

A new conditional generative adversarial neural network approach for statistical downscaling of the ERA5 reanalysis over the Italian Peninsula



Ilenia Manco^{a,b,*}, Walter Riviera^c, Andrea Zanetti^c, Marco Briscolini^d, Paola Mercogliano^b, Antonio Navarra^b

^a Department of Physics and Astronomy, University of Bologna, Bologna, Italy

^b CMCC Foundation - Euro-Mediterranean Center on Climate Change, Italy

^c Intel Corporation, Italy

^d E4 Computer Engineering SpA, Italy

ARTICLE INFO

Keywords:

Statistical downscaling
Climate downscaling
High-resolution climate modeling
Artificial intelligence
Neural networks

ABSTRACT

State-of-the-art General Circulation Models (GCMs) typically operate at a coarse spatial resolution, requiring a refinement to assess regional climate changes and their impacts. This weakness is mainly known for representing regional-scale topography and meteorological processes, particularly those responsible for extreme events. Dynamical downscaling methods are computationally demanding. In contrast, though computationally efficient, statistical approaches often sacrifice spatial consistency. To address these limitations, this work introduces an innovative and robust Conditional Generative Adversarial Network (cGAN) architecture for statistical downscaling, discussing the methodology, advantages, and contributions to refining predictions at a finer scale. By leveraging a generator-discriminator architecture, the cGAN developed permits to downscale ERA5 reanalysis at the local scale to obtain a new high-resolution dataset (~ 2.2 km), ERA5-DownGAN. The results obtained show the cGAN's architecture presented accurately reproduces the patterns, value range, and extreme values generated by dynamical models for the 2-m temperature over the Italian Peninsula.

1. Introduction

Statistical downscaling techniques represent an alternative or complementary approach to dynamical downscaling methods for increasing the spatiotemporal resolution of general circulation models (GCMs). The impacts of climate change and adaptation strategies often require higher spatial resolution, making it crucial to bridge the gap between climate data produced by GCMs, which typically operate at low resolutions (often 100–200 km), and climate information at the local scale. Outputs derived from downscaling techniques, whether dynamical or statistical, play a central role in developing a comprehensive understanding of localized impacts of climate variability and change, particularly in the context of extreme climate events. Local climate data are essential for improving risk assessments and formulating adaptation strategies across various sectors (Kondrup C. et al., 2022). The application of downscaled climate data is crucial in multiple domains. In agriculture, downscaled precipitation and temperature projections can inform crop management strategies and irrigation planning under future climate scenarios (Komasi et al., 2024; Xiao D. et al., 2021). In hydrology, accurate

downscaling of precipitation data is vital to assess water resource availability and flood risks, as low-resolution models often misrepresent precipitation extremes (Sun Q. et al., 2022). Moreover, downscaled climate variables are increasingly used in ecological modeling to evaluate potential impacts of climate change on biodiversity and habitat suitability (Tabor K. and Williams J.W., 2010; Zhang and Georgakakos, 2012). The focus on downscaling surface temperature in this study is directly linked to the critical role this variable plays in assessing climate change impacts, particularly with regard to the intensity of heatwaves and their effects on human health and infrastructure. Sharghi A. et al., 2025 also highlighted the importance of identifying optimal delay times in groundwater level (GWL) projections using downscaled meteorological data, particularly to improve prediction accuracy and address issues such as land subsidence. Ch. Sridevi et al. (2024) emphasize the importance of high-resolution surface temperature data in capturing spatial variability in heat stress, especially in regions prone to extreme temperature events. Similarly, Schwingshackl C. et al. (2024) highlight the value of such data in improving heat stress predictions, particularly in urban areas where the urban heat island effect can amplify the

* Corresponding author. Department of Physics and Astronomy, University of Bologna, Bologna, Italy.

E-mail address: ilenia.manco@cmcc.it (I. Manco).

impacts of climate change. Furthermore, [Perkins-Kirkpatrick S. E. & Gibson P. B. \(2017\)](#) underscore the necessity of high-resolution surface temperature data for accurately projecting future heatwave occurrences, particularly in the Mediterranean, where heatwave intensity is expected to increase significantly. There are two main types of downscaling: dynamical and statistical. Dynamical downscaling involves embedding a high-resolution regional climate model (RCM) within a GCM for the area of interest ([Rummukainen M., 2010](#)). This approach relies on the physical laws governing atmospheric processes to provide a more detailed and accurate representation of climate dynamics. It is particularly effective in regions with complex terrains, as it captures intricate physical interactions. However, dynamical downscaling requires significant computational resources and may still inherit biases from the GCMs providing the initial conditions. Conversely, statistical downscaling establishes empirical relationships between large-scale and local or regional variables, enabling the creation of high-resolution climate datasets with lower computational demands ([Vandal T. et al., 2017](#); [Maraun D. and Widmann M., 2018](#)).

Due to its efficiency and ease of application, statistical downscaling is often preferred over dynamical methods, which use RCMs to simulate small-scale climate processes ([Tiwari P. et al., 2018](#)). The computational efficiency, versatility, and ability to learn complex relationships from data make the statistical approach a valuable downscaling technique, especially when implemented through machine learning algorithms (including deep learning). Various categorizations of downscaling methods exist ([Klein W.H. and Glahn H.R., 1974](#); [Wilby R.L. and Wigley T.M.L., 1997](#); [Maraun D. and Widmann M., 2018](#)), but two primary subgroups commonly divide empirical downscaling strategies: observational downscaling (or empirical statistical downscaling, ESD) and RCM emulation (regional climate models). Observational downscaling involves developing a model trained on real observational data. Specifically, according to [Rampal et al. \(2024\)](#), statistical downscaling approaches can be classified into Perfect Prognosis (PP), Super-Resolution (SR), Weather Generators (WG), and Model Output Statistics (MOS). PP methods rely exclusively on observational data for model training, offering flexibility but lacking the ability to correct biases in GCMs, unlike MOS, which applies corrections during calibration specific to each GCM. WG adopts a stochastic approach to generate synthetic weather scenarios, effectively capturing temporal variability and spatial consistency. SR, a subcategory of PP, focuses on predicting high-resolution fields directly from low-resolution inputs, enhancing spatial detail. Despite their promising computational efficiency, ESD methods are not without limitations. Firstly, ESD methods are constrained by the availability of long-term observational data, limiting their ability to produce high-resolution data only for regions and variables with sufficient local observations. ESD methods are also restricted in GCM selection due to the requirement for specific variables at certain pressure levels and with a specific temporal frequency for effective calibration. Moreover, ESD methods rely on the assumption of stationarity in the relationship between global and local scales, implying that a statistical model calibrated on past and present climate data remains reliable for future climate scenarios ([Doury A. et al., 2022](#)). Numerous studies have demonstrated that this assumption can significantly influence results ([Wilby R.L. et al., 1998](#); [Dayon G. et al., 2015](#); [Erlandsen H. et al., 2020](#)). To address the high computational cost of dynamical downscaling (DD) and the limitations of ESD, hybrid models that combine the advantages of dynamical and statistical downscaling have been introduced, known as regional climate model (RCM) emulators. Initially defined by [Maraun and Widmann \(2018\)](#), these emulators use predictors from low-resolution RCM-GCM simulations and predictands from high-resolution RCM simulations. Hence, the final high-resolution output obtained from an RCM emulator is only constrained by the resolution of the RCM output, and performance can be at most comparable to that of RCMs, as this methodology cannot surpass the intrinsic biases in RCMs ([Giorgi F. et al., 2009](#)). RCM emulators operate without observational constraints, allowing them to be trained on simulations of

both historical and future periods. These emulators can be classified into two frameworks based on their training methodology: “imperfect” and “perfect.” In the “imperfect” framework, an emulator is trained to map outputs directly from GCMs to RCMs. Conversely, in the “perfect” framework, the process starts with upscaling the resolution of the RCM to match that of the GCM, followed by training an algorithm to establish a mapping between the upscaled RCM and the original RCM ([Rampal N. et al., 2024](#)). In the imperfect framework, the emulator learns a specific relationship for each GCM-RCM pair, functioning as a Model Output Statistics (MOS) technique. In contrast, the perfect framework focuses on learning general relationships between fields at different resolutions ([Boé J. et al., 2023](#)). This framework typically handles weak correlations and a degree of “independence” between RCM and/or GCM fields, aligning with the Perfect Prognosis (PP) approach. Training within the perfect framework is more aligned with PP as it only learns general relationships between low- and high-resolution RCM pairs ([Boé J. et al., 2023](#)). The consistency between low- and high-resolution RCM pairs during training in the perfect framework simplifies the emulator’s training compared to the imperfect framework ([Rampal et al., 2024](#)). Empirical Statistical Downscaling (ESD) methods and RCM emulators share some of the most widely used machine learning algorithms (including deep learning), such as multiple linear regression (e.g., [Sharifi E. et al., 2019](#) for observational downscaling and [Holden P.B. et al., 2015](#) for RCM emulators), random forests (e.g., [Limon G. and Jablonowski C., 2023](#) for RCM emulators and [Hutengs C. and Vohland M., 2016](#) for use in ESD), generalized linear models ([Baño-Medina J. et al., 2020](#) for ESD and [Maraun D. et al., 2017](#) for RCM emulators), multilayer perceptrons (e.g., [Nishant N. et al., 2023](#) for RCM emulators and [Hobeichi S. et al., 2023](#) for PP), or short-term memory networks (e.g., [Bittner M. et al., 2023](#) for RCM emulators and [Hobeichi S. et al., 2023](#) for ESD). The difference between these two strategies lies in how these techniques are employed, what inputs are provided, and what outputs are desired. For instance, traditional PP methods may not be suitable for downscaling nonlinear variables such as precipitation, as their site-specific approach (i.e., extracting variables one “site” at a time) does not incorporate information from adjacent grid points. Additionally, these methods can be computationally inefficient, as the algorithm must be repeatedly applied point-by-point on large datasets. RCM emulators, by receiving one or more fields as input across the entire spatial grid, generate high-resolution output fields for the entire region covered by the RCM. In recent years, promising results have emerged in applying computer vision algorithms to meteorology and climatology, particularly in downscaling tasks. Specifically, generative models such as Generative Adversarial Networks (GANs) have been extensively studied, with [Leinonen J. et al. \(2021\)](#) providing a comprehensive description of their usage. Additionally, diffusion models, as described by [Mardani M. et al. \(2023\)](#), have shown promising results. These models employ various architectures, including Convolutional Neural Networks (CNNs), U-Net, Fully Connected Networks (FCNs), or Multilayer Perceptrons (MLPs). The complexity of these models entails trade-offs among computational costs, horizontal resolution, and, in some cases, domain size ([Doury A. et al., 2022](#)). Generative Adversarial Networks (GANs), introduced by [Goodfellow I. et al., in 2014](#), thus offer a promising alternative to traditional downscaling methods, overcoming some key limitations. Regression-based models, for example, estimate the conditional mean of a climate variable based on large-scale atmospheric states. While these models provide mean value predictions for each time point, they often underestimate the variability of time series, as they primarily focus on mean estimation and fail to capture fluctuations or extreme events adequately. GANs provide an innovative and flexible downscaling process. They learn complex relationships directly from training data ([Leinonen J. et al., 2021](#)), surpassing the limitations of quantile mapping, which applies fixed statistical rules and may struggle with changes in distribution shapes or extreme events. By avoiding rigid assumptions, GANs effectively adapt to complex and nonlinear temporal variations, making them particularly suitable for handling climate

change scenarios. Unlike Bayesian approaches, which require large amounts of data and specific assumptions, GANs operate in an unsupervised learning context, making them more adaptable in situations with limited labeled data. Although advanced methods like weather generators can estimate conditional probability distributions of climate datasets and offer a wider range of possible outcomes, their typically site-specific nature often limits their ability to preserve spatial consistency in downscaled climate fields. GANs address this limitation by learning complex and nonlinear relationships across the area of interest rather than relying on site-specific empirical relationships. Moreover, GANs can adopt a stochastic approach to quantify the uncertainty associated with the high-resolution dataset produced while maintaining strong spatial coherence. These capabilities enable GANs to produce realistic fields, making them particularly effective at capturing local-scale climate variability and accurately representing extreme events. Consequently, GANs are emerging as promising tools in the landscape of statistical downscaling, playing a significant role in the broader domain of machine learning. Despite their advantages, GANs face challenges such as training instability and computational complexity, which present opportunities for further research to enhance their robustness and practical applicability, thereby expanding their potential for diverse applications.

2. Data

This section provides a detailed overview of the two datasets with distinct resolutions required for the Super-Resolution cGAN developed in this work: 1) the low-resolution dataset, which serves as the starting point for the statistical downscaling process; 2) the high-resolution dataset, which is only required during the training step and represents the final resolution to achieve. In our preliminary assessment of the novel algorithm developed for statistical downscaling, we have chosen to focus on the daily 2-m temperature. This allows us to evaluate the algorithm's performance and its ability to downscale temperature data effectively. Additionally, this section describes the data preprocessing

procedures which are crucial for training cGANs. Given that cGANs rely on the comparison during the training phase between real and generated data, normalizing distributions and ensuring a coherent representation of the studied phenomena directly impacts the cGANs' ability to learn underlying patterns in the data and generate more plausible and meaningful results. Therefore, a careful preprocessing phase is essential to optimize the performance and quality of cGANs.

