



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DOTTORATO DI RICERCA IN
IL FUTURO DELLA TERRA, CAMBIAMENTI CLIMATICI E SFIDE
SOCIALI

Ciclo 37

Settore Concorsuale: 04/A4 - Geofisica

Settore Scientifico Disciplinare: GEO/12 – Oceanografia e Fisica dell'Atmosfera

**AI-ASSISTED CLIMATE DOWNSCALING FOR RAPID ASSESSMENT
OF ERA5 REANALYSIS ACROSS DIFFERENT GEOGRAPHICAL
DOMAINS**

Presentata da: Ilenia Manco

Coordinatore Dottorato:

Prof.ssa Silvana Di Sabatino

Supervisore:

Prof. Antonio Navarra

Co-Supervisori:

Dott.ssa Paola Mercogliano

Dott. Marco Briscolini

Esame finale anno 2025

Abstract

State-of-the-art General Circulation Models (GCMs) typically operate at a coarse spatial resolution, posing challenges in accurately assessing regional climate changes and their impacts. This limitation is particularly evident in representing regional-scale topography and meteorological processes, including extreme weather events. Traditional dynamical downscaling methods address these issues but are computationally intensive, while statistical approaches, though efficient, often compromise spatial consistency. This study introduces an innovative application of Conditional Generative Adversarial Networks (cGANs) for climate data downscaling to address these challenges. GANs consist of two interconnected components: a generative and discriminative models. The generative model is based on ERA5 climate reanalysis data (Hersbach et al., 2020, ~31 km resolution) and learns to produce high-resolution data. The discriminator uses the VHR-REA_IT dataset (Raffa et al., 2021, ~2.2 km resolution) to distinguish between real and generated data by the GAN, known as ERA5-DownGAN. VHR_REA-IT is a convection-permitting reanalysis (horizontal grid spacing 0.02° , ≈ 2.2 km) generated with COSMO in Climate Mode (COSMO-CLM) over a domain covering the Italian Peninsula. The generative model refines its output through an iterative adversarial process to achieve enhanced spatial coherence. This study pioneers the use of cGANs to downscale ERA5 reanalysis data to high horizontal resolution (~2.2 km) for both temperature and precipitation fields. Specifically, the model uses 2-meter temperature (T_2M) data to downscale T_2M and precipitation data (TOT_PREC) to downscale precipitation. The training phase (1990-2000) allows the cGAN to learn the high-resolution data patterns, while the testing phase (2001-2005) evaluates its performance against VHR-REA_IT. The cGAN accurately reproduces patterns and value ranges for temperature and total precipitation fields, exhibiting a slight tendency toward cooler values. Furthermore, the cGAN downscaling model maintains strong consistency across all percentile classes (from the 1st to the 99th) for temperature, and in nearly all classes for total precipitation, with a tendency to generate outliers in the precipitation fields for the extreme classes (98th and 99th percentiles). Additionally, the GAN model developed was validated in

collaboration with the National Center for Atmospheric Research (NCAR) through an added case study centered on the United States territory.

This extension assessed the model's versatility across different datasets and geographical domains, confirming its effectiveness and potential as a robust tool for high-resolution climate data generation and regional climate analysis. This research demonstrates the significant potential of GANs to address the spatial limitations of traditional climate models, offering a powerful method for high-resolution climate data generation and contributing valuable insights into regional climate dynamics.

Contents

I Aim of the Work.....	1
I.1 Introduction.....	3
I.2 The Role of Artificial Intelligence in Weather and Climate.....	4
I.3 Machine Learning and Deep Learning for Statistical Downscaling.....	7
I.4 Artificial Neural Networks (ANNs).....	11
I.4.I Neurons and Activation Functions.....	13
I.5 Generative Adversarial Networks (GANs).....	23
I.5.I Conditional Generative Adversarial Networks (cGANs).....	27
2 Designing the Super-Resolution cGAN.....	29
2.1 Objectives and Necessity of Super-Resolution for Atmospheric Fields.....	29
2.2 Statistical and Dynamical Downscaling.....	30
2.3 Datasets and <i>pre-processing</i>	32
2.3.I Domain Definition.....	32
2.3.2 Low-Resolution Dataset: ERA5.....	33
2.3.3 High-Resolution Dataset: VHR_REA-IT.....	34
2.3.4 Transformations and Data Normalization.....	36
2.4 Proposed cGAN Architecture: ERA5-DownGAN.....	39
2.4.I Overall Architecture.....	41
2.4.2 Details of the Generator and Discriminator Architectures.....	41
2.4.3 Configuration of Hyperparameters and Optimization Technique Adopted.....	43
2.4.4 Training Strategies.....	46
2.4.5 Custom Loss Functions.....	47
3 Application of cGAN for the Italian Peninsula.....	49
3.I Downscaling of 2-meter Temperature.....	49
3.I.I Results and Analysis: Comparison with Dynamical Downscaling.....	50

3.1.2	Error metrics Evaluation.....	53
3.1.3	Spatial and Temporal Consistency.....	54
3.1.4	cGAN's Ability to Capture Extreme Values.....	57
3.1.5	Computational Details.....	61
3.2	Downscaling of Total Precipitation.....	62
3.2.1	Logarithmic Transformation	64
3.2.2	Exploring the Influence of Loss Weights and d_step in ERA5-DownGAN.....	76
3.2.3	An Alternative Normalization Scheme.....	84
3.2.4	Comparison of All Configurations.....	88
3.2.5	Computational Details.....	89
3.3	Potential Enhancements.....	90
4 	Expanding the Application of cGAN Downscaling to U.S. Geographical Domain...	92
4.1	Context and Motivation for the Case Study.....	92
4.2	Datasets and <i>Pre-processing</i>	94
4.2.1	High-Resolution Dataset: WRF-NCAR.....	94
4.2.2	Assessing the Portability of The Developed cGAN Downscaling Model on Different Supercomputer Architectures.....	95
4.3	Results and Analysis: Comparison with Dynamical Downscaling.....	97
4.3.1	Overall Architecture: Upgrade of The Number of Neurons.....	98
4.3.2	The Role of Epochs in Model Training.....	106
4.4	Summary Evaluation.....	107
5 	Discussion.....	110
5.1	General Conclusions.....	111
5.2	Future Developments.....	112
A 	Investigating cGAN Downscaling Under Peculiar Synoptic Conditions.....	114
B 	Lateral work: Dynamical Downscaling over the Italian Peninsula Using WRF, COSMO, and ICON Atmospheric Models.....	123
C 	Lateral work: AI-Based Method for Identifying Historical Recurrent Precipitation Patterns and Evaluating Changes in Future Projections.....	129
	<i>Code and data availability</i>	134
	<i>Acknowledgments</i>	139

<i>Bibliography</i>	140
---------------------------	-----

List of Figures

- I.1 (a) Comparison of PP (left) and SR (right) approaches for climate downscaling in New Zealand: PP maps large-scale fields to high resolution, while SR maps coarsened fields to high resolution. (b) Comparison between perfect and imperfect RCM training frameworks, with perfect using coarsened RCM fields and imperfect using GCM fields. Adapted from Rampal et al. (2024).
- I.2 Example of a fully connected Neural Network. Reproduced by Simon Haykin (1999).
- I.3 Nonlinear model of a neuron, labeled k. Reproduced by Simon Haykin (1999).
- I.4 Network architectures. (a)-(b) feedforward network with a single layer of neurons; fully connected feedforward network with one hidden layer and one output layer, as described by Simon Haykin (1999).
- I.5 Convolutional Neural Network (CNN).
- I.6 Recurrent Neural Networks. (a)-(b), by Simon Haykin (1999).
- I.7 General structure of the GAN training process.
- 2.1 Topography. (a)-(b) ERA5; VHR_REA-IT.
- 2.2 Spatial distribution of daily mean temperature. (a)-(e) SCIA-ISPRA; E-OBS; UERRA; ERA5 and VHR-REA_IT over 1989–2020 (except for UERRA covering the period 1989–2019). Reproduced by Adinolfi M., et al., 2023.
- 2.3 Spatial distribution of daily total precipitation. (a)-(e) SCIA-ISPRA; E-OBS; UERRA; ERA5 and VHR-REA_IT over 1989–2020 (except for UERRA covering the period 1989–2019). Reproduced by Adinolfi M., et al., 2023.
- 2.1 Normalization (a) and denormalization (b) approaches with different target datasets.
- 2.5 Structure of the developed cGAN, adapted from the general architecture of the Generative Adversarial Network (GAN) for the training process in Figure I.7.
- 2.6 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.

- 3.I Temporal evolution of the 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 artificially data (ERA5-DownGAN), created by the cGAN.
- 3.2 Temporal evolution of daily differences between the dataset generated by the Conditional Generative Adversarial Network (GAN), the ERA5-DownGAN dataset, and the real VHR_REA-IT dataset.
- 3.3 Correlation evolution of the daily spatial average of the 2m-temperature during the test period. (a)–(c) Real low-resolution data (ERA5), the real high-resolution data (VHR_REA-IT) and the downscaled artificially data (ERA5-DownGAN), created by the GAN.
- 3.4 Maps of time average. (a)–(d) 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 (2001-2005).
- 3.5 Examples of one day from the test data set. (a)–(c) Real low-resolution data (ERA5), the real high-resolution data (VHR_REA-IT) and the downscaled artificial data (ERA5-DownGAN), created by the cGAN.
- 3.6 Scatter diagram for evaluating the correlation between ERA5, VHR-REA-IT and ERA5-DownGAN.
- 3.7 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.
- 3.8 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.
- 3.9 Comparison of the probability density functions (PDF) of the 2m-temperature distribution for the ERA5, VHR_REA-IT, and ERA5-DownGAN datasets.
- 3.I0 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 GAN (ERA5-DownGAN).
- 3.II Comparison of the total precipitation grouped for seasons (TOT_PREC) distribution for real low-resolution data (ERA5) with the gray rectangle, the real high-resolution data (VHR_REA-IT) with the blue rectangle.

-
- 3.I2 Comparison of the total precipitation scaled from -1 and 1 and grouped for seasons (TOT_PREC) distribution for real low-resolution data (ERA5) with the gray rectangle, the real high-resolution data (VHR_REA-IT) with the blue rectangle.
 - 3.I3 Example of one day from the test data set, generated by cGAN in EXP0.
 - 3.I4 Comparison of the total precipitation with logarithmic transformation and scaling from -1 and 1; it is grouped for seasons (TOT_PREC) distribution for real low-resolution data (ERA5) with the gray rectangle, the real high-resolution data (VHR_REA-IT) with the blue rectangle.
 - 3.I5 Examples of three random days from the test data set, generated by cGAN in EXP1.
 - 3.I6 Temporal evolution of the daily moving average of the total precipitation 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 artificially data (ERA5-DownGAN), created by the cGAN in EXP1.
 - 3.I7 Scatter diagram for evaluating the correlation between ERA5, VHR-REA_IT, and ERA5-DownGAN.
 - 3.I8 Comparison of the total precipitation (TOT_PREC) 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 cGAN in EXP1 with a white rectangle for each percentile from the 75th to the 99th.
 - 3.I9 Comparison of seasonal total precipitation (TOT_PREC) 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.
 - 3.I20 Maps of time average. (a)-(d) Mean bias (BIAS), root-mean-square errors (RMSEs), mean absolute errors (MAEs), and correlation coefficients (CORRs), comparing downscaled daily total precipitation (ERA5-DownGAN) with the high-resolution dataset (VHR_REA_IT) for the test period (2001-2005).
 - 3.I21 Frequency distribution of precipitation events across different intensity classes for datasets: VHR_REA-IT (blue), ERA5-DownGAN (grey) and ERA5 (coral). The use of a logarithmic y-axis highlights the occurrence of higher-intensity precipitation events.
 - 3.I22 Comparative analysis of performance metrics (mean, standard deviation, FID, spatial and temporal correlation) for experiments EXP1, EXP2, EXP3_1, and EXP3_2 as a function of epochs.

- 3.23 Temporal evolution of the daily moving average of the total precipitation during the test period, calculated with a 30-day window. Comparison of the mean daily spatial average: benchmark (EXP1) and real high resolution dataset (VHR_REA-IT) with different cGAN configurations.
- 3.24 Comparison of the total precipitation (TOT_PREC) distribution for real high-resolution data (VHR_REA-IT) with the white rectangle, and the downscaled artificial data (ERA5-DownGAN), created by the cGAN in EXP1, EXP2, EXP3_1 and EXP3_2 for each percentile from the 75th to the 99th.
- 3.25 Maps of time average. Mean absolute errors (MAEs), Mean bias (BIAS) and root-mean-square errors (RMSEs), comparing downscaled daily total precipitation for EXP1, EXP2, EXP3_1 and EXP3_2 (ERA5-DownGAN) with the high-resolution dataset (VHR_REA_IT) for the test period (2001-2005).
- 3.26 Comparative analysis of performance metrics (mean, standard deviation, FID, spatial and temporal correlation) for experiments EXP1 and EXP4 as a function of epochs.
- 3.27 Frequency distribution of precipitation events across different intensity classes for datasets: VHR_REA-IT (blue), ERA5-DownGAN EXP4 (green) and ERA5-DownGAN EXP1 (coral). The use of a logarithmic y-axis highlights the occurrence of higher-intensity precipitation events.
- 3.28 Maps of time average. Mean absolute errors (MAEs), Mean bias (BIAS) and root-mean-square errors (RMSEs), comparing downscaled daily total precipitation for EXP1 and EXP4 (ERA5-DownGAN) with the high-resolution dataset (VHR_REA_IT) for the test period (2001-2005).
- 3.1 Domain of the 4-km CONUS404 simulation with land surface elevation height (m). Derived by Rassussen M. L. et al., 2023.
- 4.2 Architectures of (a) generator and (b) discriminator. Size input generator 157×333 , size input discriminator 1359×1015 , 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. The architecture displayed in this figure shows an increased number of neurons in each layer compared to the configuration presented in Fig. 2.6.
- 4.3 Temporal evolution of the daily spatial average of 2m-temperature during the test period. The real low-resolution data (ERA5) is shown in red, the real high-resolution data (WRF) in blue, and the downscaled artificial data generated by the GAN (ERA5-DownGAN) in green. Panel (a) compares ERA5-DownGAN_Italy, while panel (b) compares ERA5-DownGAN_U.S.
- 4.4 Examples of one day from the test data set. (a)–(d) Real low-resolution data (ERA5), the real high-resolution data (WRF), the downscaled artificial data

- ERA5-DownGAN_Italy, and the downscaled artificial data ERA5-DownGAN_U.S.
- 4.5 Correlation maps showing the relationship between the WRF dataset and the downscaled artificial data from the GAN configurations. (a) ERA5-DownGAN_Italy; (b) ERA5-DownGAN_U.S.
 - 4.6 Scatter diagram for evaluating the correlation between WRF and the downscaled artificial data from the GAN configurations. (a) ERA5-DownGAN_Italy; (b) ERA5-DownGAN_U.S.
 - 4.7 Comparison of the probability density functions (PDF) of the 2m-temperature distribution for the ERA5 (green), WRF (blue), and the downscaled artificial data from the GAN configurations (orange). (a) ERA5-DownGAN_Italy; (b) ERA5-DownGAN_U.S.
 - 4.8 Comparison of the 2m-temperature (T_2M) distribution for real low-resolution data (ERA5) with the gray rectangle, the real high-resolution data (WRF) with the blue rectangle, and the downscaled artificial data from the GAN configurations (white). (a) ERA5-DownGAN_Italy; (b) ERA5-DownGAN_U.S.
 - 4.9 Comparison of the 2m-temperature (T_2M) distribution for real low-resolution data (ERA5) with the gray rectangle, the real high-resolution data (WRF) with the blue rectangle, and the downscaled artificial data from the GAN configurations (white). (a) ERA5-DownGAN_Italy; (b) ERA5 DownGAN_U.S.
- A.1 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).
 - A.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.
 - A.3 Map of the distribution of SCIA weather stations for tmin and tmax (2001). Sourced from <https://scia.isprambiente.it/>.
 - A.4 Maps of SCIA observed-gridded temperature dataset. (a) Maximum temperature (tmax); (b) Minimum temperature (tmin) and the derived Mean temperature (c).

-
- A.5 Comparison of 2m-temperature field only over Italian Peninsula on 24th March 2001. SCIA observed-gridded temperature dataset; ERA5-DownGAN; VHR_REA-IT and ERA5.
- A.6 Synoptic analysis at 12:00 UTC from NOAA/NCEP reanalysis on 2nd March 2004. Maps of geopotential at 500 hPa (upper) and temperature (°C) at 850 hPa (lower). Reproduced from (www.meteociel.it).
- A.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.
- A.8 Comparison of 2m-temperature field only over Italian Peninsula on 2nd March 2004. SCIA observed-gridded temperature dataset; ERA5-DownGAN; VHR_REA-IT and ERA5.
- B.1 PDFs for 2019 of T_{max} (a,d), T_{min}(b,e) and T_{mean} (c). Panels (a,b,c) identify the evaluation against ERA5-Land and panels (d,e) identify the evaluation against SCIA. Reproduced by Manco I. et al 2023.
- B.2 PDFs for 2020 of T_{max} (a,d), T_{min}(b,e) and T_{mean} (c). Panels (a,b,c) identify the evaluation against ERA5-Land and panels (d,e) identify the evaluation against SCIA. Reproduced by Manco I. et al 2023.
- B.3 Taylor Diagrams of the daily cumulative precipitation at 2m in WRF, COSMO and ICON evaluated against ERA5-Land (a,c) and SCIA (b,d) in 2019 (left column) and 2020 (right column). Reproduced by Manco I. et al 2023.
- B.4 Map of time-mean wind speed at 10 m for ERA5-Land (a,f), SCIA (b,g), WRF (c,h), COSMO (d,i) and ICON (e,l). Reproduced by Manco I. et al 2023.
- C.1 Map of Mean Precipitation on Extreme Event Days Exceeding the 90th Percentile.
- C.2 Inertia Plot to identify the optimal numbers of clusters.
- C.3 Work-flow of clustering methodology.
- C.4 K-means Clustering 1981-2020.
- C.5 K-means Clustering 1981-2020 (winter).

List of Tables

- 2.I Domain details for LW and HR dataset: geographical boundaries, horizontal resolution, number of grid points.
- 3.I Computational details for the training period (a) and test period (b) on Zeus Supercomputer.
- 3.2 Sensitivity experiments, including scaling methods, logarithmic transformation, normalization scheme, and discriminator update steps.
- 3.3 Descriptive statistics for monthly totals, medians, maximum values, and interquartile ranges (Q1 and Q3) across the ERA5, VHR_REA-IT, and ERA5-DownGAN datasets.
- 3.4 Sensitivity experiments concerning loss weights and d_step parameters.
- 3.5 Computational details for the training period on Juno Supercomputer.
- 4.I Domain details for LW and HR dataset: geographical boundaries, horizontal resolution, number of grid points.
- 4.2 Infrastructure of the Casper Supercomputer (NCAR).
- A.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.
- A.2 Comparison on 2nd 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.
- B.I The sensitivity analysis about: PBL, microphysics, surface physics schemes and geographic projection has allowed to determinate the best configuration of the WRF model, that minimizes the error metrics for the precipitation, wind speed and temperature fields (Sim 9). Reproduced by Manco I. et al 2023.

Abbreviations

GANs	Generative Adversarial Neural Networks
CGANs	Conditional Generative Adversarial Neural Networks
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

*The acronyms cGAN and GAN are used interchangeably throughout the discussion, where relevant to the architecture developed in this study, as both denote the underlying generative method upon which the architecture is based.

I | Aim of the Work

This study explores the application of a new Conditional Generative Adversarial Neural Networks (cGAN) architecture for statistical downscaling, discussing its methodology, advantages, and contributions to refining climate predictions at a finer scale. Inspired by GAN architectures used for image super-resolution, this work introduces a novel approach to enhance the resolution and quality of climatological data visualization. By leveraging a generator-discriminator architecture, these GANs generate high-resolution climate datasets that are indistinguishable from real high-resolution data, offering an efficient means to increase the spatial resolution of climatological datasets. Through adversarial training, the generator network learns intricate spatial features and patterns intrinsic to climate data, producing more detailed representations. The application of GANs for super-resolution in climatology represents a cutting-edge advancement, particularly for understanding complex climatic processes. This study evaluates the robustness and versatility of a newly developed cGAN-based statistical downscaling model across varied applications. The initial implementation focuses on downscaling the 2-meter temperature field for a test period (01/2001–12/2005), starting from low-resolution data (ERA5, ~31 km) to obtain a new high-resolution dataset (~2.2 km) consistent with the high-resolution training dataset (VHR_REA-IT). The computational domain for this application includes the Italian Peninsula, parts of Northern and Central Europe, and Northern Africa. In a second application, the model's performance was evaluated by downscaling a more complex variable, cumulative total precipitation, over the same geographic region. To further assess the model's flexibility for other geographical domains, it was applied for the downscaling of 2m-temperature centered on the United States territory. This investigation was born during an internship with NCAR (National Center for Atmospheric Research). This implicated downscaling from ERA5's low resolution (~31 km) to 4 km, achieving a resolution aligned with the high-resolution WRF-NCAR dataset used in training. Additionally, the portability of this downscaling architecture was confirmed across different high-performance computing systems at CMCC (Euro-Mediterranean Center on Climate Change), and NCAR.

These applications collectively underscore the effectiveness of the developed methodology and lay a foundation for future studies aimed at scaling to additional atmospheric variables and extended simulation periods, enhancing precision in climate prediction across diverse scales and regions.

The thesis is organized as follows:

Chapter 1 introduces the fundamentals of Generative Adversarial Networks (GANs), encompassing the evolution of artificial intelligence, neural networks, and the various types of GAN architectures, with a particular emphasis on Conditional GANs (cGANs).

Chapter 2 details the design and implementation of the ERA5-DownGAN for super-resolution in atmospheric fields, covering dataset preparation, cGAN architecture, and training procedures.

Chapter 3 applies the developed cGAN to downscale 2-meter temperature and total precipitation data for the Italian Peninsula, comparing results with traditional dynamical downscaling methods

Chapter 4 demonstrates the flexibility of the developed cGAN for applications in other domains and with different datasets, specifically extending the cGAN application to the United States in collaboration with NCAR, and evaluating its performance in downscaling using the ERA5 and WRF dataset

Chapter 5 provides final considerations and discussion, highlights open questions, and indicates directions for future research.

I.I | Introduction

Statistical downscaling techniques represent alternative or complementary methods to the use of dynamic downscaling models to increase the spatial-temporal resolution of global climate models. Climate change impacts and adaptation applications often require increasing spatial resolution, making it a priority to close the existing gap between climate data produced by General Circulation Models (GCMs), which typically operate at resolutions that are too coarse (often 100-200 km), and climate information at local levels. The outcomes derived from downscaling techniques, whether dynamic or statistical, play a crucial role in developing a comprehensive understanding of the localized impacts of climate variability and change, particularly in the context of extreme climate events. Local climate data is indispensable for enhancing risk assessments and formulating adaptation strategies across various sectors (Kondrup C., et al., 2022). The application of downscaled climate data is crucial for multiple sectors. In agriculture, downscaled precipitation and temperature projections can inform crop management strategies and irrigation planning under future climate scenarios (Xiao D. et al., 2021). In hydrology, accurate downscaling of precipitation data is essential for assessing water resource availability and flood risks, as coarse-resolution models often misrepresent precipitation extremes (Sun Q., et al., 2020). Furthermore, downscaled climate projections are increasingly used in ecological modeling to evaluate the potential impacts of climate change on biodiversity and habitat suitability (Tabor K. and Williams J.W., 2010; Zhang F. & Georgakakos A.P., 2012).

