Economics
The Open-Access, Open-Assessment E-Journal

Escape from model-land

Erica L. Thompson and Leonard A. Smith

Abstract

Both mathematical modelling and simulation methods in general have contributed greatly to understanding, insight and forecasting in many fields including macroeconomics. Nevertheless, we must remain careful to distinguish model-land and model-land quantities from the real world. Decisions taken in the real world are more robust when informed by estimation of real-world quantities with transparent uncertainty quantification, than when based on "optimal" model-land quantities obtained from simulations of imperfect models optimized, perhaps optimal, in model-land. The authors present a short guide to some of the temptations and pitfalls of model-land, some directions towards the exit, and two ways to escape. Their aim is to improve decision support by providing relevant, adequate information regarding the real-world target of interest, or making it clear why today's model models are not up to that task for the particular target of interest.

(Published in Special Issue Bio-psycho-social foundations of macroeconomics)

Authors
Erica L. Thompson, London School of Economics and Political Science, Centre for the Analysis of Time Series (CATS), UK, e.thompson@lse.ac.u
Leonard A. Smith, Mathematical Institute, University of Oxford, UK

1   Introduction

Computational simulations and associated graphical visualisations have become much more sophisticated in recent decades due to the availability of ever-greater computational resources. In part, the qualitative visual appeal of these simulations has led to an explosion of simulation-based, often probabilistic forecasting in support of decision-making in everything from the UK's GDP and unemployment to weather forecasting and American Football, to nuclear stewardship and the Earth's future climate. In each case, the outputs of these models are again used to inform real-world decision making, sometimes in public view, sometimes not. We argue that the utility and decision-relevance of model simulations must be judged based on consistency with the past, and where possible on out-of-sample predictive performance, and on expert judgement; never based solely on the plausibility of their underlying principles or on the visual "realism" of outputs.

Model-land is a hypothetical world (Figure 1) in which mathematical simulations are evaluated against other mathematical simulations, mathematical models against other (or the same) mathematical model, everything is well-posed and models (and their imperfections) are known perfectly. It also promotes a seductive, fairy-tale state of mind in which optimising a simulation invariably reflects desirable pathways in the real world. Decision-support in model-land implies taking the output of model simulations at face value (perhaps using some form of statistical processing to account for blatant inconsistencies), and then interpreting frequencies in model-land to represent probabilities in the real-world. Probabilities in model-land are what I.J. Good (1959) referred to as tautological probabilities (mathematically well defined by the statement of the problem itself). Elegant though these mathematical-models may be, something is lost in the move back to reality; very low probability events and model-inconceivable "Big Surprises" are much too frequent in applied meteorology, geology, and economics. We have found remarkably similar challenges to good model-based decision support in macroeconomics, energy demand, fluid dynamics, hurricane formation, life boat operations, nuclear stewardship, weather forecasting, climate calculators, and sustainable governance of reindeer hunting.

One can justifiably aim to transform simulations in model-land into information regarding the real world, but only where such information exists to be found in the model-land simulations. Our aim is a decision-making process that remains acceptable to all involved regardless of the outcome; ideally a process retained without modification and used again under similar conditions in the future *regardless* of the outcome, unless a deeper understanding of the system has been obtained. This cannot be accomplished in model-land. Uncomfortable departures from model-land are required for (good) decision support.

Our image of model land is intended to illustrate Whitehead's (1925) "Fallacy of Misplaced Concreteness". Whitehead (1925) reminds us that "it is of the utmost importance to be vigilant in critically revising your modes of abstraction". Since obviously the "disadvantage of exclusive attention to a group of abstractions, however well-founded, is that, by the nature of the case, you have abstracted from the remainder of things". Model-land encompasses the group of abstractions that our model is made of, the real-world includes the remainder of things.

Big Surprises, for example, arise when something our simulation models cannot mimic turns out to have important implications for us. Big Surprises invalidate (not update) model-based
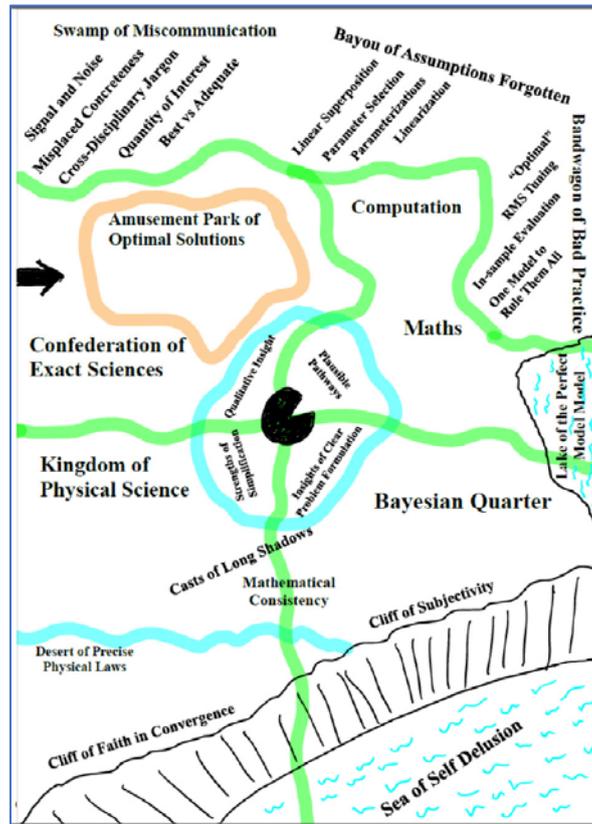
Figure 1: A map of Model-land. The black hole in the middle is a way out.