2.1. ERA5 reanalysis

In this approach, we deviated from employing random sequences for the low-resolution dataset, as often seen in prior GAN-based studies. Instead, we opted to utilize a cGAN, imposing as a condition the use of a real dataset, specifically the ERA5 reanalysis data (Fig. 1a) at a horizontal resolution of 0.25° (≈ 31 km; Hersbach et al., 2023), which can be accessed online at (<https://doi.org/10.24381/cds.adbb2d47>). The selected dataset spans fifteen years, divided into two distinct periods: a training period (01/1990–12/2000) and a test period (01/2001–12/2005). A hold-out approach was employed to test the developed architecture, using a 10-year training window and a 5-year testing window to limit computational costs. Future experiments will explore cross-validation and larger time windows to enhance performance.

2.2. VHR-REA_IT

The very High-Resolution Dynamical Downscaling of ERA5 Reanalysis (VHR-REA_IT) is chosen as the high-resolution dataset (Fig. 1b). This reanalysis is at the convection-permitting scale (horizontal grid spacing 0.02° , ≈ 2.2 km; Raffa et al., 2021) by COSMO in CLimate Mode (COSMO-CLM) on a domain covering the Italian Peninsula, described by Raffa, M. et al. (2021) available for download at https://doi.org/10.25424/cmcc/era5-2km_italy. The high-resolution dataset covers the same periods selected for the low-resolution dataset. During the training period (1990–2000) we used both the low-resolution and the

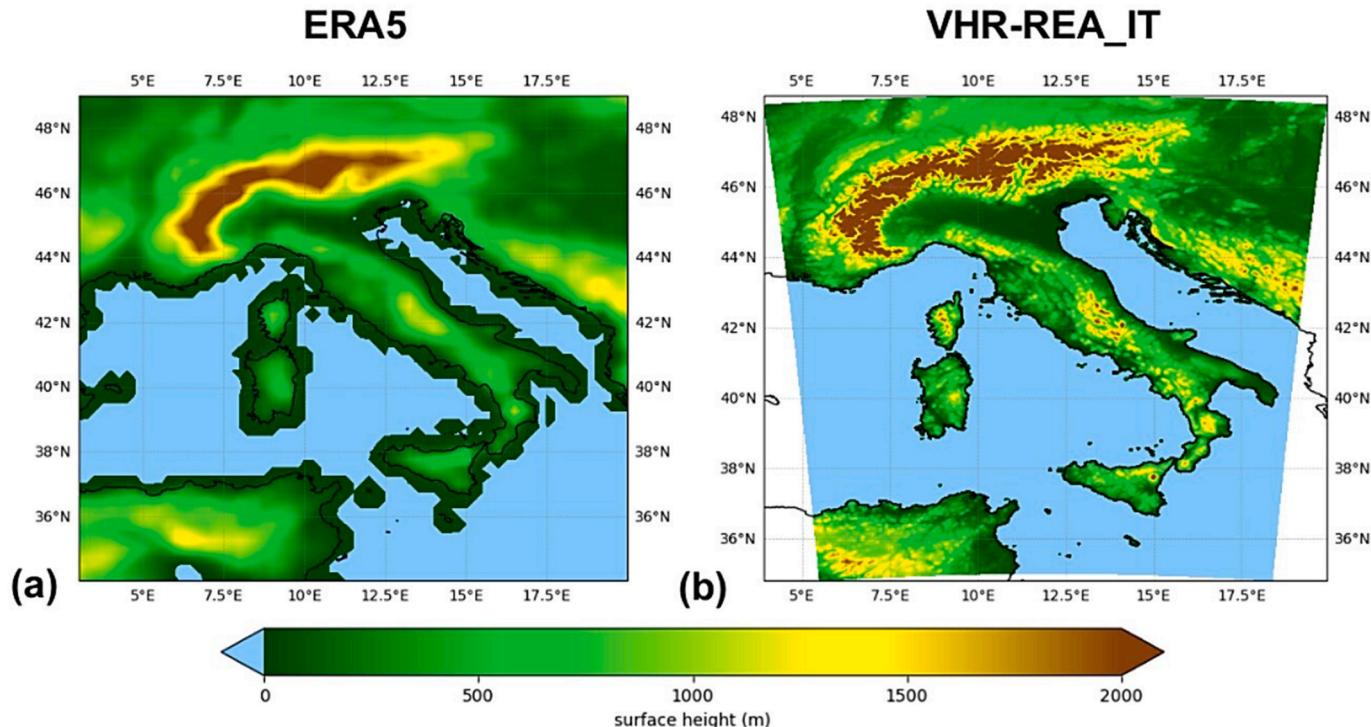


Fig. 1. Topography
(a)–(b) ERA5; VHR-REA_IT.

high-resolution dataset, while during the test period (2001–2005) we used only the high-resolution dataset in the GAN algorithm. The high-resolution dataset is subsequently taken into account to validate the results obtained during the test period by the cGAN to judge the goodness of these at the same horizontal resolution.

2.3. Preprocessing

In this section, the preprocessing procedure of the data will be detailed, which plays a significant role in training cGANs. Given that cGANs rely on the comparison during the training phase between real and generated data, normalizing distributions and ensuring a coherent representation of the studied phenomena directly impacts the cGANs' ability to learn underlying patterns in the data and generate more plausible and meaningful results. Therefore, a careful preprocessing phase is essential to optimize the performance and quality of cGANs.

2.4. Domain

The domain was chosen based on that defined in Raffa et al. (2021) for VHR-REA_IT. For the high-resolution dataset (VHR_REA_IT), its original domain was taken, and for the low-resolution dataset (ERA5), a domain was selected as close as possible to the boundaries defined by the high-resolution one, covering the Italian Peninsula (Fig. 1). The selection of a domain with similar longitudinal and latitudinal extents in both high-resolution and low-resolution real datasets is crucial, as it facilitates the downscaling process during the transition from low to high resolution (Table 1). This ensures a geographical correspondence between the datasets and, consequently, a consistent generation of the new artificial high-resolution dataset. A consistent geographical correspondence between the two datasets implies that similar points in the two datasets occupy similar positions in geographical space. This enables the model to learn spatial relationships between corresponding points, contributing to the coherent generation of high-resolution geographical details.

Furthermore, consistency in geographical space aids the cGAN in capturing relevant topographic information, and correctly positioning, for example, mountain ranges in the new artificial high-resolution dataset. This consideration is significant in geographically complex contexts such as the Italian Peninsula, characterized by diverse topography and a pronounced transition between areas with markedly different morphologies. This contributes to defining distinctive climatic features for each region, underscoring the importance of accurate domain selection to preserve geographical and climatic coherence in the results obtained through the downscaling process.

2.5. Normalization

Input normalization in a neural network, such as cGANs, is a process aimed at ensuring that the input data is standardized and consistent (Patro, S. Gopal and Sahu, Dr-Kishore Kumar, 2015). This aids the model in converging faster during the training process and enhancing overall performance. There exist several normalization techniques: standardization (Z-score normalization), which transforms variables to have a mean of zero and a standard deviation of one, thus facilitating model

Table 1

Domain details for LW and HR dataset: geographical boundaries, horizontal resolution, number of grid points.

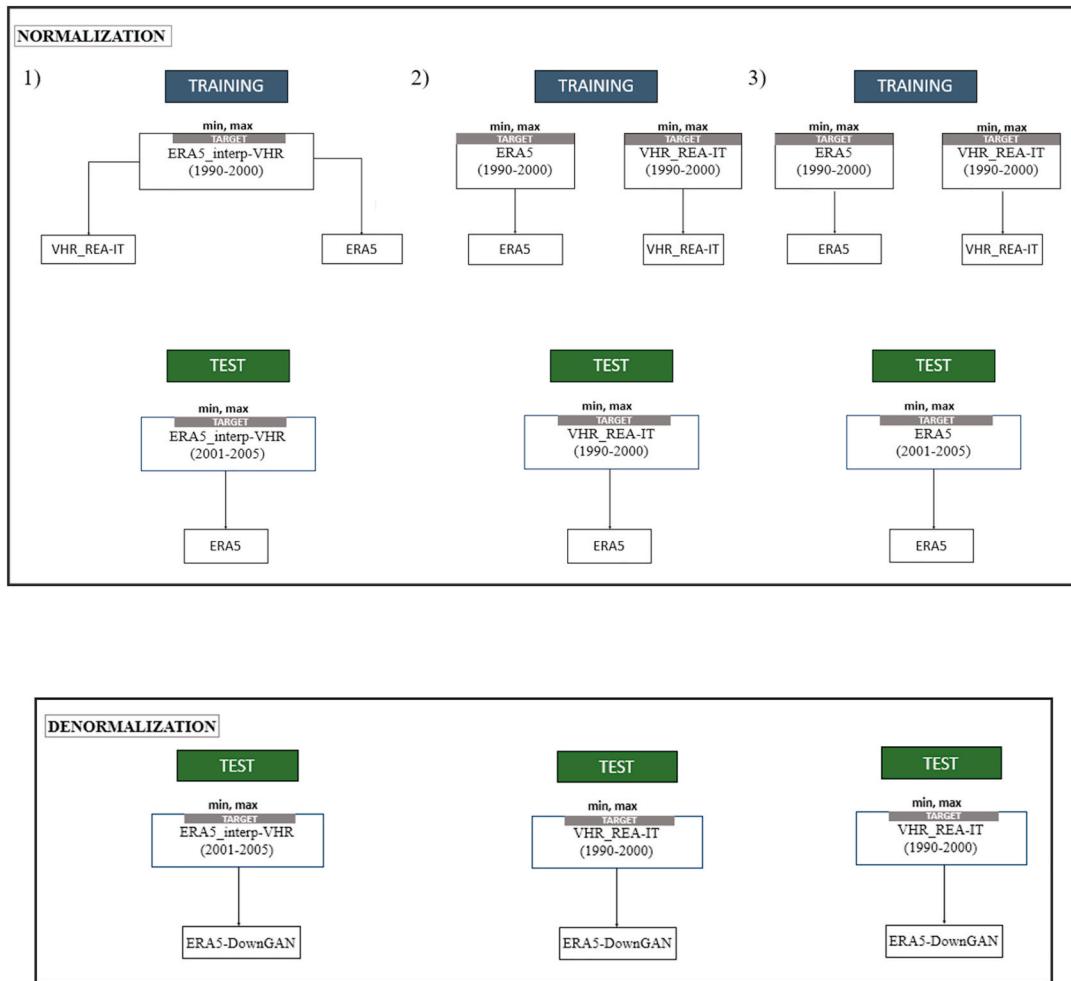
Data	longitude and latitude	horizontal resolution	n° grid points
ERA5	(Lon = 4.00° W–23.75° E; Lat = 34.25° N–48.50° N)	0.28° (\approx 31 km)	58 × 80
VHR-REA_IT	(Lon = 3.91° W–19.93° E; Lat = 34.80° N–48.59° N)	0.02° (\approx 2.2 km)	680 × 535

convergence; min-max normalization, which constrains the data range between 0 and 1, contributing to greater stability during training; Batch Normalization and Layer Normalization, which maintain a consistent distribution of inputs across each layer and promote faster convergence. This array of normalization techniques played a crucial role in shaping the behavior of our cGAN network, optimizing data representation, and enabling the generation of consistent, high-quality outputs. A targeted normalization approach was adopted to handle temperature data within our modeling framework. Before utilizing the data as input for the cGAN generator, the data were standardized to range between -1 and 1. This transformation was designed considering the natural distribution of temperatures and the expected range of variation. The normalization stabilized and balanced the data consistently, rendering them suitable for the generation process and enhancing model convergence. Subsequently, the hyperbolic tangent function was employed as the activation function in the generator's final layer, ensuring that the model's output remained consistent with the range of temperatures in degrees Celsius. The high-resolution artificial dataset generated by the cGAN (ERA5-DownGAN) was also scaled back using the same scaling. This normalization strategy proved crucial in ensuring that the generator produced coherent and meaningful data in response to a normalized input. It has been decided not to standardize the data initially, as this approach can be useful if one aims to ensure that the data is balanced concerning the mean and standard deviation. The objective is to normalize the input data, particularly both the real high-resolution dataset (VHR-REA_IT) and the low-resolution real dataset (ERA5) for both the training and testing periods. This can be achieved by selecting a target dataset for both the training and testing periods, from which the minimum and maximum values are calculated to properly rescale the input data. The sensitivity for the normalization choice has been explored, and we have applied various normalization and denormalization target datasets to optimize training and enhance the performance of our cGAN network: Rescaling all input data concerning the minimum and maximum values of a dataset obtained through simple bilinear interpolation of the ERA5 dataset onto the VHR-REA_IT grid (ERA5_interp-VHR). The calculation of minimum and maximum values is performed for temperature distributions corresponding to both the training and testing periods. The ERA5-DownGAN is denormalized for the dataset ERA5_interp-VHR from the testing period. Using the high-resolution dataset training period for both the training and testing periods, utilizing it on both the VHR-REA_IT and ERA5 datasets. In this second case, the ERA5-DownGAN is denormalized concerning the dataset VHR-REA_IT of the training period. Application of a different target for the training and test phases.