Statistical downscaling is more cost-effective and flexible compared to dynamic downscaling, allowing for faster implementation in new areas or regions. A review of current statistical downscaling techniques includes Empirical Statistical Downscaling (ESD), which maps empirical statistical relationships between large-scale climate drivers and local meteorological variables based on historical observational records (Benestad, R., et al., 2008). Various mathematical approaches, including linear regression, principal component analysis (PCA), and multiple linear regression, are applied in ESD. Weather generators (Yin, S., and Chen, D., 2020) also contribute to the field by generating synthetic meteorological data informed by statistical attributes from historical climate observations. Quantile mapping (Panofsky, H.W., and Brier G.W., 1968) adjusts the probability distributions of coarse climate model outputs using empirical observations. Additionally, Bayesian statistical downscaling techniques (Liu, J.,

Yuan, et al., 2015) combine statistical methodologies with Bayesian principles to estimate uncertainties due to the downscaling process.

Generative Adversarial Networks (GANs) offer a promising alternative by addressing many limitations of traditional methods. GANs learn directly from training data without making specific assumptions about the form of relationships or the stationarity of climatic relationships. This flexibility enables GANs to generate synthetic data that preserves the statistical characteristics of the training data (Leinonen, J. et al., 2021), adapting to complex and non-linear time variations. By avoiding rigid assumptions, GANs are well-suited to handle scenarios of climate change, capturing the complexity of meteorological phenomena, especially extreme weather conditions or non-linear climate changes. GANs represent an innovative method in the statistical downscaling landscape. Unlike Bayesian approaches, which require substantial data and specific assumptions, GANs operate in an unsupervised learning context, making them more adaptable to situations with limited labeled data. GANs do not rely on assumptions of linearity or stationarity, offering a more flexible downscaling process compared to quantile mapping, which applies fixed statistical rules and may struggle with changes in distribution shapes or extreme events. GANs address these limitations by capturing complex relationships and mitigating challenges associated with fixed-statistics methods. Beyond traditional statistical downscaling methods, Generative Adversarial Networks (GANs), particularly Super-resolution GANs (SR-GANs), have demonstrated superior performance in different applications compared to other deep learning-based approaches such as DeepSD, ConvLSTM, and U-NET. As highlighted by Kumar B., et al. 2023, SR-GANs exhibited higher correlation scores and more effectively captured fine-scale precipitation structures. According to the authors, among the techniques analysed, SR-GANs emerged as the most data-driven approach for precipitation downscaling. Despite their advantages, GANs face challenges such as training instability and computational complexity. These issues present opportunities to enhance the robustness and practical application of GANs. As a result, GANs are emerging as promising tools in the climatic context, playing a significant role in the machine learning landscape.

I.2 The Role of Artificial Intelligence in Weather and Climate

Atmospheric sciences, like many other disciplines that rely on computational methods, have been swept up in one of the greatest revolutions of our era: the development of a new set of sophisticated algorithms collectively known as Artificial Intelligence. While the origins of

such algorithms can be traced back to the mid-20th century, notably with the 1956 summer conference at Dartmouth College in Hanover, New Hampshire, as recounted by Mamalakis A. et al. (2022), their use became increasingly established from the 2000s onward, driven by the rise in computational power and the advancement of techniques such as machine learning and deep neural networks (deep learning).

Artificial intelligence (AI) classification has evolved significantly since its inception, reflecting advancements in both theoretical and practical dimensions. Initially, Russell and Norvig (1995) provided a framework for categorizing AI based on two axes: one distinguishing between thinking and acting, and the other contrasting human imitation with ideal rationality. This classification system highlights fundamental approaches to AI, including the explicit and implicit categories. The concept of explicit AI, or symbolic AI, is rooted in early work by Allen Newell and Herbert A. Simon, particularly their influential publication “Human Problem Solving”, (1972). This approach, characterized by rule-based systems and logical structures, laid the groundwork for what became known as expert systems. On the other hand, implicit AI, which encompasses modern machine learning techniques, emerged from the pioneering research of Geoffrey Hinton and Yoshua Bengio. Their seminal paper “Learning Representations by Back-Propagating Errors” (1986) marked a significant advancement in neural network approaches, underscoring the shift towards data-driven methodologies that abstract complex patterns from large datasets.

It is precisely from machine learning systems that the fields of climatology and meteorology have sought to mitigate the computational costs associated with the increasing need to improve the representation of sub-grid atmospheric and oceanic processes and to enhance model resolution. The application of machine learning in weather and climate science is remarkably versatile, encompassing everything from substituting specific sub-components of conventional models to improve their accuracy and reduce computational requirements to the complete replacement of the entire numerical modeling framework (de Burgh-Day and Leeuwenburg 2023). The versatility of machine learning algorithms allows for their application across various domains, including forecasting, data assimilation, downscaling, post-processing of model outputs, and decision support tools. For instance, in weather forecasting, machine learning algorithms have been employed to enhance precipitation predictions. A notable example is the use of convolutional neural networks (CNNs) to analyze historical weather data and improve short-term precipitation forecasts, as demonstrated by Wang J et al. (2024). This approach allows for the extraction of spatial patterns from large datasets, leading to more accurate

predictions compared to traditional methods. Similarly, AI techniques have been utilized in the post-processing of numerical weather prediction outputs to correct biases and improve forecast reliability, as highlighted by Haupt S. E. et al. (2021). In the realm of climate modeling, AI has shown promise in downscaling coarse-resolution climate model outputs to finer spatial resolutions. For example, Baño-Medina J. et al. (2020) explored various deep learning configurations for statistical downscaling, demonstrating that machine learning can effectively capture local climate variations that are often missed by traditional models. Moreover, AI has been applied to enhance the understanding of complex climate phenomena, such as the interactions between the Antarctic Ice Sheet and global sea-level rise. Sadai S. et al. (2020) utilized machine learning to analyze the future responses of the Antarctic Ice Sheet to anthropogenic warming, providing insights into potential climate feedback mechanisms.

Reichstein M. et al. (2019) utilized deep learning techniques to classify extreme weather patterns from climate simulation data, demonstrating how AI can effectively identify and categorize complex atmospheric phenomena. This clustering capability is essential for understanding the frequency and intensity of extreme weather events, which are increasingly relevant in the context of climate change. Furthermore, machine learning has shown promise in enhancing the representation of cloud processes in climate models. Barnes E.A. et al. (2019) discussed how AI can autonomously detect complex patterns in cloud formation and behavior, which are critical for accurate climate modeling. In conclusion, the integration of artificial intelligence in weather and climate science represents a transformative shift in how atmospheric research is conducted. By leveraging machine learning and deep learning techniques, researchers can improve predictive accuracy, address biases in traditional models, and enhance the understanding of complex atmospheric processes. As AI continues to evolve, its role in advancing weather and climate science is likely to expand, offering new opportunities for addressing the challenges posed by climate change and extreme weather events. These applications underscore the ability of AI to process vast amounts of data and identify critical patterns that inform climate science and the potential of AI to improve the fidelity of climate models, particularly in representing processes that are often poorly understood.

To fully grasp the application of artificial intelligence for downscaling in this study, a detailed overview of the use of machine learning techniques in climate and meteorology, specifically related to this topic, will be presented.

I.3 Machine Learning and Deep Learning for Statistical Downscaling

Computational efficiency, versatility, and the ability to learn complex relationships from data make the statistical approach a valuable downscaling method, especially when implemented through machine learning algorithms. Before providing a precise classification of the different statistical downscaling approaches, it is important to understand that, regardless of the specific application, most methods establish an empirical link between a set of predictors x and the predictands y through a statistical model $F(\cdot)$,

$$y \sim F(x).$$

The predictands commonly describe local or regional-scale surface weather or climate, whereas the predictors are generally of larger scale (Maraun D. and Widmann M., 2018).

From this formulation, it is evident that the choice of the downscaling function $F(x)$, the predictors x , and the predictands y results in different downscaling methods. There are various categorizations of downscaling methods (Klein W. and Glahn H. 1974; Wilby R.L. and Wigley T.M.L. 1997; Maraun D. and Widmann M. 2018), but two main subsets are common to divide the empirical downscaling strategies: observational downscaling (or empirical statistical downscaling) and RCM emulation (Fig. I.I). Observational downscaling (or empirical statistical downscaling) involves the development of a model trained on real-world observational data. In contrast, an RCM emulator aims to replicate the functionality of a physics-based RCM and can be trained on simulations of both historical and future climates. Although these two strategies produce downscaling outputs with different characteristics, recent years have seen an increasing use of modern machine learning algorithms (including deep learning) in both approaches, often replacing traditional statistical methods.

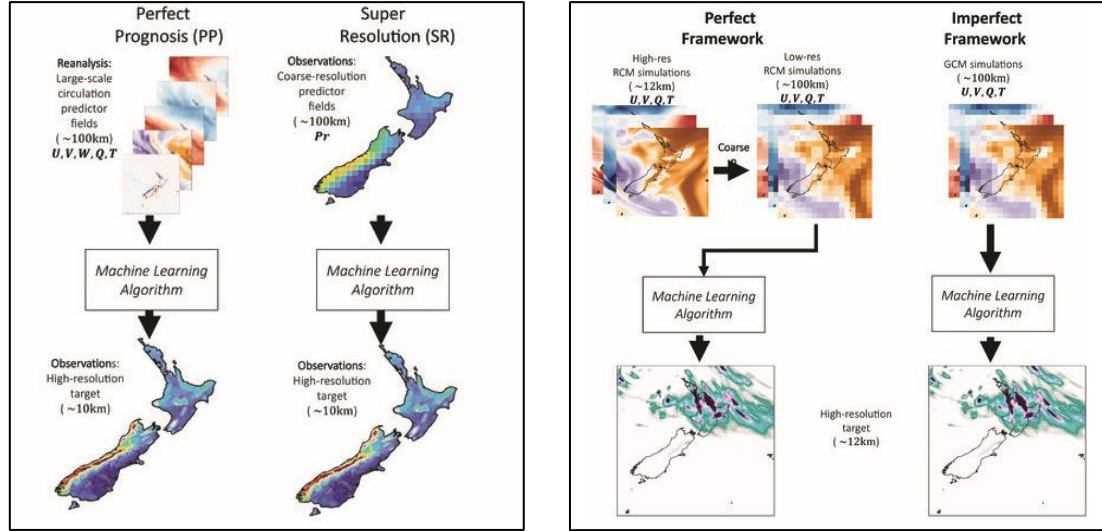


Fig. 1.1: (a) Comparison of PP (left) and SR (right) approaches for climate downscaling in New Zealand: PP maps large-scale fields to high resolution, while SR maps coarsened fields to high resolution. (b) Comparison between perfect and imperfect RCM training frameworks, with perfect using coarsened RCM fields and imperfect using GCM fields. Adapted from Rampal N. et al. (2024).

According to Rampal N. et al. (2024), one of the most recent reviews on the statistical downscaling topic, empirical downscaling approaches can be classified into four main categories: Perfect Prognosis (PP), Super-Resolution (SR), Weather Generators (WGs), and Model Output Statistics (MOS). The PP (Perfect Prognosis) statistical downscaling methods are so named because the model is trained entirely using real-world data, with both predictors and predictands sourced from observations (or observational proxies such as reanalysis data). This approach allows the model to be grounded in data that closely represents physical reality. Once trained, the model is applied to predictors generated by a GCM (Global Circulation Model) to simulate high-resolution predictands. While this methodology has been widely adopted, its performance depends heavily on the choice of the GCM that provides the predictors. PP methods are inherently unable to correct for biases present in the GCM, which may lead to distorted downscaling results. A specific application of Perfect Prognosis (PP) is Super-Resolution (SR), where a low-resolution surface field is used on its own to predict a high-resolution counterpart.

Model Output Statistics (MOS) gets its name from the way it uses statistical linking functions, identified during the calibration phase, to directly adjust the output fields of GCMs already at that stage. This approach allows biases in climate models to be structurally considered and corrected as early as the calibration process. A substantial difference between MOS and other techniques such as Perfect Prognosis (PP) concerns the applicability of the corrections: in the

case of MOS, adjustments are made on a GCM-specific basis, making them non-transferable between different models. As a result, PP is more flexible and generalisable, and more in line with the dynamic Regional Climate Models (RCMs) approach (Rampal N. et al, 2024). Another distinctive aspect is that while in the PP the predictors and predictors, both derived from observations, are temporally synchronised during the calibration phase, in the MOS this calibration does not occur in a coupled manner. This is because climate simulations are not synchronised with real observations. Consequently, the calibration in the MOS is usually based on long-term distributions (distributional calibration), which is only meaningful when predictors and predictors share the same physical dimension, a limitation not found in PP that allows for heterogeneous predictor/predictor pairs. In the climate context, therefore, the MOS almost always takes the form of a simple correction of model biases (e.g. empirical quantile mapping), rather than a calibration based on a precise temporal correspondence between simulations and observations.

Weather Generators (WG) rely on stochastic statistical algorithms, in contrast to PP, SR, and MOS, which employ a deterministic approach. The WG are designed to replicate the distribution and temporal dependence of meteorological variables, introducing variability and randomness into the simulations, even when starting from the same input data. They create synthetic meteorological scenarios based on the statistical properties of the data itself, aiming for precise replication of spatiotemporal dynamics, natural variability of the variables of interest, and correlation structures.

Despite their promising computational efficiency, Empirical Statistical Downscaling (ESD) methods are not without limitations. Firstly, ESD methods are constrained by the availability of long-term observational data, which limits their ability to produce high-resolution data only for regions and variables with sufficient local observations. ESD methods are also constrained in their selection of GCMs due to the requirement for specific variables at particular pressure levels and with a specific temporal frequency for effective calibration.

Additionally, ESD methods rely on the assumption of stationarity in the relationship between large-scale and local scales, implying that a statistical model calibrated on past and present climate conditions remains reliable for future climate scenarios (Doury A. et al. 2022). Numerous studies have demonstrated that this assumption can significantly impact results (Wilby R.L. et al. 1998; Dayon G. et al. 2015; Erlandsen H. et al. 2020). To address the high computational cost of Dynamic Downscaling (DD) and the limitations of ESD, hybrid models

have been introduced, combining the advantages of dynamic and statistical downscaling to address their respective shortcomings, known as Regional Climate Model (RCM) emulators. Initially defined by Maurun and Widmann (2018), these emulators use predictors from simulations by a low-resolution RCM-GCM and predictands from simulations by a high-resolution RCM. Thus, the final high-resolution output obtained from an RCM emulator is constrained only by the resolution of the RCM output, and the performance can be at most comparable to that of RCMs, as this methodology cannot overcome the biases inherent in RCMs (Giorgi F. et al. 2009).

Emulators of Regional Climate Models (RCMs) operate without observational constraints, allowing them to be trained on simulations from both historical and future periods. These emulators can be categorized into two frameworks based on their training methodology: "imperfect" and "perfect." In the "imperfect" framework, an emulator is trained to directly map outputs from GCMs to those of RCMs. Conversely, in the "perfect" framework, the process begins with an upscaling of the RCM resolution to match that of the GCM, followed by the training of 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 relationship specific to each GCM-RCM pair, thus functioning as a Model Output Statistics (MOS) technique. In contrast, the perfect model framework focuses on learning general relationships between fields at different resolutions (Boé J. et al., 2023). This framework typically deals with weak correlations and a degree of "independence" between RCM and/or GCM fields, aligning it with the Perfect Prognosis (PP) approach. Training within the perfect framework is more aligned with PP because 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 training compared to the imperfect framework (Rampal et al., 2024).

Empirical Statistical Downscaling (ESD) and RCM emulators share some of the most widely used ML algorithms (including DL), 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 forest (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 perceptron (e.g., Nishant N. et al., 2023 for use in RCM emulators and Hobeichi S. et al., 2023 for PP), or short-term memory (e.g., Bittner M. et al., 2023 for RCM emulators and Hobeichi S. et al., 2023 for ESD).

The difference between the two strategies lies in how these techniques are employed, what inputs are passed on, and what outputs are desired. For instance, traditional PP methods may not be well-suited 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. Furthermore, these methods can be computationally inefficient, as the algorithm must be repeatedly applied, point by point, across large datasets. RCM emulators, by receiving one or more fields as input over the entire spatial grid, generate high-resolution output fields for the entire region covered by the RCM. Recent years have demonstrated significant potential in applying computer vision algorithms to the fields of meteorology and climatology, particularly in downscaling tasks. Notably, generative models such as Generative Adversarial Networks (GANs) have been extensively studied, with Leinonen J. et al. (2021) providing a comprehensive description of their application. Additionally, diffusion models, as detailed by Mardani M. et al. (2023), have shown promising results. These models utilize various architectures, including Convolutional Neural Networks (CNNs), U-Net, and Fully Connected Networks (FCNs) or Multilayer Perceptrons (MLPs). The complexity of these models involves compromises between the computational costs, the horizontal resolution and, in some cases, the domain size (Doury A. et al. 2022).

I.4 Artificial Neural Networks (ANNs)

The architecture underlying the various Deep Learning (DL) algorithms employed for downscaling in weather and climate modeling are rooted in the framework of artificial neural networks (ANNs), commonly referred to as neural networks (NNs). These architectures are particularly suited for representing sub-grid scale processes and are therefore discussed in detail in this section. The development of neural networks was driven by the need to overcome the limitations in capacity and processing speed of traditional digital computers compared to the human brain, which, through its network of neurons, can perform highly complex and non-linear tasks, such as pattern recognition and perception, within mere *ms*. Neural networks are designed to emulate the functioning of the human brain, with several terminologies shared between the two. In the human brain, learning occurs through experience, whereas in artificial systems, this is achieved through a learning algorithm, which enables NNs to improve their performance adaptively. The basic structural components of a neural network, known as “neurons” or “processing units,” are extensively interconnected via synaptic weights. These weights encode the acquired knowledge and are adjusted by the learning algorithm to enhance the network’s performance. Neurons in NNs are organized into multiple interconnected layers

that define relationships between different parameters (Fig. I.2), thereby removing the need for explicit parameterization or numerical models to encode such relationships. Consequently, neural networks serve as universal function approximators (Lu L. et al., 2019).

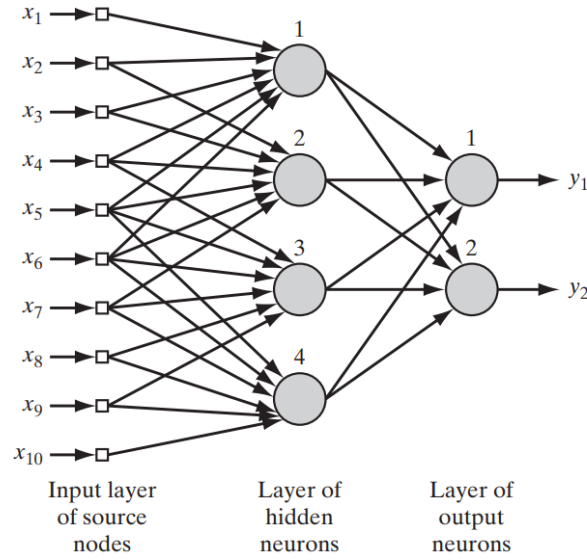


Fig. I.2: Example of a fully connected Neural Network. Reproduced by Simon Haykin (1999).

According to the definition provided by Simon Haykin (1999),

“a neural network is a massively parallel distributed processor consisting of simple processing units, with an inherent ability to store experiential knowledge and make it available for application.”

An important task for a neural network is to learn a model of the environment in which it operates and to maintain the model in sufficient coherence with the real world to achieve the specified objectives of the application. In real-world applications of "intelligent" machines, it can be asserted that a good solution hinges on an effective representation of knowledge (Wood J. and Shawe-Taylor, J., 1996). According to Fischler M. and Firschein O. (1987), this representation pertains to the information or models stored and utilized, generally by either a person or a machine, to interpret, predict, and respond appropriately to the external world.

The advantages associated with the use of neural networks are manifold, including their ability to model nonlinear processes and generalize learned patterns to previously unseen data. The capability to represent nonlinearity is crucial, particularly when the underlying physical mechanism responsible for generating the input signals exhibits intrinsically nonlinear

characteristics. This is especially pertinent in the context of weather and climate, where the phenomena involved are often highly complex and nonlinear. Another advantage of neural networks is their ability to incorporate an evidential response, which provides a confidence level along with the network's output. In a context where the difficulty of interpreting the complex processes within machine learning models and the occasional lack of transparency can undermine user trust, the evidential response allows for more informed and robust decision-making. By managing ambiguity and enhancing the reliability of the results provided by the algorithms, this feature addresses critical challenges in leveraging these advanced models effectively.

I.4.I Neurons and Activation Functions

The configuration of neurons within a neural network enables the formation of various architectures. To fully appreciate these architectures, it is essential first to understand the characteristics of a neuron, its constituent components, and its mathematical formulation. The simplest instance of a neural network comprises a single input, a single neuron, and a basic activation function. Mathematically, the operation of an individual neuron k , as illustrated in Figure I.3, can be characterized by examining the fundamental components that define its model:

- **Input:** a neuron receives incoming signals from the preceding layer of neurons or directly from raw data inputs. These signals are represented as a vector of numerical values $x_1, x_2, \dots, x_m$, where each value x_j corresponds to a specific input data feature.
- **Synaptic Weights:** each input x_j is associated with a weight w_{kj} , which signifies the relative importance of that input to the neuron. The weights w_{kj} are parameters learned during the model training phase and can either amplify or attenuate the input signals.
- **Bias:** the bias b_k is a constant additive term that allows for the shifting of the activation function, thereby enhancing the model's flexibility. Depending on the sign of the bias, the relationship between the induced local field (or activation potential) v_k of neuron k and the output of the linear combiner $\sum_{j=1}^m w_{kj}x_j$ is modified. The bias enables the neuron to produce an output even when the inputs are zero.
- **Activation Function:** after processing the weighted inputs, the neuron applies an activation function $\varphi(.)$ to the result. This function introduces non-linearity into the model, enabling the network to learn more complex relationships between the inputs and the

output. Examples of activation functions include the sigmoid, ReLU (Rectified Linear Unit), and tanh.

- **Output:** The output of a neuron, denoted as y_k , represents the value predicted by the neural network. This output can either be a continuous value in the context of regression problems or a probability in classification tasks. The final value is obtained by applying the activation function to the weighted input of the neuron.

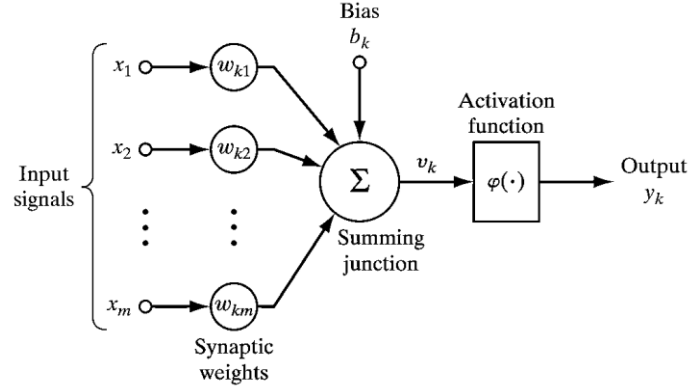


Fig. 1.3: Nonlinear model of a neuron, labeled k . Reproduced by Simon Haykin (1999).

