Supplementary material

This section provides a comprehensive overview of the cGAN architecture developed for the statistical downscaling of the ERA5 reanalysis. It includes a detailed description of the model's key components, including the neural network structures employed for both the generator and the discriminator. Furthermore, the choices regarding activation functions, loss functions, the dimensions of the gridded input and output datasets, and the adopted optimization strategy are specified. Additionally, an analysis is provided on the downscaling performance of the developed cGAN model under specific meteorological conditions, which underpin some of the findings discussed in the main manuscript. Furthermore, a detailed analysis was conducted not only for the entire Italian territory but also for the five macro-regions into which Italy is commonly divided, as described in the literature, including the study by Raffa et al. (2021). These macro-regions are: North-East, North-West, Central Italy, Southern Italy, and Insular Italy. This regional breakdown enables a more nuanced understanding of the performance of the cGAN model developed in this study, offering deeper insights into the results across geographically distinct areas, each characterized by the significant climatic variability inherent to the Italian Peninsula.

These details and specific analyses support the main manuscript, enabling a thorough understanding of the model's functionality and facilitating its reproducibility or adaptation for other applications.

A. Architecture details

Loss function

Within the intricate architecture (Fig. 4) of our dedicated cGAN tailored to the challenging task of climate downscaling, the careful selection of loss functions stands as a pivotal facet. These loss functions play a decisive role in shaping the training process for both the discriminator and generator networks, ensuring the acquisition of refined, high-resolution climate predictions that closely emulate real climate observations. For the discriminator, tasked with the critical role of distinguishing between authentic climate data (VHR_REA-IT) and the generated synthetic counterpart (ERA5-DownGAN), we have opted for the nn.BCEWithLogitsLoss() (Binary Cross-Entropy with Logits) loss function (PyTorch, 2023). This choice is underpinned by its well-recognized effectiveness within the GAN framework, serving as a robust measure of dissimilarity between the discriminator's predictions and the actual ground truth labels. Notably, the application of this loss function involves the sigmoid activation of the discriminator's

output, which, in turn, facilitates the calibration of the discriminator's acumen. Consequently, the discriminator attains an elevated proficiency in discriminating real climate data from the synthetic data generated by the GAN, a pivotal element influencing the overall authenticity of the generated climate data.

Conversely, for the generator network, which forms the core of the downscaling process, a combination of loss functions is employed. Two instances of nn.MSELoss() (Mean Squared Error) have been chosen (PyTorch, 2023). The first nn.MSELoss() instance takes the forefront by quantifying the mean squared error between the high-resolution climate data produced by the generator and their real-world counterparts. This primary loss function serves as the driver in the training process, compelling the generator to minimize the disparities between its synthetic outputs (ERA5-DownGAN) and the authentic climate observations (VHR_REA-IT).

This approach promotes the faithful replication of statistical attributes embodied by real climate data. The second application of nn.MSELoss plays an equally significant role. It acts as a mechanism to ensure the alignment of the higher-resolution climate images synthesized by the GAN with the lower-resolution inputs, often constituting the initial data. This usage of the second instance of nn.MSELoss safeguards the preservation of essential climatic attributes, spatial patterns, and structural integrity during the downscaling procedure. In parallel to these fundamental loss functions, the generator network is subjected to nn.BCEWithLogitsLoss() as part of the adversarial training process. This special loss function quantifies the differences between the discriminator's predictions on the generated climate data and the specified target labels. The process compels the generator to craft synthetic climate data that, in practical terms, becomes almost indistinguishable from authentic climate observations. This adversarial loss is a cornerstone of the GAN architecture, underpinning the generator's multifaceted mission: to emulate not only statistical attributes but also the detailed spatial and temporal patterns inherent in climatic data. The orchestrated use of these loss functions collectively delineates the training regimen of our climate downscaling GAN, culminating in the proficient generation of high-resolution climate inferences that mirror the characteristics of real climate data. This interplay of loss functions is the keystone of the convergence process during training, ultimately resulting in climate predictions that boast augmented spatial resolution and a remarkable level of authenticity, an imperative achievement for the evolving domain of climate downscaling.

Optimizer

The choice of the optimization strategy plays a pivotal role in shaping the efficacy of our cGAN. To this end, we have harnessed the power of the Adam optimizer, a venerable algorithm renowned for its adaptability and efficiency. Adam, short for Adaptive Moment Estimation (Alzubaidi L., et al., 2021), unites the virtues of both momentum-based optimization and root mean square propagation. This amalgamation empowers the optimizer to dynamically calibrate learning rates for each network parameter during the training process. In our configuration, the best results are obtained with a learning rate of 0.0001 applied to both the generator and discriminator networks. With this choice, our GAN converges toward a solution that optimally balances the interplay between the generator and discriminator networks, ultimately culminating in the generation of high-fidelity, high-resolution climate predictions. The utilization of the Adam optimizer and the specific learning rate, serves not only as a testament to the sophistication of our approach but also underscores our unwavering dedication to employing state-of-the-art optimization techniques to elevate the quality and veracity of our downscaled climate data.

The two neural networks: implementation specifics

The generator in GAN discussed in this work uses four linear layers with LeakyReLU activation functions (Dubey, A.K., Jain, V., 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. 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.

Discriminator Architecture:

The discriminator also employs linear layers with LeakyReLU activation functions, and dropout layers to introduce regularization. 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.

Input and Output Sizes:

The generator's input data consists of matrices with dimensions of 58x80, corresponding to a horizontal resolution of 31 km. The primary objective of the generator is to downscale this input to matrices of data with dimensions of 680x535, 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.

Training Data and Process:

During the training step, the discriminator uses a dataset of real high-resolution data (VHR_REA-IT). The generator aims to produce synthetic data at the same higher resolution (680x535) that resembles the real high-resolution data.

Investigating cGAN Downscaling Under Peculiar Synoptic Conditions

In this section, we investigate the behavior of the downscaling model developed in this study and, consequently, the ERA5-DownGAN dataset it produced, during complex synoptic conditions, as evident in the correlation plot in Fig. 4c. The primary aim of this analysis is to evaluate the robustness and performance of the cGAN in anomalous meteorological conditions. This investigation allows us to assess the model's effectiveness in downscaling under atypical conditions, offering a more detailed understanding of its applicability and limitations in real-world contexts where meteorological patterns deviate significantly from seasonal averages. Furthermore, to understand the behavior of the dataset produced by the GAN-based statistical downscaling in comparison with dynamic downscaling, the two datasets ERA5-DownGAN and VHR_REA-IT are compared with the observed-gridded SCIA data (Sistema nazionale per la raccolta, elaborazione e diffusione di dati Climatologici di Interesse Ambientale; Desiato F. et al., 2007). SCIA is an observational dataset derived from hundreds of weather stations covering the entire Italian Peninsula (http://www.scia.isprambiente.it/wwwrootscia/help_eng.html), shown in Fig. A.3, spanning the period from January 1961 to December 2020 on a regular 5 km grid for temperature variables. For this analysis, we used the mean temperature field derived from the variables tmin and tmax, as these are the only temperature variables available in the SCIA dataset.

