Cover letter to revision of hess-2020-128

Dear Editor, dear Referees,

We have completed a comprehensive revision of the manuscript according to our suggestions made in the replies to the referee comments and the short comments. Below we indicate the changes made with respect to each comment; the related page and line numbers refer to the revised version of the manuscript. For brevity and clarity we do not repeat the referee and short comments and our replies in full length here. Please find them in HESSD. The changes in the manuscript are marked in red.

RC 1 (Elena Toth)

Comment 1: We have rearranged and mostly re-written the entire introduction (section 1) along the suggestions made in the replies to the referee. We have also mostly re-written the abstract for the same reason, and made some modifications to the summary and conclusion (section 4). We hope the updated version of the text is easier to read now, and better conveys the main messages of the paper.

Comment 2:

We have added a better explanation of the purpose and the key messages of the model comparison to section 1.3 in the introduction ('Scope and goals of this paper').

We have replaced the previous 10-node ANN by an LSTM with a single hidden layer of 5 neurons, which is more capable of reproducing a catchments storage-release behaviour. We hope the referee finds this a fairer treatment of the ANN class of models in our study. (see Table 1 on page 8). The LSTM shows better performance than the previous ANN, but the computational effort is still high.

We have changed our model application setup. Instead of training and evaluating the models on all available data, we now follow an validation set approach, training the models on the first half of the data, and evaluating them on the second half. The setup is explained in detail in the new introduction (section 1); all results, plots etc. have been updated accordingly.

Comment 3: We have re-written most of the abstract (please see also our reply to comment 1)

Comment 4: We have maintained the detailed discussion of the 'philosophical' background of the bit by bit method as we believe it is important to put the two contributions of the paper into a larger perspective. Please see also our reply to comment 1.

Comment 5: We moved Eq. 1 into Table 1 (page 8).

Comment 6: We fit the models in a calibration period, and present results for a validation period. Please see also our reply to comment 2.

Comment 7: We have added the entropy equation and a short explanation about how it is related to information and conditional entropy at the beginning of section 2.4.1 (page 9).

Comment 8: No changes made to the manuscript

RC 2 (Anonymous)

Comment 1: No changes made to the manuscript

Comment 2: Following our suggestions in the reply to the referee, we have adjusted the abstract (last paragraph) and the introduction (page 5 line 121 pp)) to better explain what we mean by 'general'.

Comment 3: We have added to the introduction (page 4 line 80 pp) a sentence, briefly naming AIT as the basis for the claim of generalizability, and pointing to Weijs and Ruddell (2020) and references therein for more detail.

Comment 4: We have added a short description of the two possible ways to evaluate information in model predictions: As information loss against an upper benchmark (the observations), or information gain against a lower benchmark (the entropy of a uniform distribution) (page 9 lines 197-200, and page 10, lines 212-216). We also added a short explanation that we use information loss to remain consistent with Weijs and Ruddell (2020). (page 9 lines 214-216).

Comment 5: No changes made to the manuscript

Comment 6: Throughout the text, we have replaced 'universality' with 'generality', and we have adjusted the abstract (last paragraph) and the introduction (page 5 line 121 pp)) to better explain what we mean by 'general'. Comment 7: No changes made to the manuscript

Comment 8: We have rearranged and mostly re-written the entire introduction (section 1) along the suggestions made in the replies to the referee. We have also mostly re-written the abstract for the same reason, and made some modifications to the summary and conclusion (section 4). We hope the updated version of the text is easier to read now, and better conveys the main messages of the paper.

Comment 9: We have changed our model application setup. Instead of training and evaluating the models on all available data, we now follow an validation set approach, training the models on the first half of the data, and evaluating them on the second half. The setup is explained in detail in the new introduction (section 1); all results, plots etc. have been updated accordingly.

Comment 10: We removed the entire paragraph about the relation between descriptive, computational, and natural complexity at the end of section 1.3.

Comment 11: We added a short discussion about the influence of binning choices to section 2.4.1 (page 10 lines 218-224).

Comment 12: Please see our reply to comment 9.

Comment 13: We have added a new section 1.4 at the end of chapter 1, where we give a short overview on uses of the term 'complexity' and related research in the hydrological sciences, and put our uses of the term into perspective.

SC 1 (John Ding)

Comment 1: In addition to the replies given in ReplySC1, we would like to add that, based on recommendations by John Ding, referee #1 and referee #2, we have made several changes with respect to the range of models used for demonstration purposes: All models are now calibrated on the first half of the available data; results in terms of performance and computational complexity are from applying

the models on the previously unseen second half of the data. For calibration, we used the well-known Nash-Sutcliffe efficiency (NSE). We hope that these changes, which follow standard hydrological practice, will help the target audience – hydrologists – to better connect to the approach and key messages of the paper. Nevertheless, the purpose of applying a range of models in this paper is not to compare their performance; the key messages are that

- computational complexity as measured by 'strace' is sensitive to all computational aspects of a model: the amount of required forcing data, the size of the model itself, its time stepping, spatial resolution, numerical scheme, etc. ,
- 'strace' can be applied to any computer-based model, i.e. it is general within its range of application (evaluation of computer-based models).

We added a sentence explaining this in the last paragraph of the abstract, and the introduction (page 5 line 121 pp).

Comment 2: No changes made to the manuscript

Comment 3: No changes made to the manuscript

SC 3 (John Ding)

Comment 1: No changes made to the manuscript

SC 4 (John Ding)

Comment 1: No changes made to the manuscript

SC 2 (Robert Reinecke)

Comment 1: We have added to the summary & conclusions the phrase " in the general setting of incremental learning" (page 16 line 360).

Comment 2: We have added a sentence plus a reference to Hutton et al. (2016) to the summary and conclusion (page 16 lines 364-367).

Comment 3: We changed the related sentence in the abstract (page 1 lines 15 pp).

Comment 4: We changed the related sentence in the abstract (page 1 lines 15 pp).

Yours sincerely,

Uwe Ehret, on behalf of all co-authors