

Hydrol. Earth Syst. Sci. Discuss., referee comment RC1
<https://doi.org/10.5194/hess-2022-63-RC1>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on hess-2022-63

Anonymous Referee #1

Referee comment on "Integrating process-related information into an artificial neural network for root-zone soil moisture prediction" by Roiya Souissi et al., Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2022-63-RC1>, 2022

General comments:

In this manuscript, the authors estimate soil water content in the root zone (RZSM) using data-driven artificial neural networks (ANNs). While a previous study used soil surface moisture (SSM) data as input of the ANNs, the originality of this study is to test complementary input variables such as the normalized difference vegetation index (NDVI), soil temperature, and mathematical transformations of the SSM and potential evapotranspiration (PET), called soil water index (SWI) and evaporative efficiency, respectively. The results indicate that RZSM predicted with the ANNs using the complementary input data tend to be more correlated to measured RZSM than predictions using temporal integrations of SSM data alone. The ANNs were trained and validated using a large amount of observations throughout stations around the world, and the robustness of the predictions was tested against RZSM data in independent stations.

This study addresses an important topic relevant to the readership of HESS. A method proving high-throughput estimations of RZSM at large scale from satellite images could become a game-changer for global water circulation modelling and crop modelling. Overall, the figures and structure of the paper need improvement. Substantial effort would also be needed to improve the presentation of the results and develop their discussion in light of the broader literature.

I am a bit concerned that I could not find a comparison of the qualities of fit of the training and validation datasets, so that in their current form, the results cannot exclude the possibility that improvements of the quality of fit with more complex models are due to overfitting allowed by their higher numbers of degrees of freedom. I am also disappointed that time series of SSM and RZSM from only five stations at the same site are shown, which contrasts with the wealth of data used in the study. There is extensive room to display the wide range of examples of quality of fit (from poor to excellent) across stations and ANN types in the Results section and appendices. Please also consider providing scripts and data as appendices upon publication, as this has become a

widespread Open Science practice in major journals.

Specific comments:

Title: I really like the idea to use process-based models' outputs as inputs of ANNs. It seems to me that the complementary data (PET, NDVI, temperature, ...) used as novel inputs (directly or indirectly) to the ANNs could be inputs of process-based models meant to predict RZSM. However, I am less comfortable with labelling variables as "process-based" (see e.g. line 88-89), as I think such a label may characterize a model but not individual variables. Some of these variables are used to calculate indices (SWI and evaporation efficiency) related to soil water dynamics, but I have doubts that the equations behind these indices actually describe processes, though I agree that they are related to processes, just like the variables. Thus, using the term "process-based information" in the title seems misleading to me. An expression like "key hydrological indices" would probably be more representative of the content of the study.

Introduction

Line 36 (L36): Here I got confused about the meaning of RZSM, and only understood pages later that it meant a *point* observation of soil water content in the root zone, not the integral, or average of soil water content from 30 to 100 cm depth. I think that given its central role, it is very important that the authors clearly define this variable early on in the manuscript to avoid confusion. It is also unclear why say 1m as I could not find observation points reaching 1m depth in the document.

Material and Methods

Figure 1: The fact that all input arrows get into the same "scaling" circle before the ANN I found a bit confusing, as the terms "scaling" and "descaling" are not defined, and it seems like all inputs get into the same neuron. I think the graphical representation could be improved, and space could really be used to provide clarifications in the caption and in the text (the current caption is only 5 words...). The term "pheno." which I guess indicates the phenological stage is also not defined. At some point the NDVI is connected to the process of "growth" in the text. Probably worth selecting a single term to avoid confusion. Note that as discussed earlier, I would consider that the variable NDVI cannot be called "process-based", and that the process of growth, or the phenological stage change, is not modelled here. It is also worth mentioning in the caption that all the tested ANNs are variations of the one represented in Fig. 1, with all inputs or some of the inputs removed.

L100: "three SSM features" are mentioned here but I couldn't find a clear definition or equation for them in the manuscript (only a brief mention that they are rolling averages of SSM in the caption of table 1, then at line 244, which comes too late and without enough

details).

L114: It is confusing that the authors mention "a root zone depth varying between 30 and 60 cm" as "root zone depth" misleads the reader into believing that the bottom of the root zone is located between 30 and 60 cm. Please clarify that "the root zone soil moisture observation point is located between 30 and 60 cm". Also, it is unclear why 60 cm is mentioned, while the deepest observation point is 55 cm.

