Response to Referee #1: by Kuo-Chin Hsu

We thank Kuo-Chin Hsu for reviewing our manuscript and for the comments provided. We will address all aspects in the following. Please find our answers in red and the original comments in black.

The manuscript proposed to use convolutional neural network (CNN) associated with gridded meteorological data for Karst spring discharge modeling. CNN was applied to three karst spring watersheds in Europe. Results of 2D CNN model associated with gridded meteorological cells were compared to that of 1D CNN using climate station input data.

The manuscript is well written and technical sound.

General comments:

- CNN is a mature data-driven tool which highly relies on data availability and quality. The authors argue that less data is needed in the proposed approach to obtain satisficing results compared to previous deep learning approach and overcome the short of data from climate stations. The results show that 2D modeling is not necessary better than that of 1D and previous modeling in Lez spring. A question raised is that whether the key input data has been identified. For example, pumping may play an important driving factor but is not included in training and screened out by Bayesian model. Gridded meteorological data may not be enough to improve the model performance. The authors needs to address their contribution. Guide line for data preparation will be helpful for the suggestion of use machine learning.

Thank you for this thoughtful comment on data aspects. We think indeed that the 2D approach can overcome difficulties in climate station data availability. As we state in the manuscript: "climate stations are often not available within the catchment of a spring, do not match the data availability of the discharge time series (period or temporal resolution), or are more distant and thus do not truly represent the events in the catchment itself". Nevertheless, we want to clarify that we do not think that the 2D approach needs less data, instead we think that rather the amount of work necessary to collect and preprocess the data is strongly reduced. Gridded meteorological data is available online and needs only minor preprocessing in contrary to most climate station data.

We agree that the results of the 2D approach are not necessarily better, as it can be seen in the example of Lez spring that you mention. It is true that Lez spring discharge is a complex combination of natural discharge, pumping and legally regulated minimum discharge from the extracted water to protect downstream ecosystems. Nevertheless, we showed, that we were able to simulate the discharge with solely meteorological data, using both the 1D and 2D models. In this specific case of Lez spring, a lot of work was necessary to produce the 1D precipitation input time series (compare Appendix B - Lez Catchment Precipitation Interpolation) due to very patchy climate station data. It seems that these additional efforts pay off in terms of higher performance compared to the 2D model. Nevertheless, the 2D climate data seems to offer a sufficient substitution, if needed.

Regarding the consideration of other external driving factors like for example pumping, we agree that these may play an important role, and their inclusion might improve model performance. It would also be possible to combine the 2D input of meteorological parameters with a 1D-input of

additional parameters. However, the focus of our study was on the direct comparison of 1D and 2D-meteorological inputs, so we did not consider other than meteorological factors.

We do not think that we can provide a general guideline for data preparation, because this step strongly depends on the datasets that a specific user intends to use. Nevertheless, we have included a description of the input data format that is used in our published python scripts, which should enable future users to adapt and apply them.

- The modeling uncertainty is quite low to almost without uncertainty that seems abnormal. The authors may explain this.

We apologize that, given our current formulation, this aspect does not become clear. The shown model uncertainty is derived from an ensemble of 10 differently initialized models, each using Monte-Carlo dropout to produce an ensemble of 100 different forecasts (so 1000 in total). For each of the 10 models we calculated the 95% confidence interval of the 100 available forecasts (1.96 times standard deviation, because of sigma rule for 95% confidence). What is shown in the final plots, is the 95% uncertainty of the mean of all 10 model ensembles, which is indeed very small. In a revised version of the manuscript, we will shortly clarify this aspect. We want to add that the shown uncertainty does not include other sources of uncertainty of the models (such as input data uncertainty). For this reason, it might seem abnormally small to you at first glance. We also will add a clarifying statement on this aspect in a revised version of the manuscript.

Specific comments:

- Line 203, write the long short term memory for abbreviation of

Sorry, we forgot to introduce the abbreviation LSTM. We will correct that.

- Lines 213-215. Although the all programs are available from Python community. Technical functions should be described for the used library or framework should be explained.

We are not quite sure what you mean exactly. We tried to explain the most important technical functions (like CNN) in the methodology section and give the associated references. Given the limited space for the manuscript, we avoided an in-depth explanation of all details and hope that referencing the used packages and frameworks in combination with the published Python code is sufficient. Additional necessary information should become clear to the python-experienced user from our published code.

- It is not clear functions for training, validation, optimization and testing periods in Table 1. The should be explained in main text.

Yes, you are right. Sorry that we missed to explain the purpose of those sets. We will add a clarifying statement to a revised version of the manuscript in accordance with the comment on that aspect of Reviewer#2.