2003) What is field theory? *American Journal of Sociology* 109(1): 1-49.

McKenzie D (2006) *An Engine, Not a Camera. How Financial Models Shape Markets.* London & Cambridge: The MIT Press.

Merry SE (2016) *The Seduction of Quantification. Measuring Human Rights, Gender Violence, and Sex Trafficking*. Chicago: The University of Chicago Press.

Miller CA (2007) Democratization, International Knowledge Institutions, and Global Governance. *Governance* 20(2): 325-357.

Morgan MS and Morrison M (1999) *Models as Mediators. Perspectives on Natural and Social Sciences*. Cambridge: Cambridge University Press.

Nelson G, Rosegrant, MW, Palazzo A et al. (2010) *Food Security, Farming and Climate Change to 2050. Scenarios, Results, Policy Options*. Washington DC: IFPRI.

Paradeise C and Filliatreau G (2016) The Emergent Action Field of Metrics: From Rankings to Altmetrics. In: Lounsbury M, Popp Berman E and Paradeise C (eds) *The University under Pressure.* Emerald Group Publishing Limited: Research in the Sociology of Organizations, 46, pp. 87-128.

Reinecke J, Manning S and von Hagen O (2012) The emergence of a standards market: Multiplicity of the sustainability standards in the global coffee industry. *Organization Studies* 33(5-6): 791-814.

Rosenzweig C, Jones JW, Hatfield JL et al. (2013) The agricultural model intercomparison and improvement project (AgMIP): protocols and pilot studies. *Agricultural and Forest Meteorology* 170: 166-182.

Rottenburg R, Merry SE, Park SJ and Mugler J (2015) *The World of Indicators. The Making of Governmental Knowledge through Quantification*. Cambridge: Cambridge University Press.

Roueff O (2013) Les homologies structurales : une magie sociale sans magiciens ? La place des intermédiaires dans la fabrique des valeurs. In: Coulangeon P and Duval J (eds) *Trente Ans Après la Distinction, de Pierre Bourdieu*. Paris: La Découverte, pp.153-164.

Shackley and Wynne (1995) Global Climate Change: the Mutual Construction of an Emergent Science-Policy Domain. *Science and Public Policy* 22(4): 218-230.

Scoones I (2009) The politics of global assessments: the case of the International Assessment of Agricultural Knowledge, Science and Technology for Development (IAASTD). *The Journal of Peasant Studies* 36(3): 547-571.

Sjögren E and Helgesson CF (2007) Health economics and medicine in the shaping of Swedish markets for subsidized pharmaceuticals. In: Callon M, Millo Y and Muniesa F (eds) *Market Devices*, Oxford: Blackwell Publishing, pp. 215–240.

Tomlinson I (2013) Doubling Food Production to Feed the Nine Billion: A Critical Perspective on a Key Discourse of Food Security in the UK. *Journal of Rural Studies* 29: 81-90.

Winickoff DE and Bushey (2010) Science and Power in Global Food Regulation: The Rise of the Codex Alimentarius, *Science Technology & Human Values* 35(3): 356-381.

Winickoff DE and Mondou M (2017) The problem of epistemic jurisdiction in global governance: The case of sustainability standards for biofuels, *Social Studies of Science* 47(1): 7-32.

Notes

1 Models' outputs – e.g. the 70% increase in the global food production by 2050, a result of a FAO model – fuel the productionnist narrative which is instrumental to agrifood industry and export-countries (Tomlinson, 2013).

2 In 2010 the team consisted of the ten following models and institutions: AIM (developed by the NIES/Japan), ENVISAGE (FAO/World Bank), EPPA (MIT/USA), FARM, (USDA) GTEM (ANTARES/Australia), MAGNET (LEI-WUR, Netherlands), GCAM (PNNL, USA), GLOBIOM (IIASA), IMPACT (IFPRI), MAgPIE (PIK, Germany).

3 They were actually inspired by Bourdieu in this: for him, fields (e.g. the scientific field, the political field, etc.) are specialized subfields of the 'social field' (society), which is not a meta-field, but is a conglomeration of all fields.

4 It refers to a "form of embedding whereby actors that make up smaller collectivities are located within larger strategic action fields that contain larger collectivities" (Fligstein and McAdam, 2012: 59)

5 These institutions developed both *computable general* equilibrium (CGE) models (representing the whole economy) or *partial* equilibrium (PE) models (focusing on the dynamics of the agricultural sector).

6 More details are provided by Leblond and Trottier (2016) and by Joly and Lacombe (2017).

7 Four modelling traditions have been working on this difficult task since the 1970s (Leblond and Trottier, 2016): 1) *economic models* 2) *biophysical models,* representing global agricultural productivity (potential yields according to agronomic theory; actual yields according to databases), 3) *integrated models*, 4) *hybrid models,* which link socio-economic databases to agronomic databases at a pixel scale thanks to satellite-produced datasets.

8 A last evolution is that both equilibrium and integrated models benefited from the democratization of satellite datasets in the 2000s: "they integrated Geographic Information Systems and began projecting their results onto pixel grids" (Leblond and Trottier, 2016: 7), what enables them to localize even more precisely the agriculture-climate interactions.

9 Price elasticity quantifies the change in demand for a good caused by the change in its price.

10 Databases produced by the FAO or by networks of modelers follow agricultural chains, so mixed or alternate crops encouraged by agroecology cannot be represented, for example.

11 This is due to due to problems commensurating at the international scale a variety of soils, climates, etc., but also because some countries want to keep them secret for trade-related reasons.

12 Source: http://www.agmip.org/feature-video/

13 The EMF is probably the oldest model intercomparison, but there has been a rise in these projects in the context of the controversies on the validity of models used by the IPCC, as can be seen in the *Inter-Sectoral Impact Model Intercomparison Project* (ISIMIP) on the effects of climate change, or the *Integrated Assessment Modelling Consortium* (*IAMC*) on integrated models.

14 This concept refers to "the power to produce or warrant technical knowledge for a given political community, topical arena or geographical territory" (Winickoff and Mondou, 2017: 7).

Appendix 1. List of Acronyms

AgMIP: Agricultural Models Inter-comparison and Improvement Project

CAADP: Comprehensive Africa Agriculture Development Programme

CGIAR: Consultative Group in International Agricultural Research

EMF: Energy Modelling Forum

FAO: Food and Agriculture Organization

GATT: General Agreement on Tariffs and Trade

GLOBIOM (model): Global Biosphere Management Model

IAASTD: International Assessment of Agricultural Knowledge Science and Technology

IFPRI: International Food Policies Research Institute

IIASA: International Institute for Applied System Analysis

IMPACT (model): International Model for Policy Analysis of Agricultural Commodities and Trade

IPCC: Intergovernmental Panel on Climate Change

MDGs: Millennium Development Goals

MEA: Millennium Ecological Assessment

NLU (model): Nexus-Land Use

OECDE: Organization for Economic Cooperation and Development

SDGs: Sustainable Development Goals

UN: United Nations

WTO: World Trade Organization

On the Plurality of Environmental Regimes of Anticipation: Insights from Forest Science and Management

Antoine Dolez

Université Grenoble Alpes, CNRS, Sciences Po Grenoble ; Irstea Lessem, France / antoine.dolez@umrpacte.fr

Céline Granjou

Université Grenoble Alpes, Irstea Lessem / LISIS (Laboratoire interdisciplinaire sciences innovations sociétés), France

Séverine Louvel

Université Grenoble Alpes, CNRS, Sciences Po Grenoble, France

Abstract

In recent years, the social sciences have increasingly investigated ways in which futures are anticipated, fostered, and pre–empted. However, less attention has been given to how various predictive approaches inform different ways of acting in the present. Our article presents the results of an investigation into the current practices and agendas of forest scientists and managers in France. We first suggest how an anticipation of environmental futures is coming to the fore as an emerging field of expertise and practices in forest sciences, including predicting but also monitoring, preparing and adapting to projected futures. We then account for the co–existence of three 'micro–regimes' of anticipation combining a certain approach to the forest, a certain vision of the future, and a certain type of scientific predictive approach, including different anticipatory objectives, different modelling practices, and different interactions between research and management: i/ Adapting forestry to future climates; ii/ Predicting Future Tree Biology; iii/ Monitoring forests as indicators of climate change.

Keywords: regime of anticipation, climate change, forest science, foreknowledge

Introduction

In recent years, the social sciences have increasingly investigated the ways in which futures are anticipated, fostered, and pre–empted (Adams et al., 2009; Tavory and Eliasoph, 2013; Andersson and Duhautois, 2016; Coleman and Tutton, 2017; Granjou et al., 2017). A recent special issue in *the Sociological Review* is emblematic of the call for a new "engagement with and interrogation of the future in social sciences" (Coleman and Tutton, 2017: 441). It suggests "shifting the emphasis from looking into the future to looking at the future—that is, to engage with the future as an analytical object" (Coleman and Tutton, 2017: 441). This article is located at the crossroad between two streams of literature: first, the emerging scholarship which builds on cultural geography and anthropology research to question how ideas of the future inform actions in the present; second, the specific insights of Science and Technology Studies into the production and role of scientific forecasts and models in various academic fields and disciplines. In addressing how the anticipation of socio–environmental futures in a changing climate are coming to the fore as a new scientific and political agenda, this article aims to scrutinize the coexistence of various 'regimes' of predictive and anticipatory knowledge production and their embedment in several and partly conflicting politics of environmental anticipation. It follows Mike Michael's suggestion to account for the 'ecology of futures' at play in the intertwinement of both the 'Big Futures' at stake in broad societal and ecological narratives, and the 'Little Futures' pertaining to everyday social life and interactions (Michael, 2017). Our key question is: how does climate change, as the embodiment of a 'Big Future,' play out on the practices of forest modellers, and specifically on the way they anticipate the evolution of their research agendas and contributions to forest management? Our contribution shall account for the various Big and Little Futures and their relations at stake with the growing focus of environmental sciences on climate change anticipation.

Our contribution focuses on the case of forest science and forestry and unpacks how the anticipation of forest evolution in future climates is currently gaining ground in a rapidly growing field of research, expertise and management. Forest management is embedded within complex and interwoven issues of ecological sustainability and profitability. Its future prospects in a changing climate are highly uncertain and foster new concerns about how to anticipate the changing patterns of tree growth, species distribution, plant disease outbreak, forest productivity and economic profitability, as well as the overall adaptation of forestry practices. Forests are at the forefront of the "battle" against global change because they are often perceived as the "lungs of the world" and are thus a key factor in reducing carbon dioxide emissions. In this article, we document the way in which a growing range of forest scientists, field observers and managers are realigning their agendas, practices and goals around new anticipatory agendas and standards associated with the circulation of climate change projections and anticipatory concerns. By doing so, we shall highlight the plurality of anticipatory research agendas and predictive technologies that forest scientists have developed and how they are embedded within various visions of forest and forest futures as well as within contrasted relationships between research and forest management.

We will first present brief historical insights into the evolution of forests and forestry practices in France and emphasise how recent concerns about climate change have fostered new anticipatory agendas and practices in forest science and management. We shall describe how climate change topics and concerns are transforming forest science organisations, collaborations and material infrastructures of knowledge, including practices of data production, and how this process of 'climatisation' (Aykut et al.., 2017) involves increasing exchanges and collaborations between forest science, ecology and climate science and the models that were previously developed in isolation from each of those fields. Then we shall document the co–existence of three micro–regimes of anticipation in the case of French forest science and management and eventually account for their tensions and relations. Each of these micro–regimes combines a vision of the future with an approach to the forest, including a certain type of scientific predictive approach associated with modelling practices. In particular, we will discuss Aradau and Van Münster's (2013) and Amoore's (2013) idea that the rise of future–oriented knowledge agendas destabilises pre–existing scientific approaches based on the interpretation of past data and requires disruptive epistemology and practices. We argue that the future does not necessarily disrupt previous epistemic practices and organisations; instead, the production of foreknowledge is embedded in various, situated visions of the specific future and knowledge.

Empirical and theoretical background

The case of French forest science and management

This article is based on a sociological investigation into forest science and management in France, including about 30 interviews with forest scientists, forest managers and field correspondents and backed up with the reading of forest science articles, administrative literature and institutional websites. The forest scientists we interviewed are members of the main disciplines involved in forest science (mostly ecology, but also biology, genetics and computer science) and various research institutions, including France's National Center for Scientific Research (CNRS), the Research Institute of Science and Technology for Environment and Agriculture (IRSTEA) and the National Institute for Agriculture Research (INRA). Semi–structured interviews were focused on scientists' professional trajectory and career, their vision of the scientific field including modelling and predicting the evolution of forests, and their vision of the future and the type of knowledge they build on it. In the case of forest managers, interviews included a focus on management practices and potential uses of models and simulations.

What makes France an interesting example of the development of various predictive and anticipatory regimes regarding forests and climate change? It is mostly the strong interdependence between forest science and forest management which can be traced back to the French historical tradition of centralised forest science and forest management, linked to the high economic and social importance of forests to the nation (Decocq et al., 2016). Today forests cover around 30% of the French mainland with 75% being private forests and 25% public forests, a third of which are managed by the National Forests Office (ONF) while the rest is owned by local councils. The ONF, which employs around 10,000 people today, was created in 1964 and is in charge of wood production, forest protection and tourism. The Research and Development department of the ONF promotes exchanges, collaborations and knowledge transfers between forest science and forestry.

Theoretical background: 'climatisation' and the 'micro–regimes' of anticipation

The STS literature has long addressed the production of predictive knowledge and the particular status and role of predictions in science–policy interface (in systems dynamics: Bloomfield, 1986; economy: Collins and Pinch, 1996 and geosciences: Sarewitz et al., 2000; Dahan–Dalmedico, 2006; Edwards, 2010). STS scholars have notably criticized the "quest for a scientifically legitimated view of the future" through the development of scientific predictive models (Sarewitz et al., 2000: 367). Much attention has been given to why predictive models do not or cannot produce accurate predictions (Collins and Pinch, 1998) and how those predictions should be communicated, received and used (Sarewitz et al., 2000). Recent STS and post–ANT scholarship also insist on the performative role of models, scenarios and simulations and their effects on shaping the reality, for instance, how economics shape and perform economy (Callon, 1998; Mackenzie et al., 2007) and more broadly how differing scientific models and approaches shape various "ontologies" (Mol, 2002; Law and Mol, 2002), as they encompass several different ways of knowing and intervening upon it. Our point, however, is concerned less with these ontological politics than with the embedment of predictive models within various and partly conflicting politics of environmental anticipation.

STS scholars started scrutinizing the internal plurality of predictive approaches and models in the case of glaciology (yet see: Skrydstrup, 2017). However, less attention has been given to how differing predictive approaches inform various ways of acting in the present. Recent literature examined the investigation of official counter–terrorism programmes and practices and elaborated on the expansion of a 'politics of possibility' that aims "not to prevent the playing out of a particular course of events on the basis of past data tracked forward into probable futures but to pre–empt an unfolding and emergent event in relation to an array of possible projected futures" (Amoore, 2013: 9). Aradau and van Münster also depicted an all–encompassing 'Regime of Anticipation' including the development of a new "conjectural episteme" in which imagination mostly replaces the use of past data sets to attempt to make the future "knowable" (Aradau and van Münster, 2011).

We argue that those theorisations do not properly account for how scientists, experts, policymakers and managers associate practices, infrastructures and imaginaries to anticipate the 'not yet'. We propose to develop the notion of 'micro–regime' of anticipation in order to empirically account for those various assemblages and their mutual tensions. We further criticize the idea there is only one Science (Knorr–Cetina, 1999) by unpacking the diversity of foreknowledge and predictive models in the case of forests and climate change. Our notion of 'micro–regimes' of anticipation is inspired by Pestre's approach to the micro–historical embedment of the production of knowledge into socio–economic regulation (Pestre, 2003) and departs from descriptions of the historical development and succession of broad regimes or 'styles of knowing or reasoning' from past to present (Hacking, 1994; Kwa, 2011). While Pestre's notion of 'regime of knowledge' puts forward the interactions between science, politics and society on a macro–social level, the notion of 'micro–regimes' is located at the smaller level of mundane research agendas and practices (Shinn, 1999). 'Micro–regimes' of anticipation are ways of negotiating the co–production of 'Big Futures' (here embodied in narratives of climate change and its impacts on forests) together with 'Little Futures', which pertain to routine research practices and interactions, including developing models, collaborations and projects, etc. (Michael, 2017).

The construction of climate change as a global concern and expertise has been extensively documented (see in particular: Jasanoff and de Martello, 2004; Edwards, 2010). However, research is only emerging on how climate change issues and concerns are now reframed as local concerns in a wide range of sectors and activities whose practices, communities, jobs and identities are being transformed and re–aligned toward anticipatory objectives which relate to various situated activities and agendas. Sociologists Aykut, Foyer and Morena (2017) proposed the notion of 'climatisation' to depict this multi–level and highly contextual process of re–alignment of a range of agendas and practices with climate issues. We argue that 'climatisation' importantly involves the realignment of knowledge practices toward anticipatory objectives and agendas, including predicting but also monitoring, preparing and adapting to projected futures. We shall describe the 'climatisation' of forest science organisations, collaborations and material infrastructures of knowledge, including practices of data production and modelling, and how this involves increasing exchanges and collaborations between forest science, ecology and climate science—as well as between the models that were previously developed in isolation in each of those fields. Eventually, we shall document the coexistence of different competing micro–regimes of anticipation that forest scientists and managers deploy and how those anticipatory micro–regimes are embedded in various (and partly competing) processes of research agenda setting and environmental issue framing.

Anticipatory pluralism in forests science and management

The 'climatisation' of forest science and management

Almost all the forest scientists and managers we met spontaneously referred to climate change in their answers when asked about on–going changes in their research agenda and practices, indicating that aligning one's research and agenda with climate change had become necessary in order to attract funding. Beyond the rhetorical reference to climate change, forest scientists and practitioners also suggested that they were now confronted with very practical questions related to the anticipation of forest growth and productivity under future climates, such as: How will rising temperatures, soil acidification and water scarcity influence tree growth and forest species composition? Which species will be the most resilient to future droughts, heat waves or storms? When will be the best moment to harvest timber productions? Our field work thus confirms bibliometric analyses which suggest that climate change has become a central topic in forest science, along with others such as "carbon dioxide" and "adaptation" (Aleixandre–Benavent et al., 2017). However, we found that 'climatisation' (Aykut et al.., 2017) occurred not only at a discursive level—meaning the integration of the topic of climate change into forest science agendas and discourses—but also within organisations, collaborations and material infrastructures of knowledge. For instance, the ONF's R&D department was reorganised in 2005 along five topics including climate change. A range of new research and management networks were created in order to address the impact of climate change on forests and forestry, such as the GIP ECOFOR ("ECOsystèmes FORestiers") which was founded in 1993 and has been supporting research on forests and climate change since the beginning of the 2000s, including research on the future distribution of French forests in a changing climate[1]. Forest scientists and managers also started implementing new climate–related data collection and new collaborative arrangements around climate data sharing and use, for instance within the network of forest observation sites called RENECOFOR, initially created in relation to concerns over the effects of acid rain on forests. In 2012, RENECOFOR was also integrated into a Long–term Environmental Research Monitoring and Testing System (SOERE), whose goal is to produce data on the "System Earth" dynamic, illustrating how the collection of forest and climate related data becomes embedded within new collaboration arrangements between a broad range of disciplines such as ecology, forest science, genetics and population biology.

The 'climatisation' of forest science and management also fostered the development of new methods for modelling forest growth and making decisions regarding which tree species to choose and when to fell trees for timber production. Such decisions were first made on the basis of the production quotas, which calculate the annual growth of one species in a particular place, and served as decision guidelines for when to fell the trees and which tree species will grow faster. From the mid–1980s, forest engineers have developed empirical models in order to predict the timber productivity of a particular forest in a constant environment, including soil and climate. Those models are fuelled by long–term data produced by the National Forest Inventory. The 'climatisation' of forest science and management has triggered the convergence of empirical, forestry–oriented models with other types of models that have long developed in isolation from forestry concerns and practices, i.e. process–based, statistical models (Korzukhin et al., 1996; Adams et al., 2013). The latter models have been developed in biology and in ecology since the beginning of the twentieth century in the wake of the equations of Lotka–Volterra (Leslie, 1948). They aim to understand simple or fundamental biological processes such as photosynthesis (Farquhar et al., 1980) or carbon allocation and to translate them into equations. These equations can then fuel a computer program that simulates "virtual experiments" (Legay, 1997).

These two types of models have long been developed in isolation from each other, as forestry–oriented models were taught in forestry schools and process–based models in ecology and biology master's degrees. In the early 2000s, growing interest in understanding the evolution of forests in the context of climate change led to the development of new models that blur that distinction by mixing the characteristics and objectives of the two former categories of models, as a forest scientist explained:

> Process–based models attempted to summarise ecosystem functioning without any predictive objective […] however, with climate change issues we discovered that this dichotomy did not work anymore […].We understood that empirical models should also explicitly integrate climatic data into their architecture and equations; on the other hand, process–based models should also address prediction and applications…

As a result, many of the forest scientists we interviewed use climate scenarios designed by the IPCC and other climate–oriented research institutions (such as Météo–France which develops climate projections with great precision at the local level in France) and integrate them into pre–existing forestry–oriented models in order to simulate the potential growth of trees in a changing climate.

On the other hand, while climate projections are increasingly used and integrated into forest models, forest models are also increasingly integrated into climate change modelling which tends to become more and more integrative and biological (while initial climate models relied on atmosphere physics and chemistry only). Science historian Amy Dahan–Dalmedico accounted for how climate change models shifted in the 1990s from a focus on the atmosphere to a broader focus on 'Earth systems' integrating oceans and terrestrial surfaces, i.e. vegetation and forests (Dahan–Dalmedico, 2010). As a result, not only do forests scientists use climatic models in forest research, but forests models also fuel broader climate models that simulate the interactions between the atmosphere and the biosphere. For instance, the Laboratory of Climate and Environmental Sciences in Paris–Saclay, which hosts climate scientists actively involved in the IPCC, has developed the ORCHIDEE model, which simulates the role of tree development and life cycles in carbon flows in the biosphere, including the ORCHIDEE–FM submodel which integrates the effects of various forestry strategies on carbon cycle.

The 'climatisation' of forest sciences thus involves increasing exchanges and collaborations between forest science, ecology and climate science and the models that were previously developed in isolation in each of those fields. Yet far from an all–encompassing alignment towards a unique anticipatory "episteme" (Aradau and van Münster, 2013), our fieldwork also points to the co–existence of a plurality of research agendas dedicated to anticipating the future of forests under a changing climate. These research agendas differ in two ways: they develop forest–driven vs. climate–driven science on one hand. On the other hand, they handle forestry–oriented versus ecological and biological process–based research.

Three 'micro–regimes' of anticipation

This section describes three 'micro–regimes' of anticipation which became apparent during our fieldwork. Each of these combines three dimensions: a certain vision of the forest, a certain idea of the future, and a certain type of scientific predictive approach, including different anticipatory objectives, different types of models and modelling practices, and different interactions between research and management (see Table 1). The three 'micro–regimes' encompass actors' various views of the type of knowledge that matters for forest and forest management, including certain visions of the extent to which the future disrupts past and present scientific practices and technologies.

Following the analytical distinction introduced by Mike Michael (2017), each 'micro–regime' of anticipation shapes and performs both "Big" and "Little" futures including ecological futures (climate change), economic futures (forestry evolution), and academic and scientific futures (research agenda setting, maintenance or creation of collaborations, publication writing, etc.). These are ways of negotiating their coexistence and potential tensions at various levels.

First 'micro–regime': Adapting forestry to future climates

In the first 'micro–regime' of anticipation, researchers and managers seek to predict the composition and geographical distribution of forest and forest socio–ecologies in future climates. Their scientific practices are those traditionally used in forest science and engineering. They assess timber stocks using forest inventories and maps and they construct statistical models that build on correlations between a wide range of ecological, geophysical and socio–economic parameters in order to estimate the productivity of future forests, to write guidelines and design forests policies. That foreknowledge is meant to help produce guidelines for present and future forestry practices and to contribute to forestry economic planning and adaptation. The vision of the future focuses on securing future forestry activities. Forests are considered to be anthropogenic, managed socio–ecosystems.

The anticipatory logic of this 'micro–regime' is to improve forestry strategies by taking into account climate change. What matters is that the model can be applied in order to guide forestry choices. A key objective is to identify the forest practitioners' leverage actions to secure forestry in a changing climate. This first 'micro–regime' thus gathers forest scientists and engineers from various research institutes (IRSTEA, INRA), the R&D department of the National Forest office, and members of forest research networks such as GIP ECOFOR. Close relations and collaborations between forests scientists, managers and decision–makers are an

Table 1: Three 'micro–regimes' of anticipation.

	Adapting forestry to future climates	Predicting Future Tree Biology	Monitoring forests as indicators of climate change
1. Vision of forests	Forests are considered to be socio–ecosystems.	Forests are considered to be a functional system governed by ecological processes.	Forests are considered to be an observatory of the evolution of climate change.
2. Visions of the future	The future is viewed on past and present trends.	The future is disruptive. Researchers and managers assume that studying fundamental ecological processes in a changing climate is essentially different from studying them under stable climatic conditions.	The future becomes palpable and knowable, and thus governable.
3. Type of scientific predictive practices			
3–i/ Anticipatory logic and objectives	Adapting forestry strategies in a changing climate	Understanding the ecological processes at stake in the evolution of forests (such as carbon flows and water scarcity)	Producing indicators of climate change, assessing and mapping its evolution
3–ii/ Modelling practices	Statistical "meta–models" that aggregate a broad range of ecological, social and economic variables	Simple ecological and process–based models	Simple model illustrating the causal relation between the chosen indicator and climate change
3–iii/ Interactions between research and management	Strong collaborations between forest science and forest management: data sharing, collaboration in research projects, and co–production of forestry guidelines	Researchers contribute as experts to biodiversity and nature conservation international and national organizations	Contributions to the French Ministry of Environment; Co–construction of indicators between forest researchers and managers.

essential characteristic of this micro–regime; they result both from collaborations and from individual mobility between organizations dedicated to forest management (such as the National Forest Office) and organisations dedicated to scientific research and to producing technical support and advice for foresters (such as INRA and IRSTEA).

The models at stake in this first 'micro–regime' are constructed and calibrated using a large amount of data collected during forest inventories and by research teams. They integrate both ecological variables (i.e. tree growth and mortality, forests composition, light interception) and socioeconomic variables (i.e. forestry strategies, expected timber stock). The researchers' objective is to run multiple simulations with different models to foresee the consequences of their potential forestry's strategies. As a result, models tend to accumulate and integrate an ever increasing number and variety of variables, as suggested for instance by the case of the SAMSARA model, a tree–growth model whose many versions were developed over time in order to achieve the integration of an ever wider range of processes, including wind damage (Ancelin et al., 2004), colonisation processes (Cordonnier et al., 2006), intraspecific competitions (Vieilledent et al., 2010) and biodiversity's stock, ecosystems services (Courbaud et al. 2017; Lafond et al. 2017)[2] and climatic parameters (while the model was initially designed under a constant climate) (Lagarrigues, 2016). The implementation of the CAPSIS modelling platform in 1998 (Dufour–Kowalski et al., 2012) also illustrates the trend towards the increase in the number of variables as the platform integrates about 70 different forest models in only one simulation software and makes it possible to run all of them together in order to predict how any given variable will react in a changing climate. An engineer in charge of CAPSIS told us:

> In CAPSIS, we have tree–growth models. These models grow trees; they create virtual forests. Then, you can add other models to them, for example, timber quality models or risk models, to address how the forest may resist wind storms (…) or economic models. All those second–level models can be added to CAPSIS tree–growth models in the same simulator.

Accordingly, the models used in the first 'micro–regimes' must be user–friendly—the calculation speed is a key element in this: "If it takes too much time to set up the data and run the simulation, it is off–putting. It is a beautiful theoretical tool but in practice it is useless" (Forests modeller). When we asked modellers and managers about potential future improvements of the models, all of them referred to an increase in the calculation's speed. The development of remote–sensing using LIDAR (Laser Imaging Detection and Ranging) is also expected to help collect ever larger amounts of data to fuel the models with the idea that scientific progress meant both more and more past and present data for more and more anticipative models. Accordingly, the first micro–regime does not focus on knowing the ecological processes but relies on statistical methods applied to big datasets to find correlations between climatic, biological and socio–economic information. Modellers use ecological processes such as tools to parameterise the models; however, producing knowledge on those processes is not part of their objectives:

> I'm not interested in ecology as a science. I take an interest in forest as a socio–ecosystem; it's my point of view, but I can't avoid ecology because it is one of the analytical and theoretical sciences I use to study forests. (Modeller working at the National Forest Office and INRA)

As a result, in the first micro–regime, the future is mostly deduced from past and current trends detected by using ever bigger sets of data. As a modeller reported: "Yes, I'm interested in the future, but most of the time, it is the past that I study." However, both modellers and managers are aware that the future may destabilize and disrupt past and present trends. They address the disruptive character of the future by accumulating models and variables—thus, rising computing power—and by developing new statistical methods, such as the Bayesian approach, which "aims to artificially break with the linear structure of time", a modeller said. These tools are standardized for example in *R*-packages (Jabot et al., 2013).

Referring to Michael's distinction between 'Big' and 'Little' futures (Michael, 2017), one could say that the Big Future envisioned by forest researchers and managers in this micro–regime is about securing forestry under a changing climate, while the Little Futures that are at play in the everyday life of forest research teams include:

- Developing ever more sophisticated and integrative statistical models that take into account an increasing range of parameters in order to produce forestry guidelines;
- Developing new technologies of data–collection, including remote–sensing technologies, in order to improve the quantity and quality of the data available;
- Reforming forest management and the institutions in charge of it in order to secure future forestry under a changing climate.

Second 'micro–regime': anticipating future tree biology

In the second 'micro–regime' of anticipation, researchers aim to understand and predict the evolution of forest ecological functioning by modelling how future climates—including extreme events such as severe droughts (Estiarte et al., 2016; Lempereur et al., 2017)—will affect the way trees use an array of resources (water, carbon, nutrients, light). This 'micro–regime' is no longer related to utilitarian objectives (i.e. designing forestry practical guidelines and economic planning). Instead, it should be understood in the more environmentalist perspective of understanding the ecological processes at stake in the evolution of forests under future climates. Some of them contribute as experts to biodiversity and nature conservation organizations, such as the Intergovernmental Platform on Biodiversity and Ecosystem Services or the French Foundation for Biodiversity Research. This micro–regime aggregates what scientists call "simple" process–based biological modelling, *in natura* experiments and ecological theories.

Ecologists and modellers involved in this 'micro–regime' tackle ecological questions such as: What is the link between diversity and stability? (Morin et al., 2011) How resilient is the ecosystem when faced with scarcity? Which function of the ecosystems will be the first to react to a changing climate (Gustafson et al., 2015)? Compared with the first 'micro–regime', ecology—especially functional ecology—plays a central role and is no longer merely considered a tool. For the researchers involved in this 'micro–regime', understanding the dynamic of forest ecosystems is an opportunity to understand fundamental ecological processes at stake, such as photosynthesis or the allocation of carbon between plants (Gea–Izquierdo et al., 2015), atmosphere and soil and between different parts of the plants, as explained by a forest scientist whom we interviewed:

> I'm not interested in the holm oak, but rather in… how forests respond to scarcity (…) In that case, we worked on the holm oak. Yet what matters for us is really the functioning, the functional aspects in terms of ecosystems, carbon flow, growth, primary productivity …

The anticipatory logic of this second 'micro–regime' addresses climate change as a disruptive event that forces forest modellers to renew their practices and develop collaborations with experimenters and functional ecologists. Researchers assume that studying fundamental ecological processes in a changing climate is essentially different from studying them under stable climatic conditions. Climate change is thought to trigger environmental conditions that will be essentially different from the environmental past and present, as one of the forest ecologists we met explained:

> As soon as we have something calibrated to the present… I mean, it is tempting to apply the model to the future and to see what will happen. But the question is: Is the knowledge of the system in the current climate sufficient to be applied to future scenarios, including extreme conditions? … I want to know whether my little model, which is calibrated to current conditions, using 10 years of data collection, could be applied to extreme events, such as a six–month water shortage. That is to say, things that you have rarely or never observed so far…

In a similar way, another forest ecologist explained that "modelling is not interesting when everything is all right," meaning he did not expect simulation models to make a linear business–as–usual prediction based on past data, but instead to be able to integrate future extreme events and to deal with the disruptive nature of climate future. With that goal, ecologists and modellers seek to capture the non–linear responses of ecosystems by developing both modelling and *in natura* experimentation (Perez–Ramos et al., 2010). For example, the same ecologist developed an experimentation consisting of excluding rain for a few months (using artificial covers to protect plants from the rain) in order to enrich his model:

> It has allowed me to add a few modules that were not in the first version of the model. The model calibrated to the current climate works up to a precise threshold, and once this threshold is crossed, you have to add a [new] module... it is something that my models initially did not take into account, and now I am developing it in order to simulate the non–linear relationship between water scarcity and fire risk.

Field experimentations are meant to help set and observe the possible future climatic conditions and their impacts on trees and forests. The development of experimentations on the effects of climate change on ecosystem functioning is emblematic of what some interviewed forest ecologists called an "experimental turn" that dates back to the mid–2000s. Yet some forest ecologists whom we interviewed emphasized the difficulty or even impossibility of conducting experiments on trees in completely controlled 'live labs' such as the controlled experimental enclosures called 'ecotrons' that have been constructed in France in the 2000s (and elsewhere in the world[3]) because of scale issues (ecotrons are not designed to host more than two–meter–high plants) (Granjou and Walker, 2016). As a consequence, they have developed field experiments that consist of conditioning gas concentrations or simulating climates (for instance, droughts) in the field by using, for instance, flux towers which enrich the air in CO2 and measure gas concentration and temperature at the bottom and at the top of the studied trees (Misson et al., 2010).

In this 'micro–regime', relationships between researchers and managers are less close than in the first micro–regime. This is partly because researchers would rather belong to laboratories and centres favouring fundamental research over applied research (such as the French National Center for Scientific Research CNRS)—even though the institutions of belonging do not systematically determine the development of applied versus fundamental research by their members. This is also because when researchers produce knowledge on ecological processes, it makes it difficult for them to connect to forest management issues and concerns, as a forest scientists working at the CNRS (Center for Evolutionary and Functional Ecology) explained:

> Relationships with managers remain difficult because we face a cultural issue. I mean, we consider forests as an ecosystem. It is an ecological point of view, and they consider forests as a production means, something cultivated. In foresters' minds, the forest is something we cultivate, and it is not a natural ecosystem. They are concerned about productivity, plantations or species selection and not really about how trees use what they have and how they manage available resources. (…) [W]e are definitely not on it.

In the second 'micro–regime', the Big Future is about anticipating future tree biology under a changing climate. This is enacted through 'Little Futures' that include improving simple models of tree biology, developing *in natura* experimental infrastructures, improving the linkages between process–based models and field experimentations, and contributing to biodiversity conservation and management organizations.

Third 'micro–regime': Monitoring forests as indicators of climate change

In the third 'micro–regime' of anticipation, researchers and managers consider the evolution of forests as a case–study for observing and predicting the evolution of climate and its broader impacts on ecosystems and society. They monitor forest fauna and flora, collect field data and combine the data with population biology models and laboratory experiments in order to develop indicators of the growing intensity and impacts of climate change. While the models used in this 'micro–regime' are meant to be simple models, like in the second 'micro–regime', the ultimate objective is not to understand basic biology but to represent the state and evolution of climate in order to inform policy–makers, in particular from the French Ministry of Environment, in order to design environmental and climate policies. This 'micro–regime' mostly builds on scientific results produced by the first two 'micro–regimes' (as, for instance, the budburst indicator which builds on the modelling of the various steps of plant growth and life cycle, Chuine, 2000). Eventually the third micro–regime differs from the first two as it does not tackle "the forest" *per se* but focuses on climate change and only makes use of the forest as a proxy to make climate change visible: the very object of anticipation differs between the first two 'micro–regimes' and the third one.

In the third 'micro–regime' of anticipation, researchers from the National Institute for Agriculture Research (INRA) and field correspondents from the ONF monitor and record forest data in order to document and assess the ongoing evolution of climate. Here, forest data are useful to the extent that they can be directly linked to climate change, such as, for instance, shifting budburst dates which are thought to be caused by warmer springs (Chuine and Cour, 1999). Forest data are conceived as climate change indicators when the relation between the indicator and climate change is considered to be simple and almost causal. The ONERC (National Observatory on the Effects of Global Warming), which is part of the French Ministry of Environment, defines an indicator as "information tied to a phenomenon [that shows] its evolution through time in an objective way."An indicator is like a thermometer: "As the body's temperature gives a hint about a patient's health, climate change indicators tell us about the Earth's state" (National Observatory on the Effects of Global Warming, 2010). As in the first 'micro–regime', indicators are conveyed to both forest managers and politicians to inform them of the evolution of climate change and allow for designing appropriate forest policies. Therefore, the form and aesthetic of the indicator are central because this is the way in which climate change is made visible and palpable. This 'micro–regime' gathers scientists from a broad range of disciplines such as genetics, population biology or entomology. Researchers produce indicators and information for the French Ministry of Environment that are meant to be collective reference landmarks on the intensity and impacts of climate change for policy–making.

One of the best–known indicators in France is the pine processionary caterpillar (Rossi et al., 2015), which is a forest parasite that causes tree death and health problems inhumans and animals. Pine processionary caterpillars have very stinging hairs that can cause skin problems to both humans and pets. Researchers initially started studying this insect because it is one of the most dangerous European parasites for forests as they eat pine needles and cause the tree to die. The progressive change of focus in the research devoted to the caterpillar, from pest to climate indicator, gives another example of the 'climatisation' of forest research agendas and practices (Roques, 2015). Since the creation of the ONERC in 2001, the caterpillar has indeed become a central indicator for assessing the evolution of climate once the relation of its growing numbers and shifting geographic distribution with climatic parameters was established in the literature. The processionary caterpillar was thus progressively built as a "reference model" meaning "a model of response to climate change" (to quote a forest researcher whom we interviewed). This latter notion of 'model' has some similarities with the way biologists use 'model organisms' such as mice or E. Coli in order to study fundamental biological mechanisms. While the idea in both cases is to focus on a 'simplified case' in order to understand a more complicated general issue, in our third 'micro–regime', however, forest scientists' ultimate goal is not to study basic ecological mechanisms (like in the second micro–regime) but to deduce trends of on–going climate change from certain characteristics of forest evolution. Accordingly, in a similar manner to the statistical models of the first 'micro–regime', researchers and managers expect to be able to collect ever more field data in order to improve the accuracy of climate change indicators.

Results are presented under the form of a map that shows the past, present and future progression of the "colonisation front" of the caterpillar. Therefore, the anticipatory logic at stake here puts past, present and future in linear succession in order to make climate change visible and palpable. Compared to the two other 'micro–regimes', the focus of the third micro–regime is on raising the alarm and alerting managers and decision–makers on the progress of climate change with the idea that the future is already here and we have to act now in order to adapt to it. Accordingly, the vision of the future is both in continuity with past and present at the level of scientific and modelling practices that use past data of caterpillar populations and distribution, and disruptive at the political level because it invites both managers and politicians to act now in order to adapt to the future.

