Hydrol. Earth Syst. Sci. Discuss., 12, C1814–C1825, 2015
www.hydrol-earth-syst-sci-discuss.net/12/C1814/2015/

Hydrology and Earth System Sciences
Discussions

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

Anonymous Referee #2

Received and published: 27 May 2015

Overview:

This is a difficult article to review for two reasons. First, it is incredibly poorly written. Second, there is no theoretical argument in support of the proposition, which means that we have no idea whether it is a worthwhile idea.

I strongly suspect that the core argument is theoretically unsupportable (i.e., that it is simply a bad idea), but I am unwilling to do the authors' job and try to derive a proof of convergence or lack thereof. There are a few simple errors in this paper such that, at least at first glance, appear to be sufficient to invalidate the central premise. Because the authors present no meaningful justification for their idea, I won't bother to formally explore the extent to which these errors might impact convergence. Like I said, it's

not my job to present a convergence proof – if the authors were to present one, then I would be happy to review it. Since all that is offered in this manuscript is a description of some vague intuition, I will respond in kind, and it is my assessment (detailed below) that this intuition is incorrect. This superficial review is, in my opinion, more than the paper deserves, as it should be rejected simply on the grounds that it lacks any sort of substantive argument.

Related to the writing, one thing that might help would be for the authors to choose one single definition for each term, to state these definitions before the term is used in any other context, and then to not change the definition for the rest of the paper. As is, terms are used before they are defined, terms are not defined at all (e.g., "ill-conditioned", even though it is stated that we will be given a definition), and terms are explicitly re-defined partway through the manuscript (e.g., "complexity", "model structure", "model"). As is, the paper reads like a bunch of buzzwords strung together into (sometimes) grammatically correct sentences. And these words are often used in directly conflicting ways, even in the same paragraph. While reading the introduction, I wondered with some amount of sincerity whether this was submitted as some sort of test of the journal's peer review system (e.g., http://pdos.csail.mit.edu/scigen/).

Related to the content of the arguments, the authors need to sit down and think about what they want to say. Explicitly lay out some desiderata for finite inductive inference, and then prove that whatever strategy they want to propose facilitates these desiderata. Please be aware that it is insufficient to consider uncertainty as related only to the probability distribution over system response obtained by integrating (or sampling, taking the mode of, etc.) an inference posterior – we also must consider the fact that that posterior is necessarily derived from a degenerate prior, and therefore we (probably) never test a "true" model. That is, there is more to uncertainty than simply the inability to discriminate between some set of candidate models. This means that I may very well not want inference to result in a strong preference for a single model after a finite (or even infinite) amount of data (i.e., equifinality qua equifinality is not undesirable).

And I may not want inference on different data sets with different information content to result in preference for the same model (i.e., instability qua instability is not undesirable). If all models in the inference class are wrong, then we may be best off with some combination thereof, and therefore reducing what the authors here call “uncertainty” (different from a meaningful definition thereof) in the model selection process is not a desirable objective.

It is important to point out that putting a constraint on E||B|| effectively favors models with a smaller dynamic range. E||B|| is minimized when the function is a constant and generally favors models with no sensitivity to their inputs. My intuition is that this type of regularization will *increase* uncertainty (defined somewhat more comprehensively). That is, static models of dynamic systems and models that dampen the response to their inputs are incorrect, and therefore even if we were to perform inference that uniformly converged to such a model, we would not decrease overall uncertainty – we would simply decrease so-called “aleatory uncertainty” (I hate that word) at the expense of increasing epistemic uncertainty.

A single empirical analysis is presented, however it is trivial to show that this analysis is a special case. The authors make two false generalizations about nested models (see below), however even without these errors there is no way to know whether the empirical results here can be generalized. The example given represents a particularly simple type of nested structure, and so there is no chance for a different model structure to contribute different information during different periods (see, for example, [1], [2] for senses in which we may want different models for different data).