probability forecasts: the conditions I in any conditional probability P(x|I) changes. In "weather-like" tasks, where there are many opportunities to test the outcome of our model against a real observed outcome, we can see when/how our models become silly (though this does not eliminate every possibility of a Big Surprise). In "climate-like" tasks, where the forecasts are made truly out-of-sample, there is no such opportunity and we rely on judgements about the quality of the model given the degree to which it performs well under different conditions.

In economics, forecasting the closing value of an exchange rate or of Brent Crude is a weather-like task: the same mathematical forecasting system can be used for hundreds or thousands of forecasts, and thus a large forecast-outcome archive can be obtained. Weather forecasts fall into this category; a "weather model" forecast system produces forecasts every 6 hours for, say, 5 years. In climate-like tasks there may be only one forecast: will the explosion of a nuclear bomb ignite and burn off the Earth's atmosphere (this calculation was actually made)? How will the euro respond if Greece leaves the Eurozone? The pound? Or the system may change so much before we again address the same question that the relevant models are very different, as in year-ahead GDP forecasting, or forecasting inflation, or the hottest (or wettest) day of the year 2099 in the Old Quad of Pembroke College, Oxford.

2   Simulations and Model-lands: the map is not the territory

As the simulation of complex systems becomes routine in many areas of research (Petersen, 2012), the distinction between simulated variables and their real-world counterparts can become unclear (Beven et al., 2012). As a trivial example, when writing about forecasts of household consumption, energy prices, or global average surface temperature, many authors will use the same name and the same phrasing to refer to effects seen in the simulation as those used for the real world. These authors probably are not actually confused about which is which; our point is that readers of conclusions would benefit from a clear distinction being made, especially where such results are presented as if they have relevance to real-world phenomena and decision-making.

Why are we concerned about this? It is not just a philosophical worry about semantics but real implications we have observed when the consumers of this material realise just how different the model-variables are from the real-world phenomena they face. Something seen on the map may not correspond to what is in the territory; worse, something *not* seen on the map may be encountered when we explore the territory. Within model-land, we cannot even enunciate the possibility of a "Big Surprise", let alone think about the probability of such an event occurring. Yet the possibilities remain of economic surprises, previously-unseen weather events, energy price spikes, or worse-than-expected climate impacts, even where these are not simulated by today's models. Such events, in fact, happen disturbingly often. Can we escape model-land by targeting exclusively the less comfortable, but better-informed and much more relevant real-world entities in decision-making?

It is comfortable for researchers to remain in model-land as far as possible, since within model-land everything is well-defined, our statistical methods are all valid, and we can prove and utilise theorems (Judd and Smith, 2001). Exploring the furthest reaches of model-land in fact is a very productive career strategy, since it is limited only by the available computational resource. While pure mathematicians can, of course, thrive in model-land (see Box below), applied mathematicians have a harder row to hoe, inasmuch as, for large classes of problems, the pure mathematicians have proven that *no* solution to the problem will hold in the real world (Judd and Smith, 2004; Judd et al., 2008; see also, of course, the many relevant writings of Poincare and Smale on this situation.).

For what we term "climate-like" tasks , the realms of sophisticated statistical processing which variously "identify the best model", "calibrate the parameters of the model", "form a probability distribution from the ensemble", "calculate the size of the discrepancy" etc., are castles in the air built on a single assumption which is known to be incorrect: that the model is perfect. These mathematical "phantastic objects" (Tuckett and Taffler, 2008; Tuckett, 2011; Tuckett and Nikolic, 2017), are great works of logic but their outcomes are relevant only in model-land until a direct assertion is made that their underlying assumptions hold "well enough"; that they are shown to be adequate for purpose, not merely today's best available model. Of course, many assumptions are false in principle but negligible in practice and it is reasonable to ask, as we now do, whether this may not be the case here. Until the outcome is known, the ultimate arbiter must be expert judgement, as a model is always blind to things it does not contain and thus may experience Big Surprises.

Reasons for Staying in Model-land

Well-posed problems have exact solutions.