Specifically, we normalize both datasets with respect to themselves during the training phase. However, during the testing phase, we normalized ERA5 with respect to itself and denormalized the output of the cGAN, considering the dataset VHR-REA_IT from the training phase as the target. The calculation of minimum and maximum values is consequently performed on the VHR-REA_IT dataset for the training period. This is imperative as, for the testing period, we may only have access to the low-resolution dataset, from which we aim to obtain a high-resolution counterpart through the cGAN. Despite testing and evaluating all normalization and denormalization approaches (Fig. 2), we have opted to exclusively showcase the outcomes of normalization/denormalization 3). This specific approach is deemed the most rigorous within the climate change context and demonstrates highly promising results when scrutinizing error metrics and statistical analyses. This decision is underpinned by the primary aim of normalization, which is to preserve the physical consistency between low-resolution and high-resolution data, a pivotal factor for accurately representing future climatic variations.

3. Architecture

The approach developed in this work for constructing a new method

**Fig. 2. Normalization and denormalization**

Normalization (a) and denormalization (b) approaches with different target datasets.

based on neural networks is grounded in the general principles of GANs and draws inspiration from the structure of Conditional GANs (Mirza M. and Osindero S., 2014). However, it is important to emphasize that the construction of this method is original, featuring substantial modifications compared to existing approaches. This innovation allows us to address specific scientific and practical challenges while ensuring the autonomy of our work. The core idea of the algorithm is to establish the empirical link between low-resolution and high-resolution datasets during the training period and apply this relationship to the low-resolution dataset in the test period to generate a new artificial high-resolution dataset. GANs consist of two key components: a generative model (the generator) and a discriminative model (the discriminator), both implemented as neural networks. In the context of super-resolution with GANs, the generative model's goal is to produce synthetic data samples that closely resemble high-resolution real data, while the discriminative model is trained to distinguish between real high-resolution data from the training set and synthetic high-resolution data produced by the generator, conditioned on the corresponding low-resolution data. By continuously training these two networks against each other in a feedback loop, GANs have demonstrated their ability to generate highly realistic climatic data, significantly advancing the field of climatic downscaling (reference). Specifically, the use of new cGAN enables the generation of high-resolution fields that retain the fundamental characteristics of the original data, utilizing low-resolution data as conditioning input for the generator. This approach effectively mitigates the introduction of noise or undesirable artifacts, ensuring that

authentic information serves as the foundation for the generation process. The objective of the cGAN is for the generator to produce synthetic data that challenges the discriminator to distinguish between real high-resolution data and the generated output. The discriminator learns to differentiate between actual high-resolution data and the artificial high-resolution data generated by the cGAN. Concurrently, the generator seeks to refine its output to deceive the discriminator into classifying the generated data as real. GANs, like other deep learning architectures, are structured with a directed graph that allows data to pass through multiple transformational layers until it reaches the output. The evaluation of the loss function is crucial as it measures the discrepancy between the network's predicted output and the expected result. Optimization algorithms, such as the refined Adam algorithm employed in this architecture (see details in "Supplementary Material"), use gradients from this loss function to update the weights and biases of the neural networks. Weight assignments characterize this framework to connections between neurons across different layers, which affects how data is transformed within the network. In this study's cGAN, specific activation functions were chosen for various segments of the architecture, and distinct loss functions were used for the generator and discriminator. These design choices, along with the optimizer employed, are essential for the cGAN's capacity to learn complex representations and produce high-quality outputs.

The generator in cGAN discussed in this work uses four linear layers with LeakyReLU activation functions (Dubey and Jain, 2019), except for the last layer which uses a hyperbolic tangent activation function. The

LeakyReLU helps prevent the vanishing gradient problem, introducing a small slope for negative inputs as described in Cap. I. This contributes to enhanced stability and overall learning performance of the neural network. The final tanh activation scales the output to a range between -1 and 1 . The generator's input data consists of matrices with dimensions of 58×80 , corresponding to a horizontal resolution of 31 km. The architecture we developed for the two neural networks, Generator (G) and Discriminator (D), in this new cGAN is illustrated in Fig. 3.

The primary objective of the generator is to downscale this input to matrices of data with dimensions of 680×535 , equivalent to a substantially higher horizontal resolution of 2.2 km. This enhancement facilitates a finer level of detail and precision in the generated data. In particular, the choice to use 1160 neurons in the first layer of the discriminator, as well as the decision to utilize 290 neurons in the first layer of the generator, is based on a combination of theoretical considerations and empirical results from previous experiments. This number is particularly reasonable given the input matrices' context and complexity. By employing 1160 neurons, the model can capture a significant amount of spatial features from the high-dimensional data while maintaining a balance between learning capacity and the risk of overfitting. As suggested by He K. et al. (2015), the number of neurons in a layer should reflect the complexity of the problem being addressed; in the case of matrices with dimensions of 680×535 , this configuration is justified by the necessity to learn detailed representations. Furthermore, experiments conducted in previous studies have demonstrated that an appropriate number of neurons in the initial layers of a discriminator improves performance in image classification, enhancing the model's ability to generalize to new data (Zhang H. et al., 2019). Therefore, the specific choices regarding the number of neurons utilized are not only supported by the literature but also result from an empirical optimization process. The discriminator also employs linear layers with LeakyReLU activation functions, and dropout layers to introduce regularization. The dropout rate is set at 0.3 in both the generator and discriminator networks. This value is chosen to strike a balance between preventing overfitting and maintaining sufficient capacity for learning. A dropout rate of 0.3 allows the model to learn more robust features by randomly dropping 30% of the neurons during training (Hinton G.E. et al., 2012), thus promoting better generalization to unseen data. Dropout helps prevent overfitting during training. The final layer of the discriminator utilizes a sigmoid activation function that constrains the final output between 0 and 1 .

4. Results

In this section, the cGAN's performance in the downscaling process of the $2m$ -temperature field (T_{2M}) over the Italian Peninsula is evaluated by putting side by side-the synthetic high-resolution dataset (ERA5-DownGAN) generated by the cGAN against the real high-resolution dataset (VHR-REA_IT) produced through dynamic downscaling over the specified test interval (01/2001-12/2005). The assessment of the cGAN's effectiveness is conducted during the test period, wherein exclusively the low-resolution dataset is available. The primary aim is to derive a novel high-resolution dataset utilizing the optimized generator from the antecedent training phase. The high-resolution dataset does not enter any calculation during the testing phase, exclusively employed for validation to assess the similarity between the statistically downscaled results data by cGAN compared to the dynamically downscaled counterparts. A suite of conventional error metrics (BIAS, MAE, RMSE, and CORR), combined with graphical representations, are analyzed to assess the performance of the cGAN for this specific geographic domain and meteorological field.

Within this comparison, the $2m$ temperature field of the low resolution (ERA5) is further examined to assess its deviation from the high resolution and to evaluate the cGAN's ability to transition from the distribution characterizing the low resolution to that typical of the high-resolution counterpart. This analysis aims to provide an in-depth understanding of the cGAN's behavior in handling differences between the $2m$ temperature distributions across the two resolution levels, thus contributing to the overall assessment of the model's performance in the context of climatic downscaling. The time series of the daily moving average of the $2m$ -temperature exhibit a considerable correlation among themselves, with a slight discrepancy in terms of value ranges (Fig. 4a). The real low-resolution dataset ERA5 generally records higher temperatures, especially during the colder months compared to VHR, with values oscillating in a range from ~ 6 $^{\circ}\text{C}$ to about 28 $^{\circ}\text{C}$. On the other hand, the high-resolution dataset generated by cGAN (ERA5-DownGAN) not only mirrors the trend of the reference high-resolution real dataset (VHR-REA_IT), but also records values falling within the same range, showing greater proximity to the VHR-REA_IT dataset than to ERA5, especially during the colder months, ranging from ~ 4 $^{\circ}\text{C}$ to ~ 28 $^{\circ}\text{C}$. This highlights the capability of cGAN to generate not only a temporally correlated dataset but also to maintain spatial mean values that deviate very little from the reference dataset. The analysis of the time series (Fig. 4b) of daily differences between the dataset generated by the cGAN, the ERA5-DownGAN dataset, and the real VHR-REA_IT dataset

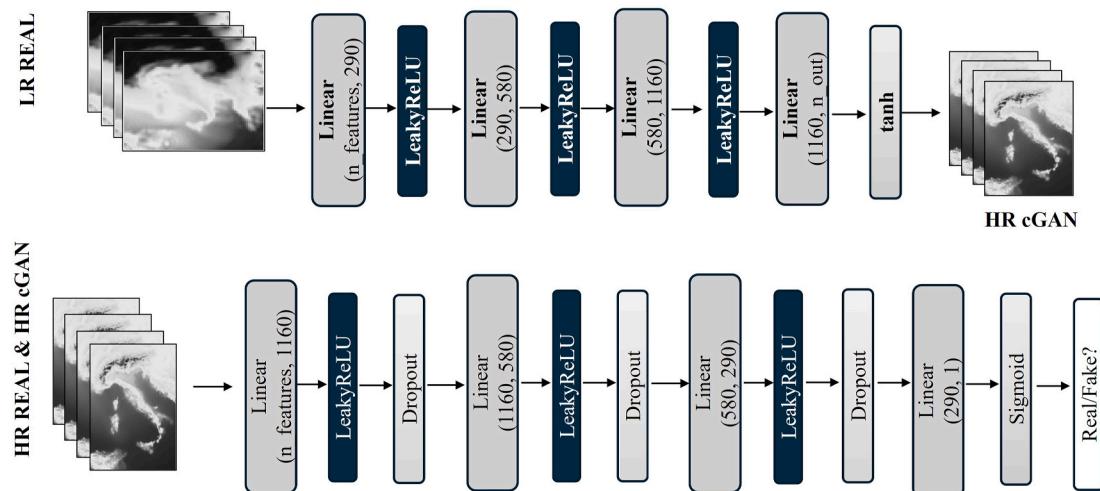


Fig. 3. Network architecture

Architectures of (a) generator and (b) discriminator. Size input generator 58×80 , size input discriminator 680×535 , the number of frames per sequence 100 (batch size = 100) for both data sets considered here. After training, the best generator is used for the test period.

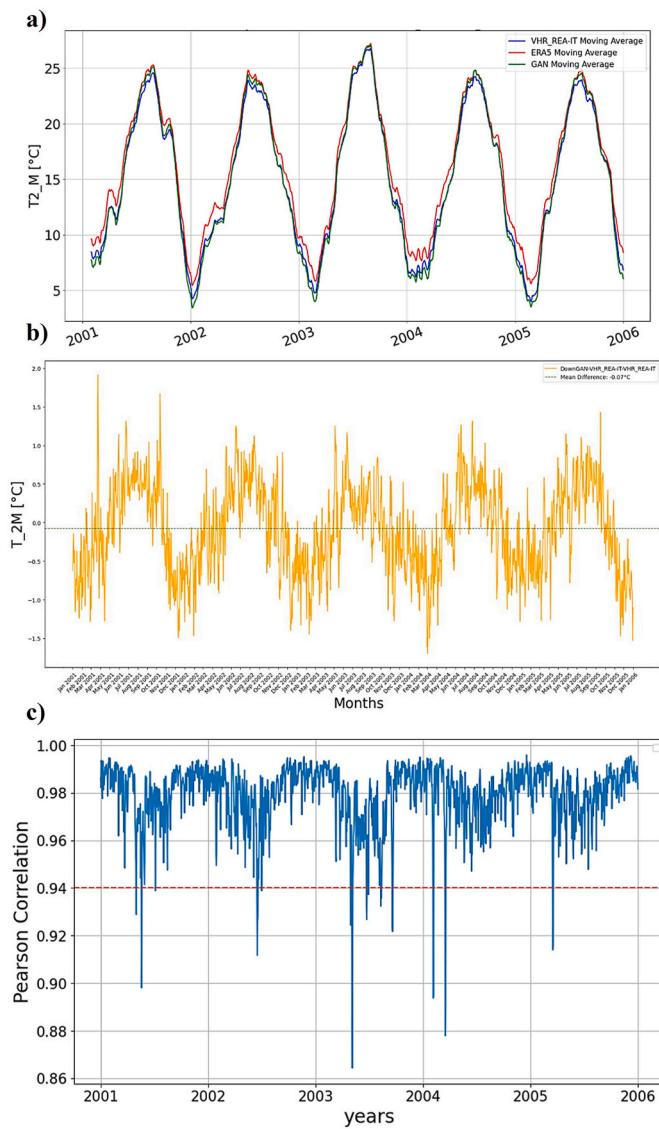


Fig. 4. Temporal evolution of 2m-temperature

Temporal evolution of 2m-temperature (T_{2M}) of: (a) Daily moving average of the 2m-temperature during the test period, calculated with a 30-day window. Mean daily spatial average of the real low-resolution data (ERA5), the real high-resolution data (VHR-REA_IT), and the downscaled artificial data (ERA5-DownGAN), created by the cGAN. (b) Daily differences between the dataset generated by the Conditional Generative Adversarial Network (cGAN), the ERA5-DownGAN dataset, and the real VHR-REA_IT dataset. (c) Daily correlation between ERA5-DownGAN and VHR-REA_IT.

