Response to Editor's Comments

Editor

Editor Comment 1.1 — Please address all comments of referees 1 and 3. Referee 3 makes comments which has been made earlier also: 1) the need to show what is novel (what is contribution to science) in this paper using LSTM in hydrological predictions if compared to other papers of this kind; 2) in this gap filling problem you are not using the "future" data (which are known) - please explain why or update your method to use it.

Response: Both reviewers' comments have been carefully addressed. We summarized the novel contribution in our conclusion section and highlighted in our abstract. The key take-away messages are: The LSTM method is able to account for both low-frequency (seasonal) and high-frequency (weekly, daily and subdaily) dynamics in data, while the traditional ARIMA type of methods focus on capturing the seasonal dynamics in data. It is also important to note that ARIMA methods fail to capture abrupt changes that are present in highly dynamic data. Wavelet analysis can be performed to understand whether high-frequency dynamics exist, which can then guide whether LSTM or other deep neural nets are needed. We also demonstrated that LSTM makes it easier to include both spatial and temporal correlations in spatially distributed data and to regress on other co-located data. LSTM may require more training data, which is another important factor when making decision. Given the importance of irregularly distributed monitoring data to understand environmental systems and to inform decision making, our work is showcasing the power of emerging deep learning methods in filling data gaps, thus improves the value of costly monitoring data.

We ran more than 400 individual models using both LSTM and ARIMA approaches to answer the following questions: (1) how is gap filling performance impacted by the length of gaps? (2) how does the amount of training data impact the model performance? (3) how important is the choice of the input time window? (4) how does the dynamics signature of the data impact the performance of both models? (4) How much value can measurements at neighboring add to the performance improvement? The answers to these questions are highly valuable in practical applications. Our discussion paper has been cited 16 times on google scholar in the past 2 years, which also reflects how much the research community values our work.

The models developed using ARIMA and LSTM in our paper were predictive models which did not explicitly include the future information. However, the future information has been implicitly used in training the spatial-temporal correlations in LSTM models. As we discussed in our response to the comments of Reviewer 3, this is a common practice in gap filling research. We added relevant statements in the introduction to further clarify. We also discuss in our conclusion section that bidirectional LSTM can be a natural choice for incorporating future information. However, developing, training and evaluating Bi-LSTM is not a trivial task, and it will overload the current paper. Thus, we would be happy to pursue this as a future publication.

Response to referee comments

Reviewer 1

Reviewer Comment 1.1 — The paper has improved significantly, yet in my opinion still the text does fail to explain data and equations were it should and some paragraphs leave uncertainty. However, the equations and answers of details appear in another section which makes it very difficult to follow. I think as it is, I suggest a las English review to see that the explanation of basic concepts are mentioned where they should. For the rest, that paper is a valuable contribution and has all information required for publication.

Response: Thanks for the positive assessment on the work we have done to improve the manuscript in the previous revisions. To further address the reviewer's comment related to clear explanation, we worked with a professional editor for a thorough edit on the English. The equations/metrics have been moved to the section 3.1 with the descriptions as suggested by the reviewer. The track changes version has been submitted as well.

Response to referee comments

Reviewer 3

Reviewer Comment 3.1 — This paper presents a method to fill data gaps in hydrological monitoring networks based on LSTM. It is of significance in hydrological applications. However, the strategy used for gap filling is the same as that for hydrological predictions. In other words, the method predicts "future" based on previous observations. In my opinion, data gap filling is different from prediction in that the observations occurring both before and after gaps can be used in gap filling. This study only uses the observations occurring before gaps, which is in essence prediction.

Response: The models developed using ARIMA and LSTM in our paper were predictive models which did not explicitly include the future information. However, the future information has been implicitly used in training the spatial-temporal correlations in LSTM models. Framing gap filling as predictive problems is a common practice when machine learning methods are used for filling gaps in time series data (see examples in Kandasamy et al. [2013], Körner et al. [2018], Chen et al. [2020], Zhao et al. [2020], Sarafanov et al. [2020], Contractor and Roughan [2021]). We added relevant statements in the introduction to further clarify. We also discuss in our conclusion section that bidirectional LSTM can be a natural choice for incorporating future information. However, developing, training and evaluating Bi-LSTM is not a trivial task, and it will overload the current paper. Thus, we would be happy to pursue this as a future publication.