The day of March 24, 2001, was characterized by a significant heatwave, attributed to the establishment of an African anticyclone from March 18 onwards. This high-pressure system facilitated the inflow of warm air into the central Mediterranean and Italy, resulting in a notable increase in temperatures. The warmest temperatures were recorded primarily on March 22, 23, and 24, with values markedly above seasonal averages. In particular, one-third of Sardinia experienced temperatures exceeding 30 °C, reaching peaks of 34.8 °C in Jerzu, 34.7 °C in Siniscola, 33.2 °C in Oliena, and 33.0 °C in Muravera. Additionally, several weather stations reported nighttime minimum temperatures around 20 °C, with 20.2 °C recorded in Iglesias and

19.7 °C in Modolo and Domus de Maria (timento.imc@arpa.sardegna.it). It is noteworthy that, according to the historical series maintained by the Regional Meteorological Service (SAR) dating back to 1928, there are no records of such high temperatures in March. The only comparable event occurred in March 1990, when select weather stations reported maximum temperatures exceeding 30 °C, peaking at 33.0 °C in Laconi.

Figure A.I illustrates the synoptic conditions for the specified date, March 24, 2001. These historical maps were sourced from www.meteociel.it and derived from NCEP reanalyses (Reanalysis 1, 2, and the 20th Century Edition), featuring a resolution of 2.5° x 2.5°.

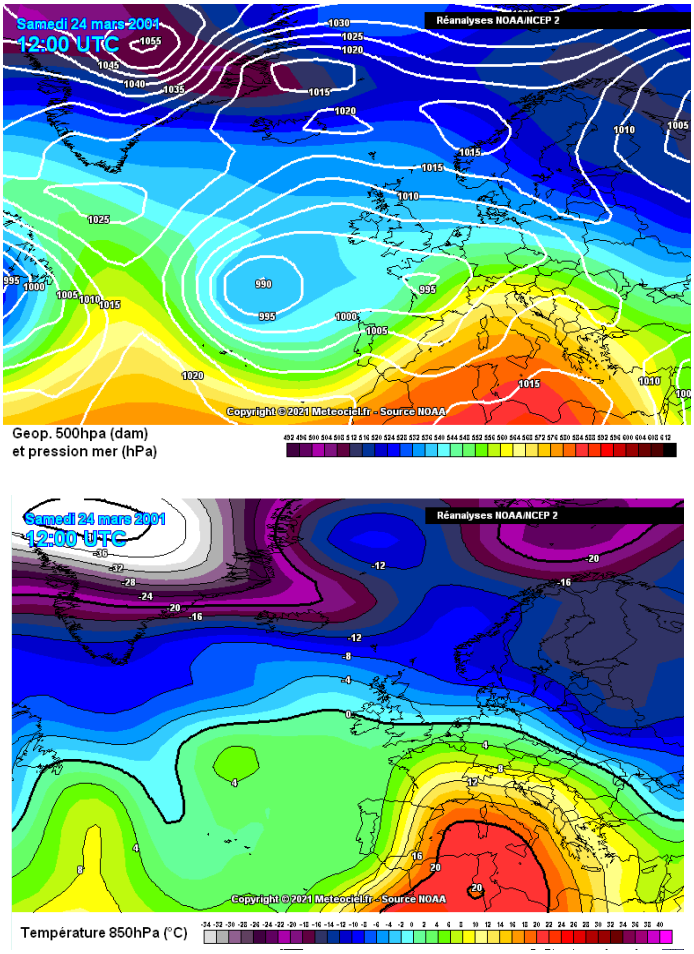


Fig. B.I: Synoptic analysis at 12:00 UTC from NOAA/NCEP reanalysis on 24th March 2001. Maps of geopotential at 500 hPa (upper) and temperature (°C) at 850 hPa (lower). Reproduced from (www.meteociel.it).

The comparison of the 2-meter temperature field for March 24, 2001, between the low-resolution ERA5 dataset, the statistically downscaled ERA5-DownGAN field, the dynamically downscaled VHR_REA-IT dataset, and the difference map between ERA5-DownGAN and VHR_REA-IT is illustrated in Fig. B.2. It is observed that the error between ERA5-DownGAN

and VHR_REA-IT shows a significantly larger discrepancy compared to the results discussed in the main manuscript, where the comparison between the two datasets indicated excellent agreement and minimal differences. In this case, however, the differences reach values up to +7.6 and –6.8 °C, with a median value of +2.2 °C (Fig. B.2, Panel a). To facilitate the analysis of the areas in the domain with the largest discrepancies, the range of the difference maps was set between -3 and +3 °C. To assess the alignment of the statistical model (ERA5-DownGAN) with the dynamic model (VHR_REA-IT) for similar period, the mean temperature field for March 24 was analyzed over the study period (2001-2005). The results show an excellent agreement between the fields produced by ERA5-DownGAN and VHR_REA-IT in this climatological analysis: the difference between the statistically and dynamically downscaled datasets varies between -0.5 and +0.5 °C, with a median value across the entire domain of approximately 0.08 °C, indicating near-zero error, with maximum differences observed only in specific locations, reaching up to +2.5 and -1.9 °C (Fig. B.2, Panel b). This finding confirms the effectiveness of the developed cGAN model in the downscaling application for the temperature field, as discussed in the section of the results in the main manuscript.

