Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2017-739-RC1, 2018
© Author(s) 2018. This work is distributed under
the Creative Commons Attribution 4.0 License.

Hydrology and Earth System Sciences
Open Access
EGU
Discussions

Interactive comment on "On the dynamic nature of hydrological similarity" *by* Ralf Loritz et al.

S.V. Weijs (Referee)

steven.weijs@ubc.ca

Received and published: 27 March 2018

The paper presents an interesting experiment in evaluating the needed spatial complexity of hydrological models, using an information-theoretical approach to compare a "full" model with a compressed one, and testing for similarity in responses to define simpler models. Generally, the paper is well-written, with interesting thoughts put forward in the discussion. I found the idea very relevant, and think there is a large potential for further work in this direction, perhaps focusing also on predictability rather than mostly on similarity. For the current paper, I have a number of comments/questions that I think would need some clarification in the paper. The comments are itemized below. I think most of the suggestions could easily be addressed/clarified in this paper, while some points may be more suitable to be addressed in further work.

1) On line 75 it is mentioned that some of the energy is dissipated in runoff concentration, and the remainder is exported as kinetic energy of streamflow. I would say that the vast majority of the energy is dissipated and only a tiny remaining fraction exported. If I am not mistaken, a back of the envelope calculation with reasonable (conservative) numbers : average height of catchment above river = 100 m, 100% runoff ratio, and 2 m/s flow velocity, gives me that 99.8% of energy is not exported as kinetic energy by stream flow. To dissipate a smaller portion, we would have to build large constructions (dams, penstocks) to get more kinetic energy and utilize for hydropower, and that is usually mainly saving the in stream dissipation, not much of the dissipation from runoff concentration.

2) 154: I would say that in hydrology, next to resistances, there is a large influence of storages (or capacitors in an electronics analogy) in the system that would make the dynamic behaviour unique/ different between two cases with the same driver/resistance combination. Could it be that the capacitances have some relation with the resistances (e.g. through porosity and path length through soil), that explains part of the similarity?

3) 222: Could you briefly state what criteria were used for the subdivision in hillslopes? I would say that could be quite relevant to understand possible redundancy/similarity. Is the stream network based on an algorithm or on direct observation in the field? (see e.g. Mutzner et al, 2016)

4) 328: Apart from the precision of the data, another very important consideration in the number of bins chosen is how the number of bins compares to the number of data points (this is the same consideration that goes into choice of bins when presenting data in a histogram). If bins are too fine for a relatively low number of data points, this will result in values always close to log(m), where m is the number of data points. If too coarse, we lose information that was in the distribution. It would be helpful to mention the number of bins used for both variables, and the number of data points, to clarify that the two are in balance. Perhaps also some discussion is warranted about using constant bins over the year vs. rescaling bins every time step using the maximum model. As noted, this choice basically reflects the question of which we quantify the

information contained in the answers to it.

5) 334: It is not completely clear to me if/how the soil moisture probe uncertainty is used in the calculation of the bin width. Is it a relevant factor in the bin width if observations are not directly used to compare with the simulations?

6) 352: The upper bound is not only related to the number of models, but also to the number of bins used to discretize the values. Now, if I understand well, the number of bins is unbounded, but in practice, there is only a maximum number of bins occupied (would be good to know how many). Depending on the range of the bins, in order to reach the maximum entropy, the models may need to simulate unrealistically high or low values. Therefore I have some problem with the interpretation of log(N) as the maximum attainable entropy / state of zero spatial organisation.

7) 348-350: I think there is also another key issue here: predictability from other hillslopes is not the same as giving identical unit runoff values. If e.g. runoff follows the same pattern with a simple scaling factor, a time shift, or any one-to-one relation, I would say the hillslopes are redundant to some extent. Giving identical runoff values is a sufficient, but not a necessary condition for providing information that is redundant.

8) 356: but attaining this maximum would also mean the rain has to be white noise spatially.

9) 357: precision: I am not sure if this statement is accurate, see point 7.

10) 359: identical : I think spatial organization/compressibility is more related to how predictable one hillslope output is from the other, rather than how identical they are, see also point 7.