Reviewer Comment 3.2 — The evaluation metrics of the prediction models seem very good. The performance of time series prediction models largely depends on the amplitude of data variation. It is recommended to present the SpC variations with different time intervals (1, 6, 12, 24, 48, and 72hours). Also, the values of RMSE are affected by the magnitude of data. It is helpful to present the mean value of SpC observations.

Response: Thank you for the suggestions. Mean and variance of the SpC from model testing period in different time intervals have been presented in two versions of boxplots: with and without outliers. The boxplot without outlier aims to reveal the distribution of majority data points (99.3% of all the data has been used in the boxplot). Figure3.1 shows the mean and variance of SpC in 24-hr duration for modeled wells. We notice that for well 1-15, both mean and variance are relative small where the predictive models have a better performance. The variance of well 2-3 contains large mount of extreme values where both tested approaches have a worse performance. The boxplots for other time intervals are also attached at the end of this response letter.

We did also include other metrics like NSE and KGE, which take into account of the magnitude of variability in data, to better compare across different data variability.

Reviewer Comment 3.3 — There are not outliers in the boxplot of Figure S4.

Response:

Figure S4 is intended to exclude outliers (approximately 0.7% of data points) because such because these extreme values significantly zoom out the plot scale and make it almost impossible to see the range in the majority of data. This is explained in the figure caption.