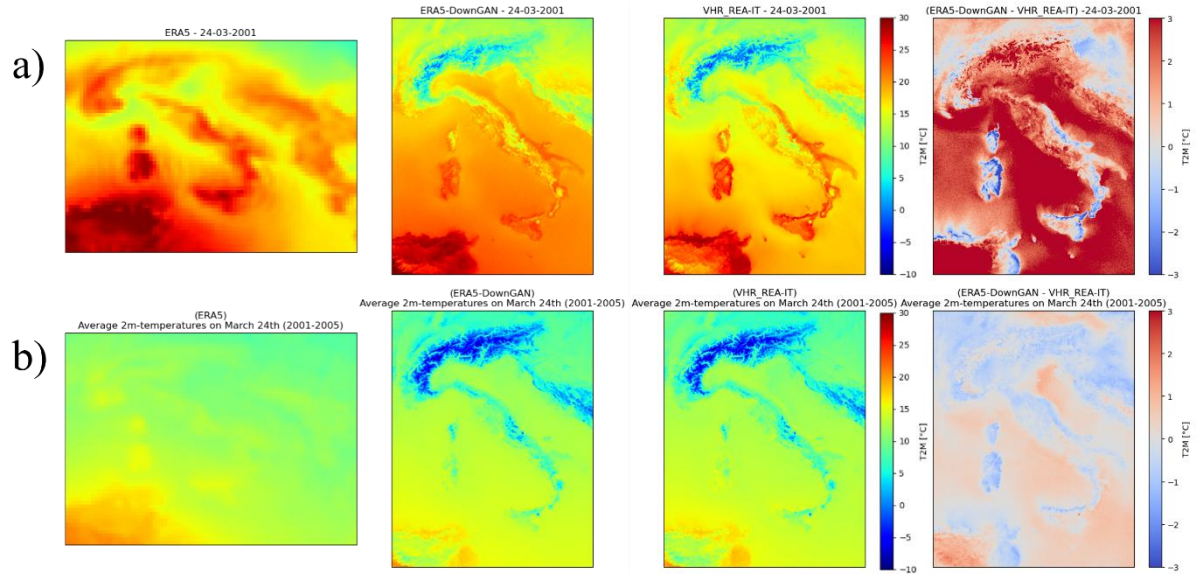


Fig. B.2:. Panel (a): Maps for March 24, 2021, for ERA5 (LR), ERA5-DownGAN (HR), and VHR_REA-IT (HR), along with the difference between ERA5-DownGAN and VHR_REA-IT. Panel (b): Average of all March 24 observations from 2001 to 2005 for ERA5 (LR), ERA5-DownGAN (HR), and VHR_REA-IT (HR), as well as the difference between ERA5-DownGAN and VHR_REA-IT.

In the context of analyzing the performance of the statistical downscaling model based on cGAN developed in this study, we have considered the VHR_REA-IT dataset, obtained through dynamic downscaling, as representative of "high-resolution" reality. This approach is justified by the model's objective to generate fields that are totally comparable to those produced by a physically-based model, which has been previously validated in the studies of Raffa et al. 2021 and Adinolfi et al. 2023.

Currently, we focus on assessing the actual performance of the two downscaling models (VHR_REA-IT and ERA5-DownGAN) under an anomalous condition, specifically the heatwave of March 24, 2001, using an observational dataset to represent reality. The observational dataset in question is the observed-gridded SCIA data (Fig. A.4), from which the surface mean temperature field is derived from the available variables of maximum (tmax) and minimum (tmin) temperatures. In this case, the grid resolution is 5 km, and the field is obtained by interpolating data from meteorological stations. Therefore, it is reasonable to expect differences arising from the varying resolutions of the two downscaling models, which operate at a resolution of approximately 2.2 km, compared to the observational SCIA grid dataset at 5 km.

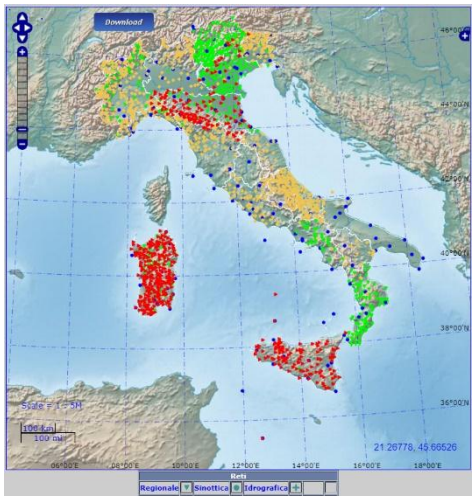


Fig. B.3:. Map of the distribution of SCIA weather stations for tmin and tmax (2001). Sourced from https://scia.isprambiente.it/.

The comparison between the surface mean temperature fields of the two downscaling datasets and those derived in the SCIA dataset is conducted by exclusively considering the values recorded on the Italian peninsula and the mainland, thereby limiting the analysis to the territory covered by the observation stations included in the SCIA dataset.

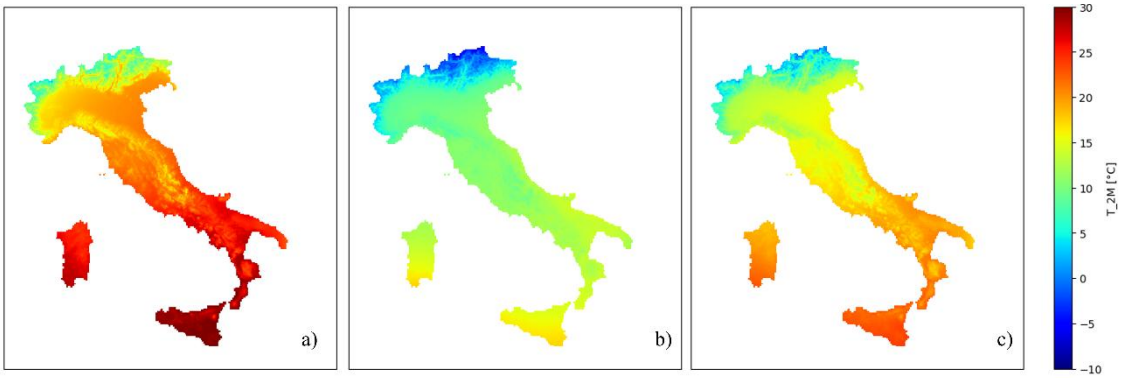


Fig. B.4:. Maps of SCIA observed-gridded temperature dataset. (a) Maximum temperature (tmax); (b) Minimum temperature (tmin) and the derived mean temperature (c).

Dataset	Temperature	Min. [°C]	Max. [°C]	Mean [°C]	Median [°C]
SCIA	derived mean temperature	-1.8	24.0	15.6	15.8
ERA5-DownGAN	Tmean	-4.3	25.5	16.9	18.3
VHR_REA-IT	Tmean	-7.3	29.3	15.3	16.2
ERA5	Tmean	12.6	28.8	19.8	18.7
SCIA	Tmax	1.0	31.0	21.1	20.9
SCIA	Tmin	-7.5	19.5	10.2	10.7

Tab. B.I: Comparison on 24th March 2001 of minimum, maximum, mean, and median values for the datasets SCIA, ERA5-DownGAN, VHR_REA-IT, and ERA5, with horizontal resolutions of approximately 5 km, 2.2 km, 2.2 km, and 31 km, respectively. Descriptive statistics for the observational dataset SCIA for Tmin and Tmax.

An analysis of the visual comparisons presented in Fig. B.5, along with the descriptive statistics presented in Tab. B.I, reveals that the datasets derived from downscaling, ERA5-DownGAN and VHR_REA-IT, representing statistical and dynamic approaches respectively, demonstrate remarkably similar patterns and comparable values.