Formally, the output y of a neuron is obtained through two main steps:

1. **Weighted sum of inputs:** Each input x_i is multiplied by its corresponding weight w_{ki} , and the bias b_k is added to the total sum. This operation can be expressed as:

$$v_k = \sum_{j=1}^m w_{kj} x_j + b_k$$

2. **Application of the Activation Function:** The result of the weighted sum z is then passed through an activation function $\varphi(\cdot)$, which transforms the neuron's output. This step is crucial for introducing non-linearity into the system, enabling the network to solve real-world problems. The final output of the neuron y is therefore given by:

$$y_k = \varphi(v_k) = \varphi\left(\sum_{j=1}^m w_{kj} x_j + b_k\right)$$

Activation functions are integral components of neural network architectures, as they facilitate the introduction of non-linearity into the model. This non-linearity enables the network to learn and accurately represent intricate relationships within the data. Mathematically, an activation function operates on the weighted sum of inputs directed toward a neuron, yielding an output signal. This output signal is pivotal in determining the activation status of the neuron and the degree of its response. In the absence of activation functions, a neural network would fundamentally reduce to a linear regression model, thereby lacking the capacity to capture complex patterns and non-linear relationships inherent in real-world datasets.

Beyond their primary function of introducing non-linearity, activation functions also serve the critical role of normalizing the output of individual neurons. This normalization process ensures that neuron outputs remain confined within a specified range, commonly between 0 and 1 or -1 and 1. Such normalization is crucial for stabilizing the learning process and mitigating issues related to the exploding or vanishing gradients, particularly in deeper network architectures. Prominent activation functions employed for this purpose include the sigmoid function, hyperbolic tangent (tanh), and softmax functions.

The distinctive characteristics of each activation function render them suitable for specific tasks within neural network architectures. Commonly utilized activation functions include:

- **Sigmoid Function:** The sigmoid function is widely regarded as one of the most fundamental activation functions in the architecture of neural networks. Characterized by its strictly monotonic increasing nature, the sigmoid function elegantly balances linear and nonlinear characteristics. A prominent example of this function is the logistic function, mathematically represented as:

$$\varphi(v) = \frac{1}{1 + \exp(-av)}$$

where a denotes the slope parameter of the sigmoid function. By manipulating the value of a , it is possible to generate sigmoid functions with varying degrees of steepness. As the slope parameter a approaches infinity, the sigmoid function asymptotically behaves like a step function. While a step function yields discrete outputs of either 0 or 1, the sigmoid function provides a continuous range of outputs between 0 and 1. This property renders the sigmoid function particularly advantageous for binary classification tasks, where the objective is to assign inputs to one of two distinct categories. However, it is crucial to acknowledge that

sigmoid functions are susceptible to the vanishing gradient problem, a limitation that can significantly impede the training process in deep neural networks.

- **Tanh:** The hyperbolic tangent function, or tanh, maps the input to a value between -1 and 1. It is an improvement over the sigmoid function because it is zero-centered, which helps the network learn more efficiently.

$$\varphi(v) = \tanh(v)$$

Activation functions are integral components of neural network architectures, as they facilitate the introduction of non-linearity into the model. This non-linearity enables the network to learn and accurately represent intricate relationships within the data. Mathematically, an activation function operates on the weighted sum of inputs directed toward a neuron, yielding an output signal. This output signal is pivotal in determining the activation status of the neuron and the degree of its response. In the absence of activation functions, a neural network would fundamentally reduce to a linear regression model, thereby lacking the capacity to capture complex patterns and non-linear relationships inherent in real-world datasets.

Beyond their primary function of introducing non-linearity, activation functions also serve the critical role of normalizing the output of individual neurons. This normalization process ensures that neuron outputs remain confined within a specified range, commonly between 0 and 1 or -1 and 1. Such normalization is crucial for stabilizing the learning process and mitigating issues related to the exploding or vanishing gradients, particularly in deeper network architectures. Prominent activation functions employed for this purpose include the sigmoid function, hyperbolic tangent (tanh), and softmax functions.

The distinctive characteristics of each activation function render them suitable for specific tasks within neural network architectures. Commonly utilized activation functions include:

- **Sigmoid Function:** The sigmoid function is widely regarded as one of the most fundamental activation functions in the architecture of neural networks. Characterized by its strictly monotonic increasing nature, the sigmoid function elegantly balances linear and nonlinear characteristics. A prominent example of this function is the logistic function, mathematically represented as:

$$\varphi(v) = \frac{1}{1 + \exp(-av)}$$

where a denotes the slope parameter of the sigmoid function. By manipulating the value of a , it is possible to generate sigmoid functions with varying degrees of steepness. As the slope parameter a approaches infinity, the sigmoid function asymptotically behaves like a step function. While a step function yields discrete outputs of either 0 or 1, the sigmoid function provides a continuous range of outputs between 0 and 1. This property renders the sigmoid function particularly advantageous for binary classification tasks, where the objective is to assign inputs to one of two distinct categories. However, it is crucial to acknowledge that sigmoid functions are susceptible to the vanishing gradient problem, a limitation that can significantly impede the training process in deep neural networks.

- **Tanh:** The hyperbolic tangent function, or tanh, maps the input to a value between -1 and 1. It is an improvement over the sigmoid function because it is zero-centered, which helps the network learn more efficiently.

$$\varphi(v) = \tanh(v)$$

- **ReLU (Rectified Linear Unit):** is a popular activation function that sets negative inputs to zero while leaving positive inputs unchanged. It has been widely adopted due to its simplicity and its ability to mitigate the vanishing gradient problem, which can occur in deep networks during training. However, ReLU suffers from the "dying ReLU" problem, where neurons can become inactive and stop learning, typically when their input values are consistently negative.

$$\varphi(v) = \max(0, v)$$

- **Leaky ReLU:** Leaky ReLU is a variant of ReLU designed to address the "dying ReLU" issue by allowing a small, non-zero gradient when the input is negative. This small slope a for negative inputs prevents neurons from becoming inactive and allows learning to continue, even for negative inputs.

$$\varphi(v) = \begin{cases} v & \text{if } v \geq 0 \\ av & \text{if } v \leq 0 \end{cases}$$

- **Softmax:** is commonly used in multi-class classification problems. It converts the raw outputs (logits) of a neural network into a probability distribution over multiple classes, ensuring that the sum of the probabilities for all classes equals 1.

$$\varphi(v_i) = \frac{\exp(v_i)}{\sum_{j=1}^K \exp(v_j)}$$

Where, v_i the raw output (logit) from the neural network for class i and K represents the total number of classes.

Once the network's final output has been obtained from steps 1) and 2), which define the *forward step*, we move on to step 3), known as the *backward step*. The combination of these three steps defines what is commonly referred to as the *backpropagation algorithm* (Rumelhart D., et al., 1986). After the output, during step 3, the error is measured against the expected output (or “ground truth”) using a loss or cost function (for example, the mean squared error or cross-entropy for classification). Specifically, backpropagation computes the gradient of the loss function for each of the individual parameters of the neural network by utilizing the “chain rule” (Child, J. M., 1916). An optimization function (gradient descent method or stochastic gradient descent method) utilizes these gradients to find and update the corresponding weights and biases of each neuron during backpropagation.

By descending along the gradient of the loss function, we minimize the loss function itself. The direction in which we should adjust the weights and biases to reduce the loss is known, as the gradient we calculated during backpropagation contains the partial derivatives for each model parameter.

The weight update can be expressed as:

$$w_{new} = w_{old} - \eta \frac{\partial Loss}{\partial w}$$

Similarly, the updated bias can be formulated as:

$$b_{new} = b_{old} - \eta \frac{\partial Loss}{\partial b}$$

Where η is the *learning rate*, which will be discussed in greater detail in section 2.4.4 (“Training Strategies”).

Thus, the weights are updated in the direction opposite to the gradient to minimize the loss, and the magnitude of the update is determined by the learning rate. The final goal of the process is to minimize the error compared to the expected output and improve the overall prediction accuracy of the network. In conclusion, backpropagation is fundamental to training neural network models, from the most basic multilayer perceptrons to the complex deep neural network architectures used for generative AI.

As previously mentioned, the configuration of neurons in a neural network defines several network architectures (structures).

Following the classification proved by Simon Haykin (1999), there are three fundamentally distinct classes of network architectures can be identified:

- **Single-Layer Feedforward Networks:** in a layered neural network, neurons are organized into distinct levels (Fig. I.4a). In its simplest form, this network comprises an input layer made up of source nodes that connect directly to an output layer containing computation nodes (neurons), with no connections in the opposite direction (feedforward network, FFNN). This architecture is referred to as a single-layer network because only the layers containing computation neurons are counted; the input layer does not perform any processing.
- **Feedforward Multilayer:** is characterized by the presence of one or more hidden layers, where the computation nodes are known as hidden neurons or hidden units (Fig. I.4b), so named because they are not directly visible from the network’s input or output. Hidden neurons serve as intermediaries between the external input and the network’s output, allowing the network to process more complex patterns and extract higher-level features from the input data. In this way, even though the network has local connections, it gains a broader, more global understanding of the problem thanks to the additional synaptic connections and layers of neural interaction (Churchland P. S. and Sejnowski T. J., 1992). The source nodes in the input layer provide the components of the activation pattern (input vector), which are the signals passed to the neurons in the second layer (the first hidden layer). The outputs from the second layer are then fed into

the third layer, and this process continues through the rest of the network. The overall response of the network to the input from the source nodes in the first layer is determined by the signals produced by the neurons in the final layer. This iterative process, where inputs are passed through multiple hidden layers, with each layer's output becoming the input for the next, can be formalized as follows:

$$v_k^{(l)} = \sum_{j=1}^m w_{kj}^{(l)} y_j^{(l-1)} + b_k^{(l)}$$

Where, as previously defined: $w_{kj}^{(l)}$ represents the weight connecting neuron k from layer $l-1$ to neuron j in layer l , $y_j^{(l-1)}$ is the output of neuron j from the previous layer and $b_k^{(l)}$ is the bias of neuron k in layer l .

The output of neuron j in layer l is obtained by applying the activation function:

$$y_k^{(l)} = \varphi(v_k^{(l)})$$

This process is repeated for each neuron in every layer, propagating the information through the network until it reaches the final output layer.

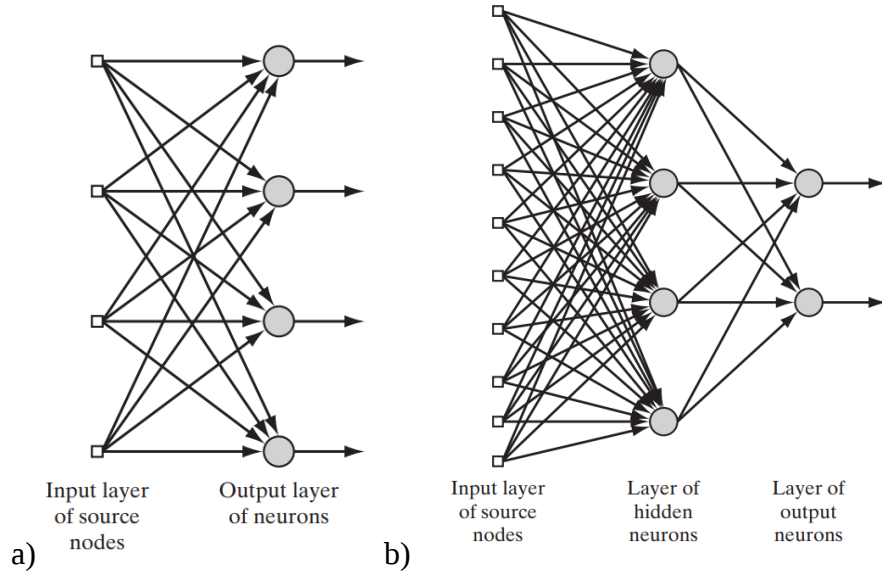


Fig. I.4: Network architectures. (a)-(b) feedforward network with a single layer of neurons; fully connected feedforward network with one hidden layer and one output layer, as described by Simon Haykin (1999).

- **Convolutional Networks (CNNs):** are characterized by the presence of one or more convolutional layers, which play a pivotal role in feature extraction from input data, particularly images. These layers utilize filters, or kernels, that traverse the input image to generate activation maps, thereby highlighting the presence of local features such as edges and corners. The convolution operation involves the multiplication of pixel values by the filter weights, enabling the network to learn meaningful representations of the data (LeCun Y. et al., 1998). Following the convolutional operations (Fig. I.5), the resulting activation maps undergo pooling processes, which reduce their spatial dimensions, thereby simplifying the data and enhancing computational efficiency. Pooling can be performed using various techniques, such as max pooling or average pooling, which select the most salient features while mitigating the risk of overfitting by retaining only the essential information necessary for final classification. Ultimately, the outputs from the convolutional and pooling layers are flattened and passed through one or more fully connected layers. In these layers, the nodes connect the extracted features, allowing the network to perform classification based on high-level combinations of the learned information. The final output is produced via an activation function, typically softmax, which generates probability distributions over the possible classes. This hierarchical structure, transitioning from the extraction of simple features to the identification of complex patterns, enables CNNs to achieve a deep and comprehensive understanding of the data, rendering them particularly effective in applications such as computer vision and object recognition.

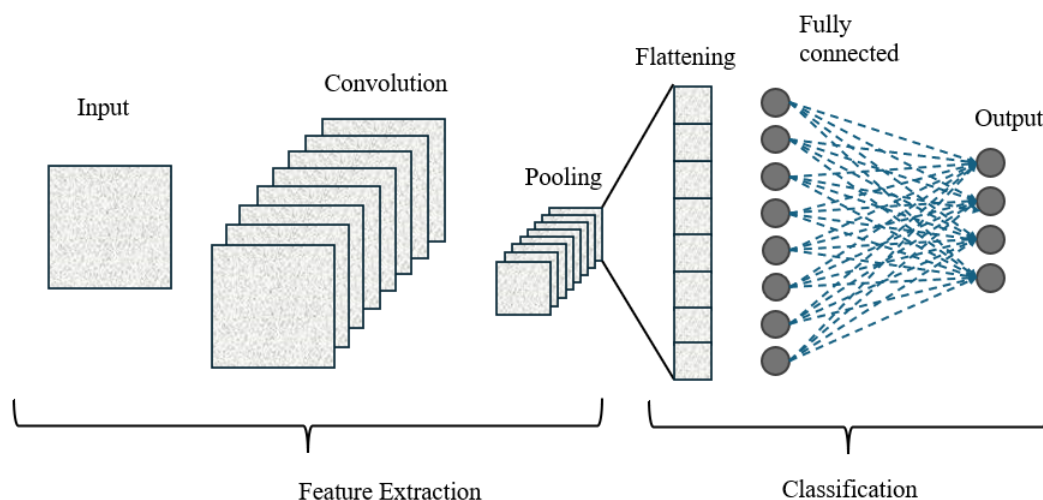


Fig. I.5: Convolutional Neural Network (CNN).

- **Recurrent Networks (RNNs):** Recurrent Neural Networks (RNNs) are characterized by their ability to "remember" information from previous inputs, allowing them to influence both current inputs and outputs through one or more feedback loops (Fig. I.6). Unlike traditional deep neural networks, which assume independence between inputs and outputs, the output of an RNN is dependent on preceding elements within a sequence. Additionally, while feedforward networks assign distinct weight parameters to each node, RNNs share the same weights across each layer. These weights are adjusted through backpropagation and gradient descent, but RNNs utilize Backpropagation Through Time (BPTT) specifically to compute gradients for sequential data. BPTT follows the same principles as standard backpropagation, where the model learns by calculating errors from the output layer back to the input; however, it differs in that it sums the errors at each time step, a requirement that feedforward networks do not have due to their lack of parameter sharing across layers. Information from the output is delayed by one time step before being used as input for the subsequent step, accomplished through the use of unit time-delay elements denoted by z^{-1} . The integration of non-linear units, such as activation functions like ReLU or tanh, introduces complexity to the dynamic behavior of RNNs. As a result, the network's response is not merely proportional to its inputs; rather, it can develop intricate patterns based on past data. This dynamic capability enables RNNs to effectively capture and model temporal dependencies and contextual relationships within sequential data, making them particularly suitable for tasks involving time series or sequential processing. However, this process introduces two potential challenges (IBM, <https://www.ibm.com/it-it/topics/recurrent-neural-networks>): vanishing and exploding gradients. In the case of vanishing gradients, the weight parameters approach zero, leading to stagnation in the learning process. Conversely, exploding gradients result in excessively large weight values, causing instability and yielding NaN values.

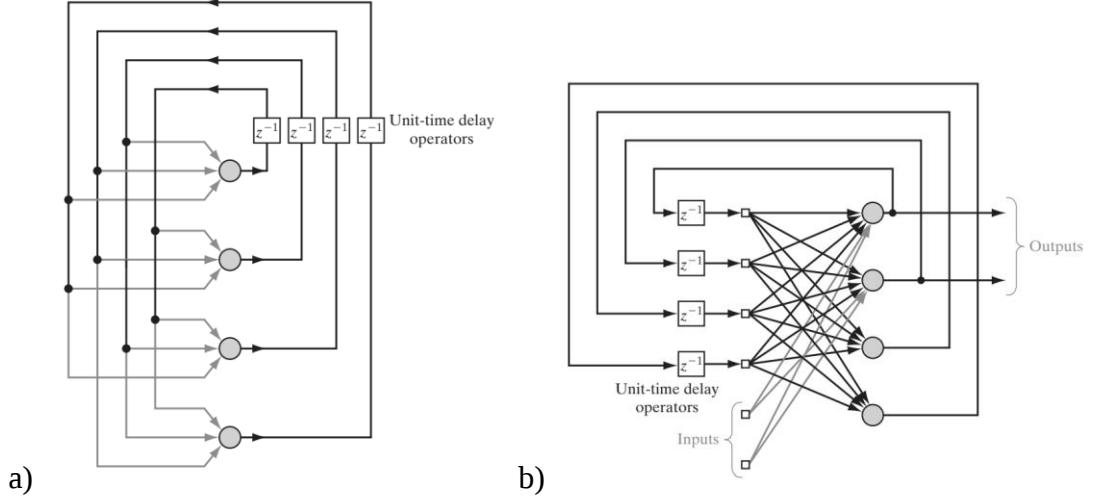


Fig. 1.6: Recurrent Neural Networks. (a)-(b), by Simon Haykin (1999).

A common solution to mitigate these issues involves reducing the number of hidden layers within the network, thereby simplifying the overall architecture and minimizing the risks associated with gradient-related challenges.

This approach closely aligns with the well-established theory of linear adaptive filters, which has been successfully applied across various fields (Widrow B. and Stearns S. D., 1985; Haykin S., 2002). However, it is also possible for a neural network to modify its topology, a concept inspired by the fact that neurons in the human brain can die and new synaptic connections can form.

I.5 Generative Adversarial Neural Networks (GANs)

The ability of Generative Adversarial Networks (GANs), particularly conditional GANs (c-GANs), to capture fine spatial details and extreme events (Mardani M. et al. 2023; Miralles O. et al. 2022; J. Wang J. et al. 2021) has led us to focus this work on developing an innovative c-GAN architecture for super-resolution. The pioneering goal is to achieve high-resolution downscaling of atmospheric variables of interest. Therefore, this section will commence with a brief overview of generative models, placing particular emphasis on Generative Adversarial Networks (GANs). This will be succeeded by an in-depth examination of conditional GANs (c-GANs).

Generative models are algorithms designed to learn the underlying distribution of data and generate new samples that resemble the originals. A generative model can be defined as a

system capable of learning and representing the probability with which data is generated (Goodfellow I. et al., 2016). These models differ from discriminative models, which focus on class recognition, as they aim to capture the intrinsic structure of the data. Among the most well-known generative models are Variational Autoencoders (VAEs), which utilize a probabilistic approach to infer latent representations; however, they may produce blurrier samples due to their approximation nature (Kingma D. P. and Welling M., 2014). Other models, such as Normalizing Flows, offer greater flexibility and generative precision through invertible transformations but require higher computational complexity and can be more challenging to train (Rezende D. J. and Mohamed S., 2015). Autoregressive models, such as PixelCNN and WaveNet, are recognized for their performance in sequential data generation, including images and audio. However, they can be slow in inference and demand significant resources due to their sequential nature (Oord, A. v. d. et al., 2016).

In this landscape of generative models, Goodfellow et al. laid the foundations for Generative Adversarial Networks (GANs) in 2014 with their seminal paper titled "*Generative Adversarial Nets*." The authors need to develop a new generative neural architecture because the most significant advancements in deep learning have predominantly involved discriminative models, employing backpropagation and dropout algorithms while utilizing piecewise linear units that exhibit well-behaved gradients. In contrast, deep generative models have had a relatively limited impact, primarily due to two factors: the complexities associated with approximating intractable probabilistic computations that arise in maximum likelihood estimation and related strategies, and the difficulty of leveraging the benefits of piecewise linear units within the generative context. In response to these challenges, Goodfellow et al. propose a novel procedure for estimating generative models that overcome these obstacles. In the framework of the presented adversarial networks, the generative model G is confronted with an adversary, namely a discriminative model D , which learns to distinguish whether a sample originates from the generated distribution or the real distribution. The training procedure for the generative model G is based on maximizing the probability that the discriminator D makes an error. This approach is structured as a minimax game between two participants (Fig. I.7).

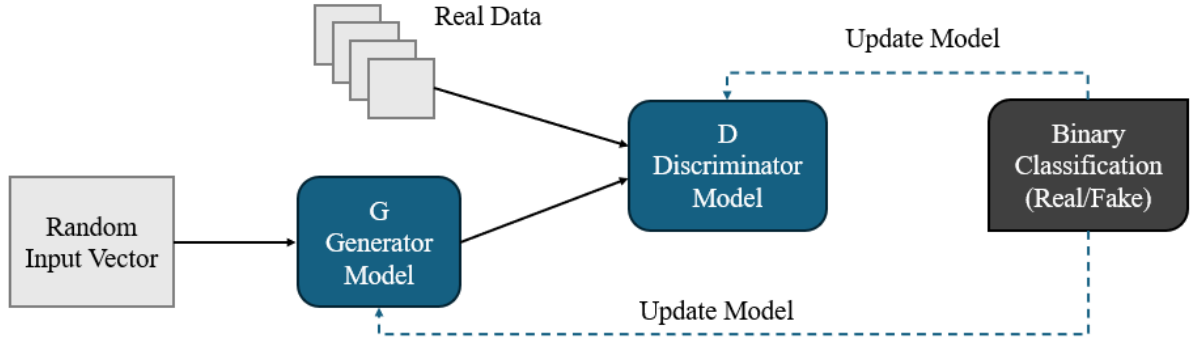


Fig. I.7: General structure of the GAN training process.

Within the space of arbitrary functions G and D , it is demonstrated that a unique solution exists, wherein the generative model G is able to reproduce the training data distribution, while the discriminator D takes on a constant value of $1/2$ at all points. In other words, training concludes when the models G and D reach an equilibrium in which G produces convincing data, and D is no longer able to distinguish between real and generated data.

The following provides a general formulation of the minimax problem that forms the foundation of Generative Adversarial Networks (GANs). In this framework, the generator takes as input a vector of random noise, denoted as z , and maps it to a data distribution that closely approximates real-world data, such as images or audio signals. This transformation aims to produce synthetic samples indistinguishable from authentic data by an associated discriminator model.

This transformation is represented as $G(z; \theta_g)$, where θ_g are the parameters of the generator (i.e., the weights of the neural network). The discriminator receives an input, which can either be real data or generated data and outputs a value that indicates the probability that the input is real. This probability is denoted as $D(x; \theta_d)$, where x is the input data (which may originate from either the generator or the real data), and θ_d are the parameters of the discriminator. The discriminator is trained to maximize its capacity to correctly distinguish real data from fake data. Conversely, the generator seeks to minimize the probability that the discriminator recognizes its outputs as fake. More precisely, it aims to minimize the expression $(1 - D(G(z)))$ which represents the probability that the discriminator makes an error.

