Article

Not peer-reviewed version

High-Resolution Early Warning Systems Using DL: Part II - Combining FourCastNet and E-TEPS for High-Resolution Climate Forecasting

Afshin Shafei * and Francesco Cioffi

Posted Date: 19 August 2024

doi: 10.20944/preprints202408.1322.v1

Keywords: Early Warning System; Temperature; Precipitation; Climate Downscaling; Machine-Learning; Super-Resolution Generative Adversarial Network; Flood

Article

High-Resolution Early Warning Systems Using DL: Part II - Combining FourCastNet and E-TEPS for High-Resolution Climate Forecasting

Afshin Shafei * and Francesco Cioffi

Dipartimento di Ingegneria Civile, Edile ed Ambientale, DICEA, "La Sapienza" University of Rome
* Correspondence: afshin.shafei@uniroma1.it

Abstract: Extreme weather events, such as heat waves and heavy precipitation, are becoming increasingly frequent due to climate change, necessitating the development of effective early warning systems (EWS) to mitigate their impacts. This study introduces an advanced EWS designed specifically for Italy, which integrates the FourCastNet global forecasting model with Elevation-integrated TEmperature and Precipitation SRGAN downscaling (E-TEPS) to enhance the spatial resolution and accuracy of climate predictions. Building on previous work that demonstrated E-TEPS's effectiveness in downscaling temperature and total precipitation variables, this research applies the integrated system to two severe weather events in Central Italy: the Emilia-Romagna floods of 2023 and the Marche floods of 2022. The system's performance was evaluated using key metrics, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Pearson Correlation, by comparing downscaled outputs from FourCastNet and ERA5 datasets against high-resolution ground truth data from the Euro-Mediterranean Center on Climate Change (CMCC) dataset. The results indicate that the integrated system offers improved predictive accuracy, particularly in capturing critical climate variables, with the entire EWS delivering high-resolution forecasts and final outputs in under one minute. Although potential limitations were identified due to biases in the underlying datasets, which could affect forecast reliability in regions with complex topography or extreme weather conditions, this research highlights the potential of combining machine learning models with downscaling techniques to enhance EWS precision, providing valuable insights for future climate forecasting and disaster preparedness strategies.

Keywords: early warning system; temperature; precipitation; climate downscaling; machine-learning; super-resolution generative adversarial network; flood

---

1. Introduction

Effective early warning systems (EWS) are crucial for mitigating the impacts of extreme weather conditions, such as heat waves and heavy precipitation events, which are becoming increasingly frequent and severe due to climate change [1]. The goal of EWS is to provides accurate, timely forecasts that enable proactive measures to protect agriculture, infrastructure, and public safety [2]. The integration of advanced machine learning models in EWS enhances their predictive capabilities, offering more precise and actionable climate forecasts tailored to specific regions [3].

This paper builds upon part I of this paper, where we developed a novel downscaling method using a Super-Resolution Generative Adversarial Network (SRGAN) to enhance the spatial resolution of climate data. In the first part of this research series, we focused on the development and validation of the SRGAN-based downscaling model, demonstrating significant improvements in capturing complex spatial patterns and fine details in climate predictions [4–7]. By integrating elevation maps as auxiliary inputs, E-TEPS achieved superior accuracy in downscaling temperature and precipitation variables [8–10]. This foundational work set the stage for creating an advanced early warning system (EWS) by combining the downscaling method with FourCastNet, a high-resolution, fast-timescale forecasting model [11].

In this second part of our research, we present the design and application of our comprehensive EWS focused on Italy. We utilize FourCastNet for global climate forecasting and employ our developed

 doi:10.20944/preprints202408.1322.v1

downscaling model, E-TEPS, to refine these forecasts to high-resolution, local-scale predictions for the Italian region.

We tested our EWS through two distinct and extreme weather scenarios in Central Italy, both of which caused significant damage to their respective regions. The first event involved the extreme precipitation event in the Emilia-Romagna region during May 2023. The second event focused on the extreme precipitation event in the Marche region during September 2022. Both scenarios aimed to evaluate the accuracy and precision of the forecasts and the downscaling model for both total precipitation and temperature during the events. These regions and events have different characteristics, which will be compared in the paper to highlight the differences in results and demonstrate the robustness of the EWS.

To enhance the performance of our downscaling model for these specific applications, a fine-tuning stage was introduced. Initially, the downscaling model was trained on the CMCC dataset [12], which provided high-resolution climate data for the Italian region. During the fine-tuning stage, ERA5 [13,14] datasets were introduced to the pretrained model. This step was essential for adapting the model to the specific climatic conditions of the tested regions and improving the accuracy of forecasts. The ERA5 data allowed the model to align more closely with the inputs from FourCastNet, enhancing its capability to downscale these outputs effectively. An elevation map, consistent with the one used during training, was also utilized to further improve accuracy. This fine-tuning process will be detailed in the materials and methods section.

To validate the results, we compared our EWS outputs against observed data using key metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Pearson Correlation. Additionally, we conducted agreement analysis for total precipitation and visual comparisons between the generated high-resolution predictions and ground truth data. These evaluations demonstrated the EWS's effectiveness in providing high-resolution, accurate climate predictions, proving its potential in enhancing disaster preparedness and response strategies.

The remainder of this paper is structured as follows: We begin with a brief review of related work in early warning systems and climate prediction. We then detail the methodology of our integrated EWS, including the architecture of FourCastNet and E-TEPS. Following this, we present the experimental setup and results for the two events, discussing the implications of our findings. Finally, we conclude with a summary of our contributions and future research directions.

2. Events

2.1. Marche Event (September 2022)

The Marche region in Central Italy experienced an extreme weather event on September 15, 2022, resulting in severe flash floods and landslides that had devastating consequences. This event is notable as one of the three millennial events in the last decade that have significantly impacted the region, alongside those in 2013 and 2015 [15]. The September 2022 event triggered 1,687 landslides, led to riverine floods, and tragically caused the deaths of 12 people [16].

On the afternoon of September 15, a torrential downpour began, leading to an unprecedented accumulation of rainfall. In just a few hours, the town of Cantiano in the Pesaro Urbino province recorded nearly 400 mm of rain [**?**]. This deluge overwhelmed the Misa, Cesano, and Esino rivers, causing them to burst their banks and flood the surrounding areas. The rapid onset of the floods caught many off guard, leaving little time for the implementation of mitigation measures, which exacerbated the damage.

In the aftermath, extensive field surveys were conducted by Italian universities and private companies to assess the damage and understand the event's causes. These surveys revealed severe damage to residential buildings, economic activities, infrastructure, and cultural heritage sites across the affected municipalities of Senigallia, Ostra, and Tre Castelli. Water levels during the flood exceeded

 doi:10.20944/preprints202408.1322.v1

1 meter in many areas, reaching up to 3 meters in some locations, leading to significant and lasting damage [17].

2.2. Emilia-Romagna Event (May 2023)

In May 2023, the Emilia-Romagna region in Northern Italy faced two consecutive severe rainfall events that led to widespread flooding and landslides. These events came on the heels of a prolonged drought that had affected the region over the previous two years, exacerbating the impact of the floods. The first event occurred between May 2-3, 2023, primarily affecting the Ravenna province. Prolonged rainfall during this period raised the level of the Po River by 1.5 meters within 24 hours, leading to significant flooding as smaller rivers overflowed their banks.

The second and more severe event, which we analyzed in this paper, driven by Storm Minerva, struck between May 15-17, 2023. During this period, an average of 200 mm of rain fell within 36 hours across parts of Emilia-Romagna, with some areas receiving up to 500 mm. The storm's impact was devastating, with 15 people killed, 36,000 evacuated, and extensive damage to infrastructure, agriculture, and homes. Over 23 rivers burst their banks, and more than 400 landslides were reported, highlighting the scale of the disaster [**?**].

