The groundwater level changes in the seasonal frozen soil region are simulated using interpretable deep learning, while the underlying mechanisms of groundwater level dynamics during the freezing and thawing periods as well as non-freezing and thawing periods are revealed. The topic is interesting and the research results can provide a reference for the assessment of groundwater resources in seasonal frozen soil regions. However, considering that the Hydrology and Earth System Sciences is the world's premier journal publishing research of the highest quality in hydrology, it could not be accepted before a minor revision.

Response: Thank you very much for your appraisal of our work and encouragement. We have studied comments carefully and have made correction which we hope meet with approval.

1）During the freeze–thaw process, the groundwater level exhibits a noticeable lag during the recovery phase. Has the author considered the physical mechanisms behind this lag, such as delayed soil thawing or the blockage effect of frozen layers?

Response: We sincerely thank the reviewer for the insightful and professional comments. As noted, we indeed observed a significant lag between the rise in air temperature and the corresponding rise in groundwater levels during the freeze–thaw period. Our analysis suggests that this phenomenon is closely related to the staged thawing process of the frozen soil in the study area and its hindering effect on vertical water movement.

Specifically, based on meteorological and soil temperature monitoring data from 2018 to 2021 in the study area, the thawing period can be roughly divided into three stages. In the first stage (around late February), air temperature first rises above 0 °C, initiating snowmelt; however, due to diurnal freeze–thaw cycles, the frozen soil has not yet completely thawed, and infiltration of liquid water is impeded. In the second stage (around mid to late March), air temperature remains steadily above 0 °C, the frozen soil gradually thaws through, and water begins to infiltrate into the aquifer, leading to a more rapid rise in groundwater levels. In the third stage (around mid-April), the frozen soil is completely thawed, and vertical water pathways become fully open. Due to the staged nature of frozen soil thawing, during the early phase of air temperature increase—i.e., between the first and second stages—although snowmelt water is present, residual frozen layers within the soil profile form a distinct “water-blocking layer” that inhibits downward infiltration and aquifer recharge. Even if liquid water appears in the shallow soil, it cannot directly recharge the groundwater. Only after the complete thawing of the frozen soil can continuous vertical flow pathways be established, allowing for a noticeable rise in groundwater levels.

Based on existing data, this study provides a preliminary explanation of the observed lag phenomenon. In future work, we plan to incorporate parameters such as frozen soil thickness, hydraulic properties, and soil profile structure to conduct quantitative analysis and further elucidate the response mechanisms of groundwater dynamics to freeze–thaw processes.

2）Line 314 mentions that "there is no significant lag between the simulated and observed values." Has any correlation or lag correlation analysis been conducted to support this statement?

Response: We thank the reviewer for the valuable comments. In response to the statement regarding the absence of an obvious lag between the simulated and observed values, we have added Pearson correlation analysis for the four representative monitoring wells shown in Figure 4. The revised content is as follows:

“The LSTM model is capable of accurately capturing the variation trend of groundwater levels, and no significant lag is observed between the simulated and observed values (Figure 4). The Pearson correlation coefficients at the four representative monitoring wells shown in the figure are 0.86, 0.81, 0.87, and 0.85, respectively. Moreover, the correlation coefficients reach their maximum values without applying any time lag, indicating that the simulated values can effectively and promptly reflect the actual variation trend of groundwater levels.”

3）Line 211 states that 150 days of meteorological variables were used as model input. What is the basis for selecting this window length? Has other time lengths been tested for their effect on model performance?

Response: We thank the reviewer for raising this professional and important question. The choice of a 150-day input window for meteorological variables was based on the following considerations:

First, the effects of meteorological changes on groundwater dynamics often require a relatively long period of transmission and accumulation, which short windows may fail to capture adequately. Based on a time-series analysis of long-term groundwater and meteorological data in the Songnen Plain, we found that a span of approximately five months (150 days) can effectively reflect the combined influence of temperature, precipitation, and other variables on groundwater level fluctuations.