revealed an overall average difference of -0.07°C across the entire domain. A maximum positive difference of 1.91°C was recorded on March 24, 2001, alongside a maximum negative difference of -1.70°C on March 2, 2004. According to the time plot of the correlation between the 2-m temperature fields from dynamic downscaling (VHR-REA_IT) and those generated by the cGAN during the testing period (Fig. 4c), a robust correlation between the datasets emerges, consistently exceeding 0.94. This suggests a substantial agreement between the predictions of the statistical model and the data from the dynamical model. Nevertheless, a more in-depth analysis reveals temporal variations. These fluctuations, exhibiting lower correlation values (as low as 0.87), may suggest the model's sensitivity in generating reliable downscaling during specific meteorological events, such as changes in atmospheric circulation or the spatial distribution of local phenomena. A future challenge will be to meticulously examine these variations to identify

underlying causes, including factors such as the sensitivity to the choice of certain parameters in the cGAN for specific atmospheric conditions, ultimately leading to the optimization of the architecture. While Fig. 4 clearly demonstrates that the ERA5-DownGAN-generated dataset is capable of accurately replicating both the values and the temporal pattern of the dynamic counterpart VHR-REA_IT, thus highlighting the cGAN's ability to reproduce the seasonal cycle, this result may be relatively straightforward to achieve with regression-based models. Therefore, to rigorously evaluate the reliability of the generated forecasts and the robustness of the developed model, the de-seasonalized correlation between the two high-resolution datasets was computed. In examining the de-seasonalized datasets, the cGAN once again demonstrates its ability to accurately reproduce the field produced by the dynamic counterpart, VHR-REA_IT, showing improved agreement in terms of

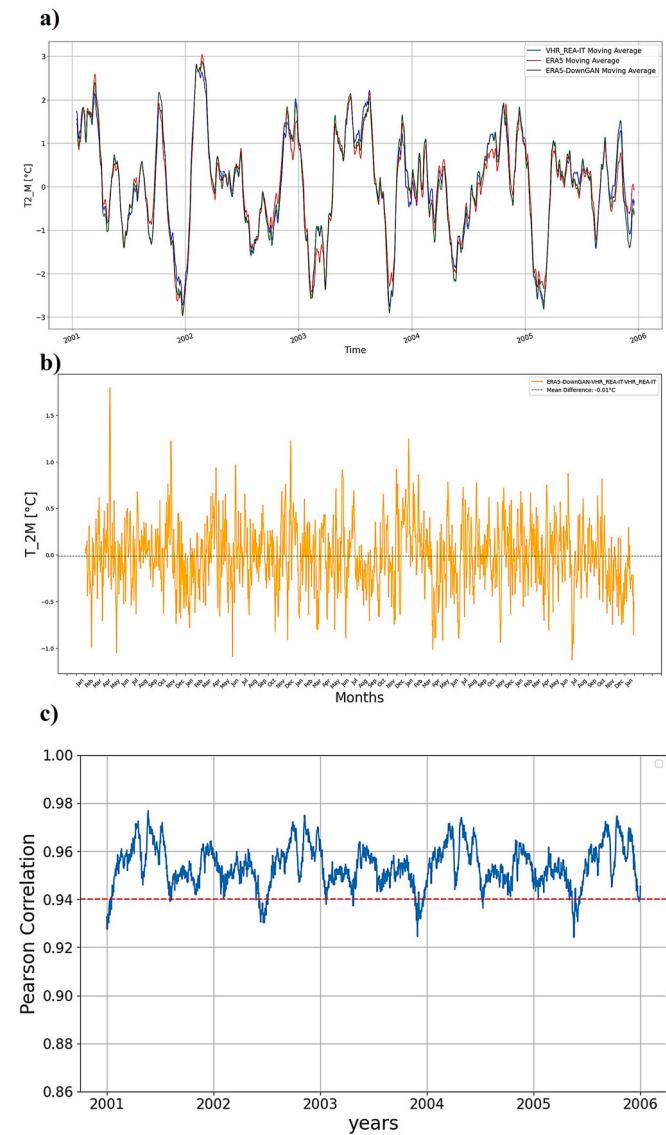


Fig. 5. Temporal evolution of 2m-temperature without seasonality

Temporal evolution of 2m-temperature (T_{2M}) of: (a) Daily moving average of the 2m-temperature during the test period, calculated with a 30-day window. Mean daily spatial average of the real low-resolution data (ERA5), the real high-resolution data (VHR-REA_IT), and the downscaled artificial data (ERA5-DownGAN), created by the cGAN. (b) Daily differences between the dataset generated by the Conditional Generative Adversarial Network (cGAN), the ERA5-DownGAN dataset, and the real VHR-REA_IT dataset. (c) Daily correlation between ERA5-DownGAN and VHR-REA_IT. The analysis is performed by removing the seasonality from the datasets considered.

values (Fig. 5b) compared to the lower-resolution ERA5 dataset. As shown in Fig. 5a, differences of up to approximately 1 °C are observed between the high-resolution datasets (ERA5_DownGAN, VHR-REA_IT) and the lower-resolution ERA5. The de-seasonalized correlation (Fig. 5c) shows an improvement over the seasonalized case (Fig. 4c), with an average value of around 0.95 and a minimum value of 0.93.

4.1. Error metrics evaluation

We assessed the performance of our cGAN using key error metrics: Bias, Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). These metrics provided valuable insights into the accuracy and consistency of the generated data compared to the high-resolution dataset during the testing period (Fig. 6).

- a. Bias: The bias (Eq. (1)), a measure of systematic error, was calculated to determine the average difference between the generated data and the high-resolution reanalysis. A lower bias indicates a more accurate representation of the target data.

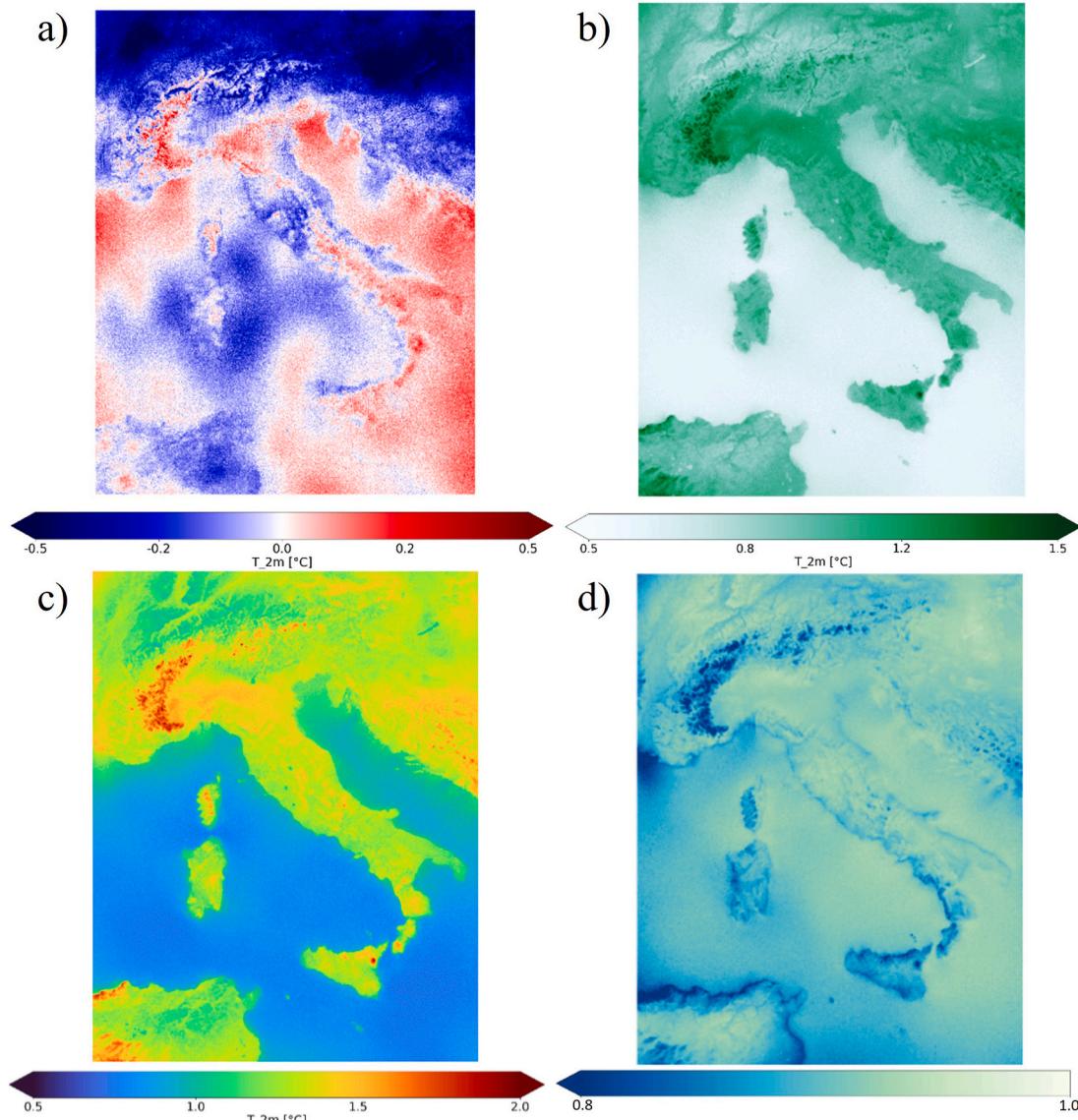


Fig. 6. Maps of error metrics

(a) Mean bias (BIAS), (b) root-mean-square errors (RMSEs), (c) mean absolute errors (MAEs), and (d) correlation coefficients (CORRs), comparing ERA5-DownGAN with VHR_REA_IT for the test period (2001–2005).

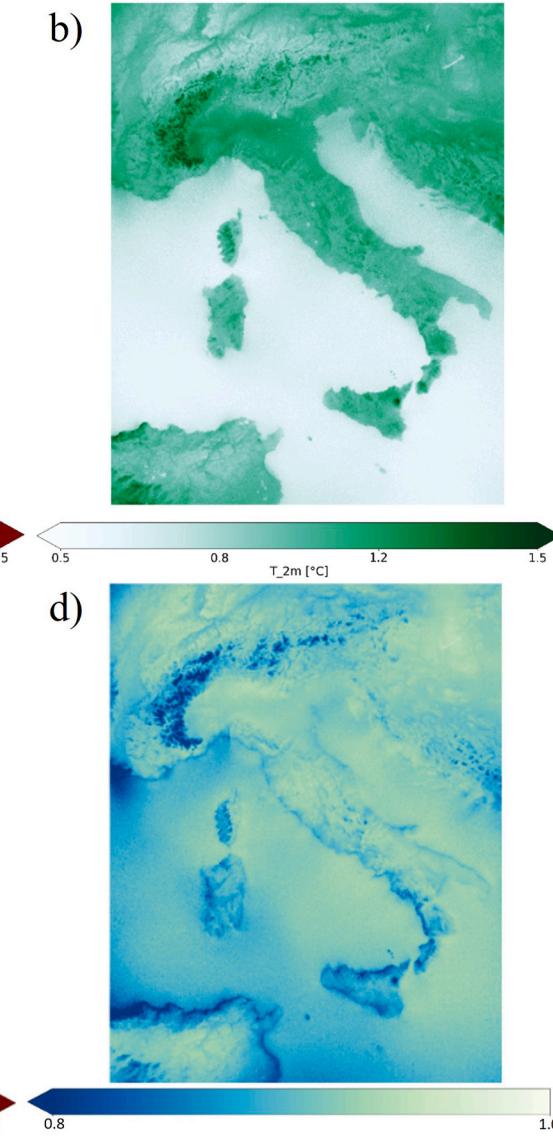
$$BIAS = x_{gen} - x_{real} \quad (1)$$

where x_{real} is the average value of the real dataset obtained from the mean calculation in time, x_{gen} is the corresponding mean value of the generated dataset.