The use of which in the real world leads to disaster.

Optimal solutions exist, and might be found.

But they fail to perform as expected back in the real world.

I can be a Bayesian, with a principled path through at all stages of a project.

The Bayesian Way is invaluable for describing the challenge, but need never lead to a complete solution outside model land.

Arguably, one can construct actionable probabilities and use them optimally. *All Probabilities are conditioned on falsehood, and the Prob(Big Surprise) is never zero.*

Nonlinear systems and Chaos pose a host of interesting mathematical problems to be solved! *Mathematicians thrive in model-land! But for large classes of problems, we can prove those interesting solutions will not hold in the real world.*

While a model is of no help in accurately forecasting phenomena it cannot simulate, it can however be useful in detecting that something has gone badly wrong in model land. One can detect that the model-state is in a region where the model has never explored or there is no data, or the model can detect that its outputs are abnormal. In our real-time forecast systems the model displays a "purple light" (Smith, 2016) to indicate it should not be interpreted as usual.

3   Structural model error and its implications: the Hawkmoth Effect

To understand the depth of the problem, it is helpful to unpick the mathematics further. Chaos is no longer as fashionable as it was a few decades ago, but most readers will be familiar with the so-called Butterfly Effect – the concept that a small difference in initial conditions (perhaps stepping on a butterfly) can result in a large difference in the outcome of a complex dynamical system over some timescale. This was noticed by Edward Lorenz (1963), coming to his attention due to slight numerical truncation error in a simple mathematical system.

In the 21st century, the Butterfly Effect is a solved problem (Judd and Smith, 2001). To account for the possibility of error in the initial conditions, instead of taking a single best-estimate of the system state, we instead use an ensemble (multiple initial conditions) to represent a probability distribution over the initial conditions consistent with both the observations and the mathematical model (Judd and Smith, 2001). This ensemble of model states is then interpreted as a probability distribution in the real world which encompasses all possible outcomes given the uncertainty in the initial conditions, parameter values, and other numbers.

This mathematical solution assumes that the equations of the dynamical system are known perfectly, as was the case for Lorenz's three-dimensional mathematical model. Where our

complex model is not an end in itself, however, but a stand-in for a complex real-world system such as the Earth's atmosphere, the economy, or the energy system, then we can say with confidence that our model is not perfect (Smith, 2002). We are then in the realm not of initial condition error but of structural model error: if our chaotic model is only slightly mathematically mis-specified then a very large difference in outcome will evolve in time even with a "perfect" initial condition (Smith, 2007). We term this the Hawkmoth Effect (Thompson, 2013).

Given a structurally imperfect model and assuming we have managed to procure the "uncertainty" at t=0 with a perfect ensemble (such a thing need not exist) , then the probability distribution that we arrive at by using this ensemble will grow more and more misleading: misleadingly precise, misleadingly diverse, or just plain wrong in general (Figure 2). If the model-state space of our model is imperfect, it is impossible to specify a perfect ensemble – doing so requires topological conjugacy (Smith, 1995). Nevertheless, it may yield useful forecast information for quite a long time (Smith, 2006). The model-state space consists of a finite number of real variables and a restricted region of model-state space in which there are ensemble members arguably consistent with both the observations and the model's dynamics. Given a perfect model with imprecisely known parameters and imprecisely known initial conditions, the challenge is merely one of finding well-defined (but imprecisely known) real numbers. Bayesian methods are effective at reducing imprecision (Berger and Smith, 2018). Structural model error is different: the Model itself is a function, not a real number. It lies in a function space, and it is not at all clear how to put a relevant probability distribution on this function space. Neither is it clear why multi-model ensembles are taken to represent a probability distribution of future states at all. The distributions from each imperfect model in the ensemble will differ from the desired perfect model probability distribution (if such a thing exists); it is not clear how combining them might lead to a relevant, much less precise, distribution resembling the real-world target of interest (assuming such a thing exists).
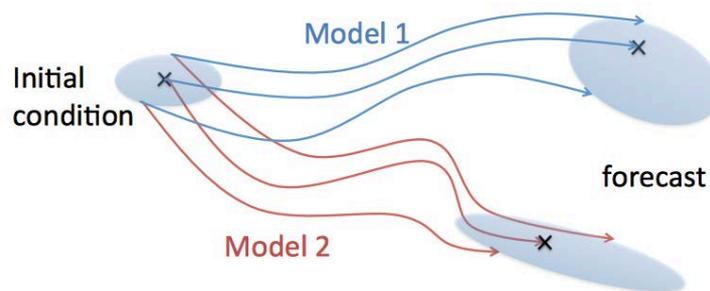
Figure 2: The Hawkmoth Effect. With a perfect estimation of the uncertainty around the initial condition, only a perfect model can result in a perfect specification of the uncertainty in the forecast. An imperfect model may be arbitrarily wrong even when initialised ideally. Note that identical initial conditions have finite separation immediately under Model 1 and Model 2; chaotic divergence identical initial conditions remain identical for all time.

