Hydrol. Earth Syst. Sci. Discuss., 12, C1671–C1678, 2015
www.hydrol-earth-syst-sci-discuss.net/12/C1671/2015/

Hydrology and Earth System Sciences
Open Access
Discussions

# *Interactive comment on* "Hydrological model parameter dimensionality is a weak measure of prediction uncertainty" *by* S. Pande et al.

**Anonymous Referee #1**

Received and published: 22 May 2015

This paper has been revised in response to earlier review comments. I have not seen those comments - nor the original draft of the paper.

I appreciate the efforts of the first (and other) author(s) to address the issue of model complexity and its relationship with parameter uncertainty and dimensionality. Th authors argue that model complexity is not synonymous to parameter dimensionallity. Parsimonious models with just a handful of parameters can have a complexity similar to models with many more parameters - depending on the assumed parameter values and their underlying uncertainty. The authors are trying to build a theoretical framework against which this can be measured, analyzed, and evaluated.

Evaluation: This paper is certainly not a mainstream paper. Many of the ideas pre-

sented herein deviate considerably from the large majority of papers published in the literature on this topic. I appreciate the innovation the authors are trying to bring to this subject - and that they are do so using appropriate theoretical (mathematical) rigor. Nevertheless, I am not persuaded by the methodology that is presented nor the arguments made - in large part because of writing and presentation. In my view the methodology that is presented is not intelligible for a sufficient readership to appreciate this work. What is more, I believe that many sentences in the paper are unnecessary difficult to follow or in fact even confusing. This really downplays the potential impact of this work - and perhaps more importantly introduces concerns. In summary, I feel the paper is not ready yet for in-depth review and would recommend a major rework of the manuscript - with particular . I will list my concerns below.

1. The manuscript is full of sentences that in my view are very unclear - or unnecessarily confusing. I list a view below.

a) Page 3: "Should not this uncertainty in assessing structure deficiency depend on the class of model structures which are used to assess deficiencies?" What is meant here?

b) Page 3: "The characteristics of uncertainty in system representation can then identify the consequences of ill-conditioned model selection problem and hence define ill-conditioned model selection."

c) Page 3: "Meanwhile, instability refers to inconsistency in process representation as more information on the underlying processes is available."

d) Page 5: "..stable system representation..." –> what does this mean? How can I assess whether my model is a stable system representation?

e) Page 5: "Often models with low parameter dimensionality (i.e. less number of parameters) are considered less complex and hence are associated with low prediction uncertainty." I disagree with the latter part of this statement. Models that are simple might not fit the data very well and hence have significant uncertainty - the uncertainty is not made up of parameter uncertainty but in large part of structural uncertainty. More complex models might better fit the data - yet their parameter uncertainty might be larger (structural uncertainty - that is model error smaller). From a statistical point of view their total uncertainty is the same - if the goal is to describe statistically the data.

f) Page 5: "...unstable model representation..." –> Antonym of what is used previously - but again this wording is new to me. What does stable/unstable refer to? A model that is numerically unstable (I know this is not what the authors are referring to) - a model that does not describe the data very well? Or?

g) Figure 1: Caption not repeated here. o_1 and p_1 are observed and predicted value - how can this be of size "N" (one would a vector then not just a scalar o_1 and p_1). What is mode output space? We will get "N" distance between the observed and simulated data. We can write this in vector notation and plot this? How does "B" (vector notation) then define the overall distance? I find this presentation to be very confusing - and because this is the underlying theory the authors are developing it is hard to assess what is going on. I would strongly recommend to keep the presentation simple - and clarify the figures.

h) Figure 2: Same issues - "also model structure output space". Why not just "model output space". This Figure generalizes Figure 1 to multiple models. How can the model structure be defined as a union of two models? Each model has its own structure and output space - the union of two models would constitute a part of the feasible model space?

i) Figure 3: Unclear as well. So model 2 can be considered to be part of the output space of model 1. OK. Thus model 2 has a large variation in the output space for given input data - OK - why would model 2 then have a higher instability in system representation? I would think this is model 1 given that this model exhibits a larger uncertainty in the output space?

j) Figure 5: Unclear (inset on right side too small. Now model 1 has highest instability? I find the wording instability very confusing. Has a negative connotation - and as I am trying to follow the logic here I am confused about the logic.

k) Page 7: "We define instability of a given model by the variability in the differences between 10 its outputs over two different realizations of data." This is confusing - two different realizations? In practice we only have one realization of the data. Or do you mean we have two or more observations of the same type but at different locations in space or time?

l) Page 7: "A model then is more unstable if it tends to have larger differences between model simulations for any given pair of data realizations. Such a definition is sufficient to encapsulate the notion of inconsistency in process representation by a model." I fail to understand this logic. I am just missing pieces here to understand the reasoning of the authors. If a model exhibits a large uncertainty in the output space, that is given some input data and prior parameter space, the model simulates a large variability in the output space - is this then equivalent to inconsistency in process representation?

Another issues that emerges here is that the output space of a model does not say much about process presentation. a model resolves many different processes (different model components – equations if you will) - the collection of which produces a model output. The focus on model output as vehicle for analysis complicates things further - because what is analyzed is a summary term of different processes. I feel that more progress can be made if authors focus on outputs that relate directly to individual processes in the model. I feel that this is a more realistic assessment of the strengths/stability/consistency of the model - rather than evaluating the range of the outputs.