This process can be viewed as a game between two players, the generator and the discriminator, each pursuing opposing objectives. This formulation can be expressed as:

$$\min_G \max_D V(D, G) = \mathbb{E}_{\mathbf{x} \sim p_{\text{data}}(\mathbf{x})} [\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim p_z(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))]$$

$D(x)$ represents the probability that x is real, while $G(z)$ denotes the data generated from the noise z .

Generative Adversarial Networks (GANs) offer several distinct advantages over traditional generative models such as Gaussian Mixture Models (GMMs), Hidden Markov Models (HMMs), Autoregressive Models (AMs), and Variational Autoencoders (VAEs). One of the primary advantages of GANs lies in their departure from probabilistic models that explicitly aim to estimate the underlying distribution of the data. Unlike such models, GANs do not attempt to learn an explicit probability density function to describe the data. Instead, the generator in a GAN framework learns to produce samples that exhibit characteristics similar to the real data through iterative feedback from the discriminator.

Moreover, in many traditional learning models, the parameters (i.e., the variables used by the model to make predictions) can retain information about the training data, which can lead to the issue of *overfitting*. Overfitting occurs when a model becomes too finely tuned to the specifics of the training data, resulting in poor generalization to unseen data. GANs mitigate this issue through their adversarial training process. Specifically, the generator does not directly copy or memorize the training data. Rather, it learns to synthesize new data instances by relying solely on the feedback provided by the discriminator, which evaluates how realistic the generated data appear. This indirect learning mechanism ensures that the generator's parameters do not encode a direct representation of the real data, but instead capture a more abstract understanding of how to produce samples that fool the discriminator into recognizing them as authentic. As a result, the generator is less prone to overfitting, as it avoids storing exact representations of the training data and focuses instead on generating plausible new examples.

Another key advantage of GANs is their avoidance of complex mechanisms such as Markov chains, which are often required in probabilistic models for data generation. Markov chains typically involve sequential dependencies that can complicate both model design and computational efficiency. In contrast, GANs rely entirely on *backpropagation*, a well-established and efficient method in neural network training, to compute gradients and update model parameters. This not only simplifies the training process but also enhances computational tractability, as no inference or sampling mechanisms from a Markov chain are required during learning. Additionally, GANs possess a high degree of versatility due to their ability to incorporate a wide range of functions into their architecture. This adaptability allows GANs to generate highly detailed and precise outputs, often surpassing traditional models that may

struggle with producing sharp or complex data samples. Models based on Markov chains or probabilistic approximations, for example, are prone to generating “blurred” or “imprecise” outputs due to their inherent need to maintain smooth transitions between modes in the data distribution. In summary, GANs offer substantial computational and statistical advantages over traditional generative models. Their ability to generate data without relying on explicit probabilistic modeling, coupled with their robust mechanism for avoiding overfitting and their computational efficiency through backpropagation, make them a powerful and flexible tool for generating highly detailed and realistic data.

I.5.I Conditional Generative Adversarial Networks (cGANs)

Conditioning in neural networks is a fundamental concept that enables the directed guidance of the learning process, enhancing both the accuracy and relevance of generated outcomes. It facilitates the integration of auxiliary information, such as class labels or specific characteristics of the dataset, enriching the representation and aiding in the generation of outputs that are coherent with the desired context (Kingma D.P. and Welling M., 2014). Conditioning has been shown to be crucial across a variety of applications, ranging from image generation to speech recognition, allowing models to produce more relevant and informative results (Salimans T. et al., 2016).

Conditional Generative Adversarial Networks (cGANs), first introduced by Mirza M. and Osindero S. (2014) in their seminal paper titled "Conditional Generative Adversarial Nets", represent a significant extension of Generative Adversarial Networks (GANs). This innovative approach enables both the generator and the discriminator to be conditioned on auxiliary information, thereby allowing for direct and precise control over the data generation modes. In cGANs, conditioning occurs through the incorporation of auxiliary variables y , which may include class labels, data from other modalities, or any relevant information about the generation context.

In the generator, the initial noise $p_z(z)$ is combined with the conditioning variable y to create an integrated hidden representation. This merging enables the model to leverage the adversarial training framework, providing significant flexibility in both the construction and application of this representation. Conversely, in the discriminator, both x (representing either real or generated data) and y are presented as inputs to a discriminative function, typically implemented through a multilayer perceptron (MLP). This configuration allows the discriminator to

effectively assess the quality of the generated data concerning the specified conditions, thereby enhancing the generator's ability to produce coherent and contextually relevant samples.

The introduction of conditional GANs (cGANs) has led to numerous advantages over traditional GAN models. By allowing the generation process to be conditioned on auxiliary variables, cGANs offer more control over the generated data, which can be particularly beneficial when the goal is to create data that adheres to specific constraints or characteristics. In climate modeling, this capacity is especially useful for downscaling, where the task involves generating high-resolution climate data from coarser input. Traditional GANs, which generate data based solely on random noise, often struggle to incorporate external information, leading to less accurate or less relevant outputs for specific tasks. In contrast, cGANs can leverage high-resolution observational data or other auxiliary variables to produce more accurate and contextually relevant downscaled outputs (Vandal T. et al., 2017; Leinonen J. et al., 2020).

Furthermore, the ability of cGANs to focus the generation process on specific attributes, such as geographic or climatic features, has demonstrated clear advantages in terms of both fidelity and precision in generated climate data. This makes them particularly suited for applications in downscaling, where preserving spatial and temporal correlations is essential for producing meaningful results. As such, cGANs have shown significant improvements in performance over traditional GANs in climate downscaling tasks, reducing the introduction of noise and artifacts, while better preserving the physical consistency of the data (Stengel K. et al., 2020; Liu Y. et al., 2020).

The demonstrated potential of Generative Adversarial Networks (GANs), particularly Conditional GANs (cGANs), in capturing complex spatial and temporal structures remains underutilized, with only a limited number of studies employing cGANs for downscaling purposes. This study is driven by the pioneering objective of harnessing this potential to develop a conditional generative adversarial network for high-resolution atmospheric variable downscaling that is flexible and versatile, applicable across various geographical domains and variables. This innovative approach aims to advance the field by addressing existing gaps in the literature and showcasing the capabilities of cGANs in generating more accurate and contextually relevant climatic data.

2 | Designing the Super-Resolution c-GAN

This chapter focuses on the development of the ERA5-DownGAN, a Conditional Generative Adversarial Network (cGAN) for super-resolution, designed to enhance the spatial resolution of atmospheric data. We start by defining the goals and importance of super-resolution for atmospheric fields, followed by a description of the GAN architecture. The chapter details the datasets used, including ERA5 and VHR_REA-IT, and explains the preprocessing and normalization techniques applied. We also cover the specifics of the network architecture, including the generator and discriminator components, as well as the training process, loss function, and optimization strategies. This chapter provides a clear view of how the new cGAN model was developed to improve high-resolution atmospheric data prediction.

2.1 Objectives and Necessity of Super-Resolution for Atmospheric Fields

In many traditional statistical downscaling approaches, a single algorithm is trained for each grid point, which results in significant time and computational resource expenditures. This methodology presents several limitations, including the inability to capture spatial and temporal interactions between points and the difficulty in generalizing learning to new data. Among traditional methods, linear regression and multiple regression are among the most widely used, allowing for the establishment of relationships between low- and high-resolution climatic variables (Maraun D. et al., 2010; Wilby R. L. et al., 2002). Other approaches include statistical transfer models, which utilize historical data from meteorological stations to create statistical relationships (Hay L.E. et al., 2002), and principal component analysis (PCA) methodologies, used to reduce data dimensionality (Eastman, J. R. and Fulk, M., 1993). Additionally, stochastic weather generators (Wilks D. S., 1992) and cumulative distribution function (CDF) matching methods (Piani C. et al., 2010) have been applied to generate more accurate estimates of precipitation and temperature. While these approaches are effective, they have limitations in their ability to model spatial complexities, paving the way for advanced techniques based on computer vision, such

as Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs), which promise to further enhance the efficiency and accuracy of downscaling (Xie L. et al., 2018).

The integration of CNNs and GANs into the downscaling process offers several advantages over traditional approaches. Firstly, these methods allow for the simultaneous processing of multiple grid points, leveraging the spatial and temporal correlations between them and improving computational efficiency. For instance, GANs can be trained on multidimensional datasets, reducing the necessity to train a model for each point (Wang Y. et al., 2018). Moreover, these methods are capable of learning directly from the data, diminishing the need for explicit models and increasing flexibility in handling variability in climatic data. The ability of GANs to generate data that captures original statistical structures makes this approach particularly promising for downscaling complex atmospheric variables, such as precipitation.

2.2 Statistical and Dynamical Downscaling

Downscaling aims to bridge the gap between the large spatial scales represented by GCMs and the smaller scales required for assessing regional climate change and its impacts.

Two major types of downscaling exist: in dynamical downscaling, a high-resolution regional climate model (RCM) is nested into the GCM over the domain of interest (Rummukainen M., 2010). In statistical downscaling, empirical links between the large-scale and local-scale climate are identified and applied to climate model output.

The first downscaling methods were invented as early as the late 1940s (Klein 1948) and became operational during the early days of numerical weather prediction in the late 1950s. Klein in 1959 utilized the extensive network of meteorological time series to infer statistical relationships between the observed large-scale circulation, described by the limited variables simulated at the time by models, and the local meteorological variables of interest. This statistical approach was then applied to downscale the actual numerical forecast of large-scale circulation, assuming the predictor was perfectly predicted by the model, into a local weather forecast. However, as a significant database of past forecasts was accumulated and analyzed, it became evident that large-scale weather forecasts were not perfect and exhibited systematic deviations from observations. Consequently, the "perfect prognosis" (PP) method was not without errors. In 1972, Glahn and Lowry developed a novel approach (Model Output Statistics (MOS) to address systematic errors (biases) inherent in large-scale models. Rather than relying

on observed data as predictors, as employed in the Perfect Prognosis (PP) method, they calibrated the statistical model using archived numerical forecasts. This methodology enabled the statistical model to adjust future predictions based on historical model performance. For each new weather forecast, the derived statistical relationships are applied to the corresponding numerical forecast, facilitating systematic bias correction during the *post-processing* stage. Simultaneously, numerical approaches were developed to enhance the resolution and accuracy of forecasts for a specific target region. The first limited-area model was created at the National Meteorological Center in the United States (Maraun D. and Widmann M., 2018) and became operational in 1971. This model encompassed the United States, Canada, and the Arctic Ocean, with a horizontal resolution of 190.5 km at 60°N. It was driven at the lateral boundaries by input from a numerical weather prediction model for the Northern Hemisphere.

Dynamical downscaling relies on the application of physical laws governing atmospheric processes to offer a more detailed and accurate representation of climate dynamics. This approach is particularly effective in regions with complex terrain, as it captures intricate physical interactions. However, it requires significant computational power and may still inherit biases from the General Circulation Models (GCMs) that provide the initial conditions. In contrast, statistical downscaling establishes empirical links between large-scale climate data from GCMs and localized climate records, allowing for the creation of higher-resolution climate projections with relatively lower computational demands (Vandal T. et al., 2017). Due to its efficiency and ease of application, statistical downscaling is often favored over dynamical methods, which use regional climate models (RCMs) to simulate smaller-scale climate processes (Tiwari P. et al., 2018). Research indicates that while dynamical downscaling can enhance the depiction of extreme weather patterns, it can also introduce uncertainties, particularly related to model parameters and boundary conditions (Maraun D. et al., 2010). Consequently, the selection between statistical and dynamical approaches depends largely on the study's objectives, including the required resolution, the climate variables of interest, and available computational resources (Xiao D. et al., 2021; Tiwari P. et al., 2018).

Since the mid-1990s, both dynamic and statistical downscaling approaches have seen significant advancements and broadening of scope. Numerous studies and reviews, notably Hewitson B. C. and Crane R. G. (1996) and Giorgi F. and Mearns L. O. (1999), have evaluated these methods. The heightened focus on model errors has spurred model intercomparison projects, such as PIRCS and PRUDENCE in Europe and NARCCAP in North America, aimed at assessing and addressing biases in climate models. The launch of the Coordinated

Downscaling Experiment (CORDEX) in 2009 marked a key milestone toward the operationalization of climate downscaling. Furthermore, in 2012, the Global Framework for Climate Services (GFCS) provided added momentum for the development of regional climate products. Initiatives such as ENSEMBLES have generated extensive databases of regional climate simulations, establishing multi-model ensembles to address model uncertainty. CORDEX-ESD, introduced in 2013, has further enhanced the coordination of statistical downscaling methodologies. Ultimately, downscaling is a vital tool in climate research, transforming GCM outputs into actionable data at regional and local scales, and supporting more effective decision-making regarding climate adaptation and mitigation across diverse sectors.

2.3 Datasets and *Pre-processing*

This section provides a detailed overview of the two datasets with distinct resolutions required for the Super-Resolution GAN 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 two crucial atmospheric fields: the daily 2-meter temperature and daily precipitation for investigation. This allows us to evaluate the algorithm's performance and its ability to downscale temperature and precipitation data effectively.

Additionally, this section describes the data preprocessing procedures which are crucial for training GANs. Given that GANs 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 GANs' 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 GANs.

2.3.I Domain Definition

The domain was chosen based on that defined in Raffa et al., 2021 for VHR_REA-IT. For the high-resolution dataset (VHR_RE-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 (Tab. 2.I). The computational domain extends in longitude from $\sim 3.90^{\circ}\text{W}$ to $\sim 19.93^{\circ}\text{E}$ and in latitude from $\sim 34.80^{\circ}\text{N}$ to $\sim 48.59^{\circ}\text{N}$. 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. 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 GAN in capturing relevant topographic information, and correctly positioning, for example, mountain ranges in the new artificial high-resolution dataset.

Data	longitude and latitude	horizontal resolution	n° grid points
ERA5	(Lon = 4° W–23.75° E; Lat = 34.25° N–48.50° N)	0.28° ($\approx$ 31 km)	58x80
VHR_REA-IT	(Lon = 3.90° W–19.93° E; Lat = 34.80° N–48.59° N)	0.02° ($\approx$ 2.2 km)	680x535

Tab. 2.I: Domain details for LW and HR dataset: geographical boundaries, horizontal resolution, number of grid points.

This consideration holds particular significance in geographically complex contexts such as the Italian Peninsula (Fig. 8), 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.

2.3.2 Low-Resolution Dataset: ERA5

In this approach, we deviated from the conventional practice of employing random sequences for the low-resolution dataset, as often seen in prior GAN-based studies. Instead, we opted to utilize a real dataset, specifically the ERA5 reanalysis data (Fig. 2.Ia) at a horizontal

resolution of 0.25 degrees (≈ 31 km), which can be accessed online at (<https://doi.org/10.24381/cds.adbb2d47>).

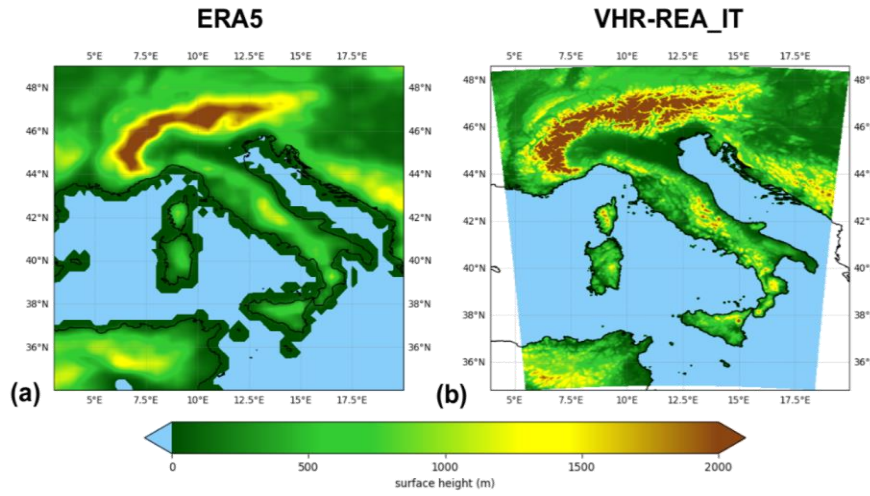


Fig. 2.1: Topography. (a)-(b) ERA5; VHR_REA-IT.

The selected dataset spans a total of fifteen years, divided into two distinct periods: a training period (01/1990–12/2000) and a test period (01/2001–12/2005).

2.3.3 High-Resolution Dataset: VHR_REA-IT

The very High-Resolution Dynamical Downscaling of ERA5 Reanalysis (VHR_REA-IT) is chosen as the high-resolution dataset (Fig. 2.1b). This reanalysis is at the convection-permitting scale (horizontal grid spacing 0.02° , ≈ 2.2 km) 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 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 GAN to judge the goodness of these at the same horizontal resolution.

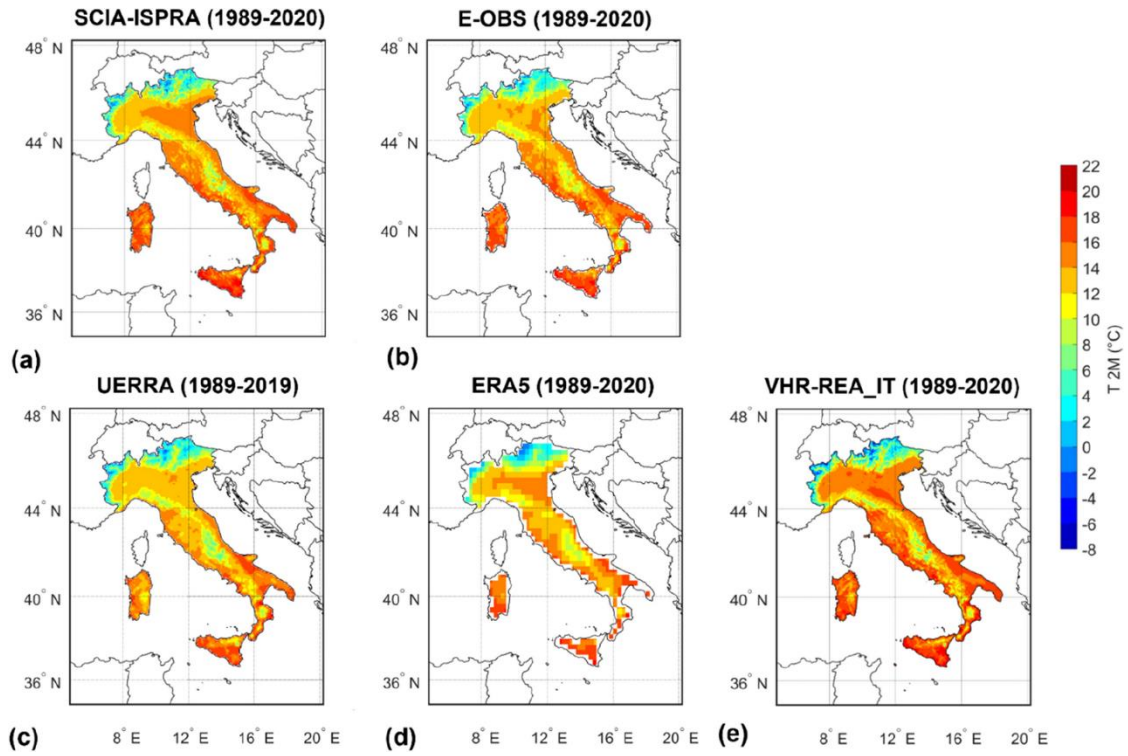


Fig. 2.2: Spatial distribution of daily mean temperature. (a)-(e) SCIA-ISPRA; E-OBS; UERRA; ERA5 and VHR-REA_IT over 1989–2020 (except for UERRA covering the period 1989–2019). Reproduced by Adinolfi M., et al., 2023.

The selection of the VHR_REA-IT dataset as the counterpart in the high-resolution model is based on several key considerations. First and foremost, VHR_REA-IT provides climate data for the examined territory with high spatial (2.2 km) over an extended period of 30 years (1989–2020), thereby overcoming the limitations of traditional approaches that rely on shorter simulation periods. Additionally, as highlighted in the study by Adinolfi et al. (2023), the model has demonstrated generally strong performance in simulating the downscaled variables using artificial intelligence techniques.

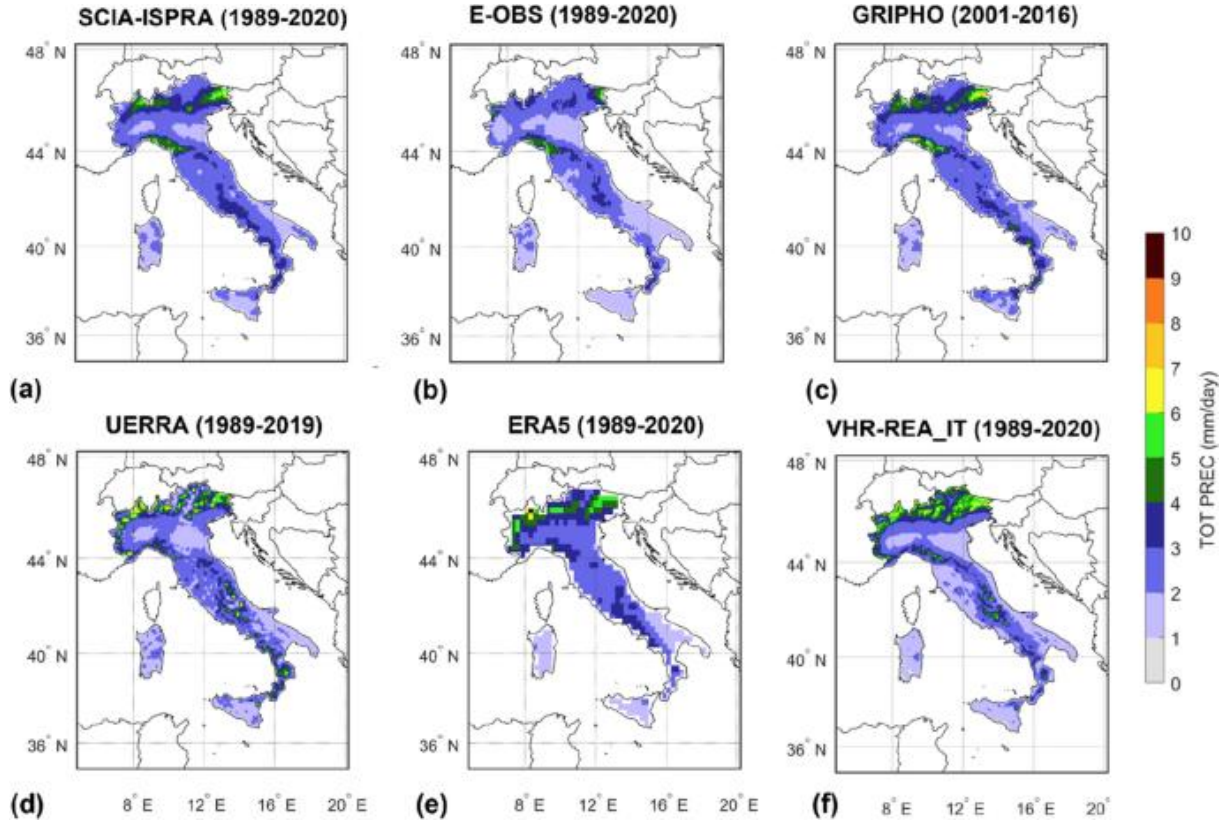


Fig. 2.3: Spatial distribution of daily total precipitation. (a)–(e) SCIA-ISPRA; E-OBS; UERRA; ERA5 and VHR-REA_IT over 1989–2020 (except for UERRA covering the period 1989–2019). Reproduced by Adinolfi M., et al., 2023.