It is sometimes suggested that if a model is only slightly wrong, then its outputs will correspondingly be only slightly wrong. The Butterfly Effect (Lorenz, 1963) revealed that in deterministic nonlinear dynamical systems, a "slightly wrong" initial condition can yield wildly wrong outputs. The Hawkmoth Effect (Thompson, 2013) implies that when the mathematical structure of the model is only "slightly wrong" then one almost certainly loses topological conjugacy (Smale, 1966). In this case, even the best formulated probability forecasts will be wildly wrong in time. These results from pure mathematics hold consequences not only for the aims of prediction but also for model development and calibration, ensemble interpretation and of course for the formation of initial condition ensembles. The limitations discussed above apply to realistic simulation with differential equations far from geophysics and economics. They suggest an ultimate barrier we can never pass if we approach by the mathematical methods of today. Both in geophysics and economics of today, there are often much harsher macroscopic errors and shortcomings that have not yet been resolved (model-mountains in climate models can be kilometres shorter than their real-world namesakes).

Naïvely, we might hope that by making incremental improvements to the "realism" of a model (more accurate representations, greater details of processes, finer spatial or temporal resolution, etc.) we would also see incremental improvement in the outputs (either qualitative realism or according to some quantitative performance metric). Regarding the realism of short-term trajectories, this may well be true! It is not expected to be true in terms of probability forecasts. And it is not always true in terms of short term trajectories; we note that fields of research where models have become dramatically more complex are experiencing exactly this problem: the nonlinear compound effects of any given small tweak to the model structure are so great that calibration becomes a very computationally-intensive task and the marginal performance benefits of additional subroutines or processes may be zero or even negative. In plainer terms, adding detail to the model can make it less accurate, less useful.

The observation that complex models may be less informative than simple models (or comparatively informative but much more costly in terms of computational resource, human resource and cold hard cash) may, paradoxically, assist decision-making by providing a stopping-point to what is otherwise a potentially endless quest for "more research", "better information" or "less uncertainty" before a decision is made. How good is a model before it is *good enough* to support a particular decision – i.e., adequate for the intended purpose (Parker, 2009)? This of course depends on the decision as well as on the model, and is particularly relevant when the decision to take no action at this time could carry a very high cost. Ideally, one would start with the decision and consider potential models in light of their ability (or lack thereof) to inform this decision. Starting in model-land, one can continue forever improving a model and exploring the implications of introducing new complexity: evaluating in model-land will no doubt show some manner of "improvement." When the justification of the research is to inform some real-world time-sensitive decision, merely employing the best available model can undermine (and has undermined) the notion of the science-based support of decision making, when limitations like those above are not spelt out clearly (Smith, 2002; Frigg et al, 2015; Smith and Petersen, 2014; Beven, 2019; Beven, 2019b). For what tasks is the model considered adequate for purpose (Parker, 2019)? Is the extent to which the model is not expected to be adequate for a range of purposes of interest presented in a clear and transparent manner?

4   Challenges for real-world decision-making

We have various illustrations of how to extract information from (ensembles of) simulations which out-perform naïve statistical model forecasts, and avoid some of the misleading assumptions that are unavoidable if one stays in model-land. These illustrations include the 2018 Pakistan heatwave (Thompson and Smith, 2019), pricing and trading in the energy market in week out to two (when constrained by regulation) (Smith (2016)), and experiments designed to explore model error in practice for nuclear stewardship.

In our work with the START Network, a group of humanitarian NGOs, we are looking at ways to streamline the use of information from weather (and other) forecasts to anticipate humanitarian crises. Following an alert, a 72-hour process decides whether to activate the release of funds and then how to allocate money to projects. In principle, for many situations it is possible to determine a timescale of applicability for the forecast. This can help both when it shows that information is available, as it allows confident use of a set of operating procedures based on the forecast, and also when it shows that relevant information is not available and the decision should be made based on other inputs. In the case of heatwave in Pakistan, it was made clear by one of us that a reasonably confident forecast can be made with sufficient lead time (several days) to follow START procedures and take actions which help to reduce the likely impact on potentially vulnerable groups. As we develop and extend this framework to other regions and hazards, such as tropical cyclones and droughts, we expect that in some cases the forecast information will be negligible and will then advise that the rapid-turnaround decision should focus more on other factors such as social, economic and practical bases for action. Taking one set of models off the table is a valuable contribution to the decision process – using a mis-informative forecast simply because it is the "best available" is a nonsense.

