

Article ID: 1006-8775(2026)02-0199-12

Temperature Bias Correction and Downscaling Forecasting Based on Numerical Weather Prediction and Deep Learning

HE Huan (何欢)¹, ZHAO Bi-yun (赵碧云)¹, ZHOU Yue (周悦)², FANG Hong-bin (方鸿斌)³, QI Hai-xia (祁海霞)²
(1. Jingmen Meteorological Bureau, Jingmen, Hubei 448000 China; 2. Hubei Key Laboratory for Heavy Rain Monitoring and Warning Research, Institute of Heavy Rain, China Meteorological Administration, Wuhan 430205 China; 3. Wuhan Central Meteorological Observatory, Wuhan 430074 China)

Abstract: This paper proposes deep learning spatiotemporal downscaling (DLSD): a deep learning model that integrates temperature bias correction and spatiotemporal downscaling based on numerical weather prediction (NWP) to enhance the accuracy and spatiotemporal resolution of temperature forecasting. The target area is partitioned into multiple subregions, thereby reducing the complexity of the model and augmenting the training dataset. By employing embedding techniques, the subregions and temporal variables are encoded in combination with diverse meteorological parameters derived from NWP for different atmospheric levels, terrain attributes, and geographic coordinates as input features. High-fidelity and high-resolution surface-temperature datasets serve as training targets. DLSD is implemented in Hubei Province to produce hourly temperature forecasts at a resolution of $0.05^{\circ} \times 0.05^{\circ}$. Upon implementation and benchmarking against Integrated Forecasting System predictions, it substantially improves evaluation metrics, including the mean absolute error (MAE), root mean square error, accuracy, and structural similarity index measure. DLSD efficiently reproduces fine spatial details and excels on regions with substantial elevation variations: the western mountainous area yields a 42.1% MAE reduction, contrasting with a 28.1% reduction observed on the eastern plains. The model also demonstrates effective temporal downscaling, narrowing the size of the hourly MAE range from 0.9°C to between 0.5 and 0.7°C , thereby enhancing the stability of hourly forecasts. Site-specific validations affirm the efficacy of DLSD in capturing the variability of daily temperatures, albeit with less prominent correction enhancements at stations than at grid points; this is partly attributable to the limited training data and observational sites available for the western mountainous terrain.

Key words: NWP; deep learning; temperature adjustment; downscaling

CLC number: P456 **Document code:** A

Citation: He, H., B. Y. Zhao, Y. Zhou, H. B. Fang, and H. X. Qi, 2026: Temperature Bias Correction and Downscaling Forecasting Based on Numerical Weather Prediction and Deep Learning. *J. Trop. Meteor.*, **32**, 199–210, <https://doi.org/10.3724/j.1006-8775.2026.004>.

1 INTRODUCTION

Numerical weather prediction (NWP) is currently the predominant method used for weather forecasting, achieving remarkable success in areas such as extreme weather warning and climate-change prediction by numerically integrating and solving atmospheric states on a grid-by-grid basis (Bauer et al. 2015). However, the relentless pursuit of higher temporal and spatial resolutions in NWP has imposed immense computational demands. Moreover, the increasing intricacies and inadequacies of microphysical processes, parameterization schemes, and discretization algorithms inevitably introduce systemic model errors (Privé and Errico 2013; Srivier et al. 2015; Yao et al. 2016; Yang et al. 2022), adding to the limitations of conventional numerical forecasting approaches.

In recent years, advances in computer science have led to the emergence of numerous novel approaches within the machine learning domain, alongside exploration for new weather forecasting paradigms. These approaches include the utilization of purely data-driven machine learning methodologies to predict future weather patterns, as these methodologies can mine and analyze data more efficiently. These capabilities have led them to showcase great potential and yield a series of breakthroughs. For example, NVIDIA introduced the FourCastNet model, which is based on Fourier transform-inspired deep-learning neural networks. This was the first approach to increase the spatial resolution of forecasts to a level comparable to that of mainstream NWP systems, enabling global weather forecasting at a 0.25° resolution (Pathak et al. 2022). Huawei Cloud introduced the Pangu model, which employs a three-dimensional vision transformer for meteorological data processing. This was the first time that an artificial intelligence (AI)-based meteorological model surpassed the most advanced NWP systems in the world in accuracy: its prediction accuracy ranged from 1 h to 10 d, which exceeds that of the Integrated Forecasting System (IFS) of the European Centre for Medium-Range Weather Forecasts (ECMWF) (Bi et al. 2023). Google's DeepMind team unveiled the

Received 2024-07-24; **Revised** 2025-08-29; **Accepted** 2025-12-17

Funding: Guangxi Key Research and Development Program (AB25069132); Natural Science Foundation of Hubei Province (Joint Fund) (2023AFD096, 2022CFD129, 2022CFD015)

Biography: HE Huan, primarily undertaking research on application of artificial intelligence method in weather forecasting.

Corresponding author: ZHOU Yue, e-mail: zhouyue8510@163.com

GraphCast model, which is based on graph neural networks and is capable of generating 10-day weather forecasts with a spatial resolution of $0.25^{\circ} \times 0.25^{\circ}$ and temporal resolution of 6 h within just 60 s, outperforming Pangu in accuracy (Lam et al. 2023). These data-driven AI models can be used to produce long-term forecasting products at extremely low computational costs. Nonetheless, the current data-driven AI meteorological models exhibit limitations such as a lack of physical consistency and a limited number of predicted elements (Chen et al. 2024). Furthermore, the advent of AI weather forecasting models is largely attributable to the availability of high-quality datasets. Notably, all of these AI models rely on the ECMWF ERA5 reanalysis dataset with a spatial resolution of $0.25^{\circ} \times 0.25^{\circ}$, inherently limiting the forecast resolutions of the models to the same level. Further increases in the resolution of AI weather models is not only constrained by the availability of datasets, but also by computational cost, as such an enhancement would also exponentially increase the computational power required for model training.

A wide range of sectors, including agriculture, transportation, energy, and daily life, demand increasingly accurate and detailed temperature forecasts. Although NWP- and AI-based meteorological models have made substantial progress in forecasting large-scale weather systems, they remain limited in the prediction of ground temperatures at a local, fine-grained scale. This limitation primarily manifests in inadequate representations of smaller-scale meteorological phenomena, such as those involving complex terrain and urban heat-island effects (Holtslag et al. 2013; Zhong et al. 2024). NWP models are constrained by the computational efficiency and sophistication of microphysical parameterization schemes, leading to limitations in the precision of temperature forecasting. Moreover, AI meteorological models face restrictions due to their computational power and the availability of training datasets, generally resulting in lower spatiotemporal resolutions (Huang et al. 2024). To promptly and effectively deliver precise temperature forecasts with high spatiotemporal resolutions, postprocessing bias correction and downscaling methods based on NWP are currently the most pragmatic and efficient approaches to address these prevailing challenges.

To obtain a more accurate forecast, bias correction needs to be applied to NWP. Scholars have proposed various methods, including model output statistics method linear regression (Glahn and Lowry 1972; Glahn 2014; Sheng et al. 2020; Wu et al. 2016), Bayesian models (Sloughter et al. 2010), Kalman filtering (Buehner et al. 2017; Pelosi et al. 2017; Yang 2019), tree-based machine learning algorithms (such as random forests, extreme gradient boosting, and the light gradient boosting machine (Fang et al. 2024; Li et al. 2019; Ren et al. 2020; Wu et al. 2023)), and deep learning methods based on neural networks such as convolutional neural networks (CNNs)

and transformers (Alerskans et al. 2022; Chen et al. 2022; Liu et al. 2023; Xiang et al. 2022; Zhi et al. 2020; Zhong et al. 2024). The main idea behind these methods is that NWP may have some biases; however, these biases are often predictable. For example, if a model systematically predicts higher temperatures under specific conditions, optimizing the model output by combining historical NWP predictions and site observational data can correct this bias. Practice has proven that these techniques can effectively improve the quality of NWP forecasts. To obtain more detailed forecasts, downscaling of NWP results is necessary. Scholars have developed many downscaling schemes, which can be divided into dynamic and statistical downscaling approaches (Meer et al. 2023). Dynamic downscaling is typically based on nested high-resolution regional models with limited target ranges (Hu et al. 2015), which can reduce the imposed computational load compared with that of global NWP. However, it still imposes large computational demands and is limited by the current understanding of atmospheric physics; thus, this approach has is limited in satisfying accuracy requirements (Yang et al. 2022). Statistical downscaling methods typically use statistical models to establish a relationship between low- and high-resolution observational data. These statistical methods include linear regression, decision-tree classifiers, and multilayer perceptrons (Ba et al. 2018; Beecham et al. 2014; Gutiérrez et al. 2019; Li and Guo 2017; Srivastava et al. 2013). With the advent of deep learning technology, researchers have begun to explore the application of deep learning models to statistical downscaling tasks by exploiting their efficiency in processing complex nonlinear relationships and large-scale datasets. Applications of deep learning to downscaling tasks have developed rapidly and achieved success. Vandal et al. (2017) first applied image super-resolution-based deep learning methods to the downscaling process, and fused high-resolution terrain data to improve the resolution of precipitation data. Since then, improved deep learning methods have been applied to meteorological element downscaling (Kuang et al. 2022; Stengel et al. 2020; Xiang et al. 2022; Zhong et al. 2024).