Specifically, VHR_REA-IT has shown significant capabilities in reproducing spatial patterns of mean temperatures (Fig. 2.2) and precipitation, although a warm temperature bias has been observed in flat areas. Furthermore, the model accurately represented precipitation patterns (Fig. 2.3), except in alpine regions, while effectively capturing the intensity and frequency of heavy precipitation events, particularly in mountainous areas.

The data used for both high-resolution and low-resolution are available in hourly aggregations, but for the sake of simplifying the computational complexity of the problem, we decided to employ daily aggregations,

2.3.4 Transformations and Data Normalization

Input normalization in a neural network, such as a Generative Adversarial Network (GAN), is a process aimed at ensuring that the input data is in a standardized and consistent form. 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 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 GAN network, optimizing data representation, and enabling the generation of consistent, high-quality outputs. Within our modeling framework, a targeted normalization approach was adopted to handle temperature data. Before utilizing the data as input for the GAN 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 GAN (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 GAN 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.

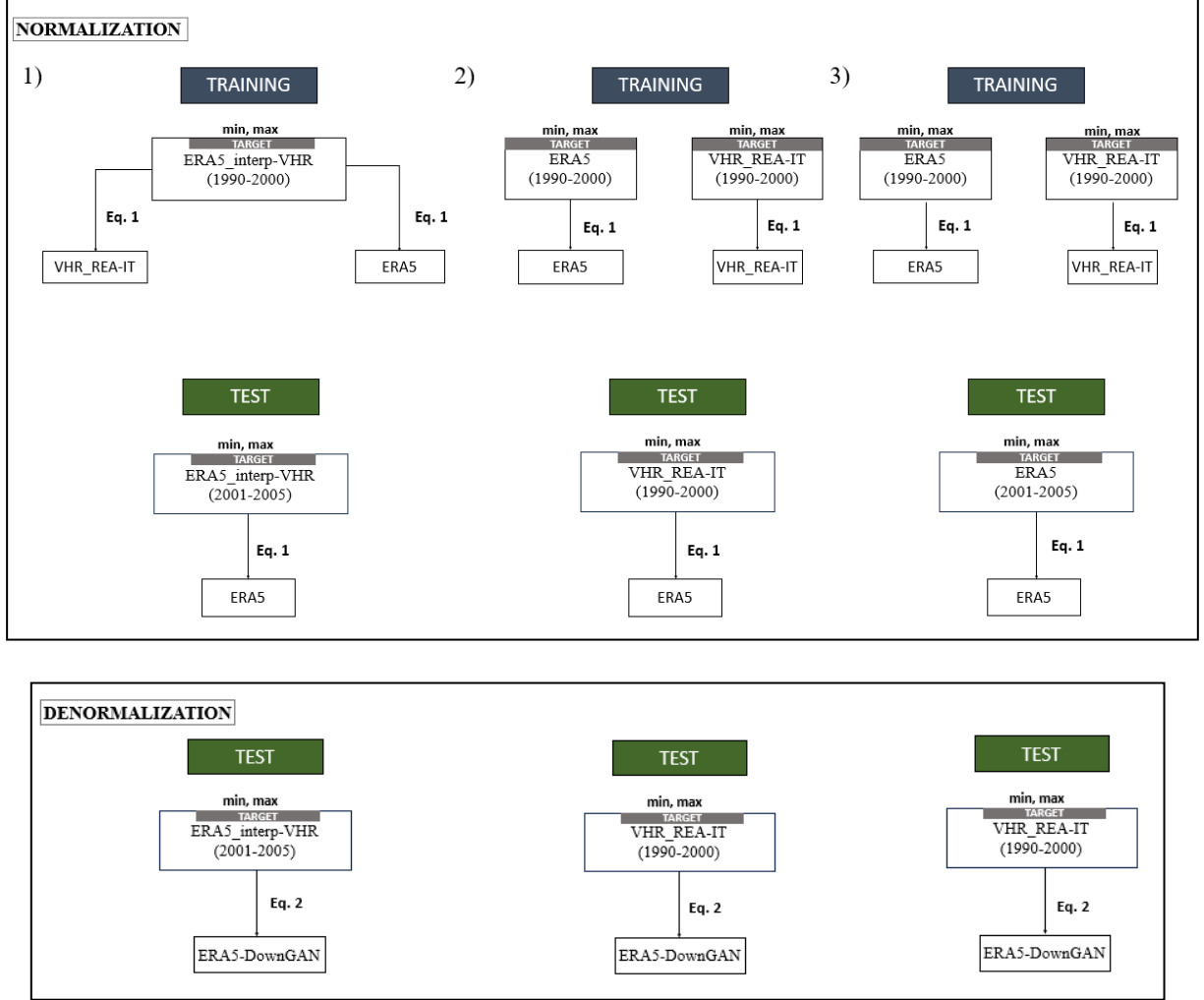


Fig. 2.4: Normalization (a) and denormalization (b) approaches with different target datasets.

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 GAN, 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 GAN.

Despite testing and evaluating all normalization and denormalization approaches (Fig. 2.4), 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.

2.4 Proposed cGAN Architecture: ERA5-DownGAN

The approach developed in this study for constructing a new method based on neural networks is grounded in the general principles of Generative Adversarial Networks (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. Generative Adversarial Networks (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.

Specifically, the use of new conditional Generative Adversarial Networks (cGANs) 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 (Fig. 2.5). 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.

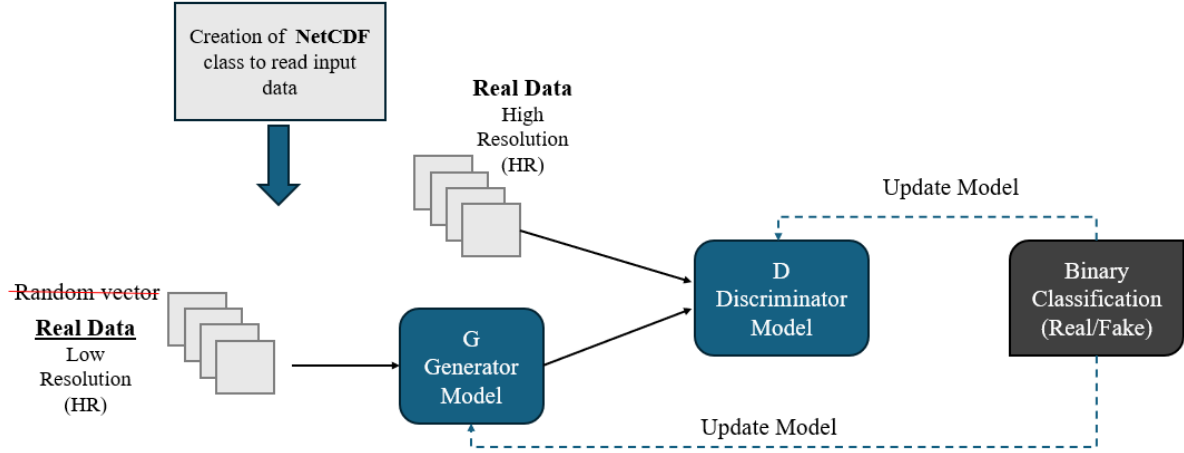


Fig. 2.5: Structure of the developed cGAN, adapted from the general architecture of the Generative Adversarial Network (GAN) for the training process in Figure I.7.

In summary, this particular cGAN employs a specific architecture with LeakyReLU and tanh activation functions for the generator, alongside LeakyReLU, dropout, and sigmoid activations for the discriminator. The training process focuses on generating high-resolution climate data that closely resembles the VHR_REA-IT dataset, to achieve convergence where the generated data is indistinguishable from the actual VHR_REA-IT data.

To implement the planned GAN, several steps were undertaken:

- **Data Pre-processing:** This phase includes quality checks of the data, domain selection, partitioning the dataset into training and testing intervals, and normalizing the data to ensure consistent input conditions.
- **Tensor Conversion from NetCDF Files:** The data is transformed into tensor format using the Pytorch library, making it suitable for processing within the GAN architecture.
- **Architectural Development:** The generator and discriminator networks are carefully designed. The generator produces synthetic data, while the discriminator learns to differentiate between real and generated data.
- **Loss Function Definition:** An appropriate loss function is chosen to guide the training process, quantifying how well the generated data matches the real data.

- **Parameter Optimization:** Hyperparameters such as learning rate and batch size are systematically adjusted to optimize the GAN's performance, ensuring effective learning and convergence.

2.4.I Overall Architecture

Generative Adversarial Networks (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, use gradients from this loss function to update the weights and biases of the neural networks. This framework is characterized by weight assignments to connections between neurons across different layers, which affects how data is transformed within the network.

In this study's GAN, 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 GAN's capacity to learn complex representations and produce high-quality outputs.

2.4.2 Details of the Generator and Discriminator Architectures

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 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 58x80, 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. 2.6.

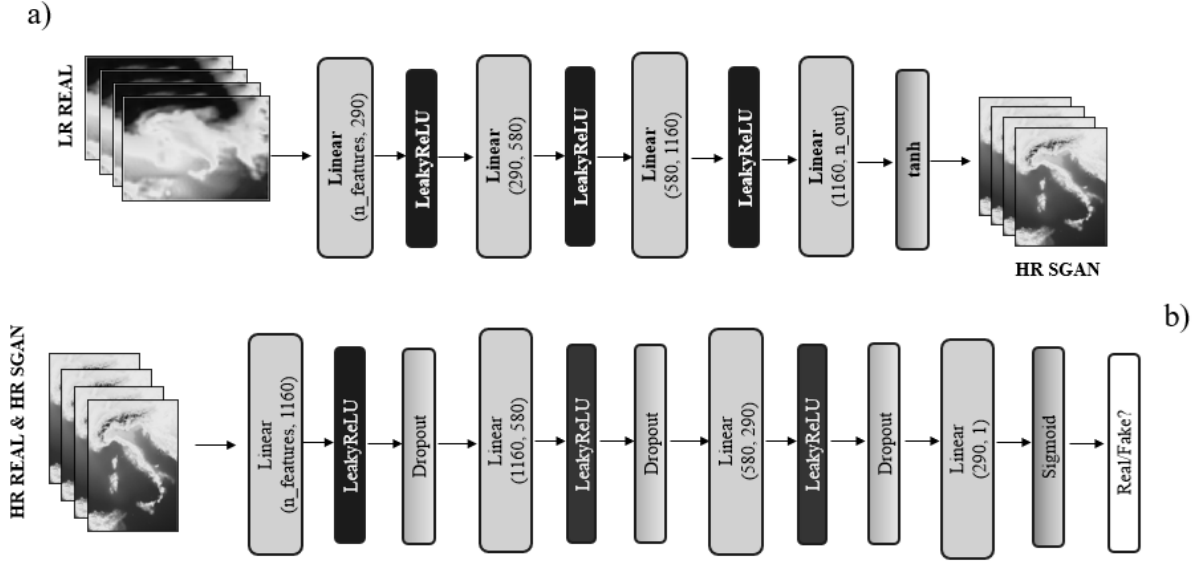


Fig. 2.6: Architectures of (a) generator and (b) discriminator. Size input generator 58×80 , size input discriminator 680×535 , the number of frames per sequence 100 (batch size = 100) for both data sets considered here. After training, the best generator is used for the test period.

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 context of the input matrices and their 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.

2.4.3 Configuration of Hyperparameters and Optimization Technique Adopted

The choice of the optimization strategy plays a pivotal role in shaping the efficacy of our Generative Adversarial Network (GAN). To this end, we have harnessed the power of the Adam optimizer, an algorithm renowned for its adaptability and efficiency.

Adam, short for Adaptive Moment Estimation (Kingma, D. P., and Ba, J., 2015), unites the virtues of both momentum-based optimization and root mean square propagation. Particularly, it combines AdaGrad, which adapts the learning rate based on the historical gradient information for each parameter, and RMSProp, which uses an exponentially decaying average of squared gradients to maintain a stable training process. This amalgamation empowers the optimizer to dynamically calibrate learning rates for each network parameter during the training process.

Mathematically, Adam operates by maintaining two-moment vectors, m_t and v_t , which represent the first and second moments of the gradients at time step t .

1. Parameter Initialization:

$$m_0 = 0, \quad v_0 = 0 \\ t = 0$$

2. Gradient Calculation. At each time step t , compute the gradient of the loss function L with respect to the parameters θ :

$$g_t = \nabla_{\theta} L(\theta)$$

3. Update Moments. Update the first and second moment estimates:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2$$

where β_1 and β_2 are decay rates typically set to 0.9 and 0.999, respectively.

4. Bias Correction. To counteract the bias introduced during initialization, Adam applies bias correction:

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t}$$

5. Parameter Update. Finally, the parameters are updated as follows:

$$\theta = \theta - \frac{\alpha}{\sqrt{\hat{v}_t} + \epsilon} \hat{m}_t$$

where α is the learning rate and ϵ is a small constant added for numerical stability.

Adam's ability to adaptively adjust the learning rates for different parameters has shown significant benefits in the context of GAN training. Evidence from several studies suggests that it effectively stabilizes the training process and mitigates issues such as mode collapse, where the generator fails to capture the diversity of the target distribution (Kingma, D. P. & Ba, J. 2015; Radford A. et al., 2015). Additionally, Adam's performance with sparse gradients enhances its utility in GAN applications, which often experience varying training dynamics between the generator and discriminator (Alzubaidi L. et al., 2021; Yu X. et al., 2018). Empirical evaluations confirm that Adam not only yields lower loss values but also produces higher-quality generated samples compared to alternative optimizers, establishing its reputation as a robust choice for GAN training due to fast convergence and reliable performance (Karras T. et al., 2019; Salimans T. et al., 2016).

In the architecture employed in this study, several critical hyperparameters were set to optimize performance and ensure an effective model training phase.

The batch size represents the number of training samples processed before the internal parameters of the model are updated. This hyperparameter is critical as it directly impacts the training dynamics and the stability of gradient estimates. In this study, a batch size of 100 was

selected after the sensitivity test. This choice balances computational efficiency with effective model training. In fact, using a batch size that is too small presents several disadvantages, including increased variability in gradient estimates, which can make convergence unstable (Bottou L. and Boucheron O., 2007). Additionally, this may lead to a longer total training time, as more iterations are required to complete one epoch. Models with small batch sizes are also more susceptible to overfitting, increasing the risk of poor generalization on unseen data (Bejani, M., and Ghatee, M., 2021). Moreover, optimizing with smaller batches may involve greater computational complexity and resource demands. On the other hand, a batch size that is too large presents its own set of disadvantages, such as reducing the model's ability to generalize to new data, as the gradient estimates become more deterministic and less variable (Hoffer E. et al., 2017; Keskar N. et al., 2016). Additionally, the computational resources required increase significantly, limiting training flexibility and making fine-tuning of parameters difficult (Kornblith S. et al., 2019). Another drawback is that larger batch sizes can lead to increased convergence times, as the model may not effectively learn the complexities of the data (Smith L. N. et al., 2018). Finally, excessively large batch sizes can cause memory issues, limiting the size of models that can be trained on hardware with reduced capacity (Hestness J. et al., 2017). The learning rate (lr) is a critical hyperparameter that governs the step size taken during the optimization process to update the model's weights. An appropriately selected learning rate can significantly enhance both convergence speed and the overall performance of the model. In this study, a learning rate of 0.0001 was employed, which is relatively low, and was chosen in consideration of using the Adam optimizer. Adam dynamically adapts the learning rates for each parameter, which promotes more stable convergence and mitigates the risks of oscillation or divergence during training. Higher learning rates can lead to overshooting minima, while lower rates can impede training progress and result in getting stuck in local minima (Kingma D.P. and Ba J., 2015). The learning rate directly influences the trade-off between convergence speed and stability, rendering it a critical factor in the training of deep neural networks (Loshchilov I. and Hutter F., 2016).

Finally, the “number of epochs” parameter is set to indicate the total number of complete passes through the training dataset during the training process. This hyperparameter is essential to ensure that the model has adequate opportunities to learn from the data. In this study, 100 epochs were utilized, allowing the model to effectively learn from the dataset without risking overfitting. The choice of the number of epochs is critical: an insufficient number may lead to underfitting, where the model fails to learn the underlying patterns, while an excessive number

can result in overfitting, where the model learns the noise and fails to generalize to unseen data (Goodfellow I. et al., 2016). This choice was made by monitoring the trends in the errors produced by the generator (“g_err”) and the discriminator (“d_err”). Through this configuration of hyperparameters, our GAN successfully converges to a solution that optimally balances the interaction between the generator and discriminator networks. This meticulous configuration culminates in the generation of high-fidelity and high-resolution climate predictions, thereby demonstrating the robustness and effectiveness of our methodological framework.

2.4.4 Training Strategies

During the training process of the cGAN model, several critical steps were implemented to ensure stable learning for both the generator and discriminator. Shuffling of the data was deliberately set to "False" to maintain the alignment between the low-resolution (LR) input data and the corresponding high-resolution (HR) target data. This alignment is crucial in a downscaling task, where each LR sample corresponds directly to a specific HR sample. If the data were shuffled, this correspondence would be disrupted, leading to mismatches between inputs and targets and, consequently, poor model performance.

The parameter “d_steps” refers to the number of updates applied to the discriminator during the training process of a GAN. In this case, setting “d_steps” = 1 means that for each update of the generator, the discriminator is updated once. This approach follows the recommendations by Goodfellow I. et al. (2014), which helps to stabilize the training dynamics between the generator and discriminator. The training of the discriminator occurs in two stages: first on real data and then on fake data generated by the generator. For the real data, the discriminator predicts its authenticity, and the loss is calculated using BCEWithLogitsLoss, where the target corresponds to the real data label. The discriminator is subsequently trained on the generated fake data, with the loss computed using the same function but employing the target for fake data.

Once the predictions for both real and fake data have been backpropagated and the gradients accumulated, the weights of the discriminator are updated. Following this, the generator is trained to minimize a weighted combination of pixel-wise loss (using MSELoss) and adversarial loss (using BCEWithLogitsLoss). The objective of the generator is to produce synthetic data that can effectively deceive the discriminator into classifying it as real. The generator's error is thus the weighted sum of these two losses, which is backpropagated, followed by an update of the generator's weights using the optimizer. To determine the optimal stopping point for the training, we plotted the performance metrics at regular intervals against

a test sample of 16 days. This sample included comparisons between the low-resolution data, the high-resolution targets, and the high-resolution outputs generated by the cGAN. The error metrics derived from these comparisons were instrumental in assessing the quality of the generated high-resolution data against the climatic downscaling reference. Additionally, we monitored the trends of the errors associated with both the generator and discriminator, which provided insights into their respective learning processes and overall model performance. This comprehensive evaluation strategy not only enhanced the reliability of our results but also ensured that the training was effectively converging toward high-fidelity outputs.

2.4.5 Custom Loss Functions

Within the intricate architecture (Fig. 2.6) of our dedicated Conditional Generative Adversarial Networks (cGANs) 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 coordinated 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. Each loss function was assigned specific weights to reflect the different roles they play in optimizing the overall performance of the network. The "pixel_weight" was set to 0.5, giving it a more substantial influence in the optimization process due to its critical role in ensuring accurate reconstruction of the spatial features of the field. This higher weight emphasizes the importance of aligning the pixel-level details of the output with the target high-resolution data, which is essential for maintaining spatial fidelity.

In contrast, the "adversarial_weight" was assigned a lower value of 0.2. This reduced weight was chosen to limit the influence of the adversarial component, as it may be more prone to instability when the adversarial loss dominates the training process (Arjovsky M. et al., 2017; Gulrajani I. et al., 2017). By assigning a lower weight to the adversarial term, we aimed to avoid overemphasizing the generation of visually plausible outputs at the expense of numerical accuracy, thereby balancing the need for realistic outputs with the requirement for spatial and temporal consistency in the reconstructed fields. This strategy enables the GAN to effectively model complex dynamics while mitigating the risks of overfitting and instability associated with adversarial training (Mescheder L. et al., 2018).

3 | Application of cGAN for the Italian Peninsula

In this chapter, after providing a clear explanation of dynamic and statistical downscaling methods, detailing their respective advantages and limitations, we present the main results of applying the Super-resolution GAN developed centered on the Italian domain. The configuration parameters for GAN-based downscaling, including batch size, number of epochs, and initial values for the loss function, were carefully optimized through sensitivity analysis to identify the most effective settings for this geographical area and the specific variable under investigation. The daily 2-meter temperature field was initially used to evaluate the performance of our Super-resolution GAN. Subsequently, we further assessed its capabilities by applying it to the daily total precipitation field, introducing a higher level of complexity to the evaluation. The chapter will also describe the preprocessing techniques employed for each variable and provide a comprehensive analysis of the results. This analysis will focus on general error metrics, as well as spatial correlation and the assessment of extreme values. This detailed examination aims to offer an in-depth understanding of the Super-resolution GAN's performance and its effectiveness in downscaling applications.

3.1 Downscaling of 2-meter Temperature

In this section, the GAN's performance in the downscaling process of the 2m-temperature field over the Italian Peninsula is evaluated by putting side by side the synthetic high-resolution dataset (ERA5-DownGAN) generated by the GAN 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 GAN'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 GAN 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 GAN for this specific geographic domain and meteorological field. Within this comparison, the 2-meter temperature field of the low resolution (ERA5) is further examined to assess its deviation from the high resolution and to evaluate the GAN'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 GAN's behavior in handling differences between the 2-meter temperature distributions across the two resolution levels, thus contributing to the overall assessment of the model's performance in the context of climatic downscaling.

3.I.I Results and Analysis: Comparison with Dynamical Downscaling

The temporal trend of the 2-meter temperature field (T_{2M}) is illustrated in Fig. 3.I over the test period (2001-2005) for both high-resolution datasets, real (VHR_REA-IT) and synthetic (ERA5-DownGAN) and the real low-resolution dataset (ERA5) across the entire considered domain.

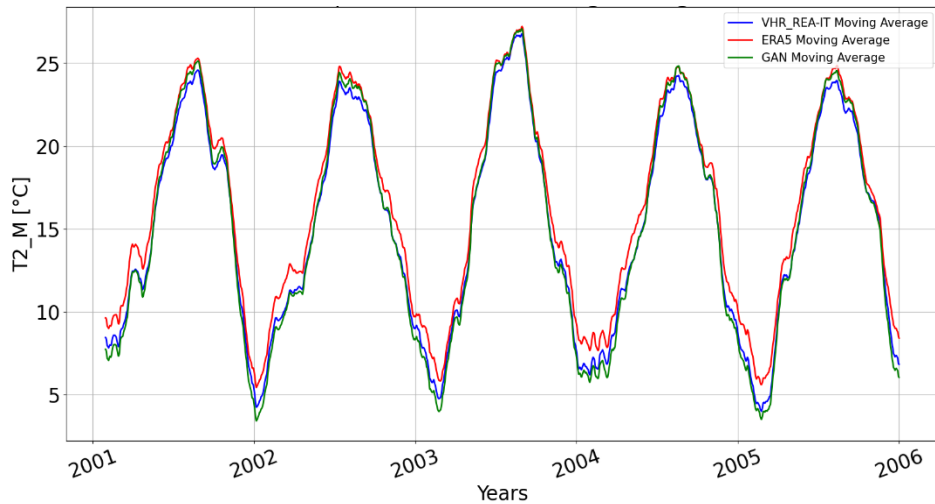


Fig. 3.I: Temporal evolution of the daily moving average of the 2m-temperature during the test period, calculated with a 30-day window. Mean daily spatial average of the real low-resolution data (ERA5), the real high-resolution data (VHR_REA-IT) and the downscaled artificially data (ERA5-DownGAN), created by the cGAN.