- b. Mean Absolute Error (MAE): The MAE (Eq. (2)), a measure of absolute error, allowed us to quantify the average magnitude of errors between the generated data and the high-resolution dataset. Lower MAE values suggest closer resemblance and precision.

$$MAE = |x_{gen} - x_{real}| \quad (2)$$

- c. Root Mean Squared Error (RMSE): The RMSE (Eq. (3)), a measure of square-rooted error variance, provided insights into the overall discrepancy between the generated data and the high-resolution dataset. Lower RMSE values reflect improved accuracy and fidelity of the generated data.



$$RMSE = \sqrt{(x_{gen} - x_{real})^2} \quad (3)$$

d. Pearson Correlation: The Pearson correlation r (Eq. (4)) quantifies the degree of similarity, ranging from -1 to $+1$, with 0 indicating no linear correlation. This metric succinctly captures the GAN's ability to replicate patterns observed in the high-resolution dataset (VHR_REA_IT).

$$r = \frac{(\sum (x_{gen}, i - \bar{x}_{gen}) * (x_{real}, i - \bar{x}_{real}))}{\sqrt{\sum (x_{gen}, i - \bar{x}_{gen})^2 * \sum (x_{real}, i - \bar{x}_{real})^2}} \quad (4)$$

The evaluation metrics, including mean bias (BIAS), root-mean-square errors (RMSEs), mean absolute errors (MAEs), and correlation coefficients (CORRs), comparing downscaled daily 2m-temperatures (ERA5-DownGAN) with the high-resolution dataset (VHR_REA_IT) for the test period (01/2001-12/2005), are shown in Fig. 6. Specifically, major differences are observed in regions characterized by complex orography. Regarding BIAS, the range spans from -0.5° to $+0.5^\circ$, with a distinct negative peak centered over Central Europe within our investigation domain. Meanwhile, MAE and RMSE exhibit values ranging from 0.5° to 1.5° and 0.5° – 2° , respectively. Throughout the test period, the mean correlation consistently approaches 1 , with lower values observed near the African, Calabrian, and Sicilian coasts, as well as in proximity to the Alps and Etna volcano.

Table 2 summarizes the error metrics (BIAS, MAE, and RMSE), as well as the values of the Index of Agreement (IOA), Willmott, C. J., 1981, and the coefficient of determination (R^2), Legates and McCabe Jr (1999); Shiru and Chung (2021), which are essential for assessing the agreement between our cGAN-based model and the dynamic model. The metrics are calculated over the entire domain, for Italy (considering only land) and for the five regions in which Italy can be analyzed based on climatic differences, as reported in Raffa et al. (2021).

In line with the work of Sun et al. (2024), R^2 enables us to quantify the explanatory power of our cGAN model in relation to the variability observed in the dataset (specifically, the VHR-REA_IT reanalysis data). R^2 values typically range from 0 to 1 , with 1 indicating perfect agreement between the model and observed data, and values closer to 0 suggesting poor explanatory power. The results presented in Table 2 highlight the excellent performance of the cGAN model in reproducing 2-m temperatures across various regions of Italy. In general, the BIAS is minimal, with values close to zero in all areas; however, slight regional differences are observed. Specifically, the BIAS for Italy (land only) is -0.03° , while for Southern Italy it is positive ($+0.01^\circ$), suggesting that the model slightly underestimates temperatures in the North and overestimates them in the South. This trend is more pronounced in Central and Insular Italy, where the bias is -0.05° , indicating a slight underestimation compared to the observed data. In contrast, in the North-East and North-West regions, the bias is negative but very small (-0.01° and -0.05° , respectively), suggesting a closer alignment with the reference data. However, these small differences may be the

Table 2

Analysis of the error and performance metrics (BIAS, MAE, RMSE, IOA, and R^2) for the ERA5-DownGAN dataset produced by the cGAN, against the VHR-REA_IT dataset. The data are aggregated over the entire domain, throughout Italy, and are further subdivided into the five regions (reported in Fig. 1b in Raffa et al., 2021). Color coding is applied to facilitate the classification of differences.

Region	BIAS [°C]	MAE [°C]	RMSE [°C]	IOA	R^2
Domain	-0.07	0.86	1.11	0.996	0.985
Italy (Land Only)	-0.03	1.02	1.29	0.995	0.984
North-East Italy	-0.01	1.07	1.36	0.996	0.985
North-West Italy	-0.05	1.10	1.40	0.996	0.985
Central Italy	-0.05	1.05	1.32	0.995	0.983
South Italy	+0.01	0.96	1.21	0.995	0.983
Insular Italy	-0.05	0.91	1.16	0.993	0.977

result of compensatory effects, as revealed by the more distinct regional differences when examining RMSE and MAE.

The MAE (Mean Absolute Error) shows a similar pattern to the BIAS, with generally low values and no significant discrepancies between the regions. In particular, the Southern Italy and Insular Italy regions exhibit MAE values of 0.96° and 0.91° , respectively, suggesting that the model performs well in predicting average temperatures in these areas. In contrast, the North-East and North-West regions show slightly higher MAE values (1.07° and 1.10° , respectively), indicating a greater discrepancy between the model's predictions and the real data. This could be attributed to the higher variability in temperatures in these areas, which makes it more challenging for the model to capture local fluctuations accurately. Regarding RMSE, the regional differences are more pronounced. The lowest value is recorded for Southern Italy (1.21°), indicating that the model performs well in reproducing temperatures in this region. In contrast, the North-West and North-East regions exhibit the highest RMSE values (1.36° and 1.40° , respectively), suggesting greater discrepancies between the model and the observed data in these areas. This may be due to the higher local temperature variability, which is more difficult to capture with the model. Overall, although the cGAN model demonstrates excellent performance at the global level (with low bias, MAE, and RMSE), regional differences suggest that the northern and central areas, with their more marked climatic variations, are the most challenging to reproduce. The Index of Agreement (IOA) is extremely high across all regions (ranging from 0.993 to 0.996), indicating a strong degree of consistency between the cGAN model and the reference dataset. In particular, the model shows the best agreement for the overall domain and in the North-East, with IOA values of 0.996 . Finally, the coefficient of determination (R^2) confirms that the cGAN model explains a significant portion of the observed variability, with values ranging from 0.977 for Insular Italy to 0.985 for the North-East and North-West, suggesting good predictive capacity across all regions examined. These results underscore the reliability of the cGAN model in accurately reproducing 2-m temperatures in various geographical areas of Italy.

The results presented in Table 3 demonstrate a high level of accuracy in the reproduction of 2-m temperatures, particularly in terms of the mean and median values, by the cGAN model (ERA5-DownGAN). Discrepancies are observed in the recorded maximum and minimum values when compared to the VHR-REA_IT reference dataset, especially in Northern Italy. The cGAN model tends to underestimate the recorded maximum temperature and overestimate the minimum compared to the VHR-REA_IT data. This is particularly evident in the northern regions, likely due to the higher thermal variability in these areas, which presents a greater challenge for the model owing to complex local factors such as topography and the presence of more intense atmospheric phenomena.

Moreover, the difference between the models in the recorded maximum and minimum values is more pronounced across the entire domain, which also includes parts of Central Europe, as well as the sea. However, these discrepancies do not significantly affect the model's ability to represent general thermal conditions, as the mean and median values, along with the distributions in terms of PDFs and percentiles (Fig. C1 and Fig. C2, supplementary material), remain well aligned. The inclusion of the sea in the domain is a crucial factor, as the cGAN model was trained using both the VHR-REA_IT and ERA5 datasets. However, the former does not incorporate marine data, unlike its low-resolution counterpart. The absence of accurate high-resolution representation of marine thermal variability may influence the temperature simulation across the entire domain, leading to greater discrepancies in the cGAN model when compared to the reference dataset. However, when considering only the terrestrial areas of Italy, the difference between the model and the VHR-REA_IT dataset diminishes significantly, with a difference of approximately 1° in the simulated maximum temperature, while the discrepancy is more pronounced for the recorded minimum temperature.

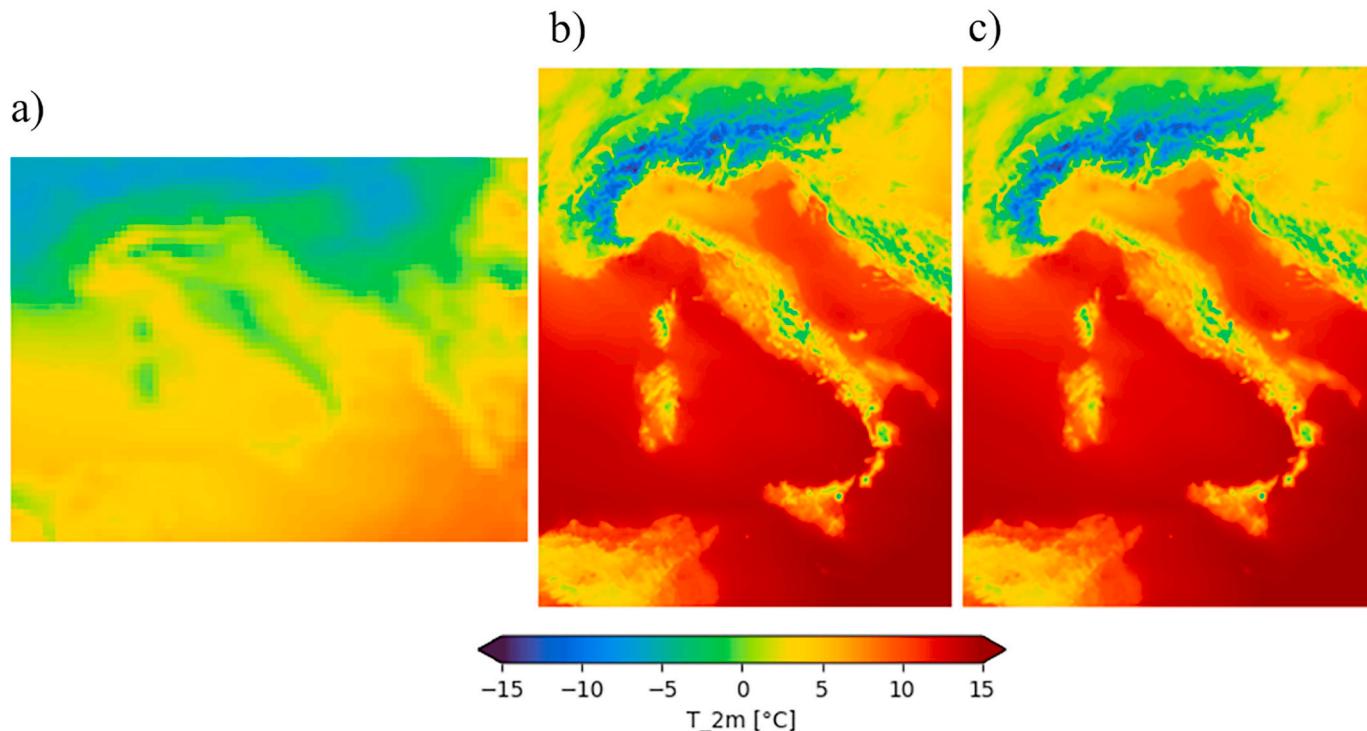
Table 3

Daily surface temperature analysis for the test period 2001–2005, as provided by VHR-REA_IT, ERA5-DownGAN, and ERA5. The data are aggregated over the entire domain, across Italy, and consider the five subregions (reported in Fig. 1b from Raffa et al., 2021). The table reports the maximum, minimum, mean, and median values, as well as the ratios of standard deviations ($\sigma_{\text{cGAN}}/\sigma_{\text{VHR}}$) and ($\sigma_{\text{ERA5}}/\sigma_{\text{VHR}}$), where σ_{cGAN} refers to the standard deviation of the ERA5-DownGAN dataset, σ_{VHR} corresponds to VHR-REA_IT, and σ_{ERA5} represents ERA5.

Region	Dataset	MAX [°C]	MIN [°C]	MEAN [°C]	MEDIAN [°C]	$\sigma_{\text{cGAN}}/\sigma_{\text{VHR}}$	$\sigma_{\text{ERA5}}/\sigma_{\text{VHR}}$
Domain	VHR-REA_IT	42.25	-30.81	14.98	15.42	1.05	0.88
	ERA5-DownGAN	38.40	-25.38	14.91	15.32		
	ERA5	41.74	-16.29	16.01	16.12		
Italy (Land Only)	VHR-REA_IT	36.66	-27.55	13.98	14.14	1.06	0.79
	ERA5-DownGAN	35.48	-24.38	13.95	13.99		
	ERA5	36.01	5.84	16.09	15.93		
North-East Italy	VHR-REA_IT	36.53	-24.46	13.18	12.97	1.05	0.76
	ERA5-DownGAN	35.48	-21.59	13.17	12.92		
	ERA5	33.34	-5.00	14.71	14.59		
North-West Italy	VHR-REA_IT	36.59	-27.55	11.08	10.62	1.05	0.72
	ERA5-DownGAN	35.44	-24.38	11.03	10.62		
	ERA5	35.02	-5.84	14.96	14.79		
Central Italy	VHR-REA_IT	35.37	-16.33	13.94	13.90	1.07	0.82
	ERA5-DownGAN	34.34	-16.28	13.89	13.74		
	ERA5	34.56	-4.19	15.90	15.74		
South Italy	VHR-REA_IT	36.66	-20.44	15.04	15.18	1.06	0.85
	ERA5-DownGAN	34.61	-18.60	15.05	15.05		
	ERA5	35.19	-3.76	16.85	16.64		
Insular Italy	VHR-REA_IT	36.24	-9.66	16.65	16.54	1.05	0.92
	ERA5-DownGAN	33.56	-10.68	16.61	16.39		
	ERA5	36.01	2.45	17.70	17.39		

It is also important to highlight that the discrepancies in the maximum and minimum values may stem from individual points that lie outside the statistical norm. While the mean, median, and percentile values are well aligned between the cGAN model and the reference dataset, the maximum and minimum values are more susceptible to being influenced by particularly anomalous local conditions. Despite these variations, the cGAN model remains highly effective in

representing general thermal conditions, providing a dataset that aligns closely with that produced by the dynamic model. This is further confirmed by the analysis of standard deviation (σ): the ratio between the standard deviation of the cGAN model (σ_{cGAN}) and that of the VHR-REA_IT dataset is on average very close to 1 (with a variation between 1.05 and 1.07), indicating that the cGAN model is able to capture the variability of temperatures simulated by the dynamic model with high

**Fig. 7. Power reconstruction with cGAN**

Examples of one day of 2m-temperature (T_{2M}) from the test data set. (a) Real low-resolution data (ERA5), (b) real high-resolution data (VHR-REA_IT), and (c) the downsampled artificial data (ERA5-DownGAN), created by the cGAN.

precision. This suggests that the model not only accurately reproduces the thermal means, but also preserves the spatial variability of the reference data, a crucial feature for a statistical downscaling model. In contrast, the ERA5 model shows a ratio of $\sigma_{\text{ERA5}}/\sigma_{\text{VHR}}$ lower than 1, indicating an underestimation of local variability relative to the reference dataset, highlighting the limited representation of local variability in ERA5 due to its low resolution.

