Dear Reviewer 1,

We sincerely thank you for your thorough review and constructive comments on our manuscript. Your insights are greatly appreciated and have helped us to further improve the clarity, transparency, and scientific rigor of our study. Below, we provide a point-by-point response to your comments. The original comments are shown in **black**, and our responses are in **blue**. Sentences intended for revision or addition in the manuscript are marked in **gold** and quotation marks, and will be formally submitted after the open discussion phase.

Comment 1:

This paper develops a long sequence dataset, and the research purpose and user target population of this dataset need to be further clarified;

Reply: Thank you for your thoughtful comment emphasizing the need to explicitly state the research objectives and define the intended user community of the dataset. In response to this comment, we will revise the concluding paragraph of the Introduction (lines 80–88 in the original manuscript) to provide a more explicit and direct statement of the dataset's research purpose and its target users. We propose to revise it as follows:

"To address the limitations of existing meteorological datasets in spatial resolution, temporal continuity, and variable completeness, this study introduces a high-resolution dataset of daily near-surface meteorological variables—including average, maximum, and minimum air temperature, atmospheric pressure, relative humidity, and sunshine duration—across mainland China. Spanning six decades (1961–2021) with kilometer-level granularity, the dataset is designed to support fine-scale applications such as land surface modeling, drought assessment, and water resource management. It is particularly suited for both scientific investigations and operational decision-making in data-sparse and topographically complex regions, such as western China. To achieve this, a hierarchical and progressive reconstruction framework is implemented to generate gridded estimates of six variables at approximately 2 meters above ground level, based on in-situ observations and a 1 km digital elevation model (DEM). A multilayer perceptron (MLP) regression model is employed in this framework to capture nonlinear relationships between station observations and topographic predictors (e.g., latitude, longitude, and elevation), enabling fine-scale reconstruction across complex terrain."

Comment 2:

"The objective of this study is to develop a high-resolution and accuracy-assessed dataset of daily near-surface meteorological variables across mainland China, suitable for applications in hydrological modeling, environmental monitoring, and climate analysis." How did the author consider the issue of temporal "homogeneity" in a long series dataset used for climate analysis?

Reply: Thank you for raising this important point. We fully acknowledge that ensuring temporal homogeneity is essential for the reliability of long-term climate analyses. In this study, rather than applying direct spatial interpolation, we adopted a deep learning–based reconstruction framework that reconstructs each meteorological variable independently on a daily basis. The core of this framework is a multilayer perceptron (MLP) model, designed to learn and reconstruct the spatial distribution characteristics of each meteorological variable independently for every single day, based on nonlinear interactions with geographic and meteorological predictors. Because each day is modeled separately, potential quality issues on a particular day do not propagate temporally, thereby preserving the dataset's temporal integrity.

The in-situ observations used for training are sourced from the China Daily Surface Climate Dataset, developed and maintained by the China Meteorological Administration (CMA). According to the dataset documentation and metadata, this dataset has undergone comprehensive quality control and homogenization procedures to ensure its temporal consistency. Prior to model training, we further applied a quality-based filtering procedure to exclude all observations flagged with quality control codes indicating suspect (code 1), erroneous (code 2), missing (code 8), or unverified (code 9) values, thereby retaining only high-confidence records (code 0). This ensures the reliability of the training dataset and minimizes the propagation of observational uncertainties in the reconstruction process.

Owing to the day-by-day modeling strategy and the exclusive use of homogenized and quality-controlled station data, the resulting gridded product structurally maintains the temporal homogeneity inherent in the original CMA dataset. This ensures that the dataset is well-suited for multi-decadal climate analyses and enables robust assessments of long-term climatic trends.

Comment 3:

What is the quality of the raw observation data used to establish 1km grid data? Has the author conducted data quality evaluation, analysis, quality control, homogenization processing, etc. on the original observation data during its use?

Reply: Thank you for bringing up this important consideration. The accuracy and consistency of the raw observational data constitute the foundation for generating reliable gridded climate datasets, particularly for long-term applications. As noted in response to Comment 2, the reconstruction relies on in-situ records from the China Daily Surface Climate Dataset, developed and maintained by the CMA, which have undergone extensive quality control and homogenization as documented. According to the official dataset documentation and metadata, this dataset has undergone extensive quality control and homogenization—particularly for the period 1951–2010, during which several nationwide campaigns were conducted to identify and correct erroneous or missing data and to ensure temporal consistency. Specifically, data from 1951 to 2010 were corrected and supplemented as part of a national data rescue initiative, involving repeated manual inspections, error correction, and recovery of missing records, resulting in a data availability rate exceeding 99% and near-perfect accuracy. From 2011 to mid-2012, a hierarchical three-level quality control system (station–provincial–national) was applied, while data after mid-2012 underwent routine station-level validation.