In this 'micro–regime', Big Futures are about assessing the progress of on–going climate change and alerting decision–makers, involving Little Futures that include finding ever more accurate and simple–to–use indicators of climate change and their impacts, collecting ever more data to fuel indicator levels and maps, and fostering policy changes and adaptation strategies.

Relations and interactions between the three micro–regimes

Our inquiry accounted for the co–existence of three 'micro–regimes' of anticipation that seek to anticipate the evolution of forests under a changing climate. We found that researchers in the three micro–regimes belonged to different scientific institutions and published in different scientific journals, suggesting that the three 'micro–regimes' operate in relative isolation from each other. The first micro–regime spans across traditional agriculture and forest science and management institutions such as INRA, IRSTEA or ONF, whereas the second one unfolds in more basic research institutions such as CNRS. The third 'micro–regime' involves both traditional forest research and management institutions (INRA, IRSTEA, and ONF) and policy–making institutions (the French Ministry of the Environment). Bibliometric analysis also shows that researchers in the first 'micro–regime' publish in academic forest science and management such as, for example, *Annals of Forest Science* or *Forest ecology and management*, while researchers in the second 'micro–regime' publish in ecology and global change journals such as *Global Change Biology* or *Ecological Letters*. In the third 'micro–regime', they publish both in genetics or population biology journals such as *Journal of Applied Entomology* or *Insect Biochemistry and Molecular Biology*; they also publish papers in more applied forest management journals and reports for the French Ministry of the Environment.

However, certain scientists moved from the first to the second 'micro–regime' as, after being initially trained in forestry schools—such as the National School of Water and Forests in Nancy—they found positions in ecology laboratories in research institutions such as CNRS. This is the case of a forest ecologist trained in the French National School of Forestry, who now develops fundamental ecological research into tree competition, which falls under the second 'micro–regime'. Other researchers also moved from the first to the second 'micro–regime' as they became more aware of the uncertainties and limits related to running big correlative models. This is, for instance, the case of a forest modeller who works on the evolution of fire risks in a changing climate. While his research was initially focused on developing correlative and statistical models, he progressively became convinced that he could not extrapolate future conditions by relying solely on past and current data, because the system would not have the same behaviour if the extreme events started to become more frequent. As a result, he stopped making predictions and started developing research on more basic processes. Another modeller started his scientific career by working on trees' large–scale distribution in various forests and using correlative models and inventory data sets—a research activity that falls under the first micro–regime of anticipation. However, during his career, his interest shifted to theoretical ecological processes such as the functioning and evolution of the diverse tree species in a given forest and he began studying the general link between an ecosystem's diversity and its stability, then meeting the approaches favoured under the second 'micro–regime'.

A number of researchers also criticized the first 'micro–regime' as 'fashionable' but not robust enough in scientific terms. They meant that too many variables and data put together as models are run in order to obtain long–term previsions with little attention to the precise biological mechanisms and diversity at stake. Importantly, while researchers in the first 'micro–regime' construct the future as the follow–up to both the past and the present, researchers in the second 'micro–regime' consider the future to be disruptive. Therefore, they doubt models and simulation based on past data series to be able to properly predict the future. They think that there are things and rules that just cannot be known in advance, also implying a different vision of which sort of knowledge is worth developing, as the head of an important research centre in ecology in France explained:

> I'm not able to model the rules of carbon allocation in a tree: how much carbon is allocated to the roots, to the trunk, or some other thing? I'm not able to give an equation and say: This is how it works. Hence, I'm unable to make a prediction. Anyone who predicts the forest productivity or the timber stock in the 2100s if rainfall drops by 30% relies on current carbon allocation rules. We know that these rules will change, but we don't know how to model it. It doesn't prevent us from making predictions, but what is their validity? I don't know.

He also suggested that the recent development of the first 'micro–regime' was strongly driven and supported by managers and decision–makers' high expectations in the capacity of forest science to produce long–term predictions on the future of forestry. In particular, he criticised the increasing development of integrative, statistical models producing maps of species distribution at very long term, for instance, 2100. To him, these correlative statistical models produce "beautiful maps" but fail to address fundamental scientific issues such as how ecosystem functioning would be impacted by a changing climate. He argued that modelling practices should not be taken as a scientific result *per se*: "So, we have a model, and that's it. We fuel it with anything and it outputs something. A model always gives you some result." Instead, models should support the search for scientific answers to problems regarding ecological mechanisms.

Researchers working under the second 'micro–regime' thought that they were "closer to the biological reality" by developing models that take into account a very limited set of variables in a very precise manner:

> I reduce everything to one parameter: quality Q. But in fact, the precision with which I calibrate my "black box" takes into account the chemical diversity of the species... while the big categories [used in statistical models] do not take it inaccount. So, they told me that my research is a "black box," but it is *their* research that is a "black box"! (A forest ecologist)

The friction between the first and second 'micro–regimes' is thus linked to their visions of which sort of scientific advancement matters, i.e. to different visions of what future knowledge agendas are worth developing and to different ideas of what "good" forest science is.

We also found a range of cases of combination and collaboration between researchers from the first and second 'micro–regimes' of anticipation, such as in the case of the CarboFor research project (2002 – 2004) (Loustau, 2004). The CarboFor project was the first French scientific project to develop integrative statistical models based on IPCC scenarios in order to foresee the impacts of climate change on the distribution of trees over the long term. As these models progressively appeared as being not precise enough, a second research project, Qdiv, aimed to improve the representation of forest ecological functioning. This project led to comparing and integrating the correlative models developed in the first micro–regime (such as BIOMOD: Thuiller, 2003) and the process–based models developed in the second micro–regime (such as Phenofit: Chuine, 2000). A third research project, Climator, was eventually developed in order to apply the results of this integrative modelling approach to the production of standardised guidelines for agriculture and forestry. Its results were translated into a Green Paper aiming to help forest and agriculture managers and policy–makers anticipate the adaption of the timber and paper production.

Overall, we found that researchers' commitments to 'micro–regimes' of anticipation are partly related to, and dependent on, the types of relations they have with forest managers. Their relations with forest managers should not be considered end–products of their research (as researchers disseminate scientific results to managers in the form of guidelines, advice or technical support) but also as determining the type of anticipatory research agenda and practices that researchers are developing (Granjou and Mauz, 2012). For instance, having close relations with research managers will provide resources (i.e. funding, project partnerships) for developing research agendas and activities aiming to predict the future conditions of forestry and to help forest managers (i.e. commitments to the first micro–regime). On the other hand, researchers committed to the second 'micro–regime' (anticipating future tree biology) and working in fundamental research institutions usually have less close relations with forest managers, who are not that interested in understanding the very basic ecological mechanisms of trees in a changing climate. Eventually, researchers committed to the third 'micro–regime' (monitoring forests as indicators of climate change) have developed close relations with policy–makers (especially from the Ministry of the Environment via the ONERC), and those relations help provide resources that fuel the agenda of monitoring forests as indicators of climate change. As a result, researchers' anticipatory agendas and practices tend to co–evolve with the nature and proximity of their relations with managers: their various relations to forest managers are an important part of the shaping, stabilization and evolution of their commitments to 'micro–regimes' of anticipation.

Conclusion

Overall, our results show that 'climatisation' (Aykut et al.., 2017) occurred in forest science and management not only at a discursive level—meaning the integration of the topic of climate change into forest science agendas and discourses—but also within organisations, collaborations and material infrastructures of knowledge, especially practices of data collecting and modelling (i.e. the network of forest observation sites called RENECOFOR, and the platform of models CAPSIS). The 'climatisation' of forest sciences also involves increasing exchanges and collaborations between forest science, ecology and climate science and the models that were previously developed in isolation in each of those fields. Yet, our fieldwork also points to the co–existence of a plurality of research agendas dedicated to anticipating the future of forests under a changing climate. These research agendas differ in two ways: they develop forest–driven vs. climate–driven science on one hand, forestry–oriented versus ecological and biological process–based research on the other hand. Our results do not only suggest how groups of environmental scientists, experts and decision makers hold various and potentially conflicting views regarding which research fields, predictive technologies and anticipatory governance are worth developing to produce sound science insights into environmental futures, they also highlight the embedment of the predictive models developed by forest scientists and experts within various and partly conflicting politics of environmental anticipation, including concerns for adapting forestry practices to future climates ('micro–regime' 1), for anticipating the ecological resilience of trees and forests ('micro–regime' 2) and for mapping and preparing the advent of climate change and its impacts ('micro–regime' 3). Forest researchers' commitments to the 'micro–regimes' and their various visions of what knowledge matters not only correspond to their professional trajectory and affiliation (i.e. belonging to fundamental or more management–oriented research institutions). They also tend to co–evolve with the nature and proximity of their relations with forest managers, as, for instance, having close relations with research managers will favour research agendas and activities aiming to predict the future conditions of forestry and help forest managers. On the other hand, researchers committed to the second regime have fewer close relationships with forest managers, while researchers committed to the third 'micro–regime' have rather developed relationships with national policy–makers in the field of forest, agriculture and environment.

Let us briefly return to the recent literature on the emergence of a 'regime of anticipation' that challenges previous models of predicting the future through the calculations of risk probability based on past data series (Amoore, 2013; see also Aradau and van Münster, 2013). Instead of the rise of a global, all–encompassing regime of anticipation accompanied by a new "conjectural episteme," our fieldwork in the case of French forest science and management suggests that the transformations of forest science and forestry practices aiming to anticipate climate change entail the co–existence of various, partly conflicting anticipatory 'micro–regimes' at work, whose goals and approaches to science, forest and the future are different. Our fieldwork suggests that Amoore's and Aradau and Van Münster's thesis does not do justice to current changes in the knowledge production practices which aim to capture environmental changes and futures. Instead of one unique way of constructing the future as an object of knowledge and action, we found several anticipatory assemblages that seek to foresee the future evolution of French forests in a changing climate, various visions of the extent to which the future is disruptive and a variety of practices and strategies for producing future–oriented knowledge. While in the first 'micro–regime', predictions rely on assembling past and present data into ever more sophisticated and integrative predictive projections, in the second regime, anticipating ecological changes requires scientists to modify the core of their models in order to predict how basic ecological mechanisms will evolve. In the third 'micro–regime', researchers aim to produce ever more accurate indicators of the on–going and future progress of climate change drawing on data collection.

Lastly, the embedment of a predictive research agenda setting within an environmental issue framework and political choices suggests that anticipatory pluralism is important in keeping a broad range of futures open to scientific and public scrutiny. In line with the new political sociology of science (Frickel et al., 2010), we argue that this plurality is essential in avoiding certain futures being completely unaddressed and unscrutinised (for instance, should the anticipation of forestry practices adaptation become hegemonic over the anticipation of tree species extinctions and 'natural' forest ecosystem destabilization). Documenting the variety of predictive scientific practices is instrumental in understanding the various and partly conflicting ways in which environmental futures are known, predicted and acted upon.

References

Adams HD, Williams AP, Xu C, Rauscher SA, Jiang X and McDowell NG (2013) Empirical and process–based approaches to climate–induced forest mortality models. *Frontiers in Plant Science* 4: 438.

Adams V, Murphy M and Clarke AE (2009) Anticipation: Technoscience, life, affect, temporality. *Subjectivity* 28(1): 246–265.

Aleixandre–Benavent R, Aleixandre–Tudó J, Castelló–Cogollos L and Aleixandre J (2017) Trends in scientific research on climate change in agriculture and forestry subject areas (2005–2014). *Journal of Cleaner Production* 147: 406–418.

Amoore L (2013) *The politics of possibility: risk and security beyond probability*. Durham, NC: Duke University Press.

Ancelin P, Courbaud B and Fourcaud D (2004) Development of an individual tree–based mechanical model to predict wind damage within forest stands. *Forest ecology and management* 203: 101–121.

Andersson J and Duhautois S (2016) Futures of Mankind: The Emergence of the Global Future. In: Van Münster R and Casper S (eds) *In the Politics of Globality: Assembling the Planet*. Oxford: Routledge, pp.106–125.

Aradau C and van Münster R (2011) *Politics of Catastrophe. Genealogies of the Unknown.* London: Routledge.

Aykut SC, Foyer J and Morena E (eds) (2017) *Globalising the Climate. COP21 and the 'climatisation' of global debate*s. London: Routledge.

Bloomfield BP (1986) *Modelling the world. The social constructions of systems.*London: Blackwell.

Callon M (eds) (1998) *The Laws of the Markets*. Oxford, UK: Oxford University Press.

Chuine I (2000) A unified model for tree phenology. *Journal of Theoretical Biology* 707: 337–347.

Chuine I and Cour P (1999) Climatic determinants of budburst seasonality in four temperate–zone tree species. *New Phytologist* 143: 339–349.

Coleman R and Tutton R (2017) Introduction to Special Issue of *Sociological Review* on 'Futures in Question: Theories, Methods, and Practices.' *The Sociological review* 65(3): 440–447.

Collins H and Pinch T (2002[1998]) *The Golem at large: What everyone should know about technology*. Cambridge: Cambridge University Press.

Cordonnier T, Courbaud B and Franc A (2006) The effect of colonization and competition processes on the relation between disturbance and diversity in plant communities. *Journal of Theoretical Biology* 243(1): 1–2.

Courbaud B, Pupin C, Letort A, Cabanettes A and Larrieu L (2017) Modelling the probability of microhabitat formation on trees using cross–sectional data. *Methods in Ecology and Evolution* 8(10): 1347–1349.

Dahan–Dalmedico A (2006) *Les Modèles du futur. Changement climatique et scenarios économiques: enjeux scientifiques et politiques.* Paris: La Découverte.

Dahan–Dalmedico A (2010) Putting the Earth System in a numerical box? The evolution from climate modeling toward global change *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 41(3): 282–292.

Decocq G, Kalaora B and Vlassopoulou C (2016) *La forêt salvatrice. Reboisement, Société et Catastrophe au prisme de l'Histoire*. Ceyzérieu (Ain): Champ Vallon.

Dufour–Kowalski S, Courbaud B, Dreyfus P, Meredieu C and de Coligny F (2011) Capsis: an open software framework and community for forest growth modeling. *Annals of Forest Science* 69(2): 221–233.

Edwards PN (2010) *A Vast machine: Computer Models, Climate Data and the Politics of Global Warming.* Cambridge and London: MIT Press.

Estiarte M, Vicca S, Peñuelas J et al. (2016) Few multiyear precipitation–reduction experiments find a shift in the productivity–precipitation relationship. *Global Change Biology* 22: 2570–2581.

Farquhar GD, von Caemmerer S and Berry JA (1980) A biochemical model of photosynthetic CO2 assimilation in leaves of C3 species. *Planta* 149: 78–90.

Frickel S, Gibbon S, Howard J, Kempner J, Ottinger G and Hess D (2010) Undone science: charting social movement and civil society challenges to research agenda setting. *Science, Technology & Human values* 35(4): 444–473.

Gea–Izquierdo G, Guibal F, Joffre R, Ourcival JM, Simioni G and Guiot J (2015) Modelling the climatic drivers determining photosynthesis and carbon allocation in evergreen Mediterranean forests using multiproxy long time series. *Biogeosciences* 12: 3695–3712.

Granjou C and Mauz I (2012) Expert activities as part of research work. The example of biodiversity studies. *Science and Technological Studies* 25(2): 5–22.

Granjou C and Walker J (2016) Promises that matter. Reconfiguring ecology in the ecotrons. *Science and Technology Studies* 29(3): 49–67.

Granjou C, Walker J and Salazar J (2017) The politics of anticipation: On knowing and governing environmental futures. *Futures* 92: 5–11.

Gustafson E, De Bruijn A, Pangle R et al. (2015) Integrating ecophysiology and forest landscape models to improve projections of drought effects under climate change. *Global Change Biology* 21(2): 843–856.

Hacking I, (1994) Styles of Scientific Thinking or Reasoning: A New Analytical Tool for Historians and Philosophers of the Sciences. In: Gavroglu K, Christianidis J and Nicolaidis E (eds) *Trends in the Historiography of Science.* Dordrecht: Springer Netherlands, pp. 31–48.

Jasanoff S and Martello M (2004) *Earthly politics*. *Local and Global in Environmental Governance.* Cambridge and London: MIT Press.

Jabot F, Faure T and Dumoulin N (2013) EasyABC: performing efficient approximate Bayesian computation sampling schemes using R. *Methods in Ecology and Evolution* 4(7): 684–687.

Knorr–Cetina K (1999) *Epistemic Cultures. How the sciences make* knowledge Cambridge and London: Harvard University Press.

Korzukhin MD, Ter–Mikaelian MT and Wagner RG, (1996) Process versus empirical models: which approach for forest ecosystem management? *Canadian Journal of Forest Research* 26(5): 879–887.

Kwa C (2011) *Styles of knowing. A new history of science from ancient times to the present*. Pittsburgh: University of Pittsburgh Press.

Lagarrigues G (2016) Effects of environmental gradients on demographic processes and the durability of the uneven–aged forest management: application of Approximate Bayesian computation to improve simulation models. PhD Thesis, Université Grenoble Alpes, France.

Lafond V, Cordonnier T, Zhun M and Courbaud B (2017) Trade–offs and synergies between ecosystem services in uneven–aged mountain forests: evidences using Pareto fronts. *European Journal of Forest Research* 136(5/6): 997–1012.

Law J and Mol A (eds) (2002) *Complexities. Social Studies of Knowledge Practices*. Durham and London : Duke University Press.

Legay JM (1997) *L'expérience et le modèle. Un discours sur la méthode.* Paris : INRA Editions.

Lempereur M, Limousin JM, Guibal F et al. (2017) Recent climate hiatus revealed dual control by temperature and drought on the stem of Mediterranean Quercus ilex. *Global Change Biology* 23(1): 42–45.

Leslie PH (1948) Some further notes on the use of matrices in population mathematics. *Biometrika* 35(3/4): 213–245

Loustau D (2004) *Final report of the project CARBOFOR. Carbon sequestration in the big forest ecosystems in France. Quantification, spatialization, vulnerability and impacts of different climatic and forestry scenarios*. INRA/CNRS, France, June 2004.

Mackenzie D, Muniesa F and Siu L (2007) *Do Economist Make Markets? On the Performativity of Economics*. Princeton: Princeton University Press.

Michael M (2017) Enacting Big Futures, Little Futures: Toward an ecology of futures. *The Sociological Review* 65(3): 509–524.

Misson L, Rocheteau A, Rambal S, Ourcival JM, Limousin JM and Rodriguez R (2010) Functional changes in the control of carbon flux after 3 years of increased drought in a Mediterranean evergreen forest? *Global Change Biology* 16(9): 2461–2475.

Mol A (2002) *The Body Multiple: ontology in medical practice*. Durham and London: Duke University Press.

Morin X, Fahse L, Scherer–Lorenzen M and Bugmann H (2011) Tree species richness promotes productivity in temperate forests through strong complementarity between species. *Ecology Letters* 14(12): 1211–1219.

National Observatory on the Effects of Global Warming (2010) *Climate Change's indicators' list*. Report to the French Ministry of the Environment.

Perez–Ramos IM, Ourcival JM, Limousin JM and Rambal S (2010) Mast seeding under increasing drought: results from a long–term dataset and from a rainfall exclusion experiment. *Ecology* 91: 3057 – 3068.

Pestre D (2003) Regimes of knowledge production in society: Towards a more political and social reading *Minerva* 41(3): 245–261.

Renecofor's assessment committee (2006) *Renecofor network's evaluation. Suggestions for the future.* 14th November 2006

Roques A (2015) *Processionary Moths and Climate Change: An Update*. Springer Netherlands.

Rossi JP, Garcia J, Roques A and Rousselet J (2015) Trees outside forests in agricultural landscapes: spatial distribution and impact on habitat connectivity for forest organisms. *Landscape Ecology* 31(2): 243–254.

Sarewitz D, Pielke RA and Byerly R (2000) *Prediction: science, decision making and the future of nature.* Washington DC: Island Press.

Shinn T (1999) Change or mutation? Reflections on the foundations of contemporary science. *Social Science information* 38(1): 149–176.

Skrydstrup M (2017) Envisioning the future by predicting the past: Proxies, praxis and prognosis in paleoclimatology. *Futures* 92: 70–79.

Tavory I and Eliasoph N (2013) Coordinating Futures: Toward a Theory of Anticipation. *American Journal of Sociology* 118(4): 908–942.

Thuiller W (2003) BIOMOD – optimizing predictions of species distributions and projecting potential future shifts under global change. *Global Change Biology* 9(10): 1353–1362.

Vieilledent G, Courbaud B, Kunstler G, Dhôte JF and Clark J (2010) Individual variability in tree allometry determines light resource allocation in forest ecosystems: a hierarchical Bayesian approach. *Oecologia* 163(3): 759–773.

Notes

1 Another example is the creation of the A-FORCE network ("Adaptation des FORêts au Changement ClimatiquE" i.e. Adaptation of Forests to Climate Change) which comprises 15 research institutes, engineering schools and the ONF, with the aim of promoting and supporting climate change knowledge transfer from research laboratories to forest managers. In particular, A-FORCE seeks to avoid misunderstandings of modelling and simulation results and their uncertainties among forest managers.

2 The founder of SAMSARA, who works at IRSTEA, is in frequent contact with the National Forest Office in order to both collect data and to transmit SAMSARA results to forest managers with the goal of improving forestry strategies. SAMSARA is also used for training forest practitioners.

3 Today, large controlled chambers for the measurement of gas exchanges between plants and the environment exist in most major universities and agronomic institutes, for example New Zealand's Biotron, the Bioklima project in Norway, the ecotron projects in Germany and in Belgium. Many more ecotron-like facilities are in progress.

Reckoning Resources: Political Lives of Anticipation in Belize's Water Sector

Sophie Haines

Institute for Science, Innovation and Society, University of Oxford, UK
sophie.haines@insis.ox.ac.uk

Abstract

Participants in Belize's water sector encounter challenges in identifying and living within shifting environments, and in conducting the work of expectation given ambiguities in rainfall patterns, historical records, institutional resources and political interests. Policymakers, scientists and practitioners generate and organise different kinds of foreknowledge as they anticipate future quantities, qualities and distribution of water, amid questions about the patterning of expertise and the nature of water as a resource. I present three ethnographic vignettes to address: the navigation of nonknowledge in water policy implementation; the frictions that arise in modelling workshops where trainees generate data-driven maps of future environments; and the situated sensing of environmental change. Building on a concept of 'reckoning' that highlights cross-cutting technical, relational, political and affective dimensions of meaning-making, I situate these foreknowledge practices in the socio-material contexts of environmental perception, economic development, and the political lives of anticipation.

Keywords: knowledge, prediction, anticipation, water, data, Belize

Introduction

This paper addresses how foreknowledge about water resources is generated, recognised and acted upon (or not) in efforts to anticipate environmental conditions and implement national policies for water resources management and climate change adaptation in a coastal country in central America. For those involved in operational decision-making and longer-term strategic planning in Belize, current negotiations of knowledge and policy around future water resources are shaped by histories of organisational development, location and responsibility. For example, the sites of weather stations and status of data archives are influenced by colonial legacies; also important are the technological capacities of meteorological and hydrological services, current political priorities, and the roles of international networks, donors and experts. I build on theoretical work in STS and anthropology to analyse the technical, relational, political and affective dimensions of quotidian knowledge practices of scientists and practitioners in public, private and NGO sectors as they craft credible futures, and encounter questions about the limits of science and the patterning of expertise and authority.

The scales and horizons of knowledge production and decision-making about environmental futures extend spatially across pixels, grid squares, communities, organisations, districts, river basins, coastlines, nations and regions, and temporally from 'real-time' to multi-decadal frames. While global infrastructures are at the heart of endeavours to model the atmosphere (Edwards, 2010), the effects of atmospheric changes are experienced as the medium of everyday life. Ingold (1993) contrasts a characterisation of the environment as a globe, separated from the observer and apprehended in a detached way (as abstracted in the techno-scientific visualisation of a general circulation model), with a perception of the environment as a sphere, known from within, for example as weather conditions experienced phenomenologically (Ingold, 2010). Other scholars have sought to complicate the duality of such models, arguing that while the 'global' pretends to a view from nowhere, or everywhere (Haraway, 1988; Hulme, 2010), it is itself constructed and stabilised through contested political-economic, infrastructural and sociocultural efforts that are locally embedded (Blok, 2010). Knowledge infrastructures can adopt different models of participation that afford different scope for empowerment and marginalisation (Jalbert, 2016). Scientists are also citizens and the boundaries between 'the lab' and 'the world' are not clear-cut (Monteiro and Rajão, 2017); we come to know the atmosphere through diverse senses, measurements, practices and comparisons (Choy, 2012).

A growing literature in STS and related disciplines has explored these imbrications of perspective and identity in the production of predictions and future imaginaries: how states mobilise anticipatory knowledge to support policy directions (Nelson et al., 2008); and how discourses of risk management and other ways of rendering the future present and actionable are interlaced with culture, society and politics (Anderson, 2010; Beck, 1992; Demeritt, 2006; Douglas, 1992; Hulme, 2009; Rayner, 2007). Ethnographic studies of environmental forecasting have addressed how meteorologists and climate scientists position themselves and their models and forecasts (and how they are positioned by others) with respect to uncertainty (Daipha, 2015; Fine, 2007; Lahsen, 2005; Shackley, 2001; Taddei, 2012); and elucidated the social lives and 'performativity' of foreknowledge, and the challenges of accountability and equity for decisions about resource allocation (Barnes, 2016; Broad et al., 2002; Hastrup and Skrydstrup, 2013; Taddei, 2013). In line with the aims of this special issue, I seek to engage the political dynamics of producing and negotiating foreknowledge; I do this by focusing on processes of 'reckoning' future quantity, quality and distribution of water to illuminate the sociality and materiality of anticipation in a context where climate change and water management are explicit policy challenges.

The paper proceeds by presenting the research setting and methodology. I then introduce 'reckoning' as a lens for exploring knowledge production and negotiation under uncertainty, before applying it to three empirical vignettes. In the first, I address the shifting political grounds on which institutional decision-makers are *reckoning with(out) data,* and how knowledge and nonknowledge for policy are organised and critiqued. I then examine experimental practices of hydrological modelling — ways of *reckoning with models* — involving the friction of negotiating standards, expectations and meanings across observed, imagined and simulated worlds. The third vignette considers *affective reckoning,* foregrounding how people orient themselves given shifting points of reference that can trouble ideals of integration, translation and management. These modes of reckoning — involving formal calculation but also opinion and judgment — are not necessarily mutually exclusive. They can however demonstrate different emphases with respect to the resources they value and put to work to anticipate future water, given incomplete knowledge of future weather and climate, limited/unevenly distributed material resources, and pressure to make timely decisions. I discuss how 'reckoning resources' can thus be understood in multiple ways: as the challenge of confronting uncertain resource futures; as the practice of calculating or estimating future water; and as the tools (including data, models, senses and expectations) that are put to work in anticipatory practice. I conclude that paying attention to modes of reckoning environmental futures, and to their cross-cutting technical, relational, political and affective dimensions, is a way to foreground the socio-material conditions under which foreknowledge may be made meaningful across contexts.

Research setting and methodology

Belize is recognised to be particularly vulnerable to impacts of climate change and variability at various timescales, with potential implications for crucial economic sectors of tourism, agriculture and fisheries (CaribSave, 2012; Richardson, 2009). Many of these anticipated impacts directly or indirectly involve changes in quantity, quality and/or distribution of water, for example sea-level rise, saline incursion, and variations in patterns of rainfall, evaporation, flooding, and coastal erosion. Although it is located mostly on the Central American mainland, the country is recognised within the UN system as one of the 'Small Island Developing States' (SIDS) that share climate change vulnerabilities, among other characteristics. With reference to Hau'ofa's (1993) vision of a 'sea of islands', Lazrus (2012) has explored how island states and communities facing climate change are not as isolated as conventionally assumed: Belize is no exception, in light of not only the transboundary watersheds it shares with Mexico and Guatemala, but also its social and infrastructural connections across and beyond the Americas and Caribbean. Belize is thus a productive site for exploring different ways of knowing and potentially acting on future water across times, spaces and institutions.

The empirical research underpinning this article was undertaken in Belize over three months in 2014, as part of a wider interdisciplinary project examining the usability of weather and climate forecasts for resource and hazard management in different national contexts and sectors. The study was designed to investigate social/institutional dimensions of forecast use and non-use, by paying attention to how (potential) forecast users situated themselves in organisations and decision processes; how they gauged success; how they did (or did not) access and use weather and climate information; and how they prepared (or did not) for future conditions. The study also sought to examine forecasters' definitions of success, and their relationships with their technical tools and with other decision-makers. To these ends, ethnographic observations and semi-structured interviews provided insights into forecasters' and users' lived experience; their views of their roles, opportunities and constraints; and the meanings they attached to their decisions and interactions. Throughout the research and analysis, relating insights from these two methods to each other and to materials including mission statements, forecast products and policy documents enabled validity checking and identification of patterns and differences in practices, perspectives and priorities.

In Belize, where I have conducted anthropological research since 2006, the ethnographic study included (participant) observation of forecasting centres, training sessions and planning activities in the water sector, including three operational shifts at the National Meteorological Service, a two-day hydrological modelling workshop, and a coastal planning seminar. This afforded practical insights into participants' applied and embodied knowledge, and opportunities for learning as situations unfolded. I also conducted interviews with 60 water sector participants, including environmental and climate scientists, weather forecasters, utility suppliers and regulators, and government/NGO staff in agriculture, natural resources, and emergency management.[1] The interviews were based on a shared but flexible protocol to enable in-depth discussion of issues important to participants, coverage of topics that may have been elusive during only three months' of ethnographic research, and opportunities for pursuing and validating emerging themes from observations and other interviews. They also offered efficient use of time when interviewing busy professionals (Bernard, 2011: 157-158). I typed daily fieldnotes, and recorded/transcribed interviews for qualitative interpretive analysis. Working with these methods helped clarify how future water quantities, qualities, and distributions are reckoned — and reckoned *with* — in Belize.

Limitations of the methods include the relatively short duration of the ethnography, which precluded following how forecasting or training endeavours developed over seasons and years. The focus on people who were willing to participate in interviews and observed events introduces potential for self-selection by those with a particular orientation to forecast production and use. I was attentive to this, particularly as Belize's low population means that there are a limited number of actors involved with relevant work; I developed a network of key informants to help me connect with participants who may have been less visible or engaged. This study was designed primarily to address forecast production/use in professional contexts; as such, it did not thoroughly examine the roles and perceptions of wider 'publics' also affected by the information and decisions in question. When interviewing people in their professional capacity, their responses may reflect official lines rather than personal perspectives. Both are interesting: the flexibility and rapport afforded by the semi-structured format, and the insights from ethnographic encounters and observations (sometimes involving the same interviewees) helped build a picture of what people do as well as what they report.

Reckoning (with) resources

In this paper, I use the notion of 'reckoning' to explore how scientists, public servants and other practitioners in Belize's water sector navigate shifting atmospheres, temporalities, values and commitments as they look to the future of water resources and hazards. I find the term's multivalence useful for thinking through different approaches to measuring and framing (im) precision or (un)certainty: reckoning formally means to count up or calculate; it also refers to estimation, expectation, trust, opinion or judgment. Along with these more or less direct paths to knowledge, its allusions to settling (accounts), tackling (challenges) and envisioning (possibilities) span temporalities and are suggestive of the resources, reputations and livelihoods that are at stake as practices of water assessment and prediction move between objectivity and interpretation, closure and ambiguity.[2] These processes of abduction — of "tacking back and forth between futures, pasts and presents… turning the ever-moving horizon of the future into that which determines the present" (Adams et al., 2009: 251)[3] — have affective power (see also Zaloom, 2009) and can influence action in the face of uncertainty, imprecision and/or ignorance.

Anthropologists Kockelman and Bernstein (2012) discuss reckoning time as a particular approach to framing temporality (distinct from metricality, performativity, or worldview) that "foregrounds the when and how long of an event [and] focuses on the social, semiotic and material resources we have for telling the time" (Kockelman and Bernstein, 2012: 324). It involves triangulation: using privileged periods of repetition and points of orientation to size and order the *event to be reckoned* relative to the *event of reckoning* (Kockelman and Bernstein, 2012: 326). Kockelman and Bernstein's (2012: 336) analysis highlights technical, relational and political dimensions of reckoning, and thus counteracts what they describe as a "pervasive theoretical insistence on independent, abstract, empty, homogeneity [that] obscures the dependent, concrete, full, heterogeneity of our actual everyday situated modes of temporal being" . This argument frames their discussion of 'portability' — the extent to which meanings produced through different semiotic technologies (such as language, clocks and calendars) can be understood and applied across historical and cultural contexts — as something that varies according to the simplicity of the technology, the knowledge shared by speaker and addressee, the relative sizes of the populations that control and reckon with privileged points, and hierarchies of credible measurements. Portability, they argue, relies not on absence of context, but on relations and mutual knowledge. The notion resonates with analyses by historians and scholars of science and technology who explore the production and circulation of immutable mobiles (Latour, 1987: 227), boundary objects (Star and Griesemer, 1989) and different modes and conditions of objectivity (Daston and Galison, 2007; Porter, 1995).

Reckoning can be applied to domains other than time.[4] While I do not apply their thorough linguistic analysis, I draw on Kockelman and Bernstein's (2012) understanding of reckoning as a useful point of entry to consider how attempts to make measurement meaningful and establish shared understandings about future water are at once technical, relational and political. This foregrounds how diverse resources are mobilised by scientists and practitioners to know about quantity, quality and distribution of water. The problem of anticipating its *future* characteristics adds further challenges. I situate these lines of enquiry in conceptual frameworks that see the implications for those reliant on the resources in question as bound up with an epistemic environmental politics whereby power dynamics among different ways of knowing can make particular futures more or less salient or imaginable (Groves, 2017; Jasanoff, 2004; Taddei, 2013), and that call attention to the role of material forces in influencing knowledge-seeking and world-making (Vaughn, 2017). Who has authority to determine legitimate units and points of reference? What instruments and processes are used to assess current and future resources? What are the implications for people whose lives are bound up with the environments in question? I see these three provocations mapping respectively onto three sites that Orlove and Caton (2010) propose for anthropological analysis of water: *water regimes* (institutions, rules and tools of water governance); *watersheds* (hydrogeographical units of assessment and intervention); and *waterscapes* (experiential entanglements of place, ideology and meaning). These respectively underpin the following three empirical vignettes, which discuss shifting governance, modelling practice, and sensory experience of reckoning resources in Belize.

Reckoning with(out) data

While environmental variability is not new to Belize, some anticipated changes are now being framed as existential threats, for example in regional assessment and policy documents that highlight rising sea levels, coastal erosion, saline intrusion, escalation in intensity of extreme weather events, and disruptions in rainfall and fresh water supplies (CCCCC and CDKN, 2012). Official classifications and measurements of baselines and extremes make — and remake — environmental resources and hazards through scientific practice and policy decisions that authorise particular technologies, motivations and objects of concern (Bond, 2013). In this section I examine how scientists and policy staff discussed data in relation to a shifting *water regime* (Orlove and Caton, 2010), paying attention to the power dynamics of knowledge and nonknowledge.

Water in Belize is vital for drinking and sanitation, for agriculture, tourism, cultural practices, place-making, fisheries, coastal development, hydropower, and conservation. It is thus a matter of concern for local and national governments, nationalised and private companies, NGOs, consultants, universities and research institutes, communities, indigenous groups, and individuals. Relevant responsibilities and knowledge are distributed among many different entities. For example, municipal supply, sewerage and some rural supply is undertaken by Belize Water Services Limited, with other rural supply managed by village boards under government oversight. Drainage infrastructure is largely the responsibility of municipalities. National electricity distributors and private dam operators manage water for hydropower production. The National Emergency Management Organisation (NEMO) oversees flood events and other emergencies. Surface water data is collected and collated by the government Hydrology Unit; the Met Service records and forecasts precipitation, and undertakes climate science-policy related duties (e.g. representing the country in UNFCCC processes) alongside the nascent National Climate Change Office, in collaboration with regional and global bodies including the Caribbean Community Climate Change Centre (CCCCC), the Caribbean Institute of Meteorology and Hydrology, and the World Meteorological Organisation (WMO).

Shifting regimes

There has to date been little control of the abstraction and use of water in Belize. The National Integrated Water Resources Management policy, presented by government in 2008, promoted a vision that formally vested water resources in the state and gathered responsibilities for water resources management under a national commission. The rationale for instituting Integrated Water Resources Management (IWRM) in Belize was that past water policies had been too "parochial" and there was minimal understanding of climate change impacts (BEST, 2008). The policy was followed in 2009 by a National Adaptation Strategy for the water sector. Citing the IPCC, the strategy's technical review identified trends in temperatures, precipitation regimes and extreme weather as motivations for improving water governance to benefit present and future generations,[5] and noted that the lack of a comprehensive water monitoring programme in Belize precluded quantification of threats to water (BEST 2009). In 2011 the government enacted a National Integrated Water Resources Act. At the time of research, however, responsibilities remained fragmented, with water regulated for example as industry by the Public Utilities Commission, as effluent by the Department of Environment, as drinking water by the Ministry of Health; regulation of water as 'raw' natural resource was expected to be the domain of the prospective National Integrated Water Resources Authority (NIWRA). The development of NIWRA (a process dating back to at least to the 2003 reactivation of a Pro Tempore Water Commission that was revived and formalised in the 2011 Act) was a significant dimension of shifts in the water regime underway during the research period (FAO, 2015).

IWRM is an internationally instituted approach that, according to the Global Water Partnership, "promotes the coordinated development and management of water, land and related resources in order to maximise economic and social welfare in an equitable manner without compromising the sustainability of vital ecosystems" (GWP, 2000). The watershed is deemed the fundamental unit of analysis and intervention, with emphasis on scientific management and the development of an international network of experts and donors (Caton, 2007). As part of efforts to institutionalise IWRM in Belize, the Hydrology Unit had been moved in 2012 from the Met Service headquarters near Belize City, to the then-Ministry of Natural Resources and Agriculture in Belmopan, where it was destined to act as the NIWRA secretariat.[6] At the time of research, the project to develop NIWRA as an autonomous body to promote stable and independent oversight for water issues was progressing under climate change funding from the European Union. A major NIWRA objective was to create datasets to reduce uncertainty about current and future water resources, and ultimately produce a national 'masterplan' based on assessments of water availability and stakeholder demand, which could be used as a management tool to inform licensing and fees.

Throughout my research, Belizean scientists and policy officials frequently framed the problem of water management in terms of insufficient data. In doing so they emphasised potentially reducible epistemic uncertainty, suggesting that water is a theoretically calculable — and thus governable — resource (Scott, 1998): a view that seems well-aligned with IWRM's science-led approach. A participant involved with NIWRA development was not the only interviewee to lament: "we can't manage what we don't know." I heard complaints that data were lacking in quantity — over time (historical records do not go very far back and there are gaps in the records) and space (there are few gauging stations, issues with accessibility for reading/maintenance, and a lack of groundwater monitoring) — and also in terms of quality, related to challenges of maintaining and calibrating instruments and relying on volunteer observers.[7] Discussing themes of evidence-based decision-making, a water expert at the utilities regulator expressed frustration about being caught between the potential costs and benefits of a precautionary approach, implying the need for more data to help overcome this impasse:

> If you don't know the amount of resource that is there, you cannot manage it properly. You probably might be over cautious, and so you're not getting what you want. Or you might be negligent and aggressive and causing it to go at-risk.

