Elevation Data Integration Approaches for Deep Learning-Based 2-m Temperature Downscaling

Antigoni Moira[1], Stelios Karozis[1], Theodoros Giannakopoulos[2], Effrosyni Karakitsou[1], Nikolaos Gounaris[1] and Athanasios Sfetsos[1]

[1]*Environmental Research Laboratory, Institute of Nuclear & Radiological Sciences & Technology, Energy & Safety, National Centre of Scientific Research Demokritos, 15341 Agia Paraskevi, Greece*

[2]*Computational Intelligence Laboratory, Institute of Informatics and Telecommunications, National Centre of Scientific Research Demokritos, 15341 Agia Paraskevi, Greece*

Abstract

Global Climate Models (GCMs) provide valuable climate projections, but operate at coarse spatial resolutions, limiting their usefulness for local-scale applications. Downscaling techniques are therefore essential to bridge this gap. This study investigates how the integration of elevation data can improve the performance of CNN-based architecture deep learning models to downscale the near-surface air temperature (T2m) from a 0.5°×0.5° grid to a 0.25°×0.25° resolution. We evaluate different elevation data integration strategies and demonstrate their impact on downscaling effectiveness, highlighting the role of terrain-related features in refining temperature estimates.

Keywords

Climate Downscaling, T2m, Elevation, DEM, Super-resolution, EDSR, Deep Learning

1. Introduction

The growing impact of climate change on various aspects of daily life underscores the increasing importance of accurate weather and climate forecasts. Weather and climate predictions can be improved by increasing their resolution in time and in space. Regional Climate Models (RCMs) are usually created by applying statistical downscaling methods to GCMs, using local observations. These methods often require a lot of input data, expert knowledge and computing resources, which makes them hard to use over large areas. On the other hand, dynamical downscaling methods use low-resolution (LR) GCMs as boundary conditions for high-resolution (HR) RCMs. These approaches rely on physical laws to simulate local climate processes and, although they are also resource intensive, they are usually better at capturing extreme events in regions with highly variable topographic characteristics [1].

Recently, Machine Learning (ML) has introduced powerful new tools. These data-driven approaches have the potential to outperform the statistical models, by modeling nonlinear patterns and taking better advantage of the provided HR data. Convolutional neural networks (CNNs), in particular, are well suited for spatial data and super-resolution (SR) tasks. They can learn complex relationships and important features, and have already shown promising results in weather downscaling [2].

The problem of downscaling climate data can be seen as a Single-Image Super-Resolution (SISR) task, which allows the use of various Deep Learning (DL) methods developed in the field of computer vision. SISR is an image processing technique aimed at reconstructing a HR image from a single LR input [3]. Recently, several studies have applied SISR techniques to tackle climate downscaling challenges [4, 5, 6, 7].

To improve the performance of these models, additional information is often included, depending on the variable being downscaled, the region of interest, and the input/output resolution. This additional

Workshop on AI-driven Data Engineering and Reusability for Earth and Space Sciences (DARES'25), co-located with the 28th European Conference on Artificial Intelligence (ECAI 2025), Bologna, Italy, October 25, 2025

a.moira@ipta.demokritos.gr (A. Moira); skarozis@ipta.demokritos.gr (S. Karozis); tyianak@iit.demokritos.gr (T. Giannakopoulos); e.karakitsou@ipta.demokritos.gr (E. Karakitsou); gounaris@ipta.demokritos.gr (N. Gounaris); ts@ipta.demokritos.gr (A. Sfetsos)

0000-0002-4914-9784 (A. Moira); 0000-0002-8727-0104 (S. Karozis); 0000-0003-1634-824X (T. Giannakopoulos); 0000-0002-3206-7010 (E. Karakitsou); 0000-0003-1906-8059 (A. Sfetsos)

data can include latitude and longitude (especially for large regions covering different climate zones), terrain characteristics (such as elevation and land use/cover), seasonal patterns, and more.

In *"Deep learning downscaled high-resolution daily near surface meteorological datasets over East Asia"* Lin etal. (2023) [8] developed a bias correction and downscaling approach, named BC-UNet, to construct a Climate Change for East Asia with Bias Corrected UNet Dataset (CLIMEA-BCUD) based on CMIP6. The BC-UNet downscaling approach firstly applied Quantile Delta Mapping (QDM) to correct CMIP6 models and then the UNet is trained for downscaling the biased corrected GCM dataset. CLIMEA-BCUD provides nine meteorological variables including 2-m air temperature, 2-m daily maximum and minimun air temperature, precipitation, 10-m wind speed, 2-m relative humidity, 2-m specific humidity, downward shortwave radiation and downward longwave radiation with 0.1° horizontal resolution at daily intervals over the historical period of 1950–2014 and three future scenarios (SSP1-2.6, SSP2-4.5 and SSP5-8.5) of 2015–2100. In their implementation, they used static elevation data from the Global 30 Arc-Second Elevation dataset as an additional input channel to the model, alongside the other input data, representing the simplest and most direct way to integrate elevation into the model.