The heavy rainfall events of May 2023 were unprecedented in their intensity and impact, ranking as the worst flooding event in the region in decades. The saturation of the soil from previous rainfall further amplified the disaster, making the region more vulnerable to the subsequent deluge [**?**]. Initial assessments estimate the economic loss to be in the billions of euros, with significant portions of the damage not covered by insurance [18].

In the wake of these events, discussions have focused on the role of climate change and land management in exacerbating the impacts. While preliminary studies suggest that the specific rainfall events were not directly influenced by human-induced climate change, the overall trend of more frequent and intense extreme weather events due to global warming remains a concern. Additionally, urbanization and changes in land use over the past decades have increased the region's vulnerability to such disasters, underscoring the need for improved flood management and resilience strategies [19].

2.3. Comparison of Marche and Emilia-Romagna Events

The Marche and Emilia-Romagna flood events, while both catastrophic, exhibited distinct characteristics due to the unique geographical and climatological features of their respective regions. These differences were a key factor in selecting these events for testing the early warning system, as they present contrasting challenges and behaviors in response to extreme weather conditions.

Marche event in Marche was highly localized, characterized by intense rainfall over a short period. The region's topography, dominated by the steep slopes of the Apennine Mountains, played a crucial role in the event's severity. The rapid runoff from these slopes funneled water into narrow valleys and river channels, leading to flash floods that were concentrated in specific areas [16]. The flash floods were sudden and intense, exacerbated by the region's shallow, rocky soils, which have limited water absorption capacity, and the confined nature of the river valleys, which accelerated the flooding process [20].

In contrast, the May 2023 flooding in Emilia-Romagna was more widespread and prolonged, impacting a much larger area over several days. This difference can be attributed to the region's extensive plains, particularly the Po Valley, which allowed floodwaters to spread over a vast area [21]. The flat topography and the presence of large river networks, including the Po River and its tributaries, contributed to the extensive nature of the flooding, as these rivers overflowed their banks and inundated the surrounding low-lying areas [22]. Additionally, the soil in Emilia-Romagna, which had been saturated by previous rainfall, had reduced capacity to absorb the intense and prolonged precipitation, further contributing to the widespread flooding [**?**].

These contrasting characteristics—localized, flash flooding in Marche and widespread, prolonged flooding in Emilia-Romagna—highlight the need for diverse approaches in flood management and

 doi:10.20944/preprints202408.1322.v1

early warning systems. By selecting these two distinct events, we aim to evaluate the effectiveness of our early warning system across different types of extreme precipitation scenarios, ensuring it can provide accurate and timely alerts regardless of the geographical and climatological context.

2.4. Dataset Acquisition for Event Analysis

To analyze the Marche and Emilia-Romagna flood events using FourCastNet and the downscaling model, specific ERA5 datasets were acquired. These datasets were carefully selected to capture the critical periods surrounding each event, providing the necessary temporal and spatial resolution to perform accurate forecasting and downscaling.

For the Marche event, which occurred in September 2022, ERA5 data was downloaded for the period of September 15-18, 2022. This dataset covers four days, with a 6-hour temporal resolution, allowing for a detailed analysis of the atmospheric conditions leading up to and during the flood. The spatial resolution of the data is 0.25 degrees.

Similarly, for the Emilia-Romagna event in May 2023, ERA5 data was acquired for the period of May 16-19, 2023, again covering four days with a 6-hour temporal resolution with he same spatial resolution of 0.25 degrees.

Further processing of these datasets, including the specifics of their preparation for use in FourCastNet and the downscaling model, as well as additional details about the data itself, will be discussed in the Materials and Methods section 3. This will include steps such as data normalization, transformation, and integration with other datasets to facilitate the subsequent analysis and modeling.

3. Material and Methods

In this section, we describe the overall framework and components of our early-warning system, which combines global predictions from the FourCastNet model with our downscaling model to achieve high-resolution forecasts for specific regions. The process begins with global climate variable predictions generated by FourCastNet, which are subsequently refined through a detailed downscaling procedure to provide localized, high-resolution outputs. This approach is designed to address the challenges of forecasting in diverse geoclimatic conditions, ensuring that both broad-scale and fine-scale features are accurately captured. The following subsections outline the key steps involved, including the fine-tuning of the downscaling model, the integration process with FourCastNet, and the specific computational resources utilized. Figure 1 provides a visual overview of the entire model system, illustrating how each component interacts within the forecasting workflow.

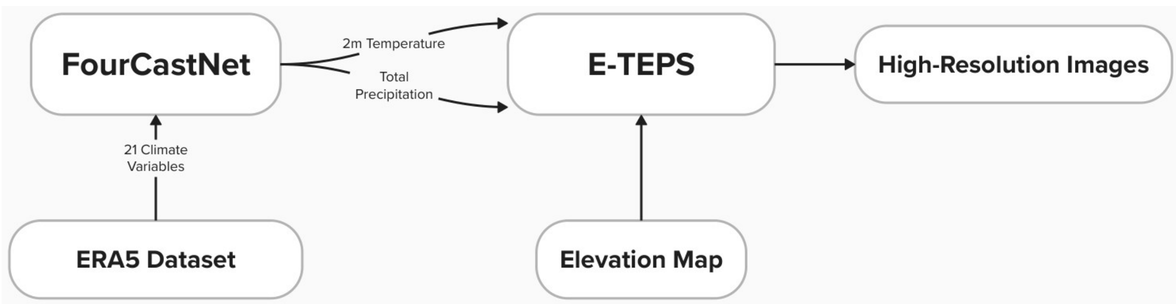


Figure 1. The diagram illustrates the data flow within the system, starting with the input of global climate variables into the FourCastNet model, followed by the integration of these outputs with an elevation map in the downscaling model. The process culminates in the generation of high-resolution climate forecasts for 2m temperature and 6-hour accumulated total precipitation.

3.1. Fine-Tuning of the Downscaling Model

Fine-tuning the downscaling model is essential for several reasons. First, the initial training of the downscaling model, although comprehensive, was tailored to a different dataset and set of climatic conditions. Fine-tuning allows the model to adapt to new, specific climatic conditions relevant to the

 doi:10.20944/preprints202408.1322.v1

application at hand. This adaptation is crucial because temperature and total precipitation, the two variables that our downscaling model focuses on, can vary significantly across different regions and datasets. By fine-tuning, we can improve the model's accuracy and reliability for the particular climatic conditions of the regions being studied.

Moreover, fine-tuning helps in aligning the downscaling model more closely with the inputs it will receive from FourCastNet. The initial training did not incorporate these specific inputs, so without fine-tuning, there could be a mismatch in the data characteristics, leading to suboptimal performance. Fine-tuning ensures that the model can effectively process and downscale the outputs from FourCastNet, thereby improving the overall quality and precision of the forecasts.

In summary, fine-tuning is a critical step to adapt the downscaling model to specific climatic conditions, ensure compatibility with FourCastNet inputs, and ultimately contribute to more accurate and reliable climate forecasts.

3.1.1. Fine-Tuning Dataset

The dataset used for fine-tuning the downscaling model includes the ERA5 reanalysis dataset and the CMCC high-resolution dataset. The ERA5 dataset, with a spatial resolution of approximately 30 km and a temporal resolution of 6 hours, was employed as the low-resolution input. In contrast, the CMCC dataset, a High Resolution Dynamical Downscaling of ERA5 Reanalysis over Italy by [12], provided high-resolution data at a 2.2 km spatial resolution with the same temporal resolution. Both datasets were used for fine-tuning the model from 2013 to 2018, with the period from 2019 to 2020 reserved for validation purposes. In this fine-tuning phase, we focused on the same variables as in the initial training phase: 2-meter temperature above the surface and accumulated total precipitation.