However, previously proposed bias-correction methods rely primarily on corrections based on observations from weather stations, without fully harnessing the potential of various meteorological factors near stations and terrain features to correct prediction biases. Furthermore, the previously mentioned deep-learning-based statistical downscaling approaches only utilize surface meteorological parameters acquired from NWP outputs, overlooking the impacts of larger-scale atmospheric circulation patterns on ground temperatures and not taking full advantage of the relatively more accurate upper-air meteorological variable predictions provided by NWP. This study introduces a novel deep-learning-based spatiotemporal downscaling and bias correction approach, which is here called the deep

learning spatiotemporal downscaling (DLSD) model, which combines NWP and deep learning techniques. The target region is initially divided into multiple subregions. This step simplifies the model structure, reduces the load on graphics processing unit (GPU) memory during training, and simultaneously increases the number of training instances. Subsequently, embedding techniques are employed to encode both the subregions and temporal variables. These encoded representations, along with a range of meteorological parameters provided by NWP derived from both the upper air and surface levels, as well as static variables such as terrain and geographical coordinates, are integrated as input features. Finally, by leveraging the strong feature extraction capabilities of CNNs, the model explores deeper connections among these various factors.

The remainder of this paper is structured as follows: Section 2 describes the study area and dataset, Section 3 introduces the methodology, Section 4 analyzes the experimental results, and the conclusion and scope for future work are presented in Section 5. By training with more precise and high-resolution surface-temperature datasets acquired from the China Meteorological Administration Land Surface Data Assimilation System (CLDAS) as calibration targets, the DLSD model performs bias correction on NWP temperature forecasts, yielding hourly temperature prediction products for the next 0–24-hour with notably enhanced spatiotemporal resolutions.

2 STUDY AREA AND DATA

2.1 Study area

The selected study area is located in Hubei Province, which is situated in the middle reaches of the Yangtze River in China, spanning the longitudes of 108.25°–116.25°E and latitudes of 28.9°–33.4°N. Predominantly characterized by a subtropical monsoon climate, the region experiences complex and variable weather conditions and possesses a rich array of geographical features. The terrain slopes from west to east—featuring a dense network of

ridges, mountains, hills, and plains—marking it as a transition zone from the second to third steps of the terrain elevation system of China, with significant altitudinal variations. This region has a well-developed meteorological observation network and features diverse terrain and variable weather patterns, rendering it an ideal experimental area for assessing the applicability of temperature correction and downscaling forecasting models. As depicted in Fig. 1, the area is divided into eastern and western regions, which are divided by a longitudinal line at 112°E for subsequent zonal evaluation and testing. A proportion of 86% of the western part has an altitude exceeding 200 m, with a standard deviation of 496 m, and predominantly comprises complex mountainous terrain. Conversely, the eastern section constitutes only 18% of the total area, has an altitude greater than 200 m with a standard deviation of 168 m, and primarily comprises plains flatlands.

2.2 Data

The ground-truth target data for model training were acquired from the CLDAS, with a horizontal spatial resolution of 0.05°×0.05° and temporal resolution of 1 h. This dataset is based on the NWP output products, which include observational data from various sources, such as ground weather stations and satellites, the space and time multiscale analysis system, physical inversion, topographic correction, and other technologies. Relevant studies have shown that this is the best high-spatial- and high-temporal-resolution temperature grid product in China (Liu et al. 2021; Liu et al. 2023; Yu et al. 2019).

The NWP forecasting data used in this study are obtained from the operational products of the ECMWF IFS. These data are released twice daily, at 00:00 and 12:00 UTC, with horizontal resolutions of 0.125°×0.125° for surface meteorological elements and 0.25°×0.25° for upper-air pressure-level elements. The temporal resolution of the forecasts is 3 h. The selected NWP forecasting variables, as detailed in Table 1, comprise 36 variables across different pressure layers, including geopotential

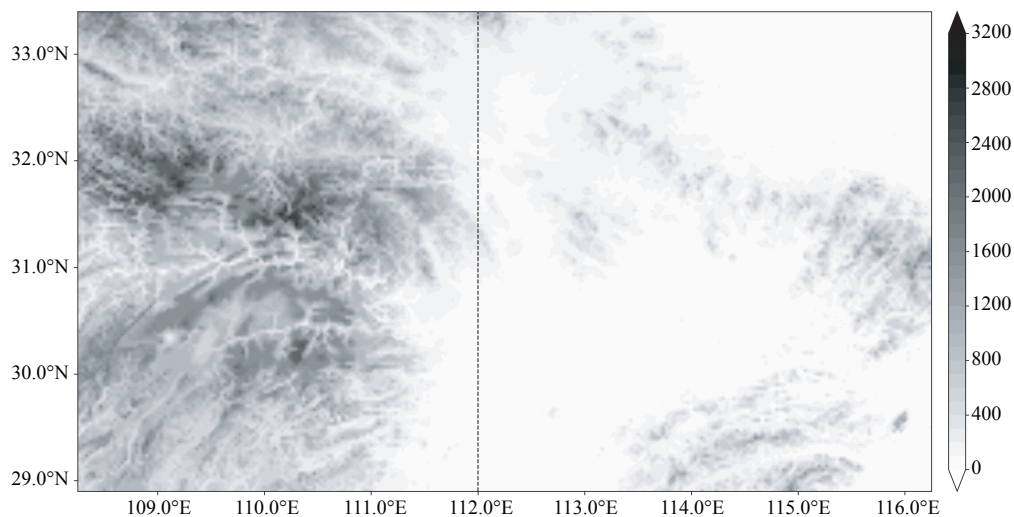


Figure 1. Study area shaded by elevation (units: m).

Table 1. Model input features.

NWP	Feature	Level (hPa)
Pressure level	Geopotential height (z)	500, 850, 900, 925, 950, 1000
	Temperature (t)	500, 700, 800, 850, 900, 925, 950, 1000
	U-component of wind (u)	500, 700, 850, 925, 950
	V-component of wind (v)	500, 700, 850, 925, 950
	Relative humidity (r)	200, 300, 400, 500, 600, 700, 800, 850, 900, 925, 950, 1000
Surface level	Mean sea-level pressure (msl)	—
	2-meter temperature ($2t$)	—
	2-meter dewpoint temperature ($2d$)	—
	0-meter U wind component ($10u$)	—
	10-meter V wind component ($10v$)	—
Others	Altitude (alt)	—
	Longitude (lon)	—
	Latitude (lat)	—

height, temperature, U- and V-components of wind, and relative humidity variables. Geopotential heights are tied to specific constant-altitude surfaces of different variables in the atmosphere. Temperature and wind parameters are indicative of the fundamental state of the atmosphere, cold and warm air advection, and the evolution of weather patterns. Relative humidity, which is closely related to cloud cover, directly influences the amount of solar radiation reaching the ground and thus has a significant effect on ground temperature. Therefore, relative humidity, across the widest range of pressure levels, is selected to represent the observed solar radiation conditions. Additionally, five surface forecast variables derived from NWP are selected, all of which directly correlate with temperature. Furthermore, three static ground-level variables—altitude (alt), longitude (lon), and latitude (lat)—are used as indicators of the geographic variations between different regions.

Air temperature data measured at 80 national meteorological ground observation stations across Hubei Province are employed as an independent third-party dataset to validate the predictive outcomes of the proposed approach. This dataset is not directly used in the training, validation, or testing phases of the model, thereby ensuring a more objective evaluation of the resulting forecasting performance.