In their study, Sha et al.(2020) [9] modify UNet, a semantic-segmentation CNN, and apply it to the downscaling of daily maximum/minimum 2-m temperature (TMAX/TMIN) over the western continental United States from 0.258° to 4-km grid spacings. They selected HR elevation, LR elevation and LR TMAX/TMIN as inputs and trained UNet using Parameter–Elevation Regressions on Independent Slopes Model (PRISM) data over the south- and central-western United States from 2015 to 2018. They found that the original UNet cannot generate enough fine-grained spatial details when transferred to the new northwestern U.S. domain. In response, they modified the original UNet by assigning an extra HR elevation output branch/loss function and training the modified UNet to reproduce both the supervised HR TMAX/TMIN and the unsupervised HR elevation. They named this improvement "UNet-Autoencoder(AE)." The UNet-AE showed better gridpoint-level performance with more than 10% mean absolute error (MAE) reduction relative to the original UNet.

In the work of Bhakare et al.(2025) [10], three machine learning models examined and specifically, Artificial Neural Network (ANN), Random Forest (RF), and Convolutional Neural Network (CNN), for downscaling of seasonal forecasts of daily minimum temperature from 12 km to 250 m horizontal resolution. The study area was part of the Trentino-Alto Adige region in northern Italy, characterized by complex terrain (elevations ranging from 166 m to 2628 m) and varied local microclimatic conditions. Data with spatial resolution 0.125° × 0.125° (approximately 12 km) from the fifth-generation Seasonal Forecast System SEAS5 operated by the European Centre for Medium-Range Weather Forecasts (ECMWF) used as predictors, while the target was the daily minimum temperature on a 250 m grid developed by interpolating more than 200 station observations in Trentino—South Tyrol. Elevation data from NASA's Shuttle Radar Topography Mission were also used as a static predictor, having been spatially aggregated to achieve approximately 250 m resolution, matching the pixel size of the target data. ANN and the RF model were provided with 9 dynamic features and 1 static feature (elevation), while the CNN model is provided with 9 dynamic features without explicitly providing elevation as an additional feature, relying on CNN's ability to understand spatial features on its own. Results suggest that CNN outperforms ANN and RF, achieving lower Mean Absolute Error (MAE) (1.59 °C to 2.03 °C). In a previous work by Bhakare et al. (2024) [11], where they attempted downscaling of daily mean temperature for the same region but with a different resolution step (from 9 km ERA5-Land reanalysis to 1 km), they also described how they derived a few more static features from the elevation, such as slope, aspect, cross-sectional curvature (C-curv), and longitudinal curvature (L-curv), in order to obtain more auxiliary predictors. The schematic of these features is shown in Figure 1.

The aim of this study is to examine whether the integration of elevation data can enhance the performance of a CNN-based DL model for downscaling T2m from a 0.5°×0.5° grid to a 0.25°×0.25° resolution. Several elevation integration strategies, derived from the related literature, as well as their combinations, were tested to identify the most effective approach for the given task.

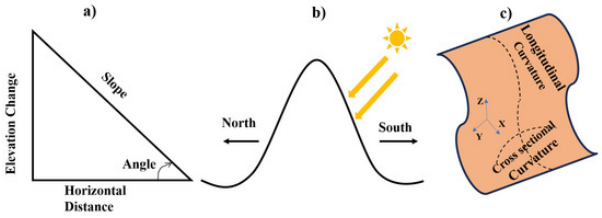

Figure 1: Schematic for static features derived from elevation, including: (a) slope, (b) aspect, and (c) longitudinal and cross-sectional curvatures (L-curv, C-curv). [11]

2. Methodology

The Enhanced Deep Super-Resolution (EDSR) network, a widely used CNN architecture for improving spatial resolution, was selected for this study. An initial baseline model, excluding elevation data, was first evaluated. Next, three elevation integration strategies were implemented and tested to assess their impact on downscaling performance and to identify the most effective architecture. Finally, an ablation study was conducted on the two selected elevation-derived features (aspect and slope) and their combinations with elevation to examine whether they improved performance.

2.1. Model Architecture

The EDSR network is a high-performing CNN model commonly applied to image super-resolution tasks. It builds upon residual learning by removing unnecessary components, such as batch normalization layers, which were shown to hinder performance in image restoration tasks. EDSR utilizes a very deep residual network with a simplified structure and greater model capacity, allowing it to learn more powerful mappings from low-resolution to HR images[12].

To conserve computational resources, a simplified EDSR model is used in this study, as the focus is on comparing elevation integration strategies rather than model architecture. The model includes four residual blocks, each consisting of two 3×3 convolutional layers with LeakyReLU activation and a scaling factor (0.1) to stabilize training. It begins with an initial convolutional layer, followed by the four residual blocks, and includes a skip connection that adds the initial features back to the processed ones. After feature extraction, an upsampling layer is applied to double the resolution of the input image. This architecture, shown in Figure 2, is used in combination with the elevation integration strategies described below.