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. 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 $\sim 6^\circ\text{C}$ to about 28°C . On the other hand, the high-resolution dataset generated by GAN (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 $\sim 4^\circ\text{C}$ to $\sim 28^\circ\text{C}$. This highlights the capability of GAN to generate not only a temporally correlated dataset but also to maintain spatial mean values that deviate very little from the reference dataset.

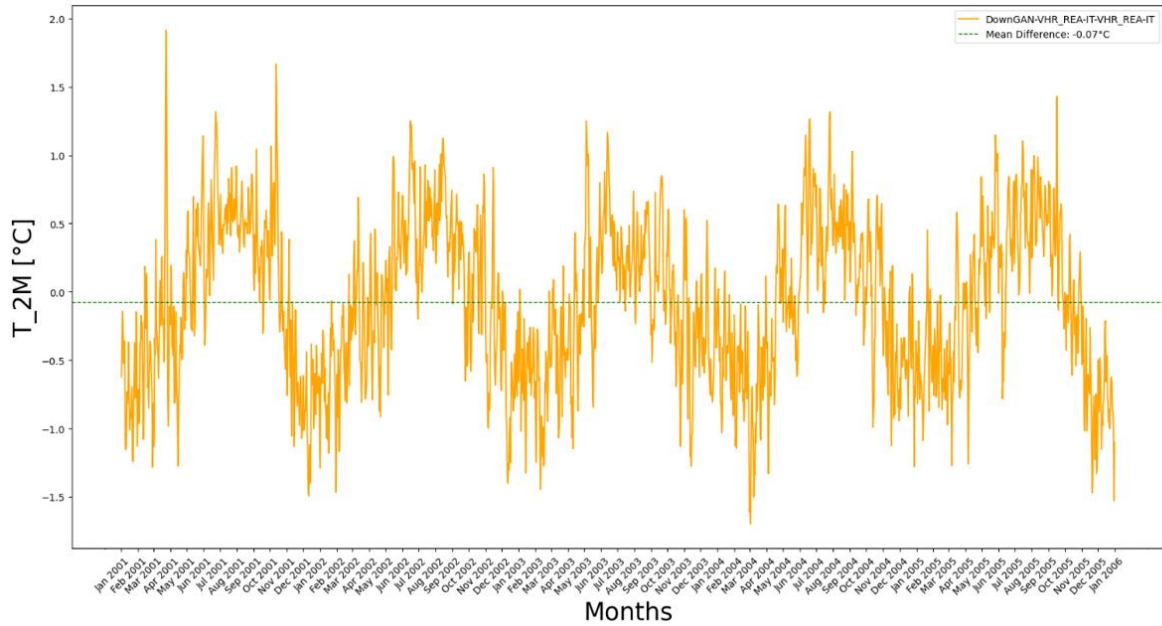


Fig. 3.2: Temporal evolution of daily differences between the dataset generated by the Conditional Generative Adversarial Network (cGAN), the ERA5-DownGAN dataset, and the real VHR_REA-IT dataset.

The analysis of the time series (Fig. 3.2) of daily differences between the dataset generated by the Conditional Generative Adversarial Network (cGAN), the ERA5-DownGAN dataset, and the real VHR_REA-IT dataset 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. These values, being significantly outside the mean range of -1.2°C to 1.2°C , prompted an in-depth analysis of these specific days. The objective is to understand the potential causes of this marked deviation from the VHR_REA-IT dataset and to assess whether there has been a deterioration in the GAN's ability to perform statistical downscaling in line with dynamic downscaling under particular meteorological conditions. Therefore, it was decided to examine the synoptic conditions associated with these anomalies in order to identify the factors influencing temperature

variations during these periods. A detailed discussion of these specific events is provided in Appendix A to avoid disrupting the main discussion of the statistical downscaling model application to the 2m-temperature, ensuring that proper attention is given to a very interesting case study in understanding the performance of the developed cGAN, particularly in anomalous contexts.

According to the time plot of the correlation between the 2-meter temperature (T_{2M}) fields from dynamic downscaling (VHR_REA-IT) and those generated by the GAN during the testing period (Fig. 3.3), 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 GAN for specific atmospheric conditions, ultimately leading to the optimization of the architecture.

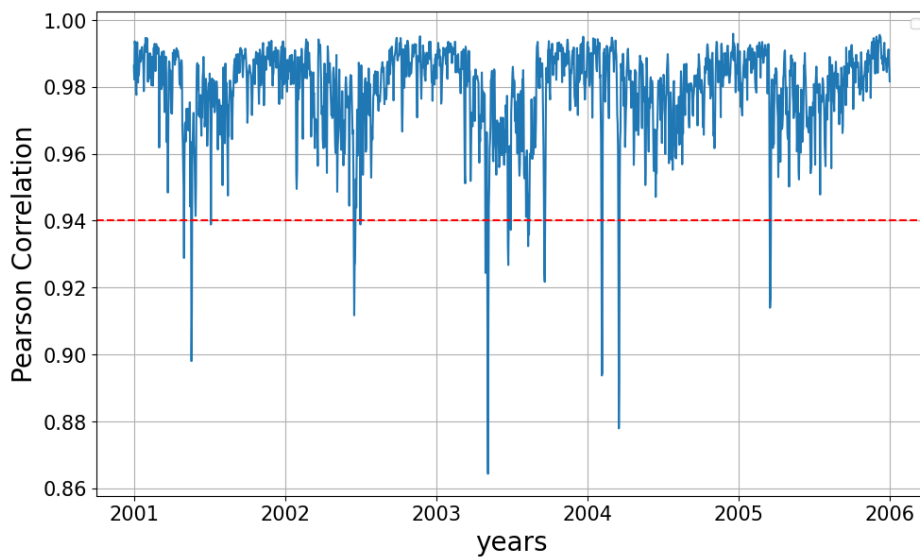


Fig. 3.3: Correlation evolution of the daily spatial average of the 2m-temperature during the test period. (a)–(c) Real low-resolution data (ERA5), the real high-resolution data (VHR_REA-IT) and the downscaled artificially data (ERA5-DownGAN), created by the cGAN.

3.I.2 Error Metrics Evaluation

We assessed the performance of our Generative Adversarial Network (GAN) 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.

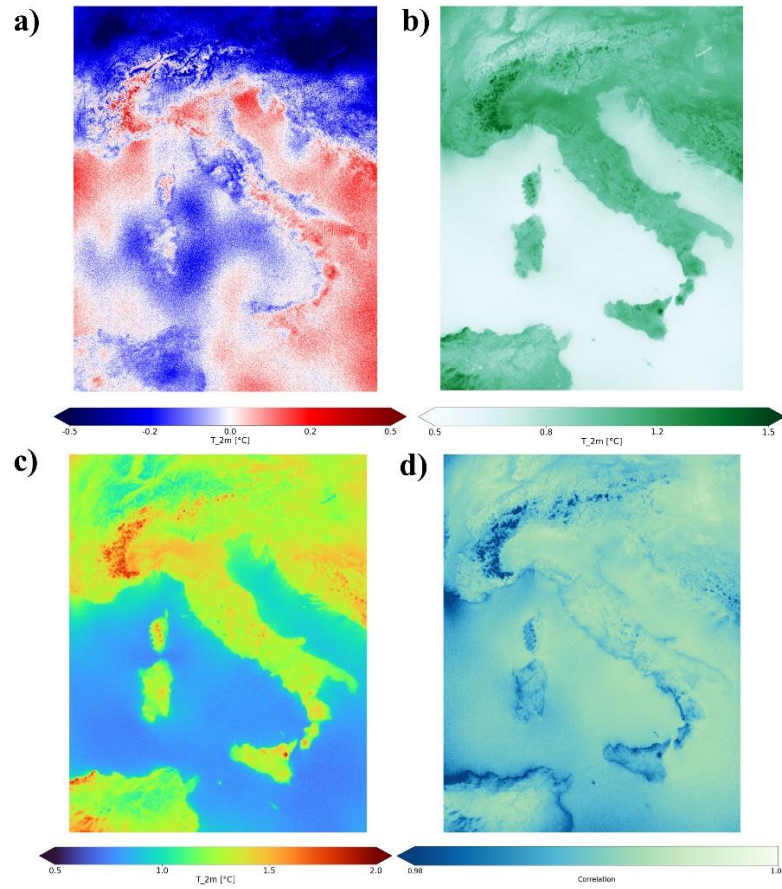


Fig. 3.4: Maps of time average. (a)-(d) Mean bias (BIAS), mean absolute errors (MAEs), root-mean-square errors (RMSEs), and correlation coefficients (CORRs), comparing downscaled daily 2m-temperatures (ERA5-DownGAN) with the high-resolution dataset (VHR_REA_IT) for the test period (2001-2005).

Bias: The bias, 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.

$$\text{BIAS} = x_{\text{gen}} - x_{\text{real}}$$

,where x_real is the average value of the real dataset obtained from mean calculation in time, x_gen is the corresponding mean value of the generated dataset.

b. Mean Absolute Error (MAE): The MAE, 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.

$$MAE = |x_gen - x_real|$$

c. Root Mean Squared Error (RMSE): The RMSE, 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.

$$RMSE = \sqrt{(x_gen - x_real)^2}$$

d. Pearson Correlation: The Pearson correlation 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).

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. 3.4. 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°C to 1.5°C and 0.5°C to 2°C , 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.

3.I.3 Spatial and Temporal Consistency

The comparison among real low-resolution (58x80 grid points), real high-resolution (680x535 grid points), and downscaled ERA5-DownGAN (680x535 grid points) high-resolution test datasets are illustrated in Fig. 3.5. This depicts 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.

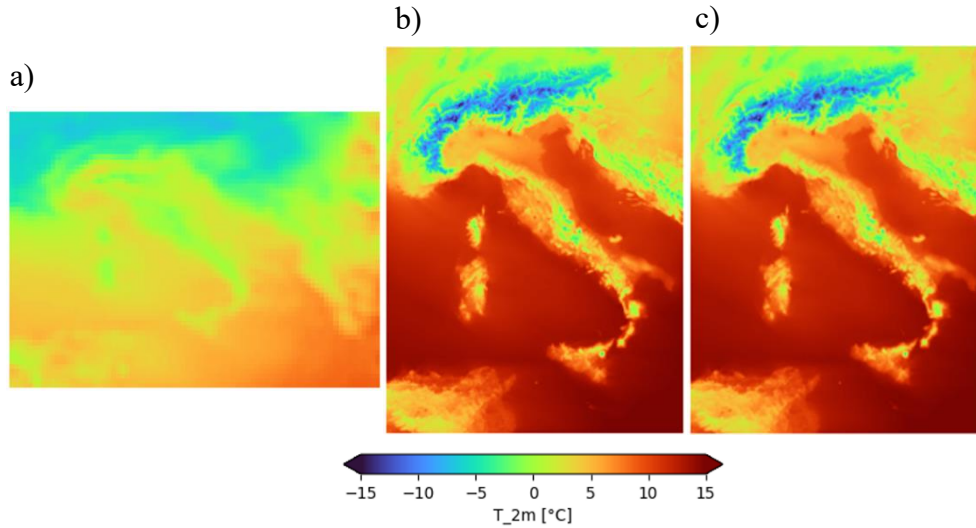


Fig. 3.5: Examples of one day from the test data set. (a)–(c) Real low-resolution data (ERA5), the real high-resolution data (VHR_REA-IT) and the downsampled artificial data (ERA5-DownGAN), created by the cGAN.

Furthermore, the GAN 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.

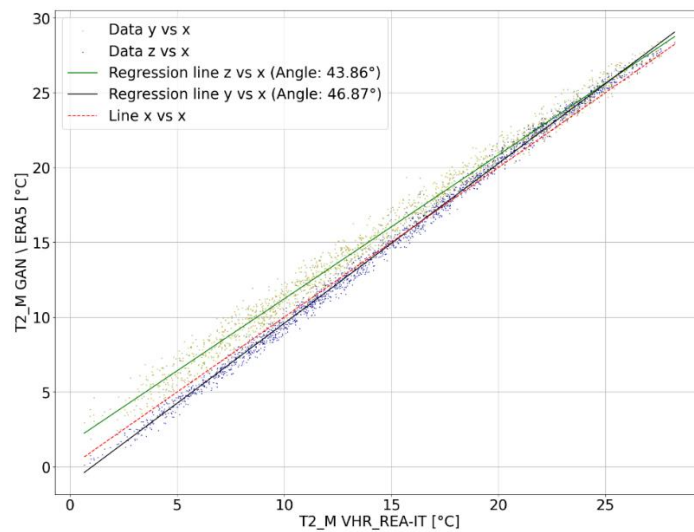


Fig. 3.6: Scatter diagram for evaluating the correlation between ERA5, VHR-REA_IT and ERA5-DownGAN.

According to Fig. 3.6, 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).

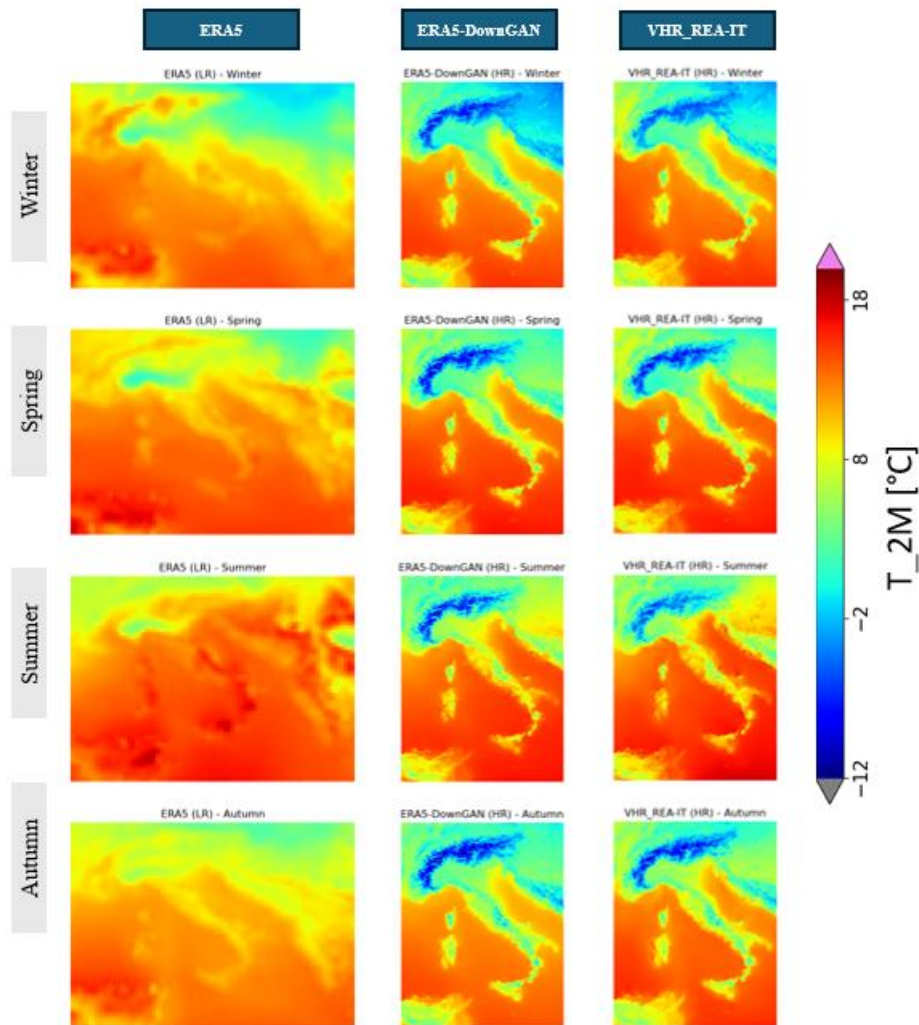


Fig. 3.7: Comparison of seasonal 2m-temperature (T_{2M}) for real low-resolution data (ERA5), the real high-resolution data (VHR_REA-IT), and the downscaled artificial data (ERA5-DownGAN), created by the cGAN.

In Fig. 3.7, the seasonal 2m-temperature (T_{2M}) maps show that the dataset generated by the GAN, 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 GAN is evident, particularly in regions of Central Europe, such as between Austria and Hungary, and in some areas of northwestern Italy.

3.1.4 cGAN's Ability to Capture Extreme Values

The assessment of percentiles ranging from the 1st to the 99th in Fig. 3.8 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 GAN's ability to accurately capture extreme events and the overall distribution of the data. For the creation of the boxplots in Fig. 3.8, 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.

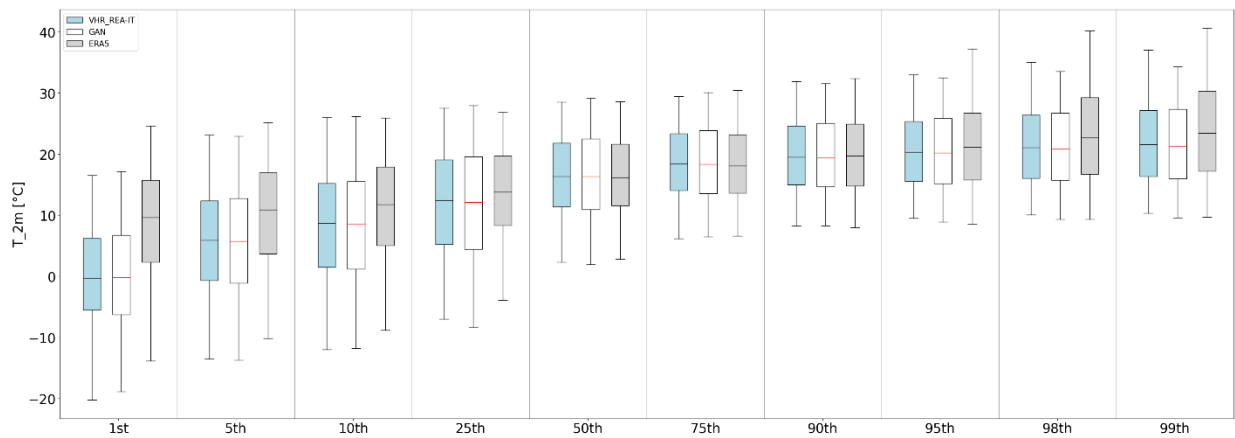


Fig. 3.8: Comparison of the 2m-temperature (T_{2M}) distribution for real low-resolution data (ERA5) with the gray rectangle, the real high-resolution data (VHR_REA-IT) with the blue rectangle, and the downsampled artificial data (ERA5-DownGAN), created by the GAN with a white rectangle for each percentile from the 1st to the 99th.

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 GAN (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 GAN and VHR_REA-IT, significant similarities emerge in the overall distribution, with the GAN 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 GAN 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 GAN-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.

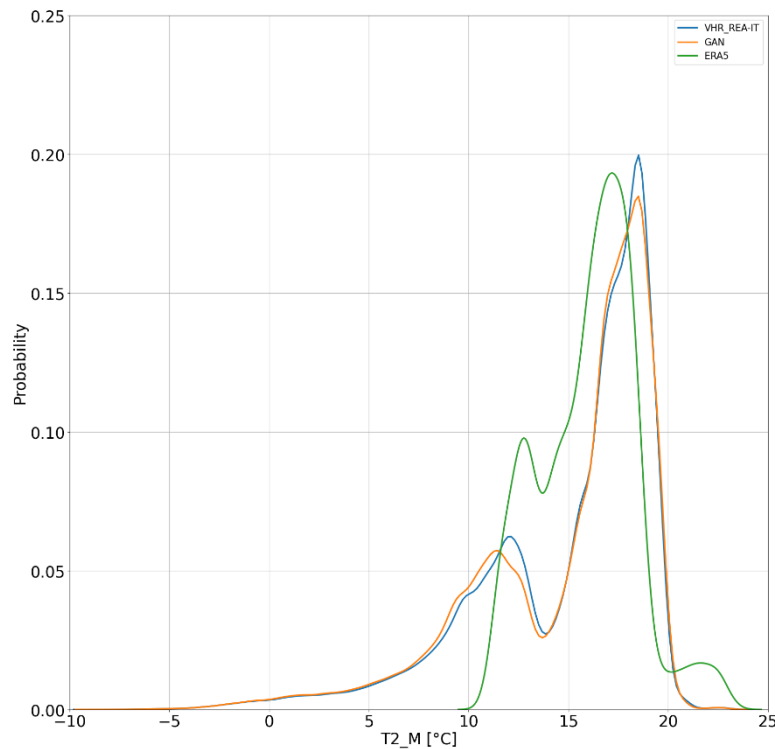


Fig. 3.9: Comparison of the probability density functions (PDF) of the 2m-temperature distribution for the ERA5, VHR_REA-IT, and ERA5-DownGAN datasets.

The analysis of the Probability Density Functions (PDFs) of the three 2m temperature datasets (Fig. 3.9), 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 GAN 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 GAN's ability to faithfully replicate the high-resolution temperature distribution. However, it is important to note that the GAN'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 GAN compared to the high-resolution dataset.

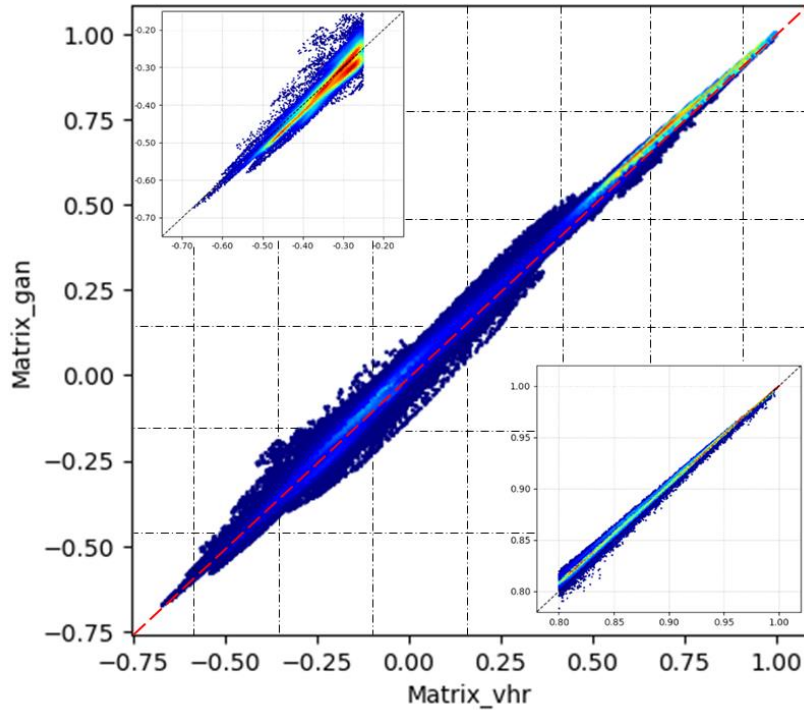


Fig. 3.10: Scatter diagram for the test period comparing the correlation of all pairs of grid points of the real dataset (VHR_REA-IT) with the correlation of all pairs of grid points of the synthetic dataset by GAN (ERA5-DownGAN).

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 GAN (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 Figure 3.I0. 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 GAN dataset (ERA5-DownGAN). The scatter plot reveals that the artificially generated high-resolution dataset by the Generative Adversarial Network (GAN) 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. 3.I0, at the top left), corresponding to the anti-correlation area between grid points, the GAN-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. 3.I0 at the bottom right, which depicts 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 GAN 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 GAN 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 GAN introduces a heightened level of coherence and structural organization in the generated data compared to dynamical downscaling. Nevertheless, these distinctions are minor when contrasted with the overarching pattern of the GAN-generated high-resolution dataset (ERA5-DownGAN), which demonstrates

a correlation between points closely mirroring that observed in the actual VHR_REA-IT dataset.