Some models are used for convenience, because they are "objective" in the sense of getting a single answer under the same input conditions regardless of user (which we note is not at all the opposite of "subjective", since the construction of any model requires expert judgement about the applicability of that model and the validity of any assumptions) and because they provide an unambiguous guide for policy-making. An example is the use of DSGE (Dynamic Stochastic General Equilibrium) models by many Western academics and Central Banks including the ECB. Lack of inclusion of a financial sector (resulting from assumptions about the efficiency of markets) was a "Known Neglected" which may well have ruled out the possibility of a banking collapse in model-land and the real-world economic consequences experienced in 2007/8. As with other models, simply "fixing the bug" or adding in the newly-identified mechanism each time something unexpected happens, is not a recipe for confident forward prediction.

5   Working in Model-land

You may be living in model-land if you…
- try to **optimise** anything regarding the future;
- believe that **decision-relevant probabilities** can be extracted from models;

o believe that there are **precise parameter values** to be found;
o refuse to believe in anything that has not been seen in the model;
o think that **learning more will reduce the uncertainty** in a forecast;
o explicitly or implicitly **set the Probability of a Big Surprise to zero**; that there is nothing your model cannot simulate;
o want "**one model to rule them all**";
o treat any failure, no matter how large, as a call for further extension to the existing modelling strategy.

You may be near the exit if you…
o are now concerned about your forecast system's Probability of a Big Surprise;
o are uncomfortable to realise that given 100x the resources, your team would build a better model rather than refine the probability estimates of your current model;
o are careful to distinguish $\mathbf{x}$ from its model-land namesake $\mathbf{x}_{model}$;
o have forsaken the quest for a model that tells you everything, to search instead for models that tell you something about your Quantity of Interest;
o are intrigued by the idea of finding your model's Relevant Dominant Uncertainty – the aspect of the model you should improve first.

It is helpful to recognise a few critical distinctions regarding pathways out of model-land and back to reality. Is the model used simply the "best available" at the present time, or is it arguably adequate for the specific purpose of interest? How would adequacy for purpose be assessed, and what would it look like? Are you working with a weather-like task, where adequacy for purpose can more or less be quantified, or a climate-like task, where relevant forecasts cannot be evaluated fully? Mayo (1996) argues that severe testing is required to build confidence in a model (or theory); we agree this is an excellent method for developing confidence in weather-like tasks, but is it possible to construct severe tests for extrapolation (climate-like) tasks? Is the system reflexive; does it respond to the forecasts themselves? How do we evaluate models: against real-world variables, or against a contrived index, or against other models? Or are they primarily evaluated by means of their epistemic or physical foundations? Or, one step further, are they primarily explanatory models for insight and understanding rather than quantitative forecast machines? Does the model in fact assist with human understanding of the system, or is it so complex that it becomes a prosthesis of understanding in itself?

6 Escaping from Model-land

There are at least two ways to escape from model-land. The first where possible, is to repeatedly challenge your model to make out-of-sample predictions and see how well it performs. This, of course, can only be done in weather-like tasks, so named because tomorrow's weather forecast is an excellent example where it can be done. The forecast lead time is much less than a model's typical life time.

Here we can not only understand in broad terms the usefulness of the forecast on different timescales, but we can also quantify the degree of confidence in its success for different weather types and in different seasons. This is precisely the information needed for high-quality decision support: a probabilistic forecast, *made complete with a statement of its own limitations*. One should not be advised to attempt to use a detailed weather forecast for this date next year, even though it is in principle perfectly possible to make today's models stable on that timescale and extend the simulation. In practice we see that today's one-year weather forecast contains less information than a "physics-free" empirical model which just forecasts the historically observed distribution of temperature for the time of year (this distribution is often called the "climatology" or climatological distribution). In addition, sufficient forecast-outcome information is generated to allow calculations which will give us an understanding of where, how, or after what lead time the model is performing poorly (Smith, 2000; Smith, 2006). This may assist us to improve the model itself.

If the model structure is perfect, the forecast will reflect what might appear to be "non-stationarity" in the observation, nevertheless "stationarity" is well defined only within model land where one can take the required limits in time. If the probabilistic structure of imprecision of the observations/measurements is known precisely, the probabilistic forecasts will mirror the outcome, and, of course, we are back in model land as we assume we understand the measurement model perfectly. Expert judgement may help us to disentangle the contributions of these and other shortcomings of our end-to-end modelling. Whether one wishes to separate the imperfections in the measurement model from structural model error is a question merely of how data assimilation (Kalnay, 2003) and forecast interpretation (Broecker and Smith, 2008) are viewed within the whole of the forecasting system.

In climate-like tasks, the lead time of interest may be far far longer than the lifetime of a model (Smith and Stern, 2011). Here, the problem as stated does not allow out-of-sample evaluation given the nature of the question for which support is requested; there is simply not enough data available (forecast-outcome pairs) either to construct or to evaluate so detailed an understanding of the limitations of a model. We call these climate-like situations. Although one may never be as confident in such cases as one is in weather-like cases, there is an alternative way to escape from model-land. Using further expert judgement, informed by the realism of simulations of the past, to define the expected relationship of model with reality (Thompson et al, 2016) and critically, to be very clear on the known limitations of today's models and the likelihood of solving them in the near term, for the questions of interest.