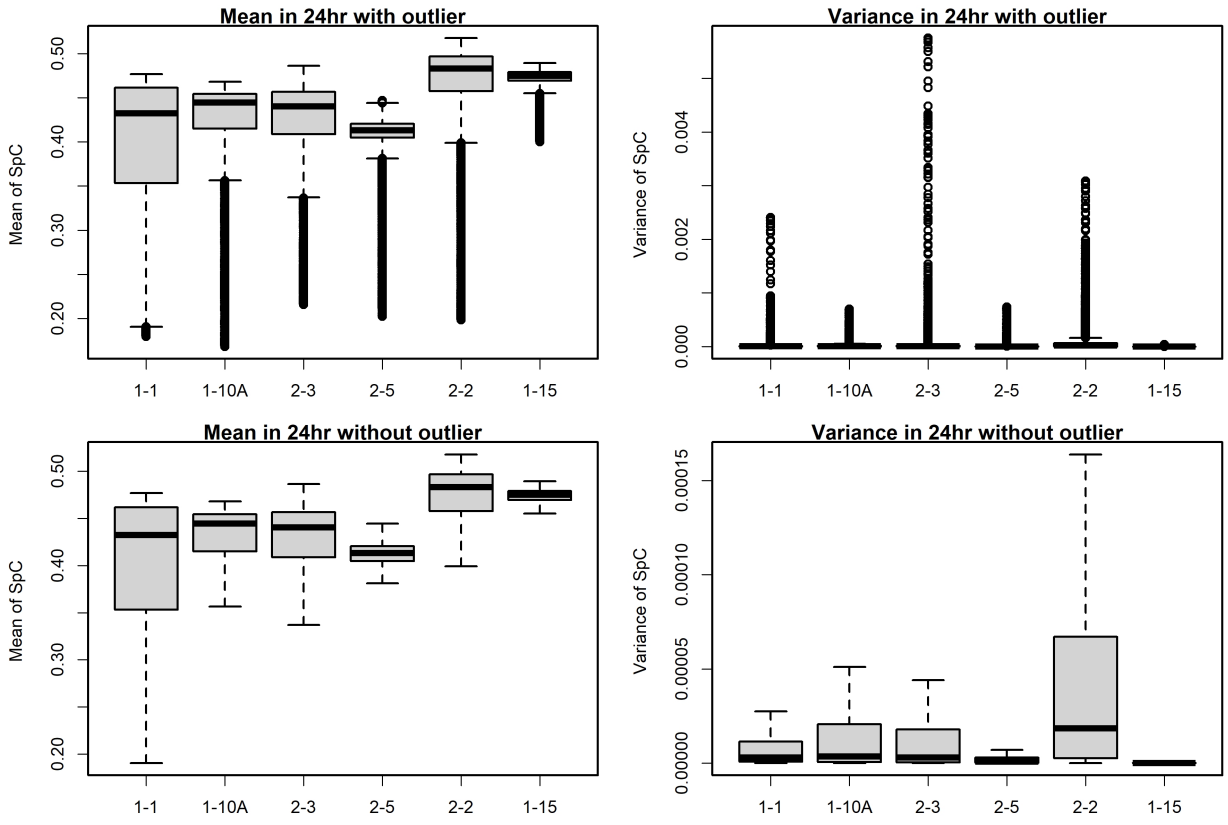

Figure 3.1: Boxplot of mean and variance of SpC at model testing period in 24-hour duration.

References

Siyong Chen, Xiaoyan Wang, Hui Guo, Peiyao Xie, and Abuobaida M Sirelkhatim. Spatial and temporal adaptive gap-filling method producing daily cloud-free ndsi time series. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13:2251–2263, 2020.

Steefan Contractor and Moninya Roughan. Efficacy of feedforward and lstm neural networks at predicting and gap filling coastal ocean timeseries: Oxygen, nutrients, and temperature. Frontiers in Marine Science, 2021.

Sivasathivel Kandasamy, Frederic Baret, Aleixandre Verger, Philippe Neveux, and Marie Weiss. A comparison of methods for smoothing and gap filling time series of remote sensing observations–application to modis lai products. Biogeosciences, 10(6):4055–4071, 2013.

Philipp Körner, Rico Kronenberg, Sandra Genzel, and Christian Bernhofer. Introducing gradient boosting as a universal gap filling tool for meteorological time series. Meteorologische Zeitschrift, 27(5):369–376, 2018.

Mikhail Sarafanov, Eduard Kazakov, Nikolay O Nikitin, and Anna V Kalyuzhnaya. A machine learning approach for remote sensing data gap-filling with open-source implementation: An ex-

