

## ***Interactive comment on “Seasonal drought prediction for semiarid northeast Brazil: About the added value of a process-based hydrological model” by Tobias Pilz et al.***

### **Anonymous Referee #1**

Received and published: 15 September 2018

#### Summary:

This study is a very comprehensive assessment comparing a process-based model of reservoir levels to a statistical regression approach for five river basins in northeast Brazil. The authors conducted substantial analysis to provide a thorough comparison for this region.

#### Major comments:

My main comment is that there is very little information regarding the statistical model used. The authors frequently reference another paper (Delgado et al, 2017), on which this work is based, however, even if the reader has not first read Delgado et al (2017),

C1

they should be able to understand this one. Adding the equation for the regression model used, how it was designed, and how well it represented these data would greatly improve the paper. In addition, I assume that Delgado et al, 2017 was for the same location but I'm not sure that that was explicitly stated in the paper.

My second main comment is that reducing and consolidate figures to the ones that are most important and illustrative of the main points could help clarify the paper. In addition the regression tree analysis is going to be included, more information is needed. Why was this approach selected? What about the limitations of decision trees, such as the fact that they can be highly sensitive to small changes in the data or decisions about the how many nodes to allow? Were random forests, which can reduce the likelihood of overfitting, considered? Also, is it common to use regression trees for outcomes which only vary between 0 and 1 (as is done here)?

#### Comments on specific sections

Abstract: Line 13 Be careful about broad recommendations here. Better to be clear that these finds are specific to these particular models applied in this study and for this region. Whether a statistical approach would be better depends substantially on the specific approach used.

4.4 Hydrological modelling: At the bottom of P 8 there are a few things mentioned briefly that seem as though they could have large implications for the results. First, for lines 28-29, It seems significant to assume that reservoir losses were set to zero. Can you explain this choice a bit more? Second, for lines 29-30, using such a different metric for the Salgado region seems like it would have a large impact. It would be reassuring to have some sort of comparison to another region illustrating that this streamflow metric can accurately represent reservoir levels. For example, when I look at figure 3, I wonder if the discrepancy between Salgado and the other sites is due to this difference.

Metrics: Can you explain why you used KGE as a metric for the simulation performance

C2

while using BE for model parameterization? Also, where does the metric in equation 2 come from? Not sure you need to add more to the paper, but it would be interesting to see a whether different quantiles (other than the 0.3 selected by decision makers) result in very different results. (Or perhaps cite any other work that might have looked at this?)

Model performance, section 5.2: It is concerning that forecast probabilities of drought are too low compared to observed occurrences. Might it be related to the issues described in Farmer, William H., and Richard M. Vogel. "On the deterministic and stochastic use of hydrologic models." *Water Resources Research* 52.7 (2016): 5619-5633.

Discussion: Would sections 6.1-6.3 make more sense in the results section? It seems like they include the presentation of new results rather than discussion. Adding a bit more about how the results fit within the context of previous literature would help to flesh out the discussion and conclusions.

Finally, it would be good to have some more substantial editing of the text to catch typos and clarify a confusing sentences. I would rethink the title which is potentially misleading. The "About the added value of a process-based hydrological model!" might imply to some that the paper will illustrate the benefits of the process-based model. Also, being more specific that the focus is on reservoir levels rather than the broad statement of "seasonal drought" would be good.

Minor comments

P1 Line 10 "associated with" instead of "towards"?

Figure 2: I don't understand the arrows going both directions between the two skill scores boxes

P 8 Line 3 what is revision 257? Maybe you can add an appropriate citation instead of mentioning this in the text because it leaves the reader wondering a bit

C3

P 9 line 15 "goodness of fit" measures, right?

P 9 line 30 – it would be good to define these variables in the text so the reader doesn't have to find the table.

Figure 3 might be nice to know the sample size represented by each bar

P 14 line 27 – why Jan and Feb?

Figure 5. Please explain the figure fully so that it can stand alone with the need to track down another reference.

Figure 7. for P12 (middle plot) It is interesting that RMSE decreases as conditions get more wet but increases for the statistical model. Do you have an idea of why this might be the case?

Figure 9. Explain significance of grey 1:1 line here

---

Interactive comment on *Hydrol. Earth Syst. Sci. Discuss.*, <https://doi.org/10.5194/hess-2018-404>, 2018.

C4