An example: the most recent IPCC climate change assessment uses an expert judgement that there is only approximately a 2/3 chance that the actual outcome of global average temperatures in 2100 will fall into the central 90% confidence interval generated by climate models (IPCC, 2013 – see footnotes c and d to table SPM.2 on page 23 of the Summary for Policymakers). Again, this is precisely the information needed for high-quality decision support: a model-based forecast, *completed by a statement of its own limitations* (the Probability of a "Big Surprise"). Structured procedures for generating and formalising expert judgements of this kind are available (e.g. Cooke, 1991).

It is worth noting here that presenting model output at face value as if it were a prediction of the real-world, or interpreting simulation frequencies as real-world probabilities, is equivalent to making an implicit expert judgement that the model structure is perfect. The IPCC does not

make this claim. The approach taken by economic modellers differs from that of climate modellers in many interesting ways. We consider one here: Climate modellers tend to present fan charts of model output (technically this is "guidance", not forecast), these simulations are then processed to yield a probability forecast, say, by shifting 30% of the probability mass from inside to outside the range of the simulation model results. In physical simulation, the fan charts are not real-world probabilities, but interpreted with additional expert judgement to inform expert probabilities of real-world outcomes. In Economics, the fan charts are the probabilities. Consider Figure 3. The original caption discusses what the probabilities are conditioned on; the report states clearly that the results of the report are believed to be robust under all Brexit scenarios. At other times specific "Known Neglecteds" are stated but not addressed as, for example, regarding the impact of Greece leaving the Eurozone.

Misplaced claims of near perfection can often be shown to be false. Indeed in-sample tests have limited power when claiming a model is adequate, but they might easily establish a high level of confidence (evidence) that the model is inadequate for a particular economic purpose. In that case unqualified presentation of model output results in illusionary accuracy.[1]

The IPCC reassignment of probability mass into the tails of the distribution referred to above is no small correction, but a first order change (of the order of tens of percentage points of probability mass) from one set of model-outcomes to another range of real-world outcomes. This is likely to have a nontrivial impact on decisions. This is an expert judgement arrived at by the expert lead authors of the IPCC chapter, themselves scientists who worked on the underlying models and simulations which went into the generation of the 90% confidence interval. But if many thousands of work-hours have gone into refining the dynamical models and statistical techniques which produced the first interval, quantifying and reducing the model-land uncertainty, their valiant efforts are *known to be* swamped by the uncertainty which exists in the abyss between model-land and the real world. A bridge to escape is sorely needed here.

In neither route to escape from model-land do we work to indefinitely increase the complexity of models. Where out-of-sample predictions can be tested, they will reveal whether each model development does or does not lead to more informative outputs. Where we rely more on expert judgement, it is likely that models with not-too-much complexity will be the most intuitive and informative, and reflect their own limitations most clearly. Even where we cannot test long-range model-based predictions, we can test how well our model can reflect (shadow) the past, and learn the phenomena with which they cope most poorly. This informs judgement as to how far in the future a given model is likely to be relevant to the evolution of the real world.

In neither route to escape from model-land do we discard models completely: rather, we aim to use them more effectively. The choice is *not* between model-land or nothing. Instead, models and simulations are used to the furthest extent that confidence in their utility can be established, either by quantitative[2] out-of-sample performance assessment or by well-founded critical expert

---

[1] We thank a reviewer for stressing this point.

