Hydrology and Earth System Sciences
Discussions

Hydrol. Earth Syst. Sci. Discuss., referee comment RC1
https://doi.org/10.5194/hess-2022-365-RC1, 2022

Comment on hess-2022-365

Anonymous Referee #1

Referee comment on "Comparison of artificial neural networks and reservoir models for simulating karst spring discharge on five test sites in the Alpine and Mediterranean regions" by Guillaume Cinkus et al., Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2022-365-RC1, 2022

General comments

This paper is addressing the comparison of an Artificial Neural Network (ANN) model with a reservoir model “KarstMod” in five contrasting karstic springs across Europe. Several papers are comparing the results of ANN models with conceptual hydrological models. However, this paper is the first paper comparing the applicability of an ANN model with a widely-used karst hydrological model "KarstMod" which is particularly developed for heterogeneous environments and successfully represents the karst system hydrodynamics. In my opinion, this paper is suitable for the audience of the *Hydrology and Earth System Sciences*. However, the paper needs methodological improvements and the strengthening of several weak spots as described below. Considering the current state of the paper, a major revision is adequate.

Kind regards,

Anonymous reviewer

Major comments

Line 290: I think the authors must explain why they used different performance metrics for the calibration of the models. Choosing the objective function(s) is the subjective decision of the modeller, however, if you want to make a comparison between the model performances, I recommend calibrating the models according to the same performance metric.

Line 294: Is this adequate for a successful calibration?

Line 298: This is not a scientific statement. Running the model for 1 hour does not tell anything about the number of simulations. Instead, I recommend applying a threshold value (e.g. NSE > 0.5) for the behavioural runs and selecting all of the simulation results above this threshold rather than selecting the top 1000 simulations. By doing this, you would probably get a different number of behavioural model runs in each catchment and you can better compare the model uncertainty bounds. (Alternatively, you may mention the stop criteria of the KarstMod model e.g. simulation-time limitation, or number of behavioural runs)

Line 305: Here you mention that you evaluated the model performances by using KGE and DE, but in **line 290**, you calibrated the models by using NSE (reservoir model) and MSE (ANN model) metrics. So, why did you change the objective function in the model evaluation phase? Please explain. As far as I experienced, the best simulation obtained by any of the performance metrics (e.g. NSE) does not guarantee the best simulation on another metric (eg. KGE). So, please calibrate and evaluate the models again by using the same performance metric(s).

Line 350: This plot shows that the reservoir model outperforms the ANN model in Spanish, French, Lebanese and Slovenian catchments except for the Austrian catchment. I think the Austrian catchment is dominated by snow (as you mentioned in **line 379**) and the reservoir model structure is not adequate to beat the ANN model. It must be discussed in the discussion part. Additionally, your snow routine may not be adequate if you did not calibrate the snow parameters. Please see the paper (ÇallÄ± et.al. 2022).

Minor comments:

Lines 39-40: In my opinion, you would have a better classification of the models. Kovacs and Sauter (2007) do not classify the models as “data-driven” or “distributed” models. Distributed or lumped, all models are somehow data-dependent. I think that would be better to classify the models as “black-box models”, “conceptual models” and “physical models" considering their complexities. You may give details about the "Machine learning models" under the "Black Box", and reservoir models (or lumped parameter models) under the "Conceptual models". You may also give some details about the advantages and disadvantages of these modelling approaches regarding the complexities and data requirements. You may mention why so many researchers choose conceptual models. Another point is that, please be more consistent about the model classification inside the paper.

Line 47: You may consider citing the paper (Addor and Melsen, 2019) about the model selection procedure (adequacy or legacy).

Line 50: You may consider rephrasing the sentence "…distributed models require a lot of data". You may alternatively say: "Distributed models require the data with high spatial resolution, however, lumped models require data in high temporal resolution." (You may cite here again Hartmann et.al. 2014 or Kovacs and Sauter 2007).

Line 51: I would remove the sentence "Both black-box and reservoir models …." to avoid repetition. The following sentences already explain the applications of the ANN and reservoir models for academic and operational purposes.