The implication here is that having more data will allow measurement and quantification of the available present resource (at the point of reckoning), which in turn would facilitate reckoning of future amounts (given an understanding of likely trajectory), and hence the ability to influence these future amounts via management of abstraction and use. The situation of not-knowing current quantities was presented as disorientating: a sense that without a stable baseline at the point of reckoning, the enterprise would be futile.

Taming nonknowledge

In conversations with government policy staff, I asked how they go about their work in this situation. The following quotation demonstrates ambivalence: the official opens with confidence in existing knowledge, before noting that there are 'known unknowns' that trouble the ability to project with precision:

> We are aware of what our aquifers are. We know where they are… there are regional maps with the transboundary aquifers, and so we know what the source is... With satellites, there are calculations that you can do that make the projections. But the calculations are based on unknowns, and it's better when they're based on knowns... So, it's really a case of we calculate based on what we know, but we know that we don't know the extent.

The official gestured across a desk piled high with papers, and explained that they use frequent reporting, data mining and iterative decision-making to face these limitations:

> We project, we document, we adjust; we project, we document, we adjust. And we go through that iterative process. And so, we're tending towards infinity. We are getting closer and closer to be able to effectively say, "OK. Well, based on our experience this is going to be equal to that."

With its heavy documentation and auditing — in the official's words, "a lot of statistics" — the process is presented as a dynamic form of reckoning that applies different technologies of observation, calculation and comparison using satellites, maps and statistics: an effort at triangulation using the situation at the 'speech act' of the projection (based on available 'current' data from e.g. gauges and satellites), and the calculations of expected change to project a future picture of water quality, quantity and distribution. This is subsequently 'ground-truthed' against observations and adjusted so that it tends towards an imagined end point where expectation meets eventuality and the problem is tamed. The government had recruited consultants to conduct a hydrological assessment of the southern portion of the country, which was intended to provide the background information that would enable projection into the future, taking into account population expansion and climate change. In these early stages of IWRM implementation, a key government goal was to develop a licensing regime (focusing on industrial abstraction), thereby instilling the IWRM principle of water as an economic good. Interviewees explained that, given the data shortage, the government was allowing a 'grace period' whereby licenses were being administered but not yet charged as part of the environmental impact assessment process, with the condition that licensees install flow meters and submit this data along with their abstraction rates to the ministry. Ministry staff would check the actuals and may change the conditions of the annual license accordingly. This can be read as a performative exercise: piloting the programme as a means to procure baseline data, which will then facilitate checking, adjustment and — ideally — scientific management. At stake is the conceptualisation of water as a calculable and governable economic good.

Political data ecologies

The scientists I encountered frequently framed the data problem as a problem of distribution (and thus politics). Their descriptions of the knowledge ecologies of sharing, concealing, overlooking or denying data and contextual information support an analysis of nonknowledge or ignorance as not a simple opposite of knowledge or lack of competence, but something that can be actively constructed, strategically managed, contested or mobilised to allow certain forms of work or life to continue (Anand, 2015; Dilley, 2010; McGoey, 2012; Power, 2007; Rayner, 2012). As Mathews (2014: 82) notes in his study of Mexican forestry, nonknowledge can be "tamed as calculable uncertainty, or alternatively transformed into ontological indeterminacy, scandals, and stories of corruption". Indeed, while policy workers — focused on turning available data into management — described processes of reckoning-as-taming, many scientists used discussions of the challenge of reckoning without sufficient data as avenues for political critique. While most stated a commitment to the theoretical calculability of water resources — and the principle of science as a "less contestable" mode of decision-making — institutional/political factors made them doubt that improved data availability and analysis would be practically possible and, even if they were, whether they would lead to real change.

Current and former public servants and scientists spoke of institutional barriers to cumulative, consistent and contextualised bodies of knowledge over both time and space. In terms of time, they alluded to losses of institutional memory when units and programmes were moved in civil service reorganisations, or when leaders were replaced according to electoral cycles. The short duration and external control of many projects impedes continuity; important documents may be removed overseas or otherwise rendered unavailable for future reference. With a shortage of trained experts employed by government departments, the small pool of available consultants relied on for much research and policy drafting raised questions for some about accountability, conflicts of interest, institutional memory and duplication of efforts. For respondents raising such objections, the problem was not so much that datasets had not been produced, but that they could be disregarded by leaders seeking to make their mark with new projects.

The spatial dimension was reflected in complaints about the reluctance or inability of organisations and individuals to share data, both within Belize and across national borders: the latter was of concern given that more than half of Belize's population rely for their potable supply on water that originates in neighbouring countries (BEST, 2009: 9). A framework was in place for cooperation with Mexico, but diplomatic relations between Guatemala and Belize remain sensitive and formal data sharing agreements were lacking. Belizean scientists and technical staff reported exchanging information through informal networks and meetings with counterparts at international conferences. A commonly cited example of data-sharing obstacles within the country was the physical and institutional removal of the Hydrology Unit from the Met Service. Several respondents expressed frustration that this had distanced the Unit from cognate work, making data sharing difficult, and flood forecasting effectively impossible. These issues were compounded by the situation that at the time of research there was no hydrologist in the Hydrology Unit — a case frequently deployed as an archetype in commentaries about the problems of managing public sector expertise and human resources, and government apathy to environmental science. Scientists and regulators raised the challenge of overcoming a "mindset" or "culture" in Belize that views data as property and political capital. In the words of a former government scientist: "Here in Belize, people tend to hold on to information as if it's gold."

Scientists and practitioners attributed political apathy about future resource pressures to a false sense of security related to Belize being a country with abundant water and low population density, but also to incentives for politicians to maintain discretion over resource allocation and keep debates about distribution out of the public sphere, particularly given connections between water and land, which is of fundamental importance in Belizean politics (Haines, 2012, 2018; Grandia, 2009; Wainwright, 2008; Wilk, 1997). Thus, there was doubt that even with data collected, made available and translated into policy, there would still be the problem of enforcement — again, frequently described as typical of Belize, or the Caribbean (see also Medeiros et al., 2011). This was seen as not only a symptom of scarce governance resources but a strategic performance of legitimacy for donors, stopping short of practical change. As an environmental scientist explained:

> We end up with these beautiful plans… That's what [funders] want to see -- policies change; they want to see legal frameworks change. But those are no good if there is no enforcement and if they are not implemented… Our policies are beautiful, actually... They say nice things. We've got to close that loop — from policy to legislation to enforcement …

In the current context, the scientist thought that taming the issue through specific and quantified economic arguments might have some influence, but in the longer term the situation would be unlikely to change without a shift in political culture from factional clientelism to more transparent democracy.

These critiques move focus from nonknowledge as calculable uncertainty towards narratives of the control of information by institutions for which such knowledge would be uncomfortable or inconvenient (Mathews, 2014; McGoey, 2012; Rayner, 2012). While conflicts about data availability and sharing are of course not restricted to this location, the common identification — by Belizeans — of such an attitude as a national phenomenon corresponds with a broader vernacular critique of political elites and their strategic relationships with (non)knowledge. It connects with melancholy expressions of nationalism, for example in newspaper editorials that speak of a sense of inferiority, rooted in colonial subjugation and maintained through reliance on external projects, and the notion of untapped potential, restricted by resource shortages and 'bad politics'. It was clear from my interviews and ethnographic encounters that many scientists experienced these frustrations as an affecting and emotional context for their work.

In this section I have discussed how a changing water regime provides the grounds for reckoning with(out) data. Resonant with the 'normal science' rationale of IWRM (Orlove and Caton, 2010: 410) and its conceptualisation of water as an economic good, water policy officials and scientists articulated an urge to quantify: mobilising an idea of certainty as theoretically attainable (Hulme, 2009: 84) and water resources as potentially knowable and manageable, if only political contingencies could be contained. For the policy staff, reckoning with different sources of data and uncertainty was presented as an iterative process working towards an eventual imagined alignment, whereby resources could be scientifically managed through a licensing system based on correctly anticipating supply and demand. The uncertainties of the current situation created the conditions for this new policy work to be done. Writing about municipal water supply in Mumbai, Anand (2015) notes that the material resistance of water and its associated infrastructures to calculation and governance is influential: it affords spaces for contestation and strategic ignorance that enable water engineers to carry out their practical work. Ballestero (2012) has examined the productive roles of faith, dissent and ruptured numeric logics that generate relations and potential in water policy processes in Brazil and Costa Rica. In Belize, too, material and social things confound certainty about water. This generates debate about not only how policy should be made and implemented but also how politics should be done. Thus, reckoning with(out) data also means reckoning with environmental and political indeterminacies.

Reckoning with models

In this vignette, I turn from the policy sphere to a training workshop concerned with reckoning future water by wrangling software and data to run models that would visualise runoff and erosion in decades to come. This hydrological modelling practice was underpinned by the topographically defined concept of the *watershed*, and the temporal scale of climate projections to 2080. The event was one component of a research project assessing the potential impacts of climate change on Belize's water resources. While, in the wider project, modelling specialists were using similar processes to produce new datasets and technical reports, this workshop was envisioned as a training opportunity for people working in different sectors: to learn about available tools; to explore 'what would happen' to water resources under different climate change and land use scenarios; and to consider the impacts for their professional domains. The international financial, technological and knowledge infrastructures behind the endeavour were made clear through the collaboration of the national university with regional NGOs and scientific institutions, funding from international agencies, the use of IPCC scenarios to frame the modelling exercise, and the venue: a Taiwan-funded computer lab in the Belizean capital city. Many of the 20 or so participants wore shirts embroidered with the logos of government ministries (e.g. agriculture and natural resources; forests, fisheries and sustainable development); others hailed from NGOs, consultancies and universities. Some had brought laptops already loaded with the ArcGIS mapping software that most participants used in their work; the rest of us logged into the lab's PCs and loaded a free alternative programme.

Over two days, the course facilitators introduced online sources of global climatological data, satellite observations, and future rainfall scenarios. On our screens, the maps — based on watersheds — projected a unit of analysis that exceeded the familiar national map of Belize, extending into Guatemala and Mexico, and drawing attention to incongruities between physical and political boundaries. The climate datasets we downloaded from *worldclim.com* comprised monthly averages for past, current and future temperature, precipitation and bioclimatic variables at resolutions of up to 1km^2, with observations interpolated for areas lacking weather stations, such as the expanse of the Maya Mountains. The future datasets had been derived by downscaling 19 general circulation models in the CMIP5 (fifth Coupled Model Intercomparison Project), run according to IPCC scenarios. Using the GIS software and the N-SPECT (nonpoint-source pollution and erosion comparison tool) extension, downloaded from the US NOAA (National Oceanic and Atmospheric Administration) website, we followed instructions to apply future precipitation and land use (i.e. deforestation) scenarios to maps of soil type and elevation. An early task was to activate the digital elevation model (DEM), which underpins the representation of the watershed as the key to water behaviour: water flows downhill, so the elevation of each grid cell relative to its neighbours determines the direction of flow. Using these resources, we would be reckoning future quantities, qualities and distributions of water. Scientific calculations would transform the baseline observations into projections using complex mathematical relationships among factors including soil type/moisture, rainfall amount/intensity, and topography; allowing us to query how variables could shift under different climate and land use scenarios.

While many of the workshop participants were familiar with using GIS software to map existing circumstances, few said they had used the demonstrated techniques to cast GIS maps into the future (or cast the future in maps). In response to prompts about how the techniques from the workshop could be transformed into practices with operational relevance, practitioners spoke of different sensitivities, information needs, technical constraints, responsibilities and capacities for enacting decisions relating to future water quality and quantity. Nonetheless, the extent of the promise of GIS felt among its proponents was characterised when, during a coffee break, one participant remarked that a lot of GIS work in Belize had started "for [biodiversity] conservation", but was now "for society".

The work of friction

Slowed by internet connection delays that disrupted smooth data downloads, we nonetheless conjured flickering visualisations of what might happen to water accumulation, runoff, and sedimentation over the next 35 years. The models ran slowly, causing some impatience. We were occasionally derailed by inconsistencies between the free and licensed software, prompting joking interjections about pirated versions, aimed at representatives from the software distributor who were present. For example, one software version automatically performed a correction to fill in 'sinks' in the DEMs that can cause problems for the next modelling stage; in the other version this had to be done manually. The question was posed: how would one know if the sink correction had been done or not? The response was to look at the model output map and check it against prior knowledge about the location of major accumulation points (e.g., river mouths).

These kinds of challenges illustrate what Edwards (2010: 83-86) terms the data and computational 'friction' of working with large data sets, multiple computer systems and diverse organisations to create and manipulate global atmospheric simulations. In his definition, computational friction is the resistance that hinders the conversion of inputs into information and knowledge; data friction is the (energy, attention, time) cost of moving data among machines, humans and organisations. In the workshop, these frictions made themselves known through the slow-moving progress bars that sent us seeking coffee ('run time is uncertain…'), in the different maps displayed on my and my neighbour's screens after attempting to perform the same operations on the same datasets, and in the frustrations of the workshop convenors who had spent time checking the data and instructions only to be faced with unexpected outcomes. Dealing with such frictions involved social and physical energies: switching file formats, converting inches to centimetres, re-running models, making jokes, offering reassurances about the validity of the methods, and advising trainees to confirm model outputs with reference to prior knowledge.

Further frictions became apparent in the course of participants' questions to the modelling experts leading the workshop, as they raised queries about data provenance, model assumptions, physical processes, and judgment of what the models were expected to reveal and what they might be useful for. There was animated discussion about the terminology and scale of watersheds, catchments, sub-catchments, and of *cuenca*, *subcuenca* and *microcuenca* in Spanish-speaking neighbouring countries. Watersheds are seen to be useful management units, but these conversations made it clear that their definition is not always a given. As we were guided through the model setup in the N-SPECT plugin, questions arose about the default inputs. The curve numbers (used to predict runoff resulting from rainfall events in particular areas) had been developed in the USA: it was suggested that a future research agenda could include developing more locally relevant parameters. Government water and environment specialists noted that many of the pre-defined water quality standards were based on those of the US Environmental Protection Agency, while best practice in Belize was guided by World Health Organisation standards. Other pollutant threshold values that had been derived from US studies (according to the user guide) were treated with some suspicion by the government environment officers who explained "the way we do things here". The facilitator emphasised the possibility of adjusting the defaults to locally relevant values in future uses of the model. (We eventually moved on, disregarding the output analyses that used the figures that the government officers professed not to trust.) There were also questions about the input of numbers of thunderstorms and intensity of rainfall, as recorded in my fieldnotes:

> We also have to add a figure in a box marked 'raining' — this stands for the number of raining days per year, defined as the average number of storms in one year in the period of interest. In Panama [where project contributors have been developing the models], they always used the number 40.[8] Some participants picked up on the Biblical reference… This number was subject to quite a lot of debate. Is 40 a good figure? How might this change? What does reducing/increasing it do?

These debates were about the degree to which people and organisations shared knowledge of — and were convinced about — the units of analysis and points of reference being used to reckon qualities and quantities of future water. As such, they concerned the portability of the modes and outputs of the reckoning process (Kockelman and Bernstein, 2012): the extent to which the validity and meaning of the parameters and modelling process could travel across geographical locations, environmental contexts, and regulatory landscapes. Dealing with the friction of standards and inputs, datasets and software took time and social work, including that which had gone in to designing and justifying parameters and software to account for (some) differences (as evidenced in the tools and user guides), and also the facilitators' efforts at clarification. These translations sometimes succeeded and sometimes did not succeed in convincing workshop participants about applicability across cultural and environmental contexts; they generated lively discussions about model assumptions, limitations and potential.

Modelling interdependence

To correct for model processes that made water seem to pool 'unrealistically' in output maps, we were instructed to perform a function that 'burned' digitally into the underpinning topographical layers. The facilitator's explanatory simile — that this was "like digging a ditch in your land to make the water flow where it should flow" — collapsed the divide between the model and the physical world, bringing earthy realism and physical labour to bear on the pixelated layers before our eyes, inviting us to craft a more realistic version of the model by figuratively carving into its representational landscape: the DEM. As for the sink-filling operation mentioned above, the implication was that knowledge of the material world had to be mobilised to check and correct the visuals being called forth on our screens.

The model of nature built in the software's equations seemed to have what Munk (2013), in an account of his own flood modelling apprenticeship, calls "its own anticipations": "it exacts a certain demeanour on behalf of its modellers; it expects us to feed it with a world rendered in specific and digestible formats" (Munk, 2013: 145) — for example the input parameters mentioned above. Munk (2013) argues that the interdependence of the model and modeller in scenario generation both requires and produces a hybrid through which uncertainty and surprise may proliferate, thus 'emancipating' nature from bounded assumptions, not merely working as a tool to anticipate a nature already 'out there'. This resonates with an 'abductive' mode of anticipatory reckoning, open to surprise and undetermined outcomes, and operating through a 'workmanship of risk' (Hallam and Ingold, 2007; Pye, 1968) which bears the possibility of failure. Adams et al. (2009: 255) describe computer modelling as a "standard means of abduction", and abduction as a core dimension of anticipation which focuses attention on how the present can and should be influenced by particular futures. Responding to challenges about the input values, and making the case for the usefulness of the approach in a range of management contexts, the instructor emphasised that a key benefit of the N-SPECT tool was the capacity to 'tweak' settings and inputs to experiment with different possibilities and decisions, for example to separate different kinds of land use and rainfall trajectories to make different relationships and possibilities visible. The training thus highlighted the constrained manipulability of the model as a "mutable mobile" (Morgan, 2012: 398) — a tool for reasoning and imagination, in which the barriers and frictions that trouble the portability of technologies and meanings (Kockleman and Bernstein, 2012) can at the same time open debate about environmental and political uncertainties.

Reckoning with models is thus a relational anticipatory practice: it relies not only on hydrodynamic equations, interpolation techniques, and infrastructural data connections, but on experiences, regulatory contexts, and discursive explanations relating virtual to physical, watershed to polity, baseline to scenario, and model output to human expectations. Highlighted through its status as a training session, the modelling practice during this workshop was less about calculating specific outputs than about experimentation: visualising multiple alternative futures; asking participants to consider potential uses of the tools; and situating us in relational interdependence with the model. Through this, our ability to affect the mapped outcomes — in intended and unintended ways — served as an analogy for the sensitivity of water to climate and land use change; demonstrated the fragility of the model itself; and invited us to adjust its parameters to more closely match our perceptions, knowledge and expectations of the material world. The experimental framework and its associated frustrations revealed the work of making translations that (partially) stabilise the models to the extent that they can act as a shared resource for meaningful negotiation. While the process evoked by the policy worker in the previous section was one of narrowing towards an ideal alignment of expectation and eventuality, mediated by iterative interventions and feedback, the workshop conjured a proliferation of futures that caused participants to question what kinds of interventions might have different effects.

Affective reckoning

In this final empirical vignette, I address facets of reckoning future water that connect foreknowledge practices to the material and meaningful affordances of the *waterscapes* that people inhabit. If (as Kockelman and Bernstein (2012: 326) set out for spatial and temporal reckoning using maps and calendars) a requirement for triangulation is being able to relate one's own position to a privileged point of reference/point to be reckoned, it is relevant to ask: how do people working with water policies and models position themselves within the environment as they reckon present and future water? The preceding vignettes have touched on some ways of identifying circumstances at the point of reckoning (for example through baselines/current datasets); I now focus on examples of how phenomenological and narrative knowledges are negotiated in (dis)connection with scientific understandings.

Strange weather

One theme that emerged strongly during my fieldwork was how people reckoned time and climate in seasonal terms, and the conundrum of doing this when expectations did not match experience. What does it *mean* to say we are in the wet season? When is a wet season not a wet season? Forecasters, agriculturalists and dam operators pondered aloud the "strange" weather conditions throughout my research starting in August 2014. At that point, we were either experiencing an exceptionally dry wet season, or a prolonged dry season. The dissonance between expectation and experience was interpreted and articulated in different ways. For most respondents, including the meteorologists from whom I sought professional opinions, definitions of wet and dry seasons were aligned with certain months, based on historical trends. In this understanding, the rainy season (usually preceded by a short rainy spell and brief pause) is typically defined as coincident with the hurricane season (June-November): as such the season is not defined based on a trigger or threshold in observed conditions. It is thus possible that material conditions do not fit the seasonal description (as in the observation of a "dry wet season").[9] An alternative perspective was articulated by the sugar industry workers who told me, in a dusty Orange Walk town in September, that the rainy season "would soon come", and by an agriculture officer who noted that one year, when it started raining in October and continued for several months, there was "almost no dry season." These alternative modes of orientation situate the speaker within material conditions rather than calendar-based seasons, and highlight narrative tensions between the two.

These different attempts to account for dissonance between expectation and experience communicate the widespread notion that atmospheric conditions had recently become unpinned from established points of reference. They draw attention to the relational work of reckoning water in the present, let alone the future: both in terms of the challenge of communicating meaning when expectations based on shared seasonal calendars are destabilised (as in apparently oxymoronic descriptions of 'dry wet seasons'); and the translations that are possible based on shared cultural understandings of what different seasons have meant in the past, and of (sensed and/or narrated) slippage. These tensions, sensations, opportunities and emotions, mobilised in discourses about weather and climate change, are part of the context in which scientists and policymakers operate. Rather than always existing at a remove from technically-mediated measurements and trends, the sensory dimensions of weather and water knowledge are often implicated in scientists' narratives and justifications. For example, a facilitator of the modelling workshop referred to the colour of the Belize River as viewed from his plane window on landing as an indicator of the anomalous current season. This observation inspired a workshop exercise using satellite data to compare this year's rainfall readings to an ostensibly 'normal' historical year, thus bringing different knowledge sources into conversation.

It is not my intention here to discuss evidence of shifting patterns, but to reflect on the ways in which people framed their interpretations of weather as experience and climate change/variability as a domain of knowledge extending into more or less distant pasts and futures.[10] Perceived shifts were attributed to different physical factors including El Niño, longer-term climate change, and land use practices. For example, a member of technical staff at the Agriculture Department mixed personal and professional experiences as he noted that, while not everyone is familiar with 'climate change' as a concept, changes are registered through situated awareness, memory and comparison:

> I remember growing [up], my grandfather saying the first of May he will plant because he knows rain is coming. [It] worked, yeah… Maybe a farmer might not know what climate change is, but he is aware that the surrounding is different… I've heard a lot of farmers saying it's hotter… Along the highway I used to see those streams yearlong with water… But now as dry season comes they are dry. That means somehow upstream they have cleared the land, and there's not much water to run into the stream again.

The sense of disorientation presents a practical as well as epistemic-ontological dilemma, as the transition between wet and dry seasons is the most sensitive period of the year for many decisions. Sugar and hydropower workers pointed out erratic curves in rainfall graphs over recent years, explaining how decisions based on reckoning using historical trends had led to losses when environmental conditions did not conform to climatological expectations. In the previous seasonal cycle, the rains had continued beyond their usual terminating point, causing upset at harvest time for sugar farmers and refiners as fields were waterlogged and cane diluted. Recalling 2012 as "a horrible year", hydropower planners explained that they had drawn down reservoir levels late in the dry season, expecting plenty of rain as normal in July. When this did not materialise, they had to turn to more expensive power sources. As Vaughn (2017) argues for climate change adaptation projects in Guyana, "unruly" worlds can push experts to reconfigure their knowledge-seeking behaviours.

Reckoning otherwise

With the past thus destabilised as a reliable frame of reference, some decision-makers were seeking alternative foundations for foreknowledge. Hydropower managers and large-scale farmers were researching the use of dynamic forecasting models at daily to monthly timescales, and/or 'real-time' information from Met Service radar or private weather stations; smaller-scale farmers spoke of using near-term, situated indicators of rainy season onset such as animal behaviours: a more intimate form of reckoning in terms of both sensory and temporal proximity. Some eschewed the pursuit of more reliable information in favour of possibilities to reduce sensitivity to variability, for example through index-based crop insurance, soil and water conservation, crop diversification, and/or moral economies of collective support. In agriculture — as in other sectors — sensitivities to atmospheric conditions and to information vary, across crops, locations, scales and styles (e.g. mechanised or manual, irrigated or rainfed), and the social relationships in which they are embedded. Predicting future water resources is but one consideration: the question of what can be done about variable conditions is entangled with the capacity and values of the individual or organisation (electricity distributors worried about value for money; dam operators worried about infrastructure failure; water supply managers prioritised quality). A high-profile industrial dispute in the sugar sector during my stay demonstrated that while the timing of rains is important for the sugar harvest, political contingencies and negotiations of quotas, prices, and farmer autonomy are crucial (Haines, 2019). Anticipation as an affective state (Adams et al., 2009; Zaloom, 2009) may be experienced and addressed very differently according to how individuals and collectives are positioned and oriented in relation to environmental conditions, more-or-less shared systems of reckoning, decision-making processes, emotional engagements, and capacities to act on information.

This section has documented narratives of disorientation and anticipation that characterise efforts to reckon future water in the face of unruly points of reference. Weather and water resources emerged in these narratives less as external objects to be known and potentially managed, more as ontologically unstable — and potentially unknowable — atmospheres and waterscapes in which people and decisions are embedded. Reckoning future water involves relating points of orientation and reference: the work of communicating meanings often involves placing sensory experience in relation with shared cultural understandings, privileged units of measurement and narrated memories in the process of reckoning with environments as well as with politics and technology.

Discussion: Political lives of anticipation

In the first vignette, I described how water policy officials and scientists reckon with(out) data as they confront the political contexts of managing nonknowledge. The theoretical calculability and governability of future water quantity and quality are conjured by narratives that emphasise epistemic uncertainty (incomplete knowledge as a result of insufficient data), and express an urge to quantify, objectify and manage water via bureaucratic instruments such as assessments, masterplans and licenses. At the same time, many practitioners acknowledge that figuring current water, let alone anticipating what climate change/variability might do to it is a political problem, unlikely to be resolved by scientific data alone. The second account — of reckoning with models — drew attention to frictions underpinning ostensibly calculative modes of anticipation. Trainee modellers queried and translated the inputs to a model, producing multiple simulated visions of the extended future (via representations of quantitative calculations solved in each grid square). The experimental adjustments of inputs catalysed discussions about assumptions, expectations and entanglements of world, model and modeller; and about the relative portability of information conveyed by scientific calculations, observations from different locations, and personal experience. The significance of experience extends into the third vignette, which addressed the affective reckoning of people trying to orient their experiences, narratives and decisions within an atmospheric context perceived to be unstable. This further exceeds the 'formal' definition of reckoning as counting or calculation, raising questions about dealing with the ontological uncertainty of chaotic atmospheric systems and the reflexive uncertainty of human responses to information that may in turn influence atmospheric outcomes (Dessai and Hulme, 2004). As such, it draws attention to the definition of reckoning as challenge, opinion and judgment, and to the work involved in situating oneself in orientation to points of reference that are more or less socially salient.

These different modes of reckoning are not separate; indeed, affective dimensions resonated through all the situations described above — from the emotional frustrations of scientists feeling their work to be constrained, to policy workers' attempts to control unknowns of water and human behaviour, to the modellers' reflexive concerns about manipulability and urgency (and their instructions to 'dig' into the model landscape), as well as the disconcerting temperatures and colours sensed in environmental surroundings. Socio-material data and computational frictions (Edwards, 2010) also draw attention to frictions between worldviews (Tsing, 2005) that emerge in processes of reckoning uncertain futures and conveying their meaning, and which can cause discomfort, anxiety, disorientation, confidence, and excitement (Adams et al., 2009; Zaloom, 2009) as they draw participants into reflexive relation with technologies, environments, people and organisations.

Notwithstanding recognition of the difficulties of knowing current and future water, I encountered many people strongly invested in the promise of assessments, maps, and models for resource management, notably among a cross-sector community of GIS workers and enthusiasts who were active in workshops and on social media, sharing maps and promoting their benefits not only for water management but also for agriculture, forestry, health and journalism. The aspiration to scientific management draws attention to how relationships between the real and the virtual are imagined and managed, for what purpose. While policy ideals promote an integrated vision of watersheds as social as well as ecological systems, the focus on addressing these through data-led interventions risks overlooking the diversity and friction of political struggles and interpretive meanings of the future, and valorising frameworks that fund and legitimise only particular projects and principles — for example the principle of water as an economic good (Orlove and Caton, 2010).

Although some forecasters and scientists expressed interest in the knowledge that small-scale farmers could contribute to water management planning, others engaged in more defensive discourses, positing a hierarchical distinction between science and the knowledge of groups often described as less-educated farmers who 'plant by the moon'.[11] A few participants cited concerns about the predominance of inputs, instruments and infrastructures originating elsewhere in the world: models are often calibrated for particular locations; external donations and tools produced in distant 'centres of calculation' (Latour, 1987) bear the weight of historical relations and legacies of coloniality (Escobar, 2004; Quijano, 2007). As such, the world as political and contingent impinges on the view of the 'globe', even as the latter pretends to detachment: "each view contains the seeds of the other" (Ingold, 1993: 41). Real and perceived power imbalances within and between sectors or between governments and publics complicate efforts to map and manage: for example, attempts by sugar industry researchers to collect data on multiple variables for growers' fields were not universally welcomed by farmers (Haines, 2019); projects to demarcate land use in southern Belize have encountered and created complex political-ontological struggles

(Wainwright, 2008). Nonknowledge is threaded throughout these narratives — sometimes as potentially reducible epistemic uncertainty, but also as ontological indeterminacy and political critique (Mathews, 2014). It may be wielded as a resource by those in positions of authority; it can also create possibilities for considering multiple water futures and re-embedding water, weather and climate knowledge into social and political lives (Hulme, 2009).

Conclusion

In their discussion of reckoning, Kockelman and Bernstein (2012: 336-337) argue that creating knowledge claims that are portable across cultural and historical contexts often involves "long chains of responsibility and right, truth and justification, evidence and inference, technologies and techniques, everydayness and expertise, as well as modes of theoretical and practical agency". Latour (1999: 58), commenting on the durability of 'things', argues that it is through a "regulated series of transformations, transmutations and translations" that acts of reference work to ensure and maintain coherence of meaning. In this article, I have shown how different technologies, senses and expertise are put to work to compare past, present and future; to map and imagine different possible futures; and to influence (or hinder) policies and actions that may usher these futures into existence. I have drawn attention to the roles of *nonknowledge* and *friction*, and the socio-material dimensions of multiple modes of anticipation that craft water resources as temporal, relational, political and also *affective* phenomena, known and debated through 'abduction' (Adams et al., 2009: 255) — an orientation to the future that lies between 'is' and 'ought'; a condition of striving to know what to do under pressures of time.

Contested values and knowledge-making practices trouble the 'integration' promoted in contemporary global frameworks for water management. This is particularly salient when resources are contentious: water may be abundant in Belize now, but its deep connection with land, in a context where land is closely aligned with power, increase its potency as an object of politicised reckoning. Technical limitations, divergent values, intractable politics, and unstable environments are challenges for the relational work of reckoning, which is social and cultural given that its ability to convey meaning relies on shared understandings. These are of course crucial for wider publics as well as the professionals whose practices and perspectives have been the focus of this article: future extensions of this work could engage with reckoning practices of wider groups, and investigate change over time as water assessments and management interventions are enacted and socio-ecological settings continue to shift. As Nelson (2009) notes in her work on the aftermath of war in Guatemala, the notion of reckoning holds the promise of accountability, but also the power to unsettle objectivity as people and institutions struggle to produce 'facts', or — in the terms explored above — meanings that make sense across contexts. The modes of reckoning described here are anticipatory practices with political effects that stem from their capacity to orient themselves in the present while rendering certain visions of the future more or less imaginable (Taddei, 2013). Thinking in the multivalent terms of reckoning, then, draws attention to the inseparability of facts, values, and consequences in attempts to navigate human-environmental relationships in past, present and future. The notion of 'reckoning resources' points both to the socio-material *practices of reckoning* future water resources using different technologies, senses, inputs, standards and understandings, and also to the ways that these *tools for reckoning* are themselves resources that may be mobilised to bring (un)certain futures into view and possibly into being. As such, these reckoning resources hold potential as catalysts and vectors for political imagination.

Acknowledgements

I wish to thank all the research participants in Belize, without whom this study would not have been possible. I am grateful to the Belize Institute for Social and Cultural Research for permission to conduct this study and the opportunity to present preliminary findings at the Belize Archaeology and Anthropology Symposium in 2015. A short summary report appeared in Volume 4 of Belize

Research Reports (Encalada et al., 2016). The Caribbean Community Climate Change Centre kindly hosted a seminar in 2015 to communicate and discuss my early results. I am grateful to Steve Rayner, Javier Lezaun, Sara de Wit, Ana Lopez and Liz Stephens, for many valuable discussions that helped to inform this paper; and to panellists and attendees of sessions at meetings of the American Anthropological Association 2015 and Royal Anthropological Institute 2016, and an InSIS seminar in 2018, at which earlier versions of the article were presented. Many thanks also to the guest editors Stefan Aykut, Bilel Benbouzid and David Demortain, and my anonymous reviewers for their constructive feedback along the way.

The research for this study was supported by the Oxford Martin School [Programme on Resource Stewardship]; writing was also supported by the Economic and Social Research Council [grant ES/N016084/1]. Because of the confidential nature of some of the research materials supporting this publication, not all of the data can be made accessible to other researchers. Please contact the author for more information.

References

Adams V, Murphy M and Clarke AE (2009) Anticipation: technoscience, life, affect, temporality. *Subjectivity* 28(1): 246–265.

Anand N (2015) Leaky states: water audits, ignorance and the politics of infrastructure. *Public Culture* 27(2): 305–330.

Anderson B (2010) Preemption, precaution, preparedness: Anticipatory action and future geographies. *Progress in Human Geography* 34(6): 777–798.

Ballestero A (2012) The productivity of non-religious faith: openness, pessimism and water in Latin America. In: Tucker C (ed) *Nature, Science and Religion: Intersections Shaping Society and the Environment*. New Mexico: School of Advanced Research, pp. 169–190.

Barnes J (2016) Uncertainty in the signal: modelling Egypt's water futures. *Journal of the Royal Anthropological Institute* 22(S1): 46–66.

Beck U (1992) *Risk Society: Towards a New Modernity*. London: Sage.

Bernard HR (2011) *Research Methods in Anthropology* (5th Edition) Lanham: Altamira.

BEST (2008) *National Integrated Water Resources Management Policy (including climate change) for Belize*. Belmopan: Caribbean Community Climate Change Centre. Available at: http://dms.caribbeanclimate.bz/M-Files/openfile.aspx?objtype=0&docid=2867 (accessed 4 October 2018).

BEST (2009) *National Adaptation Strategy to Address Climate Change in the Water Sector*. Belmopan: Caribbean Community Climate Change Centre. Available at: http://dms.caribbeanclimate.bz/M-Files/openfile.aspx?objtype=0&docid=2735 (accessed 4 October 2018)

Blok A (2010) Topologies of climate change: actor-network theory, relational-scalar analytics, and carbon-market overflows. *Environment and Planning D: Society and Space* 28(5): 896–912.

Bond D (2013) Governing disaster: the political life of the environment during the BP oil spill. *Cultural Anthropology* 28(4): 694–715.

Broad K, Pfaff A and Glantz M (2002) Effective and equitable dissemination of seasonal-to-interannual climate forecasts: policy implications from the Peruvian fishery during El Niño 1997–98. *Climatic Change* 54(4): 415–438.

Burke L and Sugg Z (2006) *Hydrologic Modeling of Watersheds Discharging Adjacent to the Mesoamerican Reef*. World Resources Institute Report. Available from: https://www.wri.org/publication/watershed-analysis-mesoamerican-reef

CaribSave (2012) *Climate Change Risk Profile for Belize*. Christ Church, Barbados: CaribSave. Available from: http://www.caribbeanclimate.bz/closed-projects/2009-2011-the-caribsave-climate-change-risk-atlas-cccra.html (accessed 17 February 2016).

Caton SC (2007) Yemen, water, and the politics of knowledge. Available at: http://globetrotter.berkeley.edu/bwep/colloquium/papers/Yemen%20Water%20Politics%20of%20Knowledge.pdf (accessed 7 October 2015).

CCCCC and CDKN (2012) *Delivering Transformational Change 2011–12: Implementing the CARICOM 'Regional Framework for Achieving Development Resilient to Climate Change'*. Full Report, Belmopan: Caribbean Community Climate Change Centre.

Choy T (2012) Air's substantiations. In: Rajan KS (ed) *Lively Capital: Biotechnologies, Ethics and Governance in Global Markets*. Durham: Duke University Press, pp. 121–152.

Daipha P (2015) *Masters of Uncertainty: Weather Forecasters and the Quest for Ground Truth*. Chicago: University of Chicago Press.

Daston L and Galison P (2007) *Objectivity*. New York: Zone.

Demeritt D (2006) Science studies, climate change and the prospects for constructivist critique. *Economy and Society* 35(3): 453–479.

Dessai S and Hulme M (2004) Does climate adaptation policy need probabilities? *Climate Policy* 4: 107–128.

Dilley R (2010) Reflections on knowledge practices and the problem of ignorance. *Journal of the Royal Anthropological Institute* 16: S176–S192.

Douglas M (1992) *Risk and Blame: Essays in Cultural Theory*. London: Routledge.

Fine GA (2007) *Authors of the Storm: Meteorologists and the Culture of Prediction*. Chicago: University of Chicago Press.

Edwards PN (2010) *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*. Cambridge, MA: MIT Press.

Encalada N, Cocom R, Solis S and Pinelo G (2016) *Research Reports in Belizean History and Anthropology Vol. 4*. Belize: NICH.

Escobar A (2004) Beyond the Third World: imperial globality, global coloniality and anti-globalisation social movements. *Third World Quarterly* 25(1): 207–230.

FAO (2015) AQUASTAT: Belize. Available at: http://www.fao.org/nr/water/aquastat/countries_regions/BLZ/ (accessed 9 June 2017).

Government of the Republic of Trinidad and Tobago (2016) MET Office declares start of 2016 Rainy Season for T&T. *news.gov.tt*. Available at: www.news.gov.tt/content/met-office-declares-start-2016-rainy-season-tt (accessed 30 April 2017).

Grandia L (2009) Milpa matters: the Maya Community of Toledo versus the Government of Belize. In: Johnston BR and Slyomovics S (eds) *Waging War, Making Peace: Reparations and Human Rights*. Walnut Creek: Left Coast Press, pp. 153–182.

Groves C (2017) Emptying the future: on the environmental politics of anticipation. *Futures* 92:29–38.

GWP (2000) *Integrated Water Resources Management*. TAC Background Papers, Stockholm: Global Water Partnership.

Haines S (2012) Meaningful resources and resource-full meanings: spatial and political imaginaries in southern Belize. In: Janowski M and Ingold T (eds) *Imagining Landscapes: Past, Present and Future*. Farnham: Ashgate, pp. 97–120.

Haines S (2018) Imagining the highway: anticipating infrastructural and environmental change in Belize. *Ethnos* 83(2): 392–413.