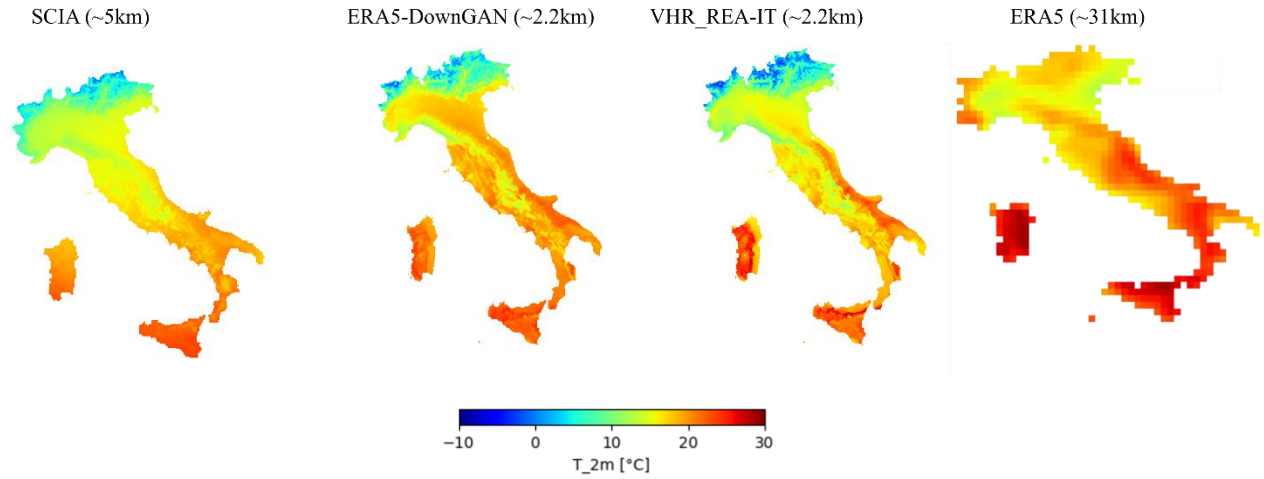


Fig. B.5:. Comparison of 2m-temperature field only over Italian Peninsula on 24th March 2001. SCIA observed-gridded temperature dataset; ERA5-DonwGAN; VHR_REA-IT and ERA5.

However, the GAN-generated dataset exhibits higher temperatures in the regions of Emilia-Romagna, Tuscany, and Abruzzo compared to those obtained through dynamic downscaling and the observational dataset. This temperature elevation is particularly pronounced in the Po Valley regions. Conversely, an inversion of this trend is noted in the island regions, where VHR_REA-IT is identified as the warmer dataset, displaying a more significant deviation from the observational dataset SCIA relative to ERA5-DownGAN. While the mean and median values suggest a stronger concordance between the dynamically downscaled dataset and the SCIA observational data, the dataset produced via conditional Generative Adversarial Networks (cGAN) demonstrates a closer alignment with the minimum and maximum values. As delineated in Table A.I, the maximum temperature recorded for ERA5-DownGAN is 25.5 °C, in contrast to 29.3 °C for VHR_REA-IT and 24.0 °C for SCIA. Regarding the minimum values documented in alpine areas, the discrepancies are even more pronounced: SCIA records -1.8 °C, while ERA5-DownGAN and VHR_REA-IT register -4.3 °C and -7.3 °C, respectively. In these alpine regions, the overall tendency for higher temperatures in the ERA5-DownGAN dataset exhibits a closer alignment with the SCIA observations.

Additional case study of particular relevance for assessing the performance of our statistical downscaling method, as highlighted in the analyses conducted in Chapter 3, is March 2, 2004, a day when we observed a discrepancy between the dataset generated by the conditional Generative Adversarial Network (cGAN) and that produced by dynamic downscaling. On March 2, 2004, Europe was affected by a complex synoptic system characterized by the presence of a broad low-pressure area that was deepening and translating toward the eastern sectors. This low-pressure area had dominated much of Europe in the preceding days, due to the deepening of a jet stream trough that was situated between two high-pressure systems. The interaction between these structures generated a flow of cold air of Arctic origin toward Central and Western Europe, promoting atmospheric instability and resulting in a decrease in temperatures (Fig. A.6), including over the Italian Peninsula.

Figure B.7 illustrates the temperature field at 2 meters for March 2, 2004, comparing the low-resolution ERA5 dataset, the statistically downscaled field using ERA5-DownGAN, the dynamically downscaled VHR_REA-IT dataset, and the map of differences between ERA5-DownGAN and VHR_REA-IT. For this case study, the results corroborate the observations made in Chapter 3 for this date, indicating a significantly greater discrepancy between ERA5-DownGAN and VHR_REA-IT compared to other dates in the reference period. This suggests that the GAN demonstrates a high level of consistency with the high-resolution VHR_REA-IT

dataset. Notably, the differences reach extreme values of +2.0 °C and -6.8 °C, with a median value of -1.64 °C (Figure B.7, Panel a).

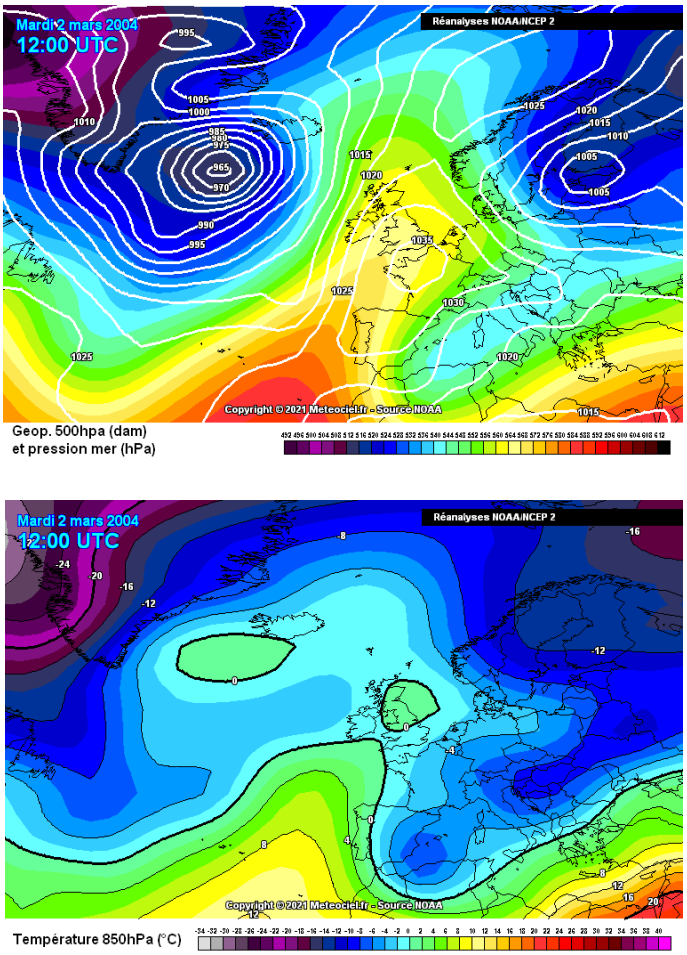


Fig. B.6: Synoptic analysis at 12:00 UTC from NOAA/NCEP reanalysis on 2sd March 2004. Maps of geopotential at 500 hPa (upper) and temperature (°C) at 850 hPa (lower). Reproduced from (www.meteociel.it).