Second, we conducted comparative experiments using different window lengths of 90, 120, 150, and 180 days for model training and testing. The results showed that the 150-day window outperformed the others in terms of root mean square error (RMSE), coefficient of determination (R^2), and Nash–Sutcliffe efficiency (NSE). It also avoided the computational inefficiency and potential overfitting issues associated with excessively long windows.

Based on the above analysis, we selected a 150-day input window to strike a balance between physical interpretability and model performance, thereby ensuring the scientific validity and stability of the prediction results.

4） How is the early stopping strategy for the LSTM model set?

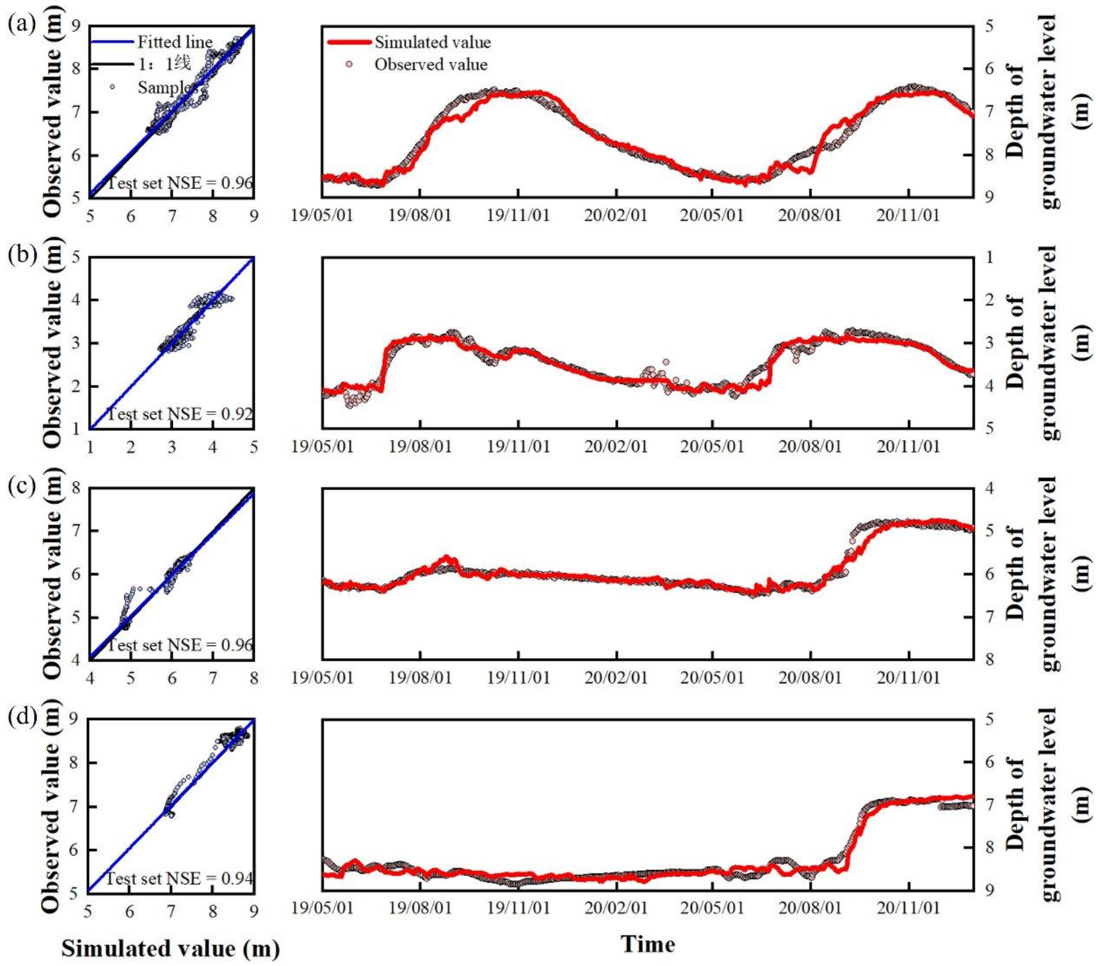
Response: To further improve training efficiency and effectively prevent overfitting, two strategies were adopted during the training of the LSTM model: early stopping and adaptive learning rate adjustment. The coefficient of determination (R^2) on the validation set was uniformly used as the performance evaluation metric.