During the initial training phase, we reduced the resolution of the CMCC data by a factor of eight to create low-resolution input images, while the original high-resolution images were used as ground truth. This approach enabled the model to learn to map low-resolution images to their high-resolution counterparts effectively. The reduced resolution closely matched the resolution of the ERA5 data, facilitating the subsequent fine-tuning process.

In the fine-tuning phase, we introduced the ERA5 data directly to the model as low-resolution inputs, paired with the corresponding high-resolution CMCC data for the same time steps. This approach allowed the model to adapt to the new data and align more closely with the resolution of the output data expected from FourCastNet. By exposing the model to these new datasets, we aimed to enhance its capability to downscale the outputs from FourCastNet, thereby improving the accuracy and precision of the climate forecasts.

There is an important point to mention regarding the compatibility of these two datasets, which influenced our decision not to train the model directly using ERA5 as low-resolution images and CMCC as high-resolution images. Although the CMCC dataset is derived from dynamically downscaling ERA5 data, discrepancies between the datasets, especially in total precipitation, posed challenges. For example, as shown in Figure 1 (to be cited and included in the article), there are instances where precipitation events are present in the ERA5 dataset but not in the CMCC dataset. Such inconsistencies made initial training with these datasets together problematic, as the model struggled to reconcile the differences in data representation at certain time steps.

Consequently, we opted to initially train the model exclusively with CMCC data, then fine-tune it by incorporating ERA5 data. While the non-compatibility of the two datasets did introduce some negative effects during fine-tuning, these were significantly less problematic compared to the issues encountered during initial training with both datasets combined. This phased approach allowed us to mitigate the impact of data inconsistencies while still leveraging the broader coverage of the ERA5 dataset to enhance the model's performance.

 doi:10.20944/preprints202408.1322.v1

3.1.2. Fine-Tuning Process

The fine-tuning process for our downscaling model involved specific adjustments to the model's hyper-parameters, layer configurations, and training procedures. The overall architecture of the generator and discriminator models remained unchanged from the initial training phase. However, we made targeted modifications to enhance the model's performance with the new datasets.

To balance the training stability and convergence speed of the two models, we adjusted the learning rates. We employed transfer learning techniques by freezing certain layers in the generator. Specifically, the initial convolutional layer and the residual blocks were frozen. This approach allowed the model to retain the knowledge learned during the initial training phase while focusing on fine-tuning the remaining layers to adapt to the new data. This strategy helps prevent overfitting and ensures that the model does not forget the previously learned representations.

The fine-tuning process began by loading the pre-trained weights for both the discriminator and generator models. These weights were obtained from the initial training phase, ensuring a solid foundation for further refinement. The training loop involved alternating updates to the discriminator and generator. The discriminator was trained to distinguish between real high-resolution images and generated images, while the generator was trained to produce high-resolution images from low-resolution inputs that could fool the discriminator.

Gradient accumulation was used to manage memory usage and improve training stability. Additionally, mixed precision training was employed to accelerate the training process and reduce memory consumption. This was achieved using PyTorch's automatic mixed precision (AMP) feature, which allows for efficient and stable training by dynamically scaling the loss values.

The loss functions for both the discriminator and generator remained unchanged from the training phase. The discriminator loss was based on the ability to differentiate real and fake images, while the generator loss combined content loss and adversarial loss to ensure both the visual quality and realism of the generated images.

In summary, the fine-tuning process involved careful adjustments to hyperparameters, strategic freezing of certain layers, and a robust training loop to adapt the downscaling model to new climatic data. This approach ensured the model's enhanced performance and compatibility with the inputs from FourCastNet, leading to more accurate and reliable climate forecasts.

3.2. Integration of FourCastNet with the Downscaling Model

The integration of FourCastNet with the downscaling model forms a seamless chain for high-resolution climate forecasting. FourCastNet serves as our global forecasting system, capable of predicting 21 climate variables at a coarse resolution. Once FourCastNet performs inference, we extract the specific variables of interest—temperature and total precipitation—from its outputs.

These extracted variables are then clipped to focus on the geographic area that the downscaling model was trained on. This step ensures that the subsequent processing is relevant and tailored to the region of interest. The clipped outputs, representing low-resolution climate forecasts, are combined with the high-resolution elevation map used during the model's training.

The integrated system feeds these low-resolution images, along with the elevation data, into the downscaling model. The downscaling model, fine-tuned to adapt to these specific inputs, processes this data to produce high-resolution forecasts. This final output achieves a spatial resolution near 3 km for both temperature and total precipitation.

3.2.1. Inference and Data Preprocessing for FourCastNet

In the preprocessing stage for FourCastNet, we begin by downloading data for 20 climate variables from the database. These variables are essential inputs for the global forecasting system. The preprocessing involves organizing and structuring this data for efficient processing. For each variable, the data is read from its source file, typically in NetCDF or HDF5 format, and then written into a consolidated HDF5 file. This file serves as the input for FourCastNet's inference process. Each variable

 doi:10.20944/preprints202408.1322.v1

is processed in batches to manage memory usage and improve computational efficiency. The data is then normalized and structured consistently to ensure compatibility with FourCastNet's architecture (see Figure 1 for 2-meter temperature).

For the total precipitation variable, a different preprocessing approach is required due to its unique handling within the FourCastNet architecture. After downloading the data, we accumulate the 6-hour precipitation values. This accumulation is necessary to align with the temporal resolution required by the model. The total precipitation data is reshaped into 6-hour intervals and summed accordingly. This reshaped and accumulated data is then saved into a new NetCDF file, ready for inference.

Once the preprocessing steps are completed, we proceed with the inference phase. The pretrained weights for FourCastNet are loaded, and the model performs inference on the preprocessed data. The outputs for the 20 variables and total precipitation are generated separately. These results are then saved into new HDF5 files for subsequent processing (see Figure 2 for total precipitation).

3.2.2. Inference and Data Preprocessing for Downscaling

The integration process for preparing FourCastNet's outputs for the downscaling model involves a series of post-processing steps, which serve as the preprocessing for the downscaling model. This ensures that the global forecasts generated by FourCastNet are compatible and optimized for the high-resolution downscaling process.

The first step in this process is the denormalization of the temperature variable from FourCastNet's outputs. This step converts the temperature values back to their original scale, making them suitable for subsequent processing. For the precipitation variable, we convert the units from meters, which is the standard output for both ERA5 and FourCastNet, to millimeters. This conversion is crucial because the downscaling model was trained using the CMCC dataset, where precipitation is recorded in millimeters. Ensuring consistent units across all datasets prevents discrepancies and enhances the accuracy of the downscaled forecasts.

After unit conversion and denormalization, we extract the temperature and precipitation data along with the other 20 climate variables from the FourCastNet outputs. These are then combined into a new dataset that represents the global forecast for temperature and total precipitation. This combined dataset forms the input for the next stage of processing.

Next, a transformation is applied to the data to ensure geographical consistency. The final step in this preprocessing sequence involves cropping the global forecasts to focus on Italy, the region of interest for the downscaling model.

By carefully managing these preprocessing steps, we ensure that the global forecasts from FourCastNet are optimally prepared for the downscaling model. This preparation is essential for achieving high-resolution, accurate climate forecasts that are geographically relevant and scientifically robust.

After these preprocessing steps, the prepared data, along with the elevation map, are introduced to the inference stage of the downscaling model. This process generates high-resolution downscaled maps for both temperature and total precipitation across all 16 time steps, representing a 96-hour forecast for the events. These outputs provide detailed and accurate predictions, essential for understanding and responding to localized climate phenomena.