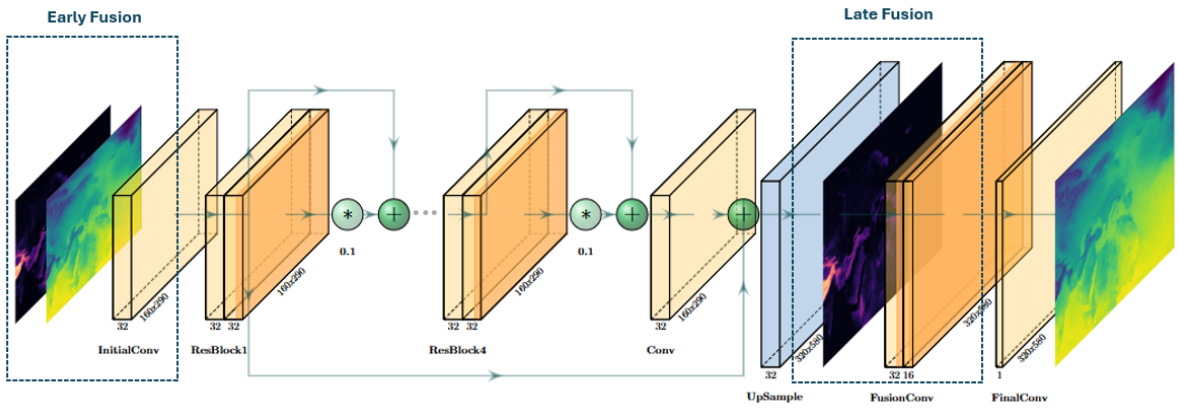

Figure 2: The architecture of the EDSR model with elevation integration methods.

2.2. Elevation Integration Strategies

Informed by the literature review presented in the Introduction section, which suggests that incorporating elevation information from Digital Elevation Models (DEMs) can positively influence prediction accuracy, the following approaches were explored to further enhance the performance of the EDSR model:

1. **Early Fusion:** The first approach involved adding elevation data as a second input channel alongside the primary input, allowing the model to process spatial and elevation features in parallel. However, this method required the DEM data to be upscaled to match the resolution of the input grid, which may result in a loss of fine-grained elevation details.
2. **Late Fusion:** In an attempt to address the low-resolution limitation of early fusion, a second approach was tested. In this case, HR DEM data was introduced after the upsampling stage of the model as a second channel to the intermediate output. This was followed by two fusion convolutional layers that incorporated the elevation information into the final prediction. This method aimed to preserve the spatial richness of the DEM and enable more meaningful integration into the model's output.
3. **Early & Late Fusion Combination:** Since each of the above strategies has its own limitations, a third model was developed that combines both approaches simultaneously. The architecture of this hybrid model is illustrated in Figure 2. This integrated design yielded the most significant improvement in performance, as demonstrated in the Results section.
4. **Elevation & Elevation-derived Features Fusion:** Finally, inspired by the approach in the literature review study by Bhakare et al. (2024) [11], additional static features were extracted from the available DEM data. Specifically, slope and aspect were derived and incorporated as extra input channels alongside elevation, both in the early and late fusion stages.

3. Experimental Setup

This section outlines the procedures and resources used to conduct the experiments. It describes the dataset and study area, the preprocessing steps applied to the input and target data, the evaluation metrics used to assess model performance and the implementation details of the deep learning models.

3.1. Dataset

The source of ground truth data used to train the DL models is the ERA5 dataset [13], available through the Copernicus Climate Data Store. ERA5 is the fifth-generation reanalysis product, by the ECMWF, providing global climate and weather data from 1940 to the present at hourly temporal resolution. Specifically, the T2m variable is used, representing the air temperature at 2 meters above the surface of land and sea. In addition, the Land-Sea Mask parameter is employed to compute certain evaluation metrics over land areas, where temperature variability is more noticeable on both a daily and seasonal basis. This parameter indicates the proportion of land within each grid cell (ranging from 0 to 1) and is a dimensionless value that distinguishes land from ocean or inland water bodies such as lakes, rivers, reservoirs and coastal waters.

The present study focuses on a broad region of the Northern Hemisphere, selected to encompass all four major climate zones (tropical, subtropical, temperate, and polar), as well as a range of topographic complexity. The study domain extends from the vicinity of Greenland to the equator and the Indian Ocean, covering mountainous regions and relatively flat terrain. GRIB files were obtained from the Copernicus Climate Data Store for the period 2000 to 2020, with a temporal resolution of 6 hours (00:00, 06:00, 12:00 and 18:00). The spatial extent spans latitudes from 80.0°N to 0.0°N and longitudes from 60.0°W to 85.0°E, capturing substantial climatic and topographic variability across the region.

The training input consists of ERA5 reanalysis data paired with altitude information from the U.S. Geological Survey 3D Elevation Program DEM [14], integrated using various strategies to enhance the

performance of the deep learning models. The original spatial resolution of 0.25° × 0.25° was upscaled to 0.5° × 0.5° using bicubic interpolation for the T2m and DEM grids, in order to generate the LR dataset. This LR dataset serves as the input for the deep learning model, which is trained to reconstruct the corresponding HR target data from the original ERA5 reanalysis.

Finally, two additional static features were derived from the available elevation data: slope and aspect. Aspect refers to the compass direction a slope faces and is measured in radians from north. These features were included at both low and high resolutions alongside elevation, aiming to enhance the model's performance.

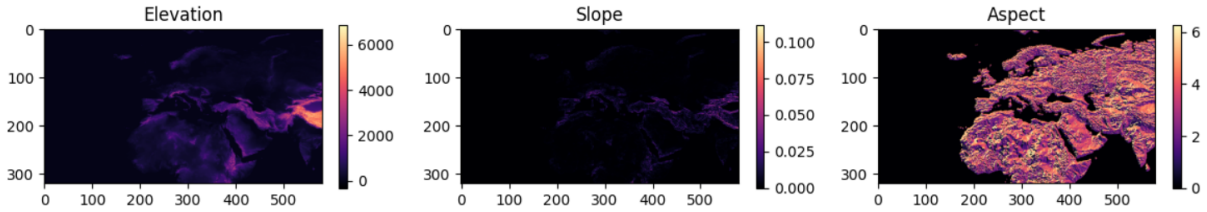

Figure 3: Elevation-derived Features.