3 METHOD

3.1 Data processing

Because of the substantial reach of the Hubei Province, which spans longitudes of 108.25°–116.25°E and latitudes of 28.9°–33.4°N, directly outputting data at a resolution of 0.05°×0.05° would result in a matrix with a size of 91×161. This would lead to high model complexity that, because of the high GPU memory demands of deep learning models, makes it challenging to achieve model convergence without advanced computer hardware. To address this challenge, the target region is divided into 299 subregions, each with a 7×7 grid

(determined as the optimal grid size through multiple experiments) at a 0.05°×0.05° resolution. This approach not only reduces the complexity of the model to facilitate convergence but also increases the number of training samples.

To ensure consistency with the hourly resolution of the CLDAS reference data, the atmospheric and surface meteorological elements of the ECMWF IFS product are interpolated hourly using cubic spline interpolation. Because of the resolution mismatch between the NWP pressure levels and the surface layer, the NWP outputs of the deep learning model are unified to a resolution of 0.25°×0.25°. This adjustment facilitates capturing the broader weather circulation pattern in the region covered by the NWP data. Similarly, three static ground variables (alt, lon, and lat) are processed at a 0.25°×0.25° resolution. Considering the direct relevance of the 2-meter temperature, alt, lon, and lat to the temperature labels, their importance within the neural network is emphasized by interpolating these four factors to match the 0.05° resolution of the label data before reinputting them into the model. Notably, although the central points of the 7×7 grids produced for these four factors align with those of the 44 remaining factors in the 7×7 grids, their differing resolutions indicate that they represent different spatial extents.

Data from January 2020 to March 2023 are used for model training and validation. Given the sequential nature of this time-series dataset, random partitioning is not appropriate; instead, the first 80% of the chronologically ordered data are used as the training set, with the remaining 20% serving as the validation set. Data from April to September 2023 are exclusively set aside for testing purposes and are not included in either the training or validation processes. The IFS forecasting data are released twice daily, at 00:00 and 12:00 UTC; however, owing to processing and distributional delays, it typically takes more than 6 h for these data to reach end users. Consequently, in the testing phase, the model is fed with

NWP data intended for 12- to 36-hour forecasts, which aligns with practical operational scenarios. To expedite model convergence, prior to inputting the data into the model, Z-score normalization is applied to each of the 48 predictors in the training phase:

$$Z = \frac{x - u}{\sigma} \quad (1)$$

where x , u , and σ are the value, mean, and standard deviation, respectively, of the factor that needs to be standardized. The parameters u and σ of the training set are saved to standardize the validation- and testing-set factors.

3.2 Model structure

The mapping relationships of the model vary even under the same weather circulation pattern owing to the different locations of the subregions. Similarly, temperature exhibits distinct diurnal variations that are directly related to the time of day. Hence, the positions of the subregions and daytimes corresponding to the observations are incorporated into the model as factors. The direct use of natural sequential numbers can lead to obvious logical errors. For example, 23:00 and 00:00, although neighboring periods, are the farthest apart in terms of the numbers that represent them, requiring them to be treated as categorical variables. Traditional machine learning methods often employ one-hot encoding (Bishop 2006) to discretize categorical variables into binary vectors; however, this strategy results in sparse vectors that fail to consider the inherent relationships among different categories and are ill-suited for the convolutional modules involved in deep learning. To address this problem, embedded encoding was adopted, which is commonly used in natural-language-processing-based deep-learning methods (Mikolov et al. 2013). Embedding transforms categorical variables into continuous vector representations, which initially sparsifies the variables via one-hot encoding and then, through neural network training, increases or reduces the number of dimensions. This allows the training process to uncover underlying connections among the sparsified variables, thus generating new vectors without treating them as entirely independent categories. Prior to inputting the subregion position serial numbers and hour numbers into the model, an autoencoder employing the embedding method is

applied to the 299 subregions and 24 h, with each categorical variable encoded with a size of 49. After reshaping the outputs to $7 \times 7 \times 1$, these embeddings are concatenated alongside the aforementioned 48 standardized factors to constitute the model input.

The structure of the DLSD forecasting model is depicted in Fig. 2. The input data are structured as a matrix ($\mathbf{H}$, $\mathbf{W}$, $\mathbf{C}$), where $\mathbf{H}$ and $\mathbf{W}$ are equal to seven, indicating the number of grids per subregion, and $\mathbf{C}$ is equal to 50, representing the number of channels or predictive factors used in the convolution operation. Thus, the total input size is $7 \times 7 \times 50$. The input data are fed into a 2D convolutional layer, which employs two key characteristics: local connections and shared weights (Simonyan and Zisserman 2014). Given the weaker correlations between distant grid points, not every grid point needs to be globally connected and local connections suffice. Furthermore, assuming similar statistical properties among grid points facilitates the sharing of learned feature parameters across different parts of the grid, thereby reducing the number of required parameters and simplifying the network. Each convolution operation captures the relationships of the temperature with the nearby meteorological elements and geographic features, collectively expressing these correlations via trainable convolution kernel parameters. Recognizing the influence of large-scale circulation patterns on local temperatures (Faranda et al. 2023; Wu et al. 2021), multiple convolutional layers are stacked to expand the receptive field by integrating both local and global features. Pooling is not performed after each convolutional layer to prevent the loss of fine details; instead, the number of convolution kernels is adjusted to maintain the output grid size. Each convolutional layer is followed by a batch normalization (BatchNorm) layer, which stabilizes the data distribution during training, mitigates variable drift issues, addresses vanishing gradients, and helps prevent overfitting. It functions analogously to Z-score normalization, albeit with its parameters learned during training rather than statically calculated. Following the BatchNorm layer, a SoftPlus function introduces nonlinearity akin to that of a smoothed rectified linear unit (ReLU), avoiding zero gradients for negative inputs and the “dying neuron”

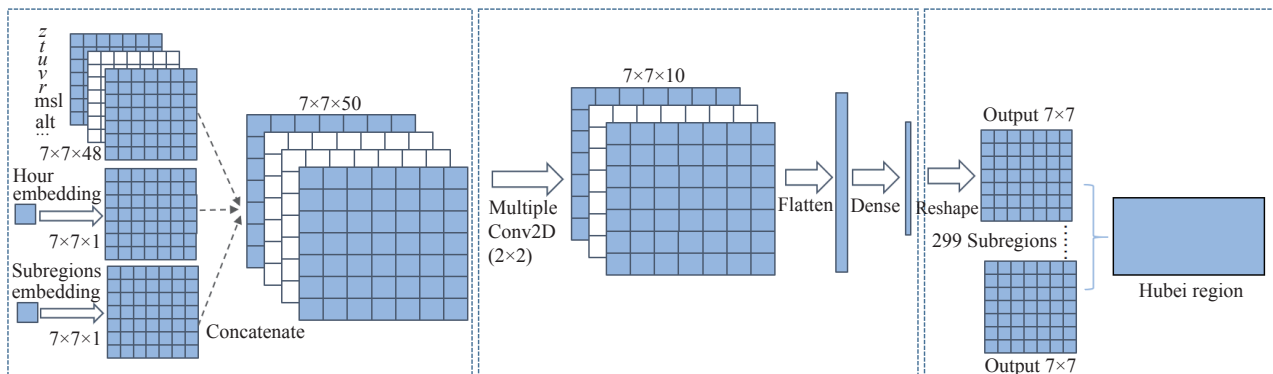


Figure 2. Overall structure of the DLSD model.

problem that is commonly encountered by ReLUs. The final convolutional layer outputs 10 feature maps ($7 \times 7 \times 10$), which are then flattened and passed through a fully connected (dense) layer. The dense layer ultimately produces 49 outputs that are reshaped back to a 7×7 grid corresponding to each subregion. This entire process corrects and downscales the 3-hour, $0.125^\circ \times 0.125^\circ$ -resolution NWP temperature data to a finer 1-hour, $0.05^\circ \times 0.05^\circ$ -resolution dataset over the 7×7 grids. Finally, 299 corrected and downscaled subregions are reassembled to form the final product, covering the entire Hubei Province.

3.3 Evaluation metrics

The model output data are validated using CLDAS data and observations acquired from meteorological station sites as ground truth. To assess the effectiveness of both bias correction and spatiotemporal downscaling, two common evaluation metrics for continuous variables are used, i.e., the mean absolute error (MAE) and root mean square error (RMSE), which are defined as follows:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |F_i - O_i| \quad (2)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (F_i - O_i)^2} \quad (3)$$

where N is the total number of predicted grid-point values (stations), and F_i and O_i are the i -th predicted and true values, respectively.