4.2. Spatial and temporal consistency

The comparison among real low-resolution (58×80 grid points), real high-resolution (680×535 grid points), and downscaled ERA5-DownGAN (680×535 grid points) high-resolution test datasets are illustrated in Fig. 7. This maps of 2m-temperature field over the entire computational domain for one random day of the test period. ERA5-DownGAN was generated utilizing the generator saved after 100 training epochs. The decision to select the generator after this specific number of epochs is based on the observed decrease in the generator's error beyond this designated epoch count. Simultaneously, the artificially high-resolution data (ERA5-DownGAN) captures the 2m-temperature pattern observed in real high-resolution data (VHR-REA_IT), especially the intricate structures along the Alps and Apennines, achieving a remarkable level of resolution with all topographical details prominently visible.

Furthermore, the cGAN output (ERA5-DownGAN) effectively mitigates ERA5's tendency to generate a field colder than VHR-REA_IT, maintaining values very close to real high-resolution data with slight underestimation, resulting in an overall slightly cooler field. According to Fig. 8, this suggests a higher correlation between the new high-resolution dataset and the real high-resolution dataset compared to the low-resolution dataset. The results clearly show that both the spatial distributions and magnitudes of the downscaled 2m-temperature (ERA5-DownGAN) are statistically like those of the real high-resolution dataset (VHR-REA_IT). In Fig. 9, the seasonal 2m-temperature (T_2M) maps show that the dataset generated by the cGAN, ERA5-DownGAN, accurately reproduces the spatial patterns and values observed in the high-resolution VHR-REA_IT dataset. The results indicate a strong ability of the generative model to capture both spatial structures and seasonal variations, with minimal discrepancies from the real data. However, a slight general underestimation of values generated by the cGAN is evident, particularly in regions of Central Europe, such as between Austria and Hungary, and in some areas of northwestern Italy.

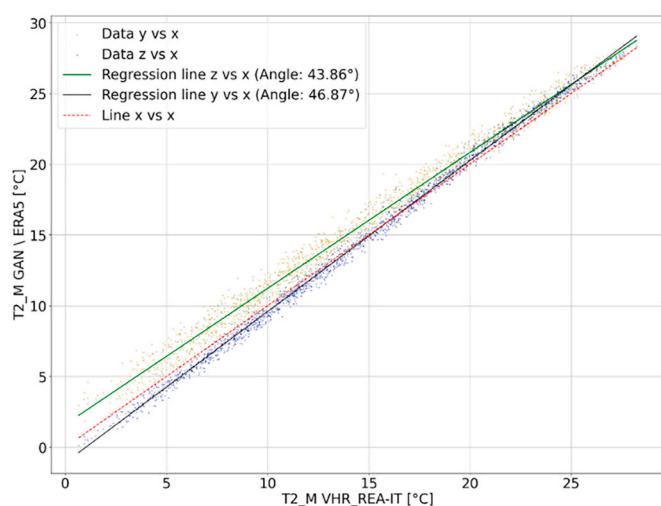


Fig. 8. Scatter diagram

Scatter diagram for evaluating the correlation between 2m-temperature (T_2M) of the ERA5, VHR-REA_IT, and ERA5-DownGAN datasets.

4.3. cGAN's ability to capture extreme values

The assessment of percentiles ranging from the 1st to the 99th in Fig. 10 facilitates an understanding of the extent to which the generated data aligns with the distribution characteristics of the high-resolution dataset. The comparison of percentile values proved valuable in evaluating the cGAN's ability to accurately capture extreme events and the overall distribution of the data. For the creation of the boxplots in Fig. 10, a rigorous and detailed methodology was implemented to capture the spatiotemporal variability complexity of the 2m-temperature. Specifically, the boxplots were derived by considering all values obtained from calculating a single value for each specific percentile (from 1st to 99th) across the entire geographical domain.

The detailed analysis of the boxplots in the results highlights significant differences in the distribution of 2m-temperature between the low-resolution real dataset (ERA5) and the two high-resolution datasets, real (VHR-REA_IT), and generated through cGAN (ERA5-DownGAN). Overall, ERA5 consistently tends to exhibit a median and distribution shifted towards higher temperature values compared to the other two datasets for all analyzed percentiles, even with notably higher positive extremes for lower percentiles (1st and 5th; from 95th to 99th) and notably lower negative extremes for percentiles from 1st to 25th. Furthermore, a wider dispersion is observed, clearly represented by the greater extension of the whiskers compared to the other two datasets, especially with the upper whisker extending more than the lower one for percentiles above the 95th; whereas a lesser extension of the whiskers is observed for the 10th and 25th percentiles. This indicates a tendency of ERA5 to predict higher temperatures compared to the other datasets (for both upper and lower extremes). For the medium percentiles, ERA5 records a median and dispersion similar to those of the other datasets. In the comparison between the results obtained from the cGAN and VHR-REA_IT, significant similarities emerge in the overall distribution, with the cGAN accurately replicating the median and whisker extension of VHR-REA_IT. However, a slight shift is observed in the limits of the upper and lower whiskers towards lower values compared to VHR-REA_IT. This may suggest that the cGAN has introduced a slight systematic discrepancy in the predicted values relative to the high-resolution dataset, shifting the extreme values towards lower values (for both upper and lower extremes). This analysis, comparing the cGAN-generated data (ERA5-DownGAN) with the real high-resolution dataset (VHR-REA_IT) for the test period, demonstrates promising results in terms of accuracy and consistency, reproducing a temperature field characterized by the same distribution, median, and approximately equal extreme values.

The analysis of the Probability Density Functions (PDFs) of the three 2m temperature datasets (Fig. 11), ERA5, VHR-REA_IT, and ERA5-DownGAN, reveals a similarity in the overall shape of the curves, indicating consistency in the distribution of temperatures. However, an interesting observation arises in the ERA5 dataset, which displays a slight modulation in the right tail of the curve, a phenomenon not observed in the other two datasets. This suggests a specific characteristic in the distribution of temperatures for ERA5. Additionally, the persistent tendency of ERA5 to be shifted towards higher values is confirmed by this analysis, with a range between 10 °C and 25 °C and a significantly lower variability, which is obviously due to the different grid point numbers resulting from the different resolutions. On the other hand, the distribution of the cGAN mirrors a range similar to that of VHR-REA_IT, ranging from -10 °C to 25 °C. The overlapping curves of ERA5-DownGAN and VHR-REA_IT at many points highlight the cGAN's ability to faithfully replicate the high-resolution temperature distribution. However, it is important to note that the cGAN's curve exhibits a slightly lower primary peak compared to VHR, centered around 19 °C, and a secondary peak around 13 °C, which is slightly lower and shifted to the left compared to VHR-REA_IT, settling around 12 °C. These details unveil subtle differences in the reproduction of temperature peaks by the cGAN compared to the high-resolution dataset.

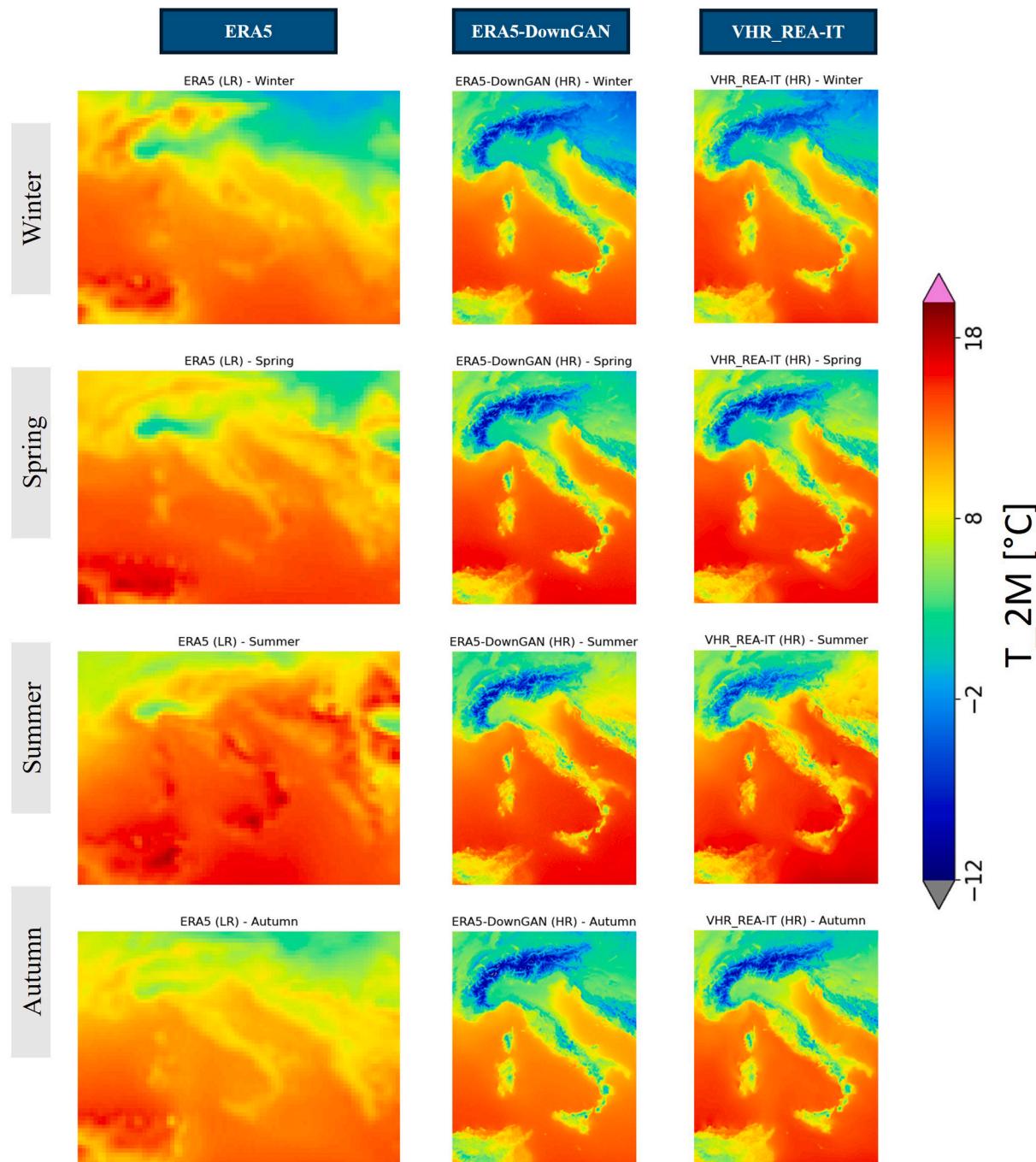


Fig. 9. Seasonal comparison of 2m-temperature

Comparison of seasonal 2m-temperature (T_{2M}) for real low-resolution data (ERA5), the real high-resolution data (VHR-REA_IT), and the downscaled artificial data (ERA5-DownGAN), created by the cGAN.

