Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2020-363-RC2, 2021

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Predicting Sediment Discharge at Water Treatment Plant Under Different Land Use Scenarios Coupling Expert-Based GIS Model and Deep Neural Network" *by* Edouard Patault et al.

Anonymous Referee #2

Received and published: 22 February 2021

General comments

The proposed article presents two topics: the first one is the realization of a deep model to estimate the Radicatel solid transport at the spring (or in boreholes?). The second is to generate different shallow solid transport scenarios from agricultural hypotheses. I personally have no knowledge of agriculture to really appreciate the interest of this approach. However, I have some doubts about the originality of the results. It seems obvious that if we want to avoid turbidity in the springs, it is better to avoid generating it on the shallow water. Concerning the quantification brought by the cascade of

models carried out (it is not a coupling), I have serious doubts about the robustness of the numerical values, because no estimation of uncertainty is carried out throughout the article, and because the description of the deep model design is not sufficiently accurate to assess its quality. Overall, the article is confused, the description of the data and of the processing carried out is scattered everywhere and it is difficult to find relevant information.

My recommendation is that the paper cannot be published in the current state because too many improvements must be done. I regret that, because the subject is interesting and challenging.

Suggested improvements are: (i) adopt a simpler and more efficient outline; (ii) given the small amount of data, use a shallow model, benchmark with a linear model; (iii) better present the model design and the model actually used in a very precise way (list of inputs, very precise architecture); (iv) do not shift the input series in relation to the output series: neural networks know how to manage this; (v) make a statistical description of the data in order to understand why the results on the test set are better than the training one. This, moreover, allows us to prejudge bad results on another test set. Other general remarks: Words have a signification: it is not really a coupling of models: one feed the other. Goal is focused on operational needs, but operational needs are not accurately described. Very few details on how the solid flow is obtained. What are the assumptions and the model considered? It is also written that "This new approach can be easily implemented". Applying a deep model is anything but easy. This may explain inconsistencies and poor explanations found in the paper.

More specific comments are hereafter

Abstract

P1 L15 "and they can be seen as a black-box due to the non-linearity of the processes generating sediment discharge. Âż there is no straightforward link between the black box property and the non-linearity of the relation. Black-box means "unknown" when

nonlinear means frequently “difficult”.

P1L23; is water extracted at sinkhole or spring? It could be interesting improving the consistency to enhance the impact of the abstract.

Introduction

P1LL 39-43. Could you please be more accurate? It is not said if the cost is linked to sediments or to other dissolved pollutants.

P2L66 “As stated by Shen (2018), DNN have now surpassed traditional statistical methods”. This sentence is too simple. It has no meaning by itself without the definition of “traditional” statistical methods, and the definition of the limitations. For example it is false for linear problems or for nonlinear relations in general. Only specific nonlinear relations need deep models. The drawback of deep models is that they need extensive database. Are they really relevant with highly noised data? This question could be discussed in the light of a comparison with the results of a linear model.

P2-L72-78. The presentation of the goals and methods is chaotic, please could you present accurately what is the goal: prediction of which variable, where and at what time-step, with or without rains?, . . . One remark about vocabulary. Usually the words prediction or forecast are used with actual data. When data correspond to scenario, it is not a prediction, maybe a prospective?

2. Study site

It is not clear if the cited catchment (106 km^2) is the underground basin, or the shallow basin. Both basins can be different in karst context. Could you also clarify the notion of "positively connected"? By tracing tests, by signal analysis?

L71, one cited goal is to study “hillslopes erosion processes into karstic transfer”, but elevation is not provided and the plateau seems very flat. Is there an explanation?

Finally the presentation of the study site could be improved. The same apply to the

hydroclimatic data. Not all types of data are presented. This section is poor and must include information provided in section 3.1 “Data handling”. A comprehensive description of data is important for data-based models.

3 Methodology

L113 what did you mean by “training limits”; is it training set?

I have a concern with the word “coupling”, or coupled models. Coupling don’t means preprocessing or cascading processes; coupling means that processes and their attached data depend each one from the other, as for example in coupled differential equations. If my comprehension is good it is not the case in the paper. Please clarify this point.

L120, I do not understand this sentence: “In accordance with previous studies by Masséi et al. (2006) and Hanin (2011) in karstic environment in Normandy, a lag of 1 day was applied to the SD time series to properly match the rainfall input”. Have you shown before that the ANN was unable to calculate the good lag? How is managed this artificial delay in the following results?

L 117, please could you explain what is the difference between calibration and training, and what is a reference period? In my opinion it is only training.

L132 “burn the stream network”?

The section “Erosion and runoff modelling” is badly organised: the description of the data should be put in a "data" chapter, the method in a "method" section, and its limitations should be given. Here everything is mixed up. You get lost in the article.

3.3 DNN Configuration First of all, as I understand it, the amount of data used for training is only 751 examples. This seems very little for a deep model, which necessarily includes a lot of parameters to adjust. Shen et al 2018 that is cited in the paper wrote also: “Despite the comparisons between DL and nondeep machine learning, this author would strongly advise against applying DL nondiscriminatively. The earlierâĂŘ-

generation methods can be highly valuable for their respective problems and situations, especially when there are limited data or relatively homogeneous data. As shown in studies, for small data sets, DL could be at a disadvantage compared to models with stronger structural assumptions."

I am sorry, but the presentation of the model is not up to international standards. It is necessary to indicate precisely how the data and related time windows are selected, how overfitting is avoided (or underfitting, given the small number of examples), and accurately what structure is obtained after these steps. Are there any regularisation methods used?

3.4 Performance evaluation

L179 There is a confusion between the coefficient of determination and the linear correlation coefficient (eq 1). The coefficient of determination is equal to the Nash efficiency. Please correct this error.

The use of the word "prediction" also needs to be discussed. Obviously the results represent what might happen in the future; but the models are fed with data that are also into the future. Actually the model only makes an estimate. If the model performs a prediction (effective anticipation from inputs in the past or present) then quality criteria specific to the prediction should be used, such as the persistence criterion. This is not the case.

3.5 Designed storm projects and land use scenarios : to put also in a data section.

4.2. DNN: Calibration and Generalization

The monthly-backward chaining nested cross-validation procedure must be defined more clearly and more accurately as this process is critical. I definitively not understand the sentence "Predicted values of runoff and sediment discharge were extracted over the connected sinkholes and summed to be used as inputs . . .", please correct it: how can future data be "extracted" from a sinkhole? An how can we sum runoff and

sediment discharge?

What is the accuracy of the turbidity/Solid-transport relation?

L 268. What does mean the sentence: "The modelling results were less efficient than for the complete dataset but overall satisfactory"

Results suggest that test data are better represented than training data. This suggest that the good quality obtained on test set will not be generalizable to the part of data used in training test. Results seem to be overestimated.

5 Discussion

Is it possible to describe the Generalized Extreme Value Distribution (GEV) in the material and method section?

L 345 : "Even if it is well known that deep learning-based methods may results in weak performance for extreme events (Zhang et al., 2019)". Then I no longer understand the coherence of the article: why did you use this method on extreme events?

5 Conclusion

L393. "This new approach can be easily implemented". Applying a deep model is anything but easy

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2020-363, 2020.