3.1.5 Computational Details

Computational results achieved through the implementation of the Generative Adversarial Network (GAN) bear paramount significance in the landscape of climate downscaling. All simulations were performed on Zeus supercomputer (<https://www.cmcc.it/research/super-computing-center>), hosted at the CMCC, equipped with Intel(R) Xeon(R) Gold 6130 CPU 2.10GHz, 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 Generative Adversarial Network (GAN) 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 GAN-based downscaling approach. The run-time for the training phase was approximately 4 hours to simulate 10 years (Tab. 3.Ia), and about 2.5 minutes to simulate 5 years during the testing phase (Tab. 3.Ib). This achievement assumes extraordinary importance when compared against the substantially protracted timescales required by conventional dynamic downscaling methodologies. 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 GAN-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.

years	lr	samples	batch size	epochs	time/training
10	0.0001	4000	100	100	~ 4 h

years	samples	batch size	time/test
5	2000	100	~ 2.5 min

Tab. 3.I: Computational details for the training period (a) and test period (b) on Zeus Supercomputer.

The training time, indicated in Tab. 3.1a, 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.

3.2 Downscaling of Total Precipitation

This section analyzes the application of the previously developed Conditional Generative Adversarial Network (cGAN), following the extremely promising results achieved in downscaling the 2-meter temperature field. These results demonstrated not only the GAN's ability to capture the mean values of the distribution but also its capability to accurately model extremes, as well as maintaining spatial and temporal consistency with the high-resolution field. The precision and reliability of the GAN in reconstructing the 2-meter temperature fields at a national scale, with very high levels of accuracy in both the training and test phases, laid the foundation for a new phase of experimentation. Thus, it was deemed appropriate to extend the application of the same GAN architecture to a more complex meteorological variable: total cumulative precipitation. Precipitation represents one of the most challenging variables to model, as it is affected by highly heterogeneous atmospheric phenomena and physical processes, influenced by topographic factors, local dynamics, and large-scale climatic conditions. Specifically, applying a statistical downscaling model like the GAN to the precipitation field is a considerable challenge, given the high spatial-temporal variability characterizing this variable, especially in a complex and diverse territory such as Italy. Indeed, Italy presents extremely heterogeneous meteorological conditions, with mountainous regions, plains, and coastal areas generating very different precipitation dynamics, making the problem of downscaling precipitation even more demanding and pioneering. Due to the complexity of the phenomenon, applying the cGAN to the precipitation field represents a crucial test of the flexibility and versatility of this architecture. While the results obtained for T_{2M} have undoubtedly been encouraging, it is necessary to demonstrate that the GAN can maintain competitive performance even for variables whose modeling is intrinsically more challenging. The ability of a GAN-based neural network used for downscaling to adapt to different contexts, such as domains and variables, is a key indicator of the robustness of the model.

The datasets used for training and testing the GAN in the precipitation field are consistent with those employed for T_{2M}: VHR_REA-IT (~2.2 km) for high resolution and ERA5 for the low-resolution counterpart (31 km). The training and testing periods are also consistent with the

previously analyzed atmospheric field, 1990–2000 for training and 2001–2005 for testing. This methodological continuity in transitioning from downscaling the 2-meter temperature field to the precipitation field allows for a comparative evaluation of the results, highlighting how the GAN performs in a variable and complex field like precipitation compared to a more regular field such as temperature. Notably, the national-scale application of the GAN model to the precipitation field has the potential to significantly improve the understanding of high-resolution precipitation dynamics, particularly in orographically complex regions such as the Alps, the Apennines, and Italian coastal zones. This provides a valuable and rapid tool for various applications, including weather forecasting, water resource management, and hydrogeological risk planning. The extension of the developed cGAN to the precipitation field is not merely a technical challenge but a critical step in verifying the model's ability to adapt to variables with complex dynamics, confirming it as a versatile and robust tool for statistical downscaling in climatology. Although the results obtained for the 2-meter temperature field (T_2M) were extremely promising, demonstrating the GAN's capacity to capture the essential characteristics of the variable with a high degree of accuracy, it is essential to proceed with caution when applying the same architecture to the precipitation field. Given the intrinsic complexity of the precipitation field, it is unrealistic to expect, from the initial experimental phases, that the GAN will achieve performance entirely comparable to those observed for T2M without an appropriate optimization process. In the case of precipitation, spatial and temporal variability is much more pronounced and nonlinear, implying that the network's sensitivity to hyperparameters could be significantly different. Consequently, sensitivity tests are essential to explore the hyperparameter space, including batch size, learning rate, loss function weights, and the number of epochs.

<i>Experiments</i>	Scaling	Logarithmic Transformation	Normalization Scheme	d_step	Pixel_loss weight	Pixel_adversarial weight
<i>EXP0</i>	-1 to 1	NO	Scheme 3 (Fig. I.2)	1	0.5	0.2
<i>EXP1</i>	-1 to 1	YES	Scheme 3 (Fig. I.2)	1	0.5	0.2
<i>EXP2</i>	-1 to 1	YES	Scheme 3 (Fig. I.2)	1	0.7	0.3
<i>EXP3</i>	-1 to 1	YES	Scheme 3 (Fig. I.2)	5	0.5/ 0.7	0.2/ 0.3
<i>EXP4</i>	-1 to 1	YES	Scheme 1 (Fig. I.2)	5	0.5	0.2

Tab. 3.2: Sensitivity experiments, including scaling methods, logarithmic transformation, normalization scheme, and discriminator update steps.

Additionally, given the more random and discontinuous nature of precipitation compared to temperature, it may be necessary to implement advanced regularization techniques to avoid overfitting, particularly in geographic areas characterized by irregular precipitation distribution. Further tests may involve temporal resolution to examine how the network responds to different temporal resolution configurations and to ensure that precipitation dynamics are captured with adequate detail, especially in areas subject to extreme phenomena or orographically complex regions. However, the primary goal of this study is to maintain the same innovative architecture developed and tested for 2m-temperature across other fields, making only modifications that do not alter the developed architecture, thereby demonstrating its versatility. Thus, this section reports on several sensitivity experiments conducted (illustrated in Tab. 3.2), regarding dataset normalization, loss function weights, and d_step to identify areas for improvement and guide the development of future versions of the network. Only through this iterative approach, combining sensitivity tests and hyperparameter optimization, will it be possible to achieve a versatile architecture while maintaining high performance even for complex fields like precipitation.

3.2.I Logarithmic transformation

The preliminary application for the downscaling of daily total cumulative precipitation (TOT_PREC) of the Conditional Generative Adversarial Network (cGAN) developed, establishes the EXP0. In this preliminary assessment, the architecture and hyperparameters were maintained in alignment with those utilized in the application for 2m-temperature. This methodological choice allowed for the assessment of the model's performance in a different data context while ensuring a consistent initial configuration. The same data normalization scheme was adopted, as illustrated in Fig. 2.4, and a scaling process was applied to constrain the values within the range of -1 to 1, ensuring the homogeneity and comparability of the input variables. The analysis of the results obtained from the EXP0, during testing phases, as reported in Fig. 3.13 revealed significant limitations in the model's performance when applied to precipitation data. Specifically, the GAN, using this scaling approach, is unable to adequately reconstruct the precipitation field even after 300 training epochs. In response to these observations, an additional step was integrated into the preprocessing process: the logarithmic transformation. This approach proved essential for the treatment of precipitation data, as these values frequently exhibit a highly skewed distribution characterized by extremely variable values, often concentrated at relatively low levels (as shown in Fig. 3.II). Unlike temperature data, for which a simple normalization between -1 and 1 is generally adequate, precipitation

data require a more sophisticated treatment. The logarithmic transformation helps to mitigate the influence of outlier values and stabilize the variance, thereby enhancing the model's ability to learn and generate meaningful patterns.

Numerous previous studies support this necessity (e.g. Obaid H.S. et al, 2019; Singh, D., and Singh, B., 2019), demonstrating that the adoption of advanced normalization techniques, including logarithmic transformation, can lead to a significant improvement in the performance of precipitation forecasting models. To address the issue of applying a logarithmic transformation to null values in the dataset, where no precipitation is recorded, this study utilizes the function $np.log1p$. This function calculates the natural logarithm of $1 + x$ for each element x in the array. Therefore, for a value of zero, the calculation becomes:

$$np.log1p(0) = \log(1 + 0) = \log(1) = 0$$

This function effectively manages zero values by transforming them into zero without the need for an additional offset. Unlike a previous study, Leinonen J. et al. (2020), where the authors added a small offset $\theta = 0.17$ to their data, found that the choice of this parameter did not significantly influence the results, our approach opts to directly use the $np.log1p$ function, which can be interpreted as applying an "offset" of 1. Furthermore, while both studies apply a logarithmic transformation, this work normalizes the data between -1 and 1, in contrast to Leinonen J. et al. (2020), who normalize their data between 0 and 1. This choice of normalization allows for a better representation of the variability in precipitation data in downscaling contexts.

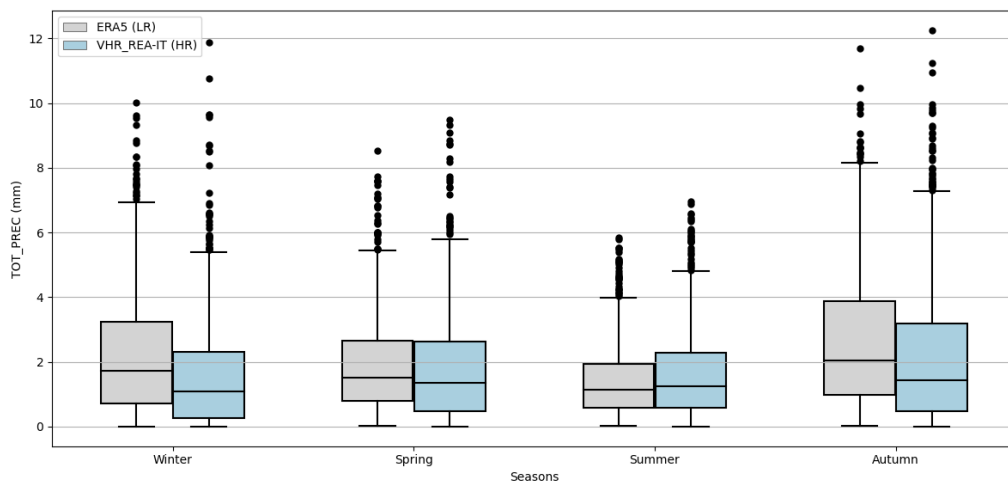


Fig. 3.II: Comparison of the total precipitation grouped for seasons (TOT_PREC) distribution for real low-resolution data (ERA5) with the gray rectangle, the real high-resolution data (VHR_REA-IT) with the blue rectangle.

The original datasets reveal distributions ranging from 0 to 7, with outliers reaching up to 12 for the VHR_REA-IT dataset, and from 0 to 8, with outliers extending to 13 for ERA5 (see Fig. 3.II). While scaling the data to a range of -1 to 1, as applied to the 2-meter temperature variable, effectively reduced the influence of outliers, it also resulted in significant compression of values near -1, thereby limiting their variability (see Fig. 3.I2). This issue became particularly apparent when evaluating the model's ability to reconstruct precipitation fields in the experiment (EXP0, Tab. 3.2), which utilized the same scaling configuration as for 2-meter temperature variable.

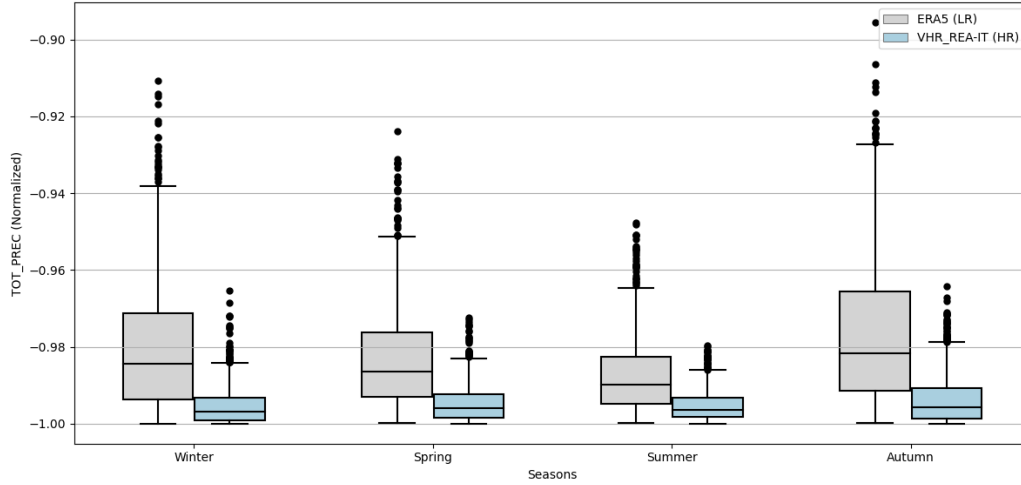


Fig. 3.I2: Comparison of the total precipitation scaled from -1 and 1 and grouped for seasons (TOT_PREC) distribution for real low-resolution data (ERA5) with the gray rectangle, the real high-resolution data (VHR_REA-IT) with the blue rectangle.

In response to the results obtained from EXP0 and the need to make the input provided to the GAN more symmetrical, we implement a logarithmic transformation on both input datasets (VHR_REA-IT and ERA5) before applying the scaling between -1 and 1. This logarithmic transformation is particularly advantageous for addressing highly skewed distributions commonly observed in precipitation data, resulting in a more balanced range of values between -1 and -0.3. Not only does this approach mitigate the impact of outliers, but it also enhances the normality of the data distribution, making it more closely resemble a Gaussian distribution.

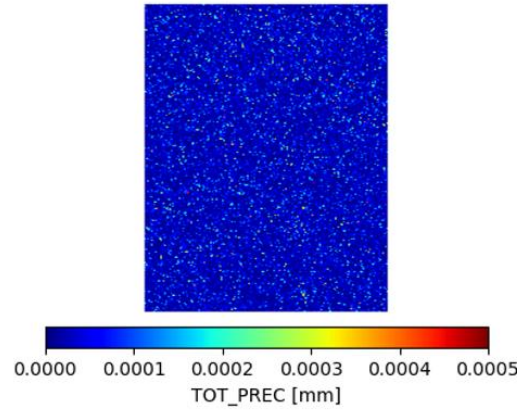


Fig. 3.I3: Example of one day from the test data set, generated by cGAN in **EXP0**.

The boxplots presented in Fig. 3.I4 clearly illustrate the marked improvement in value distribution following the combined application of logarithmic transformation and scaling to the range of -1 to 1. The boxplots in Fig 3.II, Fig. 3.I2, and Fig. 3.I4 illustrate the distribution of total precipitation (TOT_PREC) averaged spatially, obtained by aggregating daily precipitation data for each season.

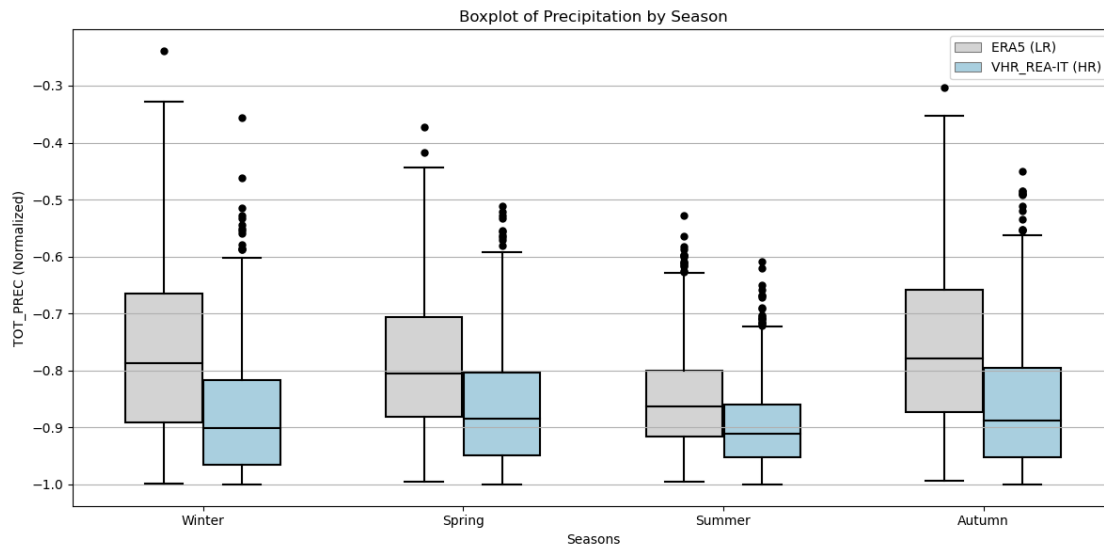


Fig. 3.I4: Comparison of the total precipitation with logarithmic transformation and scaling from -1 and 1; it is grouped for seasons (TOT_PREC) distribution for real low-resolution data (ERA5) with the gray rectangle, the real high-resolution data (VHR_REA-IT) with the blue rectangle.

The results of the EXP1, with new normalization (logarithmic transformation and scaling), are illustrated in Figure 3.15. It shows a comparison between the low-resolution real dataset (58x80

grid points), the high-resolution real dataset (680x535 grid points), and the downscaled high-resolution dataset generated by the ERA5-DownGAN model (680x535 grid points). The maps show the total precipitation field across the entire computational domain for three randomly selected days from the test period. The improvements relative to the baseline experiment (EXP0) are substantial in this new experiment (EXP 1). The GAN model is now capable of reconstructing precipitation patterns with a spatial distribution that closely mirrors the real high-resolution counterpart (VHR_REA-IT). Notably, the downscaled GAN dataset exhibits a clear enhancement in detail compared to the low-resolution ERA5 data (Fig. 3.I5). Furthermore, the values consistently fall within the range of the high-resolution real dataset, indicating a reliable representation of precipitation patterns.

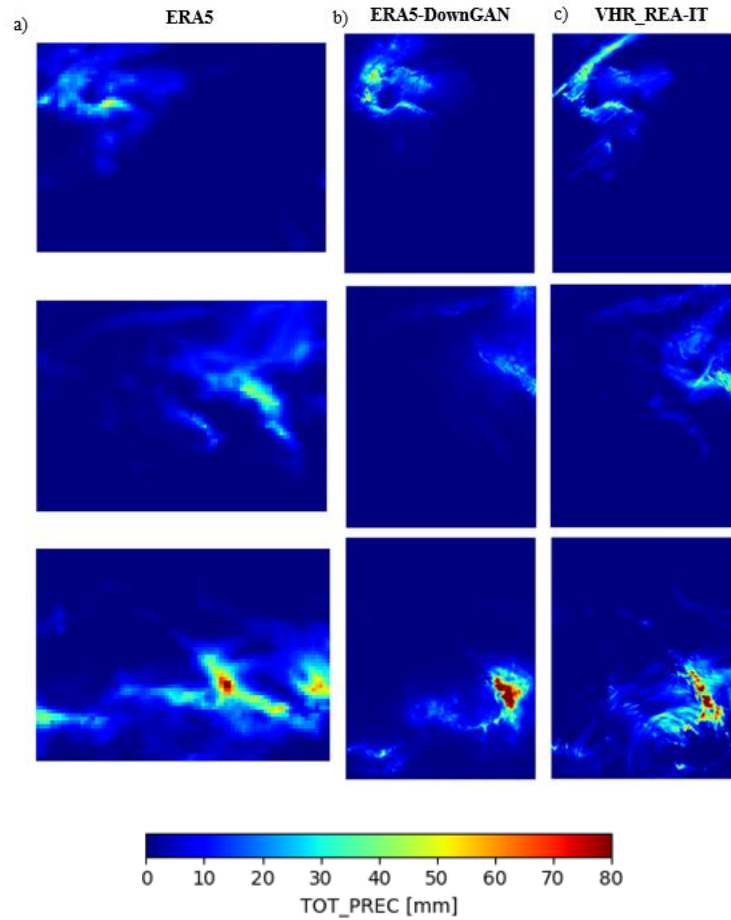


Fig. 3.I5: Examples of three random days from the test data set, generated by cGAN in **EXP1**.

The temporal trend of the total precipitation field (TOT_PREC) is illustrated in Fig. 3.I6 over the test period (2001-2005) for both high-resolution datasets, the real dataset (VHR_REA-IT)

and the synthetic dataset (ERA5-DownGAN), alongside the real low-resolution dataset (ERA5) across the entire study domain.

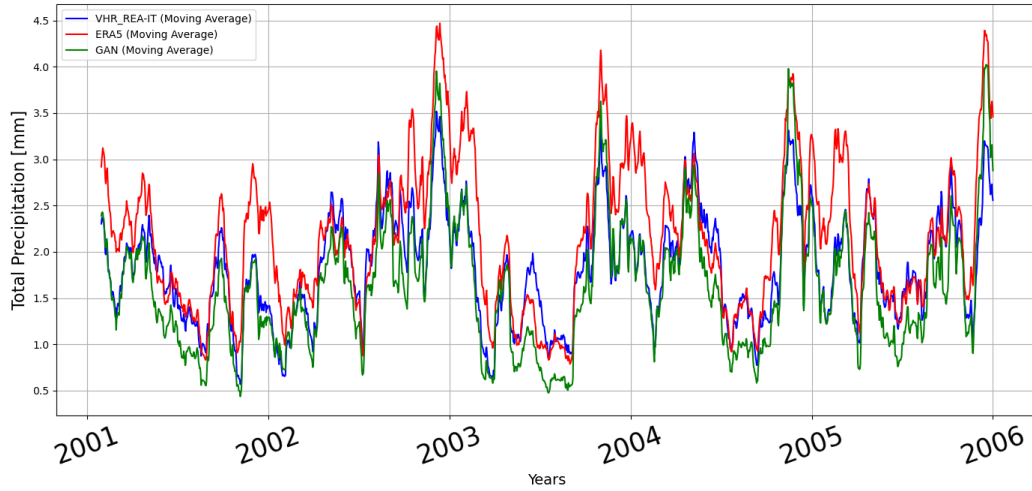


Fig. 3.16: Temporal evolution of the daily moving average of the total precipitation during the test period, calculated with a 30-day window. Mean daily spatial average of the real low-resolution data (ERA5), the real high-resolution data (VHR_REA-IT) and the downscaled artificially data (ERA5-DownGAN), created by the cGAN in EXP1.

The time series of daily total precipitation demonstrates a generally strong correlation among the datasets, albeit with some minor discrepancies in value ranges. Notably, ERA5 tends to deviate from VHR_REA-IT, either significantly or overlaps with it, often positioning itself at an intermediate level between the two datasets, with few exceptions. The displayed trends are smoothed using a 30-day moving average, which is derived from spatially averaging the precipitation data.

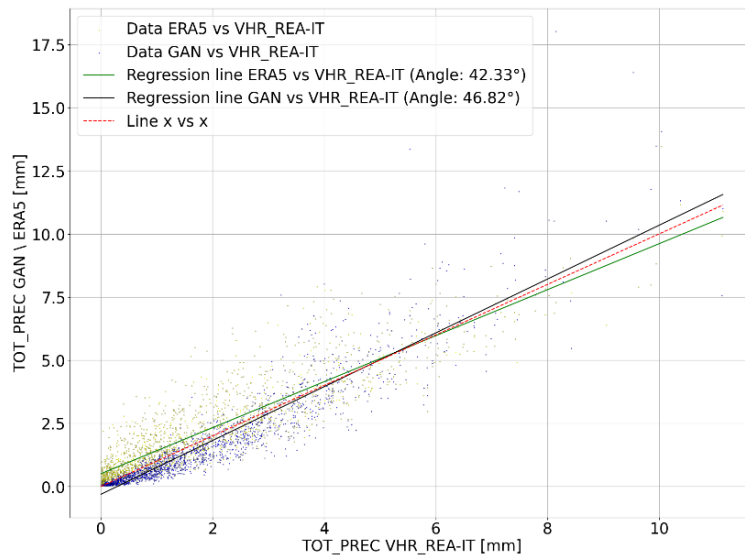


Fig. 3.17: Scatter diagram for evaluating the correlation between ERA5, VHR-REA_IT, and ERA5-DownGAN.