To verify the consistency between the statistical model (ERA5-DownGAN) and the dynamic model (VHR_REA-IT) over the same period, an analysis of the mean temperature field for March 2 was conducted for the study period (2001-2005). The results show an excellent agreement between the temperature fields generated by ERA5-DownGAN and VHR_REA-IT in this climatological context: the difference between the statistically and dynamically downscaled datasets ranges between -0.5 and +0.5 °C, with a median value across the entire domain of approximately -0.77 °C, indicating a slight tendency toward underestimation (Fig. B.7, Panel d). This result further confirms the effectiveness of the developed cGAN model in downscaling the 2m-temperature field.

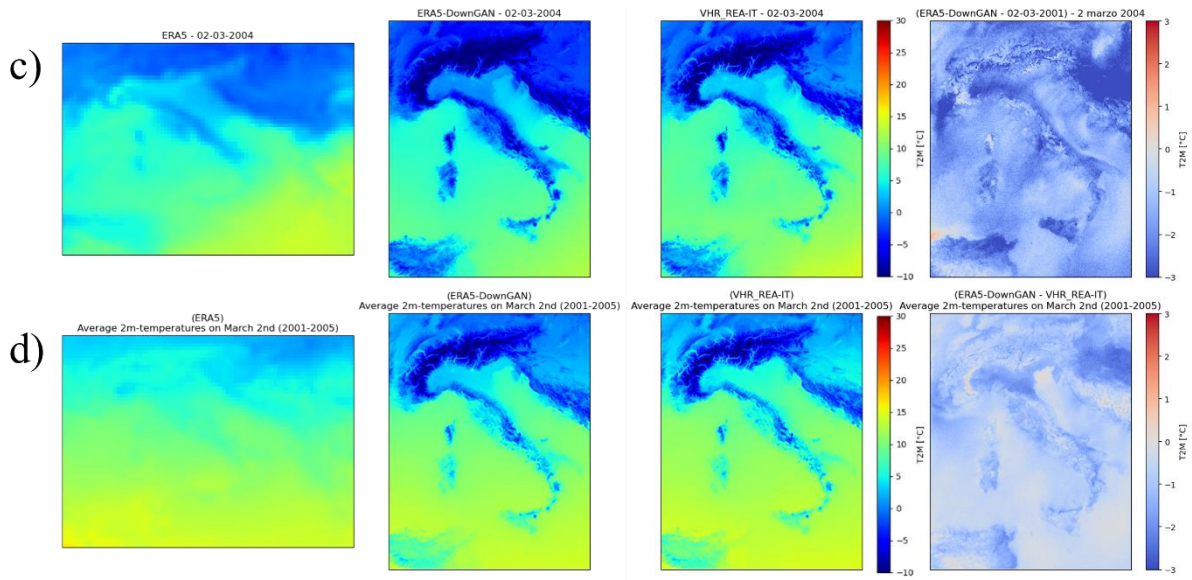


Fig. B.7:. Panel (a): Maps for March 2, 2024, for ERA5 (LR), ERA5-DownGAN (HR), and VHR_REA-IT (HR), along with the difference between ERA5-DownGAN and VHR_REA-IT. Panel (b): Average of all March 2 observations from 2001 to 2005 for ERA5 (LR), ERA5-DownGAN (HR), and VHR_REA-IT (HR), as well as the difference between ERA5-DownGAN and VHR_REA-IT.

The analysis of the temperature field at 2 meters produced by the two downscaling methodologies is further explored in this second case study, comparing the ERA5-DownGAN and VHR_REA-IT datasets with the observational gridded dataset SCIA for the Italian Peninsula. A preliminary visual analysis (see Fig. B.8) indicates that both downscaling methodologies generally maintain excellent agreement in terms of pattern localization with respect to the observational dataset, except for the Islands and the western Alps sector. An interesting difference emerges for Sardinia, where the SCIA dataset highlights lower temperatures in the eastern flank. In contrast, ERA5-DownGAN, aligning with the behavior of VHR_REA-IT, detects colder temperatures in the western sector. Conversely, ERA5, despite its significantly lower resolution compared to both the observational dataset and the downscaled products, appears to replicate the observed behavior in SCIA for Sardinia, albeit with higher temperature values due to its coarse resolution. This result emphasizes the model's tendency to align with the conceptual patterns identified in VHR_REA-IT, even when starting from the ERA5 dataset. Such alignment is crucial, as the ultimate goal is to produce a high-resolution field comparable to that generated by physics-based downscaling methods. Despite the strong agreement in terms of spatial patterns, the ERA5-DownGAN dataset shows a general tendency to underestimate temperatures compared to the observational dataset SCIA. As indicated in Table B.2, the minimum recorded in the mean temperature field in ERA5-

DownGAN is -20.9 °C, very close to the -20.4 °C observed in VHR_REA-IT but higher than the -23.0 °C reported in SCIA. The greatest discrepancies are observed in terms of the maximum value identified in the mean temperature field, with values of approximately 10.3 °C for ERA5-DownGAN, 11.0 °C for VHR_REA-IT, and 13.0 °C for SCIA. This trend is further pronounced for the median temperature, which is around 0.1 °C in ERA5-DownGAN, 2.2 °C in VHR_REA-IT, and 2.7 °C in SCIA.

Dataset	Temperature	Min. [°C]	Max. [°C]	Mean [°C]	Median [°C]
SCIA	derived mean temperature	-23.1	13,0	1.6	2.7
ERA5-DownGAN	Tmean	-20.9	10.3	-1,0	0.1
VHR_REA-IT	Tmean	-20.4	11.0	-1,0	2.2
ERA5	Tmean	-1.5	10.5	4,0	3.7
SCIA	Tmax	-19.7	16.3	5.7	6.6
SCIA	Tmin	-26.4	10.0	-2.5	-1.9

Tab. B.2: Comparison on 2sd March 2004 of minimum, maximum, mean, and median values for the datasets SCIA, ERA5-DownGAN, VHR_REA-IT, and ERA5, with horizontal resolutions of approximately 5 km, 2.2 km, 2.2 km, and 31 km, respectively. Descriptive statistics for the observational dataset SCIA for Tmin and Tmax.