For the early stopping strategy, training was automatically terminated when the improvement in R^2 on the validation set remained below a predefined threshold (0.01) for 30 consecutive training epochs, indicating that the model had converged. This approach helps conserve computational resources and avoid overfitting. After termination, the model parameters were restored to the state corresponding to the epoch with the highest R^2 on the validation set, ensuring optimal generalization performance.

In addition, to enhance precision optimization in the later stages of convergence, an adaptive learning rate decay mechanism was introduced. Specifically, when the validation R^2 did not improve by more than 0.01 for 15 consecutive epochs, the current learning rate was automatically reduced to half of its previous value, slowing down parameter updates and improving the model's ability to search near the optimum. To prevent the learning rate from becoming too small and stalling the training process, a minimum learning rate limit was set at 1% of the initial value.

5）It is recommended to include comparison plots for typical sites with NSE > 0.7 in the test set, to contrast with the low-performance sites in Figure 4 (NSE < 0.7), and to further validate the model’s applicability and stability across different locations.

Response: We thank the reviewer for the suggestion. As recommended, we have added comparison plots for typical sites with NSE > 0.7 in the test set, to contrast with the low-performance sites in Figure 4. The supplementary figure is as follows:



Comparison between simulated and observed groundwater table depths at typical sites in the study area with test set NSE > 0.7.

6）When using the EG method to calculate the importance of influencing factors, have you considered converting the EG scores into percentages to more clearly display the dominant factors and their relative contributions at different periods for the same groundwater level dynamic type?

Response: We thank the reviewer for the valuable suggestion. During our study, we did consider converting the EG scores into percentage form to provide a more intuitive representation of the relative importance of different variables. However, the EG method outputs a time series that reflects the influence of each input variable on the model's prediction at each time step. Even after converting the EG scores into percentages, they remain a time-dependent sequence rather than a single static value, making it difficult to comprehensively assess the overall importance of each variable throughout the entire prediction period. Moreover, the importance of different variables varies significantly across different time periods. Simple normalization or averaging may obscure the contribution of certain variables during critical periods. Therefore, we chose to retain the original time series of EG scores and performed qualitative analyses based on representative time intervals, in order to better reflect the stage-specific dominant mechanisms affecting groundwater level dynamics.

7）The manuscript refers to the “initial groundwater level depth at the start of the freezing period.” How is the time point of this variable consistently defined? Is it synchronized with the time when the maximum freezing depth occurs?

Response: The term "initial groundwater level depth at the beginning of the freezing period" in the manuscript refers to the groundwater level depth corresponding to the early stage of the freezing period each year—specifically, when air temperature consistently drops below 0 °C and surface freezing first occurs. To standardize the time point across different years and monitoring sites, we adopted the solar term “Lidong” (approximately November 7–8) as a unified indicator for the onset of the freezing period. The groundwater level depth on the day of Lidong was extracted and used as the initial groundwater level depth. This time point was determined based on the regional climatic patterns of the study area and does not coincide with the occurrence of the maximum freezing depth. According to previous studies and our local observations, the maximum freezing depth typically occurs in the later part of winter (e.g., from late January to mid-February), lagging behind the beginning of the freezing period.

8）In line 691, the conclusion states that a “V-shaped” groundwater level trend indicated a significant influence of the soil freeze–thaw process on the groundwater level. However, the specific causes of the V-shaped dynamics are not clearly explained.