Another thing that should be mentioned is the relationship between this idea and existing related theories. The objective of the paper is to define the concept of model complexity in a way that is measurable in model output space. The purpose is to set up a complexity-based regularization constraint on inductive inference. Of course, there already exists a definition of model complexity[3] that is used to regularize the type of inductive inference[4] known as universal induction[5]. This theory is used in hy-

drology[6]. Similarly, the ||B|| value, which is used here as the basis for a measure of “complexity”, is actually a primitive measure of sensitivity[7], not complexity. The authors are indeed correct that there is a direct relationship between the sensitivity of a function to its inputs and whether or not inverse problems related to that function are well-conditioned[8]. No criticism is offered either (a) of the theory that the paper purports to replace (complexity -> universal induction), nor (b) of the theory that it actually attempts to replace (sensitivity -> condition numbers).

Recommendation:

This manuscript is severely deficient in almost every respect: presentation, literature review, (apparent) validity of ideas, and supporting arguments. I recommend rejecting it without option for revision.

What is needed to turn this into a worthwhile contribution to the literature is not a matter of improving or augmenting what already exists, but rather a matter of constructing an entirely new argument to form the central pillar of the paper. If the authors decide to invest the time into exploring this idea in any type of rigorous or robust manner, and feel that after this type of exercise it turns out to have some validity, then I am happy to revisit.

Specific Comments: A brief collection of problems that I noticed while reading the paper. I stopped writing them down when they became too numerous. Most of this is covered more generally in the Overview section above.

Regarding the title: Does anyone argue that parameter dimensionality is a measure of prediction uncertainty? I’m not sure how this argument would proceed – the idea is prima facie absurd. Of course the two can be related, and the former can even contribute to the latter (via non-monotonicity), but to assert that one “measures” the other

is untrue. The authors claim that this is a common assertion in line 20 of page 3949, but no examples or references are given. I'm genuinely curious what this argument looks like.

———————- "This paper shows that instability of hydrological system representation in response to different pieces of information and associated prediction uncertainty is a function of model complexity"

No, it doesn't show that. It shows that instability in a performance metric is bounded by a sum that includes what the authors call "a measure of complexity" (i.e., ||B||). At best, the argument is that so-called "complexity" contributes (functionally) to a bound on the *potential for* instability. Never is it argued that instability is "a function of" ||B||, E||B||, or Φ. It is trivial to find counter-examples – for example a model of a projectile built on Newton's laws will have less variability in presumably any performance measure than, say, a linear regression, even though the latter may have very small variability in its output (i.e., small regression coefficients and small sigma).

To extrapolate on this, I wonder why we should perform inference over a prior that favors the linear regression. The objective of inductive inference is to use available information to condition our current knowledge. So what is the a priori knowledge that favors a model with smaller dynamic range? Solomonoff bases his a priori preference against complexity on Occham's principle, and then proves that convergence is guaranteed[10]. Here the concept of "instability" is used as the justification, and the rational given for this a priori preference is that instability contributes to "uncertainty". Convergence criteria are not offered, let alone any demonstration thereof.

The problem with this argument is trivial – instability (actually equifinality, which is the symptom of instability in the context of induction) contributes to uncertainty about *choosing among an a priori set of possibilities*. Really, the goal should be to find the best representation of the system (not just the best among a particular class). Since this is generally impossible in practice, we must recognize that there is always uncer-

tainty about the extent to which the best model from a particular class represents the system. So simply increasing our ability to pick a single choice from some arbitrary a priori set does not seem to be a worthwhile objective, and can easily lead to increased uncertainty.

———————- "After demonstrating the connection between unstable model representation and model complexity. . ."

What is an "unstable model representation"? Is a model represented, or does the model represent the system? If the latter (which is what I assume is meant), how can a model's representation of a system be "unstable"? As written, this sentence means that there is connection between the location of poles in the complex plane and something called "model complexity", which is not defined. This is obviously not what is intended.

———————- "Reconciling models with observations is often ill-conditioned"

The sentence is missing a subject – try: "The problem of reconciling models with observations is often ill-conditioned."

———————- "complementary pieces of information to select a better constrained model"

What is a "better constrained model"? I assume you mean that the multiple pieces of information are used to better constrain the inference procedure, however it is possible that you really mean that you will select a constrained probabilistic model. Can you elaborate?

———————- "But is the issue of ill-conditionness limited to the discourse of the number of measures used?"

The answer to this question is well known. Obviously, the extent to which an inverse problem is well conditioned depends on several things including the form of the prior (or at least the sensitivity and injectivity of the function to be inverted, if we prefer to conceptualize the problem that way), and the form of the likelihood (or optimization

objectives).