[2] As noted in the review by Arthur Petersen, expert judgement is also needed to judge that the quantitative out-of-sample performance assessment is sufficiently reliable.
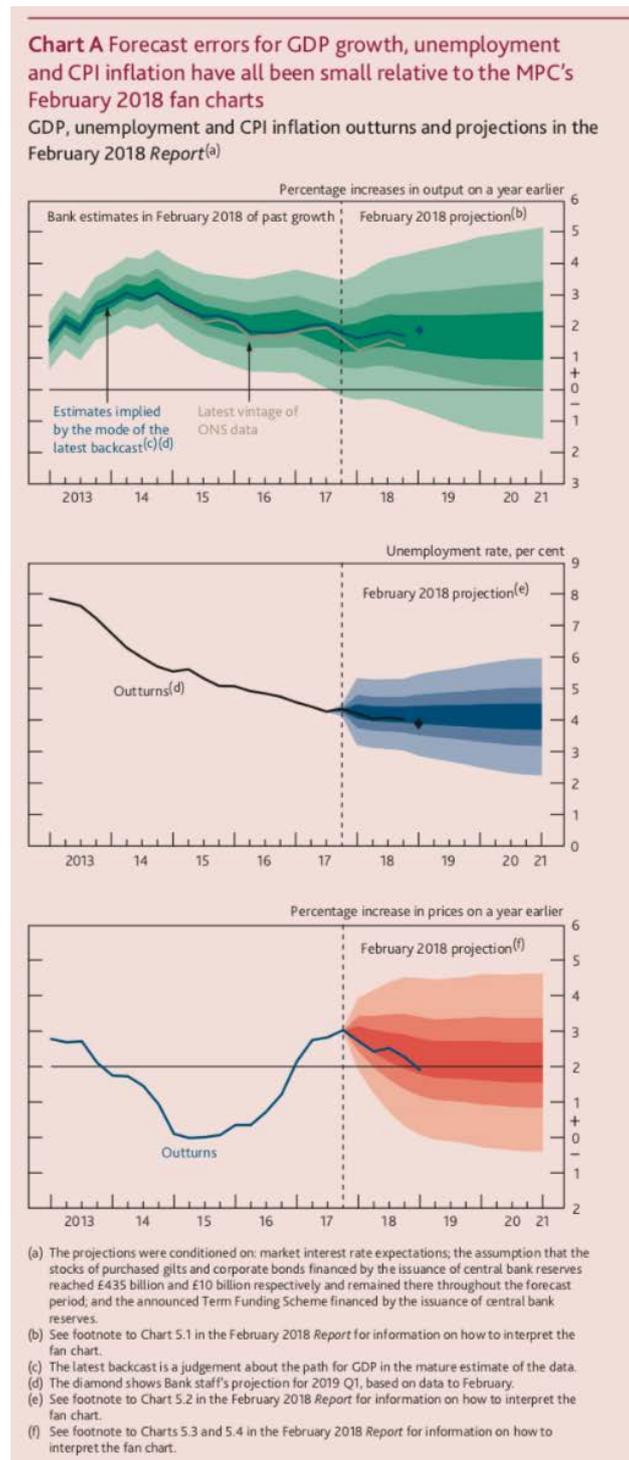
Figure 3: The latest Bank of England fan chart for GDP growth (above), unemployment rate (middle) and inflation (below). From Bank of England, Inflation Report May 2019.

judgement. A wide literature treats the use and calibration of expert judgement in such situations although there is certainly more to say about the interplay of model development and the parallel development of one's own expert judgement; and indeed more to say about the expert judgements involved in the initial construction and calibration of the model. The significant difference between the manner in which economists and physical scientists treat imprecision, and their different manner of moving between model-land and the real-world suggests the two groups might benefit from frank interaction.

More generally, letting go of the phantastic mathematical objects and achievables of model-land can lead to more relevant information on the real world and thus better-informed decision-making. Escaping from model-land may not always be comfortable, but it is necessary if we are to make better decisions.

References

Bank of England (2019). May 2019 inflation report.
https://www.bankofengland.co.uk/inflation-report/2019/may-2019

Berger, J.O., and Smith, L.A. (2018). On the statistical formalism of uncertainty quantification, *Annual Reviews of Statistics and its Application*, 6: 433–460.
https://www.annualreviews.org/doi/abs/10.1146/annurev-statistics-030718-105232

Beven, K., Buytaert, W., and Smith, L.A. (2012). On virtual observatories and modelled realities (or why discharge must be treated as a virtual variable). *Hydrological Processes*, 26(12): 1905–1908.
https://onlinelibrary.wiley.com/doi/abs/10.1002/hyp.9261

Beven, K.J., and Lane, S. (2019). Invalidation of models and fitness-for-purpose: a rejectionist approach. In Beisbart, C. and N.J. Saam (eds.), *Computer simulation validation. Fundamental concepts, methodological frameworks, and philosophical perspectives*. Springer, Cham.

Beven, K.J. (2019b). Towards a new paradigm for testing models as hypotheses in the inexact sciences. *Proceedings of the Royal Society A*, 475(2224): 20180862. https://doi.org/10.1098/rspa.2018.0862

Bröcker, J., and Smith, L.A. (2008). From ensemble forecasts to predictive distribution functions. *Tellus A*, 60(4): 663–678. https://onlinelibrary.wiley.com/doi/10.1111/j.1600-0870.2008.00333.x

Cooke, R.M. (1991). *Experts in uncertainty; opinion and subjective probability in science.* Oxford University Press, New York, Oxford.

Frigg, R., Smith, L.A., and Stainforth, D.A. (2015). An assessment of the foundational assumptions in high-resolution climate projections: the case of UKCP09. *Synthese*, 192(12): 3979–4008. https://link.springer.com/article/10.1007/s11229-015-0739-8

Good, I.J. (1959). Kinds of probability. *Science*, 129(3347): 443–447. https://www.jstor.org/stable/1757847

IPCC (2013). Summary for policymakers. In Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P.M. Midgley (eds.). *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge University Press, Cambridge, UK and New York, NY, USA.