3.3. Hardware and Computational Resources

The computational tasks for this study were carried out using Google Colab and associated cloud services. Specifically, the NVIDIA GeForce A100 GPU available through Google Colab Pro was employed, offering 40GB of GPU RAM and 80GB of system RAM. These hardware specifications were crucial in meeting the significant computational requirements involved in fine-tuning the downscaling model. The total data size for the fine-tuning process was approximately 15GB, making the advanced hardware resources provided by this platform indispensable for efficient processing and model training.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 August 2024 doi:10.20944/preprints202408.1322.v1

4. Results

In this section, we present and discuss the performance of our early warning system, using key metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Agreement Analysis for precipitation, and Pearson Correlation for temperature. Additionally, we provide a visual comparison of downscaled images for selected time steps and later we include all the time steps in the supplementary material part.

The comparison involves three datasets: the high-resolution CMCC ground truth data, the high-resolution outputs generated by downscaling FourCastNet's forecasts, and the high-resolution outputs generated by downscaling raw ERA5 data. The primary goal is to evaluate how well the downscaled forecasts from FourCastNet and ERA5 align with the CMCC ground truth data, which serves as the benchmark for accuracy. The evaluation focuses on two critical climate variables: 6-hour accumulated total precipitation, measured in millimeters (mm), and 2-meter temperature above the surface, measured in Kelvin (K), both with a 6-hour temporal resolution.

An important aspect of this comparison is the inclusion of downscaled raw ERA5 data. The reason for downscaling ERA5 data and including it in the comparison is to assess how well FourCastNet performs in forecasting these specific events. By comparing the downscaled outputs from FourCastNet with those obtained by directly downscaling ERA5 data, we can gain insights into the forecasting capabilities of FourCastNet. This comparison helps us understand whether the additional processing and prediction steps in FourCastNet improve the accuracy of the forecasts or whether the raw ERA5 data provides comparable results when subjected to the same downscaling process.

The comparison approach involves a detailed quantitative analysis using MAE and RMSE, agreement analysis for precipitation, and Pearson correlation for temperature. These metrics provide a comprehensive assessment of how well the downscaled forecasts from our early warning system and the ERA5 data match the high-resolution ground truth data. MAE measures the average magnitude of the errors between the predicted and observed values without considering their direction, with a lower MAE indicating a better fit (Equation 1). RMSE highlights the differences between predicted and observed values, with greater sensitivity to larger errors (Equation 2).

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{1}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{2}$$

The method employed in total precipitation agreement analysis involves calculating the agreement of precipitation occurrence between downscaled data and observed high-resolution ground truth data. This process is crucial for assessing the accuracy and reliability of various downscaling techniques. The agreement matrix is generated by evaluating whether both the ground truth data and the downscaled predictions indicate the presence or absence of precipitation at each time step and normalizing these instances by the total number of time steps. This matrix provides a quantitative measure of how well each method reproduces the spatial and temporal patterns of precipitation observed in high-resolution datasets. The agreement analysis methodology used in this study is inspired by approaches discussed in [23] which provides comprehensive insights into various models and methods for assessing agreement between datasets.

The Pearson Correlation Coefficient is used to evaluate the linear relationship between predicted and observed values of the temperature variable. A coefficient close to 1 indicates a strong positive linear relationship, while values close to -1 or 0 indicate weaker or no linear relationships (Equation 3).

$$r = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2 \sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2}} \tag{3}$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 August 2024 doi:10.20944/preprints202408.1322.v1

By using these metrics, we can quantitatively assess the performance of our early warning system, comparing the downscaled outputs from FourCastNet and ERA5 against the high-resolution CMCC ground truth data. This comprehensive evaluation highlights the effectiveness of our system in improving spatial resolution and predictive accuracy, particularly in the context of high-resolution climate forecasting. Additionally, the inclusion of downscaled ERA5 data in the comparison provides valuable insights into the performance of FourCastNet in forecasting these specific events.

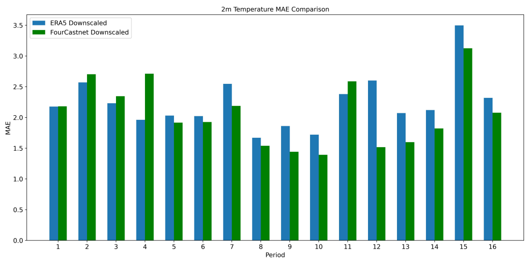
4.1. Emilia-Romagna Event

4.1.1. Quantitative Analysis

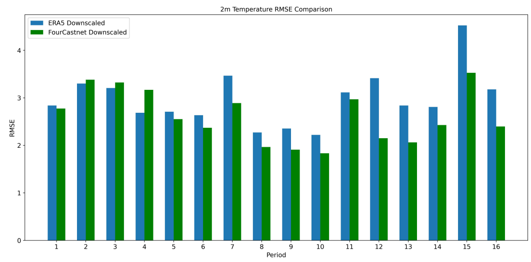
The figures present a comparative analysis between the FourCastNet downscaled outputs and the ERA5 downscaled outputs for 2m temperature and total precipitation. Both MAE and RMSE metrics are used to evaluate performance across 16 time periods.

Figure 2a shows that FourCastNet downscaled outputs generally exhibit lower MAE values for 2m temperature compared to ERA5 downscaled outputs, indicating better predictive accuracy. Similarly, Figure 2b presents the RMSE for 2m temperature, where FourCastNet again outperforms ERA5 in most time periods.

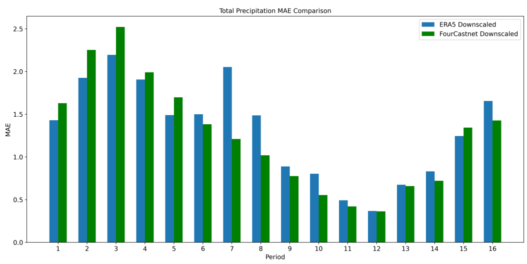
In the total precipitation analysis, Figure 2c shows that FourCastNet downscaled data generally has a lower MAE compared to ERA5, especially in the later periods. Figure 2d displays the RMSE for total precipitation, revealing a similar trend, with FourCastNet showing better performance in handling significant precipitation events.



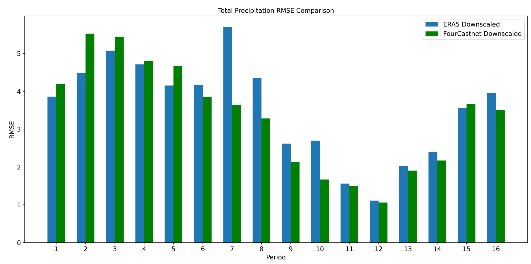
(**a**) 2m Temperature MAE Comparison



(**b**) 2m Temperature RMSE Comparison



(**c**) Total Precipitation MAE Comparison



(**d**) Total Precipitation RMSE Comparison

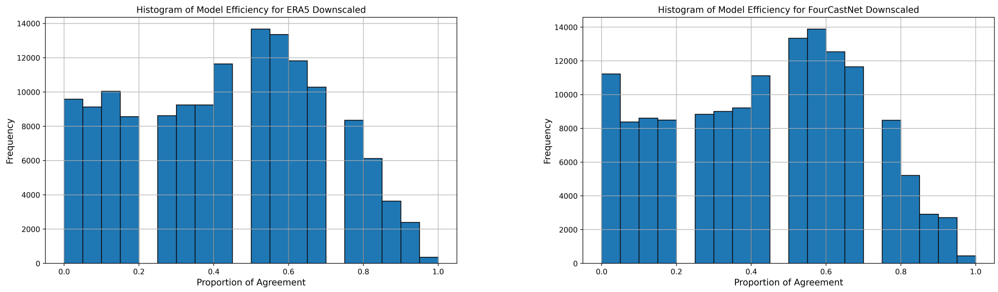
Figure 2. Comparison of FourCastNet Downscaled and ERA5 Downscaled data: (a) Mean Absolute Error (MAE) for 2m temperature, (b) Root Mean Square Error (RMSE) for 2m temperature, (c) MAE for total precipitation, and (d) RMSE for total precipitation.

