Dear Zhongbo Yu, editor of HESS,

We would like to thank you and anonymous Referees #1 and #4 for the constructive comments and suggestions on our manuscript.

We addressed all the concerns of you and the reviewers in detail, and revised the manuscript accordingly. A marked-up manuscript version is provided, showing all the modifications that we have made.

Here we respond to the comments of you and anonymous Referees #1 and #4 in detail, which is followed by information regarding the additional modifications we made in the manuscript. Your and the reviewers' comments are marked in ***black italic,*** and our responses are provided in regular font.

Response to your comment:

Two reviewers provide positive comments on the paper. Authors are required to address these comments on the revision.

Thank you for the comment. We have addressed all the comments in detail and made corresponding changes to the revised manuscript. For details, please consider our responses to the reviewers' comments below.

Response to Referee #1's comment:

Thank you for the recommendation to accept our manuscript.

Responses to Referee #4's comments:

Recommendation: Major Revisions

Thank you for reviewing our manuscript.

In this manuscript, the authors tried to connect the water table depth anomalies to precipitation anomalies over Europe using Long Short-Term Memory networks based on the daily terrestrial simulations of the Terrestrial Systems Modeling Platform (TSMP) over Europe. The proposed method has the ability to reproduce the TSMP-G2A wtda maps. The authors also hypothesize that the proposed networks could simulate the high-frequency components of wtda, This manuscript is clear in construction and easy to follow.

Thank you for the cogent summary of our study and the positive feedback.

However, there are several concerns which need to be well addressed before the paper could be accepted.

Major Concerns:
1. What are the advantages of LSTM networks over TSMP, since TSMP could generate daily wtda time series. The authors need to give more justifications on LSTM networks.

LSTM networks require less computation time and background knowledge compared to physically-based models such as TSMP, as stated in Line 145-146 in Version 3. When the

proposed LSTM networks are available, they can provide fast and reliable predictions for groundwater table depth anomalies (wtd_a) over Europe only based on data of precipitation anomalies (pr_a), which is not possible for TSMP.

In Line 467-468 in the marked-up manuscript version, we added "After training, LSTM networks could provide fast and reliable predictions of wtd_a only based on data of input variables, which is impossible for traditional physically-based models such as TSMP." for clarity.

2. The LSTM networks were trained based on the simulated outputs rather on the observed results. In my opinion, the authors should train the networks using observed data and compared the simulated results with other numerical models to demonstrate the capability of the proposed method.

We agree, but as stated in Line 48-49 in Version 3, there are no spatiotemporally continuous groundwater table depth observations available over the European continent to train the LSTM networks, therefore the need to evaluate the proposed LSTM networks based on simulation results. Because TSMP-G2A data set is in good agreement with hydrometeorological and GRACE observations in different European regions [Furusho-Percot et al. (2019) and Hartick et al. (2021)], we argue that the TSMP-G2A data set is a useful reference data set to establish the methodology (see Line 213-221 and Line 458-461 in Version 3).

In Line 460 in the marked-up manuscript version, we added "Due to a lack of spatiotemporally continuous *wtd* observations over Europe," to state the reason for evaluating the proposed methodology based on simulation results in this study.

3. The authors hypothesize that the proposed networks could simulate the high-frequency components of wtda by comparing the results from cross-wavelet transform analysis made on the outputs from both LSTM and TSMP. I am confused why the authors use cross-wavelet transform analysis to evaluate the performance, I don't think the XWT is a good method for evaluation. The authors could provide more explanations. Meanwhile, the main concern may be the ability to simulate low frequency variations of wtd caused by extreme events such as long-term drought.

XWT is not used as an evaluation method in this study. XWT was applied to visualize the pattern changes between TSMP-G2A pr_a and wtd_a in the time-frequency domain to explain the network performance C2 (i.e.., training R^2 score $\geq$ 50%, test R^2 score $\leq$ 0%) at some pixels, as mentioned in Line 88-89, Line 108-109, Line 191-195 and Line 406-408 in Version 3. Additionally, we discovered that the LSTM networks tended to gain good R^2 scores at the pixels with high power concentrated in the period from 2 to 16 months in the XWT spectra for TSMP-G2A pr_a and wtd_a series, and thus, hypothesized that LSTM networks might be frequency-aware and work well to capture the pr_a-wtd_a relationship at the monthly, seasonal and annual periods. The "high-frequency components" may cause confusion, because the period of 16 months is already longer than the annual period, and we replaced the terminology in the revised manuscript. For large periods in the XWT spectra, large areas suffer from edge effects and cannot be used for analysis.

For clarity, we made the following modifications in the marked-up manuscript version:

- Line 431-432, deleted ", which are the components of the time series with high frequency";

- Line 433-434, replaced “on high-frequency components” with “to capture the pr_a-wtd_a relationship at the monthly, seasonal and annual periods”;
- Line 455-456, replaced “on high-frequency components” with “to capture the pr_a-wtd_a relationship at the monthly, seasonal and annual periods”.

Minor coments:
1. Line 113-114: no confined aquifer?

Yes, here we only focused on unconfined aquifers where groundwater is expected to have a close connection with precipitation.

2. For Figure 5, the authors may provide a anomaly map between wtda from TSMP-G2A and LSTM as it is really hard to see the difference.

Figure 5 shows the comparison of European wtd_a maps from the TSMP-G2A data set and the LSTM networks for August 2003 and August 2015 with respect to the severity of drought. Color scales represent different degrees of groundwater drought (see Line 247-249 in Version 3). We focused on the distribution of the areas where $wtd_a \geq 1.5$ (i.e., a strong drought) and found good agreement between the wtd_a maps from the TSMP-G2A data set and the LSTM network results. Such information cannot be gained from wtd_a difference maps. In addition, Figure 7 (the map of test R^2 scores achieved by the proposed LSTM networks over Europe) already provided insights into the network performance at the spatial scale. Therefore, we argue that there is no need to provide wtd_a difference maps in the manuscript.

3. The legend of Figure 6 should be put at right position.

Thank you for the suggestion. We moved the legend of Fig.6 to the bottom.

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.

Hartick, C., Furusho-Percot, C., Goergen, K. and Kollet, S.: An Interannual Probabilistic Assessment of Subsurface Water Storage Over Europe Using a Fully Coupled Terrestrial Model, Water Resour. Res., 57(1), doi:10.1029/2020WR027828, 2021.

Additional modifications (in the marked-up manuscript version):

1. In Line 4, replaced “Institute of Bio and Geosciences (Agrosphere, IBG-3)” with “Institute of Bio and Geosciences: Agrosphere (IBG-3)”;

2. In Line 13 and Line 92, replaced “spatio-temporally” with “spatiotemporally”;

3. In Line 508-509, changed “Cell’s” and “Output” to “cell’s” and “output”, respectively.

Best regards,
Yueling MA

On behalf of all the authors