

MambaDS: Near-Surface Meteorological Field Downscaling with Topography Constrained Selective State Space Modeling

Zili Liu* *Graduate Student Member, IEEE*, Hao Chen[†], Lei Bai, Wenyuan Li, Wanli Ouyang, Zhengxia Zou *Senior Member, IEEE* and Zhenwei Shi[†] *Senior Member, IEEE*

Abstract—In an era of frequent extreme weather and global warming, obtaining precise, fine-grained near-surface weather forecasts is increasingly essential for human activities. Downscaling (DS), a crucial task in meteorological forecasting and remote sensing, enables the reconstruction of high-resolution meteorological states for target regions from global-scale forecast results. Previous downscaling methods, inspired by CNN and Transformer-based super-resolution models, lacked tailored designs for meteorology and encountered structural limitations. Notably, they failed to efficiently integrate topography, a crucial prior in the downscaling process. In this paper, we address these limitations by pioneering the selective state space model into the meteorological field downscaling and propose a novel model called *MambaDS*. This model retains the advantages of Mamba in long-range dependency modeling and linear computational complexity while enhancing the learning ability of multivariate correlation. In addition, by designing an efficient topography constraint layer, this prior information can be used more efficiently than ever before. Through extensive experiments in both China mainland and the continental United States (CONUS), we validated that our proposed MambaDS achieves state-of-the-art results in three different types of meteorological field downscaling settings. We will release the code subsequently.

Index Terms—Meteorological field downscaling, weather forecasting, state space model, super-resolution, remote sensing.

I. INTRODUCTION

IN recent decades, the increasing frequency and intensity of extreme weather events have underscored the profound impacts of climate change on human societies and natural

The work was supported by the National Natural Science Foundation of China under Grants 62125102, the National Key Research and Development Program of China (Grant No. 2022ZD0160401), the Beijing Natural Science Foundation under Grant JL23005, and the Fundamental Research Funds for the Central Universities, the National Key Research and Development Program of China (Grant No. 2022ZD0160101) and in part by the Youth Innovation Team of China Meteorological Administration (CMA2024QN02). [†]Corresponding author: Zhenwei Shi and Hao Chen (e-mail: shizhenwei@buaa.edu.cn, chen hao1@pjlab.org.cn)

Zili Liu and Zhenwei Shi are with the Image Processing Center, School of Astronautics, Beihang University, and with the State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, China, and also with Shanghai Artificial Intelligence Laboratory, Shanghai 200232, China.

Hao Chen, Lei Bai, and Wanli Ouyang are with Shanghai Artificial Intelligence Laboratory, Shanghai 200232, China.

Wenyuan Li is with the Department of Geography, University of Hong Kong, Hong Kong, China.

Zhengxia Zou is with the Department of Guidance, Navigation and Control, School of Astronautics, Beihang University, Beijing 100191, China, and also with Shanghai Artificial Intelligence Laboratory, Shanghai 200232, China.

*This work was done during his internship at Shanghai Artificial Intelligence Laboratory.

systems [1], [2]. Consequently, the need for accurate and reliable weather forecasting has never been more critical. Traditional numerical weather prediction methods [3], which have been developed over many years, along with the rapidly advancing deep learning-based forecasting models [4]–[8] in recent years, have enabled increasingly accurate global-scale weather predictions. However, due to computational resource constraints, the resolution of global-scale forecasts is limited to tens of kilometers to 100 km [9]. Such coarse spatial resolution is insufficient for the refined forecasting needs of specific regions [10] and related downstream tasks [11], [12], particularly for effectively representing extreme weather events at regional scales. As a result, using downscaling techniques to generate high-resolution weather forecasts for specific regions is essential for addressing the spatial resolution limitations of global forecast models [13]–[15].

Downscaling methods for meteorological variables can be categorized into dynamical downscaling and statistical downscaling [15]. The former, often referred to as Regional Climate Models (RCMs), uses forecast results from Global Climate Models (GCMs) as boundary conditions. It incorporates topography and other regional information to construct regional dynamical processes, solving differential equations to obtain high-resolution meteorological fields at the regional level [16]. However, dynamical downscaling methods heavily rely on differential equations to describe dynamical processes, which often results in significant biases. In addition, the numerical solution process on fine-grained regional grids is computationally intensive, requiring the support of supercomputing platforms [14]. To alleviate the computational bottlenecks of dynamical downscaling, statistical downscaling methods directly learn the mapping from coarse-resolution to fine-resolution meteorological data. Early attempts use traditional machine learning techniques, such as multiple linear regression [17], random forests [18], and support vector machines [19], to model the mapping relationship. However, the nonlinear mapping capabilities of traditional machine learning methods are relatively limited, which has constrained the performance of early statistical downscaling approaches.

In recent years, machine learning methods, particularly deep learning [20], have rapidly advanced. Due to their powerful automatic feature extraction capabilities, they have been widely applied in fields such as computer vision [21], natural language processing [22], and earth system science [23]. The downscaling task of meteorological data draws heavily

from the image super-resolution (SR) problem in the field of computer vision. The objective is to recover high-resolution images from low-resolution images, a process very similar to downscaling. Due to the vast amount of meteorological data, which aligns well with the needs of deep learning, deep learning-based image super-resolution methods have seen rapid development and widespread application in the field of meteorological downscaling [15].

Similar to the development of image super-resolution model structures [24], the structures of meteorological downscaling models are also continuously improving [15]. Convolutional Neural Networks (CNNs) and their improved methods in SR are the most commonly used model structures for meteorological field downscaling [25]–[28]. To further enhance the ability to recover fine details and textures, subsequent work has introduced generative adversarial networks (GANs) into the downscaling process [29]–[31]. While CNNs are effective at capturing local nonlinear correlations, their limited receptive field restricts the ability to capture long-term dependencies. Additionally, for large-scale data, it is challenging to continuously improve model performance using CNNs alone. Especially for large-scale meteorological data, the weather conditions of a specific region are often influenced by a broader area. Therefore, the ability to model long-term dependencies is crucial. To enhance long-term dependency modeling and improve fitting capabilities for large-scale data, Transformer-based super-resolution model [32], [33] has started to be applied to meteorological downscaling tasks [34]–[36]. However, for high-resolution meteorological fields, the quadratic computational complexity of the self-attention mechanism [37] in Transformer models leads to substantial memory consumption. Some efficient improvements to the self-attention mechanism can mitigate memory and computational costs to some extent, but they also compromise the model’s ability to capture global context [38]. Therefore, designing an efficient model with the ability to capture global context is important for meteorological downscaling tasks.

Moreover, although downscaling is quite similar to image super-resolution, directly applying SR models to downscaling tasks, as done in previous work [25], [36], has limitations. The fundamental differences between the two primarily lie in the *correlation of multiple variables* and the *influence of auxiliary prior information* [15]. The former is mainly because the downscaling process usually involves multiple variables with different distributions (such as wind speed, temperature, precipitation, etc.), and these variables often have certain physical correlations. This is distinctly different from the RGB channels in images. The latter is because fine-scale meteorological states are often influenced by factors such as topography, which can serve as prior information to guide the downscaling process. This is different from super-resolution (SR), where there is typically no obvious prior information that can be utilized. Therefore, further research is needed on how to improve super-resolution models to adapt to the characteristics of the downscaling task.

Recently, state space model (SSM), particularly its improved versions known as Mamba [39], has garnered significant attention and research in the fields of natural language pro-

cessing and computer vision due to their powerful global context modeling capabilities and efficient linear computational complexity. [40], [41]. Due to these advantages, SSMs have demonstrated immense potential in high-resolution image processing tasks. Although the original Mamba [39] was designed for modeling one-dimensional (1D) natural language sequence data, subsequent work [42], [43] has significantly enhanced its capability to process 2D image data by improving the scanning methods within the state space model (SSM). This has allowed it to surpass Transformer-based state-of-the-art (SOTA) methods in several benchmark tasks [40]. Some early attempts to use the improved Mamba architecture for image SR have also demonstrated that the SSMs could achieve SOTA performance [44]–[46]. As a result, bringing the Mamba model to the downscaling of meteorological fields to enhance performance is highly significant.

To address the aforementioned challenges and issues, we introduce *MambaDS*, which pioneers the integration of Mamba to meteorological downscaling, with specialized designs tailored to the unique requirements of downscaling. Specifically, starting from the Visual State Space Module (VSSM) [47], we first propose a *Multivariable Correlation-Enhanced VSSM (MCE-VSSM)* by introducing a channel attention branch to establish relationships between different meteorological variables and embed the MCE-VSSM into a SwinIR-like framework [32], avoiding the texture information loss typically caused by conventional UNet-like structures [33], [45], [48]. Besides, we design an *efficient topography-constraint layer* at the end of our proposed *MambaDS* to ensure physical consistency, which avoids the extra computational overhead caused by previous methods that integrated terrain information semantically into feature maps [25], [36] while ensuring the positive influence of topography priors on downscaling.

To validate the effectiveness of our method, we conducted downscaling experiments with different resolution scales using ERA5 reanalysis data [49], NOAA High-Resolution Rapid Refresh (HRRR) analysis data [50] and Fengwu forecast data [6] over the regions of mainland China and the continental United States (CONUS). The experimental results confirm the effectiveness of our proposed *MambaDS*, which outperforms previous CNN-based, Transformer-based, and vanilla Mamba super-resolution models in all experimental settings. The main contributions of the paper are as follows:

- We propose *MambaDS*, pioneering the integration of the Mamba model into meteorological downscaling, with specialized designs tailored to the unique characteristics of downscaling.
- The MCE-VSSM and specially designed topography-constraint layer improve the ability to model multivariable correlations and enhance the efficiency of utilizing topography priors during the downscaling process.
- Extensive experiments on ERA5 and HRRR data with different resolution scales demonstrate that the proposed *MambaDS* achieves state-of-the-art performance in current meteorological field downscaling tasks.

II. RELATED WORK

This section provides a brief overview of research related to meteorological field downscaling.

A. Deep Models in Image Super-Resolution

Image super-resolution, as a crucial low-level vision task, has seen rapid development in its model structures in recent years [24]. Originating from the super-resolution of natural images [51]–[53] as its core, it has extensively expanded to specific domains such as remote sensing imagery [54]–[56] and medical imaging [57].

The pioneering work Super-Resolution Convolutional Neural Network (SRCNN) [51] first demonstrated the possibility and effectiveness of deep neural networks for image super-resolution tasks. Subsequently, several improved methods [52], [58], [59] introduced more complex structures into convolutional neural networks, enhancing the modeling and learning capabilities of super-resolution models. To further enhance the model’s ability to recover detailed textures, generative methods such as Generative Adversarial Networks (GANs) [30] and Diffusion models [9] were quickly introduced into super-resolution tasks. In recent years, with the successful application of sequential modeling in the image domain, Transformer methods [32] have been extensively studied in image super-resolution and have demonstrated their superior performance across numerous downstream tasks and domains [54]. Recently, to further alleviate the quadratic computational complexity of self-attention in Transformer models, state-space models represented by Mamba [39] have begun to be applied in image super-resolution. These models have been verified to outperform previous Transformer-based super-resolution models while maintaining lower computational complexity [44], [45].

Although image super-resolution models have achieved such remarkable success and rapid development, meteorological variable downscaling has some fundamental differences from natural image and remote sensing image super-resolution tasks, which leads to limitations when directly applying super-resolution methods to downscaling tasks.

B. Deep Models in Meteorological Field Downscaling

Meteorological field downscaling can be broadly divided into dynamical downscaling and statistical downscaling based on modeling approaches [15]. This subsection will focus on the development of deep learning-based methods in statistical downscaling.