Overall, both downscaling products exhibit an underestimation compared to the observational dataset SCIA, with a more pronounced trend observed in ERA5-DownGAN. It is important to note that, despite this underestimation in the dataset produced by downscaling via cGAN, there is no exacerbation of extreme values related to the minimums. This is a significant aspect in the context of downscaling using GANs, as one of the common issues associated with the use of GANs in climate science is the generation of anomalous and extreme values, a behavior that is, however, mitigated by the use of a cGAN (Leinonen et al., 2019, 2021) as in this work.

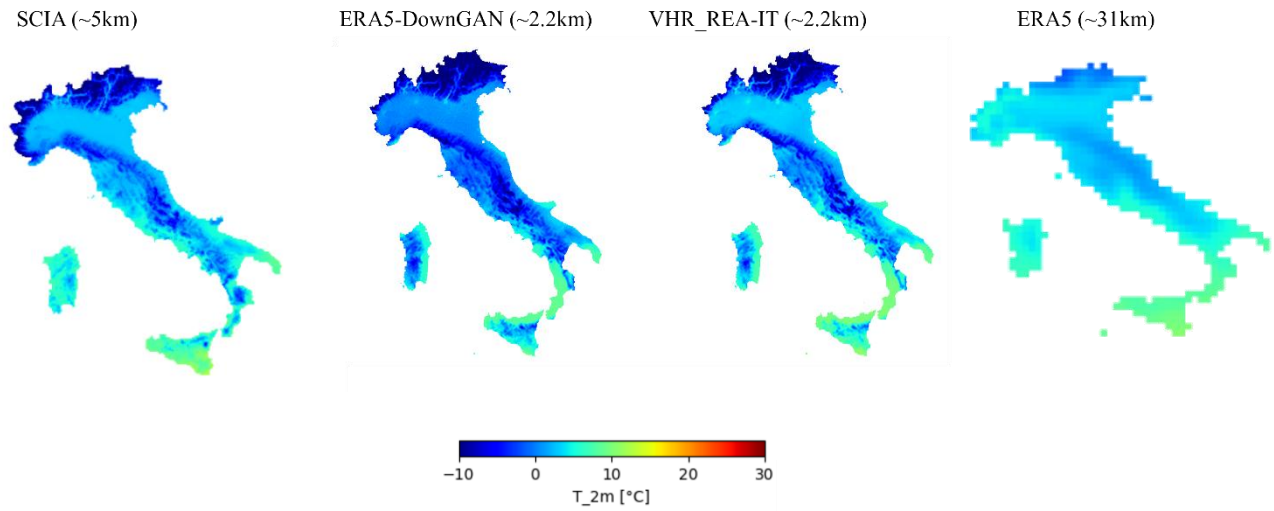


Fig. B.8:. Comparison of 2m-temperature field only over Italian Peninsula on 2sd March 2004. SCIA observed-gridded temperature dataset; ERA5-DonwGAN; VHR_REA-IT and ERA5.

In conclusion, the analysis of anomalous weather conditions reveals that the developed cGAN architecture effectively maintains spatial coherence in the representation of patterns compared to the dataset produced by the dynamic downscaling method VHR_REA-IT. Although the values generated by ERA5-DownGAN remain within a similar range, there is a noticeable tendency to exaggerate anomalous conditions relative to the reference dataset VHR_REA-IT. Specifically, during the heatwave of March 24, 2001, ERA5-DownGAN demonstrates an overestimation of temperatures compared to VHR_REA-IT, while aligning more closely with the observational dataset SCIA, particularly in the regions of the Alps and the islands. This alignment, however, highlights an extremization of the trends identified in VHR_REA-IT and reflects a broader tendency among convection-permitting models to overestimate surface temperature fields in the Po Valley. Importantly, despite these tendencies, ERA5-DownGAN does not generate anomalous maximum values; rather, it tends to mitigate those found in VHR_REA-IT, aligning them more closely with the SCIA observational dataset. Conversely, in conditions characterized by colder temperatures relative to the seasonal average, ERA5-DownGAN registers an underestimation compared to both the SCIA observational dataset and the VHR_REA-IT dataset, without producing anomalous minimum values. Therefore, even under anomalous weather conditions, the downscaling architecture based on cGAN effectively replicates the fine patterns and details observed in VHR_REA-IT, demonstrating a general tendency to exaggerate anomalous conditions that does not impact the generation of outliers, but rather appears to highlight a possible exemplification of meteorological dynamics in such peculiar contexts.

B. Enhanced Regional Characterization of Italy's Climatic Variability

In this section, we have expanded our analysis to offer a more detailed regional assessment. The analyses presented here complement the main manuscript, where the evaluation is conducted across the entire domain centered on Italy. Specifically, the evaluation presented herein is based on Italy and on the division and analysis of five distinct regions: North-West, North-East, Central Italy, South Italy, and the Insular Italy, as defined by the first level of the Nomenclature of Territorial Units for Statistics (NUTS 1) for Italy. This refined regional perspective is crucial for accurately capturing the inherent heterogeneity of Italy’s climate. Moreover, it further demonstrates the robustness of the cGAN model developed for downscaling, highlighting its reliability through a comparison of the dataset produced by ERA5-DownGAN with the dynamically simulated VHR-REA_IT dataset, both at a similar resolution of approximately 2.2 km.