3.2. Data Preprocessing

Before evaluating the elevation integration strategies described in the Methodology section, the dataset was uniformly preprocessed to ensure a fair comparison of results. The data was first standardized and then partitioned into three subsets: 70% for training, 15% for validation, and 15% for testing.

For standardization, the z-score normalization technique was applied, transforming the data, so that most values falling within the range of -1 to 1. This required calculating the global mean and standard deviation across the entire dataset, yielding a mean of 286.30798 and a standard deviation of 64.25680. The z-score was computed for each grid cell using the formula:

$$z = \frac{(X - Mean)}{Std}$$

Finally, to eliminate temporal dependencies and ensure randomness, the data was shuffled prior to splitting, disrupting its original chronological order. A fixed random seed was also used to guarantee reproducibility, enabling consistent data partitioning across different experiments and ensuring a fair comparison of integration strategies.

3.3. Evaluation Metrics and Loss Function

To evaluate the performance of the investigated elevation integration methods for temperature grid downscaling, four statistical metrics were employed: Mean Absolute Error (MAE), Mean Squared Error (MSE), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index (SSIM). Since this task can be framed as a single-image super-resolution (SISR) problem, these metrics effectively capture both pixel-wise accuracy and structural fidelity.

MSE, the most commonly used full-reference quality metric, quantifies the average squared differences between predicted and actual temperature values, emphasizing larger errors. PSNR measures the ratio between the maximum possible signal and the level of reconstruction error, providing a logarithmic assessment of prediction quality. To assess perceptual quality, SSIM is included, which evaluates structural similarity based on luminance, contrast, and spatial texture [15]. Finally, MAE offers an intuitive and interpretable measure of the average absolute error, making it especially useful for real-world temperature applications. Together, these metrics provide a comprehensive assessment of numerical accuracy and spatial structure preservation.

For model training, MAE was selected as the loss function instead of MSE. Also known as L1 loss, MAE computes the average of the absolute differences between predicted and actual values [16]. Unlike

MSE, which squares these differences and disproportionately penalizes larger errors, MAE treats all errors linearly, making it less sensitive to outliers [17]. This characteristic is particularly beneficial in temperature downscaling, where localized extremes or noise could otherwise distort the learning process. Moreover, MAE aligns more closely with real-world temperature interpretation, as it retains the original units (e.g., kelvin) and directly reflects the average deviation [18]. This improves the interpretability of the model's performance and ensures that the objective remains focused on minimizing overall error across all regions, rather than overemphasizing extreme discrepancies.

3.4. Implementation Details

During this study, DL models were optimized using the Adam optimizer. This choice is based on the fact that Adam is computationally efficient, has little memory requirement, is invariant to diagonal rescaling of gradients, and is well suited for problems that are large in terms of data or parameters. Although Adam dynamically adapts the learning rate during the training process, training can sometimes stall or converge slowly. Therefore, a ReduceLROnPlateau scheduler was applied to decrease the learning rate when the loss function stops improving for 70 epochs. A batch size of 32 was used and the models were trained for 250 epochs. Finally, the early stopping technique was employed, triggered after 100 epochs without improvement in model performance.

All data preprocessing and the training of the DL models were performed using Google Colab notebooks, utilizing the A100 GPU with the high-RAM runtime option enabled. The implementation of the models was carried out using the TensorFlow [19] framework, while for data preprocessing tasks and data splitting, the scikit-learn (sklearn) library [20] was employed. Visualization of training progress, model performance metrics and output comparisons was performed using Matplotlib [21].

3.5. Post-Processing and Model Evaluation

To effectively evaluate each method, Land MAE was calculated in both normalized units and degrees Celsius to provide a meaningful interpretation of the temperature differences. Since temperature is strongly affected by terrain and can vary significantly over land throughout the day, MAE was calculated specifically for land areas to provide a more meaningful evaluation. This approach helps avoid bias introduced by large water bodies, such as oceans, which have high heat capacity and tend to moderate temperature changes[22]. Including these regions in the evaluation could underestimate errors, as temperature fluctuation over water is generally much lower than over land.

4. Results

In this section, the results obtained from each approach in the study are presented and analyzed. The performance of the different methods is evaluated using the selected metrics, enabling a direct comparison. First, a simple EDSR model was trained without incorporating any DEM data, serving as the baseline. Subsequently, each of the elevation integration methods described in the Methodology section was tested to enhance the baseline model's performance. Specifically, the Early Fusion model integrates elevation as an additional input channel alongside the low-resolution T2m data. The Late Fusion model tests the effect of incorporating high-resolution elevation after upscaling, by fusing it as a second channel just before the final output. The Early & Late Fusion approach combines both strategies simultaneously.

Figure 4 presents the training history and validation loss for each of the four models, providing a visual comparison and facilitating interpretation of their effectiveness. The baseline shows relatively stable convergence but higher loss compared to all elevation-integrated strategies. This indicates that incorporating elevation data improves model performance. Among the tested strategies, Early & Late Fusion achieved the lowest overall loss curves. In this case, the learning rate decreased later (at epoch 212) compared to the other strategies, suggesting that the model continued to learn effectively for more epochs before requiring a reduction. This delayed activation of the learning rate scheduler is

advantageous, as it indicates that the model maintained steady improvements over a longer period, ultimately leading to better convergence.