4.1.2. Total Precipitation Agreement Analysis

The histograms in Figure 3 show the distribution of the proportion of agreement for 6-hour accumulated total precipitation between the downscaled outputs (ERA5 and FourCastNet) and the ground truth data.

Figure 3a shows that the ERA5 downscaled data has a broad distribution with moderate levels of agreement with the ground truth. On the other hand, Figure 3b indicates that FourCastNet downscaled data generally has a higher concentration of agreement values, suggesting better performance in predicting precipitation events.

 doi:10.20944/preprints202408.1322.v1



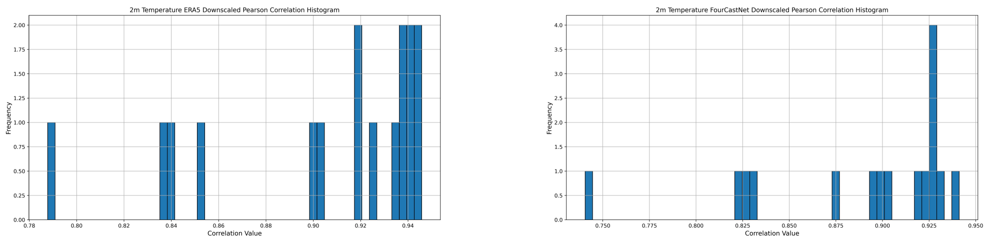
(**a**) ERA5 Downscaled (**b**) FourCastNet Downscaled

Figure 3. Histogram of Model Efficiency for Total Precipitation Agreement: (a) ERA5 Downscaled, (b) FourCastNet Downscaled.

4.1.3. Pearson Correlation For Temperature

The histograms in Figure 4 display the Pearson correlation coefficients for 2m temperature between the downscaled outputs (ERA5 and FourCastNet) and the ground truth data.

In Figure 4a, the ERA5 downscaled data shows high correlation values, particularly between 0.92 and 0.94, indicating a strong linear relationship with the ground truth. Figure 4b shows that FourCastNet downscaled data also exhibits high correlation values, with a notable peak around 0.925, though it shows a slightly wider distribution compared to ERA5.



(**a**) ERA5 Downscaled (**b**) FourCastNet Downscaled

Figure 4. Pearson Correlation Histogram for 2m Temperature: (a) ERA5 Downscaled, (b) FourCastNet Downscaled.

4.1.4. Visual Comparison

In this section, we present a visual comparison of the downscaled outputs for both 2m temperature and total precipitation at a selected timestep. This comparison includes three plots for each variable: ground truth, FourCastNet downscaled, and ERA5 downscaled. These visualizations provide a clear assessment of how well each model replicates the spatial distribution and intensity of the observed climate variables. As seen in the figures 6 and 5, the FourCastNet downscaled outputs generally capture the spatial patterns more accurately and consistently compared to the ERA5 downscaled outputs when compared to the ground truth.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 August 2024 doi:10.20944/preprints202408.1322.v1

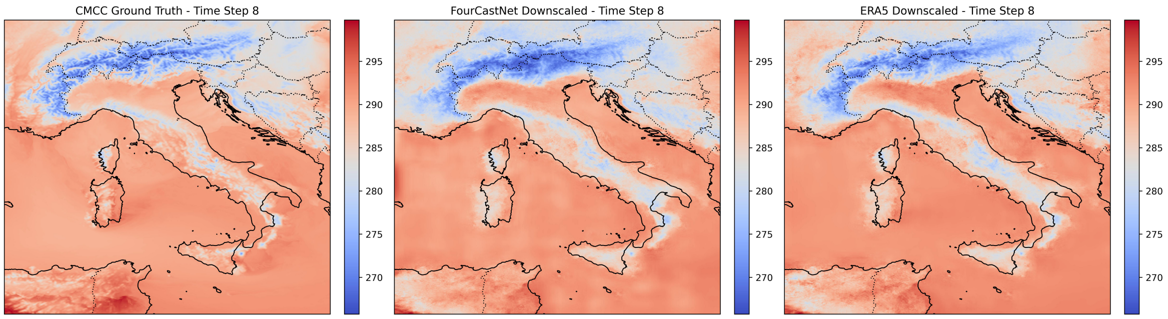


Figure 5. Visual comparison of 2m temperature at a selected timestep. Left: Ground Truth, Middle: FourCastNet Downscaled, Right: ERA5 Downscaled. Units: Kelvin (K). This figure highlights the differences in spatial distribution and intensity between the models.

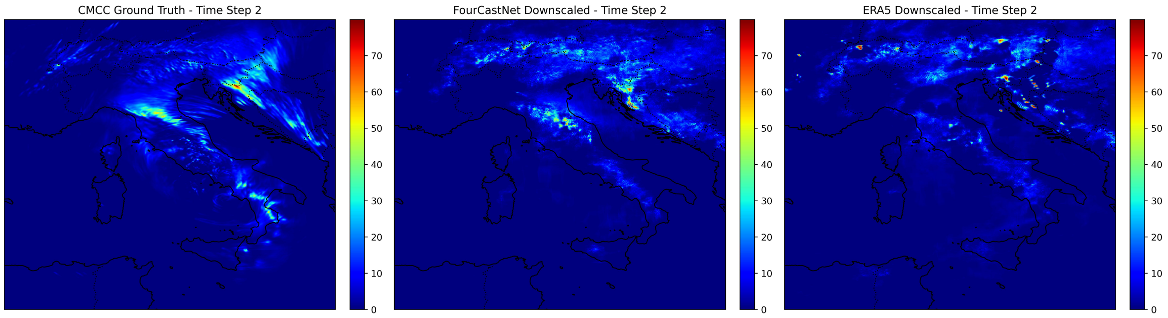


Figure 6. Visual comparison of total precipitation at a selected timestep. Left: Ground Truth, Middle: FourCastNet Downscaled, Right: ERA5 Downscaled. Units: Millimeters (mm). The figure provides insight into how well each model captures precipitation patterns.

4.2. Marche Event

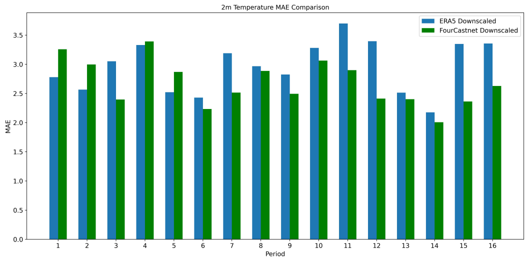
4.2.1. Quantitative Analysis

The quantitative analysis for the Marche event, shown in Figure 7, highlights the performance of the FourCastNet downscaled outputs compared to the ERA5 downscaled outputs for 2m temperature and total precipitation across 16 time periods.

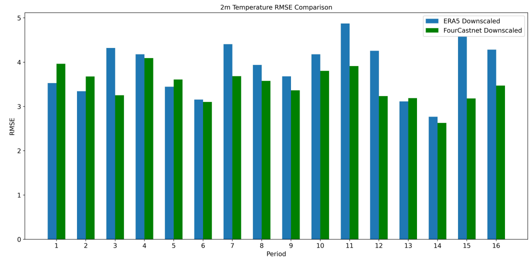
Figure 7a illustrates that FourCastNet generally maintains lower MAE values for 2m temperature, indicating a more accurate representation of temperature compared to ERA5. Similarly, Figure 7b shows that FourCastNet achieves lower RMSE values across most periods, suggesting better overall handling of temperature errors.