In the deep learning modeling paradigm, the downscaling process of meteorological fields is modeled as a nonlinear mapping from low-resolution meteorological grids to high-resolution grids, which is very similar to image SR. Therefore, the development of downscaling models has greatly drawn on the model structures of image SR. CNN-based SR models are the earliest and most common in the application and research of downscaling methods. Representative work includes DeepSD [25], which utilizes the SRCNN [51] to achieve the mapping from low-resolution to high-resolution meteorological fields. Subsequent improvement methods mostly

utilize advanced super-resolution models like VDSR [52] and encoder-decoder-based UNet [60] to achieve downscaling for meteorological variables such as wind [26], temperature [61], [62], and precipitation [63], [64]. Additionally, to further enhance the recovery of high-frequency texture information, generative adversarial networks (GANs) [53] have been widely used in downscaling, mainly targeting high-ratio downscaling tasks [29] and meteorological variables with richer texture information, such as precipitation [31]. Despite these advancements, CNN models have limited capability in modeling global context, which restricts the performance of CNN-based downscaling models.

To further enhance the global context modeling capability of downscaling models, recent work has started using Transformer-based SR models to improve downscaling performance. Zhong utilizes SwinIR [32] and Uformer [33] models to achieve downscaling and bias correction for near-surface temperature and wind speed forecast fields, with model performance significantly surpassing CNN methods [36]. Other Transformer-based downscaling works have also achieved excellent model performance in areas such as climate variables [34], spatiotemporal downscaling [35], and station-scale downscaling [65], [66]. However, the quadratic computational complexity caused by the self-attention mechanism in Transformer models requires significant computational power and memory when handling high-resolution images. Some efficient Transformer improvements [38] can alleviate this issue but at the cost of losing the ability to model long-term correlations.

To address the issues of CNN and Transformer models in meteorological field downscaling, we propose *MambaDS*, pioneering the integration of the Mamba [39] into downscaling tasks with special design and improvements. The introduction of the Mamba ensures the ability to maintain long-term dependence while managing model complexity, thereby enhancing the overall performance of the downscaling model.

C. Auxiliary Data for Meteorological Field Downscaling

Although meteorological field downscaling has borrowed model structures from image SR, the downscaling process is not simply a mapping from low-resolution images to high-resolution images. It is influenced by various physical auxiliary information, such as topography and other meteorological variables [15]. Among these, topography information is the most significant factor affecting the meteorological downscaling process. Therefore, it is common in previous work to incorporate it as auxiliary information into the downscaling process. For instance, early precipitation downscaling efforts, such as DeepSD [25] and Nest-UNet [61], utilized digital elevation model (DEM) data as auxiliary information. They integrated high-resolution DEM information into feature maps through addition or concatenation, guiding the restoration of detailed texture information. However, such simple integration lacks semantic feature extraction of auxiliary information, limiting the effectiveness of DEM data. Recent improvements [36] address this issue by designing specialized encoder structures for topography data to extract features at different scales

and integrate them into the downscaling process. Although experimental results show that this method can enhance downscaling performance, the additional feature encoder inevitably increases the overall model parameters and computational complexity. This makes it difficult to determine whether the performance improvement comes from increased model capacity or the incorporation of geographical information.

To more efficiently utilize topography information for enhancing and guiding downscaling, we design an *efficient topography constraint layer*. This layer imposes a hard constraint on the downscaled output through topography weighting at the final stage of the model. This approach ensures the effective use of topography information while avoiding additional parameters and computational complexity.

D. Visual State Space Model

Recently, state space models represented by Mamba [39] have gained significant attention and are widely applied in various fields such as computer vision and natural language processing [40], [41]. Its linear complexity and ability to model long-term dependence help alleviate the inherent issues present in CNN and Transformer models. Although Mamba was initially proposed for sequence data, recent work has demonstrated its strong performance in the visual and image domains [43]. Early works of visual Mamba primarily focus on image classification tasks [42], [47], [67], addressing the issue of position sensitivity in image information by introducing positional encoding and more complex scanning mechanisms. Building on this foundation, Mamba has been applied to a wider range of dense prediction tasks, including semantic segmentation [68], [69], remote sensing image change detection [70], [71], medical image segmentation [72], and even video tasks [73]. Besides, as an important low-level vision task, Mamba-based image SR works have recently begun to emerge. VMambaIR [45] embedding an improved SSM module into the UNet structure, achieving state-of-the-art results in image restoration and super-resolution tasks. Similarly, MambaIR [44], a simple but effective baseline, enhances the performance of vanilla Mamba using local enhancement and channel attention mechanisms. Xiao et al. achieved super-resolution of remote sensing images by introducing an additional frequency selection framework [74].

Although Mamba has been widely studied in various visual and image domains, research in the meteorological field is still limited. Particularly for the critical task of meteorological downscaling, it is meaningful to explore how to leverage Mamba's strengths to enhance model performance, considering the specific characteristics of the task.

III. MAMBADS

This section will provide a detailed introduction to the proposed *MambaDS* structure.

A. Preliminaries: State Space Models

Original State Space Models (SSMs) [75] are mathematical representations used to describe dynamic systems, extensively

applied in control theory, economics, signal processing, and more. They employ a set of first-order differential or difference equations to capture the system's dynamics. A common type of state space model is the Linear Time-Invariant (LTI) system [76], which uses a linear ordinary differential equation (ODE), as shown in Eq. 1, to describe the mapping relationship from the stimulation $x(t) \in \mathbb{R}$ to the response $y(t) \in \mathbb{R}$ through a latent state $h(t) \in \mathbb{R}^N$.

$$\begin{aligned} h'(t) &= \mathbf{A}h(t) + \mathbf{B}x(t), \\ y(t) &= \mathbf{C}h(t) + \mathbf{D}x(t), \end{aligned} \quad (1)$$

where $\mathbf{A} \in \mathbb{R}^{N \times N}$, $\mathbf{B} \in \mathbb{R}^{N \times 1}$, $\mathbf{C} \in \mathbb{R}^{1 \times N}$, and $\mathbf{D} \in \mathbb{R}$ represent the parameters of the ODE, N indicates the state size.

To further discretize the continuous differential equation for adaptation to deep learning algorithms, a timescale parameter Δ is used to discretize the continuous parameters \mathbf{A} and \mathbf{B} using the Zero-Order Hold (ZOH) criterion. The discretized parameters can be expressed as:

$$\begin{aligned} \bar{\mathbf{A}} &= \exp(\Delta \mathbf{A}), \\ \bar{\mathbf{B}} &= (\Delta \mathbf{A})^{-1}(\exp(\Delta \mathbf{A}) - \mathbf{I}) \cdot \Delta \mathbf{B}. \end{aligned} \quad (2)$$

Therefore, the discretized version of Eq. 1 can be expressed as:

$$\begin{aligned} h_k &= \bar{\mathbf{A}}h_{k-1} + \bar{\mathbf{B}}x_k, \\ y_k &= \mathbf{C}h_k + \mathbf{D}x_k, \end{aligned} \quad (3)$$

which is consistent with the representation of a recurrent neural network (RNN) [77]. However, the inherent structure of RNNs poses challenges for parallel processing. Additionally, the parameters of traditional SSM models are independent of input data, making it difficult to adapt to more complex dynamic systems.

To address the issues mentioned above, recent advancements in state space models, such as S4 and Mamba [39], have transformed the RNN-based representation into a parallelizable CNN form:

$$\begin{aligned} \bar{\mathbf{K}} &\triangleq (\mathbf{C}\bar{\mathbf{B}}, \overline{\mathbf{C}\bar{\mathbf{A}}\bar{\mathbf{B}}}, \dots, \mathbf{C}\bar{\mathbf{A}}^{L-1}\bar{\mathbf{B}}), \\ \mathbf{y} &= \mathbf{x} * \bar{\mathbf{K}}, \end{aligned} \quad (4)$$

where $\bar{\mathbf{K}} \in \mathbb{R}^L$ denotes the structured convolution kernel, L denotes the length of input sequence, $*$ is the convolution operation. Additionally, Mamba modifies the parameters $\bar{\mathbf{B}}$, \mathbf{C} , and Δ to be input-dependent, enhancing the modeling capability for complex systems. Parallel scan algorithms and hardware-aware algorithms have also been proposed to improve computational efficiency on GPUs.

B. Overall Structure

As shown in Fig. 1, our proposed *MambaDS* consists of three main stages: a shallow feature extraction, a residual hierarchical Mamba-based encoder, and a topography-constrained reconstruction, which is inspired by the overall architecture of SwinIR [32]. Specifically, given the input low-resolution meteorological fields $\mathcal{F}_{\text{input}} \in \mathbb{R}^{h \times w \times V}$, where h and w indicate the number of vertical and horizontal grids, V represents the number of variables. We first utilize a 3×3

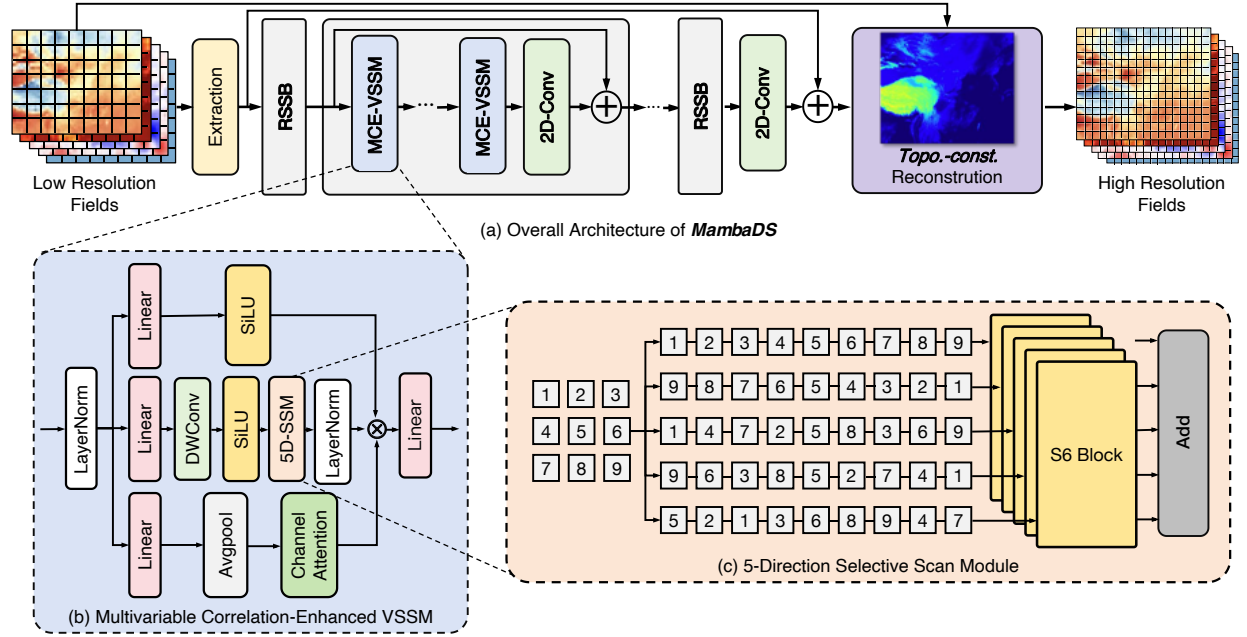


Fig. 1. Illustration of our proposed *MambaDS* (a), which include three stages. The first shallow feature extraction projects the low-resolution fields into the embedding domain. Then, the stacked Residual State Space Blocks (RSSBs) consisting of multiple Multivariable Correlation-Enhanced VSSMs (MCE-VSSMs) (b) with 5-Direction Selective Scan Module (5D-SSM) (c) are used for deep feature extraction. Finally, the efficient topography-constraint layer is proposed for high-resolution field reconstruction.

convolutional layer to embed the input field to the dimension of $\mathcal{F}_{\text{embed}} \in \mathbb{R}^{h \times w \times C}$, where C is the embedding dimension. Subsequently, $\mathcal{F}_{\text{embed}}$ extracts deep features through multiple stacked Residual State Space Blocks (RSSBs), each containing several Multivariable Correlation-Enhanced Visual State Space Modules (MCE-VSSMs) and a final convolutional layer to refine extracted features. For each RSSB, the extracted feature is denoted as $\mathcal{F}_{\text{inter}}^l \in \mathbb{R}^{h \times w \times C}$, $l \in \{1, 2, \dots, L\}$. The final deep feature obtained could be computed by $\mathcal{F}_{\text{feature}} = \mathcal{F}_{\text{inter}}^L + \mathcal{F}_{\text{embed}}$ and is combined with the input low-resolution field $\mathcal{F}_{\text{input}}$ to be fed into the topography-constrained reconstruction to achieve a high-resolution meteorological field $\mathcal{F}_{\text{output}} \in \mathbb{R}^{H \times W \times V}$ with additional topography information.

