Dear Anonymous Referee #2,

Hereby the authors of the revised paper hess-2020-382 take the opportunity to thank you for the useful comments and suggestions for improving our manuscript.

Please find below our responses to each of your comments. Your comments are marked in ***black italic,*** and our responses are provided in regular font.

In their paper, Ma et al proposed to investigate the link between water table depth anomalies (wtda) and precipitation anomalies (pra) using Long Short-Term Memory networks (LSTM). To test the proposed approach, they use a dataset generated with the Terrestrial System Modelling Plateform (TSMP) over Europe and compare the results provides by both approaches (TSMP and LSTM). The effect of several factors on the performance of the approach are also investigated. Cross-wavelet transform are also used to analyze the response of the network regarding time frequency.

Thank you for the cogent summary of our study and the thorough review.

Overall, the paper is well written and organized. The approach proposed is interesting and its novelty is clearly explained in the introduction as this type of networks used is not commonly used to examine the response of groundwater. The study has a specific focus on response to drought which is of importance for groundwater management.

Thank you for the positive feedback.

The results presented are promising but the presentation/discussion should be improved. In my opinion, the results are not discussed thoroughly especially when the performance of LSTM is not so good. My issues with the paper, and some additional minor comments and corrections, are detailed in the following.

In the discussion of the results, we explained the low LSTM performance through the weak physical link between pr_a and wtd_a and the time-varying pattern between them. We will further improve the presentation and discussion by incorporating your comments and suggestions in the revised version to address your concerns.

-The dataset generated with TSMP is the foundation of the proposed methodology as the evolution in time and space of all the variables used in the study are simulated ones. Although the reader is sent to relevant references to have further information, I think some key features need to be presented to make the paper self-consistent.

We agree and, therefore, will include key features of the TSMP in the revised manuscript.

I especially would have liked to know how the TSMP was calibrated (or not) against observed values to have an idea of how reasonable or relevant the simulated evolutions are.

The TSMP simulation results, especially anomalies, show high agreement with in situ and satellite observations (i.e., E-OBS v19 and GRACE datasets), as presented in Furusho-Percot et al., (2019). We had a sentence in our manuscript (Line 200-204) to address your concern about the performance of TSMP, and the reader is referred to related references [e.g., Furusho-Percot et al., (2019)] for detailed information.

- I think that the figures with maps are very hard to interpret owing to the extension of the study area and the spatial resolution of the approach proposed. The authors state that the agreement is good visually (Line302) which is in my opinion not so evident and not enough. I would have liked (if possible) some indicators to be presented – maybe for each PROVIDENCE regions – to have a quantitative diagnostic rather than a visual one.

The maps (e.g., Figure 5) show the spatial extent and severity of groundwater drought over Europe in a specific month, which is indicated by different colors. This type of presentation is typical for the spatial analysis of a drought event [see e.g., Shukla and Wood (2008), Gumus and Algin (2017), and Van Loon et al. (2017)]. In the interpretation, we focused on the agreement of the spatial patterns with respect to severity. The visual comparison between the two maps in Figure 5b shows good agreement (Line 302). The plots of R^2 values as a function of local factors such as yearly averaged *wtd* (Figure 6) have provided a quantitative diagnosis of the network performance for each PRUDENCE region.

- Overall, the performance of the LSTM approach is not discussed with enough details. Especially when the performances are poor. Line334–335 is an example where some more details are needed. Table 4 demonstrates that the agreement is not good in some specific PRUDENCE regions (for instance MD or IB) and no specific explanations are provided. The same goes for the discussion of Figure 8.

Thank you for pointing this out. We will improve the discussion by adding more details, especially for the scenarios having poor performances.

- The conclusion is a bit misleading, as it may convey the message that the LSTM approach is relevant all over Europe when the results are very good only in specific conditions (as specified line 406-407). Some rephrasing may be needed here.

We agree and will rephrase the related section in the revised manuscript for clarity.

Specific comments:
- Line 84: Should be RNN and not ANN here

LSTM networks are a special type of RNNs, and RNNs belong to ANNs, so it is also correct to state that LSTM networks are ANNs. Further, “The consistency of the temporal pattern between input and target variables” is required by all ANNs for good performance. Therefore, we stated ANNs rather than RNNs in Line 84. We will clarify the relationship among ANNs, RNNs and LSTM networks in the revised manuscript.

- Line 110: It is mentioned here that “Areas with surface water are not taken into account”. I wonder if or to what extend this assumption could impact the results of the study.

This assumption does not affect the results in this study, because we constructed the proposed LSTM networks only at pixels without rivers and lakes.

- Figure 6 can be improved: the color legend that specifies the PRUDENCE regions should be bigger and placed elsewhere.

We will improve this figure following your suggestions.

Reference:

Furusho-Percot, C., Goergen, K., Hartick, C., Kulkarni, K., Keune, J. and Kollet, S.: Pan-European groundwater to atmosphere terrestrial systems climatology from a physically consistent simulation, Sci. data, 6(1), 320, 515doi:10.1038/s41597-019-0328-7, 2019.

Gumus, V. and Algin, H. M.: Meteorological and hydrological drought analysis of the Seyhan−Ceyhan River Basins, Turkey, Meteorol. Appl., 24(1), 62–73, doi:10.1002/met.1605, 2017.

Shukla, S. and Wood, A. W.: Use of a standardized runoff index for characterizing hydrologic drought, Geophys. Res. Lett., 35(2), 1–7, doi:10.1029/2007GL032487, 2008.

Van Loon, A. F., Kumar, R. and Mishra, V.: Testing the use of standardised indices and GRACE satellite data to estimate the European 2015 groundwater drought in near-real time, Hydrol. Earth Syst. Sci., 21(4), 1947–1971, doi:10.5194/hess21-1947-2017, 2017.