In the case of total precipitation, Figure 7c shows that FourCastNet also produces lower MAE values, especially during later periods, indicating superior precipitation forecasting accuracy. Figure 7d further supports this with lower RMSE values for FourCastNet, especially in periods with significant precipitation events.

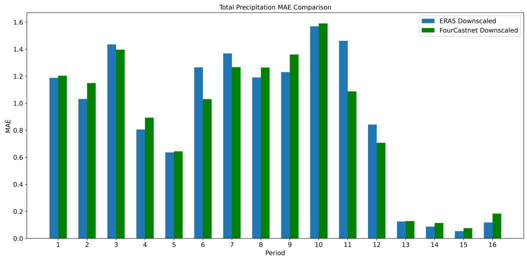
Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 August 2024 doi:10.20944/preprints202408.1322.v1



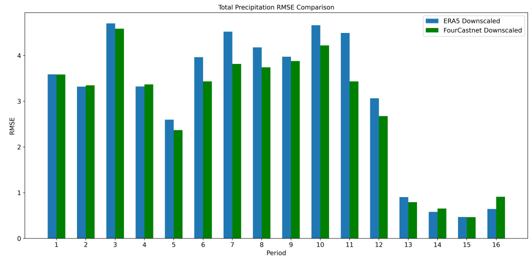
(**a**) 2m Temperature MAE Comparison



(**b**) 2m Temperature RMSE Comparison



(**c**) Total Precipitation MAE Comparison



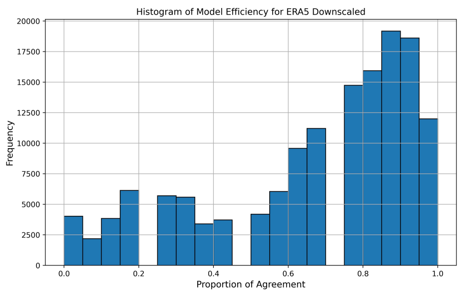
(**d**) Total Precipitation RMSE Comparison

Figure 7. Comparison of FourCastNet Downscaled and ERA5 Downscaled data for the Marche event: (a) Mean Absolute Error (MAE) for 2m temperature, (b) Root Mean Square Error (RMSE) for 2m temperature, (c) MAE for total precipitation, and (d) RMSE for total precipitation.

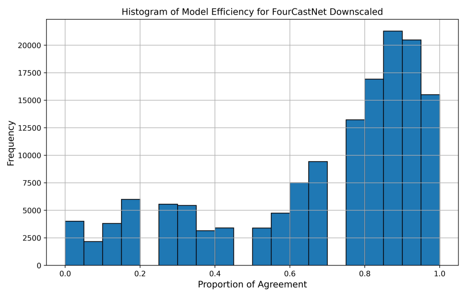
4.2.2. Total Precipitation Agreement Analysis

The agreement analysis, depicted in Figure 8, compares the alignment of the FourCastNet and ERA5 downscaled outputs with the ground truth for total precipitation.

Figure 8a indicates that the ERA5 downscaled data shows a strong level of agreement with the ground truth, particularly in higher agreement ranges. However, Figure 8b reveals that FourCastNet exhibits a more consistent alignment with the ground truth, particularly in the higher agreement ranges, indicating its superior capability in predicting precipitation events accurately.



(**a**) ERA5 Downscaled



(**b**) FourCastNet Downscaled

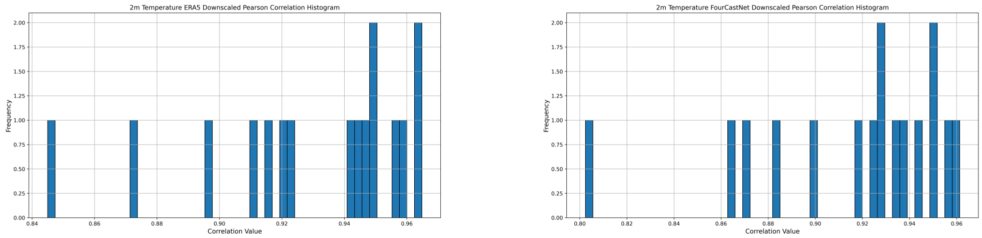
Figure 8. Histogram of Model Efficiency for Total Precipitation Agreement in the Marche event: (a) ERA5 Downscaled, (b) FourCastNet Downscaled.

4.2.3. Pearson Correlation For Temperature

Figure 9 shows the Pearson correlation histograms for 2m temperature, comparing the FourCastNet and ERA5 downscaled outputs with the ground truth data.

The ERA5 downscaled data, represented in Figure 9a, displays strong Pearson correlations, indicating a reliable linear relationship with the ground truth. FourCastNet, shown in Figure 9b, also maintains high Pearson correlations, though with a slightly broader distribution, suggesting good performance with some variability.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 August 2024 doi:10.20944/preprints202408.1322.v1



(**a**) ERA5 Downscaled (**b**) FourCastNet Downscaled

Figure 9. Pearson Correlation Histogram for 2m Temperature for the Marche event: (a) ERA5 Downscaled, (b) FourCastNet Downscaled.

4.2.4. Visual Comparison

For the Marche event, a visual comparison is provided in Figure 10 for 2m temperature and Figure 11 for total precipitation, showing the ground truth, FourCastNet downscaled, and ERA5 downscaled outputs for a selected timestep.

These figures demonstrate that the FourCastNet downscaled images generally offer a closer visual match to the ground truth, particularly in capturing finer spatial details and the intensity of the climate variables, compared to ERA5 downscaled images.

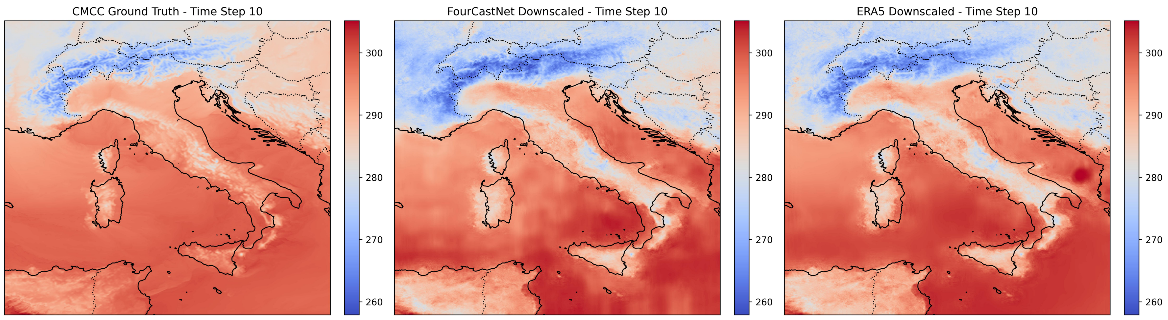


Figure 10. Visual comparison of 2m Temperature for the Marche event at a selected timestep. Left: Ground truth (CMCC dataset), Middle: FourCastNet downscaled, Right: ERA5 downscaled. Units in Kelvin (K).