Prediction-accuracy metrics are typically used in classification tasks. Temperature predictions can be converted to classification values by setting a threshold β . If the absolute error between a prediction and the corresponding true value is less than or equal to β , the prediction is a positive sample; whereas if greater than β , it is a negative sample. β is set as 2°C . The temperature prediction accuracy (Acc) is defined as

$$\text{Acc} = \frac{P}{G+P} \quad (4)$$

where P and G are the numbers of positive and negative samples, respectively.

The structural similarity index measure (SSIM) is a metric used to quantify the similarity between two images (Wang et al. 2004) and is based on the premise that human eyes perceive the structural information contained in images. Compared with traditional metrics, the SSIM aligns better with human visual perception and is widely used in the fields of computer vision and image processing. The SSIM takes into account the brightness, contrast, and structural texture of images. By substituting temperature values for image color values, the SSIM can be employed to evaluate the model, ensuring that the improvements reflected by evaluation metrics such as the MAE and RMSE are not merely the result of the model smoothing out spatial details and losing intricate features.

$$\text{SSIM} = \frac{\sigma_{of}}{\sigma_o \sigma_f} \cdot \frac{2\mu_o \mu_f}{\mu_o^2 + \mu_f^2} \cdot \frac{2\sigma_o \sigma_f}{\sigma_o^2 + \sigma_f^2} \quad (5)$$

where symbol “o” indicates the ground-truth field; “f” denotes the prediction field; μ_o and μ_f are the mean values of the ground truth and prediction, respectively; σ_o and σ_f are the corresponding standard deviations; and σ_{of} is the covariance of the ground truth and prediction. The three parts in the equation for the definition of the SSIM represent, from left to right, the correlation coefficient, brightness consistency term, and contrast consistency term. The larger the SSIM value, the more similar the compared temperature fields.

4 RESULTS

The forecast outcomes produced for the test dataset spanning April to September 2023 (which are derived from the IFS and provide forecasts daily at 00:00 and 12:00 UTC) are inspected. These forecasts are subsequently fed into the DLSD model to generate corrected, spatiotemporally downscaled temperature forecasts with a lead time of 0 to 24 h. The DLSD and IFS products are subsequently compared using various evaluation metrics.

4.1 Grid point inspection

CLDAS grid point data are used as ground truth for the test dataset, and the evaluation results are listed in Table 2. The IFS performs poorly because of its coarse resolution and the more complex terrain in the western mountainous areas, yielding an MAE of 1.801°C ; this value is slightly better than that produced in the eastern plains, where the MAE is 1.349°C . Compared with those of the IFS, the MAEs of DLSD in the western and eastern areas are reduced by 42.1% and 28.1%, respectively. DLSD produces especially great improvements in the west; it narrows the gap between the eastern and western areas, and reduces the MAE to approximately 1°C . The RMSE and Acc exhibit similar trends to that of the MAE. A notable difference is observed between the east and west before applying the correction strategy, and after the correction step they show considerable agreement.

From the perspective of image similarity, DLSD does not reduce the MAE, RMSE, or other evaluation indicators through fuzzification, and the SSIM in each region significantly improves compared with that achieved by the IFS. Notably, the SSIM values produced by DLSD and the IFS for the mountainous areas in the midwestern region are greater than those for the eastern plain, which

Table 2. Evaluation metrics produced for different regions from the test dataset with a 0–24-hour forecast lead time.

Region	Model	MAE	RMSE	Acc	SSIM
Total	DLSD	1.003	1.358	0.888	0.917
	IFS	1.559	2.03	0.713	0.8
West	DLSD	1.042	1.397	0.879	0.922
	IFS	1.801	2.318	0.64	0.753
East	DLSD	0.97	1.323	0.896	0.817
	IFS	1.349	1.74	0.776	0.65

contrasts with the other metrics. This is due to the large regional differences represented by the graph, and it is meaningless to compare the SSIM values of different graphs, as this metric can only be compared between different models applied to the same region (Wang et al. 2004). The range of spatial variation in the temperature in the eastern plain is small, and a small temperature prediction error may lead to a large decrease in the SSIM.

4.2 Spatial downscaling by calculating differences

To further study the effectiveness of the model in spatial downscaling tasks, MAEs are obtained by comparing DLSD and IFS outputs to CLDAS temperature products under a 0–24-hour lead time from April to September 2023. Subsequently, the spatial distributions of the MAEs and their relationships with altitude are investigated, as illustrated in Fig. 3. The spatial distributions of the MAEs produced by DLSD and the IFS, combined with the regional boundaries of Hubei Province (Fig. 1), are shown in Figs. 3a and 3b, respectively. The IFS products mainly produce large errors in areas with large altitudinal variations. The MAEs in some areas exceed 6 °C, and the maximum grid-point MAE reaches 9.05 °C, whereas the performance achieved in flat areas is relatively good, ranging from 1 to 2 °C. After applying the correction scheme of the DLSD model, the MAE decreases by 35.7%. A total of 57% of the grid-point MAE is maintained within 1 °C. Significant corrections are observed in the maximum area of the MAE yielded by

the IFS, and the maximum grid point MAE of the DLSD model is 2.07 °C, which is a substantial decrease.

The distributions of MAE in different altitude ranges are shown in Figs. 3c and 3d. In areas higher than 200 m above sea level, the average MAE of the IFS is 1.81 °C, and the average MAE of DLSD is 1.04 °C. In the areas below 200 m above sea level, the average MAE of the IFS is 1.30 °C, and the average MAE of DLSD is 0.97 °C. It appears that DLSD has strong correction capabilities for complex mountainous areas. In addition, the spatial distribution of the DLSD errors is more uniform than that of the IFS errors, and only a few areas have slightly larger values. According to the topographic map, the terrains of these areas are extremely complex, with crisscrossing gullies and isolated peaks. Temperature and terrain data with a 5-km resolution still cannot perfectly describe the microclimates in mountainous areas, and more refined target CLDAS grid data are needed to achieve better results.

Figure 4 shows a comparison between the results of the 6-, 12-, 18-, and 24-hour lead-time temperature forecasts from 00:00 UTC on September 1, 2023, and the grid-point values of the CLDAS. The IFS roughly reflects the spatial distribution of the temperature; however, the temperature biases at some time stamps are large, and the temperature refinements produced in high-elevation areas are insufficient and lack detailed structural characteristics. After revision of the DLSD model,