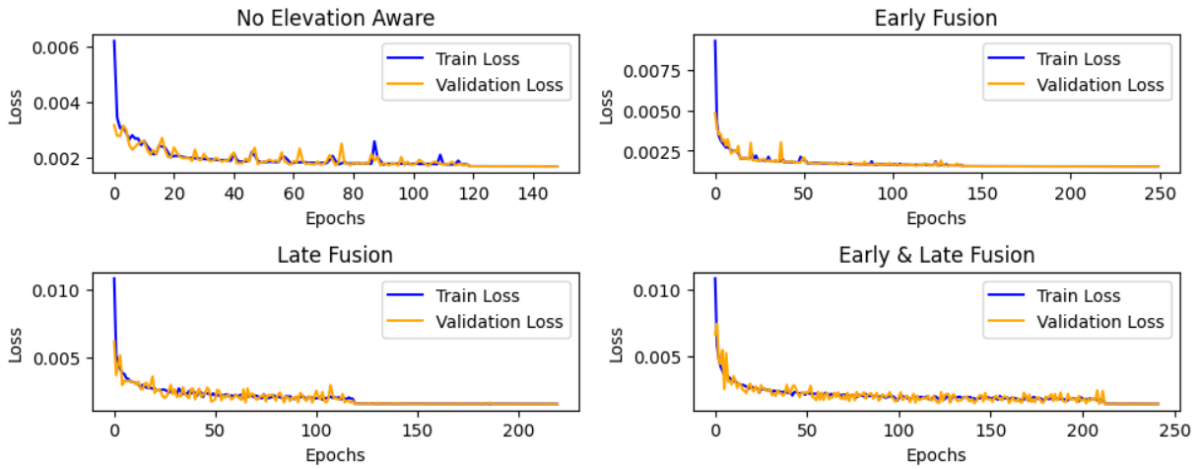

Figure 4: Training history validation loss for each Elevation Integration Strategy.

The results of the baseline and elevation integration approaches are summarized in Table 1. While the differences in the metrics are not very large, it is clear that the integration of elevation has a positive effect on performance. The small variations in the metrics are likely due to the fact that the study area is very large (all metrics are averaged over the 185,600 cells of the grid), and a significant portion of it consists of sea, where, as shown in Figure 6, the MAE is close to zero. For this reason, it was also deemed necessary to calculate the MAE over land only. In the land-only metrics, the improvement is more noticeable, with the mean MAE decreasing by 0.03°C when using the Early & Late Fusion method. Of particular importance is the 0.3°C reduction observed in the maximum MAE. Qualitatively, this indicates that even in the worst-case scenarios, this architecture deviates from the true value by only about 1°C on average—significantly lower than the roughly 1.3°C deviation observed in the No DEM variant.

Table 1
Comparison of Elevation Integration Strategies (the arrows show the trend).

	No DEM	Early Fusion	Late Fusion	Early&Late Fusion
MAE ↓	0.00167	0.00153	0.00156	**0.00141**
MSE ↓	0.000012	0.000010	0.000010	**0.000008**
PSNR ↑	49.3268	49.9457	49.9583	**50.8118**
SSIM ↑	0.9947	0.9955	0.9953	**0.9962**
Land Mean MAE ↓	0.00266	0.00241	0.00249	**0.00223**
Land Mean MAE in °C ↓	0.17062	0.15454	0.15987	**0.14298**
Land Max MAE in °C ↓	1.32484	1.44314	1.15182	**1.01760**
Training time per epoch ↓	**56s**	57s	76s	76s
Inference time per sample ↓	4.22ms	**3.04ms**	3.37ms	3.35ms
Trainable params ↓	**93,089**	93,377	107,105	107,393

After establishing Early & Late Fusion as the most effective architecture, additional static features derived from the DEM (slope and aspect), were tested to further improve performance. In this analysis, elevation was first substituted with aspect and slope individually, followed by an evaluation of their combined use with elevation. Figure 5 presents the training history and validation loss for each of the four models. All feature configurations display stable convergence with similar loss trajectories. However, the Elevation & Slope combination achieves slightly lower overall losses and triggers a reduction in the learning rate at a later stage, indicating that the model sustained effective learning over a longer training period.

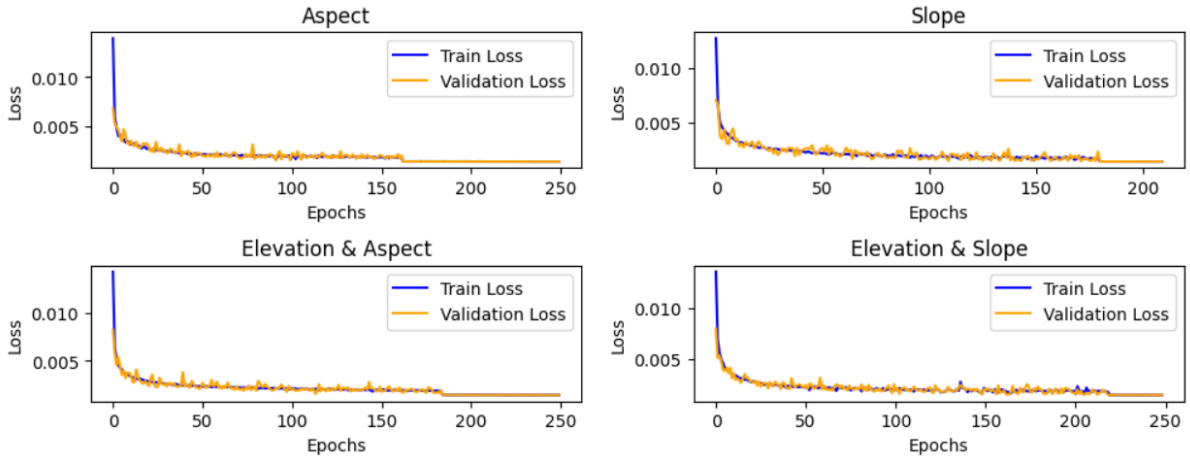

Figure 5: Training history validation loss for each Elevation-Derived Feature Integration.

