

Keith Beven, Lancaster University, UK - Comment

March 14, 2019 - 11:20

I have read the paper through and think it is one of best expressions of the issues of using models in the face of epistemic uncertainties I have read. It should therefore be published.

The only thing missing (though not necessary in the context as it would add ...[more] ... too much length) is a discussion of how to decide whether models are informative in decision making. The authors rightly refer to out-of-sample testing – but a failure can be a result of the epistemic uncertainties and non-stationarities in the data available as well as the model structure and parameters. So in the same way that statistical theory may not necessarily apply in such cases, new ways of measuring information and added information may need to be considered that allow for the qualitative subjective judgements that will often be involved.

I can add the following references to their list that address some of the same issues (and can supply copies to the authors if required).

Beven, K J., 2016, EGU Leonardo Lecture: Facets of Hydrology - epistemic error, non-stationarity, likelihood, hypothesis testing, and communication. *Hydrol. Sci. J.* 61(9):1652-1665, DOI: 10.1080/02626667.2015.1031761

Beven, K J, 2018a, On hypothesis testing in hydrology: why falsification of models is still a really good idea, *WIRES Water*, DOI: 10.1002/wat2.1278.

Beven, K. J., 2019b, Towards a new paradigm for testing models as hypotheses in the inexact sciences, *Proceedings Royal Society A*, submitted

Beven, K J, Aspinall, W P, Bates, P D, Borgomeo, E, Goda, K, Hall, J W, Page, T, Phillips, J C, Simpson, M, Smith, P J, Wagener, T and Watson, M, 2018, Epistemic uncertainties and natural hazard risk assessment – Part 2: What should constitute good practice?, *Natural Hazards and Earth System Science*, , 18(10): 2769-2783, <https://doi.org/10.5194/nhess-18-2769-2018>

Beven, K. J. and Lane, S., 2019, Invalidation of models and fitness-for-purpose: a rejectionist approach, Chapter 5 in: Beisbart, C. & Saam, N. J. (eds.), *Computer Simulation Validation - Fundamental Concepts, Methodological Frameworks, and Philosophical Perspectives*, Cham: Springer, to appear 2019

And for an alternate view:

Nearing, G.S., Tian, Y., Gupta, H.V., Clark, M.P., Harrison, K.W. and Weijs, S.V., 2016a. A philosophical basis for hydrological uncertainty. *Hydrological Sciences Journal*, 61(9): 1666-1678.

Also – to be submitted shortly.

Keith Beven and Stuart Lane, 2019, On (in)validating environmental models. 1. Principles for formulating a Turing Test for determining when a model is fit-for purpose. *Hydrological Processes*, to be submitted

Keith Beven, Stuart Lane, Trevor Page, Ann Kretzschmar, Barry Hankin and Nick Chappell, 2019, On (in)validating environmental models. 2. Implementation of the Turing Test to modelling hydrological processes, Hydrological Processes, to be submitted

Leonard Smith - Reply

September 09, 2019 - 14:42

| Author's [Homepage](#)

Thank you for your comments and suggestions, we largely, perhaps completely agree. We would be happy to write another paper on “how to decide whether models are informative in decision making.” We agree that “good practice” differs in weather-like tasks and climate-like tasks, as you say it is a ...[more]

... question of which statistical theory applies when climate-like tasks are attempted. Traditional statistical good practice has the ASA tea shirt slogan: “Friends don’t let friends extrapolate.” Alternatives to this “don’t do it” suggestion will not have as strong a foundation here as it does in weather-like tasks, but can still lead to more informed decision making, if only in that our ignorance is made more explicit.

Anonymous - What about macroeconomics

March 14, 2019 - 16:41

This paper appears to me to offer a rather general discussion of problems of system estimation quite remote from macroeconomics (and also from biology, psychology, and sociology). It is to my mind much too general. I miss an example of a macroeconomic problem that can be tackled better by the ...[more]

... proposed approach (which is not quite clear to me). The problems raised, such as structural instability, seem to be well-known.

The authors seem to be unaware of positions that see macroeconomic models as more general and more stable than disaggregated models, that not all macroeconomic models involve optimization, and that economic structure changes all the time -- as Alfred Marshall emphasize a long time ago.

Leonard Smith - Reply

September 09, 2019 - 15:11

This paper appears to me to offer a rather general discussion of problems of system estimation quite remote from macroeconomics (and also from biology, psychology, and sociology). It is to my mind much too general. I miss an example of a macroeconomic problem that can be tackled better by the ...[more]

... ... [more] ... proposed approach (which is not quite clear to me). The problems raised, such as structural instability, seem to be well-known.