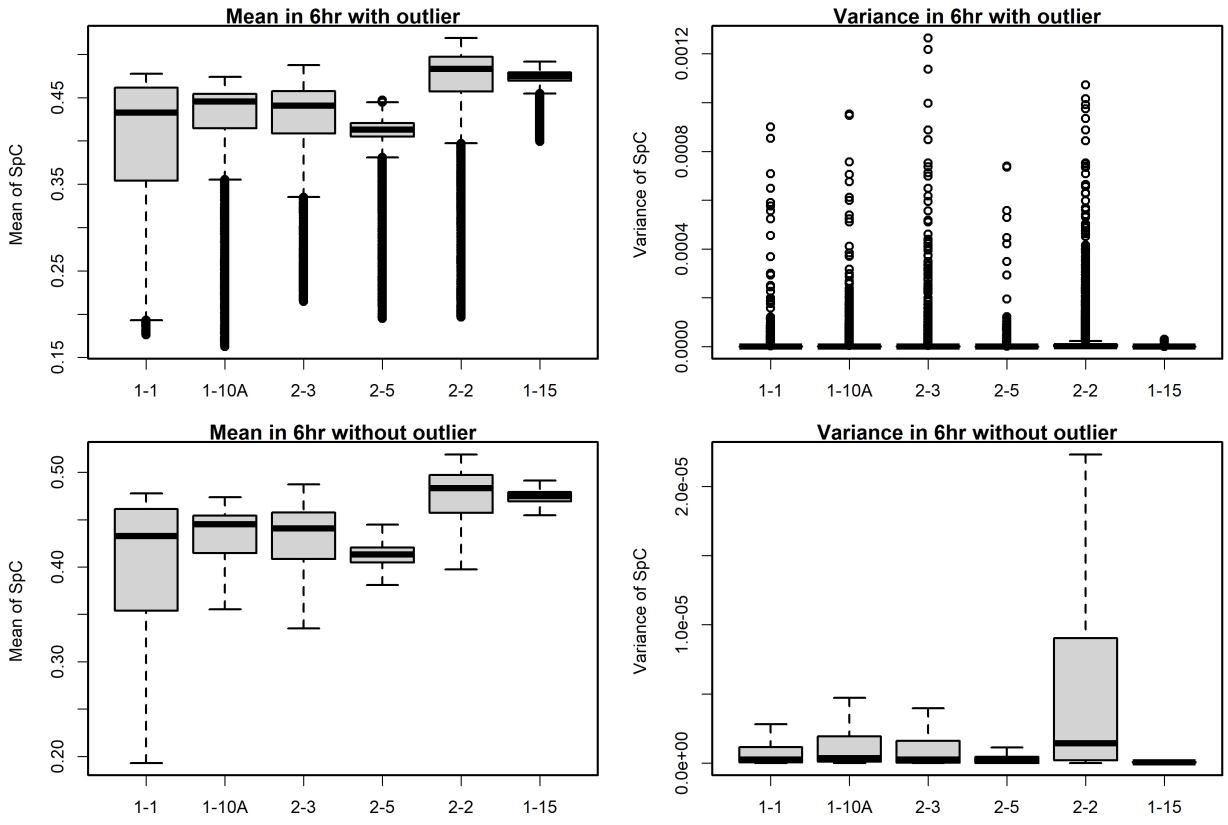

Figure 3.2: Boxplot of mean and variance of SpC at model testing period in 6-hour duration.

ample regarding land surface temperature, surface albedo and ndvi. Remote Sensing, 12(23): 3865, 2020.

Junbin Zhao, Holger Lange, and Helge Meissner. Gap-filling continuously-measured soil respiration data: A highlight of time-series-based methods. Agricultural and Forest Meteorology, 285:107912, 2020.