C. Multivariable Correlation-Enhanced VSS Module

According to the description above, different meteorological variable fields are treated as channel dimensions of a natural image and stacked together. This approach is also widely used in other AI for meteorology fields, such as weather forecasting [6], bias correction [36], and assimilation [78]. However, unlike similar spectral channel distributions in images like RGB, different meteorological variables typically exhibit completely different distribution characteristics. Therefore, directly applying image super-resolution methods to multivariable meteorological downscaling tasks, as done in previous downscaling works [25], [36], is limited because it overlooks the different distribution characteristics of each variable. Additionally, there are relationships between variables, often described by atmospheric differential equations.

Therefore, to enhance the modeling capability of correlations between different meteorological variables, as shown in

Fig. 1(b), we propose a Multivariable Correlation-Enhanced Visual State Space Module (MCE-VSSM). Specifically, given the internal features of the extracted field $\mathcal{F}_{\text{inter}}^l \in \mathbb{R}^{h \times w \times C}$, we apply the LayerNorm (LN) [79] first and then followed by a three-branched architecture. The output features can be computed by:

$$\begin{aligned}
 \tilde{\mathcal{F}}_{\text{inter}}^l &= \text{LN}(\mathcal{F}_{\text{inter}}^l), \\
 \tilde{\mathcal{F}}_{\text{inter}_1}^l &= \text{SiLU}(\text{Linear}(\tilde{\mathcal{F}}_{\text{inter}}^l)), \\
 \tilde{\mathcal{F}}_{\text{inter}_2}^l &= \text{Linear}(\text{Conv}(\text{SiLU}(\text{SSM}(\text{LN}(\tilde{\mathcal{F}}_{\text{inter}}^l))))), \\
 \tilde{\mathcal{F}}_{\text{inter}_3}^l &= \text{Linear}(\text{Avgpool}(\text{CA}(\tilde{\mathcal{F}}_{\text{inter}}^l))), \\
 \tilde{\mathcal{F}}_{\text{inter}_o}^l &= \text{Linear}(\tilde{\mathcal{F}}_{\text{inter}_1}^l \odot \tilde{\mathcal{F}}_{\text{inter}_2}^l \odot \tilde{\mathcal{F}}_{\text{inter}_3}^l),
 \end{aligned} \quad (5)$$

where, $\text{SSM}(\cdot)$ indicates our proposed 5-Direction Selective Scan Module, $\text{CA}(\cdot)$ represents the channel attention operation. \odot indicates the Hadamard product. In the above operation, this can be seen as integrating the correlation information between variables into the vanilla VSSM module [45], [47] in a weighted form through $\tilde{\mathcal{F}}_{\text{inter}_3}^l$, thereby enhancing the model's ability to handle multivariable data.

D. 5-Direction Selective Scan Module

The vanilla Mamba [39], designed for one-dimensional sequence data in natural language, uses only two directions (forward and backward scanning) to capture the correlation between tokens. However, for image-like 2D grid data with a non-casual nature, both temporal and spatial correlations need to be captured. Therefore, the improved method uses four-directional scanning (top-left to bottom-right, bottom-right to top-left, top-right to bottom-left, and bottom-left to top-right)

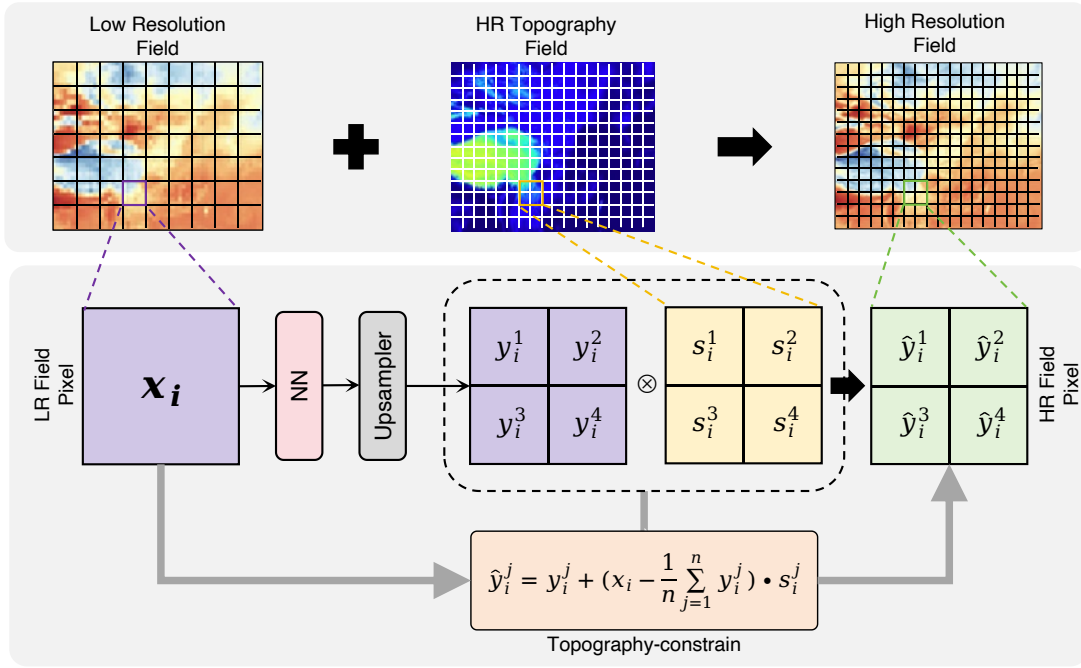


Fig. 2. Illustration of our proposed efficient topography constraint layer. For each low-resolution pixel x_i , predicted high-resolution subpixels $\{y_i^j, j = 1, \dots, n\}$ can be obtained through a downsampling model NN and an upsampler, and then topography constraints on the high-resolution subpixels are achieved using weighted constraints based on topography information.

to enhance the model’s spatiotemporal modeling capability [47].

Building on this, to address the chaotic nature of meteorological processes, and inspired by [67], we added a random scanning branch to further enhance the model’s ability to capture chaotic dynamic systems (as shown in Fig. 1(c)). Specifically, we predefine a random index list, shuffle the input tokens, and fed them into the Mamba module for feature extraction. Then, we restore the original order using the index list and integrate it with other scanning branches.

E. Efficient Topography-Constraint Layer

Topography is an important factor and prior information in the downscaling process of meteorological variables. Effectively utilizing high-resolution topographic data to guide the texture restoration of meteorological fields is key to enhancing downscaling performance [15]. This characteristic largely reflects the difference between downscaling and general natural image super-resolution, as natural images lack explicit prior texture information that can be utilized. In previous downscaling works [25], [36], [61], DEM and other topographic data have often been integrated as auxiliary information in the process. However, most studies incorporate topographic information as a separate additional input, directly integrating the raw or independently encoded topographic data into the feature vectors of the hidden layers. Although this approach can implicitly encode geographic information in the feature maps, the spatial information of high-resolution topography is lost during the feature map alignment process. This requires the model to recover detailed textures through encoded semantic features, which increases the learning difficulty to some

extent. Additionally, the extra encoding and feature extraction of topography increase the number of parameters and computational complexity of the entire downscaling model.

Therefore, inspired by previous work on hard constraints on downscaling results [80], we propose an *efficient topography constraint layer* for better preserving the detailed textures of the topography and avoiding additional computational overhead. As shown in Fig. 1(a), we placed it directly at the end of the MambaDS model to use topography information to constrain the downscaling model’s output. Therefore, the proposed layer can be applied to any other downscaling model in the future to efficiently utilize geographic information and enhance downscaling performance.

As illustrated in Fig. 2, the overall constraint process can be viewed as an additive hard constraint layer [80] weighted by DEM data. Specifically, we use high-resolution DEM data from ETOPO [81] as topography information and interpolate it to match the spatial resolution of the target meteorological field. for a given pixel x_i from low-resolution field $\mathcal{F}_{\text{input}}$, the downscaling model could produce the corresponding predicted high-resolution pixel $\{y_i^1, y_i^2, \dots, y_i^n\}$, where n represents the number of high-resolution pixels contained in x_i . For example, for 2x downscaling, $n = 4$. For x_i , we can consider its value as the average of the meteorological states at all positions covered by that pixel, i.e. $x_i = \frac{1}{n} \sum_{j=1}^n y_i^j, n \in \mathbb{N}^+$. Furthermore, topography information can be used as a weighting factor in the averaging process to make it more accurate. Therefore, the general downscaling result constrained by topography

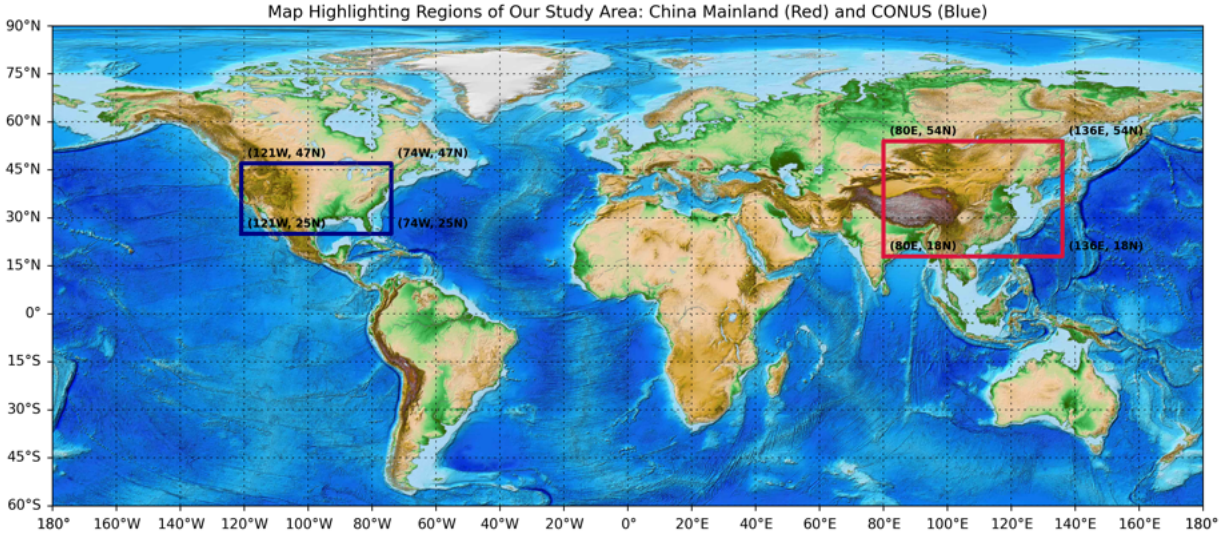


Fig. 3. The study area in this paper includes two main regions. The red box represents mainland China, with boundaries from 80°E to 136°E and 18°N to 54°N. The blue box represents the CONUS, with longitudes from 74°W to 121°W and latitudes from 25°N to 47°N.

information can be calculated as:

$$\begin{aligned} \tilde{s}_i^j &= \text{Proj}(s_i^j; \theta), \\ \hat{y}_i^j &= y_i^j + (x_i - \frac{1}{n} \sum_{j=1}^n y_i^j) \cdot \tilde{s}_i^j, \end{aligned} \quad (6)$$

where s_i^j represents the raw pixel at location (i, j) from DEM field $\mathcal{S} \in \mathbb{R}^{H \times W \times 1}$. $\text{Proj}(\cdot; \theta)$ indicates the projection operation (a CNN layer for implementation) with learnable parameter θ , which provides more flexibility for topography constraints and aligns with the dimension of the number of variables. It should be noted that to ensure that the mean x_i remains unchanged across sub-pixels, \tilde{s}_i^j needs to be normalized: $\tilde{s}_i^j = \frac{s_i^j}{\sum_{j=1}^n s_i^j} \cdot n$. Through the above operation, the model adjusts the output values based on high-resolution topography information and the mean of coarse-resolution pixels, obtaining a weighted constraint value. This efficiently uses topography information to guide the reconstruction of high-resolution textures. This allows the downscaling model to be learned in a smaller optimization subspace, thereby accelerating the convergence of the model.