Haines S (2019) Managing expectations: articulating expertise in climate services for agriculture in Belize. *Climatic Change* Epub ahead of print DOI: 10.1007/s10584-018-2357-1

Hallam E and Ingold T (2007) Creativity and cultural improvisation: an introduction. In: Hallam E and Ingold T (eds) *Creativity and Cultural Improvisation*. Oxford: Berg, pp. 1–24.

Haraway D (1988) Situated knowledges: the science question in feminism and the privilege of partial perspective. *Feminist Studies* 14(3): 575–599.

Hastrup K and Skrydstrup M (2013) *The Social Life of Climate Change Models: Anticipating Nature*. New York and Abingdon: Routledge.

Hau'ofa E (1993) Our sea of islands. In: Waddell E, Naidu V and Hau'ofa E (eds) *A New Oceania: Rediscovering Our Sea of Islands*. Suva: School of Social and Economic Development, USP, pp. 2–16.

Hulme M (2009) *Why we Disagree about Climate Change: Understanding Controversy, Inaction and Opportunity*. Cambridge: Cambridge University Press.

Hulme, M (2010) Problems with making and governing global kinds of knowledge. *Global Environmental Change* 20: 558–564.

Ingold T (1993) Globes and spheres: the topology of environmentalism. In: Milton K (ed) *Environmentalism: The View from Anthropology*. London: Routledge, pp. 31–42.

Ingold T (2010) Footprints through the weather-world: walking, breathing, knowing. *Journal of the Royal Anthropological Institute* 16: S121–S139.

Jalbert K (2016) Building knowledge infrastructures for empowerment: a study of grassroots water monitoring networks in the Marcellus Shale. *Science and Technology Studies* 29(2).

Jasanoff S (2004) Ordering knowledge, ordering society. In: Jasanoff S (ed) *States of Knowledge: the Coproduction of Science and the Social Order*. London: Routledge, pp. 13–45.

Jennings S and Magrath J (2009) *What Happened to the Seasons?* Oxford: Oxfam GB. Available at: https://www.oxfam.org.au/wp-content/uploads/2012/02/oaus-whathappenedtoseasons-0110.pdf (accessed 26 August 2015).

Kockelman P and Bernstein A (2012) Semiotic technologies, temporal reckoning, and the portability of meaning. Or: Modern modes of temporality — just how abstract are they? *Anthropological Theory* 12(3): 320–348.

Lahsen M (2005) Seductive simulations? Uncertainty distribution around climate models. *Social Studies of Science* 35(6): 895–922.

Latour B (1987) *Science in Action: How to Follow Scientists and Engineers Through Society*. Cambridge: Harvard University Press.

Latour B (1999) *Pandora's Hope: Essays on the Reality of Science Studies*. Cambridge: Harvard University Press.

Lazrus H (2012) Sea change: island communities and climate change. *Annual Review of Anthropology* 41(1): 285–301.

Macours K, Premand P and Vakis R (2012) Transfers, diversification and household risk strategies: experimental evidence with lessons for climate change adaptation. *World Bank Policy Research Working Paper* (6053). Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2045949 (accessed 26 August 2015).

Mathews A (2014) Scandals, audits, and fictions: linking climate change to Mexican forests. *Social Studies of Science* 44(1): 82–108.

McGoey L (2012) The logic of strategic ignorance. *The British Journal of Sociology* 63(3): 533–576.

Medeiros D, Hove H, Keller M, Echeverría D and Parry J (2011) Review of current and planned adaptation action: the Caribbean. *Report for the International Institute for Sustainable Development*. Available at: https://www.iisd.org/project/review-current-and-planned-adaptation-action-developing-countries-supporting-adaptation (accessed 19 March 2019).

Monteiro M and Rajão R (2017) Scientists as citizens and knowers in the detection of deforestation in the Amazon. *Social Studies of Science* 47(4): 466–484.

Morgan M (2012) *The World in the Model: How Economists Work and Think*. Cambridge: Cambridge University Press.

Munk AK (2013) Emancipating nature: what the flood apprentice learned from a modelling tutorial. In: Hastrup K and Skrydstrup M (eds) *The Social Life of Climate Change Models: Anticipating Nature*. New York and Abingdon: Routledge, pp. 144–162.

Nelson D (2009) *Reckoning: the Ends of War in Guatemala*. Durham: Duke University Press.

Nelson N, Geltzer A and Hilgartner S (2008) Introduction: the anticipatory state: making policy-relevant knowledge about the future. *Science and Public Policy* 35(8): 546–550.

Orlove B and Caton SC (2010) Water sustainability: anthropological approaches and prospects. *Annual Review of Anthropology* 39(1): 401–415.

Peirce CS (1929) Guessing. *The Hound and Horn* 2: 267–282.

Porter T (1995) *Trust in Numbers: the Pursuit of Objectivity in Science and Public Life*. Princeton: Princeton University Press.

Power M (2007) *Organized Uncertainty: Designing a World of Risk Management*. Oxford: Oxford University Press.

Pye D (1968) *The Nature and Art of Workmanship*. Cambridge: Cambridge University Press.

Quijano A (2007) Coloniality and modernity/rationality. *Cultural Studies* 21(2): 168–178.

Rayner S (2007) The rise of risk and the decline of politics. *Environmental Hazards* 7(2): 165–172.

Rayner S (2012) Uncomfortable knowledge: the social construction of ignorance in science and environmental policy discourses. *Economy and Society* 41(1): 107–125.

Richardson R (2009) *Belize and Climate Change: The Costs of Inaction*. Human Development Issues, UNDP. Available at: http://www.undp.org/content/dam/belize/docs/UNDP%20BZ%20Publications/Belize-and-Climate-Change-The-Costs-of-Inaction.pdf (accessed 17 February 2016).

Scott JC (1998) *Seeing like a State: How Certain Schemes to Improve the Human Condition Have Failed.* Durham: Duke University Press.

Shackley S (2001) Epistemic lifestyles in climate change modeling. In: Miller CA and Edwards PN (eds) *Changing the Atmosphere: Expert Knowledge and Environmental Governance*. Cambridge MA: MIT Press, pp. 107–133.

Star SL and Griesemer J (1989) Institutional ecology, 'translations' and boundary objects: amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907–39. *Social Studies of Science* 19(3): 287–420.

Taddei R (2012) The politics of uncertainty and the fate of forecasters. *Ethics, Policy and Environment* 15(2): 252–267.

Taddei R (2013) Anthropologies of the future: on the social performativity of (climate) forecasts. In: Kopnina H and Shoreman-Ouimet E (eds) *Environmental Anthropology: Future Directions*. London: Routledge, pp. 246–65.

Tsing AL (2005) *Friction: an Ethnography of Global Connection*. Princeton: Princeton University Press.

Vaughn S (2017) Disappearing mangroves: the epistemic politics of climate adaptation in Guyana. *Cultural Anthropology* 32(2): 242–268.

Wainwright J (2008) *Decolonizing Development: Colonial Power and the Maya*. Malden, MA: Blackwell.

Wilk RR (1997) *Household Ecology: Economic Change and Domestic Life among the Kekchi Maya in Belize*. DeKalb: Northern Illinois University Press.

Zaloom C (2009) How to read the future: the yield curve, affect, and financial prediction. *Public Culture* 21(2): 245–268.

Notes

1 Participants have been anonymised.

2 Nelson (2009) also notes reckoning's multiple meanings, using the term to think through the difficulties of making sense of loss and accountability in the aftermath of war in Guatemala.

3 They derive their use of the term from Peirce (1929).

4 Kockelman and Bernstein (2012) suggest e.g. velocity, price, temperature and information.

5 According to the technical review, trends already recorded include: rising frequency of warm days and nights (with night-time temperatures contributing more to the overall increase in average temperatures — approximately 1 degree in the previous 39-45 years); and changes of variation in precipitation regimes (BEST, 2009).

6 Mandates have been reorganised since 2014; at the time of writing the Unit is in the Ministry of Natural Resources.

7 The Hydrology Unit relies heavily on local volunteer observers to collect river level data (of twenty-nine stations being monitored in 2014, three were automatic).

8 This number of rainy days has been used in applications of the N-SPECT model across the Mesoamerican Reef region, based on calibration by scientists working in partnership with the World Resources Institute (Burke and Sugg, 2006).

9 In contrast, Trinidad and Tobago's Met Office has declared the start of the rainy season based on assessments of rainfall events and tropical wave development. For example, in 2016 the rainy season (usually expected to start in June) was declared on May 2nd following an "uncharacteristically early influence from the Inter-Tropical Convergence Zone" on May 1st (Government of the Republic of Trinidad and Tobago, 2016).

10 See Jennings and Magrath (2009) for a report on farmers' perceptions of changing seasons across the world, and Macours et al. (2012) for an example from Nicaragua supported by longitudinal meteorological data.

11 Shorthand for planning agricultural activities according to the lunar cycle.

Values and Consequences in Predictive Machine Evaluation. A Sociology of Predictive Policing[1].

Bilel Benbouzid

Laboratoire Science Innovation et Société (LISIS), France/ bilel.benbouzid@u-pem.fr

Abstract

Predictive policing is a research field whose principal aim is to develop machines for predicting crimes, drawing on machine learning algorithms and the growing availability of a diversity of data. This paper deals with the case of the algorithm of PredPol, the best-known startup in predictive policing. The mathematicians behind it took their inspiration from an algorithm created by a French seismologist, a professor in earth sciences at the University of Savoie. As the source code of the PredPol platform is kept inaccessible as a trade secret, the author contacted the seismologist directly in order to try to understand the predictions of the company's algorithm. Using the same method of calculation on the same data, the seismologist arrived at a different, more cautious interpretation of the algorithm's capacity to predict crime. How were these predictive analyses formed on the two sides of the Atlantic? How do predictive algorithms come to exist differently in these different contexts? How and why is it that predictive machines can foretell a crime that is yet to be committed in a California laboratory, and yet no longer work in another laboratory in Chambéry? In answering these questions, I found that machine learning researchers have a moral vision of their own activity that can be understood by analyzing the values and material consequences involved in the evaluation tests that are used to create the predictions.

Keywords: predictive policing, machine learning, algorithm, algorithmic accountability, crime prediction, sociology of quantification

Introduction

Predictive policing is a field of research whose principal aim is to develop machines for predicting crimes, drawing on machine learning algorithms and the growing availability of a diversity of data (Perry, 2013). In the United States, predictive policing is part of a longstanding project of policing reform by research (Walker, 2004) that seeks to create a pro-active police force which acts on its own initiative to prevent crime, rather than simply reacting in emergencies when called to do so by citizens. Since the 1970s, this reform project has in large part been driven by researchers critical of a situation where the police act mainly in contexts of crisis and drama, and are detached from concerns with preventing delinquency (Weisburd and Braga, 2006). In this context, over the last four decades, a large proportion of policing research budgets has been devoted to experimentation on tactics that might allow the police to anticipate and precede the commission of offenses, rather than simply reacting to them. In 2012, when PredPol, Inc. put a predictive analysis platform on the market in the form of downloadable software, offering a dashboard that displays risks of crime in real time at a precision on the order of 200 meters (Figure 1), the dream of the American police reformers of the 1970s seemed to have been given concrete form as a machine.

With the deployment of this type of analytical platform in public action, a new form of quantification began to progressively spread through public administrations. Classical public statistics, based on the law of large numbers and the associated notions of norms and means, now had to compete with these algorithmic practices, whose main objective is prediction through the automated production of classes, clusters, or patterns. Statistical learning was now liberated from the need for a fixed system of categories: "Rather than stable, permanent, structuring variables, which fixed statistical objects within categories, digital algorithms prefer to capture events that they record on the fly in order to compare them to other events, without first categorizing them. Instead of weighty variables, they seek to measure signals, behaviors, actions, performances" (Cardon, 2016: 49). In recent years, as these changes have upset traditional reference frameworks in standard statistics, many initiatives have been undertaken to make algorithms a specific research object in the social sciences (Dourish, 2016). In this literature, algorithms are seen as powerful mechanisms, with a growing role in all sectors of society and a subtle, discreet, and dissimulated power over individuals (Beer, 2009). Denouncing their intrusive, discriminatory, and underhanded nature (O'Neil, 2016), researchers and activists have demanded a politics of algorithms (Crawford, 2016). Accountability (Diakopoulos, 2014), transparency (Zarsky, 2016), and audit (Sandvig et al., 2014) have become watchwords in these public debates on the algorithm (Dourish, 2016).

The Foucauldian analyses of Rouvroy and Berns (2013) amplify what Ziewitz (2016) called an 'algorithmic drama'. Rouvroy and Berns (2013) criticize the profound transformations in the exercise of power enabled by machine learning. They argue that the normativity of the law, in its discursive and explicit form, allows individuals the choice to obey or disobey and offers them the right to a fair trial which extends the possibility of dialogue; whereas machine learning imposes a "tyranny of the real" that neutralizes critique by producing normative devices based on strict descriptions of individuals' activity, or at least of their relations with their social and material environment. In other words, through the process of statistical learning, the "social norm emerges from the real itself". This algorithmic governmentality is characterized by its capacity to make all forms of resistance schizophrenic: discrimination, exclusion, and the unethical distribution of visibility are not directly produced by the classifications of the algorithms, but by the social reality on the basis of which the algorithms take form. Quantification is no longer the operation of institution of reality and transformation of the world that sociologists have sought to reveal, but instead an operation of conservation and reinforcement of that reality and of the flagrant injustices associated to it (Anderson, 1990; Desrosières, 2002; Hacking, 1999; Porter, 1996). In this context, actors' critical sense focuses on this algorithmic reinforcement of existing realities: that is, on the feedback effects of the computation, and not on the forms of computation themselves, with respect to which, according to Rouvroy and Berns (2013), they lack all critical sense.

If actors lack any critical sense in these algorithmic contexts, neither the pragmatist sociology of social critique (Boltanski and Thévenot, 2006) nor a sociology of controversies (Latour, 1987) can be applied to machine learning algorithms. Does this mean the project of a sociological analysis of machine learning should be abandoned? Pointed as it is, Rouvroy's and Berns's analysis in terms of critical dispossession limits the concrete possibilities for emancipation in relationship to algorithms of government, by confining actors, and sociologists themselves to a stance of powerlessness. If the sociology of science and technology is to contribute to the study of algorithmic prediction, and at the same time to justify its relevance and usefulness in this context, it must develop a specific art of inquiry that allows it to create critical tests specially designed for the purpose. Here, I borrow this art of inquiry from Tim Ingold (2013). To understand algorithms and their predictions, we must enter into a process of correspondence with them—touching them, manipulating them, and subjecting them to various operations.

I undertook such an inquiry between June 2013 and March 2017 in a study on the algorithm of PredPol, the best-known startup in predictive policing. The team of California researchers behind it (Mohler et al., 2011) took their inspiration from an algorithm created by a French seismologist, David Marsan, a professor in earth sciences at the University of Savoie. As the source code of the PredPol platform is kept inaccessible as a trade secret, I contacted Marsan directly in order to try to understand the predictions of the company's algorithm . Marsan tested his algorithm on the same open-access crime data from the city of Chicago that the California researchers used in their own publication. Using the same method of calculation on the same data, he arrived at a different, more cautious interpretation of the algorithm's capacity to predict crime . Unexpectedly, I created a situation of controversy concerning knowledge of the technical properties of the algorithm (Mackenzie, 2004). By confronting a physicist specialized in earth sciences with researchers in applied mathematics who are focused on developing predictive machines, I created an opportunity to take the beings who sustain the algorithm's existence and make them visible, and to focus my full attention on the specific associations that PredPol's algorithm is composed of (Latour and Venn, 2002).

How were these predictive analyses formed on the two sides of the Atlantic? How do predictive algorithms come to exist differently in these different contexts? How and why is it that predictive machines can foretell a crime that is yet to be committed in a California laboratory, and yet no longer work in another laboratory in Chambéry? To answer these questions is to describe a controversy that will allow us to explore the workings of the algorithm from the inside—to "unfold" it, in a Deleuzian term (Deleuze, 2006) that Latour (1987) adopted for use specifically with technologies[6]. This procedure of unfolding revealed that machine learning researchers have a moral vision of their own activity (Daston, 1995) that can be understood by analyzing the *values* and *material consequences* involved in the evaluation tests that are used to create the predictions. To analyze the moral dimensions of prediction is not to study this or that usage of machine learning, but to investigate the transformations undergone by predictive categories as they move from one social context to another. By the end of this article, prediction should be clearly understandable as a moral problem that is indissociably at once cognitive and material.

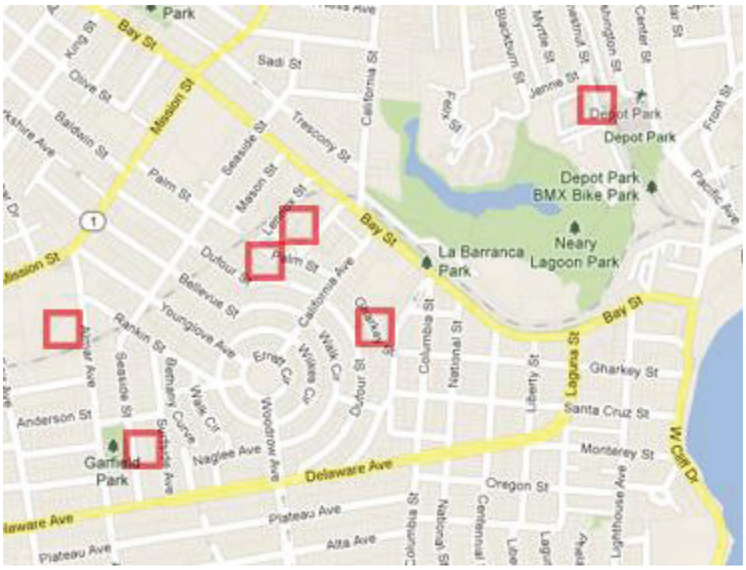

Figure 1. Screenshot of the map on the dashboard of the PredPol platform, indicating upcoming crimes at a precision of 200 metres.

An algorithm predicting earthquakes and crimes without a priori hypotheses

What makes this California startup a reference in the domain of predictive policing is its slogan, "More Than A Hotspot Tool." Since the early 1990s, the urban cartography of "hotspots"—heatmaps of the distribution of crime in the city—has been the main tool for strategic intervention in areas where crime is concentrated (Weisburd et al., 2009). PredPol claims to do better than these classical crime maps thanks to a predictive method used in the field of earthquake prediction. A similarity observed by these Los Angeles researchers between the dynamics of the propagation of crimes and that of earthquakes, they say, means that the geographical dynamics of criminality can finally be characterized mathematically (Mohler et al., 2011). PredPol uses a method of calculation taken from stochastic point processes, a branch of statistical physics. This is a classical approach to modelling the distribution of a set of events (considered as pointlike entities) in a finite space of arbitrary dimensionality (in the models discussed here, two spatial dimensions and one temporal dimension). Point processes are used to identify the formal mechanisms that produce these events in several dimensions, by modelling how they are distributed in time and space.

Producing spatiotemporal clusters by combining concentration and contagion

The question behind this statistical operation is whether the events are distributed randomly or in some more regular fashion, and in particular whether the points cluster around particular locations. The choice of the type of process depends on the researcher's hypotheses regarding the form of mechanism involved. The PredPol researchers started with a classical hypothesis in predictive crime analysis: rather than occurring randomly, crimes are concentrated in space and spread through a local neighbourhood. The repetitive structure of the events themselves is enough to model them (without drawing on external variables). In other words, the best predictor of a future crime is a past crime. I will return to this hypothesis below, but note first of all that the PredPol researchers were interested in self-exciting point processes because they represent a way of modeling interactions between events that takes the history of previous events into account: the occurrence of future events (crimes or earthquakes) depends on the history of the process. The formula below is the mathematical representation of this process, the predictive algorithm used by Marsan (Marsan and Lengliné, 2008) and PredPol (Mohler et al., 2011). It calculates a probability that, in a manner of speaking, represents an idealization of the mean number of crimes or earthquakes on a surface:

$$\lambda(\mathrm{x,y,t}) = \mu(x,y) + \sum_{i,ti<t} g(x-xi\,, y-yi, t-ti, Mi)$$

In this formula, the probability of the occurrence of an event, at a given moment in the process, contributes to an overall calculation of risk intensity per unit surface and per unit time. In the language of statisticians, it is said that this equation "describes an inhomogeneous Poisson process of intensity λ(x,y,t)." This intensity is a calculation of risk interpreted as a density that depends on both space and time. It is obtained by taking the sum of the two components of the formula: first, the function μ (x,y), known as the background rate, depends only on space, and represents a probabilistic calculation of the spatial concentration of risk in general; second, the function g(x-xi,y-yi,t-ti,Mi), known as the contagion kernel, models the spread of series of events whose occurrence depends on previous events and on the parameter Mi (the magnitude of the event). The algorithm models risk intensity at each location in a map by adding together these two components. Note that the contagion model is linear: if this were not the case, computation (simulation, optimization) would be very difficult, and would make the algorithm practically unusable. It is hypothesized, for example, that the contagion g(1,2) following two distinct events 1 and 2 is simply the sum of the individual contagion effects of g1 and g2 (hence the sum in the equation).

A model that adjusts its own form

Marsan uses this method to model how the main seismic events in an earthquake set off aftershocks, which in turn set off their own sequences of earthquakes.[8] The Los Angeles criminologists' interest in Marsan's algorithm was spurred by what they saw as a similarity of form with criminologists' characterization of crime dynamics: just as earthquakes are followed by aftershocks, so, they thought, crimes are followed by "aftercrimes." Since the 1990s, a field of research on the repetition of crimes has developed in the United States and Great Britain, after a multitude of criminological analyses converged on the conclusion that most crimes repeatedly target a small number of victims, and propagate through their immediate spatial neighbourhood. In the model of 'repeat and near-repeat crime concentration' (Pease and Tseloni, 2014) that was proposed to account for this pattern, crime can be seen either as the signal of a relatively stable risk in a given area, or as an indication that incidents of victimization reinforce the probability of the occurrence of later incidents: In other words, crimes are repeated in or near the same location, and spread by "contagion."[9] These two hypotheses are present in the calculation of risk intensity presented above: the relatively stable risk in an area corresponds to the concentration , and local reinforcement to contagion .

With this algorithm, Marsan seeks to show that the structure of cascades of events can be modeled probabilistically, without any particular hypothesis about the underlying mechanisms, and without the need to test model parameters first. Herein lies the contribution of Marsan's algorithm in seismology. While most existing seismological models are parametric (see below for an explanation of this term), with the parameters set on the basis of empirical data, Marsan and his collaborator made their mark in the field by showing that this parameterization can be dispensed with completely. The PredPol researchers made the same argument to justify the value of their own research: the first statistical approaches used in criminology, notably in the study of the spread of crimes, were parametric. This means that they required a hypothesis on how crimes propagate. But how exactly do parametric and non-parametric approaches differ?

In parametric models the form of the model is imposed and its parameters optimized. In nonparametric models, in contrast, an estimate is calculated of the optimal size of the diameter of the circular moving window (smoothing window) that records the number of points in each cell in a virtual grid projected on the map, and the number of parameters varies, increasing with the number of observations. To estimate the parameters ("nonparametric" does not mean free of parameters), Marsan uses the expectation-maximization algorithm, a classical method that, in an iterative procedure, repeatedly alternates two steps (the calculation of expectation and the calculation of maximum likelihood), in order to arrive at the estimator of the model. Marsan refers to another better-known method, artificial neural networks,[10] to explain how this non-parametric method follows in the spirit of machine learning (Domingos, 2017) :

> It's a little like a neural network. We put in bricks that depend on parameters, but the final product is not, or is very little constrained at the outset. The model adjusts its own form. For us, it was above all a way to show that a model with the fewest possible assumptions could converge toward laws (forms) that are very close to the empirical laws conventionally injected at the beginning in stochastic approaches. The fundamental difference with neural network approaches is that they're often used as a black box with a strictly predictive goal, whereas that's not at all what we had in mind. Instead we're trying to understand what a "good" contagion kernel is and how it can emerge naturally from the data analysis.

According to Marsan, this opposition between understanding to construct theories and predicting to act without necessarily having a complete understanding of the phenomenon—which is well known to machine learning specialists (Hofman et al., 2017; Shmueli, 2010)—explains the difference between his approach to modelling and that of the PredPol mathematicians. Here we will look at the practical consequences of this opposition for the way the algorithm is evaluated.

Revealing the sumptuous opacity of the algorithm

Through this simple surface description of the algorithm, we have surpassed the barrier of its mathematical formalism and glimpsed at the hypotheses that its predictions depend on. Here I will move on to a sociology of the knowledge of the technical properties (Mackenzie, 2004) of PredPol's algorithm. To do so, I will consider the algorithm as a being whose dominant mode of existence is technological (Latour, 2012; Simondon, 2016). As Latour (2010: 26) clearly showed, "The technological object is opaque, and—to put it bluntly—incomprehensible [...] in that it can only be understood provided that we add to it the invisibles that make it exist in the first place, and that then maintain, sustain, and sometimes neglect and abandon it." From this perspective, it can be clearly seen that opacity, which is now a commonplace in public debate on algorithms, is not a problem specific to machine learning: all technological beings, generally speaking, "like to hide." There is no use hoping that the PredPol algorithm will become transparent, that its developers will make it public in order to clarify and make it easier to master: like all technical beings, algorithms are fundamentally opaque. Nevertheless, an appropriate method of inquiry can allow us to get into the workings of the algorithm, revealing its "sumptuous opacity" (Latour, 2011: 22).

Touching the algorithm

To unfold the PredPol algorithm, we must know how to use its language—not the specific language of code, but the more general language of *the technical*: detours, zigzags of ingenuity, ruses (Latour, 2012). Faced with the opacity of the PredPol algorithm, we reacted in something like the way an archaeologist might when faced with an ancient object whose meaning escaped her: turning it around to view it from various angles, simulating it, and reproducing it. This stance is an unfamiliar one for the sociologist: handling the object of inquiry, squeezing it, fiddling around with it, hacking it. I thus asked David Marsan to do this on my behalf, and discussed it with him on several occasions. He tested the algorithm that I have just presented on open-access crime data from the city of Chicago—which, as noted above, is the same data that the PredPol researchers used in one of their own publications. I thus asked Marsan not only to explain to me how the algorithm works, but to run it on crime data and share the results with me. Unexpectedly, in so doing, I created a situation of controversy around knowledge of the technical properties of the algorithm. Nothing like a good controversy to get inside the workings of a machine.

To describe this controversy, I will follow the three steps that structure the work of modelling itself: the justification of the choice of point process (here, self-exciting), and then the modelling strategy and the associated model estimation techniques (the expectation–maximization algorithm), and finally, the evaluation of the model. The first two steps depend on analysts' beliefs about the nature of the problem, whereas the last depends on their moral vision of their own activity. In the experiment that I proposed to Marsan, the type of point process was imposed, and his modelling strategy did not significantly differ from that of the PredPol researchers. Marsan thus set out to evaluate a model applied to a phenomenon about which he knows nothing, using an algorithm similar to PredPol. Nevertheless, points of divergence appeared when it came time to interpret the results. Marsan expressed numerous doubts on the capacity of his algorithm to do better than classical maps of crime hotspots. In the note that he wrote on the analysis of the Chicago crime data, he concluded:

> The results obtained offer good reasons to doubt the capacity of the proposed models to do better than simple hotspot maps. The contribution of contagion (the triggering contribution) to explaining the occurrence of future events is small (it represents only 1.7% in the best model). The role of "memory" in the process can thus make no more than very modest contribution to the efficiency of the prediction system. More importantly still, the assumption is that the dynamic of the process remains the same over time. The possible non-stationarity of the process is clearly a problem, because it limits the use of past information to predict the future.
> In 2015, burglaries were not distributed (in time and space) in the same way as in 2014. This non-stationarity is probably due to uncontrolled changes in how criminal acts are carried out. It could also be due to the deployment of new predictive algorithms: as police patrols use them, they might provoke

> reactions among burglars. Contrary to natural processes such as earthquakes, analyses like the ones presented here could change the observed process, which makes correctly predicting future events more difficult (personal note from David Marsan, sent to George Mohler in September 2015).[12]

To understand Marsan's conclusion, recall that the algorithm calculates the intensity of risk in space and time by adding together two elements: concentration (space) and contagion (space- and time-dependence).David Marsan's note indicates that contagion does make a contribution to the process, but it is extremely small—in fact, negligible. And yet, this is the dimension emphasized by the promoters of PredPol in their slogan "More Than a Hotspot Tool." Could it be that the PredPol scientists altered the results to make them more favorable to their commercial project? According to Marsan, the answer is no: the PredPol researchers did honest work. Moreover, Marsan wrote to George Mohler, who responded as follows:

> Thanks for your email and sending along the analysis. I have found your work on nonparametric point processes quite interesting and influential! We have certainly seen the branching ratio vary quite a lot from city to city and crime type to crime type (from 0 to .5). As you point out, it is important to pick such parameters using cross validation in which case it is certainly possible that a simpler model may be favored. It also may be the case that the nonparametric model you are using is over-parametrized (it looks like it has over 30 parameters), so it may be over-fitting the training data. You might need more regularization, or you might want to use a semi-parametric model (you mention using an exponential smoothing kernel, which is essentially a parametric Hawkes process without the background rate). Another thing you bring up is the non-stationarity of the process. I think this is important and something we tried to estimate in the JASA paper (Mohler et al., 2011) (where the background rate \mu depends on time). Disentangling endogenous contagion from exogenous fluctuations in the intensity is a somewhat open problem, though I have done a little work in this area. The non-stationarity of the background rate is one big difference between crime and earthquakes, and you often try to factor in seasonality and other explicit exogenous predictors. (Email from George Mohler to David Marsan, 3 September 2015).

There is no reason to question the honesty of the scientists who worked to develop PredPol. In his response, Mohler shows he is conscious of the many limitations of the PredPol algorithm, and offers a defense against Marsan's critique, recalling that they sought to deal with the problem of non-stationarity by adding a time variable to the background rate. The PredPol equation thus becomes. How did Marsan react to Mohler's response? To answer this question we have to follow in the steps of Marsan's critical analysis:

> The little work I did on this - well, it took two or three weeks of work, that's not nothing - showed me that there was a problem in the data between 2014 and 2015. I took a look in a very simple way to see how they behaved, and in fact they aren't similar at all.

Marsan then invites us to look at the two figures below. On the left are earthquakes; on the right, crimes.

These two graphs (Figure 2), which represent simple descriptive statistics on change over time in the mean distance between pairs of events separated by n-1 events, led Marsan to say that the data do not behave in at all the same way between 2014 and 2015. According to Marsan, the memory effect is very weak for crime. While the distance between pairs of events increases with the number of interposed events for earthquakes, for crimes it is not apparent that there is any such trend. The most surprising thing for Marsan was that for crime, the mean distance was quite different for different years. The phenomenon is not stationary.

Accuracy vs. precision

How is it that these simple descriptive statistics were enough to convince Marsan that his algorithm is not particularly applicable to the Chicago data, but that they did not concern Mohler, who—as his response to Marsan shows—was not surprised by this difference between seismological and criminological phenomena? Marsan suggested a partial response in our interview:

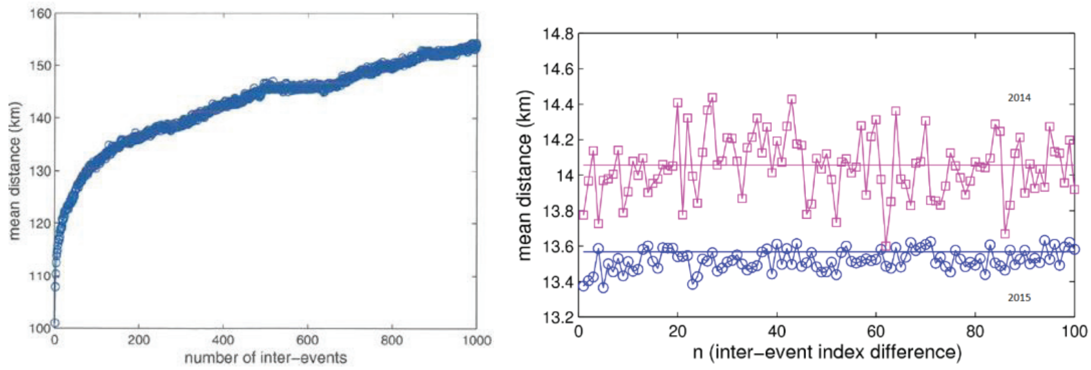

Figure 2. Graphs showing the evolution of the mean distance between pairs of events, from Marsan's notes. These graphs illustrate the "memory effect" and the stationarity problem. The x-axis is exactly the same on both graphs.

> Bilel, you have to understand. You're a statistician, you don't know much about the problem you're being given, and they say to you, "We'll pay you, we'll give you the data, give us the best possible model." You go to work and you realize that your model behaves well one year and then a year later it doesn't. You're a stats guy, you don't know much of anything about the problem. What do you do? He as a statistician says to himself, my model isn't flexible enough, I'm going to make it a little more flexible, I'm going to add . Personally, I'd rather go and talk to the Chicago police to try to understand what happened, what changed. Why is it different in 2015 than it was in 2014? Is it a counting problem? Did the police officers change their habits? Basically you try and understand what made it change from year to the next. Maybe Mohler tries to understand, but his attitude suggests to me that that's not what he does. He tries to improve the predictive power of his algorithm. But because that doesn't work too well, he tries to make it a little more flexible so it works better. His model isn't flexible enough, so he says "I'm going to take my μ(x,y) and make it a bit more flexible so it works better by adding temporal variation to the background rate".

Thanks to his intimate knowledge of the algorithm, Marsan was able, in a manner of speaking to get inside his own algorithm, putting himself in Mohler's skin. Before Marsan looked into the Chicago data, the PredPol algorithm remained invisible not because it was protected as a trade secret, but because everything that would make it possible to follow the algorithm's course of action remained hidden. Marsan allowed us to bring out some of the invisibles (Latour, 2010) that the algorithm depends on. In doing so, the seismologist revealed the ingenuity with which his research was diverted, transformed, and translated to become usable in a police officer's smartphone. He discovered, with the same stupefaction as me, what had become of his algorithm in the hands of a team of audacious mathematicians. Nevertheless, Marsan was critical of the way in which the PredPol developers deployed his machine:

> It might be that that's not the right approach. It might be that it's even the contagion that's different from one year to the next. You'd have to switch out the contagion kernels. But that's the hardest part to adjust. It's simpler to just add a time variable. What he does is really basic. In seismology, we do things that are much more complex to get the background rate to change over time, to take non-stationarity into account. The essential step after the PredPol article would be to understand the non-stationarity. But they're driving blind. Personally I think you can't analyze your data without asking questions about the reality they represent. If you like, the two of us aren't driven by the same engine. What interests us in seismology isn't doing prediction, it's understanding the form of the kernel. Contagion interests us because it gives us clues about the mechanisms that make it so that one earthquake sets off another. It interests us because it tells us something about the seismogenic process. We're not going to impose an a priori form, because the form is what interests us. He's not interested in the form of the contagion. He doesn't want to understand how the contagion happens. He wants to make a prediction. It's totally

> different. In our field you can find the same kind of researcher. There are people who do prediction, but who don't want to understand the process. And there are many of us who think this leads you into a dead end.

Here, Marsan is no longer discussing the effectiveness of the machine in relationship to the crime data, he is offering a *moral condemnation* of the work of the PredPol developers. In his view, he as an earth sciences professor in France does not share the same values as the generalist mathematicians in California. According to Marsan, prediction poses a "basic" research problem, in that what is in question are the theoretical foundations of seismology in a context where, as theoretical knowledge currently stands, predicting earthquakes is impossible. He repeated the point several times in our interviews:

> Short-term predictions (from a few hours to a few days) are rarely successful. Most times they're cruelly disappointing. Our failures at prediction regularly raise the question of whether predicting earthquakes is fundamentally impossible.

Marsan thus sees basic research as a tactical retreat: the idea is not to claim to produce pure and autonomous research, sufficient unto itself, but to take a step to the side into more theoretical research in order to overcome the problem of earthquake prediction. The community of seismologists to which Marsan belongs argues that a comprehensive approach to the phenomenon is needed. They oppose exclusively probabilistic and predictive analyses, which are defended by researchers who consider that theoretical research in seismology has reached its limits (cf. the debate in *Nature* [Main, 2017], and notably the opposing positions of Pascal Bernard and Didier Sornette, which Marsan highlighted in our interview). Marsan analysed the crime data in the spirit of his work on earthquakes. On his view, a nonparametric model is appropriate if the patterns yielded by this statistical learning approach raise important research questions. This principle implies an evaluation of the "accuracy" of the algorithm: what is evaluated is the capacity of the calculation to reveal a close link, or a certain degree of accuracy, in the match between the mathematical model and a coherent conception of the phenomenon under study. The PredPol researchers judge the algorithm according to different criteria: if it improves the "precision" of prediction scores, then the algorithm is satisfactory.[13] Marsan and the PredPol developers do not subject the algorithm to the same tests. In California, the crucial test is performed through a type of lift curve, which is a tool for comparing the performance of different algorithms (Figure 3). It was in this spirit that the PredPol developers turned to point process statis-

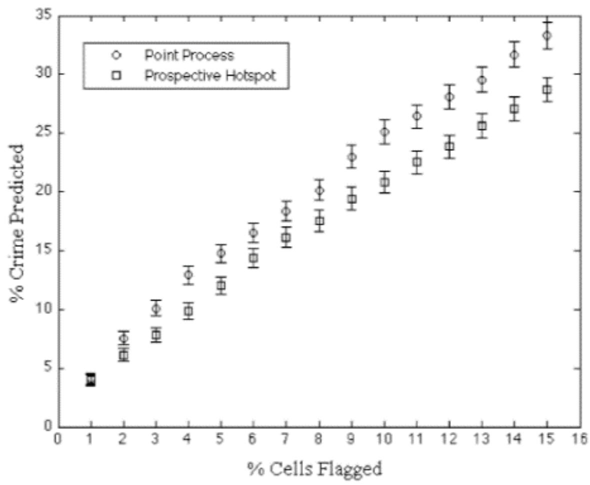

Figure 3. An example of a curve of predictive efficiency used in the article of Mohler (2011), showing the quantity of successfully predicted crimes in relationship to the number of cells flagged. The graph demonstrates the predictive superiority of PredPol with respect to Promap, the pioneering predictive policing tool developed in England in the 1990s by the criminologist Ken Pease and his collaborators.

tics, testing the algorithm on a criterion of competition with other algorithms on the prediction market. These California mathematicians aspire to other principles, basing their research practice on objectives of precision, efficiency, simplicity, and the ability to bring a predictive solution to market—all values that may be totally detached from the accuracy and correctness of the results produced by the calculations (Daston, 1995).

The robustness of a prediction is inversely proportional to its practical consequences

One invisible that would have been difficult to detect without the help of David Marsan is the fact that the PredPol developers added flexibility to the algorithm by integrating a time variable to the background rate, thereby resolving the thorny problem of non-stationarity in the simplest possible way. Through this unfolding of the algorithm, Marsan freed us from the widespread obsession with the question of predictive efficiency alone. Algorithms are technological beings that can offer the opportunity for a much richer debate. The situation of controversy that I created from scratch (a useful way of demonstrating the sumptuous opacity of the algorithm) revealed two different moral visions of predictive activity: one focused on the correctness (or accuracy) of models, the other on the precision of risk scores. Now what is needed is to follow the lines of the network that is laid out on the basis of these two different ways of assigning value to predictions.