In addition to leveraging this homogenized dataset, we implemented a strict pre-filtering protocol prior to model training, excluding observations flagged with quality control codes indicating suspect (code 1), erroneous (code 2), missing or unmeasured (code 8), or unverified (code 9) values. Only high-confidence observations (code 0) were retained for use, thereby ensuring the integrity and robustness of the training dataset and minimizing the propagation of uncertainties in the final reconstruction.

Furthermore, recognizing that surface meteorological stations in China have experienced relocations and metadata updates over the years, we incorporated time-specific station coordinates—including dynamic longitude, latitude, and elevation values—for each daily observation. This design avoids the spatial inaccuracies that could arise from using static station metadata and ensures that the model learns spatial relationships that faithfully represent the actual observational context on each day. This treatment helps maintain spatiotemporal consistency across the training samples and enhances the accuracy of fine-scale spatial reconstruction

Comment 4:

Compared to a grid spatial resolution of 1km * 1km, using over 2000 observation data from China is relatively insufficient, especially in the sparse observation areas of western China. How does the author consider this issue?

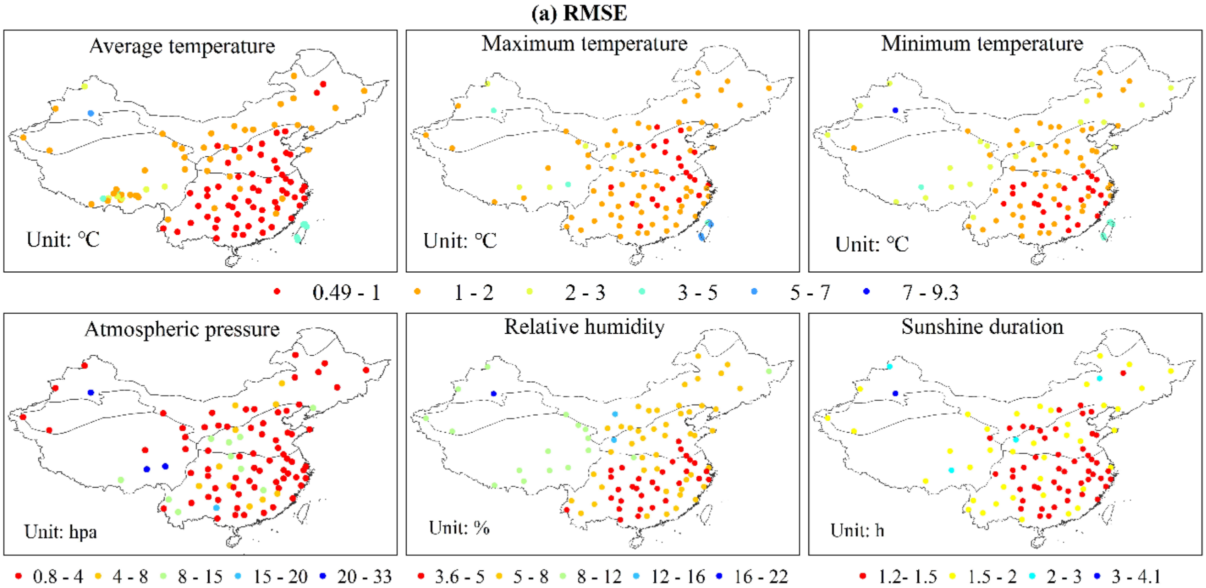
Reply: Thank you for your insightful comment. We fully acknowledge that the sparse distribution of meteorological stations—particularly across western China's complex terrain—poses a significant challenge for generating high-resolution (1-km) gridded meteorological products. Traditional interpolation methods, which often rely on linear assumptions and cannot fully incorporate topographic heterogeneity, tend to underperform in such data-scarce and topographically diverse regions.

To address this limitation, we adopted a model-driven reconstruction framework based on MLP, rather than relying on direct spatial interpolation. This framework is specifically designed to capture the nonlinear spatial distribution of each meteorological variable through a sequence of physically and statistically informed steps. By leveraging the point-wise spatial characteristics of each target variable, the model effectively learns fine-scale spatial structures across the domain. For example, air temperature is reconstructed solely from geographic predictors (latitude, longitude, elevation), while subsequent variables—such as atmospheric pressure, relative humidity, and sunshine duration—incorporate previously reconstructed variables as auxiliary inputs. This hierarchical structure enables the model to learn inter-variable dependencies and propagate spatial information from observation-rich regions to data-sparse areas.