IV. RESULTS

This section introduces the study area and dataset used in this paper, as well as the experimental setup and results.

A. Study Area

As shown in Fig. 3, to fully validate the effectiveness of our proposed method, we selected two regions for our study: China mainland and the Continental United States (CONUS). Specifically, for the mainland China region, we selected the research area with a boundary of 80°E to 136°E and 18°N to 54°N, aligns with the study area of our previous work [65]. For the CONUS region, we selected the area with longitudes from 74°W to 121°W and latitudes from 25°N to 47°N. This

area is the usable region of the HRRR CONUS data [50] in Lambert Conformal projection after geometric correction to a latitude-longitude grid.

We chose two study regions for two main reasons. First, we used different datasets and corresponding resolutions for experiments. These experiments roughly correspond to single-image super-resolution and real-world super-resolution tasks in image SR [44]. This helps to better validate the method's effectiveness. The second reason is that the two regions have significant differences in terrain and location. Testing the downscaling performance in different areas can better demonstrate the method's generalization capabilities.

B. Dataset Description

In this paper, we use three different datasets to validate our method. Table I summarizes our dataset.

1) *ERA5 Reanalysis*: ERA5 is a comprehensive reanalysis dataset produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) [49]. It provides detailed climate data, combining past observations with models to produce consistent and accurate records of atmospheric conditions. ERA5 covers the global climate from 1950 to the present, offering high-resolution data on various parameters such as temperature, precipitation, wind speed, and more. This dataset is widely used for climate research, weather forecasting, and environmental monitoring, offering valuable insights into long-term climate trends and variability.

In this paper, we select data from 5 years and crop the global ERA5 data for the regions of interest and selected 5 commonly used near-surface variables as the focus of the research, which are the wind speed of the u/v component (u_{10}, v_{10}) , surface pressure (sp) , air temperature at 2m (t_{2m}) and total precipitation in 1 hour (tp_{1h}) . These five variables are closely related to human activities and industry, and they are also the most widely studied variables in downscaling [15].

TABLE I
DATASET AND EXPERIMENT SETTING IN THIS PAPER. DATASET PERIOD: 2017-01-01 TO 2021-12-31.

Data Name	Data Type	Resolution	Variables	Experiment Setting	Area
ERA-5 [49]	Reanalysis	0.25° / 1 hour	$u_{10}, v_{10}, sp, t_{2m}, tp_{1h}$	4×: 1° ERA5 → 0.25° ERA5	China
HRRR [50]	Analysis	0.03125° / 1 hour	$gust, sp, t_{2m}$	8×: 0.25° ERA5 → 0.03125° HRRR	CONUS
Fengwu [6]	Forecast	0.25° / 1 hour	$u_{10}, v_{10}, sp, t_{2m}$	8×: 0.25° Fengwu → 0.03125° HRRR	CONUS

2) *NOAA High-Resolution Rapid Refresh*: The High-Resolution Rapid Refresh (HRRR) is a weather prediction model developed by the National Oceanic and Atmospheric Administration (NOAA) [50]. It provides high-frequency, short-term forecasts with a focus on the United States. Updated hourly, HRRR offers detailed data on various meteorological parameters, such as temperature, wind speed, precipitation, and cloud cover, with a high spatial resolution. This model is particularly useful for applications requiring precise, near-term weather predictions, including aviation, severe weather monitoring, and energy management.

In this paper, we use the 0-hour analysis field from the High-Resolution Rapid Refresh (HRRR) model at a resolution of 0.03125 degrees (approximately 3 km) as the downscaling target for the CONUS region. We focus on three near-surface variables: wind speed ($gust$), surface pressure (sp), and 2-meter air temperature (t_{2m}). It is important to note that the HRRR data does not provide wind speed components or accumulated precipitation data. Therefore, we have excluded these two variables in the CONUS region analysis.

3) *Fengwu Forecast*: To further validate the effectiveness of the downscaling model in practical applications, we conducted experiments on the downscaling performance for forecast fields. Specifically, we selected the Fengwu medium-range forecast model [6], developed by the Shanghai Artificial Intelligence Laboratory, as the generator of the forecast results.

The Fengwu model uses ERA5 reanalysis data as the background field and outputs global 0.25-degree hourly forecasts for the next 10 days through an autoregressive approach. To match the HRRR data, we selected the 6-hour forecast result as input for the downscaling model. Since the Fengwu forecast results are based on ERA5 as the background field, we selected four meteorological variables as input information: u_{10} , v_{10} , sp and t_{2m} . We used the forecast data for 2022 to verify the model’s generalization ability for Fengwu forecast results.

C. Experiment Setup

Building on the proposed MambaDS and the dataset mentioned above, as shown in Table I column four, we design three types of experiments to evaluate our proposed method: (1) ERA5 reanalysis downscaling, (2) HRRR analysis downscaling, and (3) Fengwu forecast downscaling. In the first experiment, the shape of input and output fields is $5 \times 36 \times 56$ and $5 \times 144 \times 224$ respectively, and sample label pairs were created by downsampling high-resolution fields. For the second and third experiments, the shape of the input and output fields is $3 \times 88 \times 188$ and $3 \times 704 \times 1504$, and the inputs and labels came from different meteorological models.

This roughly corresponds to single-image super-resolution and real-world image super-resolution tasks in the image super-resolution field [45].

1) *Architecture Details*: As mentioned in the above section, our proposed MambaDS consists of three stages. During the shallow feature extraction stage, we use a convolutional layer with a kernel size of 3 and a stride of 1 and embed the channel dimension from the number of variables V into $C = 240$. Then in the residual hierarchical Mamba-based encoder, we follow the previous Mamba-based image super-resolution work VMambaIR [45] and use four RSSBs with the depth of $\{14, 1, 1, 1\}$, and the number of channels after each RSSB remains unchanged. Furthermore, the depthwise separable convolution is used in MCE-VSSM to reduce the parameters. Finally, during the reconstruction stage, we upsample the extracted deep features with the pixel shuffle operation together with a convolutional layer.

2) *Training Details*: According to the previous works [36], we train our proposed MambaDS with the robust Charbonnier loss [82]:

$$\mathcal{L} = \sqrt{(y - \hat{y})^2 + \epsilon^2}, \quad (7)$$

where y and \hat{y} represent the predict and ground truth fields, the constant ϵ is set to 10^{-3} . This setup allows the loss function to retain the advantages of both L1 and L2 losses. We train the model for 120 epochs and optimized the model using the Adam [83] with $\beta_1 = 0.9, \beta_2 = 0.999$, employing a step learning rate initialized as $1e-4$ and decreased by half once the number of epochs reaches one of the milestones (60, 96, 108, 114). We trained our proposed model using 4x NVIDIA A100-40G GPUs, setting the batch size to 8 and 2 per GPU on experiments for China mainland and CONUS.

3) *Evaluate Metrics*: To evaluate the performance of our MambaDS, we follow the previous works [15], [25], [36] and employ four metrics: mean square error (MSE), mean absolute error (MAE), peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM), which are also widely used in image SR. It is important to note that for PSNR and SSIM, we need to adjust the maximum value range of meteorological variables to fit the different value ranges. Therefore, for the China mainland region, we modified the maximum values for different variables to $u_{10} = v_{10} = 25m/s, t_{2m} = 330K, sp = 120000Pa, tp_{1h} = 50mm$. For CONUS, $gust = 50m/s, t_{2m} = 330K, sp = 120000Pa$.

D. Performance comparison

1) *ERA5 Reanalysis Downscaling*: Downscaling the ERA5 reanalysis field is a simple baseline for assessing the performance of downscaling models and has recently been widely

TABLE II

ERA5 REANALYSIS DOWNSCALING RESULTS FOR U/V COMPONENT WIND SPEED (u_{10} , v_{10}), SURFACE PRESSURE (sp), 2M TEMPERATURE (t_{2m}) AND TOTAL PRECIPITATION IN 1 HOUR (tp_{1h}) OF VARIOUS METHODS.

Metric \ Method	EDSR [58]	RCAN [59]	SwinIR [32]	VMambaIR* [45]	MambaDS (Ours)	
Params. (M)	10.79	15.89	13.91	10.20	14.96	
FLOPs (G)	25.51	37.13	28.05	10.85	21.65	
u_{10}	MSE↓	0.1117	0.1004	0.0979	0.1169	0.0935
	MAE↓	0.2255	0.2145	0.2086	0.2308	0.1951
	PSNR↑	37.5357	37.9688	38.1699	37.4885	38.5232
	SSIM↑	0.9302	0.9354	0.9385	0.9297	0.9432
v_{10}	MSE↓	0.1170	0.1137	0.1038	0.1208	0.0989
	MAE↓	0.2314	0.2287	0.2156	0.2411	0.2046
	PSNR↑	37.3616	37.5688	37.9736	37.5786	38.3097
	SSIM↑	0.9263	0.9289	0.9341	0.9258	0.9404
sp	MSE↓	4813.1685	3824.2467	2508.9448	3424.2675	968.9953
	MAE↓	51.1968	42.9804	34.5225	39.2877	21.8904
	PSNR↑	64.8282	65.8998	67.9306	65.9887	72.0241
	SSIM↑	0.9698	0.9745	0.9887	0.9768	0.9958
t_{2m}	MSE↓	0.3710	0.3654	0.3208	0.3775	0.2998
	MAE↓	0.3795	0.3628	0.3443	0.3813	0.3126
	PSNR↑	55.0898	55.3287	55.8428	54.2487	56.3302
	SSIM↑	0.9772	0.9814	0.9876	0.9759	0.9946
tp_{1h}	MSE↓	0.0624	0.0644	0.0607	0.0622	0.0598
	MAE↓	0.0544	0.0556	0.0534	0.0542	0.0515
	PSNR↑	49.6796	49.4785	50.1078	49.6822	50.6245
	SSIM↑	0.9848	0.9837	0.9854	0.9843	0.9886

* Ours implementation.

used in various meteorological benchmark tasks [84], [85]. Although ERA5 data is difficult to obtain in real-time for practical applications, its widespread use and accessibility make experiments based on it meaningful for research and application. Therefore, as shown in Table I, following our previous work, we selected five important near-surface variables as our research subjects and conducted experiments in the China mainland region. For comparison, we selected four different downscaling models, covering CNN-based [58], [59], Transformer-based [32], and the recent Mamba-based super-resolution model [45].

Table II displays the test metrics for five variables using different methods. The results show that the proposed MambaDS method outperforms other methods across all metrics. The results also indicate that different downscaling methods effectively recover known downsampling processes. For instance, with the t_{2m} variable, various models stabilize the MAE around 0.3K. This suggests that the experimental setup is relatively simple. Although different models still show some performance differences, this indicates that even simple tasks can effectively validate the performance of various models. It should be noted that another Mamba-based super-resolution model, VMambaIR [45], performed worse than the CNN model in this task. We believe that this is because the VMambaIR model is based on the UNet structure, which can improve the efficiency of calculations and reduce the complexity of calculations under the premise of the same number of parameters, but also reduces the model capacity.

In addition, the downsampling of feature maps in the UNet structure in the encoder also leads to the loss of detailed texture information. Fig. 4 shows the visualization results of downscaling by different methods and the absolute error with the true value. It can be seen from the figure that the proposed MambaDS performs the best. Additionally, the table compares the parameter counts and computational complexity of different methods. The results show that our proposed MambaDS method has significant computational advantages over other CNN-based and Transformer-based downscaling methods while achieving better downscaling performance with lower computational costs.

TABLE III

HRRR ANALYSIS DOWNSCALING RESULTS FOR WIND SPEED ($gust$), SURFACE PRESSURE (sp), AND 2M TEMPERATURE (t_{2m}) OF VARIOUS METHODS.