Conceptualizing the process underlying a phenomenon or capturing the largest possible proportion of events?

Recall this fundamental principle in the sociology of the sciences: phenomena are defined by the response they give to the tests that scientists subject them to in their laboratories (Latour et al., 1992). In the Chambéry earth sciences laboratory, what Marsan calls an "aftershock" acts as a specific being:

> When a seismologist is analysing aftershocks, he doesn't content himself with counting them. First of all, periods of high activity are the ones when he has the best chance of catching a large earthquake in the net of his measurement networks. If there are enough recordings of good enough quality, he'll be able to establish a tomography of the rupture of the fault. Even without major aftershocks, he'll learn a lot from the small ones, particularly on the directions of tectonic constraints, which he can deduce from their mechanisms. The bulk of the analysis work consists in localizing earthquakes: based on the arrival times of the P waves and S waves from each earthquake at each seismometer in the network, you draw the cloud of aftershocks point by point. It looks heterogeneous, but it includes calm areas and swarms, whose distribution changes over time: the subterranean cloud moves progressively further away from the epicentral zone. The mechanical disturbances of the main fault decrease quickly with distance, so that aftershocks are mainly only observed in its neighborhood, at distances equivalent to the length of the fault itself. By studying the aftershocks you can draw a circle around the main area of movement and localize the fault that's responsible, even if it doesn't break all the way to the surface. Even better: the form of the cloud that they draw out in the opacity of the earth can sometimes be used to determine the orientation of the fault. In certain cases, an abnormal concentration of aftershocks reveals the beginnings of the destabilization of neighboring faults. In the early 1990s, it seemed like everything had been done or said with aftershocks. Post-earthquake field studies became a matter of routine, a well-oiled machine, with ever more effective standard analyses. However, while the images of aftershocks were becoming more precise, detailed interpretation seemed to be impossible, as it depended on uncontrollable parameters linked to resistance and to the—unknowable—state of tension of peripheral faults." ((Bernard, 2003), reading suggested by Marsan.)

Seismologists see aftershocks as a chance to understand what happened, to "catch a large earthquake." In Marsan's more technical formulation, aftershocks serve to "filter the signal." In seismic catalogs, all seismic waves have been recorded jointly by the seismographs. This is why seismologists need to isolate independent earthquakes (for example, those linked to the secular movements of plate tectonics, also known as mainshocks from earthquakes that depend on one another (foreshocks, aftershocks, multiplets). Marsan's algorithm is one of a range of methods known as *declustering* methods, which were developed to try to capture independent earthquakes when analyzing catalog data, distinguishing them from all other seismic events, notably those that correspond to aftershocks. The main challenge is to enrich the catalog in order to model seismicity as a process, wherein the occurrence of one earthquake alters the surrounding tension field and the capacity of nearby faults to generate other earthquakes. To model seismicity as a continuum of earthquakes, Marsan must be able to isolate classes of earthquakes, in order to integrate the fact that the tensions released by small earthquakes may be as large as those resulting from larger earthquakes in the locations where the seismicity occurs. It is in this spirit of isolating classes of earthquakes that Marsan observes the contribution of "memory" to seismicity in an area. In a critical methodological article on declustering algorithms, Marsan and coauthors surveyed the advantages of statistical learning, which free seismologists from the need to define a priori the statistical characteristics of the classes of earthquakes that seismicity consists of. Marsan is interested in artificial learning because it allows him to challenge the system of categories that he uses to investigate seismicity. He approaches the existence of the three classes of earthquakes—foreshocks, mainshocks, and aftershocks—with a certain methodological nominalism. For him, declustering must be used to test whether the conventional forms of earthquake classification are well founded. He closes the article with the following lines on this epistemic opening:

> Even though great progress has been made in the last decade, there are still many open questions, i.e., starting with the physical triggering of earthquakes (aftershocks), effects of uncertainties in the catalog on the results of declustering, or the effect of censored data (selection in time, space and magnitude range) on the outcome. In summary, care should be taken when interpreting results of declustering or results that depend on a declustered catalog, because these results cannot reflect the exact nature of foreshocks, mainshocks and aftershocks; indeed the exact nature of these events may not exist at all. (van Stiphout et al., 2012)

What if foreshocks, mainshocks, and aftershocks did not exist before being modelled? Marsan takes such an "agnostic" approach to modelling. He proposes to suspend beliefs regarding earthquakes, abandoning the idea of a pre-data structure that can simply be observed in the catalogs. In other words, Marsan expects machine learning to be able to be placed *underneath* categorical forms of seismicity (Cardon, 2016). At no point did Marsan see the algorithm that he programmed as a method of predictive analytics, because predicting aftershocks is not an end in itself in his research.[14] Aftershocks interest Marsan because they have the power to help him conceptualize the process of seismicity differently.

In their applied mathematics laboratory in Los Angeles, the PredPol developers use aftershocks in a different way. In their article they suggest that declustering methods can offer a means of enriching "crime catalogs," but do not expand on the point. Using prediction as a declustering method, as Marsan does for earthquakes, could contribute to research on the modelling of crime in general—a subject I have written about elsewhere, and will not pursue further here (Benbouzid, 2015)—which poses basic research problems no less complex than those of seismology. But the PredPol researchers are interested only in the "aftershocks" for the possibility they offer of adding an additional process to hotspot maps to incorporate regularities (repetitions). They use "crime aftershocks" (near repeat crime or near repeat victimization) for their capacity to capture the largest possible proportion of events.

The value of a prediction is inseparable from its practical consequences

Thus, on the spatiotemporal projection traversed by the algorithm, repetitions are what they do as a function of what scientists try to make them do. Between California and Chambéry, the status of repetition of crime changed, because they took up a place in two different institutional environments. In Chambéry, measured classes of entities exist in a domain where predictions lead to demonstrable consequences, which is not true in the case of policing. To illustrate this situation, in our interviews, Marsan often opened a historical parenthesis on the problem of demonstrable consequences in earthquake prediction. He mentioned the example of Parkfield as a symptomatic case of this problem in seismology:

> Parkfield is a little village on the segment of the San Andreas Fault that seismologists have transformed into an observation site, which is now considered the most instrumented place on the planet. An earthquake was predicted there in 1988, but it happened in 2004, 16 years late. Failures of prediction of this kind are not rare in seismology. They pose particularly serious problems.

In the politics of earthquake prediction, scientists are held directly responsible for false positives and false negatives. The experience of false positives (earthquakes that do not happen) leaves inhabitants with a feeling of generalized anxiety and causes large economic losses. The experience of false negatives, as in L'Aquila, Italy in 2009, clearly explains seismologists' reserved attitudes on their own capacity to provide robust predictions. Predicting an earthquake implies evacuating entire cities, which carries a considerable cost and can provoke dangerous panic reactions at the scale of the road network of an urban agglomeration.

In crime prediction policies, the PredPol platform works as a tool for the management of police action. PredPol's research has shown that by spending just 5% of their available time in the areas identified by the algorithm, police patrols are twice as effective (in terms of crime reduction) as when they patrol the hotspots classically identified by analysts. The accuracy of PredPol's claims is not very important. What counts is to be able to optimize, and above all to *precisely* control, this tactical allocation of police time to presence in high-risk space. By integrating data from the GPS tracking systems installed in police vehicles, the algorithm optimizes the dosage of the presence of police patrols in different sectors of the city: the predictive square remains red on the map as long as the police have not patrolled there, turns blue during their first movements through the area, and then green when the officer has spent the optimal period of time as calculated according to available resources (for example, 5% of a police officer's working day). For a sector manager, PredPol appears to be a good tool to ensure that police officers play their preventive role, often simply by way of their dissuasive presence, distributed randomly, but for an optimized duration, in the areas where risk is estimated to be greatest. The task of prediction is the management of the public supply of day-to-day police presence, while minimizing the need for change in police organization. While earthquake prediction has profound effects on the material and social structures of a city, crime prediction, as PredPol sees it, involves a minimal transformation in how policing is organized. David Marsan's meticulous, reserved, and prudent attitude can be understood as a habit developed in a field where researchers are held responsible for predictions that may have serious consequences. In contrast, the runaway success of PredPol can be attributed to the limited practical consequences of its predictions—hence the relatively casual manner in which the California researchers claim to predict crimes. This observation fits well with what sociologists have shown in studies of "theories, machines, and technology": that "their robustness, their solidity, their truth, their efficiency, and their usefulness depend less on formal rules or on their own characteristics than on their local and historical *context*—this independently of the various ways that there are of defining that context (Teil and Latour, 2017: 4). The robustness of earthquake or crime prediction is not the result of a rational calculation, validated by neutral researchers and integrated into a machine. It is a solidity composed of the actions targeted by the prediction and the network of the material elements that they imply. The moral of this controversy is that the robustness of a prediction is inversely proportional to its practical consequences.

The divinatory aspect of machines (conclusion)

To go further in this examination of relations between predictions and their consequences, a good source of inspiration is an article by the anthropologist Joel Robbins (2010) on deontological and consequentialist styles of reasoning. Robbins (2010: 124) refers to this style of reasoning, based on "appropriate rules and not on the consequences of one's rule-governed actions," as deontological. He contrasts this approach to morality with consequentialism, where "actions are judged by their results, not by how closely they conform to a given rule." Robbins (2010) then deepens this analysis of deontological reasoning, drawing on an article by Jane I. Guyer (2007), "Prophecy and the Near Future," on what she calls "the evaporation of the near future in theory and public representations" (Guyer, 2007: 410). Guyer shows how, in both contemporary economic policies and Evangelical discourse, the dual focus on the immediate present and the very long-term has taken the near future out of play as a temporal frame. Robbins (2010: 125) adds to Guyer's analysis that "[w]hat is lost in this move is the provision of a temporal space for [...] consequentialist reasoning," in favour of "deontological forms that do not need to refer to the near future world of demonstrable consequences to reckon the value of actions." According to Robbins, the success of evangelicalism can be attributed to the way in which the contemporary period creates uncertainty about the near future, which can no longer be predicted at all, leading individuals to concentrate on the present of their actions and to project themselves into a distant and mystical future. Here, respect for principles wins out over the anticipation of effects:

> Different styles of moral reasoning are embedded in different kinds of social circumstances, and [...] forms of moral reasoning only flourish in those social circumstances that are well suited to them. Consequentialist moral reasoning, for example, only works where people have a sense that the social world they inhabit is relatively predictable, such that the probable consequences of an action appear relatively easy to gauge with certainty. Where such conditions do not hold, deontological approaches make much more sense—even in situations in which one cannot control the consequences of one's actions, one can control whether or not they conform to a rule or set of rules (Robbins, 2010: 124).

This distinction is of interest here, as it analyzes two different moral approaches in relationship to forms of prediction, in direct parallel to the differences between the approaches of Marsan and PredPol. The social circumstance of the unpredictability of earthquakes might seemingly favour deontological approaches, but in reality Marsan's stance is consequentialist: Marsan pursues his research in the aim of making earthquakes more predictable, and thus to confer intelligibility on public announcements of the probabilistic theoretical construction of a phenomenon. PredPol gives the police the feeling of working in a more predictable world, but situations where police officers can directly observe criminals in the act are rare, even during discreet undercover patrols in the areas indicated with a precision of 200m x 200m. How, then, can PredPol claim to "predict" crime? A remark made by Sean Malinowski, the first Los Angeles Police Captain to experiment with the PredPol platform, offers a glimpse into what "prediction" means for the police: "If honestly done, there are no bad predictions in crime control."[15] Contrary to the seismologist, police officers cannot experience "failed" predictions, because in their practice, prediction is expressed not in terms of truth or falsehood, but in terms of "good" and "bad." The problem is not to believe or disbelieve in the machine's predictions, but to do something rather than nothing, following the machine's recommendations.[16]

It could be argued in response that the PredPol researchers adopted a consequentialist ethic, as they implemented a system evaluating the effectiveness of the algorithm's recommendations (Mohler et al., 2015) but the extent to which police can disrupt dynamically changing crime hotspots is unknown. Police must be able to anticipate the future location of dynamic hotspots to disrupt them. Here we report results of two randomized controlled trials of near real-time epidemic-type aftershock sequence (ETAS). But this evaluation bears only on very short-term consequences, and does not test the statistical significance of the measured decreases with respect to the general trends in crime over the long term.[17] It is difficult, or even impossible, for the police to assess the practical effects of their daily activities on long-term trends in crime. The PredPol software allows them to optimize their attempts to control an overwhelming social phenomenon (Manning, 2008). It is simpler for the police to rely on the dosages recommended by the machine: through the statistical learning procedure, crime takes form in a machine that produces rules to be followed by the police. PredPol remains a self-enclosed automaton, and like all automatons, it can provide only summary results (an analytical dashboard indicates upcoming risks by simply adding an additional process to hotspot maps to incorporate regularities (repetitions)). The seismologist, in contrast, makes use of the inductive logic of machine learning to take on a role as the continuous "regulator" and "organizer" of the predictive mechanism. In Marsan's hands, the algorithm became an *open machine* (Simondon, 2012), whose functioning can be deliberately modified, and which is used to *understand*.

To summarize, what distinguishes Marsan's approach from that of the PredPol developers is that the seismologist conceives prediction in terms of its practical consequences, and the developers conceive it in terms of an absolute duty to act. Predictive policing is deontological insofar as the principal question that it asks of the algorithm is "What must I do?" and not "What is the best possible world with respect to the consequences of my actions?" (Ogien and Tappolet, 2009). These two moral approaches to prediction are applied in two different practical temporal spaces. Marsan conceives of prediction on the scale of the near future, the time needed to provide supplies to an area or to evacuate a city, a time frame that requires him to conceive the moral dimension of his research activity in terms of foreseeable consequences. His ethic of responsibility pushes him to say that "We're incapable of prediction." PredPol's predictions are focused on the immediate present—real-time analysis—and not on the long-term consequences of the actions that are organized by those predictions. The attitudes of the Evangelicals[18] in Robbins's study and the integration of predictive machines into police organization may, in this way, be more similar than it seems. The former locate the future in the hands of God, the latter in the hands of a machine that police leaders hope will lead to salvation. When they work according to this deontological style of moral reasoning, predictive artificial learning machines are made not only of technology, science, and organization, but also of an element of divination.

References

Anderson MJ (1990) *The American Census: A Social History*. New Haven and London:Yale University Press.

Beer D (2009) Power through the algorithm? Participatory web cultures and the technological unconscious. *New Media & Society* 11(6): 985–1002. DOI: 10.1177/1461444809336551.

Benbouzid B (2015) From situational crime prevention to predictive policing. *Champ pénal/Penal field* (Vol. XII). DOI: 10.4000/champpenal.9066.

Bernard P (2003) *Qu'est-ce qui fait trembler la terre ? A l'origine des catastrophes sismiques*. Les Ulis, France: EDP Sciences.

Boltanski L and Thévenot L (2006) *On Justification: Economies of Worth*. Princeton University Press.

Cambrosio A, Keating P, Schlich T and Weisz G (2006) Regulatory objectivity and the generation and management of evidence in medicine. *Social Science & Medicine (1982)* 63(1): 189–199. DOI: 10.1016/j.socscimed.2005.12.007.

Cardon D (2016) Deconstructing the Algorithm : four types types of digital calculations. In: Seyfert R. and Roberge J *Algorithmic Cultures*. Routledge.

Crawford K (2016) Can an Algorithm be Agonistic? Ten Scenes from Life in Calculated Publics. *Science, Technology, & Human Values* 41(1): 77–92. DOI: 10.1177/0162243915589635.

Daston L (1995) The Moral Economy of Science. *Osiris* 10. 2nd Series: 2–24.

Deleuze G (2006) *The Fold*. London andNew York: Continuum.

Desrosières A (2002) *The Politics of Large Numbers: A History of Statistical Reasoning*. Cambridge and London: Harvard University Press.

Diakopoulos N (2014) Algorithmic Accountability Reporting: On the Investigation of Black Boxes. Available at: https://academiccommons.columbia.edu/catalog/ac:2ngf1vhhn4 (accessed 28 September 2017).

Domingos P (2017) *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*. 01 ed. New York, NY: Basic Books.

Dourish P (2016) Algorithms and their others: Algorithmic culture in context. *Big Data & Society* 3(2): 2053951716665128. DOI: 10.1177/2053951716665128.

Hacking I (1990) *The Taming of Chance*. Cambridge and New York: Cambridge University Press.

Hacking I (1999) *The Social Construction of What?* Cambridge and London: Harvard University Press.

Hofman J, Sharma A and Watts D (2017) Prediction and explanation in social systems. *Science* 355. Available at: https://www.microsoft.com/en-us/research/publication/prediction-explanation-social-systems/ (accessed 2 January 2018).

Ingold T (2013) *Making: Anthropology, Archaeology, Art and Architecture*. London and New York: Routledge.

Latour B (1987) *Science in Action: How to Follow Scientists and Engineers Through Society*. Cambridge, Mass.: Harvard University Press.

Latour B (2010) Prendre le pli des techniques. *Réseaux* 163(5): 11–31. DOI: 10.3917/res.163.0011.

Latour B (2012) *Enquête sur les modes d'existence: une anthropologie des modernes*. Paris : La Découverte.

Latour B and Venn C (2002) Morality and Technology. *Theory, Culture & Society* 19(5–6): 247–260. DOI: 10.1177/026327602761899246.

Latour B, Woolgar S and Salk J (1992) *Laboratory Life – The Construction of Scientific Facts*. Reprint. Princeton, N.J: Princeton University Press.

Mackenzie D (2004) *Mechanizing Proof – Computing, Risk and Trust*. New Ed. Cambridge, Mass.: MIT Press.

Main I (1999) Is the reliable prediction of individual earthquakes a realistic scientific goal? 397(1). Available at: http://www.nature.com/nature/debates/earthquake/index.html?foxtrotcallback=true (accessed 29 September 2017).

Manning PK (2008) *The Technology of Policing: Crime Mapping, Information Technology, and the Rationality of Crime Control*. New York: NYU Press.

Marsan D and Lengliné O (2008) Extending Earthquakes' Reach Through Cascading. *Science* 319(5866): 1076–1079. DOI: 10.1126/science.1148783.

Mittelstadt BD, Allo P, Taddeo M, Wachter S and Floridi L (2016) The ethics of algorithms: Mapping the debate. *Big Data & Society* 3(2): 2053951716679679. DOI: 10.1177/2053951716679679.

Mohler GO, Short MB, Brantingham PJ, Schoenberg FP and Tita GE (2011) Self-Exciting Point Process Modeling of Crime. *Journal of the American Statistical Association* 106(493): 100–108. DOI: 10.1198/jasa.2011.ap09546.

Mohler GO, Short MB, Malinowski S, et al. (2015) Randomized Controlled Field Trials of Predictive Policing. *Journal of the American Statistical Association* 110(512): 1399–1411. DOI: 10.1080/01621459.2015.1077710.

Ogien R and Tappolet C (2009) *Les concepts de l'éthique : Faut-il être conséquentialiste ?* Paris: Editions Hermann.

Olazaran M (1996) A Sociological Study of the Official History of the Perceptrons Controversy. *Social Studies of Science* 26(3): 611–659. DOI: 10.1177/030631296026003005.

O'Neil C (2016) *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown.

Pease K and Tseloni A (2014) *Using Modeling to Predict and Prevent Victimization*. London: Springer Science & Business Media.

Perry WL (2013) *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*. Santa Monica, CA: Rand Corporation.

Porter TM (1996) *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, New Jersey: Princeton University Press.

Robbins J (2010) On the Pleasures and Dangers of Culpability. *Critique of Anthropology* 30(1): 122–128. DOI: 10.1177/0308275X09360136.

Rouvroy A and Berns T (2013) Gouvernementalité algorithmique et perspectives d'émancipation. *Réseaux* 177(1): 163–196. DOI: 10.3917/res.177.0163.

Sandvig C, Hamilton K, Karahalios K and Langbort C (2014) Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms. *Data and Discrimination: Converting Critical Concerns into Productive Inquiry* 22.

Shmueli G (2010) To Explain or to Predict? *Statistical Science* 25(3): 289–310. DOI: 10.1214/10-STS330.

Simondon G (2012) *Du mode d'existence des objets techniques*. Paris: Editions Aubier.

Simondon G (2016) *On the Mode of Existence of Technical Objects*. Minneapolis: Univocal Publishing LLC.

Teil G and Latour B (2017) The Hume machine. Can association networks do more than formal rules ? *Stanford Humanities Review* 4(2). Available at: http://prodinra.inra.fr/?locale=fr#!ConsultNotice:153889 (accessed 9 February 2018).

van Stiphout T, Zhuang J and Marsan D (2012) Seismicity declustering. Available at: http://www.corssa.org (accessed 11 April 2019).

Walker S (2004) Science and Politics in Police Research: Reflections on Their Tangled Relationship. *The ANNALS of the American Academy of Political and Social Science* 593(1): 137–155. DOI: 10.1177/0002716203262699.

Weisburd D and Braga AA (2006) *Police Innovation: Contrasting Perspectives*. Cambridge and New York: Cambridge University Press.

Weisburd D, Bernasco W and Bruinsma GJN (2009) *Putting Crime in Its Place: Units of Analysis in Geographic Criminology*. New York: Springer.

Zarsky T (2016) The Trouble with Algorithmic Decisions. *Science, Technology, & Human Values* 41(1): 118–132. DOI: 10.1177/0162243915605575.

Ziewitz M (2016) Governing Algorithms: Myth, Mess, and Methods. *Science, Technology, & Human Values* 41(1): 3–16. DOI: 10.1177/0162243915608948.

Notes

1 This research was supported by a grant of the French Agence Nationale de la Recherche to the project "Innovation in Expertise. Modeling and simulation as tools of governance" (*ANR-13-SOIN-0005),* coordinated by David Demortain.

2 PredPol is a small Santa Cruz start-up developed along the same path as many other California businesses. In 2010, two entrepreneurs — Caleb Baskin and Ryan Coonerty (also Third District Supervisor for Santa Cruz County) — approached two California researchers, George Mohler (associate professor of applied mathematics) and Jeffrey Brantingham (an archaeologist specializing in the Upper Paleolithic in northern China, and the son of two well-known criminologists who did pioneering work on the geography of crime), with a view to converting the fruits of their research into a profitable business with a strong growth potential. Although the research that underpinned PredPol was publicly funded, the start-up was created with 1.3 million dollars invested in 2012 by a handful of business angels. Following a business process that proved itself in spectacular fashion within two years, owing in particular to the efforts of its lobbyists operating in the Democratic networks of California, the firm was launched in a second round of venture capital fundraising (2.4 million dollars raised in 2014) in order to extend its commercial activity. At the time of writing PredPol is a commercial web-based system deployed in a number of policing departments in the United States and the UK.

3 David Marsan's professional web page: https://www.isterre.fr/annuaire/pages-web-du-personnel/david-marsan/?id_auteur=131

4 I exchanged with David Marsan several times between February 2015 and March 2017. The quotations are drawn from my fourth interview with him in Chambéry in April 2016.

5 I made Marsan's analysis public in an article published on the web site of the magazine Vie des idées (Benbouzid, 2016), as subsequently reported by Mediapart (Hourdeaux, 2016).

6 Echoing Deleuze, Latour encourages us to *unfold* the technical action: "I would like to define the regime proper to technology by the notion of fold, without giving it all the Leibnizian connotations that Gilles Deleuze (1993) has elaborated so well. What is folded in technical action? Time, space and the type of actants" (Latour, 2002: 248). In a methodological point of view, it means observing and following empirically what the PredPol's algorithm is concretely made of, in order to make its components visible.

7 PredPol's launch strategy was largely based on this slogan. On the basis of this marketing slogan, the public relations arm of PredPol sought to win over police leaders by convincing them that the PredPol system represents an improvement on "hotspot policing," one of the ways that proactive policing has been labeled since the 1990s. On their website (www.predpol.com), trials of the program are systematically associated to a decrease in crime of around 20%, a larger decrease than in sectors where the program is not used.

8 Marsan and his collaborator Lengliné published a widely noted article in statistical seismology in Science in 2008 (Marsan and Lengliné, 2008). This is the article that the PredPol mathematicians cite in their own article (WHO?, 2011). The statistical method that Marsan developed was integrally

transposed, aside from the elements of translation that were essential to adapt the algorithm to the constraint of operationalization.

9 To represent the spread of victimization in statistical terms and to identify the more or less repetitive spatio-temporal configurations on which prevention strategies could be built, researchers apply spatial analysis statistical tools from epidemiological research. The analogy of contagion comes in 1990s with the notion of "communication of the risk of victimization". For example, the mechanism of contagion corresponded fairly well to the results of qualitative surveys run on burglars. Burglars had told researchers that they regularly returned to burgle the same house when it was easy to burgle and they had not been able to take everything the first time around. Burglars moreover operate by neighbourhood. The notion of "infectious burglaries" has been used to explain why victimization spreads in time and space. (Pease and Tseloni, 2014)

10 For a history of the controversial origins of neural networks, see Olazaran (1996). I analyze the controversy within Artificial Intelligence (AI).

11 The opacity of algorithms has become a commonplace observation: not only is the source code of machines usually protected as a trade secret, but it also describes a process of artificial learning that is so complex and that involves so many variables that the results are difficult to interpret even for the specialists themselves. Whether intentional or not, opacity is understood as a central problem in current public debate in all countries where the problem of algorithms has reached the political agenda (Mittelstadt et al., 2016).

12 Marsan's complete critical note can be consulted through the article published on online news web site *Mediapart*, "Police prédictive: deux chercheurs démontent l'algorithme" [Predictive policing: two researchers take apart the algorithm], 13 September 2016.

13 As Daston emphasizes, whereas "accuracy concerns the fit of numbers or geometrical magnitudes to some part of the world and presupposes that a mathematical model can be anchored in measurement [...] precision concerns the clarity, distinctness, and intelligibility of concepts, and, by itself, stipulates nothing about whether and how those concepts match the world." (Daston, 1995: 8).

14 Although these aftershocks threaten the safety of rescue workers searching through ruins for survivors, Marsan's research does not aim to improve the prediction of their occurrence.

15 Interview with Sean Malinowski, August 2013.

16 The scientists at the startup undertook a serious evaluation of efficacy of the algorithm's algorithms, in the tradition of quasi-experimental methods (Mohler et al., 2015)but the extent to which police can disrupt dynamically changing crime hotspots is unknown. Police must be able to anticipate the future location of dynamic hotspots to disrupt them. Here we report results of two randomized controlled trials of near real-time epidemic-type aftershock sequence (ETAS), which may seem to suggest that PredPol's approach is not deontological. Measuring efficacy is indeed a way of judging police actions in terms of results. However, the experimentation test of PredPol was punctual: The cities that use the platform do not systematically carry out a randomized trial experiment. The assessement appears in the continuity of PredPol's marketing plan, rather than in a logic of "regulatory objectivity" (Cambrosio et al., 2006) as in seismology with the program "Collaboratory for the Study of Earthquake Predictability". http://www.cseptesting.org/.

17 The assessment implemented by PredPol appears more in the continuity of the start-up's marketing plan than in a logic of "regulatory objectivity" (Cambrosio et al., 2006) as in seismology with the "Collaboratory for the Study of Earthquake Predictability (CSPEP)" programme. http://www.cseptesting.org/, accessed 25.10.2017.

18 Robbin's analysis sheds light on the profession of "technological evangelist" (or Chief Evangelist Officer), which emerged in the world of information technology.

Inventing Prediction for Regulation: The Development of (Quantitative) Structure-Activity Relationships for the Assessment of Chemicals at the US Environmental Protection Agency

Henri Boullier

INSERM, Institut de Recherche Interdisciplinaire sur les enjeux Sociaux (IRIS), France/henri.boullier@inserm.fr

David Demortain

INRA, UMR 1326 Laboratoire Interdisciplinaire Sciences Innovations Sociétés (LISIS), France

Maurice Zeeman

Retired Senior Scientist, U.S. Environmental Protection Agency, USA[1]

Abstract

In policies targeting environmental and health hazards, an effort is frequently made to anticipate and avert more or less probable adverse events. In this context, computerized models are often portrayed as superior knowledge tools, for their capacity to extrapolate from existing data and predict hazards. This paper looks at the historical development and use of such models in regulation, with the specific example of structure-activity relationships (SARs) in the regulation of new industrial chemicals at the US Environmental Protection Agency (EPA). It asks how evidential culture(s) in a regulatory organization change, in particular how new methods and forms of knowledge find their place alongside others to forge regulatory decisions. The development and application of, first, a qualitative approach to structure-activity relationships, and then of quantitative models, show that the EPA had the necessary autonomy to imagine and adjust a method emerging in the research environment to respond to regulatory needs. This can be understood from a coproductionist perspective, if adjusted to take into account the bureaucratic knowledge that mediates the imagining and application of prediction in regulatory practice.

Keywords: Structure-activity relationships, QSAR, SAR, modelling, prediction, Environmental Protection Agency, regulatory knowledge, industrial chemicals, regulation

Introduction

In policies targeting environmental and health hazards, an effort is frequently made to anticipate uncertain adverse events, calculating the probability of their occurrence in the future (Sarewitz and Pielke Jr, 1999; Nelson et al., 2008)[2]. A number of scientific disciplines have taken on this ambition to predict risks, developing such tools as computational models, and associated software technologies, to do so. In the past decades, these tools and the underpinning practices have expanded, to be more and more routinely incorporated in the processes of risk assessment. Computational knowledge is used alongside experimental evidence and situated observations of risk, to better extrapolate from existing data and predict safety issues. Computational models help bring the future of human risk to bear on present decisions (Adams et al., 2009; Montgomery, 2017; Rajan, 2009).

This paper asks how model-based predictions come to constitute a routine form of knowledge for regulatory agencies. The use of computational models to predict risks is a case of change in the forms of evidence that a regulatory agency uses. This change affecting the way in which an organization knows risk and make decisions is puzzling in itself. First, regulatory agencies are generally constrained to use certain kinds of evidence, under the influence of legal frameworks and of representations of what is credible scientific knowledge. There are norms that define what counts as regulatory knowledge (Demortain, 2017). Second, preferred forms of knowledge tend to institutionalize in the organization. They materialize by roles, identities and boundaries that are difficult to change thereafter. Third, the knowledge that regulatory agencies generate is generally subjected to rigorous trials of credibility and deconstruction games, in courts and in other arenas (Jasanoff, 1990; Hilgartner, 2000). An agency is seldom in a position to impose the validation of new methods or claims itself – validation being the key question when it comes to using models to make predictions (Oreskes, 1998). Because of these circumstances, one can assume that a science-based regulatory organization has a limited capacity to choose and evolve new forms of regulatory knowledge, or to shift from one kind of science to another. How then have computational methods emerged as a form of science-for-policy in an organization marked by high level of constraints on the demonstration of risk? How have these computational methods grown into an element of the evidential cultures practiced in a regulatory agency?

These questions are applied to the case of models of structure-activity relationships (SAR) at the US Environmental Protection Agency. In the past three decades, such models have been developed and applied in greater amounts in EPA's regulatory assessment of chemicals. In SAR, 'structure' refers to the molecular structure of the chemical under consideration, while 'activity' stands for the biological activity, including toxicity that the substance may cause in the body. The correlation between the structure of a molecule and the toxicity it causes in the human body can be established qualitatively by a chemist or toxicologist with experience in toxicity testing, simply by looking at the structure of the substance, identifying within it a particular element or 'fragment' that, in an experimental study that was seen in the past, caused some kind of toxicity. A quantitative structure-activity relationship, or QSAR model, is a statistical correlation between a chemical property that is common to a class of chemical substances (e.g. solubility) and a frequent biological effect of that class, as established in animal experiments. Once a correlation is established, it can be used as a benchmark to infer the potential toxicity of a chemical for which no test data is available, without further (or with limited amount of) animal experimentation, if this chemical has the same structure-related property as the class of chemicals for which the correlation has been established. Such 'in silico' methods (as opposed to studies performed 'in vivo' or on animals[3], and those conducted 'in vitro', on cultured cells) are often presented as an alternative to animal experiments because modelling is future-oriented, predictive, and is not affected by uncertainties surrounding extrapolation of results in animals to future, human conditions. It is the US Environmental Protection Agency (EPA) that early on invested most resources in the development of SAR tools and models, turning it into an accepted and credible way of knowing the hazards of new industrial chemicals, now used across the world and particularly in Europe, as Laurent and Thoreau (2019) show in this very issue. The agency formalized these methods as part of the implementation of the Toxic Substances Control Act (TSCA)

– an act that did not foresee the use of computational methods, but only authorized the agency to conduct or request "epidemiologic studies, serial or hierarchical tests, in vitro tests, and whole animal tests" (Anonymous, 1976: 2007)[4].

The history provided in this paper shows that the agency did not only import quantitative modelling from the outside, namely from the pharmaceutical industry, but actively constructed an original, qualitative technique of anticipation of risk based on the consideration of molecular structures by experts. It made use of quantitative models, derived from large, validated sets of experimental data produced in house only gradually, after designing this qualitative way of making structure-activity correlations. So, predictive modelling gained credibility as a regulatory tool very progressively and only because the agency designed an initially limited, contextual use of predictions to prioritize substances, as a complement to other kinds of data and studies used to assess risks. This invention can only be understood if we take into account the organizational context by which the emerging scientific order – the supposed capacity to predict safety quantitatively, without experiment — meets and interacts with the political or legal order — the requirement to make decisions without data imposed on the EPA by the law-makers.

This research makes use of three different sources of information. The first are the official numerical archives of the EPA, from which we retrieved several dozens of documents produced by the agency about structure-activity relationships. Second, we interviewed nine officials of the agency who were, or still are, in charge of structure-activity assessment in its Office of Toxic Substances (OTS), or of research on structure-activity relationships on the side of the Office of Research and Development (ORD). Third, we relied on the deep knowledge of the agency and of relevant events, methods, decisions and people of one of us (Maurice Zeeman), who was in charge of the Environmental Effects Branch (EEB) in the Office of Pollution Prevention and Toxics (OPPT) between 1988 and 1997, with responsibility for the supervision of the group of QSAR specialists of this office. In the remainder of the paper, we begin by giving some background concerning today's importance and use of SAR. We then move to the history of what happened at the EPA during the 1980s and 1990s, before engaging in a discussion and analysis of this history, bringing out the forms of bureaucratic knowledge pertaining to the EPA, which explain how it could shape and apply structure-activity thinking in regulatory work.

Evidential cultures and the regulation of risk

The rise of modelling as a practice underpinning judgement about chemical hazards is a problem in regulatory knowledge that concerns the formation and use of particular bodies of evidence to justify a decision and intervention. There are different ways of producing evidence of a risk, and of the need for intervention, just like there are different ways of producing a proof in any "pure" scientific discipline or in fundamental research. One way to characterize this conflict is to distinguish between the various ways of producing regulatory evidence of a risk, or 'evidential cultures' that coexist or compete within the same regulatory regime to constitute the norm of objective regulatory knowledge, particularly in controversial policy environments (Böschen, 2009; Demortain, 2013). The notion of 'evidential culture', first articulated by sociologist Harry Collins (1998) as part of his ethnographic study of gravitational wave research, broadly refers to strategies and criteria that frame the collective validation of knowledge.

For our purpose, we will draw from the framework advanced by Böschen (2009, 2013), who distinguishes among four different evidential cultures in chemical regulation — restrictive, holistic, instrumental and evaluative. According to Böschen, a restrictive evidential culture, first, rests primarily on experimental methods and on the possibility, in controlled laboratory settings, to verify toxicity at a given endpoint in an animal model, and establish causality between a dose of chemical and that endpoint. That culture takes form in the context of a regulatory regime which names and concentrates on these endpoints and individual chemical objects. The ambition to establish proofs for such causalities, and the high evidence threshold[5], has its drawback, namely the reduction of the phenomenon being evaluated (e.g. toxicity in a human population is reduced to toxicity measured in a limited population of rats). In the holistic culture that is typical of ecotoxicology, experiments may be combined with other tests and knowledge evaluating the phenomena at other scales of biological organization. The interest is less in discovering simplified causal relationships than in capturing complex interactions between elements of an ecosystem – which means that this culture is less operational, and its agents less frequently consulted for making individual regulatory decisions on a given chemical, than for framing, delimiting an issue in the first place. The third, instrumental culture is oriented towards the development and use of instruments to detect and produce data to estimate the existence of a problem in a new context (embodied by analytical chemistry in the case of environmental chemicals). The disciplines of environmental medicine, or the exercise of hazard assessment, embody a fourth, evaluative kind of culture, where the epistemic goal is less to explain and precisely predict – the evidence threshold is not high – but to address practical problems as they arise, evaluating them against the background of other problematic situations to determine a level of response (Böschen, 2013: 77-80).

Historians, philosophers and sociologists of science alike agree on the fact that modelling is a practice of mediating between experimentation and theory, of creating a fit between the data that surge from experimentation or observation, and available theory. They allow "experimenting on theory" (Dowling, 1999: 261) with constantly renewed sets of data (Morgan, 1999), gluing one and the other (Sismondo, 2006) to be able to use conclusions emerging from experiments to learn about untested situations. This is what makes models more or less useful, possibly a more relevant benchmark for evaluating them than "truth" (Box, 1979; Sismondo, 1999; Zeeman and Mayo-Bean, 2009; Wambaugh, 2014). Models thus have a possible role in each culture. A restrictive culture, for instance, incorporates statistical causal models, to establish links between the experimentally measured variables and the tested object. Analogical models are necessary to apprehend the complexity of systems, and contemplate the relationships between parts of this system. Instruments of detection and measurement cannot function outside ontological models and classifications, that define, delimit or demarcate the thing being measured. Finally, an evaluative culture oriented towards the definition of practical solutions, will employ analogical, physical or statistical models to be able to simulate the effects of a given change in the system. The introduction of structure-activity thinking at the EPA means that evidential cultures in use in the organization evolve either in the direction of a restrictive culture (using quantitative models to find correlations between two reductively considered thing, a molecule and a given toxicity endpoint), or an evaluative one (using analogies to produce signals of safety, and justify a pragmatic decision to further test a chemical).

The coproduction perspective developed by Sheila Jasanoff (2004) offers a way to analyze such change. From that perspective, the evidential culture evolves under the influence of two mutually influencing dynamics: changes in what defines the regulatory order, or more simply the regulatory regime that the agency operates; changes in what counts as valid scientific knowledge in the corresponding scientific fields. In other words, changes in the evidential culture of an organization like the EPA is coproduced by two emerging changes in the realm of the law and in the realm of scientific knowledge. The making of structure-activity correlations at the EPA would be the result, from this perspective, of an emerging legal order in which the agency is requested to make decisions about risks in the absence of data, interacting with an emerging scientific discipline of quantitatively predicting toxicity problems in chemicals. Both comforted one another in the constitution of a credible regulatory knowledge culture.