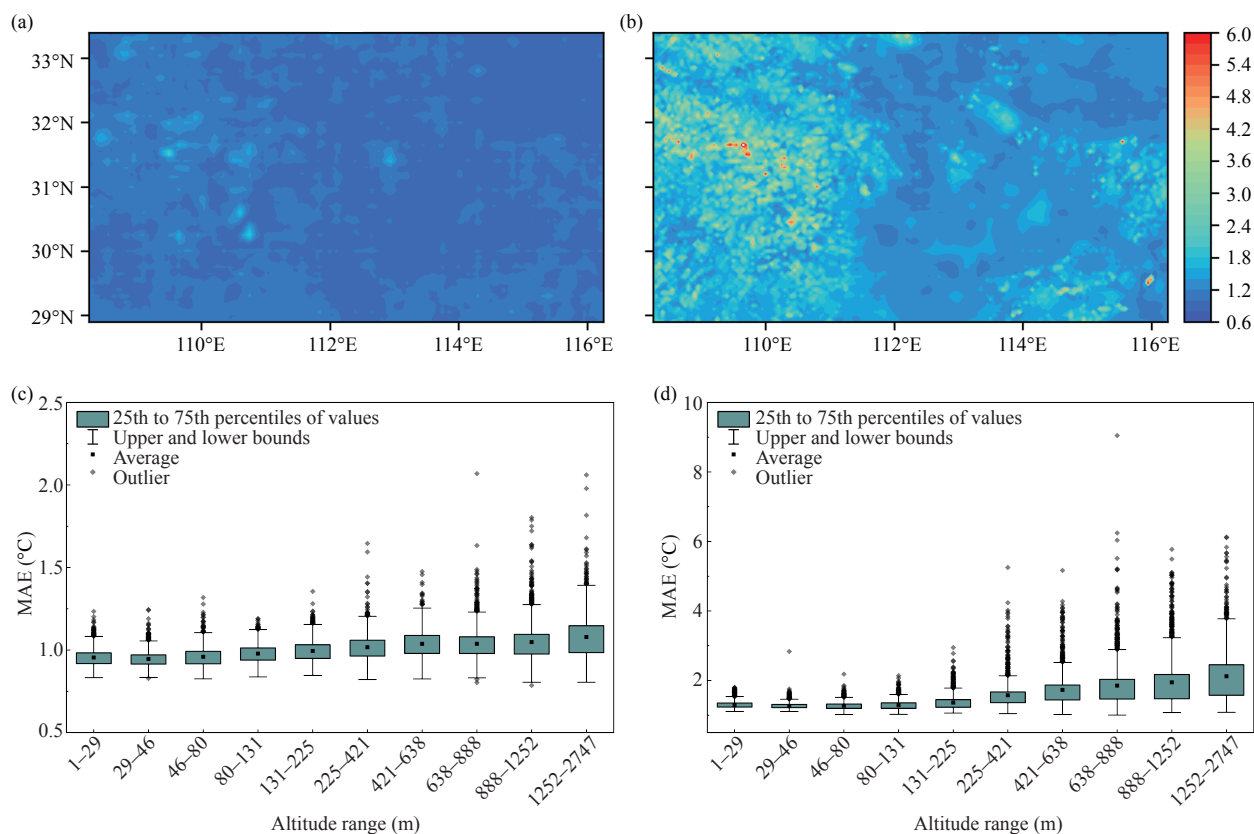


Figure 3. (a, c) Spatial distributions of the MAEs and their relationships with altitude produced by the differences between DLSD and CLDAS, and the differences (b, d) between IFS and CLDAS under 0–24-hour forecast lead times from April to September 2023 (units: °C).

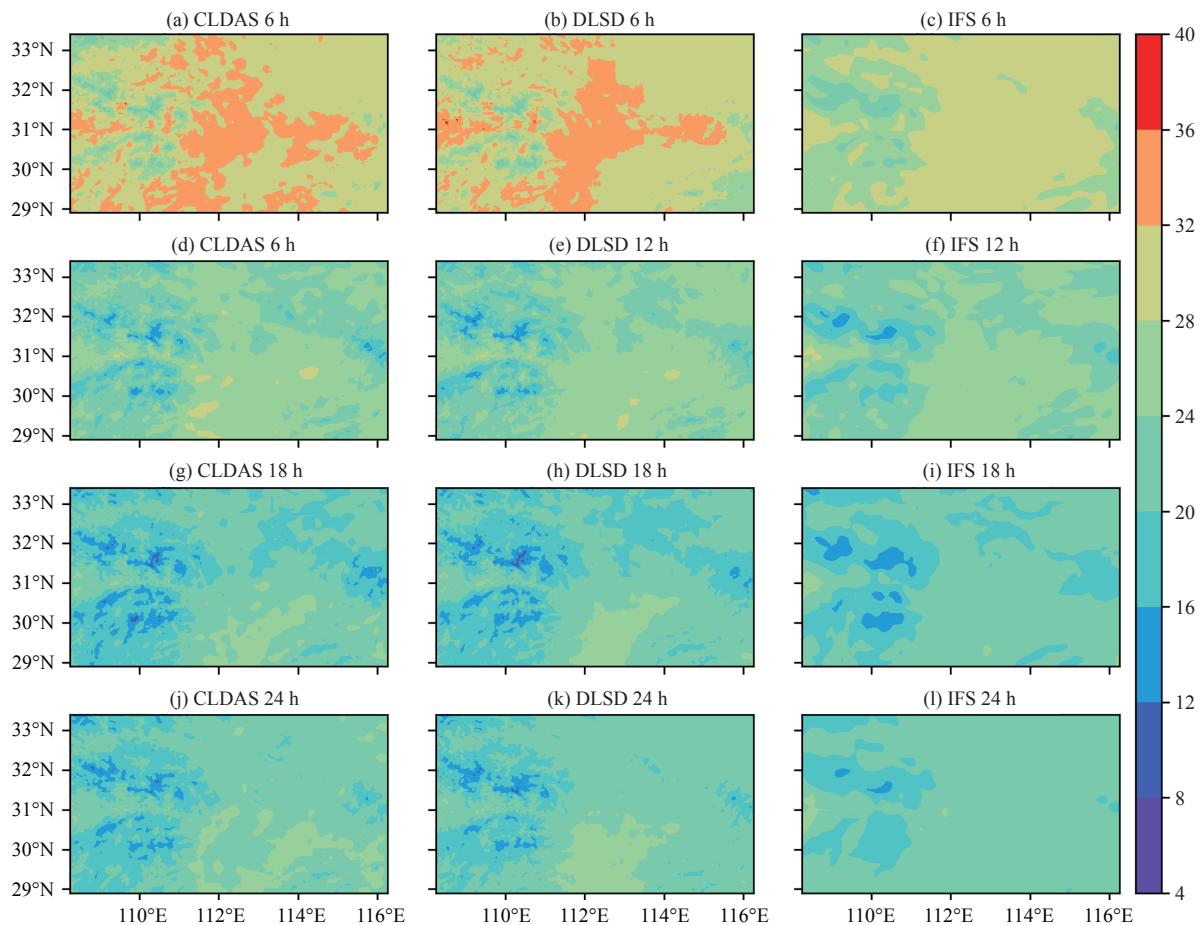


Figure 4. Comparison among the results obtained by DLSD, IFS, and CLDAS (units: $^{\circ}\text{C}$) with (a–c) 6-hour, (d–f) 12-hour, (g–i) 18-hour, and (j–l) 24-hour forecast lead times from 00:00 UTC on September 1, 2023.

predictions at all four forecast lead times effectively reflect the refined temperature structure, indicating that spatial downscaling does not fuzzify the forecasting products. This is confirmed by the SSIM values (Ding et al. 2024; Hess and Boers 2022). Compared with those of the CLDAS, the SSIMs of the IFS at the four time steps are 0.744, 0.786, 0.771, and 0.758, respectively, while DLSD improves the SSIMs to 0.849, 0.945, 0.952, and 0.934. The results show that the DLSD model can reconstruct small-scale spatial details, and the errors between the downscaled results and true values are smaller.

4.3 Temporal downscaling

The DLSD model framework incorporates a temporal encoding feature; therefore, the temperature bias correction and spatial downscaling processes are accompanied by temporal downscaling. The temporal resolution of the IFS prediction product is 3 h, which is increased to 1 h by using the DLSD model. The daily variational characteristics exhibited by the MAEs obtained by comparing the DLSD and IFS outputs with the CLDAS are compared under 0–24-hour forecast lead times from April to September 2023 at 00:00 UTC and 12:00 UTC, as shown in Fig. 5. The forecast lead times of 0, 3, 6, ..., 18, 21, and 24 h are the original outputs of the IFS, and the remaining IFS lead-time data are obtained via cubic spline

interpolation. Within a 24-hour period, the hourly MAEs of the two products are larger in the daytime and smaller at night. This is because factors such as convective weather and solar radiation lead to greater uncertainty during daytime, as NWP models are typically limited in accuracy when simulating these phenomena; thus, this occurrence is commonly observed (Chen et al. 2022). DLSD has an obvious corrective effect on the IFS, and its results are better than the hourly IFS results of day. The IFS performs especially poorly during the high-temperature daytime period (5–8-hour lead times from 00:00 UTC and 17–20-hour lead times from 12:00 UTC), with its maximum MAE reaching 2.2°C , whereas the maximum correction range of DLSD at this time is $0.7\text{--}0.9^{\circ}\text{C}$. The difference between the maximum and minimum hourly MAE of the 0–24-hour IFS data is 0.9°C , and DLSD reduces it to between 0.5 and 0.7°C , which reduces small time differences and enhances the stability of hourly temperature predictions.