Metric \ Method	EDSR [58]	RCAN [59]	SwinIR [32]	VMambaIR* [45]	MambaDS (Ours)	
$gust$	MSE↓	3.4764	3.2453	3.3408	3.5433	2.9848
	MAE↓	1.3534	1.3088	1.3256	1.3457	1.059
	PSNR↑	28.8034	29.0973	28.9535	28.8523	29.4590
	SSIM↑	0.8596	0.8515	0.8585	0.8588	0.8675
sp	MSE↓	125528.79	150519.64	123167.52	126434.65	10474.476
	MAE↓	171.7839	203.2089	171.3848	172.2387	124.3965
	PSNR↑	50.6081	49.8205	50.6904	50.5886	50.9918
	SSIM↑	0.9776	0.9763	0.9877	0.9777	0.9924
t_{2m}	MSE↓	3.9440	3.9130	3.7390	3.9385	3.4432
	MAE↓	1.3778	1.3715	1.3506	1.3724	1.3047
	PSNR↑	44.0257	44.8907	45.0114	44.2584	45.8654
	SSIM↑	0.9817	0.9828	0.9833	0.9824	0.9846

* Ours implementation.

2) *HRRR Analysis Downscaling*: The downscaling of the HRRR analysis field is a process of learning from the low-resolution ERA5 reanalysis field to the high-resolution analysis field. The two data are obtained from two different modes, so the degradation process of the meteorological field is unknown, which is similar to the real-world image super-resolution task in image super-resolution. It should be noted that since the HRRR analysis field only provides the absolute value of wind speed, but does not provide the component data of wind speed, we merged the wind speed components in ERA5 and obtained the wind speed information at the corresponding location as the input of the model, i.e. $gust = \sqrt{u_{10}^2 + v_{10}^2}$.

Table III shows the experimental results of different downscaling methods. From the results, we can see that the error of this experiment is significantly increased compared with the ERA5 downscaling experiment with known degradation processes. For example, for the temperature variable, the MSE increases by about 10 times (from about 0.3K to about 3K). We believe that this is due to the significant differences between the regional assimilation model of HRRR and the global-scale assimilation model of ERA5 in terms of parameterization and discrete grid settings. In addition, as the spatial resolution increases to the kilometer scale, the state of meteorological variables will be more affected by fine-grained scale factors such as topography, thus generating some high-frequency

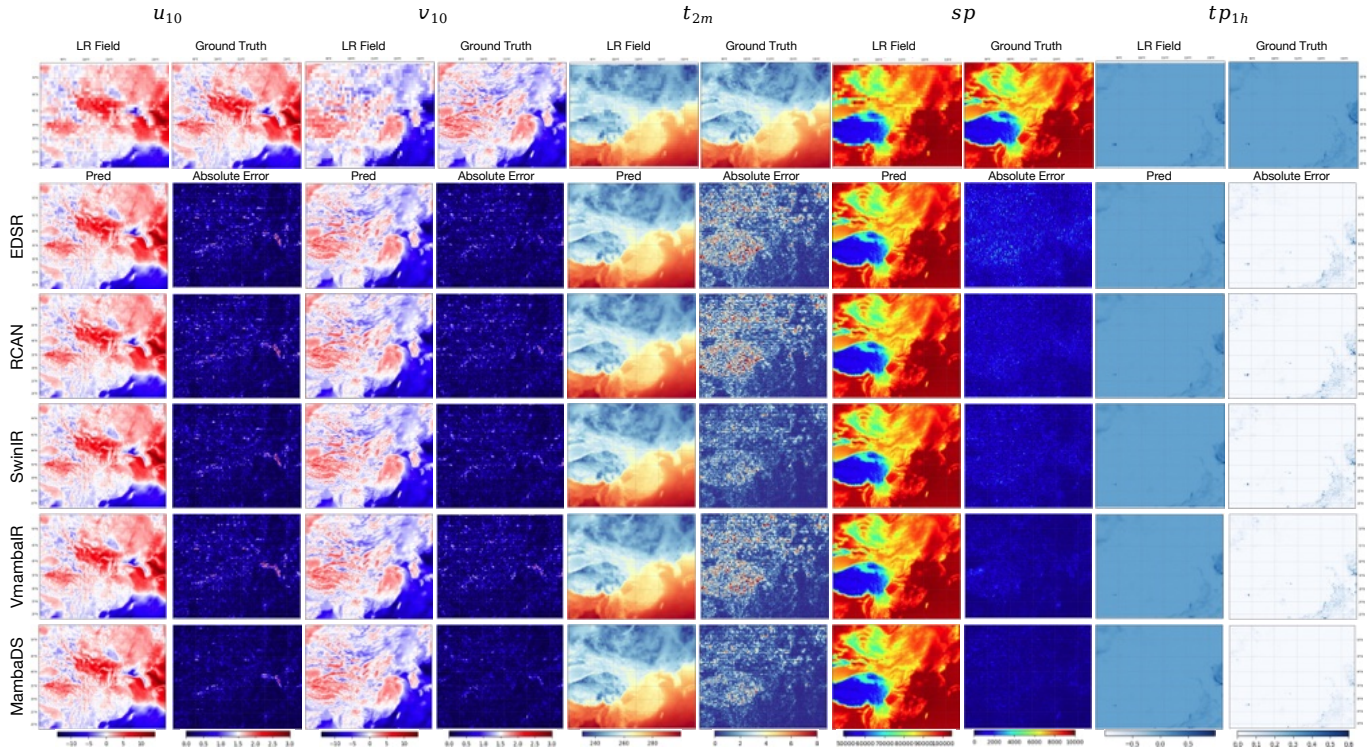


Fig. 4. Visualization comparison of ERA5 reanalysis downscaling using different methods. The first row contains the low-resolution meteorological fields of each variable (obtained by downsampling the high-resolution meteorological fields) and the high-resolution GT fields. The following lines show the input results of different downscaling methods and the absolute error map with GT. As can be seen from the figure, the MambaDS proposed in this paper shows the smallest error for all variables, that is, the downscaling performance is optimal.

texture information, which is mostly smoothed at the global scale. Although the indicators have changed significantly, the relative performance ranking of different downscaling models remains unchanged. Our proposed MambaDS method still significantly outperforms other methods in all metrics.

Fig. 5 shows the downscaling results of different methods. From the figure, we can see that the output results (LR field and ground truth) of different scale modes are obviously different. This is mainly because the input LR field smoothes most of the high-frequency textures, and the state values at some positions are significantly lower than the HRRR analysis values. Especially for wind speed variables, the ERA5 reanalysis field has obvious underestimates. Therefore, the reconstruction process is more complicated than the downscaling process of ERA5. It can also be seen from the results in the figure that the error distribution range of different variables is closely related to terrain factors. For example, for temperature variables, there is obviously a large error at the junction of land and sea. For pressure, it is more concentrated near the lake (upper right corner). Therefore, it is very important to include terrain data in the downscaling process.

3) *Fengwu Forecast Downscaling*: In order to further verify the effectiveness of the proposed method and get close to the actual downscaling business application scenario, we designed a downscaling task for the forecast field. Instead of using the existing numerical forecast results, we chose Fengwu [6], an artificial intelligence medium-term forecast model that has developed rapidly in recent years, as our forecast model.

TABLE IV
FENGWU FORECAST DOWNSCALING RESULTS FOR WIND SPEED (*gust*), SURFACE PRESSURE (*sp*), AND 2M TEMPERATURE (*t2m*) OF VARIOUS METHODS.

Metric \ Method	EDSR [58]	RCAN [59]	SwinIR [32]	VMambaIR* [45]	MambaDS (Ours)	
<i>gust</i>	MSE↓	3.9195	3.8374	3.7461	3.8137	3.7031
	MAE↓	1.4411	1.4027	1.4050	1.4298	1.3924
	PSNR↑	28.2947	28.6243	28.7245	28.5749	28.7425
	SSIM↑	0.8524	0.8544	0.8564	0.8523	0.8573
<i>sp</i>	MSE↓	94139.38	121930.62	91899.40	101245.76	80295.97
	MAE↓	162.9230	195.6916	162.2511	163.5641	135.3585
	PSNR↑	51.4932	49.2467	50.2414	50.2453	52.3823
	SSIM↑	0.9777	0.9734	0.9755	0.9786	0.9793
<i>t2m</i>	MSE↓	5.3664	4.6805	6.1633	6.8214	4.5310
	MAE↓	1.7275	1.5707	1.8705	1.9156	1.5080
	PSNR↑	42.5323	43.4257	42.1425	42.2465	43.6783
	SSIM↑	0.9828	0.9822	0.9824	0.9812	0.9884

* Ours implementation.

Specifically, we used ERA5 as the forecast initial field of Fengwu and selected the 6-hour forecast results as the input of the downscaling model. It should be noted that we directly use the HRRR analysis field downscaling model to perform zero-shot inference instead of training from scratch, so as to verify the generalization ability of the model to a certain extent. There are three reasons for this setting: First, we believe that the previous two experiments have fully demonstrated

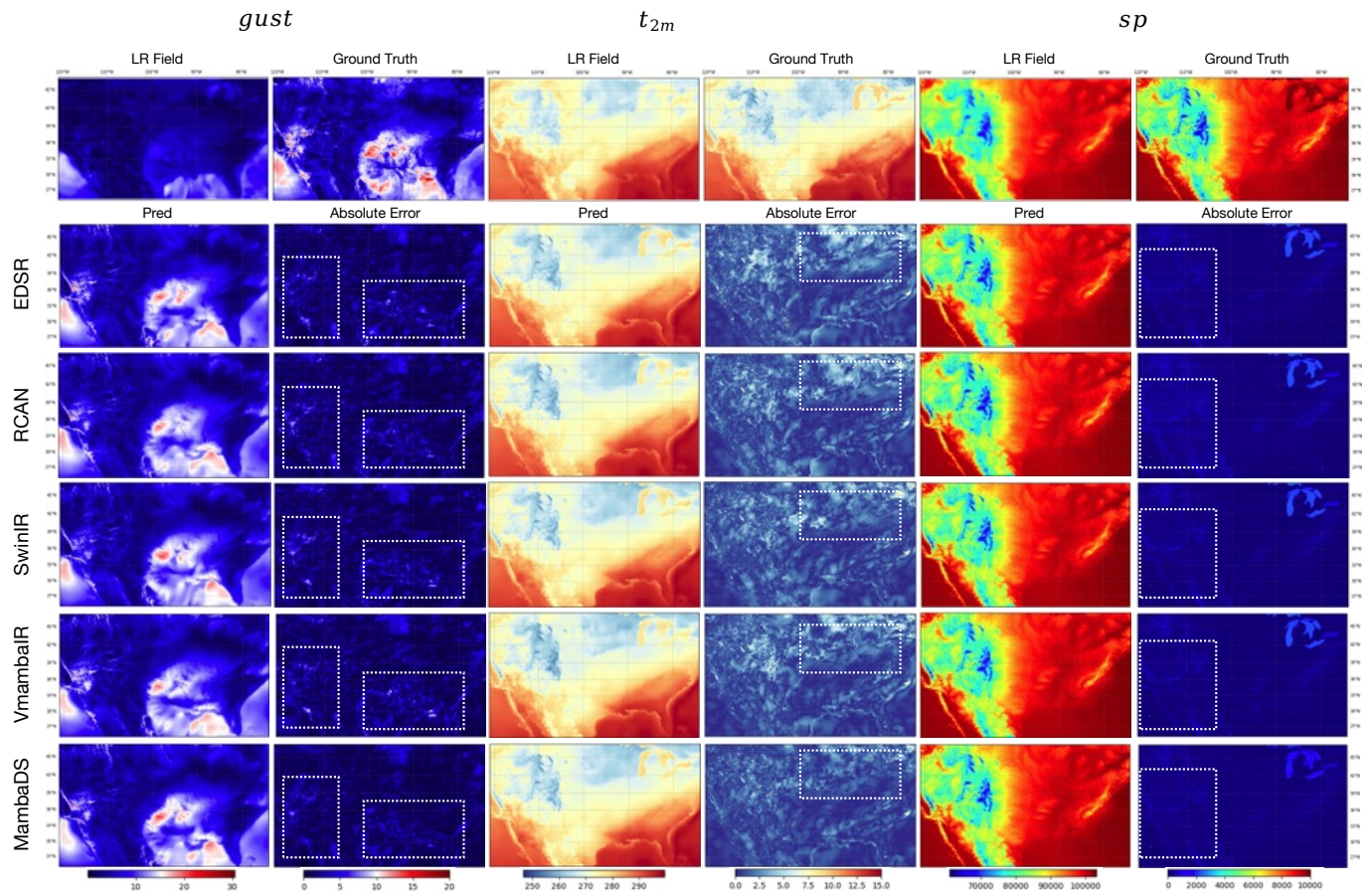


Fig. 5. Visualization comparison of HRRR analysis downscaling using different methods. The first row contains the low-resolution meteorological fields of each variable (obtained by downsampling the high-resolution meteorological fields) and the high-resolution GT fields. The following lines show the input results of different downscaling methods and the absolute error map with GT. As can be seen from the figure, the MambaDS proposed in this paper shows the smallest error for all variables, that is, the downscaling performance is optimal.

the ability of different methods to reconstruct the target input field, and it is redundant to train or fine-tune from scratch based on the Fengwu forecast results. In addition, in actual applications, we usually face forecast results from different models, and their distributions are also different. Therefore, the generalization ability of different downscaling methods for different inputs is more important. Finally, compared with multi-temporal forecast results, the acquisition and storage of the analysis field at time 'f00' is more convenient and economical. A large number of previous works [5], [6] are also mostly based on analysis/reanalysis fields for pre-training to adapt to different downstream inputs.