Coproduction, however, does not take place in the abstract, but in concrete organizational conditions, that need to be taken into account to understand the particular kind of modelling that was identified, chosen and encultured at the EPA. In one of the rare references to this organizational dimension in the coproductionist literature, Jasanoff mentions that the "making of institutions" — the emergence of "tried-and-true repertoires of problem-solving", or "administrative routines" that provide ready-made solutions to political problems and controversies — is one of the "pathways" of coproduction (Jasanoff, 2004: 40; see Waterton and Wynne, 2004 also Hunt and Shackley (1999) had a few years earlier noted that what they call "bureaucratic knowledge" — the heuristic guides, aids or frameworks that help an organization achieve politically feasible and legitimate outcomes — plays an important role in defining what emerges as science-for-policy, at the junction of science and law (that they called, respectively, academic knowledge and fiducial knowledge). Bureaucratic knowledge designates the heuristics that are shared across the organization, to know how to form a final and credible decision of the organization as a whole, beyond and above the boundaries that separate the members of the organization, notably the specialists of various kinds of science, lawyers, political decision-makers (Bijker et al., 2009). It is specific to a regulatory organization, whose main ambition and challenge it is to precisely turn out credible decisions with a variety of knowledge bases and criteria, observable inside and outside the organization. This bureaucratic knowledge, existing or emerging, influences the definition of appropriate science-for-policy. It mediates the interpretation of scientific affordances and legal mandates to orient the definition of what kind of evidence is most appropriate.

In summary, we assume here that the forms of knowledge that are incorporated in the practice of regulatory organizations are the product of an organizational interpretation of legal constraints and scientific capacities; a process in which the bureaucratic knowledge of the organization in charge – its formalized experience of the coordination among participants in the formation of a decision — plays a key role. This bureaucratic knowledge was particularly important in forging a particular kind of structure-activity reasoning in the agency, mostly evaluative, and distinct from the kind of restrictive QSAR that was then emerging in the field of quantitative drug design. We now turn to the history of the development of structure-activity reasoning and computation models in the EPA. The discussion section then returns to the descriptions of these three orders of change — in the legal order, in science and in the bureaucratic knowledge pertaining to the EPA — to explain how structure-activity reasoning has become regulatory knowledge at the EPA.

Qualitative and quantitative structure-activity relationships at the EPA

Preparing for the review of new industrial chemicals without data (1973–1979)

A QSAR model is a statistical analysis (by regression or classification or else) of the biological activity of a group of two or more chemicals that have some structural similarity, as captured through a chosen descriptor of the chemical[6]. The modelling of causal relations between chemical properties and biological impacts is rooted in fundamental chemistry (Crum-Brown and Fraser, 1868; Meyer, 1899; Overton, 1899). The *quantitative* approach towards these correlations was pioneered by a Professor of Chemistry at Pomona College in California, Corwin Hansch, now known as "father" of computer-assisted molecule design[7].

The interest for QSAR modelling at EPA emerged in the mid-1970s, thanks to connections between the agency and this emerging work of computer-assisted drug design. At that time, the passage of the future TSCA was already under discussion. The proposition to have a dedicated status for the control of chemical substances emerged in the 1960s under the pressure of public interest groups, to comprise what will soon be known as the new social or risk regulation: regulatory regimes dedicated not to the control of markets and economic activities, but to the improvement of health, environment and working conditions (Harris and Milkis, 1989). The new Act was designed to cover the kinds of chemicals that were not already regulated via provisions applying to food additives, pesticides or medicines. A whole continent of industrial chemicals, many suspected to be toxic, had escaped the legislation in place (Vogel and Roberts, 2011). The Council on Environmental Quality suggested in a report in 1971 to develop new legislation to cover all of these chemicals, and to generate information about them in the first place, as many were simply not known or registered. It started being discussed soon after the establishment of a federal environmental agency, the EPA, was decided, in 1970. It was one of the first Acts that the agency would be entirely in charge of, from the start, and applied the sort of holistic perspective that inspired the creation of a dedicated environmental agency.

During the final years of the discussion of the Act — it was finally adopted in 1976, after six years of negotiation — the EPA's newly OTS was starting to realize that it would eventually have to handle the rapid evaluation of large amounts of unknown, new chemical substances. During the negotiation of the Act, the chemical industry succeeded in convincing Congress and the executive to withdraw most of the requirements for mandatory testing from it. At the end of the day, implementing the Act appeared as a challenge for the agency that had to prove scientifically the existence of risks, to regulate products, without any possibility to execute or require scientific studies from the industry, even though tests were available and already routinely applied by corporations of the sector (Craeger, 2018). The officials of the EPA had to very quickly operationalize an approach to deal with the evaluation of new chemicals for which no testing and no data were going to be available (since there was no obligation for companies to do testing). The discussions revolved around the need for "identification" of chemicals and methods for "early warning". By the end of 1973, the EPA's toxic substances staff had already identified a significant body of scientific literature concerning structure-activity correlations and methodologies, without making any clear-cut decision as to the potential which these methods might have for helping EPA in its early warning activities (FRI, 1975, 1976). At the beginning of 1974, the view according to which key properties of substances, notably their toxicity can be derived from their structure was accepted (EPA, 1975).

Structure-activity interests in the EPA's Office of Toxic Substances (OTS) and Office of Research and Development (ORD)

The interest for the modelling of structure-activity correlations crystallized simultaneously in two separate places in the EPA. The benefits of using structure-activity correlations to formulate judgments about the safety of new chemicals became clearer as Joseph Seifter, a medical doctor by training and a pioneering pharmacologist, joined the EPA's OTS in 1978 from the George Washington University Medical School. Seifter had approached structure-activity work as part of his research in pharmacology and drug development. He contributed to establish the so-called 'structure-activity team' (SAT) in the OTS in 1979, to prepare for the incoming of the first new chemical 'pre-manufacture notifications' (PMN). These notifications applied to any new chemical, meaning a chemical substance that was not on the inventory of existing chemicals prepared by the OTS and published in July 1979 (Hepler-Smith 2019). After that date, there could now be a 'new chemical PMN' submitted to OTS's New Chemicals Program (NCP) by industry.

The interest for QSAR modelling also crystallized in another place in the EPA, the Office for Research and Development (ORD)[8]. Gilman Veith, a scientist in one of the laboratories of the ORD, the Environmental Research Laboratory (located in Duluth), got interested as early as in the mid-1970s by this notion of prediction of toxicity from chemical structures, and devoted enormous efforts to the development of that science[9] (Veith, 1981). In 1975, at about the same time as the EPA's toxics staff started to conceive of SAR as a possible approach to deal with the chemicals 'data gap' in the upcoming TSCA, he initiated the 'QSAR research program' of his laboratory (Bradbury et al., 2015: 17). He developed a clear vision of the necessary tools for the QSAR approach to become applicable and had the necessary leadership skills to have people work together, both in his lab and between his lab and the EPA's new OTS, to develop these tools.

One of the early projects that Veith and his group launched was the development of high-quality databases of experimental results, on which to compute correlations and estimations of toxicity. Two databases were developed. The first came to be known as ECOTOX, and was essentially a collection of experimental results presented in scientific journals. The second developed from dissatisfaction with literature-derived databases. The results collected from the literature are never fully comparable: even where two experimenters test the same substance using a similar protocol (e.g. administration of a 50% concentrated dose of a substance to rats for 28 days), these protocols necessarily differ in some dimensions. The resulting database of toxicity measures is not of sufficient quality to compute robust mathematical correlations with chemical structures. Veith thus came to the conclusion that he would need to do the experiments at home, replicating a strictly identical protocol on a large number of substances, and collecting the result in an operational database. To do so, a large testing program was set up in 1981, in close cooperation with, and funded by, the OTS. It consisted in performing a short-duration toxicity test on a fish (96 hours, on fathead minnows), to derive the LC50 value of several hundred of chemicals.

Structure-activity in the regulatory practice of OTS

Only eight PMN were submitted to OTS in 1979, but their numbers grew quickly. There were almost 1,000 PMNs submitted to OTS by September 1981. 900 more PMNs had been submitted by September 1982, and another 1,400 PMNs by September 1983. The PMN process consequently resulted in an average of about 1,600 submissions per year, in 32 years of application[10].

The Act imposed a 90-day limit to the agency but did not impose the industry to provide any data to the EPA for it to do that estimation (other than data the company already has). By law, a PMN submission dossier includes the name of the chemical, a description of its structure, the production volume, methods of uses and disposal, estimates of human exposure, and any extant test data obtained. Nevertheless, in approximately 65% of cases, submissions by the industry did not include any substance-specific experimental data. The information the EPA got was the name of the chemical, a description of its molecular structure, the volume of production, the uses and disposal methods, and estimates of the number of people in the general population that will be in contact with the substance ('human exposure'). Only 45% of the dossiers included health test data, mainly for acute toxicity endpoints, genotoxicity test results or local irritation studies (EPA, 1984). There was little, if any, ecotoxicological or physical/chemical fate data submitted (e.g., Auer et al., 1990; Auer et al., 1994; Zeeman et al., 1993; Zeeman, 1995; Zeeman et al., 1995). If the EPA wants to get test data from the manufacturer, it has to make a risk-based case for it, an obviously difficult thing to do with minimal information in hand or, in regulatory science terms, in a 'data-poor' situation. The sheer volume of substances to assess, coupled with the absence of data, rendered the perspective of making predictions from structure-activity correlations in similar chemicals, particularly attractive, if not a necessity.

The first step in the PMN process designed by the OTS was determining that all necessary information has been included in the notification. This was followed by a series of three meetings (1. Chemistry Review and Search Strategy; 2. Structure Activity Team; 3. Exposure Analysis Meeting) which bring senior level expertise to bear on the questions of chemistry, hazard, and exposure within the first 15 days of the 90-day period available to EPA for the assessment of each new chemical. The 'Chemistry Review' meeting would be held between from days 8–12. 30% of substances, on average, would be left off the hook at this stage[11] during which the chemical identity of the substance is considered, the methods by which it is synthesized and the feedstocks used for the process, the physico-chemical properties of the substance. The remaining 70% would then be considered in the 'structure-activity meeting' between days 9 and 13 of the process.

The structure-activity work was prepared within the Health and Environmental Review Division (HERD) of the OTS, which provided the scientific and technical support for chemical assessment[12]. In the terminology of the paradigm set in the National Research Council report on risk analysis (NAS, 1983), then being institutionalized in the agency (Demortain, 2019), the structure-activity work contributed to the first step of the process, 'hazard assessment' – that is, the mere identification of toxicity problems, without measurement of their gravity, frequency or probability of apparition (the heart of the exposure assessment and risk characterization stages) (Zeeman and Gilford, 1993).

In structure-activity meetings, one chemist was assigned the task of summarizing the profile of the substance. The physico-chemical properties were then considered, followed by the environmental fate of the substance, the health issues (metabolism, mutagenicity, neurotoxicity…), to conclude by consideration of the ecotoxicity of the substance. Inferences from what was known of the properties of a given chemical structure infused the work of the whole team.

Structure-activity meetings were not a meeting of modelers discussing mathematical models and numerical estimates. These were "professional judgment" meetings (EPA, 1984: 4), where people used their experience of the toxicity typically associated with a kind of chemical structure, to anticipate the safety issues that might arise from exposure to a new chemical with a structure that they deemed comparable, or sufficiently similar, to those for which they had prior experience of toxicity. These meetings served to elicit the views of experts. They were patterned, in effect, after the Delphi method — a collective forecasting method based on successive rounds of questioning of a group of experts, developed by the US Army forces — in which OTS people found inspiration. Auer, who headed the structure-activity team from 1979 to 1986, described it in this way:

> We had the first meeting. You have the [chemical's] structure and you have the little or no data that were available on it, and you just started going around the room. What do you think? You know, in your area of expertise, what can you offer about this chemical? Over time that evolved into a very regularized approach to decomposing the chemical, and then through SAR [structure activity relationships], putting it back together to tell the story. […] Pretty quickly, within probably the first year of the operation of this program, you had a regimen in place where you had done preliminary chemistry analysis. So, what kind of chemical is it? How does [the chemical] function in its use? … (CHF, 2010).

Of all the necessary resources to do QSAR modelling (availability of test data on the substance being examined, data on analogous substances, statistical methods to analyze them and 'professional judgment'), the knowledge and professional judgments of scientific assessors in the interpretation and integration of available information, was "the most critical in terms of the overall success of the evaluation effort" (EPA, 1984: 13). In regulatory practice, then, modelling was a mode of reasoning applied to accumulated experience, to form a hypothesis about the lack of safety of a substance in anticipation of any experiment or observation. This was, in essence, a qualitative kind of structure-activity analysis, based on knowledge gained from reading masses of experimental data published in the literature, an experience that toxicologists classed by chemical categories, themselves defined by molecular structures. This knowledge was deposited in people, and exercised by them during meetings in what became a particular kind of competence in making analogies between substances. Auer, again, recollects that

> there were smart people on the team who could say 'Jeez, this substance looks a lot like that case we had a year ago'[13]

such as Joe Seifter, recalled as

> one of the early practitioners of the concepts of forming categories of chemicals [and] looking for ranges of toxicity across a category, being sensitive to where the toxicity shifted in a category, and then, attempting to understand mechanistically what was going on to cause that shift. He was just a remarkable guy, encyclopedic knowledge. You could show him a structure and he could just tell you what kinds of things it was likely to do to a human. (CHF, 2010).

Charles Walker was another of these experts that came to the EPA from the *U.S. Fish and Wildlife Service* to help apply this practical analysis (Lipnick, 1998). According to their colleagues, the scientists of the OTS, Robert Lipnick, Richard Clements and Vincent Nabholz in particular, were said to have develop a great ability in that exercise of inferring possible toxicity issues from reading chemical structures over time. More people soon joined the team, extending this repository of embodied knowledge of structures and toxicity that compensated for the absence of experimental data in industry notifications. Paul Bickart, a Harvard-trained chemist with a very broad background, contributed his capacity to characterize chemicals. Joseph Arcos, a university-based chemist had joined the SAT in 1979 already, bringing his vast knowledge of chemical carcinogenesis to the team. Adrian Albert, an Australian professor in medicinal chemistry, spent the summer of 1982 working with the SAT too.

Structure-activity considerations were not equally useful for all kinds of toxicity and endpoints, however. They were mostly useful for the environmental fate or ecotoxicological issues, for one simple reason: these were the issues for which experimental results or what Auer calls a "base set of data[14]" (generated through acute tests, such as fish tests) could be generated more quickly, assays being short and relatively less expensive than the tests on rodents used for human health outcomes. This analog chemical assessment was *not* the basis of final regulatory decisions. The final decisions resulted from a more complete risk assessment process and the vast majority of such decisions could be made without any testing needed. Some of the more difficult decisions required the consideration of data produced by the company *after* the initial structure activity meeting, and following the indications of the structure-activity team. But chemical analogues served to anchor the assessment that problems may arise from exposure to that substance. It was sufficient to meet the standard of proof established in the Act to justify requiring data from a company: that is, that it "can reasonably be determined or predicted" that the substance "may present an unreasonable risk of injury to health or the environment" (Anonymous, 1976: 2006).

Criticism, doubts and progress towards quantitative SARs

Although the structure-activity team grew over time, as notifications started to pour in, from around six to a dozen people, the (Q)SARs approach was not consensual. Skeptics of structure-activity inferences could be found either in the agency or outside, among academics of corresponding fields for instance. Several reviews of EPA work on structure-activity emerged just a few years after the initiation of the structure-activity team. EPA's use of SAR in reaching PMN hazard assessment conclusions soon started to be questioned by Congress, environmental groups and others (OTA, 1983; GAO, 1984; ACS, 1984) who point out the many uncertainties associated with the approach (EPA, 1984).

Adrien Albert, a professor at Australian National University specialized in structure-activity relationships who participated in several of the team meetings in 1982, undertook a review of the work of the EPA on structure-activity correlations. In his report, he noted that the EPA heavily relied on the professional expertise of scientific assessors, and on the exercise of relating a whole molecule to a class of chemicals for which adequate biological data exist, much more than on quantitative structure activity relationships, which was limited by the lack of toxicity data and was based on physical chemical property data, and QSAR descriptors (EPA, 1984).

At about the same time, a paper appeared in the journal Environmental Science and Technology, quoting Corwin Hansch questioning the use of SARs in a regulatory environment. The pioneer and "father" of quantitative SARs thought that the approaches employed by the EPA differed from his own for two reasons. Contrary to their applications in the pharmaceutical industry, the EPA did not focus on a single endpoint but multiple pathways, which made the objectives and contents substantially different. The lack of data also made the EPA approach very different: both experimental data for the health effects of PMN chemicals, but also more basic data of physical chemical properties were missing. As a consequence, Corwin Hansch judged that EPA's approach was useful but insufficient:

> While SARs can be very helpful to regulatory agencies in deciding which chemicals should be subject to special testing—EPA is doing this now—I believe that you cannot yet base regulations on SARs. In other words, SARs are not yet ready to use for confirming or denying market access to any given chemical, but they are of use, and are being used to guess which may be especially toxic or relatively safe (Anonymous, 1984).

Debates emerged as to whether the chemicals between which comparisons were made were sufficiently analogous, but also about the extent to which SARs could ever predict other forms of toxicity than acute toxicity, for which there was more biological data available than the rest. Criticism over EPA's PMN review process also stemmed from the Government Accounting Office and the Congress' Office of Technology Assessment, for which uncertainties were pervasive in toxicity assessment and in experimental tests "on the exact product, not closely related chemicals, are necessary to ban or restrict production" (OTA, 1984: 77).

These reports, while critical of quantitative structure-activity modelling, were not altogether depreciative of the use of analogies between groups of chemicals to screen large numbers of chemicals, and select those on which more data would be requested, and that would undergo a closer review. They also confirmed that the approach was promising, and that it could be refined and reinforced in order to go beyond its initial uses. This is what happened subsequently in the OTS and in the lab of Gilman Veith. Besides routine evaluation work in the structure-activity team, the "QSAR folks" of the OTS, as they were sometimes called, put forces into the development of quantitative SARs and in tools to develop them faster. Robert Lipnick, a chemist turned "QSAR scholar" (Lipnick, 1985, 1991) started using QSAR models as early as in 1981. With the support of the chief of the Environmental Effects Branch (EEB) of the OTS, he set out to compare the results of the screening of 55 alcohols with a known QSAR model for narcosis that was published in the literature. The model allowed generating a value for the level of narcosis that predictably occurs for certain chemical structures. Comparing the alcohols with these chemical structures, a possibility emerged to actually say whether these alcohols would themselves produce narcosis (Lipnick et al., 1985). Lipnick was known as a more theory-oriented person, very much interested in researching and validating models and applying them to new substances.

Two other scientists in the EEB worked to develop quantitative structure-activity models. Despite this actual lack of ecotoxicological (and chemical fate) test data, they worked continuously for several years to increasingly develop and to then make use of many individual SAR and QSAR estimates. This team had managed to develop 13 QSAR models by the early 1980s (see Clements et al., 1993). By 1988 there would be 49 different QSARs for estimating the aquatic toxicity or bioconcentration potential for about 30 classes or subclasses of industrial chemicals produced (Zeeman et al., 1993), the majority of which by EEB scientists. All of them were published in the so-called QSAR manual in 1988 and 1994 (over 120 QSARs by then). Structure activity work could be refined by intense collaborations between ORD and OTS. ORD's "fathead minnow studies" continued in parallel and eventually covered 617 industrial chemicals in total[15]. The programmatic offices extended funds for the ORD to develop its databases, with the ORD making these results accessible to their scientists so that they could identify structure-activity correlations and computer programs for each correlation, to generate predicted toxicity values. Thanks in part to this database, yet more quantitative models were produced[16], all of which fed the structure-activity meetings.

A more systematic use of quantitative SARs models implied some tension among scientists of the EEB. Some of them, such as Robert Lipnick, seemed not willing to accept the use of QSARs that had not been somewhat rigorously evaluated. Others, like Vincent Nabholz or Richard Clements, were more the reluctant sponsors of creating and using whatever QSARs they found or developed that were able to provide them with some of the numerical ecotoxicity answers that were needed for use under the circumstances of the OTS PMN review process. The EEB's leadership managed the opposition by functionally separating these scientists that were more oriented to the pragmatic use of QSARs in regulatory practice, from the more theoretical QSAR folks, who were actually getting in the way of efficiently performing the ongoing regular NCP chemical assessments via QSAR. The regulation-oriented scientists got major control of the QSAR hazard assessment process of new chemicals, while the more theoretical-inclined ones were allowed to invest in the publication and further development of models.

These tensions were not necessarily sensed outside the EEB. The new chemicals program was not subject to intense political and legal scrutiny. Auer recalls that in normal days, as director of OTS, he would only meet with the people in charge of PMNs three or four times a year, which denotes a low priority[17]. This relative protection of the space of EEB, coupled with the support from its hierarchy and the commitment of its scientists to improve structure-activity inferences, meant that progress towards increased quantification of SARs and consideration in regulatory evaluation of dossiers, did continue. Vincent Nabholz and Richard Clements made a continuous effort to revise EEB's structure-activity correlations, as valid new data were found. They assembled their models into the various QSAR Manuals (EPA, 1988; 1994b) and also created ECOSAR. The initial one was an internal notebook of 13 QSARs and it could be considered as the first QSAR Manual (EPA, 1988[18]). The second manual had 49 QSARs and it was published by OTS as an EPA document (EPA, 1988). The hope of the OTS, at this point in time, was to make the tool circulate. A revised version of the QSAR manual was completed in 1993, and this one constituted an important turning point. It was published by the toxics office as an EPA document (EPA, 1994b), and it stated that it was published to accompany the EPA release of the SAR software program called ECOSAR (see EPA, 1994a: 1). Until 1993, the QSAR manual had only been available in paper version. In 1994, the ECOSAR, which was a PC version of the manual, was then made available to the public. This revised version contained 42 chemical classes along with 120 QSARs.

In the 1990s, structure-activity work on hazard assessment still progressed in these circumstances, although the NCP resources declined[19]. After 1996, there was a hiatus in that EPA QSAR development. The EEB was dissolved in the 1997 reorganization of OPPT. The old EEB staff were then distributed amongst the various new Product Line Branches in the newly organized Risk Assessment Division[20], and thus many of the QSAR staff were now basically spread out and were basically on their own, with less support from management. The investment in SAR/QSAR continued, but those scientists that had been instrumental in its development no longer had a direct access to a management support system.

By that time, however, the New Chemicals Program, including in its use of QSAR, had been hailed a success by the chemical industry in the US. The review process had been formalized, with highly detailed manuals, guidelines and software tools, that helped understand how the process unfolded, and how each sort of information – chemical, toxicological or exposure-related – was used (EPA 1997). Many of the developments of the EPA/OTS of the preceding years were taken up in the OECD. When the OECD initiated the harmonization of SAR tools and models, it first undertook a validation exercise, to verify the accuracy of predictions that were made by the EPA staff. The conclusion of that exercise was that the SAR methods of the EPA "performed extremely well in predicting acute toxicity to fish and daphnia" (OECD 1994, cited in Zeeman, 1995: 712). In general, model-based predictions generated estimates of toxicity that were within an order of magnitude of those that were observed in animal experiments. The OECD validation of EPA's predictions was very much an endorsement of the work performed there since the 1970s to develop these tools, the "careful development and analysis of chemical categories", "the thoroughness and diligence in adding new data points to established categories", the high level of "refinement" of its predictive capabilities (OECD, 2007a: 28). Several important tools for SAR work, notably the concept of 'chemical category', were taken up by the OECD, recognizing the usefulness and applicability of what the EPA had developed internally (see EPA, 1993; cited and described in Zeeman, 1995, OECD, 2007; OECD, 2009).

Organizing the coproduction of science and law

The above history shows that the EPA did a lot around structure-activity in the area of chemicals hazard assessment, much more than any other organization, and earlier than anyone else too. It conceptualized the use of structure-activity correlations and judgments as a method to evaluate the hazardousness of chemicals, accumulated the experience in making such judgments, practicing them day by day on a very large number of products over a long period of time, integrating those judgments into a concrete decision-making machinery. It invested in a massive testing program to generate a database of experimental results, to perform the statistical analysis necessary to the production of more models. It formalized and put into circulation several important tools for other organizations to be able to use these estimations or develop new ones in turn. Altogether then, the EPA brought much credibility to the translation of an initial chemistry theory into the practical regulatory assessment of chemicals.

The striking aspect this history is that computerized models were not abruptly introduced as a regulatory tool, and certainly not approached as a tool for definitive decision-making through prediction. Rather, it is structure-activity thinking that was first introduced as an element of knowledge in the gradual formation of decisions, buying time for the more progressive development and subsequent consideration of validated statistical models in the regulatory process. The agency initially fell back on a more modest approach to knowing and predicting the future, embodied by qualitative, non-statistical structure-activity thinking. It practically restricted the validity of the approach of establishing relationships between structure and activity, to the screening of large sets of substances, and to the triaging of those that needed further investigation from those that could be deemed reasonably safe. Making precise, quantified predictions on any one of these substances was a horizon, and promise, that EEB scientists and regulators in the Office of Toxic Substances gradually came closer to[21].

Several elements concur in explaining why the EPA emerged as a site of formalization of a new kind of regulatory knowledge, and of invention of ways of applying structure-activity thinking. Based on the discussion of science-policy coproduction dynamics in the first part of this article, it appears that an evidential culture takes form, first, in response to the political stabilization of criteria of decision and proof. An evidential culture is not made of "research science", but of this particular kind of science that is believed to be appropriate to inform a decision criteria (a given definition of what counts as a risk) and a standard of proof set in the law. Those criteria are defined not by the regulatory organization itself, but emerge from the power relationships among the actors that take part in the construction of the regulatory framework, and in its implementation and subsequent evolution. From this perspective, an evidential culture changes where and when risk criteria and standards of proof evolve, under the pressure of principals (Congresspersons that design the Act that the agency must implement), courts (who review decisions and confirm or change the actual criteria of safety that the agency is supposed to apply) and/or of the regulated industry.

In that case, it seems clear that an evolution in the ways of establishing a proof of the existence of chemical risks was in order, given the particular regulatory design of TSCA. As the first leader of the structure-activity team recalled ex-post facto, "necessity is the mother of invention" (CHF, 2010: 4): the EPA had no alternative in the face of the double-bind in the implementation of TSCA — an ambitious mission to rapidly review a great number of different chemicals, with limited scientific and legal means to do so. Computerized QSAR models and expert systems simply filled the gap left after the Act was emptied of any requirement for testing (Mayo et al., 2012; Craeger 2018). The particularly large number of substances to review created a strong pressure to apply new methods, even though reliable QSAR models were not yet in sight. Structure-activity thinking thus became the immediate solution, even though it was a new, embodied expertise only practiced by a handful of specialists who the EPA had managed to recruit or attract. At the same time, the criterion that was chosen to define safety (an "unreasonable risk" of injury), coupled with a low evidentiary threshold (the Act authorizes the agency to act if there is a "reasonable basis" to conclude about the existence of an "unreasonable risk") meant that the agency could use emerging, judgmental (as opposed to formal and quantitative) methods in its work. No one among courts, environmental groups or regulated businesses, contested the interpretation that the agency made of this criterion. In other words, the agency was both constrained and given some autonomy to search for new methods to document the risk. The context gave weight to non-testing approaches emerging in the world of pharmaceuticals development.

Second, an evidential culture is built on a set of methods, and representations of what can be known, and what may not or is not interesting to know. In this sense, changes in evidential cultures depend on the production and availability of new research outside the organization, and on the capacity of the organization to translate and incorporate this research into its own expertise for its needs. What is noticeable in the present case is the network of relationships between the agency and external scientific groups, and the capacity of the EPA to attract experienced scientists from the world of computational drug design, and even more so, of qualitative structure-based toxicity prediction. Another noticeable aspect is that there were protected spaces in the organization, in which the necessary work to make modelling function could be envisioned and deployed. The group of QSAR folks inside OTS, alongside the ORD's ERL in Duluth and private scientific service companies, which the EPA intensely used, all dedicated to bettering the approach. Those different spaces were inter-linked by a network of QSAR people who consistently cooperated, sometimes in productive tensions, e.g. over the relative importance of the administrative imperative of availing of usable tools for deciding on substances, or screening them, and the more scientific, long-term ambition to produce reliable, fully validated statistical models. The OTS had among its staff the necessary level of scientific skills to manage these collaborations and profit from external productions. This means that the Office was, almost from the start of its existence, in a position to understand the challenges and difficulties of modelling, and aware of the technical developments to perform before embracing models fully.

One should add another factor, namely the autonomy of the new chemicals program. The very design of the NCP program in TSCA was a recipe for failure (thousands of substances to review, with little possibility for EPA to neither request data nor obligation to the industry to provide some), and the EPA leadership did not expect much from this program to start with. The New Chemicals program was also much less of a threat for companies than the Existing Chemicals program[22]. So, both the legal challenges and the political supervision from the higher echelons of the organization were limited, granting autonomy to the people inside the OTS to forge common rubrics of information and judgment, and ways of evolving decisions.

However, coproduction does not occur in the abstract, and structure-activity methods did not emerge spontaneously. The concrete form that modelling and prediction took in the agency was not quantitative modeling to start with, but a collective, human judgment about similarities between structures and of toxicity associated with structures. This collective judgement was formed during dedicated meetings, conceived of as a step in a regulatory sequence, organized around a dedicated team. This organizational materialization of structure-activity thinking can only be understood taking into account the autonomous bureaucratic knowledge present in the agency, which mediated both the legal requirements and scientific affordances.

Bureaucratic knowledge, in this case, covers several things: a form of procedural rationality, by which objective decisions are the product of a sequence of judgments formed on discrete bodies of information, applying different criteria (chemical analysis, then structure-based hazard assessment, followed by risk), some scientific, others more readily political. This, in essence, is the heart of the "risk analysis framework" that the agency and a suite of expert bodies started to formalize and apply in those years, faced as it was with massive controversies about the proper use of science, and suspicions about the distortion of evidence by political appointees (Demortain, 2019). There was, in those years, an important kind of bureaucratic experience of what were the correct ways of coordinating scientists and decision-makers in the organization, across the boundary that separates or should separate them (Bijker et al., 2009). EPA's specific bureaucratic knowledge also includes the notion that, in the presence of enduring uncertainties and disagreements — uncertainties that no single method or discipline could lift — regulatory decisions would only seem objective if they were based on a set of converging expert judgments, a style of decision formation that was captured in the Delphi method, that inspired the people of the OTS. It covers the infrastructural knowledge of individual chemicals, of their uniqueness and mutual resemblances, accumulated in people who make up this "molecular" bureaucracy (Hepler-Smith 2019). Bureaucratic knowledge is, finally, the experience of the credibility of this mode of making decisions in the interaction with audiences that evaluate the agency – scientists, courts and regulated businesses among others, as opposed to the making of decisions based on non-validated quantitative models. Each of these elements demonstrate the existence of an autonomous bureaucratic knowledge of the people of the Office of Toxic Substances, that decisively influenced the interpretation of the constraints posed by TSCA, and the capacities emerging from the scientific world, to give shape to, and anchor, an original and credible way of making structure-activity correlations.

Conclusion

In this paper, we have looked into a case of formalization of a kind of predictive knowledge in policy and regulatory practice. The main aspect of the history of the use of structure-activity thinking and predictions at the EPA is that it was a history of experimenting, ascertaining the value of this kind of knowledge for implementing a complex, challenging public program of reviewing hundreds of substances at once. Structure-activity predictions at the EPA was a developing discipline throughout the period, only getting to validated models after applying a qualitative form of structure-activity method, and investing in the gradual development of a database of good enough quality to produce reliable, usable quantitative models in a not too distant future.

Today, the development and use of models like structure-activity relationships in chemical regulation has become almost systematic. For many commentators, their adoption is the logical result of recent advances in biology and biotechnology, along with numerous controversies over the relevance of animal models (NRC, 2007). Such interpretations, however, fail to take into account the set of determinants that are necessary to make sticky epistemic and evidential cultures evolve in an organization faced with multiple constraints. It overlooks the particularity of this history: namely the fact that the EPA, as an organization that needs to forge credible demonstrations of the risk to convince audiences that are necessarily critical of its assessments and decisions, first opted more pragmatically for an evidential culture, in which prediction takes for the form of an analogical reasoning that helps define objects that can legitimately be subjected to further review and investigation. It approached the promises of predicting risks that underpin the development of quantitative, computational models of toxicity with great caution.

We have outlined a framework to analyze this particular case, expanding the coproductionist perspective to include an organizational component that seemed crucial to make sense of the particularity of this case: the fact that the change in the way of knowing and proving risks for regulatory decision, emerged from the capacity of the EPA to articulate a new method, adapted to its constraints, and in a sense to innovate. New methods and ways of proving the existence of a risk emerge as a new culture in a regulatory organization, if, first, a regulatory regime emerges in the environment that commands new practices and ways of making decisions. In the present case, the sub-regime of the new chemicals program, with its distinct and relatively weak criteria of decisions and demonstration, delimited a new space of regulatory work. Second, this culture will take form in the presence of a capacity to connect with the research environment to find methods that may respond to regulatory needs and uncertainties. Third, and perhaps most importantly, the methods that are chosen and the form they take – in this case, a qualitative form before a quantitative one – derives from the need to foster a mode of cognitive coordination in the agency. The knowledge that is incorporated in the agency must provide references of the risks and the uncertainties that various members of the decision process can share and build on, sequentially. In this case, we find no hierarchy of knowledge, or dominant standard of proof. No superiority is granted to modelling techniques or computational models, as a tool to handle uncertainties and compensate for the limits of animal experiments. We see an articulation of evidences, invented in situ.

The particular historical case of predictive, structure-activity knowledge at EPA, teaches us something more general about future-oriented expertise. Foreknowledge, or in this case scientific methods that claim to have a greater predictive power, is specific in this that its credibility is more difficult to establish. It is knowledge of uncertain, yet-to-come objects. Its credibility and value can only be established in the longer run, as and when these objects finally take form and make the demonstration of its relevance. As a form of knowledge that is both highly pertinent for policy-making, yet also less immediately authoritative than more realist knowledge, its production depends on the autonomy of that a given organization can find, to be able to try, verify and evaluate different ways of making decisions. The paradox of predictive organizations is that of being under a lot of constraints, not least to have to face many uncertainties, and of having the capacity to gain autonomy from these constrains, to design forms of knowledge and decision-making. Ways of predicting risks emerge from what science allows and what the law requests, decisively mediated by the knowledge of how to coordinate people and their knowledge to produce a decision, in an at least partly autonomous agency.

References

Adams V, Murphy M and Clarke AE (2009) Anticipation: Technoscience, life, affect, temporality. *Subjectivity* 28(1): 246–265.

Anonymous (1976) An Act To regulate commerce and protect human health and the environment by requiring testing and necessary use restrictions on certain chemical substances, and for other purposes. Oct. 11, 1976. Public Law 94-469, 94th Congress.

Anonymous (1984) Structure activity relationships: Are they predictive of toxicity and potential regulatory tools? *Environmental Science and Technology* 18(9): 285–286.

Auer CM, Nabholz JV and Baetcke KP (1990) Mode of action and the assessment of chemical hazards in the presence of limited data: use of structure-activity relationships (SAR) under TSCA, Section 5. *Environmental Health Perspectives* 87: 183–197.

Auer CM, Zeeman M, Nabholz JV and Clements RG (1994) SAR—the US regulatory perspective. *SAR and QSAR in Environmental Research* 2(1–2): 29–38.

Bijker W, Bal R and Hendriks R (2009) The paradox of scientific authority: The role of scientific advice in democracies. Cambridge and London: The MIT Press.

Bowker G and Star SL (1999) Sorting things out: classification and its consequences. Cambridge and London: MIT press.

Box GEP (1979) *Robustness in the Strategy of Scientific Model Building*. Wisconsin University Madison Mathematics Research Center.

Bradbury SP, Russom CL, Schmieder PK, Schultz TW, Diderich R and Auer CM (2015) Advancing Computational Toxicology in a Regulatory Setting: A Selected Review of the Accomplishments of Gilman D. Veith (1944–2013). *Applied In Vitro Toxicology* 1(1): 16–25.

Boullier H (2019) Toxiques légaux. Comment les firmes chimiques ont mis la main sur le contrôle de leurs produits. Paris : La Découverte.

Böschen S (2009) Hybrid regimes of knowledge? Challenges for constructing scientific evidence in the context of the GMO-debate. *Environmental Science and Pollution Research* 16(5): 508–520.

Böschen S (2013) Modes of Constructing Evidence: Sustainable Development as Social Experimentation—The Cases of Chemical Regulations and Climate Change Politics. *Nature and Culture* 8(1): 74–96.

CHF (2010) *The Toxic Substances Control Act from the perspective of Charles M. Auer*. Oral history program, Chemical Heritage Foundation.

Clements RG, Nabholz JV, Johnson DW and Zeeman M (1993) The Use and Application of QSARs in the Office of Toxic Substances for Ecological Hazard Assessment of New Chemicals. In: Landis W, Hughes J and Lewis M (eds) *Environmental Toxicology and Risk Assessment*. ASTM STP 1179, Philadelphia: American Society for Testing and Materials, pp. 56–64.

Collins HM (1998) The Meaning of Data: Open and Closed Evidential Cultures in the Search for Gravitational Waves. *American Journal of Sociology* 104(2): 293–338.

Craeger A (2018) *Testing Toxics: Tools, Standards, and Corporate Data in the History of Chemicals Regulation. Conference Pervasive Powers: Corporate Authority in the Shaping of Public Policy*. University Paris-Dauphine, Paris, 14-16 June 2018.

Crum-Brown AC and Fraser TR (1868) On the Connection between Chemical Constitution and Physiological Action; with special reference to the Physiological Action of the Salts of the Ammonium Bases derived from Strychnia, Brucia, Thebaia, Codeia, Morphia, and Nicotia. *Journal of Anatomy and Physiology* 2(2): 224–242.

Demortain D (2013) Regulatory Toxicology in Controversy. *Science, Technology & Human Values* 38(6): 727–748.

Demortain D (2017) Expertise, Regulatory Science and the Evaluation of Technology and Risk: Introduction to the Special Issue. *Minerva* 55(3):139–159.

Demortain D (2019) The Science of Bureaucracy. Risk Decision – Making and the US EPA. Cambridge and London: MIT Press.

Dowling D (1999) Experimenting on Theories. *Science in Context* 12(2): 261–273.

EPA (1975) Papers of a Seminar on Early Warning Systems for Toxic Substances. Conference Report. Office of Toxic Substances, US Environmental Protection Agency, Washington D.C. EPA-560/1-75-003. July 1975.

EPA (1981) *State-of-the-art report on structure activity methods development.* Prepared by Environmental Research Laboratory, for the Office of Toxic Substances, US Environmental Protection Agency, Washington D.C.

EPA (1984) *Proposed Design for a Retrospective Study of PMN Health Hazard Predictions.* Memorandum from John A. Moore, Assistant Administrator for the Office of Pesticides and Toxic Substances, to Terry Yosie, Director of the Science Advisory Board, US Environmental Protection Agency, Washington D.C.

EPA (1988) *Estimating Toxicity of Industrial Chemicals to Aquatic Organisms Using Structure-Activity Relationships, Volume 1*. Office of Toxic Substances, US Environmental Protection Agency, Washington, DC. EPA-560/6-88-001. July 1988.

EPA (1993) Categories Provide Guidelines for Premanufacture Notice Submitters. *OPPT's Chemicals in Progress Bulletin* 14(2): 35-36. US Environmental Protection Agency, Washington D.C. EPA-745-N-93-001. June 1993.