Figure 2: Words and points are too small to read.

L124-147: Please improve the descriptions for the sites to make them easier to compare. In one case, coordinates are mentioned, in another one yearly rainfall and PET, while the specific soil type at the observation site is only mentioned for one site, etc. Also, ideally a few descriptors (e.g. climate and soil type) for the other sites used in the training and validation steps should be provided in appendix, not necessarily individual descriptors, but at least "population" descriptors giving an idea of the frequency of specific soil types across sites. Climates are mentioned with acronyms that are not defined. They can be found in appendices, but the reference to the appendices is missing.

L162-164: Could you explain in more details why the pixel with highest NDVI value is selected among 16 days? Could you specify if it is at a single location, or if the pool of pixels to choose from covers a wider area than a single pixel?

L165: Throughout the manuscript, the terms evapotranspiration and evaporation are used in alternance. It is often unclear if only soil evaporation or soil-plant evapotranspiration is concerned in the analysis. For instance, potential "evapotranspiration" is used to calculate "evaporation" efficiency, which is counter-intuitive. Could you clarify that and adjust the text accordingly? Do you separate "E" from "T"?

L166-168: The expression "evaporation efficiency" has been mentioned a few times already, but it is unclear what it is as it has not been defined yet, which contributes to the overall confusion. Maybe a definition in a few words would be good early on, for instance in the introduction, which is supposed to introduce important concepts for the reader to understand them before going through the bulk of the manuscript. This remark could apply to other important concepts in the manuscript.

L174: If the raw PET value is the "sum" of PET over 8 days, do you turn them into daily values by dividing them by 8? Please clarify the integration and deconvolution process.

L182: The SWI_m index is calculated as a recursive series, but it is not mentioned how

the first value of SWIm is calculated in the time series. Please explain it.

L182-189: It is unclear what the subscript "m" stands for in SWI_m, and is it necessary? Overall many of the symbols feel a bit obscure or poorly chosen. A few examples include the symbol β_3 (Why is there a subscript 3, and is it necessary? Is it the same variable as B₃? If not, please choose symbols that are easier to distinguish), A₃ (Why is there a subscript 3, and is it necessary?), the variables that have a temporal component (The time indices alternate between "t_n" and "n", please be more consistent), θ_L (Please do not use soil layer thickness as a layer index as two layers with the same thickness would not be distinguishable).

L182-189 and throughout the manuscript: Please introduce new variables with their units.

L195-200: This paragraph should be located in the Results or Discussion section.

L201-208: These paragraphs should be located in the Discussion section.

L214-229: I find the presentation of the evaporative efficiency equations confusing. The text suggests that β_3 is the evaporation efficiency index, but it also says that it is the ratio of actual to potential soil evaporation, which does not seem to be the content of the equations. Please clarify directly that equation 4 is not used or even better, put it in appendix (maybe describe the connection between the standard equation of P and yours in appendix, in order to go to the point in this part of the Methods). If P* is used in the calculation of β_3 , then please use the symbol P* in equation 3. Also, there is no need to provide two definitions of θ_{max} then to say that you are not using the first one. Only mention in the Discussion section that another definition exists if you think it is worth discussing, or else send this information to appendices. Same story for the choice between LE_p and PET.

L250: Please clarify in this part of the Methods how you are using the additional independent datasets from Italy, India and France.

Results

Figure 3: The caption does not specify if the results displayed concern the training or validation dataset (or both). The evaluation of the quality of the prediction should be done on the validation dataset (please display these results independently in parallel to the results of the training dataset to ease the comparison), whose quality of fit should remain similar to the quality of fit of the training dataset. If the quality of fit in the validation stage is substantially lower than in the training stage, the ANN is considered to "overfit"

the training dataset. I do not think that this comparison is provided in the current version of the manuscript, while it is central for the evaluation of the results. There is an evaluation of the quality of fit in an independent dataset (for Italy, France and India) but the quality of fit of this "second validation" step is not compared to the quality of fit during the training stage. Also, the correlation between measurements and predictions as an indicator of quality of fit has limitations. One of them is that it is "blind" in case of predictions proportional to measurements. That is why root mean square errors are provided systematically when conducting such a model validation. Please systematically provide results for this metrics too.