Notably, in western China—characterized by rugged topography and limited station coverage—the model exhibits strong generalization capacity. Its ability to accurately reproduce fine-scale spatial patterns in these high-elevation regions provides empirical validation of the framework's robustness under sparse observational constraints. Moreover, although no meteorological data from Taiwan were included during training, validation results in this region reveal high reconstruction accuracy, further underscoring the framework's ability to generalize learned spatial representations to previously unseen areas.

As part of our validation, we evaluated model performance across different regions using in-situ station data that were intentionally excluded from model training and reserved exclusively for validation purposes. The spatial distributions of key evaluation metrics (RMSE, ME, and CC) are presented in Figure 6 and discussed in detail in Lines 307–322. These results indicate that, even in high-elevation and data-scarce regions such as the Tibetan Plateau and Taiwan—where no training stations were included—the reconstructed variables maintain reasonably high accuracy. This provides empirical support for the model's generalization capability and its applicability beyond the original training domain.

(a) RMSE



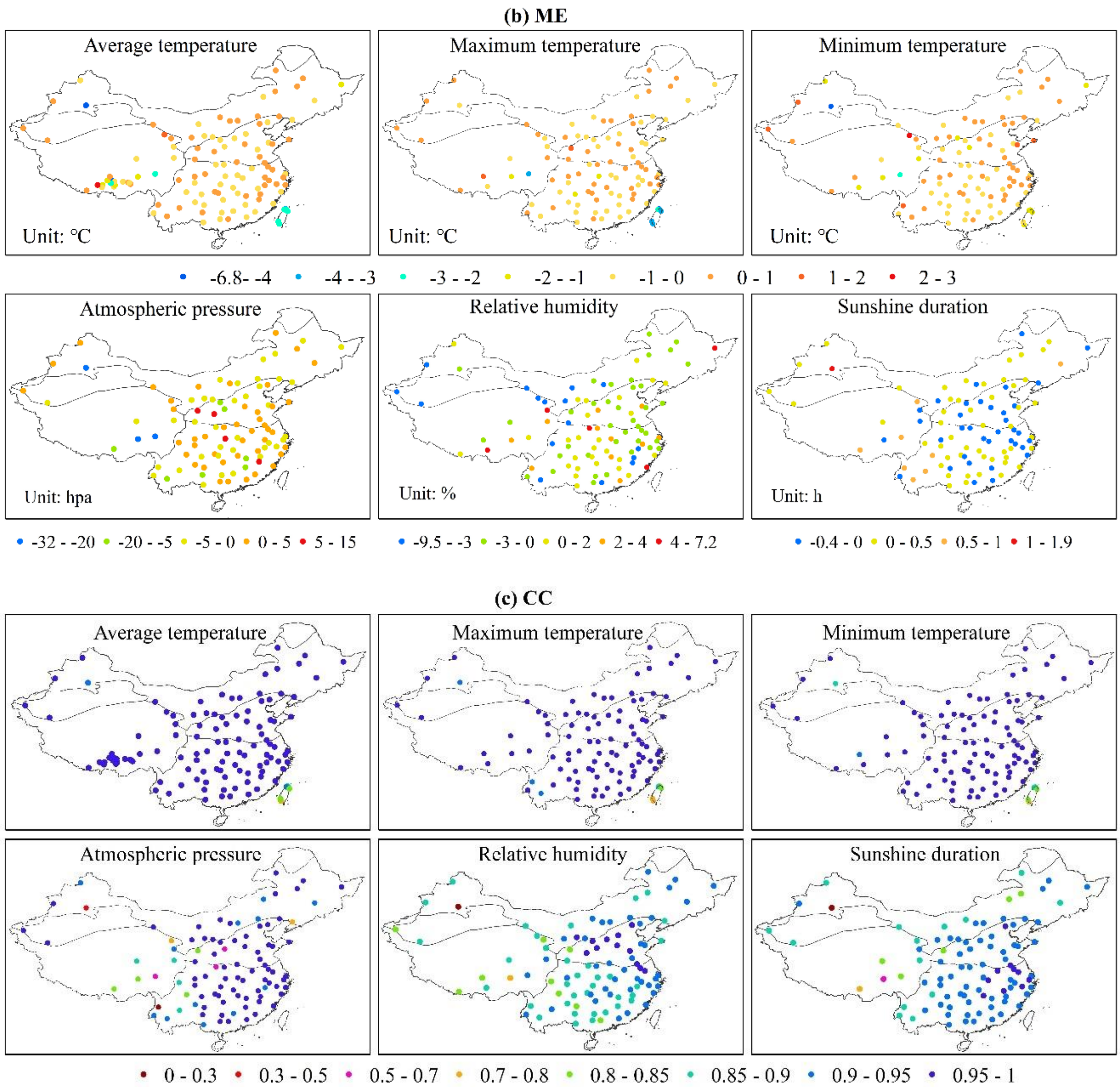


Figure 6: Distribution maps of RMSE (a), ME (b) and CC (c) between grid-modelled data of six meteorological element products and in-situ data.

Comment 5:

How is the " day boundary issues" handled? The ground meteorological observation in China adopts "20:00 Beijing time" as the boundary point of the day, which means that the observation day is from 20:00 to 20:00 the next day. This standard is applicable to daily value statistics of factors such as precipitation and temperature. Prior to the 1980s, some stations had a phenomenon of inconsistent day boundaries (such as a few stations using 08 or local time), which led to a decrease in comparability between early data and other stations

Reply: Thank you for your thoughtful comment regarding the potential inconsistency in day boundaries in Chinese meteorological data. We fully agree that variations in the definition of daily observation periods—if present—could affect the accuracy of reconstructed spatial patterns.

As stated in the documentation of the official daily meteorological dataset provided by the CMA, daily mean values are computed using observations recorded at 02:00, 08:00, 14:00, and 20:00 Beijing Time. Specifically, daily mean air temperature, relative humidity, and ground surface temperature are calculated as the average of these four values. For station pressure and wind speed, the same four-time averaging is generally applied; however, in the case of manual stations without automatic instruments, daily means are computed from three observations (08:00, 14:00, and 20:00). If any scheduled observation required for averaging is missing, the daily mean for that variable is flagged as missing. This standardized method ensures consistent temporal boundaries and comparability across stations.

