Response to the short comments

Reviewer 1

Reviewer Comment 1.1 — This paper uses long short-term memory (LSTM) neural networks to fill in gaps in spatially distributed time-series data. The performance of the LSTM-based gapfilling method is compared to that of a traditional, popular gapfilling method: autoregressive integrated moving average (ARIMA). Overall, this paper is well written, structured and results seem sufficiently justified and useful. However, this paper is very technical and there is no physical insight beyond just feeding data into a standard code. I think this paper should be published as technical note (not as research article). Several aspects could be further improved in order to having it published in this journal.

Response: We thank the reviewer for overall positive assessment of our manuscript. While we emphazised the scientific importance of gap-filling the spatio-temporal data to capture and understand dynamic behaviors of complex systems, we also agree with the reviewer that our primary focus was to introduce a technical method that can fill data gaps by capturing the dynamic features using LSTM. We will resubmit this manuscript as a technical note as suggested.

Reviewer Comment 1.2 — Would you guarantee the LSTM method in your paper can achieve the same excellent performance in other areas of the whole world? Is it possible that the good performance of the LSTM model is just applicable for the case given by the manuscript? The authors should include one more test for another area (maybe not in the text but in the supporting materials).

Response: There is unfortunately no guarantee for any model to have the same performance in other applications, which is the case for all data-driven and physics-based models. The performance of data-driven models can be optimized by iterating on various model configurations based on data types and characteristics of the relations between predictors and desired responses, as we have demonstrated in our study case when comparing ARIMA with LSTM. Sometimes the ARIMA works better, and sometimes the LSTM works better, and we believe that we have started down the path of predicting which model will work better for a given case. The LSTM model we adopted in our study has the same chain-like nature as other recurrent neural networks, meaning that this architecture lends itself well to sequences, so it will often be a useful (if not the best) approach for dynamic system behaviors [Karim et al., 2017] [Malhotra et al., 2015] [Kratzert et al., 2018] [Malhotra et al., 2016] [Wang et al., 2017] [Reddy and Prasad, 2018][Lipton et al., 2015]. The optimal model configuration and performance we can achieve would be case by case, and our focus of this technical note in to introduce a general method that can be broadly applied to other systems and be evaluated similarly. We will emphasize this aspect of transferability in our discussion and conclusion sections.

Reviewer Comment 1.3 — LSTM model is only compared to ARIMA. Why not compare LSTM with other widely-used methods (such as Kriging interpolation and Gaussian process)? Furthermore, are the authors familiar with DIEOF (Data Interpolating Empirical Orthogonal Functions) which are proposed by Beckers and Rixen (2003)? I think that DIEFOF is powerful and useful for filling temporal and spatial gaps in geophysical datasets. Maybe the authors can compare LSTM with DIEOF.

Response: There are many interpolation approaches that are commonly used, including the EOF-based approaches and kriging (based on Gaussian processes) that the reviewer mentions. We did discuss that kriging and the Gaussian processes are mostly used for spatial interpolation rather than spatio-temporal interpolation. We will add more discussions on the EOF related interpolation methods, such as least squares EOF (LSEOF), data interpolation EOF (DINEOF), and recursively subtracted EOF (REEOF), which are widely used to fill in missing data from geophysical fields such as clouds in sea surface temperature datasets or other satellite-based images with regular gridded domains. However, as discussed by Beckers and Rixen (2003), "For the method to have a chance to work, one needs, for each moment, at least a sufficiently large number of data points (otherwise one should drop the whole picture) and for each spatial point a sufficient amount of data in time (otherwise one should discard the point from the analysis)", which is a challenge for most of the monitoring networks that are sparsely distributed. Therefore, we would keep using ARIMA as the benchmark since it is the most commonly used method in time series analysis (a conclusion based on reviewing the hydrological literature) and gap filling. While we acknowledge that we did not (and could not) explore every possible interpolation method, we feel that by choosing such a representative approach, our study would not suffer greatly from loss of generality. We will also acknowledge this aspect in our discussion and conclusion sections.

Reviewer Comment 1.4 — The present title "Using Deep Learning to Fill Spatio-Temporal Data Gaps in Hydrological Monitoring Networks" are inaccurate. I suggest new title like "Using Long Short Term Memory Neural Network Model to Fill Spatio-Temporal Data Gaps in Hydrological Monitoring Networks"

Response: We agree that specifying LSTM in the title could be more accurate, so we will change the title as suggested.

References

Fazle Karim, Somshubra Majumdar, Houshang Darabi, and Shun Chen. Lstm fully convolutional networks for time series classification. *IEEE access*, 6:1662–1669, 2017.

F. Kratzert, D. Klotz, C. Brenner, K. Schulz, and M. Herrnegger. Rainfall–runoff modelling using long short-term memory (lstm) networks. *Hydrology and Earth System Sciences*, 22(11):6005–6022, 2018. doi: 10.5194/hess-22-6005-2018. URL `https://www.hydrol-earth-syst-sci.net/22/6005/2018/`.

Zachary C Lipton, David C Kale, and Randall C Wetzel. Phenotyping of clinical time series with lstm recurrent neural networks. *arXiv preprint arXiv:1510.07641*, 2015.

Pankaj Malhotra, Lovekesh Vig, Gautam Shroff, and Puneet Agarwal. Long short term memory networks for anomaly detection in time series. In *Proceedings*, volume 89. Presses universitaires de Louvain, 2015.

Pankaj Malhotra, Anusha Ramakrishnan, Gaurangi Anand, Lovekesh Vig, Puneet Agarwal, and Gautam Shroff. Lstm-based encoder-decoder for multi-sensor anomaly detection. *arXiv preprint arXiv:1607.00148*, 2016.

D Sushma Reddy and P Rama Chandra Prasad. Prediction of vegetation dynamics using ndvi time series data and lstm. *Modeling Earth Systems and Environment*, 4(1):409–419, 2018.

Qianlong Wang, Yifan Guo, Lixing Yu, and Pan Li. Earthquake prediction based on spatio-temporal data mining: an lstm network approach. *IEEE Transactions on Emerging Topics in Computing*, 2017.