11) 442: The median of all realizations is not a compressed model, since it would need calculation results from all catchments. Please clarify whether you calculated: a) the KGE of the median of the simulation outputs from the 1000 bootstrapped models or b) the median of the KGE's calculated from each bootstrapped model simulation output

individually. From the text, it seems a) is done, but figure 3 suggests b). Please clarify. There are actually several points in the text where this is not completely clear (e.g. also 528), and I think it is relevant to know.

12) 457 -459: I don't think correspondence with KGE or any other metric should be used as an argument for using NMI, because then why not just use KGE as similarity metric? Ideally, arguments for suitability should come from some desiderata for the metric, based on what you set out to measure.

13) 465-468: I think another reason for decreasing entropy is the negative feedback of decreasing evaporation with decreasing storage, which stabilizes the system and drives it to some equilibrium. Any changes in storage tend to dampen out over time because of this mechanism.

14) 478-479: See point 6/7.

15) 596: Given points 6/7 This interpretation should be checked after comparing against the max given by the log(nr of bins).

16) 620-623: When using mutual information as a basis for clustering, it means that hillslopes outputs within one cluster are predictable from each other, not per se that they are similar. This may be another explanation for the large differences between the bootstraps.

Specific minor comments:

15 : Implies -> assumes?

144-145 either 0 or 2 commas

148: what is meant by “free water”?

184: I would suggest trying to slightly rephrase this to make it clearer of what you calculate the entropy. Maybe adding “distribution of outputs”. Shannon entropy of the ensemble sounds too vague.

188-190: this is some repetition of 110-114. I agree it is worth emphasizing, but maybe rephrase and change “state” to “reiterate”

304: Bit = abbreviated from “binary digit”

310-311: Be more specific of what the Shannon entropy is calculated. This needs to be a probability distribution.

387: Maximum mutual info depends on the minimum of two entropies rather than the joint entropy.

392: Scaling with the maximum of the 2 entropies rather than the minimum seems somewhat strange, as it is actually bounded by the minimum. On the other hand, it can still be interpreted as giving the proportion of predictable behavior in the higher entropy variable, given the lower entropy one. This is however, strongly influenced by the choice of binning, while scaling with the min(H) would be less sensitive to that. I would say it is worth discussing the choice briefly, as I think it is more conventional to use the min(H) for normalizing.

395-396: Also without the normalization, MI is symmetric. Before and after the normalization, it does not satisfy the axioms of a distance function (specifically, it does not satisfy the triangle inequality).

Figure 3: bottom panel legend: “Compressed models” plural would be better fitting description of 1st line. Vertical axis should be labeled “cumulative frequency”

503: NMI efficiency: I would recommend not to use this term (first appearance in manuscript), because efficiency suggests a positively oriented score: higher=better, like NSE. For 1-NMI, lower is better.

516: consistently -> consistent?

534: I don’t think a factor 17.5 would define as 2 orders of magnitude. You can leave out the last sentence as it is redundant.

540 tow -> two

Figure 6: It would be helpful to also indicate the KGE for the full model (area weighted mean of the hillslopes). I am not sure if this is anywhere in the paper. But that seems to be a good upper benchmark.

560: “information content of the hillslope responses” sounds too vague, too loosely defined

562: Similarity is here defined as having the same value, but many readers would associate similarity of hillslope response with moving in a similar direction / pattern. Please add some words of context to clarify this.

563: What is meant with the absolute minima and local maxima? In what?

633-634: high spatial organization -> high spatial complexity (I think the current wording is the opposite of what you want to say).

655: I think “without the necessity of normalization” is overselling the approach somewhat, because you still have to deal with two discretizations that determine how your bits of entropy in 2 variables compare to each other.

I would add in paragraph 49 some statement about the relevance of the nr. of data points in the choice of binning.

673: what is meant with “the other way around” ?

725: maybe add the exact equations used for easier reproducibility?

729: equally distant -> equidistant

Figure 8: in panel B, the fact that NMI=0.65 for a perfect relation is illustrative of the dependence on binning, that is introduced by normalizing for max(H) rather than min(H).

Reference:

Mutzner, R., Tarolli, P., Sofia, G., Parlange, M. B., and Rinaldo, A. (2016) Field study on drainage densities and rescaled width functions in a highâĂŘaltitude alpine catchment. Hydrol. Process., 30: 2138–2152. doi: 10.1002/hyp.10783.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-739, 2018.