The results of this ablation study are summarized in Table 2. According to the evaluation metrics, the differences between the approaches are minimal, with the Elevation & Slope combination showing a slight improvement in most cases. All configurations achieve results comparable to those reported in Table 1 for elevation alone, suggesting that the model extracts similar information from elevation and its derived features. Nevertheless, the Elevation & Slope combination yields marginal but consistent improvements in most metrics and is therefore recommended as the most effective approach.

Table 2
Comparison of Elevation-Derived Features Integration (the arrows show the trend).

	Aspect	**Slope**	**Elevation &Aspect**	**Elevation &Slope**
MAE ↓	0.00141	0.00142	0.00142	**0.00140**
MSE ↓	**0.000008**	0.000009	**0.000008**	**0.000008**
PSNR ↑	50.7447	50.7145	50.7910	**50.8345**
SSIM ↑	**0.9962**	0.9961	**0.9962**	**0.9962**
Land Mean MAE ↓	0.00223	0.00225	0.00224	**0.00222**
Land Mean MAE in °C ↓	0.14352	0.14457	0.14364	**0.14242**
Land Max MAE in °C ↓	1.04202	1.09875	**0.86288**	1.03497
Training time per epoch ↓	**76s**	77s	**76s**	81s
Inference time per sample ↓	3.44ms	3.41ms	**3.38ms**	3.41ms
Trainable params ↓	**107,393**	**107,393**	107,969	107,969

Figure 6 shows the average per-pixel MAE in °C on the test dataset for both the No Elevation Aware baseline model and the best-performing Elevation and Slope Fusion model. In both cases, errors are more prominent in temperate zones and regions with complex terrain, where temperature variations tend to be greater. The baseline model, in particular, exhibits higher errors in mountainous areas such as the Alps, Caucasus, Ethiopian Highlands, and Himalayas, highlighting its difficulty in capturing temperature variability in such regions. While the proposed method does not eliminate these errors entirely, the figure demonstrates a noticeable reduction in their value.

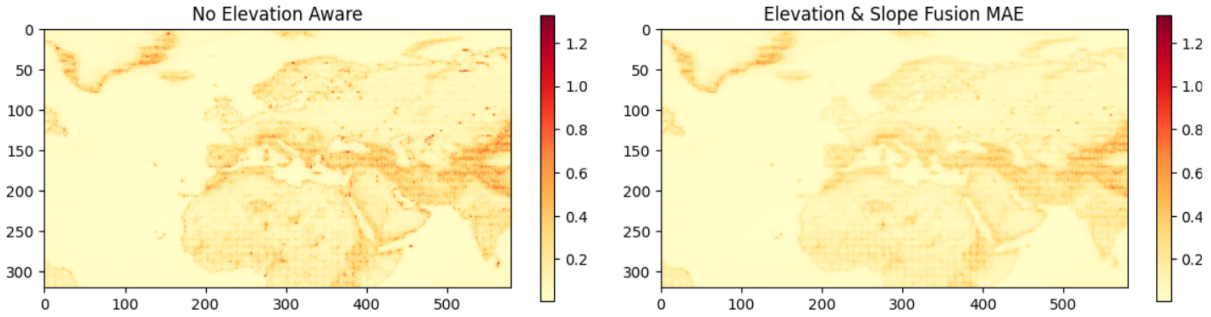

Figure 6: Comparison of MAE (°C) between the No Elevation Aware baseline model and the best-performing Elevation and Slope Fusion model.

Finally, Figure 7 illustrates an example output from best-performing Elevation and Slope Fusion model, alongside the LR input data and the expected HR target for comparison. To highlight the differences more clearly, a zoomed-in view of the wider European region has been provided.