Judd, K., Reynolds, C.A., Rosmand, T.E., and Smith, L.A. (2008). The geometry of model error. *Journal of Atmospheric Sciences*, 65(6): 1749–1772.

Judd, K., and Smith, L.A. (2001). Indistinguishable states I. Perfect model scenario. *Physica D*, 151(2–4): 125–141. https://www.sciencedirect.com/science/article/pii/S0167278901002251

Judd, K., and Smith, L.A. (2004). Indistinguishable states II: the imperfect model scenario. *Physica D*, 196(3–4): 224–242. https://doi.org/10.1016/j.physd.2004.03.020

Kalnay, E. (2003). Atmospheric modeling, data assimilation and predictability. Cambridge University Press.

Lorenz, E.N. (1963). Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*. 20(2): 130–141. https://journals.ametsoc.org/doi/pdf/10.1175/1520-0469%281963%29020%3C0130%3ADNF%3E2.0.CO%3B2

Mayo, D.G. (1996). *Error and the growth of experimental knowledge (science and its conceptual foundations series).* University of Chicago Press.

Parker, W. (2009). Confirmation and adequacy-for-purpose in climate modelling. *Proceedings of the Aristotelian Society, Supplementary Volume,* 83(1): 233–249. https://www.jstor.org/stable/20619137

Parker, W. (2019). Model evaluation: an adequacy for purpose view. *Philosophy of Science*. Forthcoming

Petersen, A.C. (2012). *Simulating nature: a philosophical study of computer-simulation uncertainties and their role in climate science and policy advice*. Chapman and Hall/CRC.

Smale, S (1966). Structurally stable systems are not dense. *American Journal of Mathematics*, 88(2): 491–496.

Smith, L.A. (1995). Accountability and error in ensemble forecasting. In 1995 ECMWF Seminar on Predictability. Vol. 1, 351–368. ECMWF, Reading.

Smith, L.A. (2000). Disentangling uncertainty and error: on the predictability of nonlinear systems. In Mees, A.I. (ed.), Nonlinear dynamics and statistics. Boston: Birkhauser.

Smith, L.A. (2002). What might we learn from climate forecasts? *Proceedings of the National Academy of Sciences of the United States of America*, 4(99): 2487–2492. https://www.pnas.org/content/99/suppl_1/2487

Smith, L.A. (2006). Predictability past predictability present. In Palmer T., and R. Hagedorn (eds.), *Predictability of weather and climate*. Cambridge University Press, Cambridge, UK.

Smith, L.A. (2007). *Chaos: a very short introduction*. Oxford University Press, Oxford.

Smith, L.A. (2016). Integrating information, misinformation and desire: improved weather-risk management for the energy sector. In Aston, P.J., A.J. Mullholland, and K.M.M. Tant (eds.), *UK success stories in industrial mathematics*. Springer.

Smith, L.A., and Petersen, A.C. (2014). Variations on reliability: connecting climate predictions to climate policy. In Boumans, M., G. Hon, and A.C. Petersen (eds.), *Error and uncertainty in scientific practice*. Pickering & Chatto, London.

Smith, L.A., and Stern, N. (2011). Uncertainty in science and its role in climate policy. *Philosophical Transactions of the Royal Society A,* 369(1956): 4818–4841.
https://doi.org/10.1098/rsta.2011.0149

Thompson, E. (2013). Modelling North Atlantic storms in a changing climate. PhD thesis, Imperial College, London.

Thompson, E., Frigg, R., and Helgeson, C. (2016). Expert judgment for climate change adaptation. *Philosophy of Science* 83(5):1110–1121.
https://www.journals.uchicago.edu/doi/abs/10.1086/687942

Thompson E.L., and Smith L.A. (2019). Informing anticipatory humanitarian action: a framework for using forecasts effectively. *In preparation.*

Tuckett, D. (2011). *Minding the markets: an emotional finance view of financial instability*. Springer.

Tuckett, D., and Nikolic, M. (2017). The role of conviction and narrative in decision-making under radical uncertainty. *Theory & Psychology*, 27(4): 501–523.
https://journals.sagepub.com/doi/10.1177/0959354317713158

Tuckett, D., and Taffler, R. (2008). Phantastic objects and the financial market's sense of reality: A psychoanalytic contribution to the understanding of stock market instability. *The International Journal of Psychoanalysis*, 89(2): 389–412.
https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1745-8315.2008.00040.x

Whitehead, A.N. (1925). Science and the modern world: Lowell lectures, 1925. New American Library, (1956).

Please note:

You are most sincerely encouraged to participate in the open assessment of this article. You can do so by either recommending the article or by posting your comments.

Please go to:

http://dx.doi.org/10.5018/economics-ejournal.ja.2019-40

The Editor