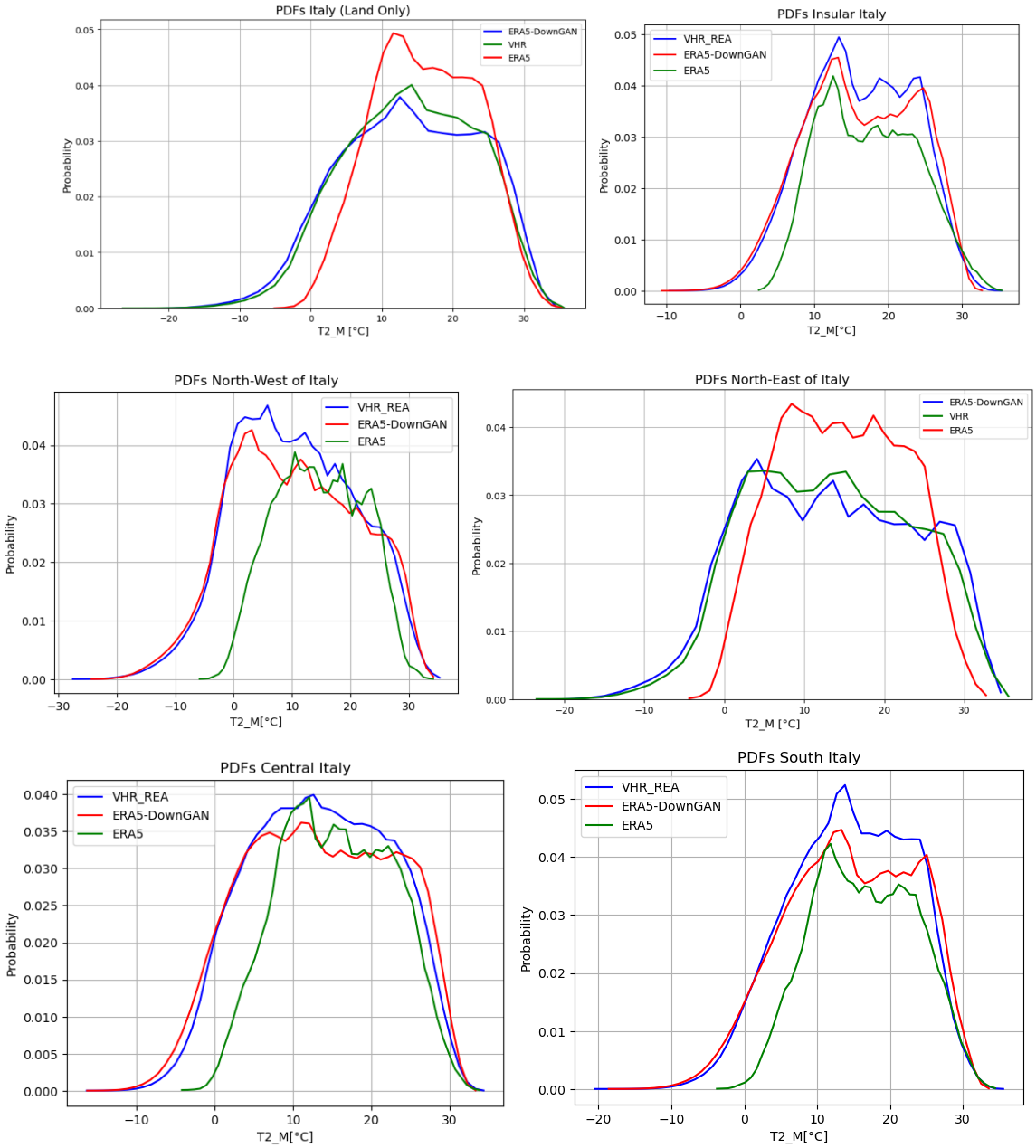
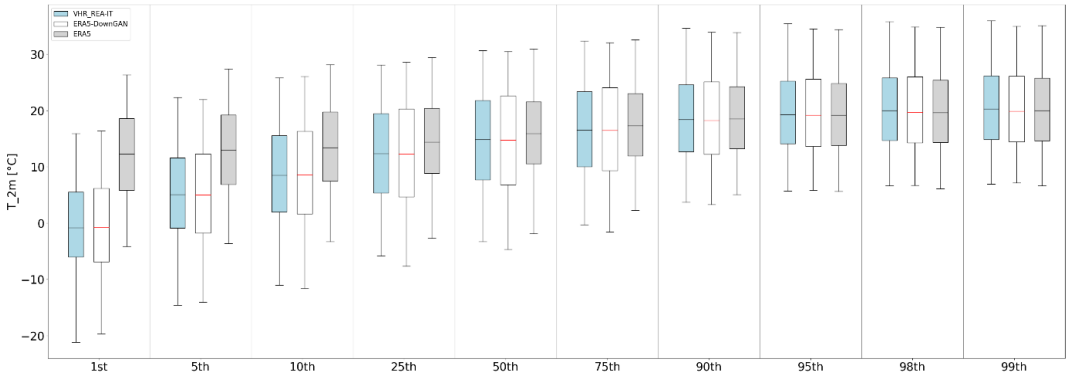


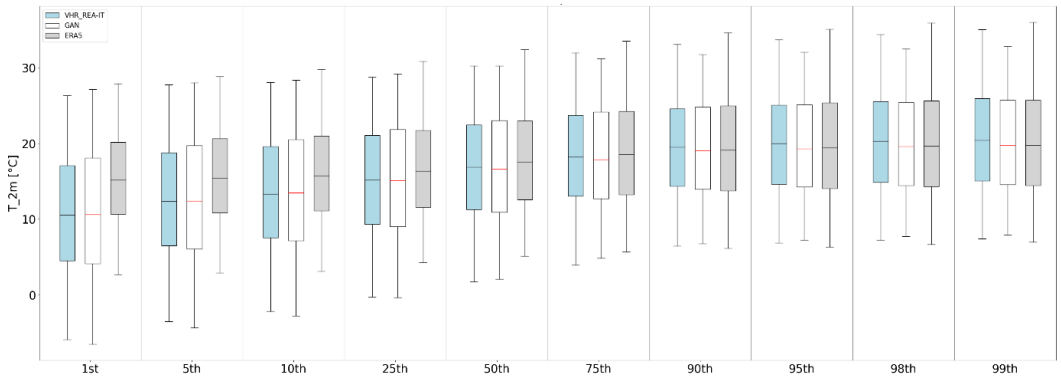
Fig. C.1: Probability density functions (PDFs) for VHR-REA_IT (blue line), ERA5-DownGAN (red line), and ERA5 (green line), considering Italy and five distinct regions: North-West, North-East, Central Italy, Southern Italy, and the Islands.

The Probability Density Functions (PDFs) show strong agreement in both distribution and value range between the high-resolution dataset produced by the dynamical model (VHR_REA_IT) and the downscaled dataset generated with our cGAN-based approach (ERA5-DownGAN). The most notable differences are observed in southern Italy, where probabilities are lower and the cGAN-derived distribution appears slightly flatter. Additionally, in the northwestern region, the minimum value reached in the ERA5-DownGAN dataset is approximately -28°C, whereas the dynamical model reaches around -30°C. Although relatively minor, these discrepancies highlight regional variations in downscaling performance.

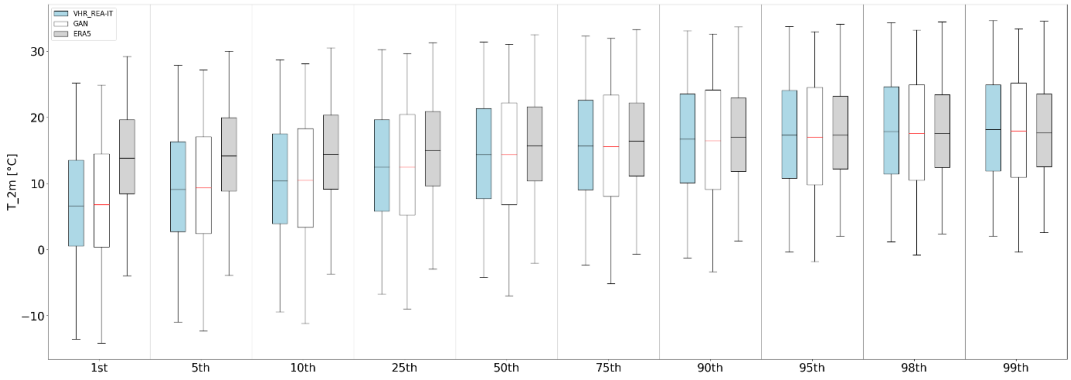
Box-Plot Italy (Land Only)