In addition, discontinuous sections are observed at 22 and 23 h (00:00 UTC) and at 10 and 11 h (12:00 UTC) for the IFS products because this is the local time of sunrise. Generally, the temperature continues to decrease at night and starts to increase at this time, but the data from the interpolation scheme do not show this feature, resulting in a sudden increase in the MAE of the IFS at this time,

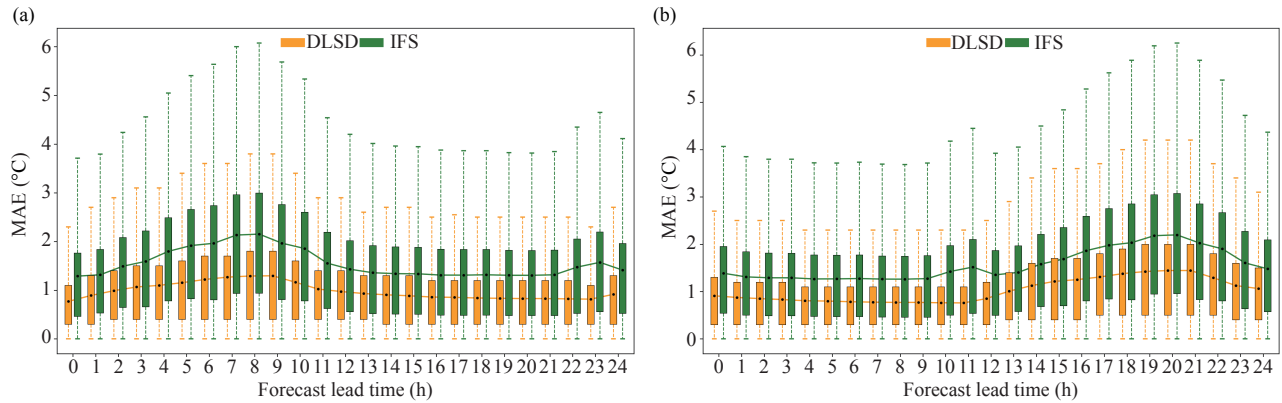


Figure 5. MAEs resulting from comparison of DLSD and IFS with CLDAS under forecast lead times of 0–24 h from (a) 00:00 UTC and (b) 12:00 UTC between April and September 2023.

whereas DLSD still provides very stable corrections. Moreover, at the interpolated NWP time points, the DLSD results are relatively smooth compared with the adjacent time points, with no sudden MAE increases, which is mainly due to the addition of temporal encoding features. The proposed model can minimize the error introduced by direct temporal NWP interpolation, indicating the stability of the DLSD model and effectiveness of its temporal downscaling.

4.4 Station inspection

To further verify the effectiveness of the proposed model, the observational data from the third-party stations are used to independently test the model forecasting results—these data are not directly involved in the model training, validation, and testing processes. Furthermore, the DLSD and IFS results are interpolated to the meteorological stations via the bilinear method to evaluate their forecasts more objectively. Fig. 6 shows the spatial distributions of the MAEs from DLSD and the IFS relative to the observed values at the meteorological stations. The differences in the MAEs of the IFS among the stations in the eastern region are small, and the model performs well; however, obvious errors occur at many stations in the western region, for which the maximum MAE is 2.87 °C. By contrast, the MAEs of the DLSD model at most stations are substantially reduced, and the effect in the western region is more notable. After the correction step, the difference between the eastern and western regions is substantially reduced. DLSD produces a negative correction for only one station in the northwest (lon: 110.75°E, lat: 32.05°N), where the MAE increases from 1.679 to 1.721 °C, which can be considered equivalent to no correction. DLSD has an obvious effect on the IFS in terms of station forecast correction. The MAEs of all stations decrease from 1.497 to 1.157, with a reduction of 22.7%; those of the eastern station decrease by 19.2%, whereas those of the western station decrease by 27.2%, which is less than the MAE reductions obtained in the grid test. The main reason for this is that the training labels of the DLSD model use the grid point live test, which has several advantages. In addition, for both the grid

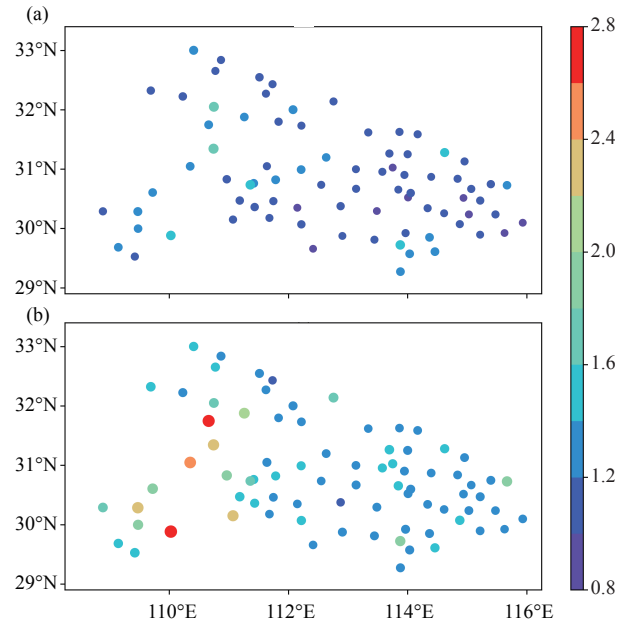


Figure 6. Spatial distributions of the 0–24-hour lead-time MAEs produced by the (a) DLSD and (b) IFS relative to the true station temperatures from April to September 2023 (units: °C).

and station tests, the largest reduction is observed in the western region. However, because relatively few stations are located in the western mountainous region, the reductions achieved for all station tests are not as good as those achieved in the grid test.

Figure 7 shows the hourly Acc values of DLSD and the IFS relative to the observed values of the meteorological stations from April to September 2023; these results are consistent with those of the grid test. DLSD leads in terms of Acc each time. The accuracies of DLSD and the IFS are characterized by daily variations in a single-valley distribution, and the accuracies are lower at 15–19 h (Beijing Time); that is, the nadir values for Acc occur from the afternoon to the evening. During this period from April to September, many afternoon showers occurred in Hubei Province (Fang et al. 2024; Lin et al. 2016), and the temperature fluctuated greatly, resulting in a

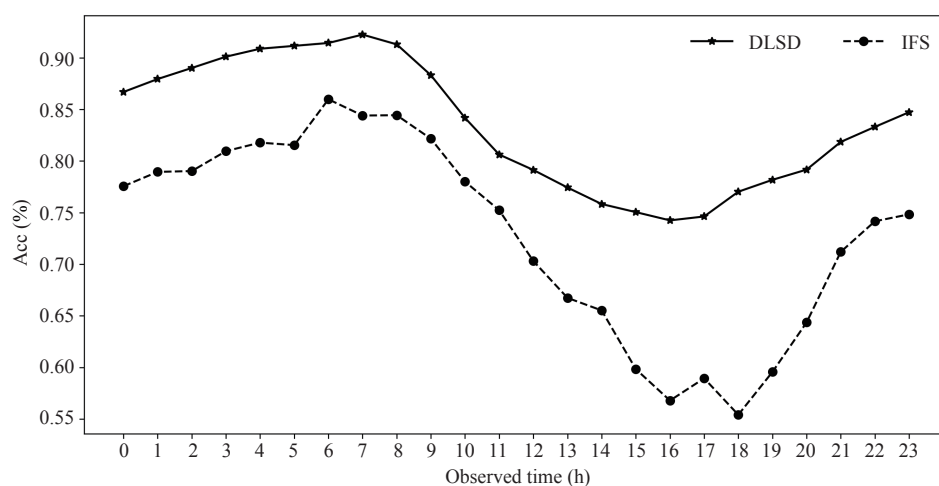


Figure 7. Hourly accuracies (Acc) of DLSD and the IFS with respect to the observed values of meteorological stations from April to September 2023.

reduction in accuracy. Although both DLSD and the IFS yield relatively low values at this time, the correction range of DLSD relative to that of the IFS reaches its maximum. DLSD can learn the diurnal variation characteristics of temperature through training and can be effectively corrected.

5 CONCLUSION

In this study, a deep learning model based on NWP, called DLSD, is proposed; this approach is a fusion scheme consisting of temperature bias correction and spatiotemporal downscaling. Based on a CNN, the model divides the research area into several subregions to reduce model complexity and increase the number of training samples. The subregions and time variables encoded by embedding and static variables such as high-altitude and ground-level meteorological elements, terrain, longitude, and latitude output by NWP are input as features. Using CLDAS grid temperature data with greater accuracy and higher resolution as training targets, the bias of the NWP temperature forecasts is corrected, and temperature forecasting results with higher temporal and spatial resolutions are obtained. The model is applied to Hubei Province and compared with IFS products. The results show the following:

(1) The temperature prediction product with a 1-hour temporal resolution and a $0.05^{\circ} \times 0.05^{\circ}$ horizontal resolution provided by the DLSD model is significantly better than that provided by the IFS model in terms of the MAE, RMSE, Acc, and SSIM evaluation. In a grid test, the MAE is reduced by 35.7% overall, indicating strong applicability in operational meteorological services, and the SSIM shows that the reductions in the MAE, RMSE, or other evaluation metrics brought about by DLSD is not the result of fuzziness.

(2) The DLSD model can reconstruct small-scale spatial details and has a stronger forecasting capability and larger correction range in areas with large altitudinal variations. The maximum MAE of the IFS predictions is

9.05 °C, and the maximum MAE of DLSD is 2.07 °C in areas with complex terrain, representing a substantial decrease. Overall, the MAE decreases by 42.1% in the west, which has complex terrain, and by 28.1% in the east, which has flat terrain.