L255-258 and other places in the results: If the validation and training sets seem are merged in the histograms, it is possible that the improvement of the quality of fit is due to stronger "overfitting" allowed by the larger number of ANN parameters as more input variables types are added from the simplest ANN_SSM to the most complex ANN_SSM_NDVI_EVAP-EFF-B60_EXP-FILT_T5 model. If that is the case, from what is shown the authors cannot exclude that the improved quality of fit is due to overfitting allowed by the increasing degree of freedom. Could you specify the degree of freedom of each ANN model, and show results for training and validation stages separately?

L288-311: It is interesting that for each ANN, the set of parameters remains the same across stations while also allowing seemingly good predictions of RZSM. I think it is something to discuss, particularly because it does not have to be so. The authors could have separated the datasets in sub-groups corresponding to a few major climate types or major soil types. Instead, they merged them all, seemingly trying to derive "general relations" between RZSM and a few variables, regardless of the soil and climate types. I think it would be of interest to develop the analysis (or at least the perspectives) in this direction. Though it is not within the scope of the objectives of the manuscript.

Figure 4: This is the only time SSM data (measured) and RZSM data (measured and predicted with only one of the ANNs) is shown. This is far too limited to have an idea of the quality of predictions. Please provide at least a comparison of ANN_SSM and ANN_SSM_NDVI_EVAP-EFF-B60_EXP-FILT_T5 models within the Results section, and more diverse examples (e.g., training vs validation datasets, a few typical fits, good and bad) in the Results and/or appendices.

Figure 5: It is unclear to me why the bars corresponding to "all stations" do not have the same length across panels (except for panel d, ANN_SSM_TEMP, which does not have temperature data for all stations). I think this should be clarified. Also, it is unclear what the numbers in blue are on top of the bars. This should appear in the caption and body of the text.

L337-364: These are new results. They should appear in the Results section even if they are a transverse analysis of results already shown. The associated methods (including equation 6) should be detailed in the Material and Methods section. There is substantial space to develop a discussion in view of the existing literature, and a few perspectives.

Figure 6: The climate types in the legend should be defined through a reference to the table in appendix. The darkest colours do not allow reading the overlapping symbols for the types of ANN. Please display results in a way that makes it comfortable to read them, possibly by duplicating panels. It would also be interesting to see where the stations are located, using colours to show which input variable was the most critical for each site, and compare with the "climate-type-associated-critical-input" already there in the figure. Please make the caption more descriptive (in general across figures). A first point is to mention in the caption that what is shown is a "World map of (...)".

L344-364: This part of the analysis uses substantial extrapolation spatially-speaking, which I think would be ok if it was shown that within each climate type, one of the input variable type clearly stands out as a key predictor. Given the results shown in Figure 5, it seems that it is not the case, and that various stations allocated to the same climate type may most rely on diverse input variables. Thus, I am not really convinced about the added value of this part of the analysis (at the end of the discussion, the authors also acknowledge that this classification suffers limitations), and I wonder if the authors could make it more convincing using complementary analyses, or else I think it would be better to remove it from the paper.

L377-383: Please acknowledge in the text that the quality of fit decreased in Italy and India. This is also worth developing in the Discussion section.

Typos and minor details:

L35, 81, 88 & 334: The acronyms ECV and ML are introduced but I couldn't find them in the rest of the text. The acronym MLP hasn't been defined yet when first introduced and is only used once after being defined. I could not find the definition of the acronym LST. More generally I would avoid introducing unnecessary acronyms, and remove the ones listed in this comment.

L41: I don't get why the terms mission is there twice, while it seems the sentence describes sensors. Please correct or clarify.

L43: has -> have

L68, 69, 74 and many other instances: When the authors of the cited paper are the subject of the sentence, or any kind of complement within the sentence, please use the following form "author et al. (year)" rather than "(author et al., year)".

L77: Please clarify that these are correlations between series of measured and predicted

RZSM and whether it concerns the training stage or validation on independent data.

L87-90: In point (1), the temperature input is missing in the list.

L176: The sub-title "Methods" in the section Material and Methods is not very specific. Could you modify it to make it more specific?

L186: "which occurs in [0, 1]" -> "which ranges between 0 and 1".

L191-192: The expression "given (...) T values" appears twice in the same sentence.

L210: It seems to me that "two" ANN models include the evaporative efficiency input.

L255-257: No need for 2 digits in the numbers. One or none would seem sufficient to me given the large differences.

Throughout figures: Please use letters to refer to the panels within individual figures.

L342: There is a "*" symbol at the start of equation 6, which seems to be a typo.