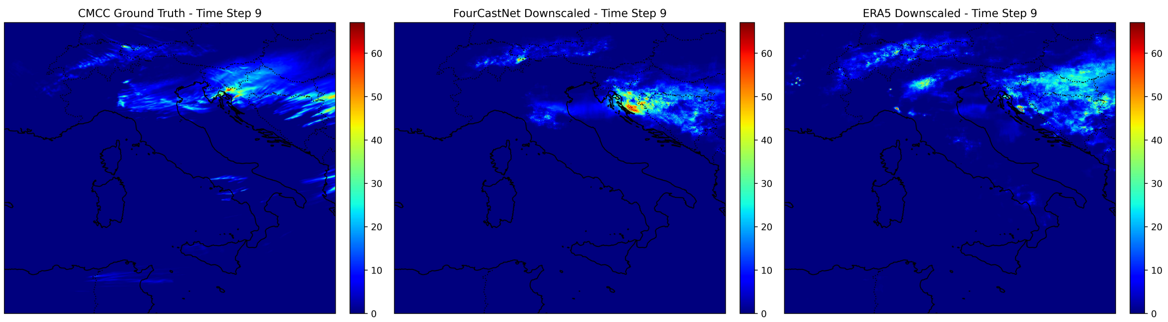


Figure 11. Visual comparison of Total Precipitation for the Marche event at a selected timestep. Left: Ground truth (CMCC dataset), Middle: FourCastNet downscaled, Right: ERA5 downscaled. Units in millimeters (mm).

5. Discussion

The integrated forecasting and downscaling model demonstrates strong performance in improving the resolution and accuracy of climate variables, particularly temperature and precipitation, across different types of extreme weather events. By leveraging the FourCastNet global forecasting system in conjunction with our downscaling model, we achieved significant improvements over traditional methods, as evidenced by lower MAE and RMSE values and better alignment with high-resolution ground truth data.

One of the key successes of E-TEPS is its ability to capture both extended and localized extreme events effectively. The results for the Emilia-Romagna event, which was widespread and prolonged, show that the FourCastNet downscaled outputs maintained lower error metrics compared to ERA5 downscaled data, particularly in handling total precipitation across a broad area. This is crucial for accurately forecasting large-scale events where the impact is spread over a significant region. Similarly, for the Marche event, which was more localized and intense, our model demonstrated superior performance in predicting sharp variations in precipitation and temperature, as shown by the strong agreement and high Pearson correlation values. This ability to handle the distinct characteristics of these two extreme events validates the robustness and flexibility of our early warning system.

The analysis also reveals that the model maintains a high level of accuracy throughout the entire 96-hour lead time for both 6-hour accumulated total precipitation and 2m temperature, even in the final stages of the forecast. This is particularly noteworthy given the distinct characteristics of the Emilia-Romagna and Marche events—one being widespread and prolonged, the other more localized and intense. The system was able to accurately capture the unique dynamics and evolution of both events, demonstrating its versatility in handling different types of extreme weather scenarios. This ability to adapt to and predict the diverse nature of these events underscores the model's robustness and its value as a reliable tool for early warning systems.

The successful application of the model to these two distinct events underscores the importance of selecting them for our study. The Emilia-Romagna and Marche events provided a comprehensive test of the model's capabilities, allowing us to demonstrate its effectiveness across different geoclimatic contexts. The model's consistent performance in both cases suggests that it is well-suited for a wide range of scenarios, from broad, regional-scale disasters to more concentrated, high-impact events. This versatility is a significant advantage, offering a more reliable tool for early warning systems in diverse environments.

However, the approach is not without limitations. A significant concern is the potential bias in the datasets used for training both the FourCastNet model and the downscaling model. The ERA5 reanalysis data, which serves as the foundation for both models, has known biases, particularly in its precipitation data. Research by [24] identified that ERA5 precipitation data tends to underestimate extreme precipitation events in the Tropics and over complex terrains. Similarly, studies by [25] and [26] highlight that ERA5 often overestimates light precipitation and underestimates heavier events, particularly during extreme weather conditions like typhoons. These biases in ERA5 data could propagate through the FourCastNet and downscaling models, affecting the accuracy of the forecasts, especially in regions where ERA5's limitations are most pronounced.

Similarly, the CMCC dataset, used for downscaling, while providing detailed regional forecasts, also inherits some of the biases present in ERA5 data. Although dynamically downscaling to a finer resolution offers detailed insights, it may also amplify certain biases, particularly in areas with complex topography or during extreme weather events, as noted by [12]. Despite these challenges, the overall accuracy of the forecasts remains high, and the ability to handle such extreme events effectively demonstrates the model's resilience.

The compatibility between the ERA5 and CMCC datasets also poses a potential limitation. Discrepancies between these datasets, such as those observed in precipitation events where ERA5 shows rainfall in areas where the CMCC dataset does not, as illustrated in Figure 12, suggest that the fine-tuning process may introduce additional errors. These discrepancies can lead to less reliable forecasts, particularly in regions where the datasets do not align well. This issue highlights the challenges of using multiple datasets with inherent biases and the importance of carefully considering these factors during model development and fine-tuning.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 August 2024 doi:10.20944/preprints202408.1322.v1

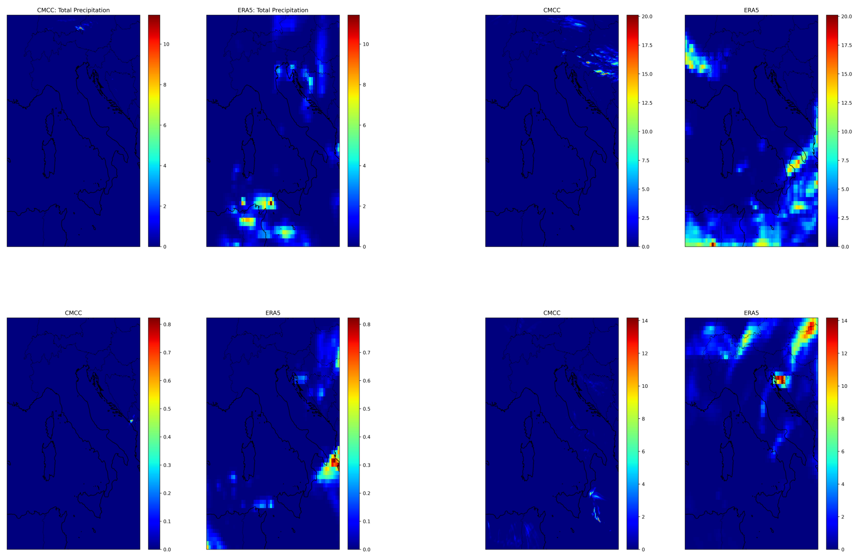


Figure 12. Comparison of total precipitation between the CMCC dataset (top) and ERA5 dataset (middle), both measured in millimeters (mm). The bottom image shows the discrepancies where ERA5 indicates precipitation events not present in the CMCC dataset. These comparisons are made at the same coordinates and for the same timesteps in both datasets. This highlights the potential errors introduced during the fine-tuning process due to the incompatibility between the two datasets, particularly in regions where the datasets do not align well.

In summary, the integrated model offers substantial advantages in enhancing the resolution and accuracy of climate forecasts, demonstrating its effectiveness across diverse and challenging weather events. While some limitations in the underlying datasets exist, they provide valuable opportunities for further refinement. By addressing these challenges, we can continue to improve the reliability and applicability of the model, ensuring its ongoing success in future climate forecasting efforts.

6. Conclusions

In conclusion, this study demonstrates the effectiveness of integrating advanced machine learning models, specifically FourCastNet and E-TEPS, into an EWS for improving climate forecast accuracy and resolution. The comprehensive evaluation of this system across two extreme weather events in Central Italy—Emilia-Romagna and Marche—highlights its capability to provide more precise and actionable climate predictions compared to traditional methods. The lower errors and higher alignment with high-resolution ground truth data validate the advantages of this integrated approach, particularly in capturing critical climate variables like temperature and precipitation.

Furthermore, the integrated system is capable of delivering high-resolution outputs and final forecasts in under one minute, ensuring that timely and accurate information is available for decision-making during extreme weather events. This rapid processing time is crucial for enhancing the effectiveness of early warning systems and improving disaster response.