Table IV and Fig. 6 show the zero-shot inference results of Fengwu forecast results using different methods. From the results, we can see that the relative performance rankings of different methods for different variables have changed to a certain extent. For example, for the t_{2m} , the CNN-based methods EDSR [58] and RCAN [59] outperform the Transformer-based SwinIR [32], which is different from the results in the first two experiments. Despite this, the MambaDS model proposed in this paper still shows a stable advantage in all variables, especially for sp , which is more obvious. We believe that this is also due to the positive effect of adding

topography information on surface pressure, a variable that is highly correlated with terrain. The results of this experiment show that the MambaDS proposed in this paper has a stronger generalization ability on the basis of better reconstruction ability than other models. This ability is inseparable from the addition of geographic information and the advantages of the model structure.

E. Ablation Studies

To further analyze the impact of different components of the proposed MambaDS on the model performance, we conducted ablation experiments. The experimental results listed in Table 3 show that each component in MambaDS improves the performance of the model to varying degrees, and adding all innovative components to MambaDS will achieve the best downscaling performance. Specifically, it can be seen that the addition of the MCE-VSSM module has improved all variables to varying degrees, which proves that compared with the vanilla VSSM module, the enhanced branch of multivariate correlation does provide additional modeling capabilities and model capacity for the downscaling process. As for the addition of the scanning mechanism of 5D-SSM, it can be seen that although it does not show a positive effect on all variables

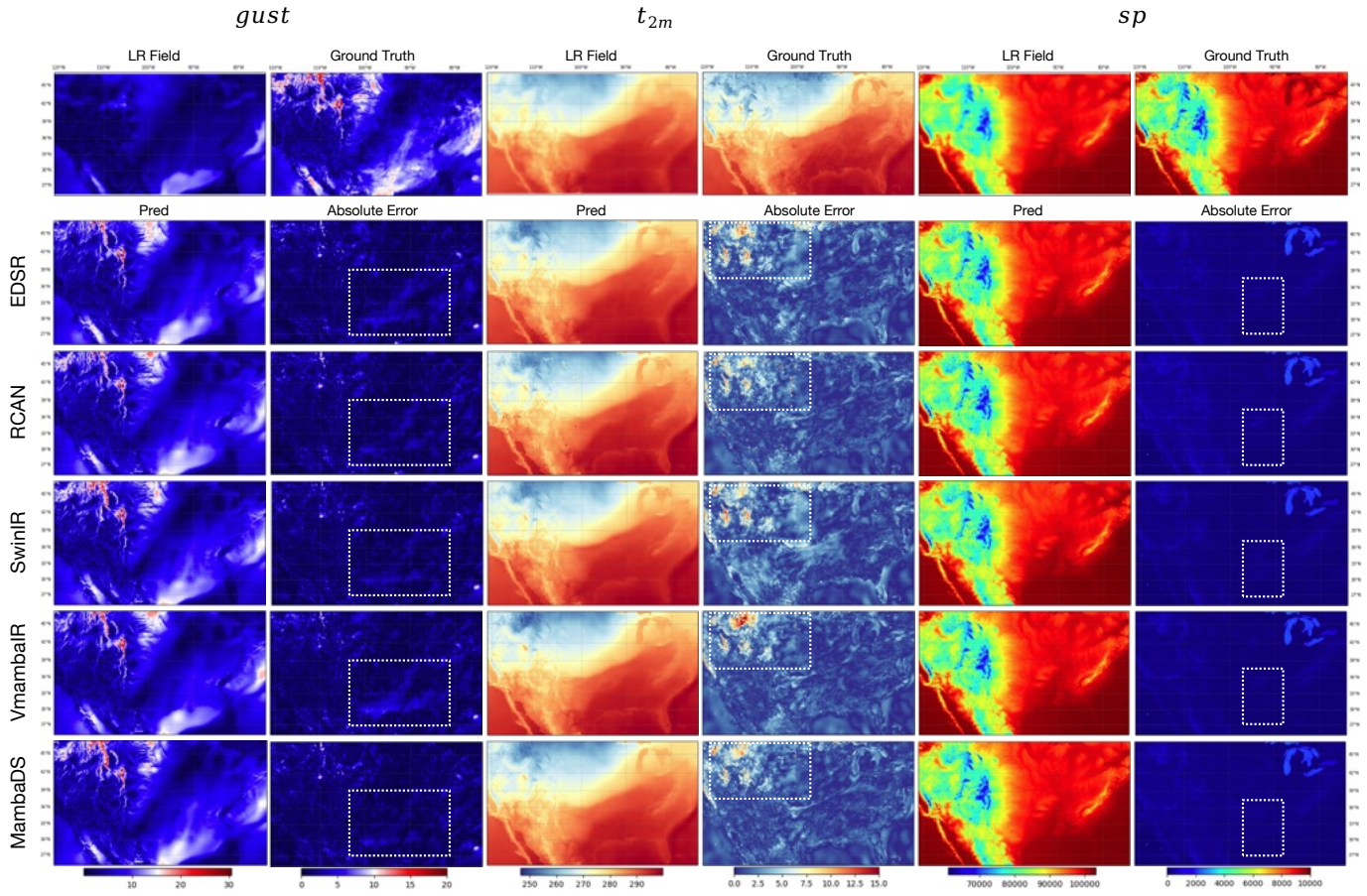


Fig. 6. Visualization comparison of Fengwu forecasts downscaling using different methods. The first row contains the low-resolution meteorological fields of each variable (obtained by downsampling the high-resolution meteorological fields) and the high-resolution GT fields. The following lines show the input results of different downscaling methods and the absolute error map with GT. As can be seen from the figure, the MambaDS proposed in this paper shows the smallest error for all variables, that is, the downscaling performance is optimal.

TABLE V
ABLATION STUDY RESULTS OF MAMBADS.

Method	Settings			t_{2m}		sp		u_{10}		v_{10}		tp_{1h}	
	MCE-VSSM	5D-SSM	Topo.	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
MambaDS	✗	✗	✗	0.3406	0.3552	2137.8452	32.9356	0.1016	0.2173	0.1048	0.2217	0.0612	0.0539
	✓	✗	✗	0.3256	0.3462	1844.8749	28.8756	0.1008	0.2025	0.1044	0.2145	0.0602	0.0541
	✓	✓	✗	0.3148	0.3414	1749.3832	27.9723	0.1012	0.2137	0.1037	0.2128	0.0606	0.0533
	✓	✓	✓	0.2998	0.3126	968.9953	21.8904	0.0935	0.1951	0.0989	0.2046	0.0598	0.0515

(such as u_{10}), it has a positive effect on other variables. This also reflects to a certain extent that this module helps to improve the global modeling capabilities for meteorological downscaling tasks. Finally, the addition of the topography constraint layer can be seen to have the most significant effect on the performance improvement of the model. In particular, for temperature and pressure variables, which are closely related to topography, the addition of terrain information can greatly improve the downscaling performance.

F. Comparison of different topography prior integration methods

In the downscaling of meteorological variables, topography data is important prior information. To verify the effectiveness of the topography constraint layer ('soft-Topo.' in Table VI) proposed in this paper, we compare our approach with previous topography fusion methods based on feature extraction [36] ('hard-Topo.' in Table VI). Previous feature extraction-based methods use an independent CNN encoder to extract multi-scale topography features and fuse the features of the corresponding scales into the downscaling model. This can be seen as a soft constraint to achieve the constraints on the topography. We compare the downscaling performance of

the two methods under the ERA5 downscaling setting and further compare the performance of using the two methods simultaneously ('hard&soft-Topo.' in Table VI).

From the results in Table 3, it can be seen that compared with the model without incorporating topography information ('wo Topo.'), the introduction of terrain information can significantly improve the downscaling effect of different variables. For the a priori integration methods of different approaches, it can be seen that the proposed topography constraint layer achieves downscaling results that are similar to or even exceed those of other methods without a significant increase in the number of parameters. This also further verifies the effectiveness and efficiency of using topography data as a hard constraint weight on the output.

TABLE VI
ERA5 FORECAST DOWNSCALING RESULTS FOR WIND SPEED ($gust$), SURFACE PRESSURE (sp), AND 2M TEMPERATURE (t_{2m}) OF VARIOUS TOPOGRAPHY PRIOR INTEGRATION METHODS.

Metric \ Method					
Variable		MambaDS wo Topo.	MambaDS w/ soft-Topo.	MambaDS w/ hard-Topo. (Ours)	MambaDS w/ hard&soft-Topo.
Params.(M)		14.96	17.04	14.96	17.04
u_{10}	MSE↓	0.1012	0.0997	0.0935	0.0933
	MAE↓	0.2137	0.2015	0.1951	0.1947
	PSNR↑	38.0154	38.2467	38.5232	38.5217
	SSIM↑	0.9362	0.9402	0.9432	0.9433
v_{10}	MSE↓	0.1037	0.1004	0.0989	0.0993
	MAE↓	0.2128	0.2085	0.2046	0.2050
	PSNR↑	37.9488	38.1958	38.3097	38.3846
	SSIM↑	0.9369	0.9396	0.9404	0.9411
sp	MSE↓	1749.3832	1032.8832	968.9953	988.2949
	MAE↓	27.9723	24.5729	21.8904	22.4638
	PSNR↑	68.4298	70.2472	72.0241	72.3533
	SSIM↑	0.9913	0.9923	0.9958	0.9965
t_{2m}	MSE↓	0.3148	0.3073	0.2998	0.2977
	MAE↓	0.3414	0.3156	0.3126	0.3122
	PSNR↑	56.0368	56.2465	56.3302	56.3422
	SSIM↑	0.9903	0.9924	0.9946	0.9953
tp_{1h}	MSE↓	0.0606	0.0593	0.0598	0.0588
	MAE↓	0.0533	0.0524	0.0515	0.0517
	PSNR↑	49.9935	50.2567	50.6245	50.6105
	SSIM↑	0.9842	0.9857	0.9886	0.9893

V. DISCUSSION

This work is the first to introduce state-space models like Mamba into meteorological variable downscaling tasks. Through task-specific design for downscaling, our proposed MambaDS method surpasses previous CNN-based and Transformer-based downscaling models in both performance and efficiency. In this section, we will further analyze the limitations of our work and discuss future directions for Mamba-based downscaling methods.

The overall framework of our proposed MambaDS method follows SwinIR [32], which maintains the spatial dimensions of input images without downsampling, rather than using the encoder-decoder structure of UNet which is more common in the image super-resolution field. While this approach maximally preserves spatial texture information and thus improves downscaling performance, it increases computational

complexity. Therefore, further research is needed on how to effectively embed the Mamba structure into UNet-like architectures to further reduce model computational complexity while maintaining model performance. Although there have been some early attempts in natural image restoration and super-resolution tasks [45], based on our reproduction of these related methods, we found that model performance still has certain limitations. Furthermore, for state-space models, the scanning pattern is crucial. Currently, most improvements to Mamba in computer vision-related work are primarily focused on enhancing scanning methods [47], [67], [70]. Therefore, how to design scanning patterns suitable for multivariate meteorological fields is another direction that needs further research.

