This manuscript applies a machine learning (ML) approach to predict time-varying groundwater levels in seasonally freezing regions of China. The topic is timely and of high importance for groundwater resource management and environmental protection. However, the study overlooks several critical factors that could significantly influence the results and interpretations. By incorporating additional hydrogeological and environmental variables, the model's accuracy could be greatly improved, leading to a more comprehensive understanding of groundwater dynamics.

Response: Thank you for your valuable comments on our study. We have carefully reviewed your suggestions and made corresponding revisions, and we hope these modifications meet your expectations. We agree that key factors such as hydrogeological conditions and environmental variables may significantly influence the model outputs and their interpretation. However, the core focus of this study is on building LSTM models for each monitoring site individually, aiming to simulate the temporal variation of groundwater level at the point scale. Within this framework, spatially fixed attributes such as aquifer properties and topography remain relatively stable over the short term and are unlikely to exert dynamic influence on the time series at a single site. Additionally, factors such as vertical leakage and surface water interactions are difficult to quantify due to limited data availability. In future work, if data conditions permit, we will consider incorporating these variables to enhance the physical interpretability and predictive accuracy of the model.

Specific comments:

Line 25: Please define NSE upon first mention to ensure clarity for readers unfamiliar with the metric.

Response: Thank you for pointing out this issue. We have added the full term “Nash-Sutcliffe Efficiency” when “NSE” first appears in the abstract.

Line 39: Provide more detailed justification of why monitoring groundwater levels is crucial, not only for managing water resources but also for protecting ecological systems. Additionally, consider using the ML-predicted results to present a case study with quantitative analysis to better illustrate the implications.

Response: Thank you for the valuable comments from the reviewer. Following the suggestions, we have comprehensively revised the relevant parts of the manuscript to further elaborate on the importance of groundwater level depth, especially emphasizing its role in water resource management and ecosystem protection. Additionally, we supplemented the citations with the study by Liu et al. (2022), which used machine learning to predict groundwater level depth in the lower Tarim River, providing a quantitative case validation of the practical significance of groundwater level prediction. The revised content is as follows:

“Groundwater level is a crucial indicator reflecting the water balance status of groundwater systems, and its dynamic changes reveal the evolving trends of regional hydrological processes. In terms of water resource management, monitoring groundwater level depth helps managers understand changes in groundwater storage, optimize water extraction schemes, and prevent resource depletion caused by overexploitation (Hao et al., 2014; Yang, 2012). Regarding ecosystem protection, fluctuations in groundwater level depth directly affect regional ecological patterns. Excessively low water levels may lead to wetland desiccation and biodiversity loss, while rapid rises can cause soil salinization and vegetation degradation (Singh et al., 2012). Relevant studies have also practically validated the significance of groundwater level prediction. For example, Liu et al. (2022) demonstrated in the lower Tarim River that machine learning–based groundwater level prediction models can quantitatively reveal current and future groundwater changes, clarifying the critical role of ‘ecological water conveyance’ in regional ecological restoration. Therefore, in-depth identification of the controlling mechanisms behind groundwater level depth variations and achieving high-precision spatiotemporal simulation are of great significance for promoting sustainable groundwater resource utilization and ecological environment protection.”

Lines 62–67: The key disadvantage of physical models, compared to ML models, lies in their time-consuming setup, calibration, and validation processes. However, physical models have the advantage of offering more mechanistic insight into underlying hydrological processes, which ML models often lack.

Response: We thank the reviewer for highlighting the insufficient discussion on the comparison between physical models and machine learning models in the current manuscript. In response to your suggestion, we have revised and supplemented the relevant content accordingly. In the updated version, we have clearly stated the advantages of physical models in revealing the physical mechanisms of hydrological processes, while also acknowledging their limitations in regions with complex geological conditions due to high modeling complexity and substantial data requirements. The revised content is as follows:

"Current models used for simulating groundwater level dynamics can generally be categorized into two groups: physical models and machine learning models (Ao et al., 2021). Most physical models are based on hydrodynamic processes and water balance principles, and are capable of accurately representing the physical mechanisms of groundwater systems. Therefore, they possess irreplaceable advantages in characterizing groundwater flow and uncovering hydrological processes such as recharge, runoff, and discharge. However, in areas with complex geological structures or highly heterogeneous aquifer systems, the construction, parameter calibration, and validation of physical models typically require large amounts of high-resolution geological, hydrological, and hydraulic data. These requirements make physical modeling challenging to implement and time-consuming (Raghavendra N and Deka, 2014)."