EPA (1994a) ECOSAR: *Computer Program for Estimating the Ecotoxicity of Industrial Chemical Based on Structure Activity Relationships, User's Guide*. Office of Pollution Prevention and Toxics, US Environmental Protection Agency, Washington D.C. EPA-748-R-93-002. February 1994.

EPA (1994b) *Estimating Toxicity of Industrial Chemicals to Aquatic Organisms Using Structure Activity Relationships 2nd Edition.* Office of Pollution Prevention and Toxics, US Environmental Protection Agency, Washington D.C. EPA-748-R-93_001. February 1994.

EPA (1997) *Chemistry Assistance Manual for Premanufacture Notification Submitters*. Office of Pollution Prevention and Toxics, EPA, Washington DC. EPA 744-R-97-003. March 1997.

EPA (2015) *Summary of Accomplishments in the New Chemicals Program*. Available at: http://www.epa.gov/oppt/newchems/pubs/accomplishments.htm (accessed 15.3.2019).

FRI (1975) Structure activity correlation bibliography, Prepared by the Franklin Institute Research Laboratories for the Office of Toxic Substances, PB-240-658, Philadephia, Pennsylvania.

FRI (1976) *Analysis and trial application of correlation methodologies for predicting toxicity of organic chemicals*. Technical report requested by the Office of Toxic Substances from the Franklin Institute Research Laboratories, EPA/560-1-76-006.

GAO (1984) A*ssessment of New Chemical Regulation Under the Toxic Substances Control Act*. General Accounting Office, Washington, DC (USA). Office of the Comptroller General.

Harris R and Milkis S (1989) *The politics of regulatory change: a tale of two agencies*. London and New York: Oxford University Press.

Hilgartner S (2000) *Science on stage: Expert advice as public drama*. Stanford: Stanford University Press.

Hepler-Smith E (2019) Molecular Bureaucracy: Toxicological Information and Environmental Protection. *Environmental History* 0(2019): 1-27.

Hunt J and Shackley S (1999) Reconceiving Science and Policy: Academic, Fiducial and Bureaucratic Knowledge. *Minerva* 37: 141–164.

Jasanoff S (2004) *States of Knowledge: The Co-Production of Science and Social Order*. London and New York: Routledge.

Knorr-Cetina KD (2011) Financial analysis: Epistemic profile of an evaluative science. In: Lamont M (ed) *Social knowledge in the making.* Chicago: University of Chicago Press, pp. 405-442. Lipnick RL (1985) A perspective on quantitative structure-activity relationships in ecotoxicology. *Environmental Toxicology and Chemistry* 4(3): 255–257.

Laurent B and Thoreau F (2019) Situated expert judgment: QSAR models and transparency in the European regulation of chemicals. *Science & Technology Studies* 32(4): 158-174.

Lipnick RL (1991) Outliers: their origin and use in the classification of molecular mechanisms of toxicity. *The Science of the Total Environment* 109–110: 131–153.

Lipnick RL (1998) *Correlative vs Mechanistic QSAR Models for Review of Industrial Chemicals Under TSCA*. Conference, Dayton, Ohio.

Lipnick, RL, Bickings CK, Johnson DE and Eastmond DA (1985) Comparison of QSAR predictions with fish toxicity screening data for 110 phenols. In Aquatic Toxicology and Hazard Assessment: Eighth Symposium. ASTM International.

Lipnick RL, Johnson DE, Gilford JH, Bickings CK and Newsome LD (1985) Comparison of fish toxicity screening data for 55 alcohols with the quantitative structure-activity relationship predictions of minimum toxicity for nonreactive nonelectrolyte organic compounds. *Environmental Toxicology and Chemistry* 4(3): 281–296.

Mayo-Bean K, Moran K, Meylan B et al. (2012) *Methodology Document for the ECOlogical Structure-Activity Relationship Model (ECOSAR*) Class Program: Estimating Toxicity of Industrial Chemicals to Aquatic Organisms using ECO- SAR Class Program (Ver. 1.1). Washington, DC: Office of Pollution Prevention and Toxics, Office of Chemical Safety and Pollution Prevention, U.S. Environmental Protection Agency.

Meyer H (1899) Zur Theorie der Alkoholnarkose - Erste Mittheilung. Welche Eigenschaft der Anästhetica bedingt ihre narkotische Wirkung? *Archiv für Experimentelle Pathologie und Pharmakologie* 42(2): 109–118.

Montgomery CM (2017) From standardization to adaptation: Clinical trials and the moral economy of anticipation. *Science as Culture* 26(2): 232-254.

Morgan MS and Morrison M (1999) *Models as mediators: Perspectives on natural and social science.* Cambridge: Cambridge University Press.

Nelson N, Geltzer A and Hilgartner S (2008) Introduction: the anticipatory state: making policy-relevant knowledge about the future. *Science and Public Policy* 35: 546–550.

NRC (2007) *Toxicity Testing in the 21st Century: A Vision and a Strategy*. Committee on Toxicity, National Research Council, Washington D.C.: National Academies Press.

NTP (2002) *Current Directions and Evolving Strategies*. Washington D.C.: National Toxicology Program, US Department of Health and Human Services.

OECD (1994) *US EPA/EC Joint Project on the Evaluation of (Quantitative) Structure Activity Relations (QSARS)*. OECD Environmental Monographs N°88.

OECD (2007a) Report on the regulatory uses and applications in OECD member countries of (quantitative) structure-activity relationship [(Q)SAR] models in the assessment of new and existing chemicals, ENV/JM/MONO(2006)25, Feb. 15, 2007.

OECD (2007b) *Guidance on Grouping of Chemicals.* OECD's EHS Series on Testing and Assessment No. 80. ENV/JM/MONO(2007)28; Sept. 26, 2007.

OECD (2009) *Guidance Document for Using the OECD (Q)SAR Application Toolbox to Develop Chemical Categories According to the OECD Guidance on Grouping of Chemicals*. OECD's EHS Series on Testing and Assessment No. 102. ENV/JM/MONO(2009)5; Feb. 24, 2009.

Oreskes N (1998) Evaluation (not validation) of quantitative models. *Environmental Health Perspectives* 106: 1453–1460.

OTA (1983) *The Information Content of Premanufacture Notices*. Background paper, Congress of the United States, Office of Technology Assessment.

Overton E (1899) Über die allgemeinen osmotischen Eigenschaften der Zelle, ihre vermutlichen Ursachen und ihre Bedeutung für die Physiologie. *Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich* 22: 189–209.

Powell MR (1999) *Science at EPA: Information in the Regulatory Process*. Washington D.C.: Resources for the Future.

Rajan KS (2006) *Biocapital: The constitution of postgenomic life*. Durham: Duke University Press.

Sarewitz D and Pielke RJr (1999) Prediction in science and policy. *Technology in Society* 21: 121–133.

Schultz TW (2014) Gilman D. Veith (1944–2013) – A dedication. *SAR and QSAR in Environmental Research* 25(4): 249–251.

Sismondo S (1999) Models, Simulations, and Their Objects. *Science in Context* 12(02): 247–260.

Veith G (1981) *State of the art of structure-activity methods development*. Environmental Research Laboratory, Office of Research and Development, US EPA.

Vogel SA and Roberts JA (2011) Why the toxic substances control act needs an overhaul, and how to strengthen oversight of chemicals in the interim. *Health affairs* 30(5): 898–905.

Wambaugh J (2014) "Building fit-for-purpose pharmacokinetic models", In Vitro to In Vivo Extrapolation for High Throughput Prioritization and Decision Making Webinar Series.

Waterton C and Wynne B (2004). Knowledge and political order in the European Environment Agency. In: Jasanoff S (ed) *States of Knowledge: The Co-Production of Science and Social Order*. London and New York: Routledge, pp. 87-108.

Zeeman M (1995) Ecotoxicity Testing and Estimation Methods Developed Under Section 5 of the Toxic Substances Control Act (TSCA). In: Rand GM (ed) *Fundamentals of Aquatic Toxicology: Effects, Environmental Fate And Risk Assessment*. Boca Raton: CRC Press, pp. 703–715.

Zeeman M, Auer CM, Clements RG, Nabholz JV and Boethling RS (1995) US EPA regulatory perspectives on the use of QSAR for new and existing chemical evaluations. *SAR and QSAR in Environmental Research* 3(3): 179–201.

Zeeman M and Gilford J (1993) Ecological Hazard Evaluation and Risk Assessment under EPA's Toxic Substances Control Act (TSCA): An Introduction. In: Landis W, Hughes J, Lewis M (eds) *Environmental Toxicology and Risk Assessment – 1st Volume*. ASTM STP 1179. Philadelphia: American Society for Testing and Materials, pp. 7–21.

Zeeman M and Mayo-Bean K (2009) *Development and Use of (Q)SAR Tools by U.S.* EPA/OPPT to Assess Industrial Chemicals, Invited talk, Annual SETAC congress, New Orleans.

Zeeman M, Nabholz JV and Clements RG (1993) The Development of SAR/QSAR for Use Under EPA's Toxic Substances Control Act (TSCA): An Introduction. In: Gorsuch J, Dwyer J, Ingersoll C and La Point TW (eds) *Environmental Toxicology and Risk Assessment - 2nd Volume*. ASTM STP 1216. Philadelphia: American Society for Testing and Materials, pp. 523–539.

Notes

1 Disclaimer: the views expressed by Dr. Zeeman are those of the author and this article has not been reviewed by the U.S. EPA and it should not be considered to reflect the views or policy of the U.S. EPA.

2 This article is based on empirical research performed in the framework of the project INNOX - Innovation in Expertise. Modeling and simulation as Tools of Governance, supported by the French Agence Nationale de la Recherche (Grant n°ANR-13-SOIN-0005), between 2014 and 2017, and coordinated by David Demortain (innox.fr).

3 For the last 40 years, the 'gold standards' of toxicological testing for humans have been the 13-week and the 2-year rodent bioassays along with other whole-animal studies (NTP, 2002).

4 The main use of QSAR modeling is performed within the pharmaceutical industry, as part of the early screening of candidate molecules or leads for toxicity problems. This aspect is outside the scope of this paper.

5 The level of certainty that is to be reached before one goes out making a claim or publishing something presented as true (see Collins, 1998).

6 As an example, we can cite an early EPA publication that introduced a QSAR model for estimating the LC50 for industrial alcohols, ethers, alkyl halides and benzene derivatives (EPA, 1981). At the heart of the model, one finds the following equation: Log 1= 1.17 + 0.94 log P, where log P is the logarithm of the n-octanol/water partition coefficient, the structure-related property of the chemical with which the toxicity in question (LC50, the concentrations of the chemical in air or water that kills 50% of the test animals during the observation period) is correlated. Running such a formula produces a numerical estimation of toxicity. Structure-activity theory allows inferring that a chemical that is the member of/ included in the class of substances for which the formula has been developed (similarity according to the chosen descriptor for constituting the class; in the above case log P or the n-octanol/water partition coefficient), is likely to have a similar level of toxicity.

7 His equations summarizing structure-activity relationships are frequently nicknamed "Hansch equations". Hansch and other chemists have developed these equations and pushed for their use in the pharmaceutical industry throughout the 1970s and 1980s, leading to a situation where pharmaceutical companies R&D departments now routinely have "computational drug design" units that collaborate with other groups such as toxicologists, who perform the initial tests of the toxicity and effects of a substance on animals.

8 The EPA is organized in offices. These offices are headed by an "assistant administrator" who, like the EPA Administrator, is a presidential appointee. There are two types of offices: programmatic offices, which are created to implement a particular act (on toxic chemicals; on air quality; on water quality; on pesticides; and so on); and non-programmatic offices or services, such as the office of the general counsel or the Office for Research and Development, which is basically the scientific arm of the agency. ORD counts several dozen laboratories and more than 500 staff. It develops science for programmatic offices, but also follows its own internally defined research programs. Coordination between program offices and the ORD, or the responsiveness of the latter to the needs of program offices, is a recurrent issue in the history of EPA (Powell, 1999).

9 Veith had a PhD in water chemistry from the University of Wisconsin. He joined the Environmental Research Laboratory of the EPA in 1972, developing work on bioaccumulation of chemicals and pesticides in the environment. Veith did show some understanding of the specific goals and constraints of the regulatory work, specifically the needs and challenges of implementing the new chemicals program for the OTS, and of the need for applicability of methods in regulatory evaluation of products (as opposed to ever more refined and sophisticated methods and results) (Schultz, 2014).

10 As of September 2010, the OTS had received a total of 50,449 submissions, more than the total number of substances included in the EU REACH program (EPA, 2015).

11 This means that they could in effect legally be manufactured. However, there was always a significant proportion of new chemicals that made it through the entire NCP PMN process, but that for a variety of reasons were apparently never actually manufactured (i.e., no notice of commencement of manufacture was received by the EPA, and thus they were not put on the TSCA inventory as an existing chemical in commerce).

12 OTS, HERD and EEB no longer exist. OTS was renamed the Office of Pollution Prevention and Toxics (OPPT) in around 1992. The major reorganization of that office in 1997 resulted in the morphing of HERD into what became the Risk Assessment Division (RAD).

13 Source: interview with the authors.

14 Source: interview with the authors.

15 On this, see Bradbury et al. 2015: "its usefulness in QSAR modelling can mainly be credited to the strategic approach taken in the development of the database. The express purpose of the fathead minnow database was to build relevant and reliable QSAR models based on data that covered a wide range of structure space and thereby a wide range of possible modes of toxicity. All toxicity tests were conducted in the same laboratory following standard test methods. Both the dilution water and fish used were from a single source. Chemicals used were of the highest purity, with all treatment concentrations measured under stringent data quality objectives. By controlling for these factors, variability in the test results was minimized and thereby increased confidence that variation in toxicity was related to variation in chemical structure and associated toxicological properties." (Bradbury et al., 2015: 19)

16 There is a list of the 49 SARs in the OTS QSAR Manual (EPA, 1988: ix). Thirty-one (31) of them cite their "Source" as being developed by EEB scientists (Vincent Nabholz and/or Richard Clements, etc.), and four of them list their "Source" as publications of the laboratory of Gil Veith in ORD.

17 Source: interview with the authors.

18 Two manuals developed in 1984 and 1996 were never published.

19 There was a decrease of almost 40% in NCP funding and a decrease of about 33% in NCP staffing between 1990 and 1995, even though the number of PMNs received seemed to trend upward.

20 Previously called HERD

21 When one launches the ECOSAR program on a computer, a special warning appears in a window saying "it is a screening-level tool", and that "Estimated values should not be used when experimental (measured) values are available".

22 The New Chemicals program eventually managed to process dozens of thousands of PMN, but the most profitable chemicals were the existing ones, for which TSCA is often analyzed as a failed statute in terms of decisions actually made on controlling existing toxic chemicals (Vogel and Roberts, 2011; Boullier, 2019).

Situated Expert Judgment: QSAR Models and Transparency in the European Regulation of Chemicals

Brice Laurent

MINES ParisTech, PSL Research University, CSI – Centre de Sociologie de l'Innovation, i3 UMR CNRS, France/ brice.laurent@mines-paristech.fr

François Thoreau

Spiral research centre, University of Liege, Belgium

Abstract

This paper discusses the kind of expert judgement demanded by the development of a particular class of models. It analyses the case of 'Quantitative Structure-Activities Relationship' (QSAR) models, used to predict the toxicity of chemical substances, for regulatory and other purposes. We analyse the production of these models, and attempts at standardizing them. We show that neither a technical nor a procedural standardization is possible. As a consequence, QSAR models cannot ground a production of knowledge along the lines of 'mechanical objectivity' or 'regulatory objectivity'. Instead, QSAR models imply that expert judgement is situated, re-worked for each new case, and implies an active intervention of the individual expert. This has important consequences for risk governance based on models. It makes transparency a central concern. It also means that new asymmetries emerge, between companies developing sophisticated models and individual experts in regulatory agencies in charge of assessing these models.

Keywords: expertise, models, objectivity, regulation, transparency

Introduction

Computer simulation and computer modelling are being used to govern a growing share of social activities. A recent evolution has made computer models a tool for evaluating and controlling the health and environmental risks raised by chemicals. Using statistical correlation, models would predict which chemicals are problematic, and complement other risk assessment methods such as *in vitro* or *in vivo* tests. In situations where scientific uncertainty is present, models would provide additional scientific elements to ensure that regulatory decisions are appropriate.

Described as such, it would be tempting to see models as ready-made scientific tools expected to provide objective descriptions of technical entities, for later use in regulatory settings. But what 'objective' means in this context is not self-evident. Works in Science and Technology Studies

(STS) have shown that objectivity is manufactured in various ways, which differs across historical and regulatory contexts, and which directly impacts how expert judgment is conducted (Cambrosio et al., 2006; Cambrosio and Keating, 2009; Daston and Galison, 2007; Jasanoff, 2011). One of the important insights of STS works on objectivity is that the production of objective knowledge implies that the human subjects expected to produce or witness objective knowledge are shaped in particular ways. In regulatory settings, this means that the production of objective knowledge also defines the type of expert judgment at stake.

We follow this inspiration in this paper. We examine the use of models for regulatory purposes by analysing the expert judgment that it entails. We focus on models known as 'Quantitative Structure-Activities Relationship' (QSAR), designed to predict the toxicity of chemical substances. These models are based on statistical correlations between a set of physicochemical descriptors that characterize a substance (e.g. chemical composition, morphology, …) and its biological activity, including its potential toxicity. In other words, QSAR models are based on the hypothesis that relevant knowledge regarding the toxicity of a chemical can be inferred from its structure. Diverse actors are developing QSAR models and produce a multiplicity of different QSAR models for different purposes (Lo Piparo and Worth, 2010). They thus embody the diversity and complexity of foreknowledge used in policy. Like other models in various technical areas, QSAR models are used by policy-makers to inform regulatory decisions. And like other models, they raise a series of uncertainties that have political consequences (see e.g. Edwards, 1999, 2010 about climate modelling).

Our objective in this paper is to analyse the political issues raised by QSAR models, particularly focusing on the ways by which they challenge the practice of public expertise. We argue that QSAR models are empirical entry points to reflect on risk governance based on models, and in particular the type of expert judgment that this approach entails. We demonstrate that the expert judgment that these models require cannot be tied to the use of ready-made technical tools providing stable scientific evaluations (as in situations of 'mechanical objectivity', see Daston and Galison, 2007), nor to procedures and standards framing the appropriate mode of action (as in situations of 'regulatory objectivity', see Cambrosio and Keating, 2009). Instead, QSAR models imply that expert judgement is situated, re-worked for each new case, and implies an active intervention of the individual expert. This has important consequences for risk governance based on models. It makes transparency a central concern. It also means that new asymmetries emerge, between companies developing sophisticated models and individual experts in regulatory agencies in charge of assessing these models.

Echoing Boullier, Demortain and Zeeman who look at the beginnings of QSAR modelling in chemicals regulation at the US Environmental Protection Agency (Boullier et al., 2019), our focus is on European institutions, and how they use or plan to use QSAR models for the regulation of chemicals, possibly by using international standards. Chemicals are regulated in Europe within the REACH regulation (Registration, Evaluation, Authorization and Restriction of Chemicals; see European Commission, 2006). Within this framework, companies have to demonstrate to public expert bodies that they are able to evaluate and manage the risks of the substances they produce. This requirement results in a large number of toxicological tests. The use of computational models could appear as a means to mitigate this trend.

Following an approach undertaken by scholars who have examined the use of models in policy arenas (Edwards, 1999; Heaphy, 2015; Fisher et al., 2010), we examine the making of QSAR models and the debates about them in European institutions, as well as in the Organisation for Economic Cooperation and Development (OECD) where such discussions are held and European actors are involved. We base our reflection on three sets of empirical material. First, we use observations from a research project that developed QSAR models for nanomaterials. This research project involved material scientists and toxicologists, and was conceived as a demonstration of the interest of the QSAR approach for regulatory purposes. We were involved in the project for a year in 2014-2015[1]. We observed research meetings and conducted interviews with the leaders of the toxicology and materials science teams, as well as with the post-doc researcher and the engineer involved. Second, we examine standardization attempts at the OECD, which were expected to help public bodies evaluate the use of QSAR models by private companies. We use the OECD literature on the topic, as well as two interviews with participants in the OECD working groups. Third, we build on five interviews with experts working in public organisations in charge of evaluating what private companies submit to register the chemicals they produce within the REACH framework. We use this qualitative empirical material to infer how QSAR models are expected to function with risk governance frameworks. Taking inspiration from Science and Technology Studies (Jasanoff, 2004), we discuss the type of technical knowledge that QSAR models are expected to provide, and the expert judgment that this knowledge requires.

The progression of our argument mirrors the list of our empirical sites. First, we situate our approach in a more general debate about expert judgement. We then examine the practices of QSAR model-making and elaborate further about QSAR models as outcomes of trial-and-error processes, unfit for reaching definitive closure, as indicate the efforts coordinated by the OECD in order to standardize processes of validation. We then analyse the "QSAR toolbox" developed by the OECD. The toolbox provides an evolving and flexible tool amenable to public experts uses and appropriations. Lastly, we finish by analysing the consequences of the previous considerations for the works of experts working in public agencies in charge of evaluating the use of QSAR models. We show that QSAR models require expert judgment to be defined in situated ways, and that this situatedness makes transparency a key component of the risk governance framework, in turn producing new asymmetries between public bodies and private companies.

Expert judgment and the problem of QSAR models

Risk governance relies on the ability of public institutions to mobilize technical expertise for decision-making. Expertise has been famously problematized by Sheila Jasanoff (2005) as a 'three-body problem', in that its legitimacy is the outcome of a subtle articulation between a public body organized to deliver expertise, a body of knowledge stable enough to provide grounded facts, and the body of the expert as an individual expected to provide consequential advice (Jasanoff, 2005). This perspective shows that the form of this articulation may vary. Throughout Jasanoff's works, the American case appears as a particularly interesting illustration of the importance of the 'view from nowhere' in defining expert legitimacy. The 'view from nowhere' points to the set of mechanisms whereby expert advice is disconnected from the particularities of its conditions of production, whether related to situated technical choices or to the individualities of the experts themselves.

Problematizing expertise as the outcome of a view from nowhere has consequences for both the organization of public institutions and the type of expert judgment. First, it implies that risk assessment (as an outcome of expert judgment) is carefully separated from risk management (where decisions can be related to particular decision-makers and political stakes). Second, experts ground the legitimacy of their interventions on their ability to ensure a form of 'mechanical objectivity' (Daston and Galison, 2007) whereby instruments can stabilize descriptions of the technical world purified from human intervention. Despite this importance in the organization of American expert bodies, this configuration is only painfully and temporarily stabilized, as regulators themselves acknowledge the inter-relatedness of risk assessment and risk management, experts' political motivations are questioned, and the very ability to operate the view from nowhere in practice is questioned (Hilgartner, 2002; Jasanoff, 1990). When expert judgment is framed as the outcome of the view from nowhere, experts are expected to disappear behind the instruments they mobilize. Mechanical objectivity relies on instruments that can travel in a stable way, and ensure robust fact-making because of their stability. As such, they are black-boxes in the Latourian sense (Latour, 1987). This does not mean that experts as human beings are no longer individuals on their own, but that public institutions define the legitimacy of their interventions in their ability to make their selves independent from the production of facts.

The debates about European expertise can be read as a variation on the three-body problem of expert legitimacy. They display a pervasive tension between attempts at reproducing the 'view from nowhere' and the political negotiations at the heart of the European regulation of technical objects. Thus, when European institutions responded to food crisis by the creation of a centralized expert body (the European Food Safety Authority) expected to operate independently from political pressure (Demortain, 2009), they were caught in pervasive tensions about whether or not the experts of the agency were actually free from private interests (Vos, 2000), and about whether the agency could provide technical advice expected to ground decisions for all member states (Wickson and Wynne, 2012). The difficulties that an expertise body such as EFSA has encountered can be interpreted as outcomes of a pervasive tension within the European expertise institutions. In the European context, manufacturing expert judgment is directly connected with the negotiations between member states and stakeholders (see: Saurugger, 2002). This makes the call to reproduce the 'view from nowhere' highly problematic, since this configuration might neglect the specificities of the European political landscape and modes of negotiation.

The case of chemicals however seems to provide an illustration of a successful stabilization of European expertise. Within the REACH regulation, the European regulation of chemicals is based on the coordinated action of the European Chemicals Agency (ECHA) and national expert bodies, the former acts as a centralizing body able to leave room for national variations (Boullier, 2016). Techniques through which assessments can be conducted in uncontroversial ways are therefore even more important. Some of them are procedural (as registration dossiers codified in European regulations), while others are based on standardization. Examples of the latter include the methods described in the technical guides published by the ECHA to help operationalize REACH, and standardized testing methodologies produced by the OECD, intended to be technical tools neatly distinguished from the regulatory choices that sovereign members of the international organizations might make (Salzman, 2005: 203).

For all their diversity, these tools have a similar role in the REACH risk governance framework, namely to provide the European experts with stabilized tools able to ensure the technical validity of risk assessment as they examine registration dossiers for chemicals. They serve as instruments through which experts working at ECHA can evaluate the dossiers submitted by private companies as they ask to register the substances they produce. These tools are expected to ensure that the outcome of expert advice only depends on the instruments being used and not on the individuality of the expert conducting the evaluation. Eventually, they allow these experts to separate the technical phase of risk assessment from the political phase of risk management.

QSAR models have been promoted by regulatory agencies for over twenty years, but have recently gained momentum in Europe, in the wake of the REACH regulation. As the regulation on chemicals is becoming more constraining for private companies, usual experimental approaches raise many concerns. Testing methods are lengthy, costly and often require animal testing. REACH is gradually extended to larger families of materials, which implies that even more tests need to be performed to ensure that chemicals can circulate on the European markets. In this context, QSAR models could appear as an alternative. They could provide knowledge about the potential risks of a given substance without conducting any test. In practice, this means that a company wishing to register a new substance could argue, based on models, that this very substance has a risk profile similar to other substances already registered.

Models could provide an additional resource for European expertise to ensure its technical validity, and its ability to be distinguished from regulatory decisions. However, the ECHA experts do not present QSAR models as technical black boxes that could ground a mechanical objectivity expected to make subjective interventions disappear. Consider the ways in which ECHA presents QSAR models in one of its 'guidance documents':

> The process of (Q)SAR acceptance under REACH will involve initial acceptance by industry and subsequent evaluation by the authorities, on a case-by-case basis. It is not foreseen that there will be a formal adoption process, in the same way that test methods are currently adopted in the EU and OECD. In other words, it is not foreseen that there will be an official, legally binding list of (Q)SAR methods. (ECHA, 2008: 27)

The contrast with the OECD tests is interesting since the latter are a good illustration of internationally agreed-upon methods that can act as resources for the technical validity of the assessment. By contrast, in regulatory decision making, QSAR models cannot be seen as ready-made instruments with unambiguous consensus on their scientific validity.

The ECHA (2008: 26-27) document quoted above states that the "use of (Q)SAR predictions in an automatic way" is "not recommended". Instead, it asks experts to consider "validation results, regulatory purpose and use of weight of evidence" (ECHA, 2008: 26-27). Rather than offering ready-made instruments providing scientific evidence for the technical phase of risk governance, independently from particular regulatory choices, QSAR models seem to be far from universal acceptance, and to be effectively tied to particular regulatory considerations. The specialists of QSAR modelling whom we met concurred. Many of them saw QSAR models as tools for conducting a preliminary selection of potentially problematic substances (or "screening," as they would say), while being extremely wary of a potential use that would go beyond providing additional evidence to that produced through standardized testing.

Thus, when discussing the use of QSAR models within the ECHA, specialists of the methods explain that:

> Under the coordination of the Chemicals Agency, the regulatory bodies in the EU will then make case-by-case decisions on the acceptability of any (Q)SAR models and estimates used, taking into account the regulatory context and the availability of other information. (Worth et al., 2007: 116)

Such wording seems to imply that QSAR methods are not expected to become black-boxes ready to be used as proof- making devices, but are tied to local conditions of use. This, we contend, prevents expert judgment from relying on a form of mechanical objectivity, and entails new political challenges, in terms of the identity of the actors involved in risk governance, the possibility of publicly controlling them, and the nature of public proof. To understand these challenges, we need to demonstrate that the impossibility to black-box QSAR models is not a mere incidental and preliminary situation before their eventual stabilization, but part of their very nature, and of what makes them of interest to industrial producers and public experts in the first place. To do so, we need to delve into the mechanisms of model-making. This requires that we temporarily leave the world of experts working in public agencies such as ECHA and follow other actors, namely specialists in materials science, toxicology and computer science as they attempt to craft QSAR models.

Unstable categories, unstable models

QSAR models are based on statistical correlations between a set of physicochemical descriptors which characterize a substance (e.g. chemical composition, morphology…) and its biological activity, which includes its potential toxicity. In other words, QSAR models are based on the hypothesis that relevant knowledge regarding the toxicity of a chemical can be inferred from its very structure. QSAR models are developed using a limited number of substances that serve as reference points, so that the properties of other chemicals could later be predicted by the model, according to their proximities to the reference points.

One of the main interests of QSAR models for regulatory purposes lies in their ability to re-group chemicals across existing categories and according to similar structure-activity profile. Instead of the existing classifications (such as those based on substances' atomic compositions), substances would be grouped according to their hazard profile. A telling illustration of this point is the case of nanomaterials.

Attempts at regulating nanomaterials within the European institutions have been caught in a tension between two opposite approaches (Laurent, 2017). On the one hand, the European Commission argues for a case-by-case approach to deal with nanomaterials. In this approach, nanomaterials could be gathered in broad categories (carbon nanotubes, titanium dioxide, etc.) each of them further broken down into smaller ones (e.g. single-walled and double-walled nanotubes, rigid single-walled and flexible sing-walled nanotubes…). On the other hand, other regulatory actors criticize this approach for failing to stabilize categories necessary for constraining legal interventions, such as labelling or control. The European Parliament added an amendment to the 2011 cosmetic regulation which introduced mandatory labelling of cosmetics containing nanomaterials. In 2012, France became the first country to introduce a mandatory declaration of nanomaterials. These initiatives require that new definitions be introduced in regulatory texts. In these regulatory texts, nanomaterials were defined using a size limit (set between 1 and 100nm), which could only partly account for the possibility of additional hazard. The antagonism between the two approaches can be summed up as follows: while the former tends to propose an endless subdivision of ever more refined categories (at the price of the postponement of regulatory decision), the latter is based on the construction of general categories, technically imperfect, and possibly arbitrary.

QSAR models can be seen as a way of escaping this quandary. Scientists propose to use QSAR model for nanomaterials, so as to group them in relation to the similarity of different substances' risk profiles. By defining "profiles" of risk more precisely, it would become possible to generate new categories. One could group together substances based on physical or chemical descriptors (e.g. their shapes), and associated expected properties (including those linked with toxicity). Accordingly, QSAR methods would provide a tool to group chemicals according to common characteristics that would generate similar physicochemical properties – including those linked to potential hazards, i.e. the properties that are particularly interesting from a regulatory viewpoint. As such, these methods offer ways of grouping chemicals without either constantly separating them in new categories or creating general and arbitrary criteria.

How is it then possible to group chemicals according to common characteristics correlated with similar properties, including above all toxicological properties? The process we observed when studying scientists developing QSAR for nanomaterials comprised:

- the choice of a set of reference substances (in the project we observed, as many as 45 different nano-substances, belonging to different chemical families such as Zinc oxides, Nickel oxides, or Boehmite);
- the definition of a list of "descriptors" such as the morphology (shape) or the size of chosen compounds, whether they come in filaments, aggregates, etc.;
- the definition of a list of "endpoints" linked to experimental test data on cell cultures in the laboratory, mostly so as to predict rates of reproduction or cell defects;
- the production of statistical correlations between descriptors and endpoints, which led to the refining of both lists.

New groups of chemicals could then be constituted according to their similarities in terms of their structures (descriptors) and correlated activity (endpoints).

The challenge, here, is to avoid two opposite problems. The first one is called *over-fitting* by QSAR specialists. It means that the model is so tailored to the substances being used to construct it that it is unable to provide any significant information about any other substance. In a case of over-fitting, any substance that is different from those used to produce the statistical correlation would be too different for the model to perform. In order to avoid over-fitting, QSAR specialists need to build statistical correlations that are *not too accurate*, in order for the model to be usable for new entry data. Over-fitting requires that one use a limited number of descriptors so that other chemicals can fit within the model. Yet this raises a second problem, namely that of using too few descriptors for the model to build significant statistical correlation, i.e. *under-fitting*. For a correlation to arise, one needs a minimal number of descriptors, various enough for statistical relationships to emerge.

Avoiding the problem of over-fitting and that of under-fitting requires that QSAR practitioners proceed with caution. The following discussion (between A, B and C, three members of the research project we observed) is about whether or not to quantify the shape of the substances being used to build the model, and then about what criteria to select in order to differentiate among substances:

A. Descriptors are not all quantitative… how will we do for the shape of substances?
B. So far, what I've done is that I have typed the number for each dimension. So if I see "first dimension equals 6"; "second dimension equals 6"; "third dimension equals 300", I know that it's a little stick, shaped as a cylinder. (…) Because all our particles have cylindrical symmetry.
A. But you could also do, "if it's a sphere then 1", "if it's a cylinder 2", "3 is a lump", etc.
B. Right, I could separate among all those… Well, what we need to differentiate is among those that are agglomerated or not. (…) There are three or four shapes that we feel like separating, when looking at the pictures.
C. We could differentiate among 4 types: isotropic isolated nanoparticles, isolated sticks, isolated bars, and formed aggregates. (…)
A. Then there is an ambiguity with boehmite, because boehmite is really bars. But we see sticks, because the bars are superposing themselves – like tiles. Somehow it's bars and sticks in the same time.
B. Yeah right, you could do both… but then the question is "what does the cell see?". And for me, the cell sees sticks. (…) We just take the situation according to the cellular cell, and then it's not bars. I agree that for a chemist, it's bars.
C. What the chemist sees, and what the biologist sees…
B. But there's no truth in itself here, we choose descriptors from the viewpoint of the cell…
C. That's why when you look at the OECD descriptors, some of them are from the viewpoint of the environment, or from the viewpoint of the river.

This somewhat long dialogue offers a window into the practical process through which developers of QSAR models choose descriptors. Here, the descriptors being discussed are related to the "shape" of the substances, and what various shapes scientists "feel like separating" from one another, so that a substance on which the model will be used will be described as "particles", "bars", or "aggregates"… Then the question relates to the number and type of these descriptors of shape.

Two remarks follow from there. First, we can see in this exchange that isolating descriptors is a process based on a variety of inputs, including references to guidelines produced by international organizations (here, the OECD), considerations about what will make a difference in toxicological effects, and expectations about the potential effects on potential endpoints. Second, the choice of descriptors is tightly connected to the choice of endpoints. The later part of the dialogue above is about the "viewpoint of the cell", "the environment" or "the river". If the endpoint is cell toxicity (as it is in the previous excerpt), then the descriptor has to be chosen "from the viewpoint of the cell". If the endpoint is aquatic toxicity, then the viewpoint will be that of the river. Accordingly, the choice of appropriate descriptors is tightly connected to the potential endpoints one needs the model to provide, themselves directly related to regulatory constraints (are the required tests related to cell toxicity? Or to environmental toxicity in aquatic environment?).

Therefore, the list of descriptors might significantly vary among QSAR models. In this respect, there is a fundamental uncertainty about the appropriate choice of descriptors, and, consequently, about the categories emerging from the grouping of substances according to descriptors. There is no such thing as "the best" category, but rather a trade-off between different descriptors and the importance granted to various criteria. Getting back to the dialogue above, the project might lead to group substances according to their shapes as "bars" or "sticks", yet will only do so in the context of an inquiry on cell toxicity.

This snapshot is only a glimpse into how QSAR models are produced in practice. One could provide other examples, related not to the choice of descriptors, but also to that of endpoints, or that of reference substances themselves. Eventually, the calculation of statistical correlations between descriptors and endpoints is itself an iterative process. The person in charge of calculating the statistical correlation between the descriptors and endpoints in the research project that we observed explained during an interview that the process of building statistical correlation (that is, the model itself) was characterized by "trials and errors" (she used this expression). If she observed "no answer" from a series of descriptors, that is, that they did not impact the value of the endpoints in statistically significant ways, then she would deduce that they were not relevant. She would eliminate them, thereby reducing an initial long list to just a few parameters.

These considerations show that the practices of QSAR modelling are not stabilized, but partly re-invented for each dataset of chemicals used to build models. For QSAR practitioners, the objective is to build models *accurate enough*. To do so, these practitioners process by trial and error, concerning the list of descriptors, the list of endpoints, and the calculation of statistical correlations. Thus, in QSAR modelling, accuracy is negotiated. As sociologists of science and technology have demonstrated, constructing accuracy is part and parcel of the making of technological systems, and impacts on / is impacted by the larger choices about their objectives and modes of functioning (MacKenzie, 1993). In this particular case, accuracy is negotiated in a way that never aims to construct the model as a settled entity. Models need to be accurate, yet not too accurate.

This characteristic might result from a more general feature of models based on the identification of statistical correlations, as opposed to models based on the application of general laws of physics or chemistry. Yet in the case of QSAR, they point to particular regulatory issues. This helps to explain the connection between the use of QSAR and considerations related to the 'regulatory context' that was drawn by European actors commenting on the use of this method. Constituting groups of chemicals with similar risk profiles depends on the choice of descriptors and endpoints, the latter being directly tied to regulatory priorities (e.g. aquatic toxicity for certain animal species). Eventually, various choices of descriptors and endpoints might lead to the crafting of various groups of chemicals, each of them tied to certain models. The possibility of re-defining the perimeters of the categories that bring chemicals together is precisely what makes QSAR models interesting in cases such as nanomaterials where substances are not covered by existing regulatory categories. But this also means that QSAR models and the group of chemicals on which they are expected to be applied are constituted in the same movement, and that, consequently, the former cannot easily be disentangled from the latter.

A procedural standardization?

The standardization of models expected to be used for regulatory purposes is a daunting task. Standards for experimental test methods are developed at the Organisation for Economic Cooperation and Development (OECD) and used in the European regulatory bodies. But QSAR models raise practical difficulties for standardization. How, for example, to define in advance the list of descriptors and endpoints without compromising the trial-and-error process that is at the heart of the construction of QSAR models? We begin here to understand the difficulty with which we started our exploration of QSAR models in European regulatory bodies. If the European Chemical Agency does not envision "a formal adoption process, in the same way that test methods are currently adopted in the EU and OECD" (ECHA, 2008: 27, see above), it might well be because of the situatedness of the elaboration of QSAR models.