VI. CONCLUSION

In this paper, we pioneer the selective state space model into meteorological field downscaling and propose a novel downscale model namely *MambaDS*. Compared to previous downscaling methods based on CNN and Transformer super-resolution models, MambaDS can model long-range dependencies while maintaining efficient linear computational complexity. Building on this, we made specific designs and improvements tailored to the characteristics of the downscaling task. Specifically, we first designed the MCE-VSSM to enhance the modeling of correlations between different meteorological variables, addressing the need for downscaling multiple variables simultaneously. Additionally, we designed a model-agnostic efficient topography constraint layer that uses hard constraints to directly guide the restoration of fine texture details in meteorological fields with high-resolution topography data, which ensures performance while avoiding additional computational overhead. Through extensive experimental comparisons of three different downscale settings and two study areas, we have verified the effectiveness of our proposed model for multivariable near-surface meteorological field downscaling.

REFERENCES

- [1] P. Stott, "How climate change affects extreme weather events," *Science*, vol. 352, no. 6293, pp. 1517–1518, 2016.
- [2] C. Lesk, P. Rowhani, and N. Ramankutty, "Influence of extreme weather disasters on global crop production," *Nature*, vol. 529, no. 7584, pp. 84–87, 2016.
- [3] P. Bauer, A. Thorpe, and G. Brunet, "The quiet revolution of numerical weather prediction," *Nature*, vol. 525, no. 7567, pp. 47–55, 2015.
- [4] R. Lam, A. Sanchez-Gonzalez, M. Willson, P. Wirmsberger, M. Fortunato, F. Alet, S. Ravuri, T. Ewalds, Z. Eaton-Rosen, W. Hu *et al.*, "Learning skillful medium-range global weather forecasting," *Science*, vol. 382, no. 6677, pp. 1416–1421, 2023.
- [5] K. Bi, L. Xie, H. Zhang, X. Chen, X. Gu, and Q. Tian, "Accurate medium-range global weather forecasting with 3d neural networks," *Nature*, vol. 619, no. 7970, pp. 533–538, 2023.
- [6] K. Chen, T. Han, J. Gong, L. Bai, F. Ling, J.-J. Luo, X. Chen, L. Ma, T. Zhang, R. Su *et al.*, "Fengwu: Pushing the skillful global medium-range weather forecast beyond 10 days lead," *arXiv preprint arXiv:2304.02948*, 2023.
- [7] T. Kurth, S. Subramanian, P. Harrington, J. Pathak, M. Mardani, D. Hall, A. Miele, K. Kashinath, and A. Anandkumar, "Fourcastnet: Accelerating global high-resolution weather forecasting using adaptive fourier neural operators," in *Proceedings of the platform for advanced scientific computing conference, 2023*, pp. 1–11.

- [8] D. Kochkov, J. Yuval, I. Langmore, P. Norgaard, J. Smith, G. Mooers, M. Klöwer, J. Lottes, S. Rasp, P. Düben *et al.*, “Neural general circulation models for weather and climate,” *Nature*, pp. 1–7, 2024.
- [9] F. Ling, Z. Lu, J.-J. Luo, L. Bai, S. K. Behera, D. Jin, B. Pan, H. Jiang, and T. Yamagata, “Diffusion model-based probabilistic downscaling for 180-year east asian climate reconstruction,” *npj Climate and Atmospheric Science*, vol. 7, no. 1, p. 131, 2024.
- [10] E. R. Rodrigues, I. Oliveira, R. Cunha, and M. Netto, “Deepdownscale: A deep learning strategy for high-resolution weather forecast,” in *2018 IEEE 14th International Conference on e-Science (e-Science)*. IEEE, 2018, pp. 415–422.
- [11] J. Younis, S. Anquetin, and J. Thielen, “The benefit of high-resolution operational weather forecasts for flash flood warning,” *Hydrology and Earth System Sciences*, vol. 12, no. 4, pp. 1039–1051, 2008.
- [12] A. d. De Wit, H. Boogaard, and C. Van Diepen, “Spatial resolution of precipitation and radiation: the effect on regional crop yield forecasts,” *Agricultural and Forest Meteorology*, vol. 135, no. 1-4, pp. 156–168, 2005.
- [13] R. L. Wilby and T. M. Wigley, “Downscaling general circulation model output: a review of methods and limitations,” *Progress in physical geography*, vol. 21, no. 4, pp. 530–548, 1997.
- [14] Z. Xu, Y. Han, and Z. Yang, “Dynamical downscaling of regional climate: A review of methods and limitations,” *Science China Earth Sciences*, vol. 62, pp. 365–375, 2019.
- [15] Y. Sun, K. Deng, K. Ren, J. Liu, C. Deng, and Y. Jin, “Deep learning in statistical downscaling for deriving high spatial resolution gridded meteorological data: A systematic review,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 208, pp. 14–38, 2024.
- [16] R. A. Pielke Sr and R. L. Wilby, “Regional climate downscaling: What’s the point?” *Eos, Transactions American Geophysical Union*, vol. 93, no. 5, pp. 52–53, 2012.
- [17] J. T. Schoof and S. Pryor, “Downscaling temperature and precipitation: A comparison of regression-based methods and artificial neural networks,” *International Journal of Climatology: A Journal of the Royal Meteorological Society*, vol. 21, no. 7, pp. 773–790, 2001.
- [18] R. J. Davy, M. J. Woods, C. J. Russell, and P. A. Coppin, “Statistical downscaling of wind variability from meteorological fields,” *Boundary-layer meteorology*, vol. 135, pp. 161–175, 2010.
- [19] H. Chen, J. Guo, W. Xiong, S. Guo, and C.-Y. Xu, “Downscaling gcms using the smooth support vector machine method to predict daily precipitation in the hanjiang basin,” *Advances in Atmospheric Sciences*, vol. 27, pp. 274–284, 2010.
- [20] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [21] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [22] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding,” *arXiv preprint arXiv:1810.04805*, 2018.
- [23] C. Irrgang, N. Boers, M. Sonnewald, E. A. Barnes, C. Kadow, J. Staneva, and J. Saynisch-Wagner, “Towards neural earth system modelling by integrating artificial intelligence in earth system science,” *Nature Machine Intelligence*, vol. 3, no. 8, pp. 667–674, 2021.
- [24] Z. Wang, J. Chen, and S. C. Hoi, “Deep learning for image super-resolution: A survey,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 10, pp. 3365–3387, 2020.
- [25] T. Vandal, E. Kodra, S. Ganguly, A. Michaelis, R. Nemani, and A. R. Ganguly, “DeepSD: Generating high resolution climate change projections through single image super-resolution,” in *Proceedings of the 23rd acm sigkdd international conference on knowledge discovery and data mining*, 2017, pp. 1663–1672.
- [26] K. Höhle, M. Kern, T. Hewson, and R. Westermann, “A comparative study of convolutional neural network models for wind field downscaling,” *Meteorological Applications*, vol. 27, no. 6, p. e1961, 2020.
- [27] L. Sun and Y. Lan, “Statistical downscaling of daily temperature and precipitation over china using deep learning neural models: Localization and comparison with other methods,” *International Journal of Climatology*, vol. 41, no. 2, pp. 1128–1147, 2021.
- [28] R. Tie, C. Shi, G. Wan, X. Hu, L. Kang, and L. Ge, “Cldassd: reconstructing fine textures of the temperature field using super-resolution technology,” *Advances in Atmospheric Sciences*, vol. 39, no. 1, pp. 117–130, 2022.
- [29] K. Stengel, A. Glaws, D. Hettinger, and R. N. King, “Adversarial super-resolution of climatological wind and solar data,” *Proceedings of the National Academy of Sciences*, vol. 117, no. 29, pp. 16 805–16 815, 2020.
- [30] L. Harris, A. T. McRae, M. Chantry, P. D. Dueben, and T. N. Palmer, “A generative deep learning approach to stochastic downscaling of precipitation forecasts,” *Journal of Advances in Modeling Earth Systems*, vol. 14, no. 10, p. e2022MS003120, 2022.
- [31] P. Hess, M. Drüke, S. Petri, F. M. Strnad, and N. Boers, “Physically constrained generative adversarial networks for improving precipitation fields from earth system models,” *Nature Machine Intelligence*, vol. 4, no. 10, pp. 828–839, 2022.
- [32] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, “Swinir: Image restoration using swin transformer,” in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 1833–1844.
- [33] Z. Wang, X. Cun, J. Bao, W. Zhou, J. Liu, and H. Li, “Uformer: A general u-shaped transformer for image restoration,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 17 683–17 693.
- [34] F. Gerges, M. C. Boufadel, E. Bou-Zeid, H. Nassif, and J. T. Wang, “A novel deep learning approach to the statistical downscaling of temperatures for monitoring climate change,” in *Proceedings of the 2022 6th International Conference on Machine Learning and Soft Computing*, 2022, pp. 1–7.
- [35] L. Xiang, J. Guan, J. Xiang, L. Zhang, and F. Zhang, “Spatiotemporal model based on transformer for bias correction and temporal downscaling of forecasts,” *Frontiers in Environmental Science*, vol. 10, p. 1039764, 2022.
- [36] X. Zhong, F. Du, L. Chen, Z. Wang, and H. Li, “Investigating transformer-based models for spatial downscaling and correcting biases of near-surface temperature and wind-speed forecasts,” *Quarterly Journal of the Royal Meteorological Society*, vol. 150, no. 758, pp. 275–289, 2024.
- [37] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” *Advances in neural information processing systems*, vol. 30, 2017.
- [38] Q. Fournier, G. M. Caron, and D. Aloise, “A practical survey on faster and lighter transformers,” *ACM Computing Surveys*, vol. 55, no. 14s, pp. 1–40, 2023.
- [39] A. Gu and T. Dao, “Mamba: Linear-time sequence modeling with selective state spaces,” *arXiv preprint arXiv:2312.00752*, 2023.
- [40] X. Wang, S. Wang, Y. Ding, Y. Li, W. Wu, Y. Rong, W. Kong, J. Huang, S. Li, H. Yang *et al.*, “State space model for new-generation network alternative to transformers: A survey,” *arXiv preprint arXiv:2404.09516*, 2024.
- [41] B. N. Patro and V. S. Agneeswaran, “Mamba-360: Survey of state space models as transformer alternative for long sequence modelling: Methods, applications, and challenges,” *arXiv preprint arXiv:2404.16112*, 2024.
- [42] L. Zhu, B. Liao, Q. Zhang, X. Wang, W. Liu, and X. Wang, “Vision mamba: Efficient visual representation learning with bidirectional state space model,” *arXiv preprint arXiv:2401.09417*, 2024.
- [43] R. Xu, S. Yang, Y. Wang, B. Du, and H. Chen, “A survey on vision mamba: Models, applications and challenges,” *arXiv preprint arXiv:2404.18861*, 2024.
- [44] H. Guo, J. Li, T. Dai, Z. Ouyang, X. Ren, and S.-T. Xia, “Mambair: A simple baseline for image restoration with state-space model,” *arXiv preprint arXiv:2402.15648*, 2024.
- [45] Y. Shi, B. Xia, X. Jin, X. Wang, T. Zhao, X. Xia, X. Xiao, and W. Yang, “Vmambair: Visual state space model for image restoration,” *arXiv preprint arXiv:2403.11423*, 2024.
- [46] Y. Lu, S. Wang, Z. Wang, P. Xia, T. Zhou *et al.*, “Lfmamba: Light field image super-resolution with state space model,” *arXiv preprint arXiv:2406.12463*, 2024.
- [47] Y. Liu, Y. Tian, Y. Zhao, H. Yu, L. Xie, Y. Wang, Q. Ye, and Y. Liu, “Vmamba: Visual state space model,” 2024. [Online]. Available: <https://arxiv.org/abs/2401.10166>
- [48] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, “Swin-unet: Unet-like pure transformer for medical image segmentation,” in *European conference on computer vision*. Springer, 2022, pp. 205–218.
- [49] H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers *et al.*, “The era5 global reanalysis,” *Quarterly Journal of the Royal Meteorological Society*, vol. 146, no. 730, pp. 1999–2049, 2020.
- [50] D. C. Dowell, C. R. Alexander, E. P. James, S. S. Weygandt, S. G. Benjamin, G. S. Manikin, B. T. Blake, J. M. Brown, J. B. Olson, M. Hu *et al.*, “The high-resolution rapid refresh (hrrr): An hourly updating convection-allowing forecast model. part i: Motivation and system description,” *Weather and Forecasting*, vol. 37, no. 8, pp. 1371–1395, 2022.