Since the dataset documentation does not report any exceptions to this protocol, we consider the issue of day boundary consistency to be sufficiently addressed. Moreover, if the documentation had indicated that specific stations used non-standard day boundaries (e.g., using 08:00 or local time as the daily cutoff), we would have excluded those stations from our reconstruction and validation datasets to prevent potential biases.

Comment 6:

Overall evaluation: This long sequence dataset did not take into account the quality of the observation data used, day boundary issues, uniformity issues, etc. during the development process. Therefore, the dataset reconstructed in this article also has day boundary and uniformity issues, which will have a serious impact on downstream user research.

Reply: Thank you very much for raising this important point. We fully understand the reviewer's concern regarding potential issues such as observation data quality, day-boundary inconsistencies, and temporal homogeneity in long-term meteorological datasets, as these aspects are indeed critical for ensuring the reliability of reconstructed products.

As elaborated in our responses to Comments 1 through 5, our reconstruction is based on the China Surface Climate Daily Dataset, which is the national benchmark product released by the CMA. This dataset has undergone extensive quality control and homogenization procedures prior to public release.

Specifically: (1) According to the official metadata, all daily values are calculated based on observations at 02:00, 08:00, 14:00, and 20:00 Beijing Time, ensuring standardized daily boundaries across the entire network; (2) The dataset documentation clearly states that data from 1951–2010 underwent repeated manual validation, correction, and gap-filling, resulting in >99% data availability and near-100% accuracy. For data after 2010, a standardized multi-level quality control procedure was applied, including station-, provincial-, and national-level checks, ensuring consistency and reliability across the full observation period. (3) For stations lacking full observations, daily means are flagged as missing, thereby preventing the inclusion of inconsistent data in our training or validation samples.

We believe these procedures reflect a robust national-level quality assurance protocol, which addresses many of the concerns raised regarding early inconsistencies and observation practices. We suggest that the issues raised in Comment 6—such as the quality of the observation data, day boundary definitions, and data homogeneity—have already been addressed in detail in our responses to Comments 1 through 5. In those responses, In those responses, we provided detailed explanations based on the official documentation and metadata descriptions of the CMA dataset, to clarify how such issues are handled in the source data. We sincerely hope those clarifications help to resolve any remaining concerns.

Finally, we fully appreciate the reviewer's attention to the integrity of long-term climate data, and we will consider adding a brief sentence in the manuscript to explicitly state that the input dataset follows a nationally standardized observation protocol with unified day boundaries and homogenized records.