Yet the regulation of technological innovation provides numerous examples of standardization and/or regulatory interventions that are designed for their ability to cope with the local adaptation of technical tools. Commenting on such processes, Cambrosio and Keating (2009) speak of 'regulatory objectivity'. By contrast with 'mechanical objectivity' (Daston and Galison, 2007), based on stable technical instruments, 'regulatory objectivity' refers to situations within which public and private institutions need to agree on procedures according to which various regulatory entities can be crafted. Regulatory objectivity "consistently results in the production of conventions, sometimes tacit and unintentional but most often arrived at through concerted programs of collective action" (Cambrosio et al., 2006: 190). Describing various standardization and/or regulatory interventions related to biomedicine, Cambrosio and Keating analyse the ways in which public and private actors coordinate in order to produce procedural instruments ('conventions' or 'protocols') allowing them to stabilize the use of technological tools that might otherwise vary across the local sites where they are applied. Cambrosio and Keating point to a configuration whereby expert judgment may rely on stable tools: where there is no technical black-boxes (e.g. a testing method), then at least a set of agreed principles offers common references for experts to base their actions on. Thus, even if the diversity of QSAR models prevents them from being used as stable instruments that would ensure the production of mechanical objectivity, a procedural approach could be seen as an answer. Since the expert judgment about the hazards of a substance implies a judgment about the validity of the QSAR model being used, then standardized procedures for crafting valid models could be valuable resources. Would an approach based on the standardization of procedures offer a path forwards for experts working in public agencies to use QSAR models?

This directly echoes some of the propositions made at the OECD, where the significant variation of QSAR uses across countries was tied to an issue of harmonization:

> The regulatory use of (…) (Q)SARs varies considerably among OECD member countries, and even between different agencies within the same member country. This is partly due to different regulatory frameworks, which impose different requirements and work under different constraints, but also because an internationally harmonised conceptual framework for assessing (Q)SARs has been lacking. The lack of such a framework led to the widespread recognition of the need for an internationally-agreed set of principles for (Q)SAR validation. The development of a set of agreed principles was considered important, not only to provide regulatory bodies with a scientific basis for making decisions on the acceptability (or otherwise) of data generated by (Q)SARs, but also to promote the mutual acceptance of (Q)SAR models by improving the transparency and consistency of QSAR reporting. (OECD, 2007: 15)

In this quote, "the development of a set of agreed principles" can be read in the terms of regulatory objectivity. It proposes international coordination for producing conventions. Within the international organization, this objective is directly connected to a boundary work, between internationally harmonized procedures that could guarantee the validity of the modelling approach, and the technical content of the model, which could be adapted to local situations according to regulatory choices (Thoreau, 2016). The task of the international organization, here, is to define generic principles of use, defined in such ways that they do not cross the perimeter of states' regulatory choices. Distinguishing international principles from (nationally-produced) technical content is both a way of standardizing QSAR models through conventions and ensuring international agreement without delving into potentially contentious regulatory choices.

The principles that the OECD released were the following:

> To facilitate the consideration of a (Q)SAR model for regulatory purposes, it should be associated with the following information:
> 1. a defined endpoint;
> 2. an unambiguous algorithm;
> 3. a defined domain of applicability;
> 4. appropriate measures of goodness-of-fit, robustness and predictability;
> 5. a mechanistic interpretation, if possible (OECD, 2007: 14).

These guidelines offered a way of ensuring international agreement about QSAR validation processes. Yet these principles had to do so without entering the domain of regulation, which is that of sovereign policy choices, and outside the scope of OECD intervention. Thus, instead of stating which endpoints or which algorithms should be used (choices potentially related to regulatory decisions), the guidelines stated that the two had to be identified in unambiguous ways. For the OECD intervention to be acceptable, QSAR validation principles had to be framed in a very general way.

The attempt to craft principles according to which the quality of QSAR models could be assessed is directly connected to a crucial issue for model-making, namely validation. Validating a model is both a technical task, tied to the scientific value of the model, and a political one, as it must be decided whether or not the model is robust enough to ground policy action (Edwards, 1999). While the OECD principles only considered the validation of QSAR models in general terms so that the international organization would not enter the perimeter of states' regulatory actions, the European institutions undertook an explicit reflection about whether and how QSAR models could be validated.

Validating QSAR models can be carried out by processing the data that have been used to construct the statistical correlations (this is described as "internal validation"), or other data (e.g. chemicals of known risks, on which the model will be run, and its predictions checked against the known risks of the tested chemicals). The latter approach is called "external validation" and is deemed more robust for regulatory choice by QSAR specialists (Gramatica, 2007). Yet external validation also requires additional data, and additional testing to check whether the predictions according to the model are correct, and yet another validation process for the choice and use of these additional data.

Considering the diversity and permanent evolution of statistical tools, Andrew Worth, QSAR specialist and Senior Scientific Officer at the Joint Research Centre (JRC) of the European Commission, concludes that the validity of a given model cannot be "set in stone":

> There should be nothing to fear from this process, since no conclusion on the validity of an experimental test or a (Q)SAR model is ever set permanently in stone — scientific and technical developments should always be taken into account. The question will always be when should the validity of a (Q)SAR (or a test method) be reviewed, either due to an adaptation of the model (test) itself, or because a new assessment (e.g. statistical) method is developed, or because new information (e.g., test data) becomes available. (Worth et al., 2004: 356)

In practical terms, this means that the standardization of validation processes can only take the form of general principles, leaving the practical conduct of validation to the particularities of the regulatory and technical situations at stake. Depending on the type of chemicals and models, internal or external validation processes will be used, and in ways that will differ from one case to the next. Thus, QSAR practitioners and regulators need to re-examine the appropriate validation methods for each new situation.

Situated expert judgment and the QSAR toolbox

Validation processes can only take the form of general prescriptions. This makes it impossible to consider QSAR models as stable black-boxes that could circulate straightforwardly across various domains of application. This does not mean that standardization is impossible, but that this standardization cannot take the form of technical harmonization (if, for instance, descriptors or endpoints were predefined) or procedural harmonization (if widely applicable validation principles were identified). Both types of harmonization (technical and procedural) require a certain stability of the technology being standardized, whether a stable instrument turned into a black-box circulates across various sites of application, or stable principles define procedures expected to be generally applicable. This means that QSAR models cannot be grounded on mechanical objectivity and the accompanying 'view from nowhere', or on regulatory objectivity and the coordinated approach on which it relies. How then can we understand the type of expert judgment at play when QSAR models are used? Another OECD initiative, the "QSAR toolbox" developed in partnership with the European Chemical Agency (ECHA), can help us to understand how experts working in public agencies are expected to use QSAR models.

Developed at the OECD and supported by ECHA since 2008, the QSAR toolbox is a free software application designed to "identify and fill (eco)toxicological data gaps for chemicals hazard assessment" (ECHA, 2011). It is intended to be used by private companies seeking to evaluate the hazard of the substances they produce, by experts working in public agencies and in charge of evaluating companies' propositions, and by other stakeholders[2]. Contrary to what its name seems to indicate, the QSAR toolbox does not provide a ready-made QSAR model fit for application on any given chemical. Rather, it brings together:

- databases with results from experimental studies;
- accumulated knowledge for structural characteristics (alerts) that can indicate the presence of hazards and other properties, and
- tools to estimate missing experimental values by read-across, by trend analysis (i.e. interpolating [preferred] or extrapolating from a trend [increasing, decreasing, or constant] from tested to untested chemicals within a category) and/or by (Q)SAR models. (ECHA, 2011)

Thus, QSAR models are one component of a more general platform. This platform is fed with experimental data, some of which are related to the physical causality between "structural characteristics" and hazards (second bullet point in the previous quote), and comprise modelling tools, some quantitative (as QSAR models are), and others based on statistical approaches that do not use quantitative predictive modelling. An example of the latter in the quote above is "read-across", which consists in using available empirical data to estimate the missing ones. The QSAR toolbox does not attempt to deliver ready-made risk assessments for a user (whether a regulator or a scientist) eager to know the toxicity of a given chemical. Rather, it offers a way "to systematically group chemicals into categories according to the presence or potency of a particular effect for all members of the category." (ECHA, 2011). "A particular effect" relates here to the particular endpoint that the user might want to test, and which requires the mobilization of various experimental data and instruments, comprising QSAR models and other, non-quantified, statistical tools.

Rather than providing a neatly defined quantitative instrument to which the technical task of risk assessment could be delegated straightforwardly, the QSAR toolbox is a platform that demands a reflective and cautious intervention by users, as they work on its many components to gather a set of indications about whether a chemical could be grouped with others, and how so. The OECD (2007: 92) gives the example of choosing a "no-observed-effect" as an endpoint. It asserts that while such a level may be relevant for policy-making purposes, it may as well be irrelevant for the purpose of generating scientific knowledge, i.e. "referring to a specific effect within a specific tissue/organ under specified conditions" (OECD, 2007: 92). One sees here that an active uptake about the very purpose of choosing the endpoint will affect its relevance.

Thus, the QSAR toolbox can only be used by an informed user, who has particular regulatory objectives in mind. This informed user is able to identify the scope of the evidence provided, and its limitations. This means that the QSAR toolbox can in no way be mobilized as a black-boxed instrument that could be used without opening up its inner mechanism. It follows that the concern for the transparency of the platform is constant among both the designers and users of the QSAR toolbox. Allowing regulators to access the characteristics of databases has become a necessary condition for the platform to function, as an OECD official told us during an interview:

> What we're also going to develop in the new version is to have a kind of reliability score related to the database and the profile so that at least they are all well documented. (...) we are very transparent on how these databases or profiles are constructed, what kind of chemicals have been used to develop – which are included in the database. So, if you go to the Toolbox, you also have an "about" section. You select a database and click on the "about" section then you will get information on the database. (interview, OECD)

Being transparent about the toolbox is about making its inner mechanism visible. It is also about making it possible for users to contribute, by providing new experimental data that could refine the existing correlations. The toolbox is indeed designed to be fed on an on-going basis with new experimental data and refined statistical correlations. Such a development implies enrolling more and more users, so as to ensure both the collective legitimacy and the technical validity of the instrument. This enrolment process is driven by the constitutive process of the toolbox itself as depicted above. It follows that it cannot be considered as a mere "beta testing" phase after which the toolbox would be closed and remained unchanged. Instead, openness, try-outs and transparency are inherent to the exercise of QSAR modelling.

The case of the QSAR toolbox is particularly interesting to further our understanding of the difficulty related to the use of QSAR models. Many ECHA documents state the impossibility of envisioning a formal adoption process of QSAR models within the European regulation of chemicals (see section 1). It is a consequence of the approach lying at the heart of the QSAR approach, and, eventually, a consequence of the particular type of standardization that can be pursued. Rather than standardizing a technical content or a procedure, the OECD and ECHA proposed a constantly evolving platform expected to help its users group chemicals together, along lines that are permanently subject to change.

The QSAR toolbox is meant to make QSAR models usable. Examining how it does so, as we have just done, is a way of better identifying the characteristics of the QSAR models, and the ways in which they are expected to contribute to risk governance within the QSAR toolbox:

- QSAR models are constituted at the same time as the groups of chemicals which they are expected to govern, and cannot easily be disentangled from these groups;
- Their scientific and regulatory value can only be assessed according to general criteria, which then require case-by-case assessment of models;
- QSAR models are not stable entities circulating across situations of use. Rather, they are meant to be articulated with one another and with other methods (as in the QSAR toolbox), so as to be refined as new experimental data are produced;
- Therefore, their potential users are not expected to apply them as ready-made instruments that operate autonomously, but need to mobilize their informed judgment to assess the ways in which they can provide relevant information for a given regulatory purpose. This results in an emphasis on transparency.

All these characteristics made QSAR models unfit for standardization as black-boxed instruments. Private companies and public experts can use them in coordination with other approaches. A platform such as the QSAR toolbox is therefore better defined as a 'grey box', which is mobilized in different ways according to particular situations of use, and never meant to be closed to external examination. Eventually, the QSAR toolbox cannot serve as an unproblematic coordination device, which could guarantee the value of the risk assessment performed by private companies and could be used by public experts to validate it. The toolbox example provides an illustration of how expert judgment is expected to be exercised in the case of QSAR. Rather than grounding the expert intervention on the ability to mobilize stable instruments that make the individual characteristics of the expert disappear (as when mechanical objectivity is the objective) or on the possibility to refer to common procedures (as in a regulatory objectivity framework), QSAR models require expert judgement to be situated locally, and discussed in relation with particular regulatory objectives. This has consequences for risk governance, which the next and last section discusses.

What risk governance in the world of QSAR models?

So far, we have discussed how QSAR specialists craft their models, how the OECD proposes only general principles of validation and a QSAR toolbox that is neither the provider of ready-made instruments nor the vehicle for common and operational procedures. What about the work of people in charge of evaluating the proposals of companies attempting to register the substances they produce? This is the task of public experts working at the European Chemicals Agency, and at national agencies in charge of risk assessment. Our reflection started with the consideration of the practical difficulties that these actors encountered when using QSAR models. These experts working in public agencies do not develop QSAR models. Nor are they in charge of standardizing their use[3]. Instead, they need to evaluate the ways in which companies describe the risk profile of the substances they wish to register. The impossibility of using QSAR models as black boxes, and the mobilization of grey boxes such as the QSAR toolbox, has consequences on how they can assess the validity of companies' claims.

First, public agencies constantly need to examine the QSAR models used by companies. Consider for instance how members of the French public agency for environmental safety describe their roles in assessing how companies use QSAR models:

> - And I think that the challenge for us is to identify the limits and confront the companies. (…) If we are not able to deconstruct the reasoning and know what there is in black boxes, then we can't argue with what companies propose! We can't say that we don't accept because we would have checked the domain of application, or whatever. That's why we need internal competencies for that… for a counter-expertise really. (interview, ANSES)

This quote points to an important consequence of the use of QSAR models for risk assessment purposes. Because of the complexity of these methods, and the diversity of actors producing them (in various ways according to the particularities of the situation), public experts might find themselves in a position of weakness - as they need to assess pieces of evidence produced by non-standardized and ever more complex tools. This asymmetry is only made more acute by the diversity of actors producing QSAR models. In addition to public research centres, many companies and open-source communities also develop their own QSAR software, either licensed or not, for profit or not (Lo Piparo and Worth, 2010). Datasets to inform the models are compiled by many different actors, including scientists for knowledge-production purposes, but not only. Many statistical techniques or mathematical models can be tailored to the creation of a particular QSAR. Various heuristic tools and different classes of algorithms are designed as a means to browse through the diversity of data and gather different sorts of results, including a wealth of machine-learning techniques (Lavecchia, 2015).

Second, the nature of expert intervention evolves, as neither the delegation to a trusted instrument (as in a regime characterized by mechanical objectivity) nor the mobilization of collectively produced conventions (as in a regime of regulatory objectivity) are possible. When assessing the dossiers submitted by companies to apply for the registration of chemicals, officials at ECHA will examine the models by opening them up, and comparing them with experimental data, as one of them told us during an interview:

> If I know that for example the prediction is backed up by some solid hypothesis which is confirmed by for example different in vitro observations or other observations in vitro from similar substances, this is for me something much more important than just predictions generated by super duper fancy logic, for example neural networks. (Interview, ECHA)

This quote explicitly connects the diversity of the methods used to produce evidence that require transparency (the expert needs to know what is inside the models) with the possibility for the expert working in public agencies to draw on other sources of information. The same official eventually referred to experts' "own experience" in assessing the use of QSAR models:

> Regulators are not looking for the tool which will give you the smallest possible error in predicting something on your validation set; regulators are more keen on something which they can understand how it works and they can extrapolate it to the normal – **their own experience**. It's even easier to accept the tool which gives you some error, like for example a few units plus or minus, but you know that this is really more or less what's going on and this sounds reasonably good, rather than using some very advanced mathematical model which you cannot really follow and you don't even know exactly how those features have been generated by the model. (Interview, ECHA, emphasis added)[4]

When confronted with QSAR models, expert judgment is based on the expert's experience, and on his ability to confront the construction of the model itself with other sources of information. This directly echoes the expected functioning of the QSAR toolbox (see above). Yet it stands in uneasy relation to the complexity of QSAR models, as the potential sophistication of the statistical approaches might well turn some QSAR models into black-boxes that are impossible to open to the gaze of the experienced public expert.

A condition for carrying out such a situated expert judgment is that public experts have the possibility to access the inner functioning of the models presented to them. During an interview, an ECHA official explained the issue this situation raised in the following terms:

> the most important, most critical element for regulators is the transparency of the model. If you have a very sophisticated statistical model (…) this is not very convincing for regulators because they don't exactly know what was exactly the training set which you used to train those networks and even if you see that they are performing very well on your test validation set, it doesn't mean that they will perform equally good on the new substance which are out of the validation set. And this is the basic problem of all those advanced QSARs, that they are not so transparent because they are very complex and regulators have always this problem in understanding what will the logic behind the tool? What kind of features were driving predictions? Interview, ECHA

Thus, the requirement for transparency makes public experts wary of overly complex instruments that they would be unable to grasp (regarding, for example, the hypothesis, the domain of applicability, or the statistical methods being used). But the requirement for transparency also impacts the institutional role of public expert bodies. We showed in that the ways in which QSAR models are constructed is directly tied to regulatory objectives (as, for instance, the set of endpoints is chosen according to regulatory requirements, or the models and the group of chemicals on which they are expected to be applied are manufactured in the same process). This means that when examining companies' use of QSAR models, public experts working in agencies such as the ECHA also need to evaluate how model-based risk assessment approaches fit with regulatory objectives.

We can now get back to the three-body problem of expertise. The examination of the practical conduct of QSAR modelling and standardization shows that QSAR models challenge the three components of expert legitimacy. Rather than grounding expert judgment in the ability to deliver a 'view from nowhere' in which the individuality of the expert disappears, QSAR makes public experts fully- fledged individuals who need to draw on their personal experience to evaluate private actors' propositions. Rather than providing a set body of knowledge, possibly formalized in black-boxed instruments or standardized by stable procedures, the use of QSAR for the regulation of chemicals requires situated examinations that need to be adapted to the particularities of every case. Rather than being a public body intended to act in isolation as a provider of scientific advice to inform risk management decisions, from which it is neatly separated, ECHA is an institution that coordinates with national agencies and international organisations in developing tools for the evaluation of QSAR models, while articulating its risk assessment mission with regulatory considerations.

Conclusion

As models are increasingly expected to contribute to regulatory decisions, understanding their political consequences it crucial. This paper has focused on models aiming to produce statistical correlation, and discussed one of these political consequences, related to the type of expert judgement that models entail. The case of QSAR models in the European governance of chemical risks illustrates a type of expert judgment that is situated, as experts in regulatory bodies cannot consider models as stable technical black boxes, and cannot rely on standardized procedure to use them. This explains why QSAR models, while being seen as a powerful alternative to animal testing, are also considered with caution in expert bodies such as the European Chemicals Agency. Examining how QSAR models are crafted in practice, we showed that this situation is not the first step before these models can act as stable instruments, but is derived from their very characteristics (and of what makes them interesting in the first place). Attempts are made to standardize principles for their evaluation, and a "toolbox" is proposed by the OECD to carry forth their validation. Yet, the use of QSAR models does not imply either mechanical objectivity or regulatory objectivity. Instead, the use of QSAR models in the European regulatory context means that expert judgment is situated, and grounded in the experience of the expert.

Facing the proliferation of information produced by models that they cannot completely rely on, public experts are confronted with an asymmetry of resources. They need to invent procedures by which they can gather enough information to make regulatory choices. For private companies, the submission of dossiers is becoming more strategic than ever, since the plurality of available models means that some of them might suit their needs and interests better than others. These companies therefore mobilize resources and develop an in-house expertise on models in their routine R&D process. This results in increasing demands on public bodies in charge of critically examining the models used by industries. That transparency becomes a growing concern follows, since public experts need to open up the models, or at least gather information about them. As new private actors enter the picture (most notably the companies producing models), producers of chemicals need to engage in new strategic activities (choosing relevant models), and public experts need to re-invent their roles so that they are able to monitor both the construction and the use of models.

The case of models in the governance of chemicals is specific, yet has value for a broader reflection on the use of models for regulatory purposes, particularly correlation-based statistical instruments, as QSAR models are. The value of these particular models is tied to the empirical data they are based on, and to the domain of use they are applied to. This paper has shown that the type of objective knowledge that these models are claimed to produce requires an active intervention of the expert in charge of interpreting it. As models are increasingly called for to settle controversies, plan long-term developments, or argue for or against policy choices, it is crucial not to see them as ready-made providers of objective knowledge, but as instruments that re-work what objectivity is, and directly constrain how experts can and should act.

Acknowlegments

A previous version of this paper was discussed during the workshop “Produire la prediction: le travail de la modélisation pour l’évaluation des risques”, held in Paris, June 19th, 2017, at the invitation of David Demortain and Henri Boullier (project INNOX). The authors would like to thank warmly the anonymous reviewers, as well as the editors of this special issue, for their insightful comments and suggestions.

References

Boullier H (2016) *Autoriser pour interdire. La fabrique des savoirs sur les molécules et leurs risques dans le règlement européen REACH. Thèse pour le doctorat de sociologie*. Université Paris-Est Marne-la-Vallée.

Boullier H, Demortain D and Zeeman M (2019) Inventing Prediction for Regulation: Modelling Structure-Activity Relationships at the US Environmental Protection Agency. *Science & Technology Studies* 34(2): 137-157.

Cambrosio A and Keating P (2009) Biomedical Conventions and Regulatory Objectivity. A Few Introductory Remarks. *Social Studies of Science* 39(5): 651-664.

Cambrosio A, Keating P, Schlich T and Weisz G (2006) Regulatory objectivity and the generation and management of evidence in medicine. *Social Science & Medicine* 63: 189–199.

Daston L and Galison P (2007) *Objectivity*. New York: Zone books.

Demortain D (2009) Legitimation by standards: Transnational experts, the European Commission and regulation of novel foods. *Sociologie du Travail* 51:2 104-116.

ECHA (2008) Guidance on information requirements and chemical safety assessment. Chapter R.6: QSARs and grouping of chemicals. *European Chemical Agency*.

ECHA (2011) The OECD QSAR toolbox for grouping chemicals into categories. *European Chemical Agency*. ECHA-11-L-08-EN.

Edwards P N (1999) Global climate science, uncertainty and politics: Data-laden models, model-filtered data. *Science as Culture* 8(4): 437-472.

Edwards P N (2010) *A vast machine: Computer models, climate data, and the politics of global warming*. Cambridge (MA): MIT Press.

European Commission (2006) *Regulation (EC) n° 1907/2006 of 18 December 2006 Concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH)*. Consolidated version of 1 June 2015, doc. 2006R1907. Luxembourg: Publications Office of the European Union.

European Commission (2016) Better Regulation: Delivering better results for a stronger Union. *Communication from the Commission to the European Parliament, the European Council and the Council*. COM (2016) 615 final.

Fisher E, Pascual P and Wagner W (2010). Understanding environmental models in their legal and regulatory context. *Journal of Environmental Law* 22(2): 251–283.

Gramatica P (2007) Principles of QSAR models validation: internal and external. *Molecular informatics*, 26(5): 694-701.

Heaphy L (2015) The role of climate models in adaptation decision-making: the case of the UK climate projections 2009. *European Journal for Philosophy of Science* 5(2): 233-257.

Hilgartner S (2002) *Science on Stage. Expert Advice as Public Drama*. Stanford: Stanford University Press.

Jasanoff S (1990) *The Fifth Branch. Science Advisers as Policymakers*. Cambridge (MA): Harvard University Press.

Jasanoff S (2004) *States of Knowledge*. Cambridge (MA): MIT Press.

Jasanoff S (2005) Judgment Under Siege: The Three-Body Problem of Expert Legitimacy. In: Maasen S and Weingart P (eds) *Democratization of Expertise? Sociology of the Sciences Yearbook*, vol 24. Dordrecht: Springer, pp. 209-224. .

Jasanoff S (2011) The practices of objectivity in regulatory science. In: Camic C, Gross N and Lamont M (eds) *Social Knowledge in the Making*. Chicago: University of Chicago Press, pp. 307-337.

Latour B (1987) *Science in action: How to follow scientists and engineers through society*. Cambridge (MA): Harvard university press.

Laurent B (2017) *Democratic experiments. Problematizing nanotechnology and democracy in Europe and the United States*. Cambridge (MA): MIT Press.

Lavecchia A (2015) Machine-learning approaches in drug discovery: methods and applications. *Drug Discovery Today* 20(3): 318–31.

Lo Piparo E and Worth A (2010) Review of QSAR Models and Software Tools for predicting Developmental and Reproductive Toxicity. *JRC scientific and technical reports* EUR, 24522.

MacKenzie D (1993) *Inventing Accuracy. A Historical Sociology of Nuclear Missile Guidance*. Cambridge (MA): MIT Press.

OECD (2007) *Guidance document on the validation of (quantitative) structure-activity relationships [(Q)SAR] models*. OECD, ENV/JM/MONO (2007)2.

Salzman J (2005) Decentralized administrative law in the Organization for Economic Cooperation and Development. *Law and Contemporary Problems* 68: 3-4.

Saurugger S (2002) L'expertise: un mode de participation des groupes d'intérêt au processus décisionnel communautaire. *Revue française de science politique*, 4:52, 375-401.

Thoreau F (2016) 'A mechanistic interpretation, if possible': How does predictive modelling causality affect the regulation of chemicals? *Big Data & Society* July-December: 1-11.

Vos E (2000) EU food safety regulation in the aftermath of the BSE crisis. *Journal of Consumer Policy* 23(3): 227-255.

Wickson F and Wynne B (2012) The anglerfish deception. The light of proposed reform in the regulation of GM crops hides underlying problems in EU science and governance. *EMBO reports* 13: 100-105.

Worth A, Hartung T and Van Leeuwen CJ (2004) The role of ECVAM in the validation of QSARs. *SAR and QSAR in Environmental Research* 15(5–6): 345–358.

Worth A, Bassan A, De Brujin J et al. (2007) The role of the European Chemicals Bureau in promoting the regulatory use of (Q)SAR methods. *SAR and QSAR in Environmental Research* 18(1-2): 111-125.

Notes

1 We were co-partners of this research project, in charge of analysing the political aspects of models meant for regulatory purposes.

2 See OECD Toolbox website: https://www.qsartoolbox.org (accessed 2017-12-05).

3 The distinction is partly arbitrary since the ECHA experts also participate in the OECD working groups.

4 In this quote, 'regulators' is used to qualify the experts working in public agencies.

Laura Watts (2018) Energy at the End of the World. An Orkney Islands Saga. Cambridge and London: The MIT Press. 440 pages. ISBN 978-0262038898

Veera Kinnunen
veera.kinnunen@ulapland.fi

Jarno Valkonen

> The energy in these islands is a literal force that changes the body, and so changes what you know about the world. (Watts, p. 50)

The saga of the Energy Islands by science and technology studies scholar Laura Watts begins with a cup of tea. The reader is offered a cup of steaming hot black tea with milk, and s/he is invited to join the author for a walk and meet the Orkney Islands in all their messy, multisensory presence. The author then reminds the reader to pull on wellingtons and woollen mittens in order to bear with the force of the harsh island winds.

The focus of the book is on the *making* of *energy futures*. Watts takes a lead from science scholar Lucy Suchman who points out that innovations are not universal: they change when they travel from place to place. This view emphasizes that place matters – place itself is an effect. Landscape has agency in future-making. The energy future of the Orkney is irreducible from the place itself, its geography, history and culture. In any other place, the future of the low-carbon renewable energy would have to be made differently. Thus, Watts turns attention to the situated practices of energy future-making. From this place-specific point of view, she weaves a story of what happens to the future when it is being locally imagined and crafted. By scrutinizing closely the present "struggling, overheating, and partial electrical flows" in the cables of Orkney Islands, Watts asks: how could energy futures be made otherwise? (p. 68).

The study is based on ten years of ethnographic fieldwork in Orkney, an archipelago of around 20 populated islands that lie ten miles off the northeast coast of mainland Britain. Over the last decades, Orkney and its open-sea marine energy test site have become the centre of the ongoing revolution in sustainable power generation. The European Marine Energy Centre, EMEC (pronounced "ee-mek" as Watts notes) located in Orkney, has tested 30 different devices since its launch in 2003. Many other marine energy test sites around the world draw from the expertise of Orkney. In this respect, the Orkney Islands undoubtedly exert a significant impact on the future of renewable energy in the whole world: A local adaptation of low-carbon renewable energy future that the contemporary world so badly needs is already being lived in the Orkney Islands. In addition, as Watts points out, the Orkney Islands have another, less visible, but no less important, impact on the energy-futures of the world. The Islands are a living example of skilful resourcefulness of common people in energy generation. While peripheries are often represented as passive producers of resources, this study emphasizes the dynamic vibrancy of the edges. Whereas cities are utterly dependent on their edges, the resourceful edges have learned to make do without the centres. Getting by independently in the remote islands even when the connection to the mainland is cut off, as it often is, necessitates becoming both interdependent and inventive. In order to cope, one needs to form alliances with *every body* – including non-human agents, such as rocks, winds and tides. Thus, the islanders constantly seek ways to ally with the plentiful natural forces. They have come up with community-owned wind turbines, self-customized electric cars turned into power-storing mobile batteries, and ferries fuelled with wave-power. This forced and learned "make do" attitude makes the islands a flexible and open-minded place where different infrastructural solutions to the emerging problems are being constantly anticipated, imagined and trialled. Watts's description of being innovative at the end of the world reminds us of Tim Ingold's definition of design as a continuous process of dwelling, of making life liveable. The islander resilience is about making do with what one's got and being resourceful in inventing ways of getting by.

The saga of electricity is a saga of infrastructures. It pays attention to the power of infrastructures to hold over time and to define certain futures. Moreover, it draws attention to the work that is needed to maintain the stability and invisibility of the infrastructures. By focusing on the resourceful coping and experimenting with energy infrastructures – and occasional failure in these attempts – Watts emphasizes that as static as they may seem, they are not eternally unchanging.

Focusing on the frictions of making energy – the situated entanglements of Orkney 'electropolitics'– illustrates how complicated issue energy is. It is an untidy bundle of technology and natural forces, science and markets, history and future, national and local politics, calculations and sensory perceptions, dreams, hopes and fears, infrastructures and improvising, and it is as much a question of high-tech innovation and resourceful tinkering. By following the sometimes painfully difficult processes of turning the overflowing 'natural' elements, such as wind and waves, into utilizable form, the reader realises that 'energy' is not a generalizable issue but rather constituted in and through specific places and times. In the Energy Islands, the question of energy is a constantly pressing issue, an inseparable part of everyday life.

Energy at the end of the world is also a writing experiment. Methodologically, Watts follows in the footsteps of Donna Haraway and other feminist science studies scholars, who have called for recognizing the coupling together of writing as creative craft and academic research. In her saga, Watts pushes the boundaries of fact and fiction ("Stories and Fables") and weaves them together into an empirical experiment on feminist science studies. Donna Haraway defines SF (meaning simultaneously Science Fiction, Speculative Fiction, Speculative Fabulation, or Feminist Fabulation) as *worlding* – writing as a knowledge-making and world-making practice. Thus, SF is less about critically observing and analysing the existing world, and more about being mindful of the worlds that the writing opens, and also about bearing its consequences. As Watts puts it: "Tales have the power to make the future – or rather, some futures for some people" (p. 11).

Albeit being speculative and fabulated, the Energy Islands saga is far from fictive. It is based on a long-term, meticulous and committed ethnographic fieldwork on the islands. The ethnographic material alone comprises over 10,000 words of field notes and 2,000 photographs. As a serious anthropologist, Watts takes the myths, gossip and other marginal forms of knowing as seriously as historical narratives, numeric information and other scientific 'facts'. Although Watts's text structurally follows the conventions of academic writing and uses a reference-based format, she plays with experimental forms of ethnography by drawing from eclectic sources, using first-person narrative, and seeking to address the creative process of research itself, rather than reporting generalized findings. The traditional research text is enriched and complemented with, for instance, an orally chanted poem, photographs, vividly narrated accounts of encounters with a female monster Electric Nemesis, and even series of comic-style illustrations of this fictional monster.

Watts composes her saga of Orkneyan electropolitics from fragments of actual and imagined events, encounters and stories and through these accounts, she shows the richness of elements entangled in the making of energy futures.

With Watts's authorial prowess, the outcome is an extremely enjoyable read. She is not just a scientist, but also a great storyteller. Energy at the end of the world succeeds in being a piece of research that is so captivating that the reader is unable to put the book down – a rare reading experience in the academic genre

Laura Watts's study is about making energy futures. However, the study itself is a contribution to future-making in that it paves the way for the future ethnographers of science and technology studies. It shows what ethnographic science and technology studies can do at their best, and how scientific texts can be enjoyable, inspiring and convincing. For those in doubt, the book is a powerful proof that the social sciences and humanities can offer valuable insights into research on technologies and energy infrastructures e.g. by bringing forth the complexity of more-than-human entanglements needed to produce energy, and by revealing the uneven power relations and multiple ontologies at play in the situated practices of making futures, and showing the efforts that it takes to work and fit these different realities together.

Ericka Johnson (2017) Gendering Drugs. Feminists Studies of Pharmaceuticals. Cham: Palgrave Macmillan. 232 pages. ISBN 978-3-319-51486-4

Irene Blanco Fuente
ireblanc@ucm.es

Gendering Drugs edited by Ericka Johnson was published in 2017 by Palgrave Macmillan. The work presented is result of a five-year group research project and increase understanding about the intersection of pharmaceuticals and gender. On the one hand, drugs can be gendered and tailored to either men or women. On the other hand, they contribute to create gendered subjectivities for us, so they can produce gender.

This work analyses how drugs can be sexed/gendered and produce sex/gender. Pharmaceuticals, as they are technologies and nonhuman elements, can be active agents involved in social and power relationships. They cannot be understood as being separated from each other, but as material-discursive entanglements. The drugs context analysis provides us the possibility of calling attention on how some discourses articulate subject positions and relations. The book has used a wide fieldwork, as drug pharmaceutical advertisements, medical guidelines and the experience of some of the people directly affected by the side effects of some of tools for health.

The work is theoretically premised in feminist technoscience studies with an interest in material-discursive bodies (Haraway, 1997; Barad, 2007) and how pharmaceuticals produce bodies and gender (Petryna et al., 2006). From a feminist critique, the concern for masculinities, non-binary sex/gender understandings, and the intersection of race, class, sexuality, and global inequalities is present. The posthuman approach extend analytical focus to incorporate nonhumans as active agents.

The book is divided into three parts and it has ten chapters in all. Part I analyses pharmaceuticals in different life phases and adjacent health concerns, such as Alzheimer's disease, prostate treatments or trans-childhood. The section starts with ethnographic observations of Tara Mehrabi of breeding transgenic Alzheimer's flies where Mehrabi worked as a laboratory assistant. The chapter shows that in Alzheimer's research not only are fly models tested upon, research also produces sex differences. The sexing processes of flies are essential for pharmaceutical development in Alzheimer well before pharmaceutical cure or treatment is manifest. As such the gendered nature of pharmaceuticals begins well before the materialisation of pharmaceuticals but a gendered binary notion of gender guides the very research methods and questions asked about diseases, in this case Alzheimer's. The next chapter is written by the editor of the book Ericka Johnson. It takes the example of alpha-blockers used to treat lower urinary tract secondary symptoms of benign prostate hyperplasia (LUTS/BPH) to develop an analysis based on Karen Barad's (2007) concepts of actant and intra-action. Based on three sets of different clinical practice guidelines the chapter concludes that alpha-blockers treat and increase the size of the prostate, creating also a pharmaceuticalized prostate. The last chapter is written by Celia Roberts and Cron Cronshaw focuses on trans-childhood and the uses of the gonadotropin-releasing hormone (GnRH) used to prevent pubertal development. The authors argue that we live in a 'pharmacopornographic era', as Paul Preciado (2013) says, in which legal minors cannot engage in adult-centric policies. They are dependent on adults to obtain pharmaceuticals. The authors advocate for the recognition of the needs experienced which are produced partially in these kinds of discourses and practices which try to solve them, and the obligation of including trans people as an active and essential part in these politics.

Part II analyses advertisements of pharmaceutical treatments, which includes commercial images and discourses and the way they also prescribe relational practices for individual subjectivities. First, Ericka Johnson and Cecilia Åsberg analyse two pharmaceuticals with the assumption that they also prescribe particular ways of becoming a healthy subject. They analyse three advertisements, two about Alzheimer's drug and one about a benign prostatic hyperplasia (BPH) drug. In all of cases, they conclude, the drug represents an important part of the relational agency between the potential users of the drugs and their partners, introducing a unique element in intimate relationships. After this, the Lisa Lindén studies Gardasil the vaccine advertisements produced by pharmaceutical company Sanofi Pasteur in Sweden against cancer caused by human papillomavirus (HPV) aimed at young girls. The advertisements show the vaccine as a product through which parents can exercise care responsibly. Advertisements were especially focused on mothers, who are made responsible for such matter, leaving fathers an invisible role: vaccines appear to be as nonhuman participant involved in the relationships of care. This second part demonstrates the importance of addressing the pharmaceutical advertisement as material-discursive elements, which take part in the construction of diseases and subjectivities.

Part III analyses throughout its three chapters gendering in HPV campaigns in three different contexts: Colombia, the UK and Austria. It commences with an analysis by Oscar Javier Maldonado of tensions around the introduction of HPV vaccines in Colombia. The vaccine represents a new gendered technology aimed at girls as sexualized subjects. The context implies that the relation between sex and pathology could reproduce a stigma. Maldonado reflects on the intersections of race, gender and class in the cancer prevention campaigns as well as the relationship between HPV and social difference. The chapter is followed by Ali Hanbury's analysis about the introduction of HPV vaccine in the UK as a possibility to make young women responsible for their own health and so their sexual partners' health, too. Hanbury takes five cases of experiences of vaccine injury to prove deficiencies in this vaccination program. Despite the fact that the vaccine is presented as being gender-neutral, vaccination programmes continue to be paternalistic especially over women's bodies and autonomy. Finally, Lisa Lindén and Sina Busse describe how Austria is the first country to offer the HPV vaccine to both girls and boys with no expense. Offering the vaccine universally changes the gendered focus; now girls are not the one and only group at risk but all children equally. The analysis illustrates how sexual health is built and how a discursive shift makes changes possible. This third part is relevant due to the fact that a posthuman approach of HPV diagnosis experience is very interesting and could open new analysis about it.

Concerning to the omissions in the book along the three parts, in the first one in the chapter about Alzheimer it would have been interesting to deepen in the relation between the construction of the pharmaceutical in the laboratory and the patients experience with the treatment. In addition, the third chapter about trans children and pharmaceuticals could have mentioned the strategies developed in some cases to obtain these pharmaceuticals in an illegal way as an evidence of the situation in terms of politics which concern trans people. The second part would have been more completed with the pharmaceuticals' advertisements analysis of other diseases or health campaign because this method of analysis is very interesting and enlightening. Finally, the third part would have been better with more information about the situation of the HPV vaccine in more countries.

In a nutshell, the book is a relevant contribution which highlight the subjective processes in science and evince how nonhuman takes part in gender construction. It throws light on the construction of bodies and subjects in relation to pharmaceuticals, and the multiplicity of such material-discursive entanglements.

References

Barad K (2007) *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning*. Durham, NC: Duke University Press.

Haraway D (1991) *Simians, Cyborgs, and Women. The Reinvention of Nature*. New York: Routledge.

Petryna A, Lakoff A and Kleinman A (2006) *Global Pharmaceuticals. Ethics, Markets, Practices*. Durham, NC: Duke University Press.

Preciado P (2013) *Testo Junkie*. New York: Feminist Press.

Science & Technology Studies

Volume 32, Issue 4, 2019