Line 118: The model would benefit from incorporating a wider range of influencing factors, such as aquifer properties, topography, hydraulic conditions (e.g., lateral flow, vertical leakage, groundwater storage, surface water interactions), and anthropogenic variables like population density. Spatial heterogeneity in evapotranspiration and precipitation should also be considered to improve model realism.

Response: We sincerely thank the reviewer for the professional and insightful comments. We fully agree that a variety of natural and anthropogenic factors—such as aquifer properties, topography, groundwater dynamics, population density, and evapotranspiration—can exert significant influence on regional groundwater level changes.

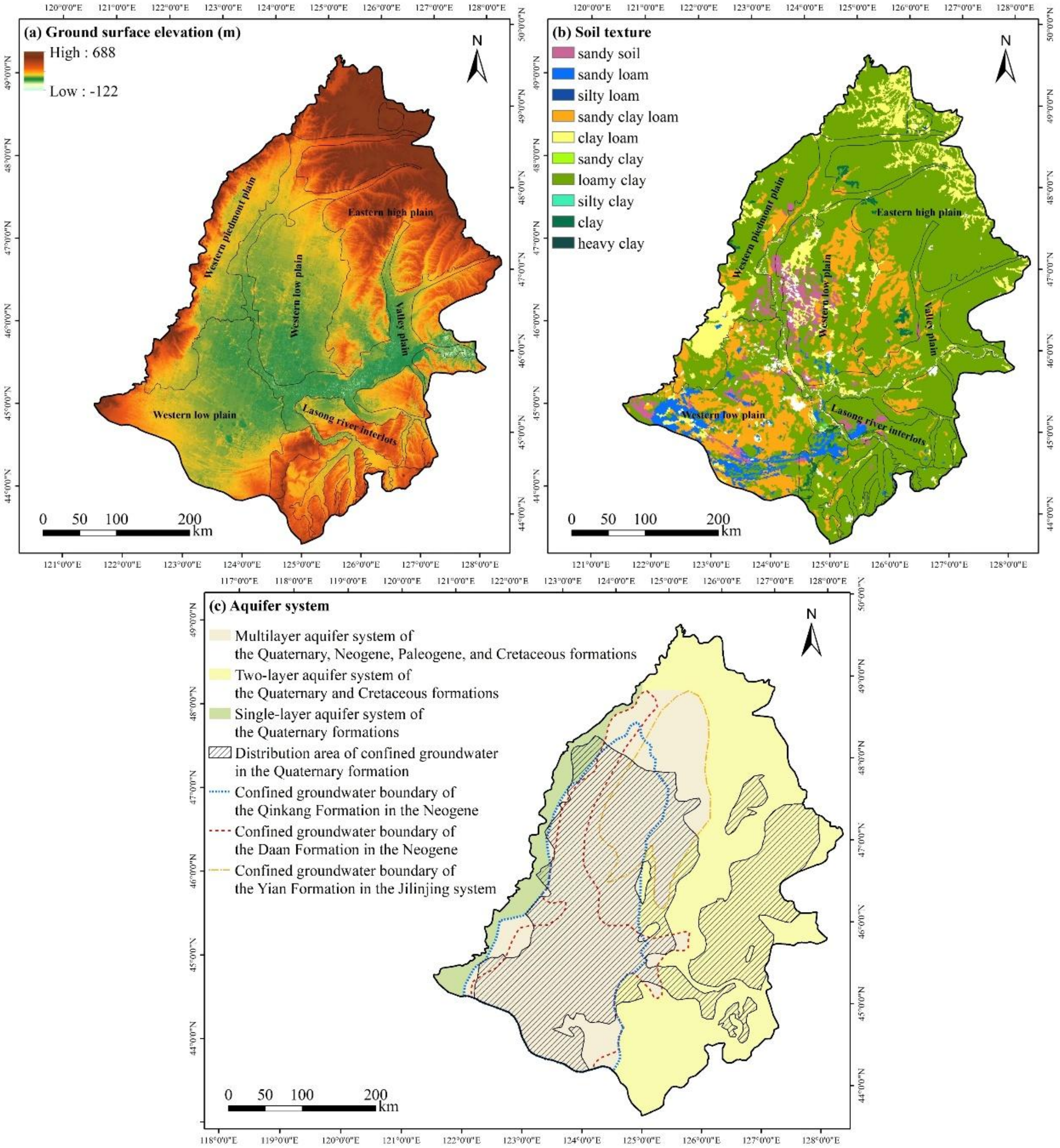
However, considering the design rationale and actual data availability in this study, we have carefully reflected on and responded to this point from the following two perspectives:

First, the core framework of our study is to independently construct an LSTM model for each monitoring well to simulate the temporal variation of groundwater level at that specific location. The model uses historical meteorological variables and anthropogenic dynamic factors (including air temperature, precipitation, snow depth, and groundwater extraction) as inputs, aiming to capture the nonlinear response relationship between these temporally dynamic factors and groundwater level changes. Under this modeling strategy, spatially static attributes such as aquifer properties and topography remain constant over short periods at a given site and thus cannot provide dynamic explanatory power for the temporal evolution of groundwater levels at that point. Additionally, the spatial heterogeneity of factors such as evapotranspiration and precipitation primarily influences regional-scale patterns or spatial distributions. Since our study focuses on site-specific time series modeling and identification of dominant influencing factors, it is relatively less dependent on spatially heterogeneous variables. We have clarified this limitation in Section 3.5 “Model Limitations” of the revised manuscript.

Second, regarding the absence of variables related to groundwater dynamics (e.g., lateral flow, vertical leakage, and surface–groundwater interactions), we fully acknowledge their critical roles in groundwater system evolution. Although in theory, groundwater flow fields could be constructed through spatial interpolation of observed water levels, in our study the groundwater level is the target output variable of the model. Thus, prior to obtaining the model predictions, it cannot serve as an input driver. Moreover, in practice, there is a lack of independent observational data (such as hydraulic gradients or recharge–discharge rates) that directly reflect groundwater dynamics, making it currently unfeasible to incorporate these factors into the model. In future work, if data availability improves, we intend to include such variables as key supplementary inputs to enhance the model's physical interpretability.

Figure 2: Consider including a geological map that shows the distribution of geological formations or aquifer types. This would help contextualize the results spatially.

Response: We thank the reviewer for the suggestion. In response, we have added a new subfigure to Figure 2 showing the distribution of the aquifer system. The revised figure is as follows:



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

Figure 4: The observed and simulated groundwater levels do not align well; the simulated series appears overly variable. Please explain the possible causes of this discrepancy, such as overfitting, lack of key input variables, or limitations in the model's temporal resolution.

Response: We thank the reviewer for the valuable comments. Although the original manuscript included an explanation for the poor model performance at certain monitoring wells, the reasoning lacked clarity and failed to accurately convey the sources of model error. In response, we have revised and reorganized the relevant paragraph to enhance its logical structure and coherence. The modified version is as follows:

“Only 18.11% of the monitoring wells in the study area had a Nash-Sutcliffe Efficiency (NSE) below 0.7 on the test dataset, and these wells were primarily located in the southern part of the western low plain (Figure 3a). In this region, the average absolute error between simulated and observed daily groundwater level depth ranged from 0.04 to 2.93 meters, although the maximum error reached as high as 11.56 meters (Figure 3c), indicating that the model exhibited certain instability in localized areas. Figure 4 compares the simulated and observed groundwater level depth series at several poorly performing wells in this region. As shown in the figure, significant discrepancies occurred during certain periods, and the fitting performance was unsatisfactory. The primary reason for this discrepancy is the large annual fluctuation in groundwater level depth at many wells in this region: 21.43% of the monitoring wells had a fluctuation range exceeding 10 meters. These extreme fluctuations posed challenges for the LSTM model’s simulation accuracy. In the training data used for the LSTM model, samples with extreme values of groundwater level depth were relatively scarce, while samples with moderate values were more abundant. Consequently, the model tended to fit the data in the moderate range more accurately, resulting in limited predictive ability for the extreme ends of the groundwater level series. Despite the reduced accuracy at certain wells, 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.”

Lines 373–376 and 557–558: These sections are overly descriptive. Instead of simply stating observations, clarify what the results reveal about the status or trends of water resources. Quantitative insights or implications for water management should be emphasized.

第 373–376 行与第 557–558 行：这两部分的描述偏于叙述性，应进一步挖掘这些结果对水资源状况或趋势的揭示意义。建议突出定量分析的结论或对水资源管理的启示。

Response: We sincerely thank the reviewer for the valuable comments. We have carefully revised the relevant sections of the manuscript.The main adjustments are summarized as follows:

1. Lines 373–376 (Analysis of frozen-thawed period dynamics):

While retaining the proportions and temporal characteristics of the three types of groundwater level dynamics during the frozen-thawed period, we further elaborated on the practical significance of their spatial differences for regional water resource management. We highlighted the indicative role of each dynamic type in reflecting groundwater recharge or depletion conditions and proposed a framework for differentiated zoning and adaptive regulation strategies based on these types to enhance the scientific basis and precision of groundwater management. The revised paragraph is as follows:

“ Freeze-thaw processes intensify the transformation between soil water and groundwater (Daniel and Staricka, 2000; Miao et al., 2017; Lyu et al., 2022). As a typical seasonally frozen soil region, the Songnen Plain in China exhibits three main types of groundwater level fluctuations during the frozen-thawed period: 'decline during freezing and rise during thawing,' 'continuous decline,' and 'continuous rise' (Figure 6). The proportions of monitoring wells corresponding to these three types are 38.4% (V-shaped), 23.2% (continuously declining), and 38.4% (continuously rising), respectively. The distribution of these types reflects the diverse responses of regional groundwater systems to seasonal freeze-thaw processes, providing a foundation for refined water resource management. On one hand, this classification can help identify potential recharge and depletion zones during spring, serving as a basis for groundwater storage adjustment, agricultural irrigation, and water resource allocation. On the other hand, it supports the development of dynamic management strategies tailored to freeze-thaw processes, enhancing responsiveness to groundwater level fluctuations in cold regions."

2. Lines 557–558 (Analysis of annual-scale dynamics):

We briefly summarized the differences in the controlling mechanisms of the three types of annual-scale groundwater dynamics. While precipitation is the dominant recharge source across the study area, the discharge pathways differ significantly. From the perspectives of both "naturally dominated" and "human activity dominated" processes, we proposed localized management strategies—such as enhancing ecological water use security and strengthening groundwater extraction control—emphasizing the value of classification results in improving the adaptability of groundwater resource regulation. The revised paragraph is as follows:

"The classification of annual-scale groundwater level dynamics identified three types: precipitation infiltration–evaporation type(29.0%), precipitation infiltration–runoff type(18.1%), and extraction type (52.9%). These results indicate that the regional groundwater system is generally controlled by precipitation recharge; however, the different types reflect distinct water level response mechanisms. The evaporation and runoff types are dominated by natural processes and exhibit groundwater fluctuations that are more sensitive to climatic conditions. In contrast, the extraction-driven type is associated with intensive groundwater use and is more responsive to changes in anthropogenic activities. These classification results provide a basis for tailored management. In areas dominated by natural processes, efforts should focus on securing ecological water demand and integrating rainwater resources to maintain groundwater system stability. In regions where extraction dominates, optimizing groundwater abstraction and improving water use structure are essential to mitigating continuous water level decline and enhancing resource sustainability."