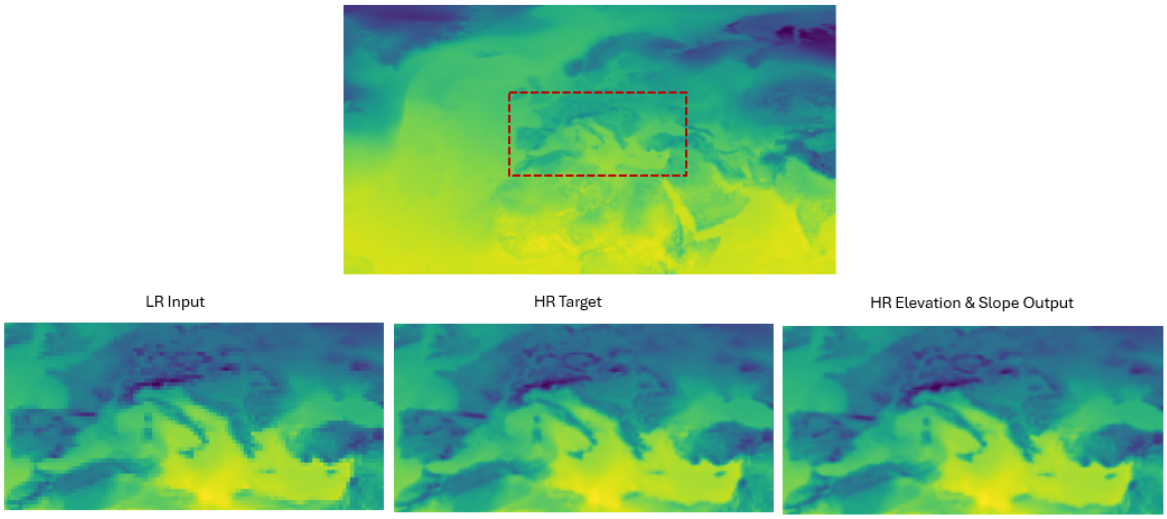

Figure 7: Comparison between the LR input, HR expected output and the best approach Elevation and Slope output zoomed in Europe.

5. Discussion & Conclusions

This study presented an effective method for downscaling climate data (both in terms of performance and inference time) capable of doubling the spatial resolution (from a 0.5°×0.5° grid to 0.25°×0.25°), while maintaining a low average error of approximately 0.14°C per cell. The main focus of the research was the integration of geospatial information (elevation and elevation-devired static features) into the DL model. This approach enhanced the accuracy of the downscaled outputs, particularly in complex terrains such as mountainous regions, where traditional methods tend to perform poorly. The study

shows that this relatively simple yet targeted intervention shift the problem from a generic SISR task toward a domain-informed climate application. By doing so, the model is better able to capture spatial variability in temperature that is influenced by underlying geographic and topographic features. This not only improves performance but also increases the physical interpretability of the results, which is considered an important consideration in climate modeling.

While the proposed strategy has demonstrated promising results in improving the spatial resolution of temperature data using DL and geospatial augmentation, several directions remain open for further development. The following key points are suggested for further investigation:

1. **Finer-scale downscaling:** The current elevation integration strategy may be even more beneficial for higher spatial resolution downscaling, as HR DEM data (which are available at scales of a few meters) can help the model capture finer local variability, particularly in areas with complex topography.
2. **Exploration and evaluation of different static feature integration across varying resolutions:** When downscaling is performed in multiple steps, different static features can be incorporated at each stage, depending on the factors that influence climate at the corresponding spatial scale. For instance, when downscaling from 250 km to 100 km, broad features such as land–sea masks may be more relevant. At intermediate resolutions (e.g., 100 km to 25 km), factors like latitude or season may play a larger role. Finally, for resolutions below 25 km, more detailed and information-rich features, such as elevation, land cover or land use, become increasingly important for capturing fine-scale temperature variations.
3. **Expansion to more climate variables:** The proposed elevation integration strategy could also be applied to other climate variables, such as precipitation and wind. Extending the approach in this way would help assess whether it can preserve physical consistency between variables, while exploring the feasibility of multi-variable downscaling.
4. **Model architecture optimizations**: As previously mentioned, this study focuses on exploring different strategies for integrating DEM data to enhance the performance of the deep learning model, rather than modifying the model architecture itself. However, architectural improvements (such as employing a deeper EDSR model) could increase the network's capacity to learn complex patterns. Additionally, testing more advanced architectures with attention mechanisms, such as RCAN, could be a potential direction to scale up the approach.
5. **Use of temporal models:** In this study, the temporal dimension was entirely ignored and training, validation and testing sets were shuffled to disrupt their temporal continuity. In future work, the integration of temporal information could be explored using models designed to capture sequential dependencies, such as ConvLSTMs [23]. Since temperature patterns exhibit temporal correlations, incorporating such models may improve downscaling accuracy by capturing seasonal or daily trends that static models do not represent.

These directions highlight the potential for both methodological and practical improvements in this climate data downscaling method.

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT in order to: Grammar and spelling check, reword, drafting content and improve writing style. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

Funding

This research was funded by the New Enabling Visions and Tools for End-useRs and stakeholders thanks to a common MOdeling approach towards a ClimatE neutral and resilient society (NEVERMORE)

project, Grant agreement ID: 101056858 and the Adaptation solutions to reduce climate change impact on health in the Mountain area (MOUNTADAPT) project, Grant agreement ID: 101155958.

References

[1] N. Politi, D. Vlachogiannis, A. Sfetsos, P. T. Nastos, High-resolution dynamical downscaling of era-interim temperature and precipitation using wrf model for greece, Climate Dynamics 57 (2021) 799–825. URL: http://dx.doi.org/10.1007/s00382-021-05741-9. doi:10.1007/s00382-021-05741-9.

[2] E. Tomasi, G. Franch, M. Cristoforetti, Can ai be enabled to perform dynamical downscaling? a latent diffusion model to mimic kilometer-scale cosmo5.0_clm9 simulations, Geoscientific Model Development 18 (2025) 2051–2078. URL: http://dx.doi.org/10.5194/gmd-18-2051-2025. doi:10.5194/gmd-18-2051-2025.