More important is the fact that we don't have a very clear idea here what you mean by "conditionedness". In the classic sense of the word, the problem is formulated around some function that must be inverted. In the more general context of inductive inference, we may want to consider including (as is done here) inference in function space. It would be good to put the content of the second paragraph before the first, as it introduces the content necessary to understand the discussion of "multiple measures", etc. More to the point, you state that "a definition of ill-conditionedness . . . [is] needed" but you don't actually offer one. Define first, discuss second.

————– "what happens when an ill-conditioned model is selected to represent the underlying hydrological system"

How can a model be ill-conditioned? Previously you discussed that "reconciling models with observations" is ill conditioned, now we are assigning that property to the model itself.

————– "Since it fails to exploit interesting information in the data, there is uncertainty in system representation."

This sentence appears to be gibberish. I'm not even going to guess. First, we don't know what these words refer to (see previous comments). Second, even if we assume that an "ill conditioned model" is a model that results in equifinal parameter distributions during calibration, then this model in no way "fails to exploit" interesting information in the data, it simply means that the information in the data does not yield a unimodal inference posterior. If we instead are talking about a model with variable performance metrics, then it is also not necessarily true that it "fails to exploit" any information in the data because we have no idea what information is in the data to begin with. Perhaps the input data simply is insufficient to predict the observed response.

————– "Should not this uncertainty in assessing structure deficiency depend on

the class of model structures which are used to assess deficiencies?"

First, what is a "class of model structures"? Second, how is a class of model structures used to assess deficiencies? Third, deficiencies in what, exactly? We have not assessed any structure deficiency up to this point, we have only discussed whether an inverse problem is ill-conditioned.

————– "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."

So now we are back to the inference problem (over models) being ill-conditioned, whereas earlier in this paragraph it was the model. Second, you have made no argument that "characteristics of uncertainty" can identify anything. I'm sorry, but this whole paragraph (and the one preceding) is gibberish. There is no discernable content here; it is just a bunch of buzzwords strung together to make semi-coherent sentences.

————– "We characterize uncertainty in hydrologic system representation as composed of non-uniqueness and instability in system representation."

Then you would be wrong. Uncertainty includes many things, one of which is the fact that any inference posterior will have finite entropy after a finite number of experiments. This fact results both in what you call equifinality and also inconsistency. In particular, both of these things result if the posterior is generally multi-modal. There are, however, other sources of uncertainty. One of which is that we may not test an accurate model, so even if we have low-entropy unimodal posteriors we may still not know whether we have an accurate system representation. Even if we had an infinite number of experiments. Also, even if we did happen to test a "true" model, we could never know that we did due to Hume's problem. Anyway, this "characterization" is false.

————– "different measures of closeness, which when orthogonal, provide complementary pieces of information to select a better constrained model (Sivapalan et al.,

2003; Winsemius et al., 2006)."

I wonder what orthogonal means here? Unless I missed something Sivapalan 2003 says nothing about orthogonal measures of closeness. Both of these references actually talk about using multiple types of data for inference, which is a particular kind of "more information", and is not the same thing as "multiple measures".

————— "Instability can then be rephrased as a changing set of good system representations (models) . . . when different data sets . . . are used."

What is a "good" system representation? Why are we talking about value judgments ("good") when we could be talking about inference posteriors? How is this definition different than before? More importantly, why is this undesirable, and why will it help anything for us to force this not to happen?

————— "This paper demonstrates that . . . instability of a given model over realizations of data can be understood and controlled by what we term as model output space."

No, it doesn't. First of all, you argue that instability can be bounded by a measure on model output space. Second, you argue that inference on models that demonstrate instability can be "controlled" using measures of model output space. But you don't actually show that this is true at all, you merely assert that it is so. The paper is a set of unjustified (and I imagine unjustifiable) assertions with minimal supporting evidence.

————— "We call the extent of model output space as the measure of complexity since its regularization would lead to a stabler representation of underlying processes"

Sure, if a model has no variability in its response to inputs, the limiting magnitude of the difference in values of a performance metric during different applications is minimal (bounded only by ||D||, which does not depend on the model). Yes, a constant is a stable system representation; it is not, however, one that offers any insight into the system response. That is, a model that returns a constant value for all inputs is favored

according to this philosophy because the bound on the "variability in the errors" is minimized (at the variability in the observations). I fail to see a priori why we should favor models that lack a dynamic response or any sensitivity to inputs.