Following Sperati et al. (2017), a quantitative assessment of spatial structure using a scatter plot that illustrates the correlation between all pairs of grid points from both the real dataset (VHR-REA_IT) and the dataset generated by cGAN (ERA5-DownGAN) is conducted. This diagram was computed for the entire test period, employing a color-encoded kernel density estimation technique (Venables and Ripley, 2002) to enhance visualization, as depicted in Fig. 12. The y-axis in the figure represents the average correlation between all pairs of grid points in the real dataset (VHR-REA_IT). On the other hand, the x-axis is determined by computing the average correlation between all pairs of grid points in the cGAN dataset (ERA5-DownGAN). The scatter plot reveals that the artificially generated high-resolution dataset by the

cGAN closely mirrors the spatial correlation structure of the reference real high-resolution dataset. The data points follow the 1:1 line, and there is a larger dispersion around it, ranging from -0.25 to $+0.25$. The higher density is observed from 0.5 to 1 , indicating a predominant positive correlation between grid points in both high-resolution datasets. Focusing on specific areas of interest in the plot, excluding regions with negligible correlation, more significant features can be identified. In the range from -0.75 to -0.25 (Fig. 12, at the top left), corresponding to the anti-correlation area between grid points, the cGAN-produced dataset exhibits a stronger negative correlation compared to the dynamically downscaled dataset (VHR-REA_IT), with a higher density around -0.3 . Similarly, in Fig. 12 at the bottom right, which depicts

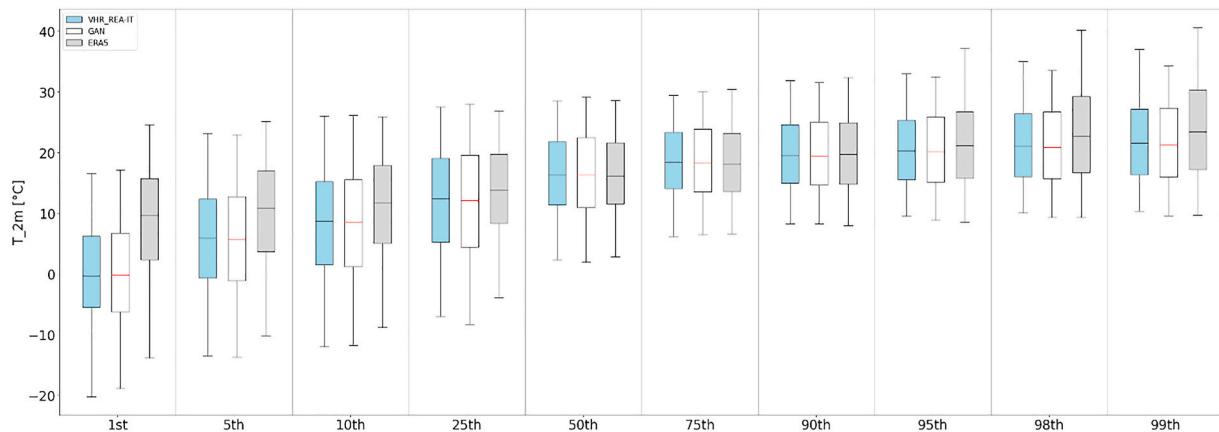


Fig. 10. Box-plot for percentile from 1st to the 99th

Comparison of the 2m-temperature (T_2M) distribution for real low-resolution data (ERA5) with the gray rectangle, the real high-resolution data (VHR-REA_IT) with the blue rectangle, and the downscaled artificial data (ERA5-DownGAN), created by the GAN with a white rectangle for each percentile from the 1st to the 99th.

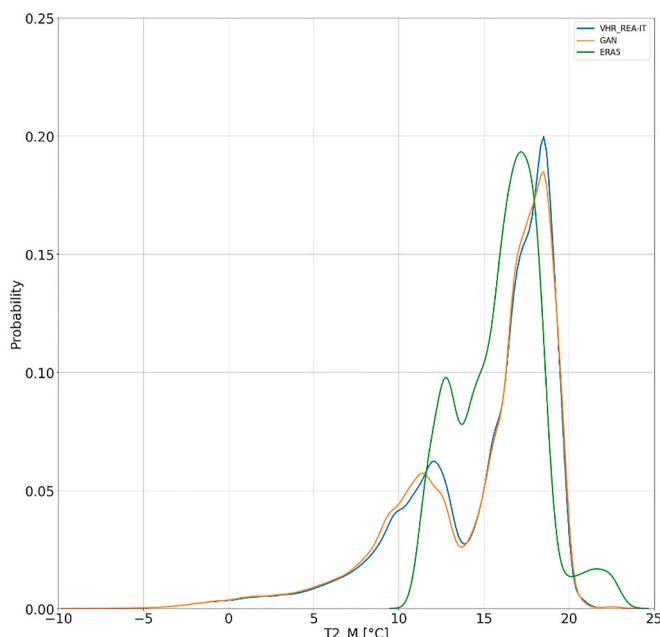


Fig. 11. Evaluation of PDFs

Comparison of the probability density functions (PDF) of the 2m-temperature (T_2M) distribution for the ERA5, VHR-REA_IT, and ERA5-DownGAN datasets.

spatial correlation values exceeding 0.8, a higher positive correlation is evident in the statistically downscaled dataset compared to the dynamically downscaled one. Finally, for values exceeding +0.95, there is a slight reversal of this pattern. Specifically, two areas with higher density can be identified, around +0.83 and +0.95. The observed spatial correlation trend in the cGAN implies a higher level of correlation among grid points, whether positively or negatively, compared to VHR-REA_IT. In practical terms, this suggests that the cGAN generates a field with more coherent temperature variations, exhibiting a more pronounced and structured pattern compared to the actual dataset, a characteristic often described as a smoother field. Positive correlations result in more similar temperature variations among neighboring points, contributing to a more homogeneous field. Conversely, negative correlations lead to opposite temperature variations among nearby points, creating a field with more distinct variations. In summary, our results suggest that the cGAN introduces a heightened level of coherence and structural organization in the generated data compared to dynamical

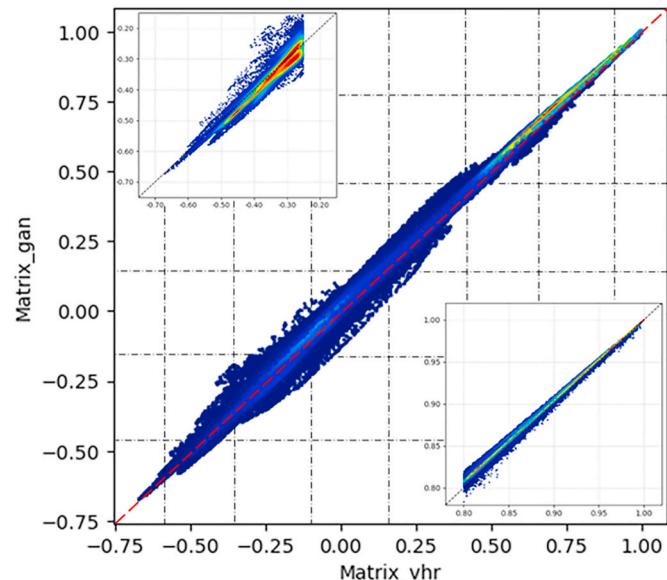


Fig. 12. Spatial correlation

Scatter diagram for the test period comparing the correlation of all pairs of grid points of the real dataset (VHR-REA_IT) with the correlation of all pairs of grid points of the synthetic dataset by cGAN (ERA5-DownGAN).

downscaling. Nevertheless, these distinctions are minor when contrasted with the overarching pattern of the cGAN-generated high-resolution dataset (ERA5-DownGAN), which demonstrates a correlation between points closely mirroring that observed in the actual VHR-REA_IT dataset.

4.4. Computational details

Computational results achieved through the implementation of the cGAN bear paramount significance in the landscape of climate downscaling. All simulations were performed on Zeus supercomputer (<http://www.cmcc.it/research/super-computing-center>), hosted at the CMCC (Euro-Mediterranean Centre on Climate Change), equipped with Intel(R) Xeon(R) Gold 6130 CPU 2.10 GHz, with 36 physical cores. This computational resource facilitated the tackling of ambitious challenges concerning scalability and computational complexity.

In the context of this research, it is crucial to highlight that the implementation of the cGAN was carried out sequentially following a

serial approach. The execution of the code was performed using 72 cores in the computational process. One salient aspect of our computational findings lies in the remarkable efficiency of the cGAN-based downscaling approach. The run-time for the training phase was approximately 4 h to simulate 10 years (**Table 4a**), and about 2.5 min to simulate 5 years during the testing phase (**Table 4b**). This achievement assumes extraordinary importance when compared to the substantially protracted timescales required by conventional dynamical downscaling methodologies (**Table 5**). This computational prowess enables fast generation of high-resolution climate predictions, endowing us with a considerable advantage in terms of timeliness for climate change analysis and response planning. The utilization of cutting-edge computational resources and the optimization of the cGAN-based downscaling framework signify a significant leap forward within the scientific climate community. They underscore the fact that innovative computational solutions can profoundly enhance our capacity to comprehend and proactively manage the intricacies of climate change.

The training time, indicated in **Table 4a**, doubles when considering the checks performed to evaluate the training at regular intervals, which involve plotting the ERA5 (LR), VHR-REA_IT (HR), and Era5-DownGAN (HR) maps, as well as the error metric maps.

5. Conclusions

This study assesses the capability of the cGAN model to perform statistical downscaling of 2m-temperature from $\simeq 31$ km to $\simeq 2.2$ km. Some studies have explored deep learning techniques in the climatological context, especially regarding statistical downscaling, but an increasing number of researchers are focusing on GAN due to its promise in maintaining spatial coherence between fields and the initial promising results it has shown. There are several innovative aspects of this study. It could represent a fundamental and flexible cGAN architecture for future several practical applications of AI-based statistical downscaling, enabling the generation of high and very high-resolution datasets across different regions of the world, starting from coarser datasets and requiring significantly reduced computational resources compared to the dynamical counterpart. Based on the promising results, the cGAN approach shows a great option for achieving a downscaling with high performance while also reducing the computational resources involved in the process. It leverages real datasets in both the training phase for high-resolution and its low-resolution counterpart, allowing for the establishment of empirical connections between the two datasets and maintaining the correlation between the two fields. This approach offers a significant advantage, more easily highlighting the existing empirical links between the high-resolution and low-resolution datasets. All of this occurs with computational efficiency, positioning the GAN as a viable alternative to dynamic models in the downscaling process. In terms of the obtained results, the developed cGAN demonstrates the ability to replicate patterns and median values almost identical to those simulated by the dynamic model, except for a tendency of this statistical downscaling to produce a slightly cooler field than the dynamic counterpart. A thorough analysis of extremes shows that the cGAN is capable of capturing extreme values, reflecting an approximately identical distribution in terms of median and dispersion to that produced by the dynamic model. Despite the promising results achieved, the complexity of this architecture opens the door to multiple possible improvements that can be made by acting on both the involved datasets, considering, for example, training windows in cross mode, and on the complexity of the architecture itself. Among the future objectives outlined by this study is the intention to test this developed architecture on other atmospheric variables, such as the wind field and precipitation field, to understand the versatility of the architecture. Moreover, it would be interesting to assess the performance of the cGAN and the quality of the results by replacing the linear layers used with convolutional layers. A convolutional architecture enables faster training and widespread model application (Park, S., et Shin, Y.-G., 2022), making the cGAN suitable for

Table 4

Computational details for the training period (a) and test period (b) on Zeus Supercomputer.

a)					
years	lr	samples	batch size	epochs	time/training
10	0.0001	4000	100	100	~ 4 h
b)					
years	samples		batch size	time/test	
5	2000		100	~ 2.5 min	

Table 5

Comparison of computational requirements for dynamical and cGAN models.

Model (HR Dataset)	Type	Cores	Core-Hours/Year
COSMO-CLM (VHR-REA_IT)	Dynamical	2160	$\sim 131,760$
cGAN (ERA5-DownGAN)	Statistical	36	Training: ~ 216 Test: ~ 22.5

downscaling over large geographic areas. Additionally, convolutional layers introduce translation invariance, allowing the model to recognize patterns regardless of their precise position in the data matrix. This is advantageous in meteorological forecasting, where the specific location of phenomena may vary. In conclusion, the cGAN emerges as a potential technique for generating climate data at significantly higher resolutions, paving the way for an improvement in climate modeling and downscaling applications.

CRediT authorship contribution statement

Ilenia Manco: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Walter Riviera:** Software. **Andrea Zanetti:** Software. **Marco Briscolini:** Supervision, Funding acquisition. **Paola Mercogliano:** Writing – review & editing, Supervision, Investigation, Funding acquisition. **Antonio Navarra:** Writing – review & editing, Funding acquisition, Conceptualization.

Funding sources

This work was financially supported by ICSC – Centro Nazionale di Ricerca in High Performance Computing, Big Data and Quantum Computing, funded by the European Union – NextGenerationEU.

Funding

This work was supported by the University of Bologna; E4 Computer Engineering SpA; and CMCC Foundation - Euro-Mediterranean Center on Climate Change.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Ilenia Manco reports financial support was provided by CMCC Foundation - Euro-Mediterranean Center on Climate Change, Italy. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The entire simulation work was carried out using the computational

resources of the CMCC (Euro-Mediterranean Centre on Climate Change) Foundation. The authors gratefully acknowledge the technical assistance provided by the supercomputer user support team. The author also wishes to acknowledge the University of Bologna and E4 Computer Engineering for their support in the development of this research.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.envsoft.2025.106427>.

Glossary

GANs	Generative Adversarial Neural Networks
CGANs	Conditional Generative Adversarial Neural Networks
DD	Dynamical Downscaling
ESD	Empirical Statistical Downscaling
ANNs	Artificial Neural Networks
RNNs	Recurrent Neural Networks
CNNs	Convolutional Neural Networks
AI	Artificial Intelligence
ML	Machine Learning
DL	Deep Learning
GCMs	General Circulation Models (or Global Climate Models)
RCMs	Regional Climate Models
MOS	Model Output Statistics
PP	Perfect Prognosis
WG	Weather Generators
ERA5	European Centre for Medium-Range Weather Forecasts Reanalysis 5
VHR-REA_IT	Very High-Resolution Reanalysis for Italy
ERA5-DownGAN	Dataset generated in this work by cGAN Downscaling Model
d_step	discriminator step
lr	learning rate
T_2M	2meter-temperature
TOT_PREC	Total precipitation

Data availability

The code developed in this study will be fully released and assigned a DOI on the CMCC's DDS (Data Distribution Service) to ensure the reproducibility of the experiments.