The authors seem to be unaware of positions that see macroeconomic models as more general and more stable than disaggregated models, that not all macroeconomic models involve optimization, and that economic structure changes all the time -- as Alfred Marshall emphasize a long time ago.

We find the bimodal responses to this manuscript interesting, and thank the reviewer for his/her observations and for raising several interesting points. During the meeting we noted that several of the speakers, referred to as macroeconomists, said that while they thought the phenomena of the crash (the topic of the meeting) was very interesting, it was not what they did. They then proceeded to give interesting

mathematical talks about macroeconomics.

<https://twitter.com/lynnyrdsmyth/status/1047149503357341696?s=20>

Our talk was about foreseeing events like the crash and dealing with them in real time; actual economics. To be honest there seems to be some disagreement as to where or not the dynamics of the real economy is or is not a target of macroeconomics. Others certainly share your view. That said, we (and a few others) feel our talk was relevant to the meeting and thus relevant for this journal.

We agree deeply that relevance of structural instability is long known (Smale in the early 60's), and we ourselves have been concerned about its impacts on modelling and simulation for over two decades. As we were asked to speak on the actual financial crisis, and feel actions during the crisis (and comments made afterward) make it clear that while they are long known, the implications of structural stability are not well known. As it happens, the same applies to real world simulation in biology, medicine, psychology and sociology.

I would suggest that had decision makers had a true picture of the known limitations of their models, specifically of their limitations and the "purple light" (Smith, 2016 and new text) regions before the beginning of the crisis, they would have acted differently. Inasmuch as the meeting focused on the real world, and the crisis itself, we do not understand how our contribution could be considered remote. We have added text to try and clarify these points. Thank you for pointing them out.

Arthur Petersen - The two uses of "expert judgement"

April 03, 2019 - 10:37

| Author's [Homepage](#)

This paper both welcome and timely. It very clearly makes the point, of relevance to all field that use models to inform decision-making, that one has to argue that one actually has derived decision-relevant information about the real world from one's model.

I have only a minor issue with ...[more]

... the way "expert judgement" is used in two ways in the paper, as is also the case in wider practice: my point is to signal the fact that this raises confusion. Like in the IPCC, the phrase "expert judgement" is used BOTH to refer to a result obtained from applying an expert elicitation technique instead of being obtained directly from observation or a model (see, e.g., the IPCC uncertainty guidance, the audience being experts) AND to indicate that all statements that experts provide to decision-makers are expert judgements, whether they derive from observations, models, or expert elicitation (e.g., the Summaries for Policymakers, the audience being decision-makers and their advisers).

On page 8, it is stated that "models and simulations are used to the furthest extent that confidence in their utility can be established, either by quantitative out-of-sample performance assessment or by well-founded critical expert judgement." This may lead the uninitiated reader to forget that experts will still need to judge that the quantitative out-of-sample performance assessment is sufficiently reliable.

Anonymous - AP

August 28, 2019 - 17:05

We have added a footnote to make this distinction clear. Thank you.

Anonymous - Referee Report

April 15, 2019 - 09:27

see attached file

Anonymous - Reply

September 09, 2019 - 14:50

We thank the reviewer for an interesting and useful review. We have added examples of the BoE fan charts, and feel that there is much more to be said on the differences in the way BoE, IPCC and typical weather “fan charts” are constructed by practitioners, and what they are ...[more]

... interpreted as representing.

We agree that “What the authors stress should indeed be well-known to any professional user of scientific models.” That said, in many fields it is not the case.

We also agree with the reviewer that “it would be useful to have more discussion of ‘model-land’ problems in economics.” We have added additional text along the lines the reviewer has suggested, and hope the current manuscript may stimulate additional discussion.

Marcus Miller, University of Warwick, UK - Comment

May 10, 2019 - 08:49

see attached file

Anonymous - Reply

September 09, 2019 - 14:56

- - -

We thank the reviewer for his helpful comments and for giving us the chance to mention Whitehead’s Fallacy, which was cut (from the talk) due to time restrictions. Given the Whitehead was criticising the negative effect that Newtonian physics had on science, I always find it odd that ...[more]

... modellers take offence when compared to Newton; we have added a reference.

We have also added as examples the fan charts of the BoE and a short discussion. I believe an extended discussion of the differences between geophysical fan charts and the BoE’s fan charts would clarify some of the challenges the discussion between economists and physical scientists face.

We value your discussion of DSGE models and their impact, but do not have space to add this topic to our paper. That said, we take to heart your point that there may be other ways to escape from model-land, or perhaps revolutionary advances in understanding (note Kelvin’s original comments supporting a young earth) that do not require us to venture so deeply into model-land. We have added brief text pointing to these two suggestions. Thank you very much for your comments.