(3) The addition of temporal encoding features can minimize the error introduced by direct temporal NWP interpolation, thereby increasing the temporal downscaling effect of DLSD. The difference between the maximum and minimum hourly MAE of the 0–24-hour IFS data is 0.9 °C; DLSD reduces this to 0.5–0.7 °C and significantly enhances the stability of the hourly temperature forecasts. In addition, the corrective effect is stronger in periods with poor IFS forecasts, especially in high-temperature periods during daytime. The maximum MAE of the IFS is 2.2 °C, whereas the maximum correction range of DLSD is 0.7–0.9 °C.

(4) A station test shows that the DLSD correction process is effective and performs well in both time and space. The daily variational characteristics of temperature can be learned through training and effectively corrected. However, because of the DLSD training tag data and the relatively small number of observation stations located in the western mountainous areas, the reduction in MAE during station correction is not as good as that observed during grid correction.

This study shows that, under NWP limitations induced by computational-power deficiencies and microphysical parameterization schemes, the bias correction and spatiotemporal downscaling strategy applied to NWP meteorological elements based on a deep learning model is effective in producing higher-resolution meteorological-element predictions. The DLSD model shows good applicability in the Hubei Province, which has complex and variable weather conditions and rich geographical characteristics. It is also expected to have good applicability in other regions. However, there is room for improvement in some areas with extremely complex terrains, which require higher-resolution grid data

as training target features. In addition, the accuracy of the NWP data, which are the inputs of the DLSD model, changes with forecast lead-time alterations, and the forecast lead time can be input as an encoding feature. These results can be used in future research to improve forecasting.

REFERENCES

- Alerskans, E., J. Nyborg, M. Birk, and E. Kaas, 2022: A transformer neural network for predicting near-surface temperature. *Meteor. Appl.*, **29**, e2098, <https://doi.org/10.1002/met.2098>.
- Ba, W. L., P. F. Du, T. Liu, A. Bao, M. Luo, M. Hassan, and C. Qin, 2018: Simulating hydrological responses to climate change using dynamic and statistical downscaling methods: A case study in the Kaidu River Basin, Xinjiang, China. *J. Arid Land*, **10**, 905–920, <https://doi.org/10.1007/s40333-018-0068-0>.
- Bauer, P., A. Thorpe, and G. Brunet, 2015: The quiet revolution of numerical weather prediction. *Nature*, **525**, 47–55, <https://doi.org/10.1038/nature14956>.
- Beecham, S., M. Rashid, and R. K. Chowdhury, 2014: Statistical downscaling of multi-site daily rainfall in a South Australian catchment using a Generalized Linear Model. *Int. J. Climatol.*, **34**, 3654–3670, <https://doi.org/10.1002/joc.3933>.
- Bi, K. F., L. X. Xie, H. H. Zhang, X. Chen, X. Gu, and Q. Tian, 2023: Accurate medium-range global weather forecasting with 3D neural networks. *Nature*, **619**, 533–538, <https://doi.org/10.1038/s41586-023-06185-3>.
- Bishop, C. M., 2006: *Pattern Recognition and Machine Learning*. Springer New York, 1122–1128 pp.
- Buehner, M., R. McTaggart-Cowan, and S. Heillette, 2017: An ensemble Kalman filter for numerical weather prediction based on variational data assimilation: VarEnKF. *Mon. Wea. Rev.*, **145**, 617–635, <https://doi.org/10.1175/mwr-d-16-0106.1>.
- Chen, H., R. H. Cai, J. J. Chen, C. H. Fu, L. Zhou, and L. Chen, 2022: Application and evaluation of temperature forecast based on deep learning method. *Meteor. Mon.*, **48**, 1373–1383, <https://doi.org/10.7519/j.issn.1000-0526.2022.070101>. (in Chinese)
- Chen, Y. T., Y. Wang, G. Huang, and Q. Tian, 2024: Coupling physical factors for precipitation forecast in China with graph neural network. *Geophys. Res. Lett.*, **51**, e2023GL106676, <https://doi.org/10.1029/2023gl106676>.
- Ding, S. Y., X. F. Zhi, Y. Lyu, Y. Ji, and W. Guo, 2024: Deep learning for daily 2-m temperature downscaling. *Earth Space Sci.*, **11**, e2023EA003227, <https://doi.org/10.1029/2023ea003227>.
- Fang, H. B., S. S. Wang, X. L. Wang, J. H. Tan, and L. B. Lu, 2024: Gridded temperature forecast correction method based on machine learning. *Meteor. Mon.*, **50**, 103–114, <https://doi.org/10.7519/j.issn.1000-0526.2023.111701>. (in Chinese)
- Faranda, D., G. Messori, A. Jezequel, M. Vrac, and P. Yiou, 2023: Atmospheric circulation compounds anthropogenic warming and impacts of climate extremes in Europe. *P. Natl. Acad. Sci. USA*, **120**, e2214525120, <https://doi.org/10.1073/pnas.2214525120>.
- Glahn, H. R., and D. A. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.*, **11**, 1203–1211, [https://doi.org/10.1175/1520-0450\(1972\)011<1203:tuomos>2.0.co;2](https://doi.org/10.1175/1520-0450(1972)011<1203:tuomos>2.0.co;2).
- Glahn, B., 2014: Determining an optimal decay factor for bias-correcting MOS temperature and dewpoint forecasts. *Wea. Forecasting*, **29**, 1076–1090, <https://doi.org/10.1175/waf-d-13-00123.1>.
- Gutiérrez, J. M., and Coauthors, 2019: An intercomparison of a large ensemble of statistical downscaling methods over Europe: Results from the VALUE perfect predictor cross-validation experiment. *Int. J. Climatol.*, **39**, 3750–3785, <https://doi.org/10.1002/joc.5462>.
- Hess, P., and N. Boers, 2022: Deep learning for improving numerical weather prediction of heavy rainfall. *J. Adv. Model. Earth Syst.*, **14**, e2021MS002765, <https://doi.org/10.1029/2021ms002765>.
- Holtzlag, A. A. M., and Coauthors, 2013: Stable atmospheric boundary layers and diurnal cycles: Challenges for weather and climate models. *Bull. Amer. Meteor. Soc.*, **94**, 1691–1706, <https://doi.org/10.1175/bams-d-11-00187.1>.
- Hu, H. Y., Y. J. Hu, Z. Zhong, and Y. M. Zhu, 2015: Double nested dynamical downscaling research on summer precipitation over China with WRF model. *J. Meteor. Sci.*, **35**, 413–421, <https://doi.org/10.3969/2014jms.0041>. (in Chinese)
- Huang, X. M., Y. L. Lin, W. Xiong, J. H. Li, J. C. Pan, and Y. Zhou, 2024: Research on international developments of AI large meteorological models in numerical forecasting. *Trans. Atmos. Sci.*, **47**, 46–54, <https://doi.org/10.13878/j.cnki.dqkxxb.20231201001>. (in Chinese)
- Kuang, Q. M., C. K. Shen, T. Z. Yu, and J. Liu, 2022: Weather statistical downscaling using a 3D multi-scale residual Laplacian pyramid network. *Trans. Atmos. Sci.*, **45**, 660–673, <https://doi.org/10.13878/j.cnki.dqkxxb.20220424002>. (in Chinese)
- Lam, R., and Coauthors, 2023: Learning skillful medium-range global weather forecasting. *Science*, **382**, 1416–1421, <https://doi.org/10.1126/science.adi2336>.
- Li, H. C., C. Yu, J. J. Xia, Y. Wang, J. Zhu, and P. Zhang, 2019: A model output machine learning method for grid temperature forecasts in the Beijing area. *Adv. Atmos. Sci.*, **36**, 1156–1170, <https://doi.org/10.1007/s00376-019-9023-z>.
- Li Y J, and Guo P W. 2017: Application of downscaling forecast for the North of Zhejiang precipitation in summer based on the BP neural network model. *Trans. Atmos. Sci.*, **40**, 425–432, <https://doi.org/10.13878/j.cnki.dqkxxb.20140324001>. (in Chinese)
- Lin, C. Z., L. Liu, W. C. Lin, Y. Q. Bai, H. X. Qi, and H. Yang, 2016: Characteristics of summer precipitation diurnal variations in Hubei Province. *Trans. Atmos. Sci.*, **39**, 490–500, <https://doi.org/10.13878/j.cnki.dqkxxb.20160130001>. (in Chinese)
- Liu, Q., and Coauthors, 2023: Deep-learning post-processing of short-term station precipitation based on NWP forecasts. *Atmos. Res.*, **295**, 107032, <https://doi.org/10.1016/j.atmosres.2023.107032>.
- Liu, Y., C. X. Shi, H. J. Wang, and S. Han, 2021: Applicability assessment of CLDAS temperature data in China. *Trans. Atmos. Sci.*, **44**, 540–548, <https://doi.org/10.13878/j.cnki.dqkxxb.20200819001>. (in Chinese)
- Liu, Y., Q. Q. Yan, and J. Yan, 2023: Analysis and assessment on the application of intelligent gridded real-time fusion analysis products during the Wuhan Military Games. *Meteor.*