References

- Baño-Medina, J., Manzanas, R., Gutiérrez, J.M., 2020. Configuration and intercomparison of deep learning neural models for statistical downscaling. *Geosci. Model Dev. (GMD)* 13, 2109–2124. <https://doi.org/10.5194/gmd-13-2109-2020>.
- Bittner, M., Hobeichi, S., Zawish, M., Diatta, S., Ozioko, R., Xu, S., Jantsch, A., 2023. An LSTM-based downscaling framework for Australian precipitation projections. *NeurIPS 2023 Workshop on Tackling Climate Change with Machine Learning*. NeurIPS, New Orleans, LA. <https://www.climatechange.ai/papers/neurips2023/46>.
- Boé, J., Mass, A., Deman, J., 2023. A simple hybrid statistical-dynamical downscaling method for emulating regional climate models over Western Europe. Evaluation, application, and role of added value? *Clim. Dyn.* 61, 271–294. <https://doi.org/10.1007/s00382-022-06552-2>.
- Dayon, G., Boé, J., Martin, E., 2015. Transferability in the future climate of a statistical downscaling method for precipitation in France. *J. Geophys. Res. Atmos.* 120 (3), 1023–1043. <https://doi.org/10.1002/2014JD022236>.
- Doury, A., Somot, S., Gadat, S., Ribes, A., Corre, L., 2022. Regional climate model emulator based on deep learning: concept and first evaluation of a novel hybrid downscaling approach. *Clim. Dyn.* 60 (5–6), 1751–1779. <https://doi.org/10.1007/s00382-022-06343-9>.
- Dubey, A.K., Jain, V., 2019. Comparative study of convolution neural network's relu and leaky-relu activation functions. In: Mishra, S., Sood, Y., Tomar, A. (Eds.), *Applications of Computing, Automation and Wireless Systems in Electrical Engineering*, Lecture Notes in Electrical Engineering, vol. 553. Springer, Singapore. https://doi.org/10.1007/978-981-13-6772-4_76.
- Erlandsen, H.B., Parding, K.M., Benestad, R., Mezghani, A., Pontoppidan, M., 2020. A hybrid downscaling approach for future temperature and precipitation change.
- J. Appl. Meteorol. Climatol. 59 (11), 1793–1807. <https://doi.org/10.1175/JAMC-D-20-0013.1>.
- Giorgi, F., Jones, C., Asrar, G.R., 2009. Addressing climate information needs at the regional level: the CORDEX framework. *World Meteorol. Organ. Bull.* 58, 175–183.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial networks. In: *Advances in Neural Information Processing Systems*, vol. 3, pp. 1–9. <https://doi.org/10.1145/3426222>.
- He, K., Zhang, X., Ren, S., Sun, J., 2015. Deep residual learning for image recognition. *arXiv preprint arXiv:1512.03385*.
- Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., Thépaut, J.-N., 2023. ERA5 hourly data on single levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). <https://doi.org/10.24381/cds.adbb2d47>.
- Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.R., 2012. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*.
- Hobeichi, S., Nishant, N., Shao, Y., Abramowitz, G., Pitman, A., Sherwood, S., Bishop, C., Green, S., 2023. Using machine learning to cut the cost of dynamical downscaling. *Earth Future* 11, e2022EF003291. <https://doi.org/10.1029/2022EF003291>.
- Holden, P.B., Edwards, N.R., Garthwaite, P.H., Wilkinson, R.D., 2015. Emulation and interpretation of high-dimensional climate model outputs. *J. Appl. Stat.* 42, 2038–2055. <https://doi.org/10.1080/02664763.2015.1016412>.
- Hutengs, C., Vohland, M., 2016. Downscaling land surface temperatures at regional scales with random forest regression. *Remote Sens. Environ.* 178, 127–141. <https://doi.org/10.1016/j.rse.2016.03.006>.
- Klein, W.H., Glahn, H.R., 1974. Forecasting local weather by means of model output statistics. *Bull. Am. Meteorol. Soc.* 55 (10), 1217–1224.
- Komasi, M., Alizadehfar, A., Ahmadi, M., 2024. Examination of players' strategies in determining the optimal groundwater exploitation by game theory. *Hydrogeol. J.* 32, 691–704. <https://doi.org/10.1007/s10040-024-02770-6>.
- Kondrup, C., Mercogliano, P., Bosello, F., Mysiak, J., Scoccimarro, E., Rizzo, A., Ebrey, R., de Ruiter, M., Jeukens, A., Watkiss, P., 2022. Springer Climate. <https://doi.org/10.1007/978-3-030-86211-4>. ISSN 2352-0698 | ISBN 978-3-030-86210-7.
- Legates, D.R., McCabe Jr., G.J., 1999. Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. *Water Resour. Res.* 35, 233–241.
- Leinonen, J., Nerini, D., Berne, A., 2021. Stochastic super-resolution for downscaling time-evolving atmospheric fields with a generative adversarial network. *IEEE Trans. Geosci. Rem. Sens.* 59 (9), 7211–7223. <https://doi.org/10.1109/TGRS.2020.3032790>.
- Limon, G., Jablonowski, C., 2023. Probing the skill of random forest emulators for physical parameterizations via a hierarchy of simple CAM6 configurations. *J. Adv. Model. Earth Syst.* 15, e2022MS003395. <https://doi.org/10.1029/2022MS003395>.
- Maraun, D., Widmann, M., 2018. Statistical Downscaling and Bias Correction for Climate Research. Cambridge University Press. <https://doi.org/10.1017/9781107588783>.
- Maraun, D., Shepherd, T., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J., Hagemann, S., Richter, I., Soares, P.M.M., Hall, A., Mearns, L., 2017. Towards process-informed bias correction of climate change simulations. *Nat. Clim. Change* 7, 362–370. <https://doi.org/10.1038/nclimate3418>.
- Mardani, M., Brenowitz, N., Cohen, Y., Pathak, J., Chen, C.-Y., Liu, C.-C., Vahdat, A., Kashinath, K., Kautz, J., Pritchard, M., 2023. Generative residual diffusion modeling for km-scale atmospheric downscaling. *arXiv*. <https://doi.org/10.48550/arXiv.2309.15214>.
- Mirza, M., Osindero, S., 2014. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*.
- Nishant, N., Hobeichi, S., Sherwood, S., Abramowitz, G., Shao, Y., Bishop, C., Pitman, A., 2023. Comparison of a novel machine learning approach with dynamical downscaling for Australian precipitation. *Environ. Res. Lett.* 18, 094006. <https://doi.org/10.1088/1748-9326/ace463>.
- Park, S., Shin, Y.-G., 2022. Generative convolution layer for image generation. *Neural Netw.* 152, 370–379. <https://doi.org/10.1016/j.neunet.2022.05.006>.
- Patro, S., Gopal, Sahu, Kumar, Dr-Kishore, 2015. Normalization: a preprocessing stage. *IARJSET*. <https://doi.org/10.17148/IARJSET.2015.2305>.
- Perkins-Kirkpatrick, S.E., Gibson, P.B., 2017. Changes in regional heatwave characteristics as a function of increasing global temperature. *Sci. Rep.* 7. <https://doi.org/10.1038/s41598-017-12520-2>. Article 12256.
- Raffa, M., Reder, A., Marras, G.F., Mancini, M., Scipione, G., Santini, M., Mercogliano, P., 2021. VHR-REA_IT dataset: very high resolution dynamical downscaling of ERA5 reanalysis over Italy by COSMO-CLM. *Data* 6 (8), 88. <https://doi.org/10.3390/data6080088>.
- Rampal, N., Hobeichi, S., Gibson, P., Medina, J., Abramowitz, G., Beucler, T., González-Abad, J., Chapman, W., Harder, P., Gutiérrez, J., 2024. Enhancing regional climate downscaling through advances in machine learning. *Artificial Intelligence for the Earth Systems* 3. <https://doi.org/10.1175/AIES-D-23-0066.1>.
- Rummukainen, M., 2010. State-of-the-Art with regional climate models. *Clim. Change* 1, 96. <https://doi.org/10.1002/wcc.8>, 82.
- Schwingshackl, C., Daloz, A.S., Iles, C., Aunan, K., Sillmann, J., 2024. High-resolution projections of ambient heat for major European cities using different heat metrics. *Nat. Hazards Earth Syst. Sci.* 24 (1), 331–354. <https://doi.org/10.5194/nhess-24-331-2024>.
- Sharghi, A., Komasi, M., Ahmadi, M., 2025. Variable sensitivity analysis in groundwater level projections under climate change adopting a hybrid machine learning algorithm. *Environ. Model. Software* 183, 106264. <https://doi.org/10.1016/j.envsoft.2025.106264>.

- Sharifi, E., Saghfian, B., Steinacker, R., 2019. Downscaling satellite precipitation estimates with multiple linear regression, artificial neural networks, and spline interpolation techniques. *J. Geophys. Res. Atmos.* 124, 789–805. <https://doi.org/10.1029/2018JD028795>.
- Shiru, M.S., Chung, E.S., 2021. Performance Evaluation of CMIP6 Global Climate Models for Selecting Models for Climate Projection over Nigeria. Springer, pp. 599–615. <https://doi.org/10.1007/s00704-021-03746-2>.
- Sperati, S., Alessandrini, S., Delle Monache, L., 2017. Gridded probabilistic weather forecasts with an analog ensemble. *Quart. J. Roy. Meteor. Soc.* 143, 2874–2885. <https://doi.org/10.1002/qj.3137>.
- Shrudevi, C., Routray, A., Ramarao, M.V.S., Dutta, S., Hari Prasad, K.B.R.R., Colón, E., Gibbs, A., Pondeca, M., Prasad, V.S., 2024. Study of heat wave using high-resolution real-time meso-scale analysis over India. *Geophys. Res. Lett.* <https://doi.org/10.1029/2024GL109310>. First published. (Accessed 5 September 2024).
- Sun, Q., Fang, J., Dang, X., Xu, K., Fang, Y., Li, X., Liu, M., 2022. Multi-scenario urban flood risk assessment by integrating future land use change models and hydrodynamic models. *Nat. Hazards Earth Syst. Sci.* 22, 3815–3829. <https://doi.org/10.5194/nhess-22-3815-2022>.
- Sun, T., Yan, N., Zhu, W., Zhuang, Q., 2024. Assessing a machine learning-based downscaling framework for obtaining 1km daily precipitation from GPM data. *Heliyon* 10 (17), e36368. <https://doi.org/10.1016/j.heliyon.2024.e36368>.
- Tabor, K., Williams, J.W., 2010. Globally downscaled climate projections for assessing the conservation impacts of climate change. *Ecol. Appl.* 20 (2), 554–565. <http://www.jstor.org/stable/27797827>.
- Tiwari, P., Kar, S., Mohanty, U.C., Dey, S., Sinha, P., Shekhar, M., Sokhi, R., 2018. Comparison of statistical and dynamical downscaling methods for seasonal-scale winter precipitation predictions over north India. *Int. J. Climatol.* 39 (1), 240–258. <https://doi.org/10.1002/joc.5897>.
- Vandal, T., Kodra, E., Ganguly, S., Michaelis, A., Nemani, R., Ganguly, A.R., 2017. DeepSD: generating high resolution climate change projections through single image super-resolution. *arXiv* 1703, 03126v1. <https://doi.org/10.48550/arXiv.1703.03126>.
- Venables, W.N., Ripley, B.D., 2002. *Modern Applied Statistics with S*, 4th ed.
- Wilby, R.L., Wigley, T.M.L., 1997. Downscaling general circulation model output: a review of methods and limitations. *Prog. Phys. Geogr. Earth Environ.* 21 (4), 530–548. <https://doi.org/10.1177/030913339702100403>.
- Wilby, R.L., Wigley, T.M.L., Conway, D., Jones, P.D., Hewitson, B., Main, J., Wilks, D.S., 1998. Statistical downscaling of general circulation model output: a comparison of methods. *Water Resour. Res.* 34 (11), 2995–3008. <https://doi.org/10.1029/98WR02577>.
- Willmott, C.J., 1981. On the validation of models. *Phys. Geogr.* 1 (2), 184–194.
- Xiao, D., Bai, H., Tang, J., De, Liu, L., Yang, Y., 2021. The role of cropping system adjustment in balancing grain yield and groundwater use across a rainfall gradient in the North China Plain under future climate scenarios. *Irrig. Drain.* <https://doi.org/10.1002/ird.2653>.
- Zhang, F., Georgakakos, A.P., 2012. Climate change impacts on the hydrology of the North American great lakes. *Clim. Change* 111 (3), 945–972. <https://doi.org/10.1007/s10584-011-0167-9>.
- Zhang, H., Goodfellow, I., Metaxas, D., Odena, A., 2019. Self-attention generative adversarial networks. In: Proceedings of the 36th International Conference on Machine Learning, vol. 97, pp. 7354–7363. <https://doi.org/10.48550/arXiv.1906.06972>.