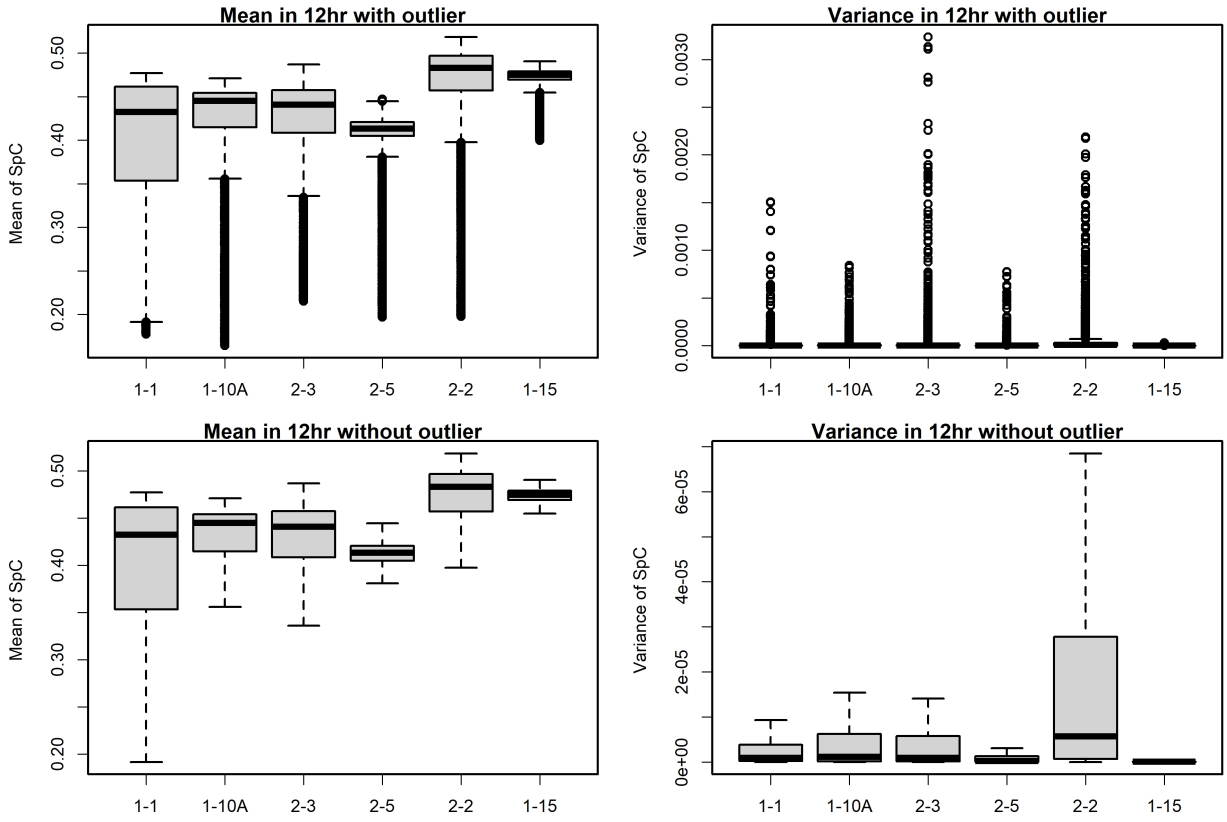

Figure 3.3: Boxplot of mean and variance of SpC at model testing period in 12-hour duration.