- Environ. Sci.*, **46**, 98–105, <https://doi.org/10.16765/j.cnki.1673-7148.2023.03.013>. (in Chinese)
- Mikolov, T., I. Sutskever, K. Chen, G. Corrado, and J. Dean, 2013: Distributed representations of words and phrases and their compositionality. *Proc. 27th Int. Conf. on Neural Information Processing Systems*, Red Hook, NY, 3111–3119.
- Pathak, J., and Coauthors, 2022: FourCastNet: A global data-driven high-resolution weather model using adaptive Fourier neural operators. *arXiv*: 2202.11214, <https://doi.org/10.48550/arXiv.2202.11214>.
- Pelosi, A., H. Medina, J. Van den Bergh, S. Vannitsem, and G. B. Chirico, 2017: Adaptive Kalman filtering for postprocessing ensemble numerical weather predictions. *Mon. Wea. Rev.*, **145**, 4837–4854, <https://doi.org/10.1175/mwr-d-17-0084.1>.
- Privé, N. C., and R. M. Errico, 2013: The role of model and initial condition error in numerical weather forecasting investigated with an observing system simulation experiment. *Tellus A*, **65**, 21740, <https://doi.org/10.3402/tellusa.v65i0.21740>.
- Ren, P., M. X. Chen, W. H. Cao, Z. W. Wang, L. Han, L. Y. Song, and L. Yang, 2020: Error analysis and correction of short-term numerical weather prediction under complex terrain based on machine learning. *Acta Meteor. Sin.*, **78**, 1002–1020, <https://doi.org/10.11676/qxxb2020.060>. (in Chinese)
- Sheng, C. Y., S. D. Fan, Y. M. Rong, and W. Q. Sun, 2020: Comparison of several objective methods and optimal consensus forecast study of temperature. *Meteor. Mon.*, **46**, 1351–1361, <https://doi.org/10.7519/j.issn.1000-0526.2020.10.009>. (in Chinese)
- Simonyan, K., and A. Zisserman, 2014: Very deep convolutional networks for large-scale image recognition. *arXiv*: 1409.1556, <https://doi.org/10.48550/arXiv.1409.1556>.
- Sloughter, J. M., T. Gneiting, and A. E. Raftery, 2010: Probabilistic wind speed forecasting using ensembles and Bayesian model averaging. *J. Amer. Stat. Assoc.*, **105**, 25–35, <https://doi.org/10.1198/jasa.2009.ap08615>.
- Srivastava, P. K., D. W. Han, M. R. Ramirez, and T. Islam, 2013: Machine learning techniques for downscaling SMOS satellite soil moisture using MODIS land surface temperature for hydrological application. *Water Resour. Manage.*, **27**, 3127–3144, <https://doi.org/10.1007/s11269-013-0337-9>.
- Sriver, R. L., C. E. Forest, and K. Keller, 2015: Effects of initial conditions uncertainty on regional climate variability: An analysis using a low-resolution CESM ensemble. *Geophys. Res. Lett.*, **42**, 5468–5476, <https://doi.org/10.1002/2015gl064546>.
- Stengel, K., A. Glaws, D. Hettinger, and R. N. King, 2020: Adversarial super-resolution of climatological wind and solar data. *P. Natl. Acad. Sci. USA*, **117**, 16805–16815, <https://doi.org/10.1073/pnas.1918964117>.
- Vandal, T., E. Kodra, S. Ganguly, A. Michaelis, R. Nemani, and A. R. Ganguly, 2017: DeepSD: Generating high resolution climate change projections through single image super-resolution. *Proc. 23rd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, Halifax, NS, Association for Computing Machinery, 1663–1672, <https://doi.org/10.1145/3097983.3098004>.
- van der Meer, M., S. de Roda Husman, and S. Lhermitte, 2023: Deep learning regional climate model emulators: A comparison of two downscaling training frameworks. *J. Adv. Model. Earth Syst.*, **15**, e2022MS003593, <https://doi.org/10.1029/2022MS003593>.
- Wang, Z., A. C. Bovik, H. R. Sheikh, and E. Simoncelli, 2004: Image quality assessment: From error visibility to structural similarity. *IEEE T. Image Process.*, **13**, 600–612, <https://doi.org/10.1109/tip.2003.819861>.
- Wu, Q. S., M. Han, H. Guo, and T. H. Su, 2016: The optimal training period scheme of MOS temperature forecast. *J. Appl. Meteor. Sci.*, **27**, 426–434, <https://doi.org/10.11898/1001-7313.20160405>. (in Chinese)
- Wu, W. C., F. Y. Wei, Y. Q. Wang, and E. D. Zhu, 2023: Machine learning-based regression analysis of the winter temperature anomalies in China and associated influencing factors. *Acta Meteor. Sin.*, **81**, 163–174, <https://doi.org/10.11676/qxxb2023.20220078>. (in Chinese)
- Wu, X. Y., Z. C. Hao, F. H. Hao, X. Zhang, V. P. Singh, and C. Sun, 2021: Influence of large-scale circulation patterns on compound dry and hot events in China. *J. Geophys. Res. Atmos.*, **126**, e2020JD033918, <https://doi.org/10.1029/2020jd033918>.
- Xiang, L., J. P. Guan, J. Xiang, L. Zhang, and F. Zhang, 2022: Spatiotemporal model based on transformer for bias correction and temporal downscaling of forecasts. *Front. Environ. Sci.*, **10**, 1039764, <https://doi.org/10.3389/fenvs.2022.1039764>.
- Yang, D. Z., 2019: On post-processing day-ahead NWP forecasts using Kalman filtering. *Solar Energy*, **182**, 179–181, <https://doi.org/10.1016/j.solener.2019.02.044>.
- Yang, S. X., F. H. Ling, W. S. Ying, S. Yang, and J. J. Luo, 2022: A brief overview of the application of artificial intelligence to climate prediction. *Trans. Atmos. Sci.*, **45**, 641–659, <https://doi.org/10.13878/j.cnki.dqkxxb.20210623003>. (in Chinese)
- Yao, S. L., G. Huang, R. G. Wu, X. Qu, and D. Chen, 2016: Inhomogeneous warming of the Tropical Indian Ocean in the CMIP5 model simulations during 1900–2005 and associated mechanisms. *Climate Dyn.*, **46**, 619–636, <https://doi.org/10.1007/s00382-015-2602-5>.
- Yu, J. W., C. Li, N. H. Cai, M. Liu, and Q. H. Zhao, 2019: Applicability evaluation of the national gridded real-time observation datasets in Jiangsu Province. *Meteor. Mon.*, **45**, 1288–1298, <https://doi.org/10.7519/j.issn.1000-0526.2019.09.009>. (in Chinese)
- Zhi, X. F., T. Wang, and Y. Ji, 2020: Multimodel ensemble forecasts of surface air temperature over China based on deep learning approach. *Trans. Atmos. Sci.*, **43**, 435–446, <https://doi.org/10.13878/j.cnki.dqkxxb.20200219003>. (in Chinese)
- Zhong, X. H., F. Du, L. Chen, Z. Wang, and H. Li, 2024: Investigating transformer-based models for spatial downscaling and correcting biases of near-surface temperature and wind-speed forecasts. *Quart. J. Roy. Meteor. Soc.*, **150**, 275–289, <https://doi.org/10.1002/qj.4596>.

Citation: He, H., B. Y. Zhao, Y. Zhou, H. B. Fang, and H. X. Qi, 2026: Temperature Bias Correction and Downscaling Forecasting Based on Numerical Weather Prediction and Deep Learning. *J. Trop. Meteor.*, **32**, 199–210, <https://doi.org/10.3724/j.1006-8775.2026.004>.