The authors continue on Page 7 with "This is because it is quite likely that a highly unstable model that appears to be a suitable representation of the underlying system on one piece of information may not be a suitable representation on another or more pieces of information." Exactly - this is a common problem - a model might be considered appropriate when evaluated against one type of data - but completely useless when asked to simulate/predict another variable. I guess my problem is with the unstable formulation the authors. I do not view a large range in simulated value as unstable. I might refer to it as uncertain. I think authors can be much clearer in their reasoning if they adopt a more logical jargon.

m) From Page 7 on - the material is excessively difficult to follow. Perhaps the earlier part is still intelligible - the later part (Page 8 middle forward) is hard to follow.

n) Page 9: structure output space?

o) Page 10:"Figure 4a also demonstrates that deviation in performance of system representations from model structure ˆ2 is often larger than ˆ1, to the extent that Pr(j kAkôĂĂĂkCk j>) is larger for nearly all >0." This paragraph is trying to say that models with more parameters (more complex) generally have a larger uncertainty in the simulated output - and hence exhibit a large simulation uncertainty?

p) Page 11: "Figs. 4 and 6 suggest that controlling for the complexity in a model selection exercise may stabilize the representation of underlying processes. This is akin to “correcting” the ill-posedness (Vapnik, 1982) of model selection problem by constraining the complexity of the model structures used. This is equivalent to regularized model 20 selection problem" OK. But this analysis is based on synthetic data? What about using real-world data - would one arrive at a similar conclusion? But why would one need to constrain the complexity for model selection? The marginal likelihood (Bayesian perspective) will pick the simplest model that still explains the data - so if complexity is inappropriate then this model will not get selected - or if the model is too complex then the integral of prior and likelihood will provide values for p(D) that are smaller than those derived for a simpler model.

Another emerging issue here: If a model is indeed very complex - has many parameters but the parameters their prior uncertainty appears relatively small. Then the model

might be better constrained (more stable in wording of authors) then a model with far fewer parameters but that exhibit a much larger prior uncertainty. All this is taken care of in Bayesian model selection - if looked beyond simple criteria such as the AIC or BIC. So why not compare the arguments made here against full (numerical) integration of the prior and likelihood? This might make arguments more compelling. Because one can view p(D) as a measure of complexity as well. One that integrates quality of fit with uncertainty.

q) Page 12: dimensions [1/T]ˆ3 - rather awkward -> each recession parameter has unit 1/T -> not [1/T]ˆ3

r) Page 18: "First we note that E[kBk] is the expected dierence in a model’s simulations for two realizations of observations." –> unclear. What is meant by two realization of observations? Which observations? Forcing data. Unnecessary difficult to follow. Again, I only highlight a few of these places - many other sentences can be found that are confusing at best.

s) Equation 4 - 11: Here things become confusing. Equations are provided but their relevance remains unclear - again first we need to understand what the expectation of B refers to? Two different realizations of input data? Precipitation data? Or all forcings combined? Dedfinition 1: Am I right that this is the difference between any simulated data point and the mean of these data points for a parameterization alpha? Why not word this - I highly recommend to explain each of the equations this way - and also to illustrate their calculation graphically in a plot. Just plot some data - calculate the mean of the data simulated by a model and then introduce Definition 1. Much easier to follow. Do so with each of the equations/definitions. Then it is much easier to follow for a reader.

t) Definition 2 is unclear. a model parameterized by alpha by gamma tilde? What is gamma and what is gamma tilde? Two different parameterizations? This is where things become rather unclear. Either use graphics - ideally combined with simple explanation in words. Do not hide behind equations - this will make things unnecessarily complicated.

u) How does the Markov Lemma come into play here? I miss the connection. What is X >= 0?

v) Page 24: "By doing so we test whether the ordering in terms of its complexity of various model structure set-ups changes with different data sets. Insensitivity of the ordering of structure complexities to the data sets used for input forcings is crucial for any robust statement about the 5 role of parameter magnitudes in determining model complexity" Difficult to follow. Many readers will have lost your arguments here - nor understand the underlying theory. Many elements need to be clarified before one can judge competence, relevance and importance.

2. How would you evaluate the complexity of an artificial neural network? If you add such model to the analysis - would the ANN then come out as most complex? I need to see more the results of more than two models to evaluate the findings.

3. The present methodology requires an ensemble of forcing data - to evaluate the range of simulated output for a given parameterization. Is sampling of the prior parameter space not sufficient? Because fundamentally the approach that is presented herein differs from Bayesian model selection in that multiple inputs are considered.

4. Going back to my earlier comment. The authors evaluate complexity by looking at the output space of the model. This is one measure of complexity - the number of parameters used can be another measure of complexity - depending on their ranges as well. Thus one can define different measures of complexity - nevertheless - I believe the authors should benchmark their findings against common complexity criteria - ideally numerical integration of the posterior distribution (marginal likelihood).

5. The definitions the authors provide use a L1 norm for the distance between the simulated and mean data. How does their analysis hold if a different norm was used?

Why use a L1 norm? Why not generalize this to any norm? L1-L2-L3-Linf.

6. I fear that focus on the model output space gives a rather limited view of complexity. I believe that author should focus on individual components of their model - and better recognize that if the goal is process understanding and analysis - that model output is not the way to go. Much better is to investigate specific metrics that are sensitive only to given components of the model. Such implementation would enhance significantly the impact of this paper. One can define summary metrics of the data and then use those to quantify complexity.

7. What is the unit of complexity the authors are proposing? Something that should be clarified. Also is this metric relative or absolute?

8. Appendix B: How does this inequality hold for N data points (N>2)?

9. Paper has many typos.

In summary - the authors are trying to do something interesting. Yet, the presentation requires much attention before the paper can be judged to making a significant contribution. I find the wording awkward (a few examples discussed above) - and the derivation of the equations rather difficult to follow. Authors should use appropriate/simple wording that does not confuse reader. Equations that are presented should be discussed in words first - then illustrated with a schematic - before proceeding to next equation. This will much enhance readability and understanding. Also this is a prerequisite for full review.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 12, 3945, 2015.