Response: We sincerely thank the reviewer for the valuable suggestions. In response, we have revised the conclusion section as follows:

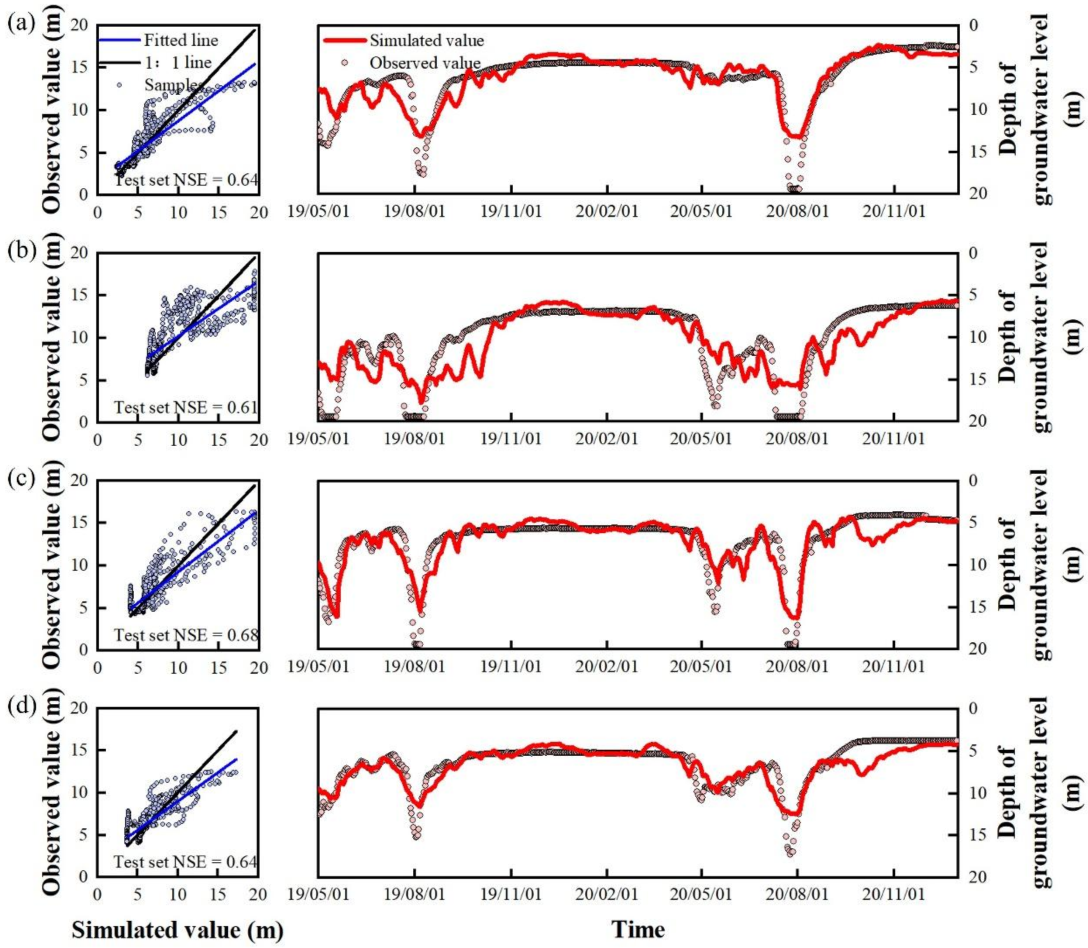
“During the freeze–thaw period, changes in the soil water potential gradient due to freezing and thawing lead to interactions between soil water and groundwater, resulting in the V-shaped variation. In contrast, the continuously declining and rising types reflect gradual water level changes primarily driven by groundwater extraction and precipitation recharge, without strong influence from freeze–thaw processes."

9）In line 222, the formula should be revised to: $f_t = \sigma\left(W_{xf}x_t + W_{hf}h_{t-1} + b_f\right)$

Response: Thank you for pointing out the error in the formula. As suggested, we have corrected the formula in line 222 accordingly.

10）It is recommended to display the specific NSE value of the representative site in the western low plain region within the test set in Figure 4.