Lines 55-60: Please be more consistent about the references in the brackets. You may use time order (Perrin et.al. 2003; Jukic and Denic-Jukic 2009; Tritz et.al 2011; Bittner et.al. 2020) or alphabetical order (Bittner et.al. 2020; Jukic and Denic-Jukic 2009; Perrin et.al. 2003; Tritz et.al 2011).

Line 81: The paper does not include any simulations by using the artificial future data (different emission scenarios) to compare the model adaptability against climate change. So this is not fair to have such an inference. If you want to declare that the models are not robust in extreme event predictions, please cite several climate-related studies (to better link the connection between the climate-change and extreme events).

Line 91: I recommend moving map A.1 to this section.

Line 174-175: In this sub-section, adding a schematic illustration of the ANN model would be very helpful to better understand the modelling architecture.

Line 180: Please simply explain why you applied the 1-D convolutional layer approach (you may consider citing previous ANN modelling studies).

Line 184: I think you can move the names of the python libraries to the appendix. It is not necessary inside the text.

Line 185: Here I see you used the library BayesOpt. Was that library used for the Bayesian uncertainty analysis? Or did you apply an uncertainty analysis to the ANN model predictions? If so, please explain which method you applied. Please give some details about the model assumptions, parameter distributions etc.

Line 189: You may consider pointing out the functionality of the conceptual models in karst water predictions. What is the main advantage of this modelling approach?

Line 240: You may consider adding a snow reservoir above the Epikarst in Fig 1a. This would be more suitable for the mountainous catchment.

Line 261: Please mention how to determine the snow routine parameters (Degree-day factor and melting temperature). You mentioned that you did not make an optimization for the snow parameters, so please cite the relevant literature (e.g. He et.al. 2014).

Line 343: The uncertainty bounds are not easy to see especially for the reservoir model in (a). I think when you select all the behavioural simulations (above the threshold), the uncertainty bound will be much more visible. Then the reader can make a visual comparison between them.

Line 347: Again the same problem. Please apply a threshold for the reservoir model, and use all the behavioural runs to obtain the uncertainty bounds.

Line 355: Please share the model calibration skills in the table.

Line 355: You mentioned that the ANN models require long time series to learn the functionality of the karst system. On the other hand, reservoir models could be calibrated for relatively shorter periods. But, there are some results to be discussed in detail as below:

Lez and Qachqouch springs simulation results support the hypothesis, but the Aubach catchment does not. We expect better calibration skills in the reservoir model in a short calibration period. However, the ANN model outperforms the reservoir model in Aubach. How do you explain this?

Line 389: You can make a representative snow routing even if you do not calibrate the snow parameters. Please cite the relevant literature.

Line 527: You may discuss the uncertainty in the temperature data. Temp data strongly affect the timing of the recharge, especially in snow-covered areas (Aubach case).

Line 547: You may add some other model optimization techniques (e.g. cross-validation, see Wilks 2011).

Line 559: Please discuss the model's structural adequacy here.

Recommended references

Addor, N., & Melsen, L. A. (2019). Legacy, rather than adequacy, drives the selection of hydrological models. *Water resources research*, *55*(1), 378-390.

ÇallÄ±, S. S., ÇallÄ±, K. Ö., YÄ±lmaz, M. T., & Çelik, M. (2022). Contribution of the satellite-data driven snow routine to a karst hydrological model. *Journal of Hydrology*, *607*, 127511.

He, Z. H., Parajka, J., Tian, F. Q., & Blöschl, G. (2014). Estimating degree-day factors from MODIS for snowmelt runoff modeling. *Hydrology and Earth System Sciences*, *18*(12), 4773-4789.

Wagener, T., Wheater, H., & Gupta, H. V. (2004). Rainfall-runoff modelling in gauged and ungauged catchments. World Scientific.

Wagener, T., & Gupta, H. V. (2005). Model identification for hydrological forecasting under uncertainty. Stochastic Environmental Research and Risk Assessment, 19(6), 378-387.

Wilks, D. S. (2011). Statistical methods in the atmospheric sciences (Vol. 100). Academic press.

Xu, T., Longyang, Q., Tyson, C., Zeng, R., & Neilson, B. T. (2022). Hybrid Physically Based and Deep Learning Modeling of a Snow Dominated, Mountainous, Karst Watershed. *Water Resources Research*, *58*(3), e2021WR030993.