[3] S. Dubey, P. P. Sahu, P. S. Bazal, Literature review on single image super resolution, International Journal of Trend in Scientific Research and Development Volume-2 (2018) 2485–2490. URL: http://dx.doi.org/10.31142/ijtsrd18339. doi:10.31142/ijtsrd18339.

[4] Y. Sun, K. Deng, K. Ren, J. Liu, C. Deng, 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 208 (2024) 14–38. URL: http://dx.doi.org/10.1016/j.isprsjprs.2023.12.011. doi:10.1016/j.isprsjprs.2023.12.011.

[5] N. Rampal, S. Hobeichi, P. B. Gibson, J. Baño-Medina, G. Abramowitz, T. Beucler, J. González-Abad, W. Chapman, P. Harder, J. M. Gutiérrez, Enhancing regional climate downscaling through advances in machine learning, Artificial Intelligence for the Earth Systems 3 (2024). URL: http://dx.doi.org/10.1175/aies-d-23-0066.1. doi:10.1175/aies-d-23-0066.1.

[6] N. M. Pawar, R. Soltanmohammadi, S. K. Mahjour, S. A. Faroughi, Esm data downscaling: a comparison of super-resolution deep learning models, Earth Science Informatics 17 (2024) 3511–3528. URL: http://dx.doi.org/10.1007/s12145-024-01357-9. doi:10.1007/s12145-024-01357-9.

[7] F. Merizzi, A. Asperti, S. Colamonaco, Wind speed super-resolution and validation: from era5 to cerra via diffusion models, Neural Computing and Applications 36 (2024) 21899–21921. URL: http://dx.doi.org/10.1007/s00521-024-10139-9. doi:10.1007/s00521-024-10139-9.

[8] H. Lin, J. Tang, S. Wang, S. Wang, G. Dong, Deep learning downscaled high-resolution daily near surface meteorological datasets over east asia, Scientific Data 10 (2023). URL: http://dx.doi.org/10.1038/s41597-023-02805-9. doi:10.1038/s41597-023-02805-9.

[9] Y. Sha, D. J. Gagne II, G. West, 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 59 (2020) 2057–2073. URL: http://dx.doi.org/10.1175/JAMC-D-20-0057.1. doi:10.1175/jamc-d-20-0057.1.

[10] S. Bhakare, M. Matiu, A. Crespi, D. Zardi, Spatial downscaling of daily temperature minima using machine learning methods and application to frost forecasting in two alpine valleys, Atmosphere 16 (2025) 38. URL: http://dx.doi.org/10.3390/atmos16010038. doi:10.3390/atmos16010038.

[11] S. Bhakare, S. Dal Gesso, M. Venturini, D. Zardi, L. Trentini, M. Matiu, M. Petitta, Intercomparison of machine learning models for spatial downscaling of daily mean temperature in complex terrain, Atmosphere 15 (2024) 1085. URL: http://dx.doi.org/10.3390/atmos15091085. doi:10.3390/atmos15091085.

[12] B. Lim, S. Son, H. Kim, S. Nah, K. M. Lee, Enhanced deep residual networks for single image super-resolution, 2017. URL: http://dx.doi.org/10.1109/cvprw.2017.151. doi:10.1109/cvprw.2017.151.

[13] C. C. C. Service, Era5 hourly data on pressure levels from 1940 to present, 2018. URL: https://cds.climate.copernicus.eu/doi/10.24381/cds.bd0915c6. doi:10.24381/CDS.BD0915C6.

[14] U.S. Geological Survey, The 3d elevation program (3dep), https://www.usgs.gov/core-science-systems/ngp/3dep, 2023. https://www.usgs.gov/core-science-systems/ngp/3dep.

[15] Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, Image quality assessment: from error visibility to

structural similarity, IEEE Transactions on Image Processing 13 (2004) 600–612. doi:10.1109/TIP.2003.819861.

[16] Y. Gong, Y. Zhang, F. Wang, C.-H. Lee, Deep learning for weather forecasting: A cnn-lstm hybrid model for predicting historical temperature data (2024). URL: https://arxiv.org/abs/2410.14963. doi:10.48550/ARXIV.2410.14963.

[17] C. Willmott, K. Matsuura, Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance, Climate Research 30 (2005) 79–82. URL: http://dx.doi.org/10.3354/cr030079. doi:10.3354/cr030079.

[18] C. J. Willmott, K. Matsuura, On the use of dimensioned measures of error to evaluate the performance of spatial interpolators, International journal of geographical information science 20 (2006) 89–102.

[19] M. A. et al, TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL: https://www.tensorflow.org/, software available from tensorflow.org.

[20] P. et al, Scikit-learn: Machine learning in python, Journal of Machine Learning Research 12 (2011) 2825–2830.

[21] J. D. Hunter, Matplotlib: A 2d graphics environment, Computing in Science & Engineering 9 (2007) 90–95. doi:10.1109/MCSE.2007.55.

[22] H. Li, J. Lu, H. Tong, Y. Liu, Impact of high temperature heat waves on ocean carbon sinks: Based on literature analysis perspective, Journal of Sea Research 198 (2024) 102487. URL: https://www.sciencedirect.com/science/article/pii/S1385110124000200. doi:https://doi.org/10.1016/j.seares.2024.102487.

[23] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-k. Wong, W.-c. Woo, Convolutional lstm network: A machine learning approach for precipitation nowcasting (2015). URL: https://arxiv.org/abs/1506.04214. doi:10.48550/ARXIV.1506.04214.