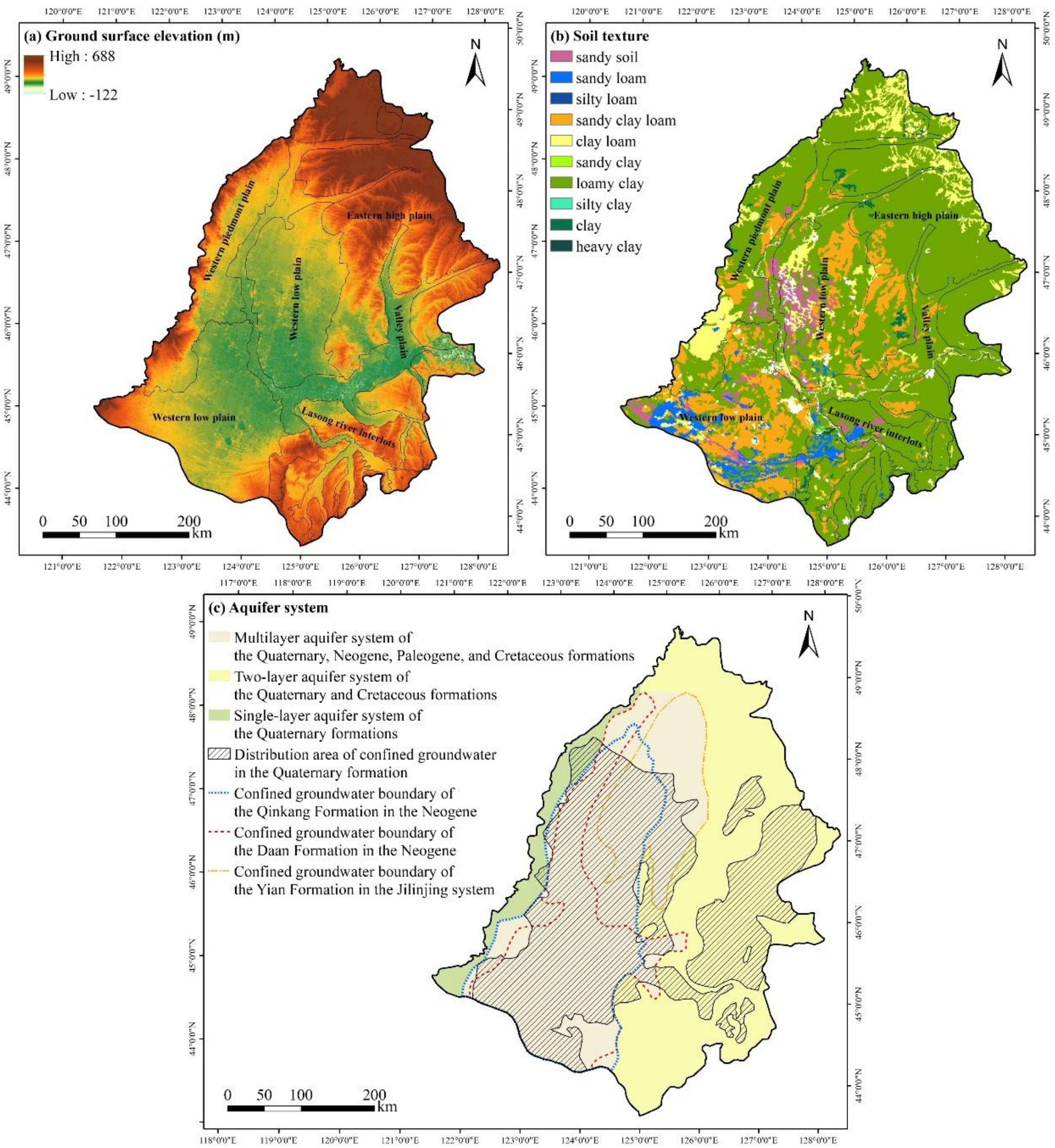
Response: We sincerely thank the reviewer for the valuable suggestion. We have added the specific NSE values for representative sites in the low western plain area within the corresponding section of Figure 4. The revised figure is as follows:



Comparison of the simulated and observed groundwater level depths at typical points in the western low plain (NSE values on the test set < 0.7).

11）It is suggested to delete Figure 2d and merge Figure 2b with Figures 2a and 2c.

Response: We sincerely thank the reviewer for the valuable suggestion. In response, we have revised Figure 2 as suggested. The updated figure is as follows:



Spatial distribution of the ground surface elevation (a), topography (b) and aquifer system (c) in the Songnen Plain, China.