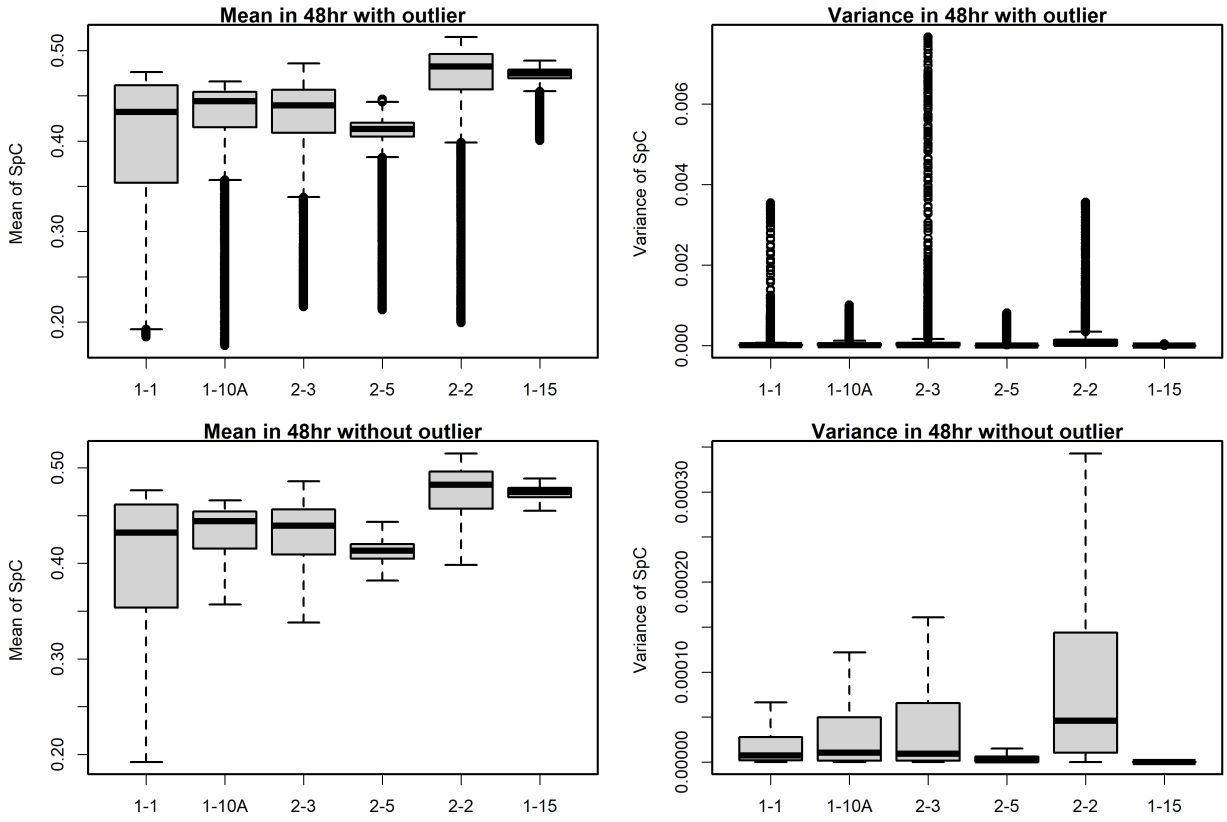

Figure 3.4: Boxplot of mean and variance of SpC at model testing period in 48-hour duration.

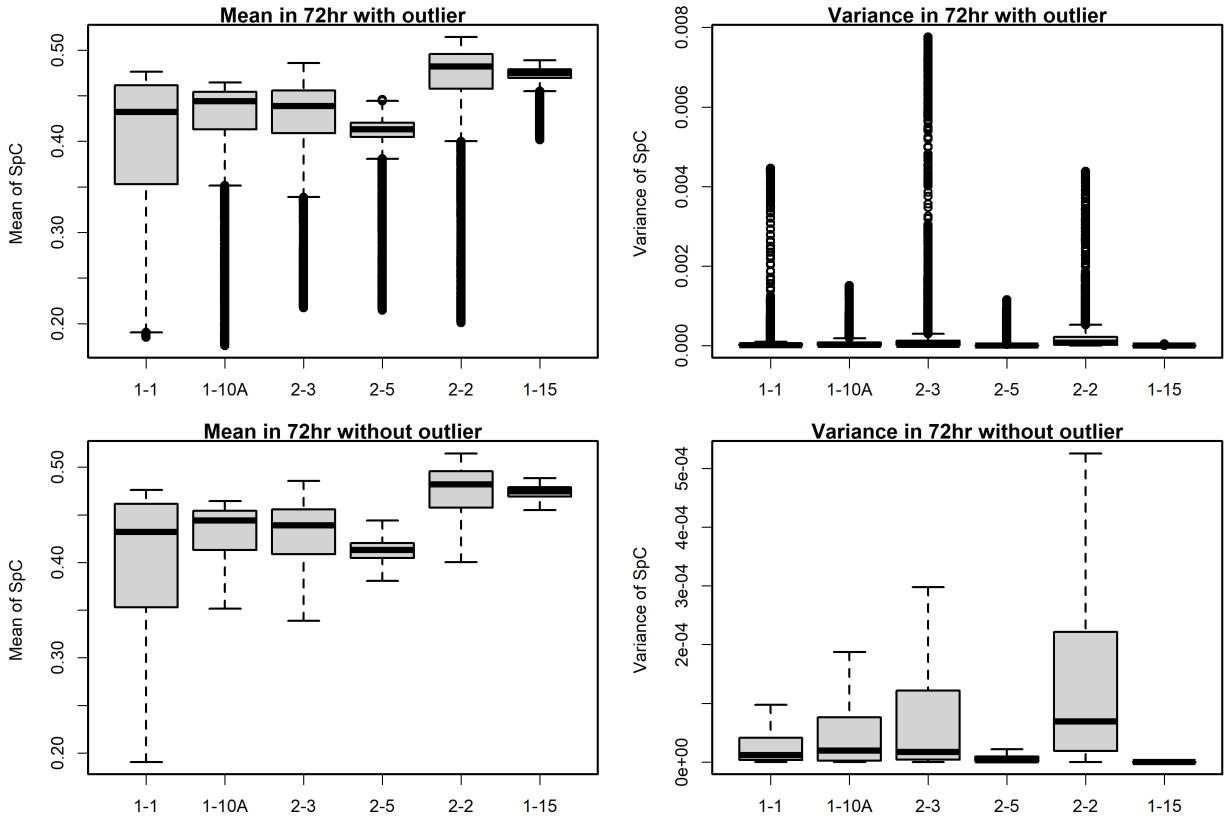

Figure 3.5: Boxplot of mean and variance of SpC at model testing period in 72-hour duration.