3rd Revision of "Technical note: Using Long Short-term Memory Models to Fill Data Gaps in Hydrological Monitoring Networks" by Ren et al.

I am happy to see that the authors tested a simpler model architecture, as suggested in previous reviews. The paper has become shorter and more concise, which is great. It is worth remembering that this is intended as a technical note, rather than a full paper, but there are still some open questions from my side.

1. Conceptually, I don't understand why the manuscript is framed as gap filling, when the setup of the model matches the common setup of forecasting (i.e. having historic observations + additional inputs to predict the next time step). From my point of view, gap filling is a task that is performed on historic data records where observations of both sides of the gap are available. This raises the question why you would not make use of this additional information, since interpolating between two points is most likely easier than predicting into the future. If the framing of the manuscript should indeed be focused around gap filling, I think it might be worth discussing the decision for running a gap filling model in this forecasting setup.

2. Why are ARIMA and LSTM not treated equally? That is, why does the ARIMA model predict all missing values at once, while the LSTM predicts only one time step at a time (and is then re-run for the next time step with the previous prediction filled into the input sequence). I think both models could be set up similarly and I wonder if you ever tested this and if the results suggested that this is the optimal setting for both models.

3. Looking at Figure 6: If you perform hyper parameter tuning and one model (almost) constantly picks the largest (or smallest) value, usually you should increase the search range, as this indicates that eventually even larger (or smaller) parameters would be better. Looking at e.g. the input window length of the ARIMA model.

4. To my understanding, part of this technical note is the benchmarking/comparison of two models, LSTMs and ARIMA. As such, I think this paper is still missing a statistical analysis of the modeling results. The entire discussion is currently focussed around a few plots and a textual description of what one can see in these figures. However, to a certain degree I would argue that this analysis/interpretation is rather subjective. I think it would benefit the paper to have a table that compares both models on a range of different metrics, including statistical tests of e.g. the robustness/significance of the results. Right now, I wonder what the takeaway message of this paper is. I would argue that it was probably known that both models, LSTMs and ARIMA, are generally capable of time series forecasting. If I would be a user with similar data or a similar problem, what is the additional knowledge that I can gain from reading this paper?

Minor comments:

- Figure 4: I think this figure is misleading to someone unfamiliar with LSTMs. You actually drew a fully connected network, rather than a recurrent (sequential) neural network. As of now, it seems like all inputs go into the LSTM at once (no time steps are visible in this figure), and the outputs of all time steps (since on the left side the timesteps are top to bottom) are used to predict the output. Figure (b) is actually the more correct depiction of the LSTM and I don't understand why both visualizations are needed.

- P 3, L 32 "*at the 300 Area of the U.S. Department of Energy Hanford site*". What is the 300 for?

- P 8, L 25: "*The well data were then pre-processed by normalizing all measurements via zero-mean and unit variance for each variable*". You do not normalize *"via"* zero-mean and unit variance, you rather normalize *to* zero mean and unit variance. Normalizing via zero-mean and unit-variance sounds like you subtract a mean of zero and divide by a variance of one, which is hopefully not what you have done.

- P9 L 7ff You use plural for "multi-well models" throughout this passage but you only trained one multi-well model to predict at well 1-1, or?

- P17 L 17: Which "auto.arima" function?