- [51] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 2, pp. 295–307, 2015.
- [52] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1646–1654.
- [53] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4681–4690.
- [54] S. Lei, Z. Shi, and W. Mo, "Transformer-based multistage enhancement for remote sensing image super-resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–11, 2021.
- [55] S. Lei and Z. Shi, "Hybrid-scale self-similarity exploitation for remote sensing image super-resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–10, 2021.
- [56] S. Lei, X. Xiao, H.-C. Li, Z. Shi, and Q. Zhu, "Exploring fine-grained image-text alignment for referring remote sensing image segmentation," *arXiv preprint arXiv:2409.13637*, 2024.
- [57] H. Yang, Z. Wang, X. Liu, C. Li, J. Xin, and Z. Wang, "Deep learning in medical image super resolution: a review," *Applied Intelligence*, vol. 53, no. 18, pp. 20891–20916, 2023.
- [58] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017, pp. 136–144.
- [59] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 286–301.
- [60] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer, 2015, pp. 234–241.
- [61] Y. Sha, D. J. Gagne II, G. West, and R. Stull, "Deep-learning-based gridded downscaling of surface meteorological variables in complex terrain. part i: Daily maximum and minimum 2-m temperature," *Journal of Applied Meteorology and Climatology*, vol. 59, no. 12, pp. 2057–2073, 2020.
- [62] A. Doury, S. Somot, S. Gadat, A. Ribes, and L. Corre, "Regional climate model emulator based on deep learning: Concept and first evaluation of a novel hybrid downscaling approach," *Climate Dynamics*, vol. 60, no. 5, pp. 1751–1779, 2023.
- [63] Y. Sha, D. J. Gagne II, G. West, and R. Stull, "Deep-learning-based gridded downscaling of surface meteorological variables in complex terrain. part ii: Daily precipitation," *Journal of Applied Meteorology and Climatology*, vol. 59, no. 12, pp. 2075–2092, 2020.
- [64] A. Y. Sun and G. Tang, "Downscaling satellite and reanalysis precipitation products using attention-based deep convolutional neural nets," *Frontiers in Water*, vol. 2, p. 536743, 2020.
- [65] Z. Liu, H. Chen, L. Bai, W. Li, K. Chen, Z. Wang, W. Ouyang, Z. Zou, and Z. Shi, "Observation-guided meteorological field downscaling at station scale: A benchmark and a new method," *arXiv preprint arXiv:2401.11960*, 2024.
- [66] W. Li, Z. Liu, K. Chen, H. Chen, S. Liang, Z. Zou, and Z. Shi, "Deep-physinet: Bridging deep learning and atmospheric physics for accurate and continuous weather modeling," *arXiv preprint arXiv:2401.04125*, 2024.
- [67] K. Chen, B. Chen, C. Liu, W. Li, Z. Zou, and Z. Shi, "Rsmamba: Remote sensing image classification with state space model," *IEEE Geoscience and Remote Sensing Letters*, 2024.
- [68] Z. Wan, Y. Wang, S. Yong, P. Zhang, S. Stepputtis, K. Sycara, and Y. Xie, "Sigma: Siamese mamba network for multi-modal semantic segmentation," *arXiv preprint arXiv:2404.04256*, 2024.
- [69] Q. Zhu, Y. Cai, Y. Fang, Y. Yang, C. Chen, L. Fan, and A. Nguyen, "Samba: Semantic segmentation of remotely sensed images with state space model," *arXiv preprint arXiv:2404.01705*, 2024.
- [70] S. Zhao, H. Chen, X. Zhang, P. Xiao, L. Bai, and W. Ouyang, "Rsmamba for large remote sensing image dense prediction," *arXiv preprint arXiv:2404.02668*, 2024.
- [71] H. Zhang, K. Chen, C. Liu, H. Chen, Z. Zou, and Z. Shi, "Cdmamba: Remote sensing image change detection with mamba," *arXiv preprint arXiv:2406.04207*, 2024.
- [72] J. Ma, F. Li, and B. Wang, "U-mamba: Enhancing long-range dependency for biomedical image segmentation," *arXiv preprint arXiv:2401.04722*, 2024.
- [73] Y. Yang, Z. Xing, and L. Zhu, "Vivim: a video vision mamba for medical video object segmentation," *arXiv preprint arXiv:2401.14168*, 2024.
- [74] Y. Xiao, Q. Yuan, K. Jiang, Y. Chen, Q. Zhang, and C.-W. Lin, "Frequency-assisted mamba for remote sensing image super-resolution," *arXiv preprint arXiv:2405.04964*, 2024.
- [75] J. D. Hamilton, "State-space models," *Handbook of econometrics*, vol. 4, pp. 3039–3080, 1994.
- [76] R. L. Williams, D. A. Lawrence *et al.*, *Linear state-space control systems*. John Wiley & Sons, 2007.
- [77] L. R. Medsker, L. Jain *et al.*, "Recurrent neural networks," *Design and Applications*, vol. 5, no. 64-67, p. 2, 2001.
- [78] K. Chen, T. Chen, P. Ye, H. Chen, K. Chen, T. Han, W. Ouyang, and L. Bai, "Fnp: Fourier neural processes for arbitrary-resolution data assimilation," *arXiv preprint arXiv:2406.01645*, 2024.
- [79] J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer normalization," *arXiv preprint arXiv:1607.06450*, 2016.
- [80] P. Harder, A. Hernandez-García, V. Ramesh, Q. Yang, P. Sattigeri, D. Szwarcman, C. Watson, and D. Rolnick, "Hard-constrained deep learning for climate downscaling," *Journal of Machine Learning Research*, vol. 24, no. 365, pp. 1–40, 2023.
- [81] M. MacFerrin, C. Amante, K. Carignan, M. Love, and E. Lim, "The earth topography 2022 (etopo 2022) global dem dataset," *Earth System Science Data Discussions*, vol. 2024, pp. 1–24, 2024.
- [82] P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud, "Two deterministic half-quadratic regularization algorithms for computed imaging," in *Proceedings of 1st international conference on image processing*, vol. 2. IEEE, 1994, pp. 168–172.
- [83] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [84] P. Ren, N. B. Erichson, S. Subramanian, O. San, Z. Lukic, and M. W. Mahoney, "Superbench: A super-resolution benchmark dataset for scientific machine learning," *arXiv preprint arXiv:2306.14070*, 2023.
- [85] S. Rasp, P. D. Dueben, S. Scher, J. A. Weyn, S. Mouatadid, and N. Thurey, "Weatherbench: a benchmark data set for data-driven weather forecasting," *Journal of Advances in Modeling Earth Systems*, vol. 12, no. 11, p. e2020MS002203, 2020.



Zili Liu (Graduate Student Member, IEEE) received the B.S. and M.S. degrees from the School of Science, China University of Mining and Technology, Beijing in 2017 and 2021. He is currently working toward the PhD degree in the Image Processing Center, School of Astronautics, Beihang University.

His research interests include deep learning, AI for meteorology, remote sensing image processing, and efficient deep learning.



Hao Chen received the B.S. and Ph.D. degrees from the Image Processing Center, School of Astronautics, Beihang University, Beijing, China, in 2017 and 2023, respectively.

He is currently a Researcher at Shanghai Artificial Intelligence (AI) Laboratory, Shanghai, China. His research interests include geospatial machine learning, remote sensing, Earth monitoring, and prediction.



Lei Bai received the Ph.D. degree from the University of New South Wales, Sydney, NSW, Australia, in 2021.

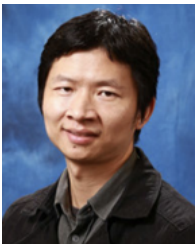
He was a Post-Doctoral Researcher with the University of Sydney, Camperdown, NSW, Australia. He is currently a Research Scientist with Shanghai Artificial Intelligence (AI) Laboratory, Shanghai, China. He has authored or co-authored a set of peer-reviewed papers in top AI conferences and journals, such as Neural Information Processing Systems (NeurIPS), Conference on Computer Vision and Pattern Recognition (CVPR), International Joint Conference on Artificial Intelligence (IJCAI), Knowledge Discovery and Data Mining (KDD), International Conference on Computer Vision (ICCV), International Conference on Ubiquitous Computing (UbiComp), IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (TPAMI), and IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS (TITS). His research interests include machine learning, spatial-temporal learning, and their applications (e.g., Earth System Science and Smart City).

Dr. Bai is or was a Program Committee Member or Reviewer for IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, NeurIPS, International Conference on Machine Learning (ICML), International Conference on Learning Representations (ICLR), CVPR, ICCV, Association for the Advancement of Artificial Intelligence (AAAI), IJCAI, KDD, European Conference on Computer Vision (ECCV), IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON MULTIMEDIA, and ACM Transactions on Sensor Networks. He was a recipient of the 2020 Google Ph.D. Fellowship, the 2020 UNSW Engineering Excellence Award, and the 2021 Dean's Award for Outstanding Ph.D. Theses.



Wenyuan Li received the B.S. degree from North China Electric Power University, Beijing, China, in 2017, and the Ph.D. degree from Beihang University, Beijing, in 2023.

He is currently a Post-Doctoral Fellow with the Department of Geography, The University of Hong Kong, Hong Kong. His research interests encompass remote sensing image processing and agricultural crop mapping.



Wanli Ouyang received the Ph.D. degree from the Department of Electronic Engineering, Chinese University of Hong Kong, Hong Kong, in 2010.

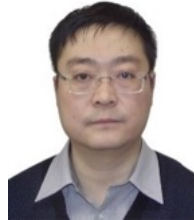
He was an Associate Professor with The University of Sydney, Camperdown, NSW, Australia. He is now a Professor with Shanghai Artificial Intelligence (AI) Laboratory, Shanghai, China. His research interests include pattern recognition, machine learning, and AI for Science.

Dr. Ouyang served as an Associate Editor for International Journal of Computer Vision (IJCV) and Pattern Recognition (PR), the Senior Area Chair for Conference on Computer Vision and Pattern Recognition (CVPR), and the Guest Editor for IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (TPAMI).



Zhengxia Zou (Senior Member, IEEE) received the B.S. and Ph.D. degrees from Beihang University, Beijing, China, in 2013 and 2018, respectively. He is currently a Professor with the School of Astronautics, Beihang University. From 2018 to 2021, he was a Postdoctoral Research Fellow at the University of Michigan, Ann Arbor, MI, USA. His research interests include remote sensing image processing and computer vision. He has published more than 60 peer-reviewed papers in top-tier journals and conferences, including Nature Communications,

Proceedings of the IEEE, IEEE Transactions on Image Processing, IEEE Transactions on Geoscience and Remote Sensing, and IEEE/CVF Computer Vision and Pattern Recognition. Dr. Zou serves as an Associate Editor for IEEE Transactions on Image Processing (TIP). His personal website is <https://zhengxiazou.github.io/>.



Zhenwei Shi (Senior Member, IEEE) is currently a Professor and the Dean of the Image Processing Center, School of Astronautics, Beihang University, Beijing, China. He has authored or co-authored over 200 scientific articles in refereed journals and proceedings. His research interests include remote sensing image processing and analysis, computer vision, pattern recognition, and machine learning.

Prof. Shi serves as an Editor for IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, Pattern Recognition, ISPRS Journal of Photogrammetry and Remote Sensing, Infrared Physics and Technology, and so on. His personal website is <http://levir.buaa.edu.cn/>.