————— "The deviation in performance of a model over two different information sources is bounded by ||B|| that measures how large is the model output space."

Strictly read, this sentence states ||A||-||C||$\leq$||B||.

————— "Since ˆ2 is nested within ˆ1, if p_1ˆ1 is not the same as p_1ˆ2 then p_1ˆ1 is closer to the observation o_1."

This is true only in special cases: when the additional processes that differentiate ˆ1 and ˆ2 can be parameterized so that they have no effect. If this is not possible then the stated conclusion is not guaranteed. It is very easy to imagine a situation where this is not the case. For example, if we have a Nash cascade with N and N+1 reservoirs respectively then this will not hold. Similarly, if we have a cascade of independently parameterized linear reservoirs treated with an explicit solver that takes the form (e.g., HyMod): x_(i,t)=(x_(i,t-1)-k_(i-1) x_(i-1,t))/(1+k_i), we must allow k_(N+1) to approach infinity for pˆ1=pˆ2 (i.e., there is no value of k_(N+1) for which this is true, and at best ˆ2 can only approach the accuracy of the less complex ˆ1 when ˆ1 is the true model). The point is that the concept of nested structures is not very illuminative.

————— "We now note that the model output of any hydrological model is continuous in its parameters."

What? No. Of infinite-capacity linear bucket models, yes. Of anything with a threshold or a classification (e.g., vegetation, soil), no. It is simply false to claim that Φ(y(α)) is continuous in α. Moreover, if we are here using the y(α) definition of a model, then this is an absurd statement, since changing parameters α changes the structure, and this change may be discrete (e.g., number of linear reservoirs, choice of infiltration equation, etc.). Also, the authors are correct that a model necessarily includes the

distributions over parameters (i.e.., a determined parameter θ is actually a Dirac distribution at x=θ). The problem with this discussion is that it only considers uniform distributions. I am not sure that this actually affects anything, but the presentation is careless.

———————- "a definitive statement on structure complexity based on parameter ranges or parameter dimensionality, i.e. without knowing their complexity in advance, can only be made if the corresponding structures are nested."

I can't tell what the point is. If we use the definition of complexity offered here, then I can make a definitive statement about complexity on any model using the proposed algorithm. Yes, if we have a (very) special case of nested structures, then we can – perhaps – say that one is more complex than the other without running the algorithm, but so what? I'm not sure what Gupta et al (2008) says about a "continuum of model structures". There is no Renard et al (2010) in the reference list.

1. Clark, M.P., et al., Hydrological field data from a modeller's perspective: Part 2: processâĂŘbased evaluation of model hypotheses. Hydrological processes, 2011. 25(4): p. 523-543.

2. Ruddell, B.L. and P. Kumar, Ecohydrologic process networks: 2. Analysis and characterization. Water Resources Research, 2009. 45(3): p. W03420.

3. Kolmogorov, A.N., On tables of random numbers. SankhyÄĄ: The Indian Journal of Statistics, Series A, 1963. 25(4): p. 369-376.

4. Solomonoff, R.J., A formal theory of inductive inference. Part I. Information and Control, 1964. 7(1): p. 1-22.

5. Rathmanner, S. and M. Hutter, A philosophical treatise of universal induction. Entropy, 2011. 13(6): p. 1076-1136.

6. Weijs, S., N. van de Giesen, and M. Parlange, HydroZIP: How Hydrological Knowledge can Be Used to Improve Compression of Hydrological Data. Entropy, 2013. 15(4): p. 1289-1310.

7. Saltelli, A., K. Chan, and E.M. Scott, Sensitivity Analysis. 2009, Chichester, NY: Wiley.

8. http://en.wikipedia.org/wiki/Condition_number.

9. Howson, C. and P. Urbach, Scientific Reasoning: The Bayesian Approach. 1989, Chicago, IL: Open Court Publishing.

10. Hutter, M., Universal Algorithmic Intelligence. 2003, Technical Report IDSIA-01-03.

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