Box-Plot Insular Italy



Box-Plot Central Italy



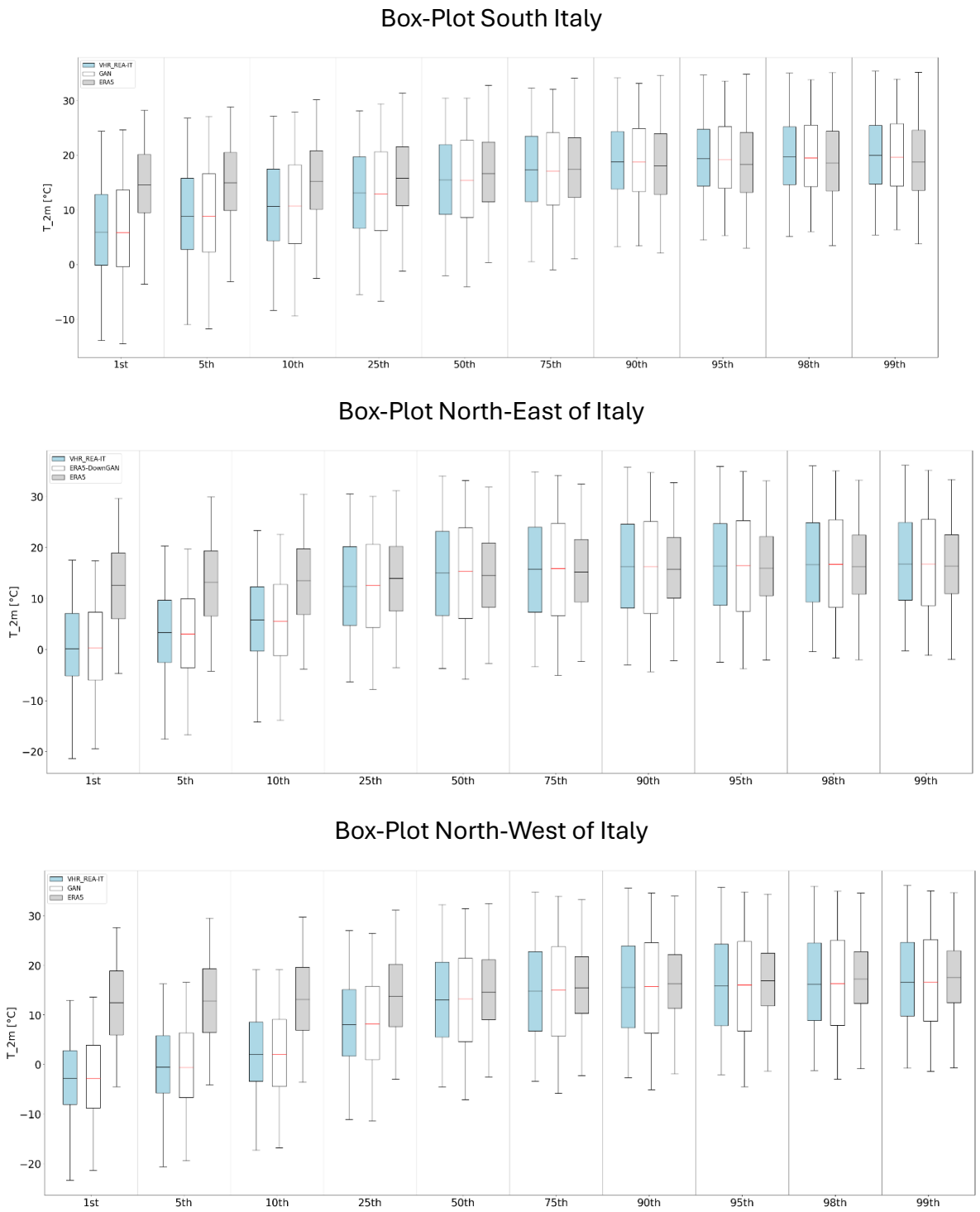


Fig. C.2: Box plot from 1st to 99th percentiles for VHR-REA_IT (blu box), ERA5-DownGAN (white box), and ERA5 (gray box), considering Italy and five distinct regions: North-West, North-East, Central Italy, Southern Italy, and the Islands.

The assessment of percentiles, ranging from the 1st to the 99th, shown in Fig. C.2, allows for analyzing how well the generated data align with the distribution of the high-resolution dataset. The comparison of percentile values proved to be a particularly effective tool for evaluating

the cGAN's ability to replicate both extreme events and the overall data distribution. The boxplots were constructed by considering all values calculated for each specific percentile (from the 1st to the 99th) across the entire Italian Peninsula and the analyzed subregions. This approach allows for examining not only the global characteristics of the distribution but also regional peculiarities, highlighting any behavioral differences between various geographical areas.

The analysis of the boxplots shows significant differences in the distribution of surface temperature between the low-resolution real dataset (ERA5) and the two high-resolution datasets, one real (VHR_REA-IT) and the other generated by the cGAN (ERA5-DownGAN). In general, ERA5 shows a median and distribution systematically shifted towards higher temperature values compared to the other two datasets for all percentiles considered. This discrepancy is particularly noticeable for the lower percentiles (from the 1st to the 25th), with a marked accentuation in the Northwest. This behavior suggests that ERA5 tends to overestimate lower temperatures relative to the high-resolution datasets, especially in the northern regions.

The comparison between the results obtained from the cGAN and VHR_REA-IT, however, reveals significant consistency in the overall distribution, confirming the strong advantage introduced by the downscaling developed in this study. The cGAN accurately replicates both the median and the whisker extension of VHR_REA-IT, confirming its ability to reconstruct the variability of 2m temperature with a high level of fidelity. However, a slight difference is observed in the width of the boxes, which is slightly larger in ERA5-DownGAN, particularly for the Northeast and Northwest.

The extension of the whiskers in the boxplots of the different subregions confirms the results obtained across the entire domain, reinforcing the hypothesis that while the cGAN ensures an excellent reproduction of the overall distribution, it may introduce slight variations in the extremes, with a slight shift towards lower values for both the upper and lower extremes. However, the comparison between the data generated by the cGAN (ERA5-DownGAN) and the real high-resolution dataset (VHR_REA-IT) for the test period generally shows excellent results in terms of accuracy and consistency across all analyzed subregions. The generative model is able to reproduce a temperature field with a distribution, median, and extreme values that are substantially equivalent to those of the real dataset, confirming its effectiveness as a tool for generating high-resolution climate data.