However, the study also underscores the challenges associated with the inherent biases in the ERA5 and CMCC datasets, which form the foundation of the forecasting and downscaling models. These biases, particularly in precipitation data, may affect the accuracy of predictions in regions with complex topography or during extreme weather events. Additionally, the fine-tuning process, while essential for adapting the model to specific climatic conditions, can introduce further uncertainties, particularly when discrepancies between datasets are evident.

Despite these limitations, the findings suggest that the proposed EWS represents a significant step forward in enhancing disaster preparedness and response strategies. Future research should focus on addressing the identified biases and improving the compatibility between different datasets to further refine the accuracy and reliability of climate forecasts. By continuing to develop and optimize such integrated systems, we can better equip communities to anticipate and respond to the growing challenges posed by climate change.

References

1. Grasso, V.F.; Singh, A. Early warning systems: State-of-art analysis and future directions. *Draft report, UNEP* **2011**, *1*, 7.
2. Quansah, J.E.; Engel, B.; Rochon, G.L. Early warning systems: a review. *Journal of Terrestrial Observation* **2010**, *2*, 5.
3. Slater, L.J.; Arnal, L.; Boucher, M.A.; Chang, A.Y.Y.; Moulds, S.; Murphy, C.; Nearing, G.; Shalev, G.; Shen, C.; Speight, L.; others. Hybrid forecasting: blending climate predictions with AI models. *Hydrology and earth system sciences* **2023**, *27*, 1865–1889.
4. Cheng, J.; Liu, J.; Xu, Z.; Shen, C.; Kuang, Q. Generating High-Resolution Climate Prediction through Generative Adversarial Network. *Procedia Computer Science* **2020**, *174*, 123–127. doi:10.1016/j.procs.2020.06.067.
5. Cheng, J.; Liu, J.; Kuang, Q.; Xu, Z.; Shen, C.; Liu, W.; Zhou, K. DeepDT: Generative Adversarial Network for High-Resolution Climate Prediction. *IEEE Geoscience and Remote Sensing Letters* **2022**, *19*, 1–5. doi:10.1109/LGRS.2020.3041760.
6. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *Communications of the ACM* **2020**, *63*, 139–144.
7. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; others. Photo-realistic single image super-resolution using a generative adversarial network. Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4681–4690.
8. Jiang, Y.; Yang, K.; Shao, C.; Zhou, X.; Zhao, L.; Chen, Y.; Wu, H. A downscaling approach for constructing high-resolution precipitation dataset over the Tibetan Plateau from ERA5 reanalysis. *Atmospheric Research* **2021**, *256*, 105574. doi:10.1016/j.atmosres.2021.105574.
9. Chiang, C.H.; Huang, Z.H.; Liu, L.; Liang, H.C.; Wang, Y.C.; Tseng, W.L.; Wang, C.; Chen, C.T.; Wang, K.C. Climate Downscaling: A Deep-Learning Based Super-resolution Model of Precipitation Data with Attention Block and Skip Connections. *arXiv preprint arXiv:2403.17847* **2024**.
10. Vandal, T.; Kodra, E.; Ganguly, S.; Michaelis, A.; Nemani, R.; Ganguly, A.R. DeepSD: Generating High Resolution Climate Change Projections through Single Image Super-Resolution. Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2017, KDD '17. doi:10.1145/3097983.3098004.
11. Pathak, J.; Subramanian, S.; Harrington, P.; Raja, S.; Chattopadhyay, A.; Mardani, M.; Kurth, T.; Hall, D.; Li, Z.; Azizzadenesheli, K.; others. Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators. *arXiv preprint arXiv:2202.11214* **2022**.
12. Raffa, M.; Reder, A.; Marras, G.F.; Mancini, M.; Scipione, G.; Santini, M.; Mercogliano, P. VHR-REA_IT dataset: very high resolution dynamical downscaling of ERA5 reanalysis over Italy by COSMO-CLM. *Data* **2021**, *6*, 88.
13. Hersbach, H.; Bell, B.; Berrisford, P.; Biavati.; Muñoz Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Rozum, I.; others. ERA5 hourly data on single levels from 1979 to present. *Copernicus climate change service (c3s) climate data store (cds)* **2018**, *10*.
14. Hersbach, H.; Bell, B.; Berrisford, P.; Biavati.; Muñoz Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Rozum, I.; others. ERA5 hourly data on pressure levels from 1979 to present. *Copernicus climate change service (c3s) climate data store (cds)* **2018**, *10*.
15. Moccia, B.; Ridolfi, E.; Mineo, C.; Russo, F.; Napolitano, F. On the Occurrence of Extreme Rainfall Events Across Italy: Should We Update the Probability of Failure of Existing Hydraulic Works? *Water Resources Management* **2024**. doi:10.1007/s11269-024-03852-6.
16. Santangelo, M.; Althuwaynee, O.; Alvioli, M.; Ardizzone, F.; Bianchi, C.; Bornaetxea, T.; Brunetti, M.T.; Bucci, F.; Cardinali, M.; Donnini, M.; Esposito, G.; Gariano, S.L.; Grita, S.; Marchesini, I.; Melillo, M.; Peruccacci, S.; Salvati, P.; Yazdani, M.; Fiorucci, F. Inventory of landslides triggered by an extreme rainfall event in Marche-Umbria, Italy, on 15 September 2022. *Scientific Data* **2023**, *10*. doi:10.1038/s41597-023-02336-3.

 doi:10.20944/preprints202408.1322.v1

17. Molinari, D.; Arrighi, C.; Da Porto, F.; Domeneghetti, A.; Scorzini, A. The record-breaking flood in Central Italy in September 2022: preliminary impacts analysis from a field survey campaign **2023**. doi:10.5194/egusphere-egu23-13743.
18. Fraser-Baxter, S. Limited net role for climate change in heavy spring rainfall in Emilia-Romagna **2023**.
19. Sabelli, C. Examining the role of climate change in the Emilia-Romagna floods. *Nature Italy* **2023**. doi:10.1038/d43978-023-00082-z.
20. Tarolli, P.; Preti, F.; Romano, N. Terraced landscapes: From an old best practice to a potential hazard for soil degradation due to land abandonment. *Anthropocene* **2014**, *6*, 10–25. doi:10.1016/j.ancene.2014.03.002.
21. Pistocchi, A.; Calzolari, C.; Malucelli, F.; Ungaro, F. Soil sealing and flood risks in the plains of Emilia-Romagna, Italy. *Journal of Hydrology: Regional Studies* **2015**, *4*, 398–409. doi:10.1016/j.ejrh.2015.06.021.
22. Di Giustino, G.; Bonora, A.; Federico, K.; Reho, M.; Lucertini, G. Spatial Analysis of the Vulnerability to Flooding in the Rural Context: The Case of the Emilia Romagna Region. *Atmosphere* **2022**, *13*, 1181. doi:10.3390/atmos13081181.
23. Choudhary, P.K.; Nagaraja, H.N. *Measuring agreement: models, methods, and applications*; John Wiley & Sons, 2017.
24. Lavers, D.A.; Simmons, A.; Vamborg, F.; Rodwell, M.J. An evaluation of ERA5 precipitation for climate monitoring. *Quarterly Journal of the Royal Meteorological Society* **2022**, *148*, 3152–3165. doi:10.1002/qj.4351.
25. Jiang, Q.; Li, W.; Fan, Z.; He, X.; Sun, W.; Chen, S.; Wen, J.; Gao, J.; Wang, J. Evaluation of the ERA5 reanalysis precipitation dataset over Chinese Mainland. *Journal of Hydrology* **2021**, *595*, 125660. doi:10.1016/j.jhydrol.2020.125660.
26. Jiao, D.; Xu, N.; Yang, F.; Xu, K. Evaluation of spatial-temporal variation performance of ERA5 precipitation data in China. *Scientific Reports* **2021**, *11*. doi:10.1038/s41598-021-97432-y.