According to Fig. 3.I7, this suggests that the new high-resolution dataset exhibits a higher spatial-temporal correlation with the real high-resolution dataset (VHR_REA-IT) for lower precipitation values compared to the correlation observed between the low-resolution dataset (ERA5) and VHR_REA-IT, while the correlation for higher precipitation values remains similar between the two datasets.

The analysis of percentiles, from the 75th to the 99th, presented in Fig. 3.I7, allows for evaluating the extent to which the data generated by the GAN replicates the distributional characteristics of the high-resolution real dataset for total precipitation. The decision to focus on higher percentiles is justified by the inherently skewed nature of precipitation distributions, where low or zero values are more frequent. As a result, the examination of higher percentiles provides a more insightful evaluation of the GAN's ability to capture extreme events, which, while rare, are of significant meteorological importance.

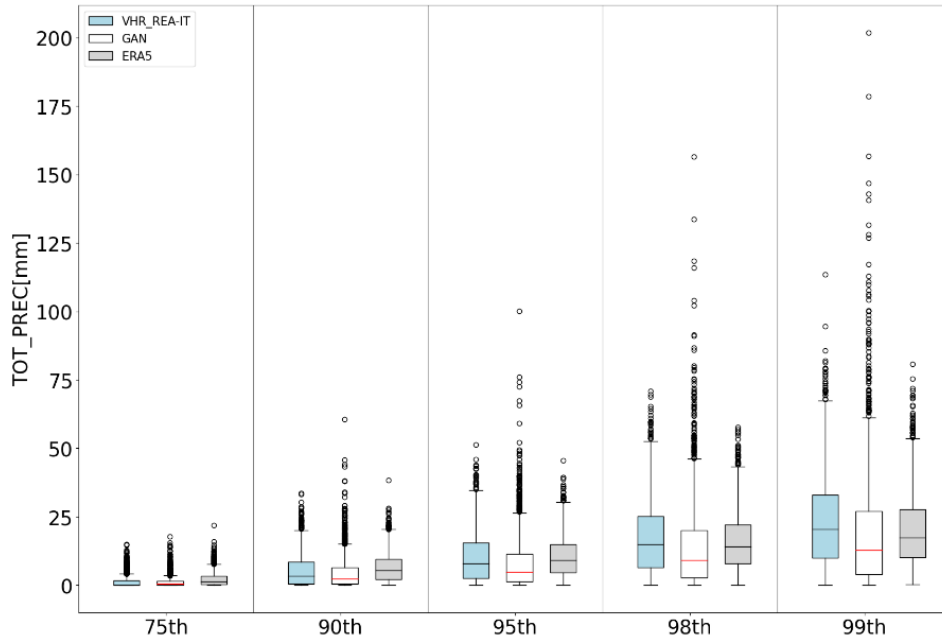


Fig. 3.I8: Comparison of the total precipitation (TOT_PREC) distribution for real low-resolution data (ERA5) with the gray rectangle, the real high-resolution data (VHR_REA-IT) with the blue rectangle, and the downscaled artificial data (ERA5-DownGAN), created by the cGAN in **EXP1** with a white rectangle for each percentile from the 75th to the 99th.

The analysis reveals that the GAN-generated dataset, ERA5-DownGAN, exhibits a lower median compared to the high-resolution real dataset (VHR_REA-IT) across percentiles 95 through 99. The data generated by the Generative Adversarial Network (GAN) model reveal a lower median in comparison to the high-resolution real dataset (VHR_REA-IT) within the 95th

to 99th percentile range. Nonetheless, the dispersion of central data, as represented by the interquartile range (IQR), is comparable to that of the VHR_REA-IT dataset. This observation indicates an enhancement in the distribution when contrasted with the ERA5 dataset, which exhibits reduced dispersion of central data. Moreover, the upper whiskers of the boxplot for the GAN-generated dataset align more closely with those of VHR_REA-IT than with those observed in ERA5. Importantly, the GAN model also demonstrates a greater prevalence of outliers compared to VHR_REA-IT, suggesting increased variability in the extreme values of the generated dataset. Several factors may contribute to this behavior. Firstly, GANs are engineered to learn the underlying distribution of real-world data; thus, if the model fails to accurately capture the patterns associated with extreme precipitation or rare occurrences, it may yield a lower median. Additionally, the presence of outliers exhibiting higher values than those in the real dataset may stem from the GAN's proactive exploration of the generated data space, resulting in the creation of rare and extreme samples. This phenomenon is particularly pronounced if the training dataset underrepresents high-intensity precipitation events, prompting the model to compensate by generating elevated values.

Finally, it is imperative to recognize the significance of hyperparameter selection, especially the weights assigned to the loss functions, as they can greatly influence the performance and outcomes of the GAN model. In light of these findings, a decision was made to further investigate the weights associated with the loss functions during the training process. Specifically, it was deemed appropriate to increase the weight of the pixel loss to encourage the generator to produce values more consistent with the actual distribution, thereby improving the median of the outputs. Concurrently, reducing the weight of the adversarial loss may contribute to stabilizing the generator's outputs, limiting the generation of extreme values and outliers. Therefore, the study proceeded with EXP2, aimed at optimizing the model's performance, ensuring greater coherence with the real data, and achieving a more accurate representation of the desired statistical characteristics.

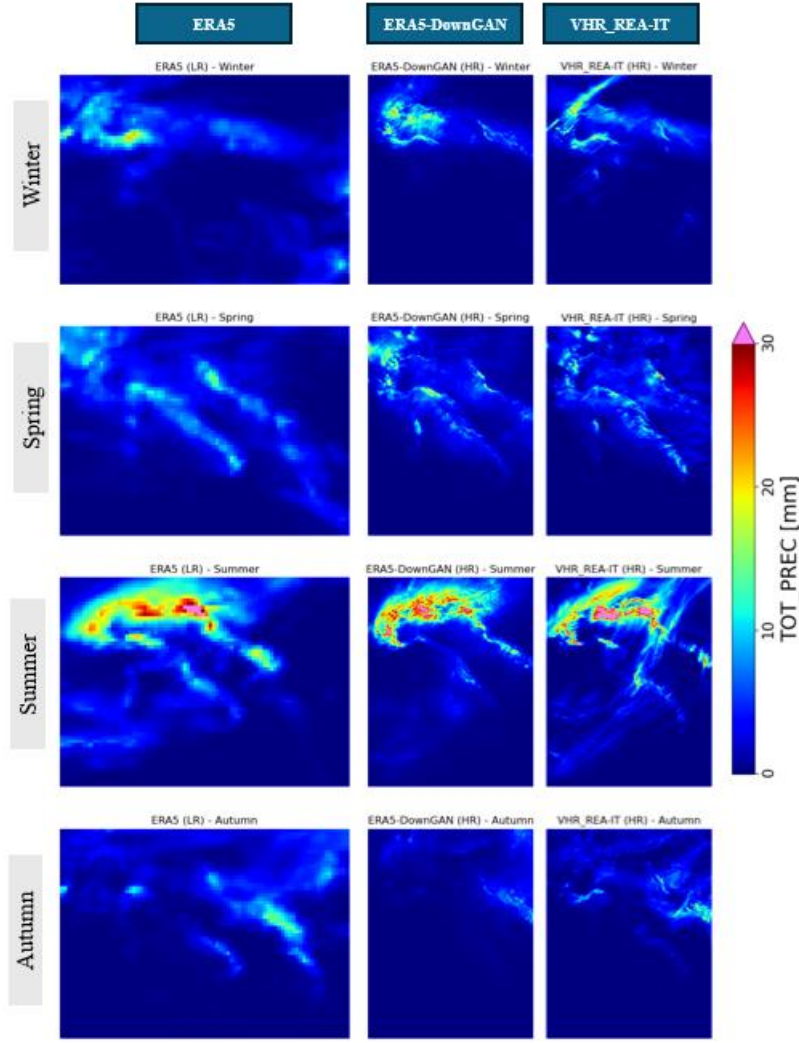


Fig. 3.I9: Comparison of seasonal total precipitation (TOT_PREC) for real low-resolution data (ERA5), the real high-resolution data (VHR_REA-IT), and the downscaled artificial data (ERA5-DownGAN), created by the cGAN.

The seasonal analysis is illustrated through maps (see Fig. 3.I9) depicting the mean precipitation values for each region across the three analyzed datasets. These maps highlight the GAN model's capability to effectively reconstruct the intricate details of precipitation patterns observed in the high-resolution dataset (VHR_REA-IT). In contrast, the ERA5 dataset, due to its lower resolution, demonstrates a limited ability to capture these details. Notably, in certain specific areas, the dataset generated by the GAN exhibits shapes that are more reminiscent of those reproduced by ERA5, albeit with precipitation values that are closer to those of VHR_REA-IT. Also in this context, increasing the weight associated with pixel loss at the expense of adversarial loss may further enhance the model's ability to accurately reproduce precipitation structures.

	Sum [m]			Median [mm]			Min [mm]			Max [mm]			Q1 [mm]			Q3 [mm]		
	ERA5	VHR_REA-IT	ERA5-DownGAN	ERA5	VHR_REA-IT	ERA5-DownGAN	ERA5	VHR_REA-IT	ERA5-DownGAN	ERA5	VHR_REA-IT	ERA5-DownGAN	ERA5	VHR_REA-IT	ERA5-DownGAN	ERA5	VHR_REA-IT	ERA5-DownGAN
January	1758,5	94852,3	94655,9	0,3	0,0	0,1	0,0	0,0	0,0	112,0	302,1	273,7	0,0	0,0	0,0	2,3	0,7	0,7
February	1424,4	80019,6	74108,5	0,2	0,0	0,1	0,0	0,0	0,0	102,8	288,1	300,7	0,0	0,0	0,0	2,1	0,6	0,6
March	1258,6	79912,2	74378,0	0,1	0,0	0,0	0,0	0,0	0,0	97,9	260,1	404,6	0,0	0,0	0,0	1,2	0,2	0,4
April	1628,9	120578,3	107603,1	0,2	0,0	0,1	0,0	0,0	0,0	79,0	270,2	414,2	0,0	0,0	0,0	2,1	1,1	0,9
May	1284,3	107369,3	82526,8	0,0	0,0	0,1	0,0	0,0	0,0	150,2	341,1	347,5	0,0	0,0	0,0	1,2	0,4	0,6
June	999,6	86589,1	59925,0	0,0	0,0	0,0	0,0	0,0	0,0	105,5	425,1	210,6	0,0	0,0	0,0	0,5	0,0	0,3
July	1064,0	82338,1	59259,9	0,0	0,0	0,0	0,0	0,0	0,0	166,2	336,1	323,6	0,0	0,0	0,0	0,5	0,0	0,2
August	1155,4	92973,5	75998,4	0,0	0,0	0,0	0,0	0,0	0,0	161,4	364,1	268,3	0,0	0,0	0,0	0,7	0,0	0,3
September	1595,5	107894,5	83958,3	0,1	0,0	0,1	0,0	0,0	0,0	186,6	467,6	349,4	0,0	0,0	0,0	1,5	0,2	0,5
October	1701,5	109808,7	107528,8	0,1	0,0	0,0	0,0	0,0	0,0	144,3	520,7	382,4	0,0	0,0	0,0	1,3	0,2	0,5
November	2170,1	124243,4	135420,6	0,4	0,0	0,1	0,0	0,0	0,0	147,2	330,5	527,2	0,0	0,0	0,0	2,9	0,9	0,9
December	2313,7	125886,6	128907,2	0,5	0,0	0,1	0,0	0,0	0,0	124,4	550,4	535,6	0,0	0,0	0,0	3,4	1,4	1,2

Tab. 3.3: Descriptive statistics for monthly totals, medians, maximum values, and interquartile ranges (Q1 and Q3) across the ERA5, VHR_REA-IT, and ERA5-DownGAN datasets.

A detailed statistical analysis of the precipitation datasets (see Tab. 3.3) reveals the significant impact of downscaling, revealing substantial differences between high-resolution datasets (ERA5-DownGAN and VHR_REA-IT) and the coarse-resolution ERA5. The results show that ERA5 consistently underestimates maximum values compared to VHR_REA-IT across all months, while recording a higher third quartile (Q3). This discrepancy can be attributed to ERA5's lower spatial resolution, which fails to capture local variations during extreme precipitation events. Consequently, maximum values are spread over larger areas, leading to an overestimation of Q3 (see Tab. 3.3) and the median compared to high-resolution datasets (particularly evident for the 90th and 95th percentiles in Fig. 3.I8). Boxplots further support this observation, showing that ERA5 exhibits reduced spread around the median for extreme percentiles (98th and 99th), in contrast to both ERA5-DownGAN and VHR_REA-IT. Conversely, ERA5-DownGAN better approximates VHR_REA-IT's maximum values, particularly in January, February, May, July, and December. Median values across all months for ERA5-DownGAN are almost identical to those of VHR_REA-IT, demonstrating greater agreement compared to ERA5.

We report in Tab. 3.3 the sum for ERA5 for informational purposes only, as the differing resolution and, consequently, the varying number of grid points render the values incomparable to those of the high-resolution datasets. Regarding cumulative precipitation sums, ERA5-DownGAN aligns closely with VHR_REA-IT, though with a slight underestimation, particularly in June and July. This improvement indicates that the GAN-based architecture successfully reconstructs precipitation fields that more accurately reflect the patterns observed in the high-resolution dataset. Low-resolution models, such as ERA5, tend to overestimate moderate or light precipitation events, contributing to an overall overestimation of total precipitation. The frequency analysis (Fig. 3.2I) further illustrates ERA5's overestimation of

light and moderate precipitation events. A logarithmic y-axis is used to better visualize higher precipitation classes, where frequency would otherwise appear small and compressed. This overestimation diminishes for more intense events (>20 mm), where ERA5 demonstrates more realistic behavior, although it still underestimates higher-intensity events. This discrepancy, as noted, is primarily due to the different resolutions of the datasets. ERA5, being a coarser-resolution model, struggles to capture localized, intense precipitation events, which can lead to overestimation of light and moderate events due to spatial averaging. In contrast, VHR_REA-IT shows greater accuracy for high-intensity precipitation events, likely due to its ability to resolve local features such as orography, which significantly affects precipitation patterns in mountainous or coastal regions. ERA5-DownGAN performs more similarly to VHR_REA-IT across different frequency classes, with the notable exception of values >200 mm, where it still exhibits some overestimation of very intense events. Not surprisingly, all three datasets show similar behavior in the low-intensity precipitation range. Events with negligible precipitation (0-1 mm) are frequent and are likely not significantly influenced by spatial resolution.

To assess the spatial distributions of the two high-resolution datasets, observed and those generated through GAN downscaling, we conducted an analysis of the averaged error metrics over the entire study period (Fig. 3.20). Our findings consistently demonstrate the tendency of the GAN model to underestimate precipitation, particularly in mountainous regions where bias and mean absolute error (MAE) can reach up to 10 mm. Specifically, the average bias is approximately -1 mm, while the mean MAE is around 1.4 mm. In terms of root mean square error (RMSE), the most pronounced values, reaching up to 15 mm, are observed in these complex terrains, with an overall average of approximately 4.5 mm. Despite this underestimation, the correlation map (Fig. 3.20) indicates a strong correlation over terrestrial areas, with values approaching unity in regions characterized by significant underestimation. This high correlation implies that, although the model consistently underestimates precipitation, the identified patterns remain coherent with the observed data. Such consistency is a positive indicator of the model's predictive capabilities, notwithstanding the negative bias. In contrast, the lowest correlation values are observed in marine areas, which presents a remarkable condition potentially attributable to the absence of sea assimilation in the COSMO-CLM model. This model generates the VHR_REA-IT data through dynamic downscaling, leading to predictions of precipitation patterns that diverge from those produced by ERA5.

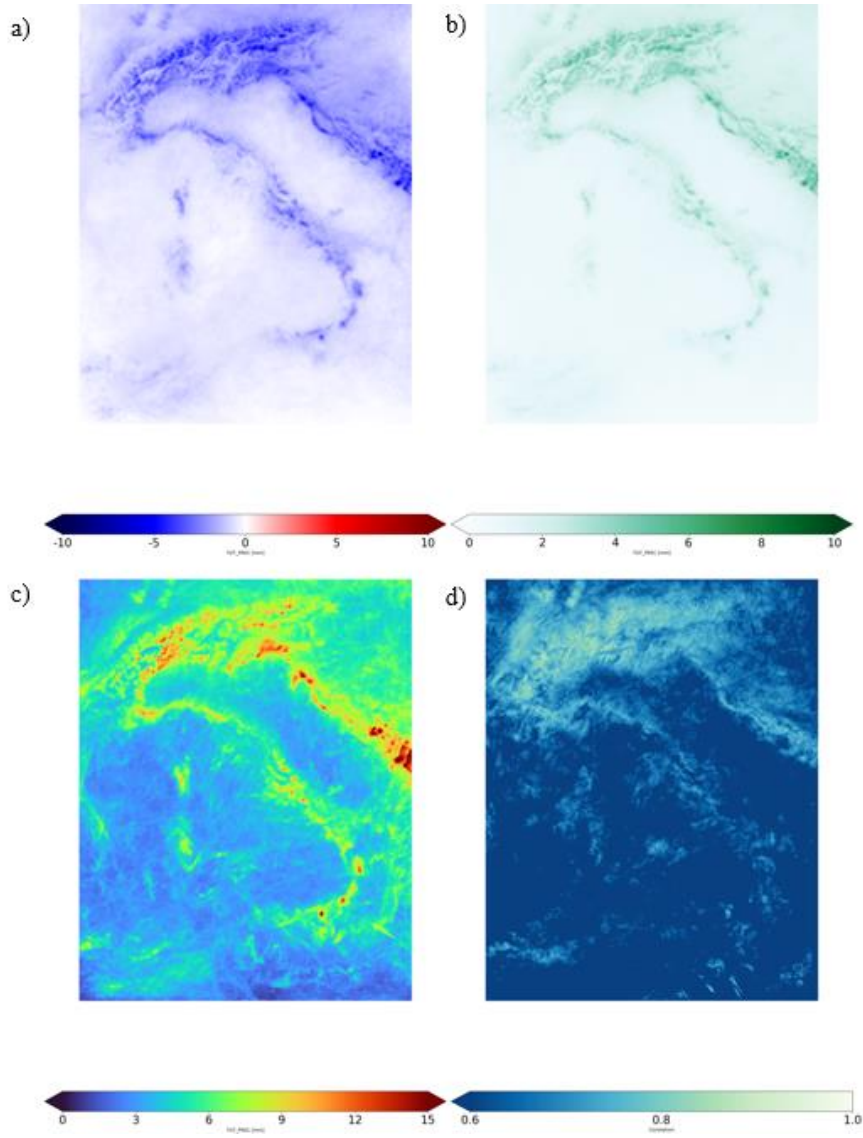


Fig. 3.20: Maps of time average. (a)-(d) Mean bias (BIAS), root-mean-square errors (RMSEs), mean absolute errors (MAEs), and correlation coefficients (CORRs), comparing downscaled daily total precipitation (ERA5-DownGAN) with the high-resolution dataset (VHR_REA_IT) for the test period (2001-2005).

The correlation between model estimates and observed data ranges from 0.6 to 1, exhibiting particularly high correlations in mountainous regions. This elevated correlation underscores the model's ability to maintain alignment with observed data, even in the face of precipitation underestimation. Such findings underscore the model's predictive capacity, suggesting that it retains a degree of robustness despite the inherent negative bias.

In conclusion, the downscaling process, particularly for precipitation fields, proves crucial in achieving a better correlation with high-resolution observational datasets compared to the coarse ERA5 field. However, the cGAN model in its current EXP1 configuration still shows

room for improvement, primarily due to its tendency to overestimate values associated with very intense events and its slight underestimation of median values relative to VHR_REA-IT.

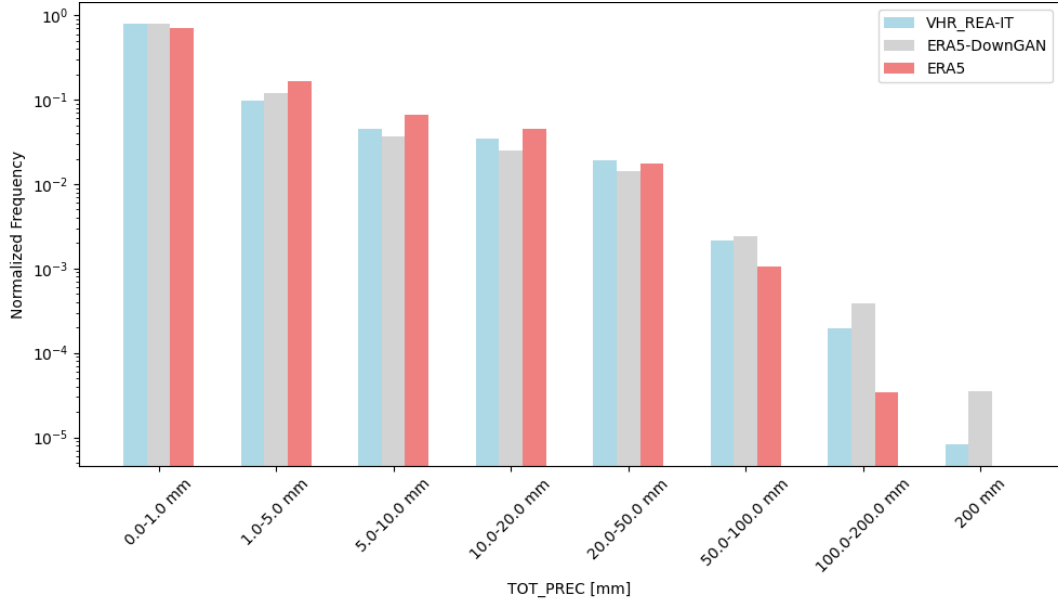


Fig. 3.2I: Frequency distribution of precipitation events across different intensity classes for datasets: VHR_REA-IT (blue), ERA5-DownGAN (grey) and ERA5 (coral). The use of a logarithmic y-axis highlights the occurrence of higher-intensity precipitation events.

This has prompted further investigations, focusing specifically on the model's hyperparameters, particularly the weights associated with the loss functions, which will be explored in the next section and examined in EXP2. Overall, the results from EXP1 demonstrate that the developed GAN is capable of reconstructing the complex patterns characteristic of the precipitation field, and that the architecture designed for 2m-temperature proves valid even for more complex variables like precipitation. Therefore, the EXP1 establishes a benchmark (EXP0) for evaluating the performance of subsequent experiments and improvements.

3.2.2 Exploring the Influence of Loss Weights and d_step in ERA5-DownGAN

Based on the results obtained from EXP1, we decided to explore different combinations of two fundamental parameters in GANs: the weights associated with the loss functions and the d_step. These combinations define three additional experiments, as detailed in the Tab. 3.5. In EXP2 and EXP3_1, both the pixel loss and adversarial loss weights are increased compared to EXP1 (from 0.5 to 0.7 and from 0.2 to 0.3), with the d_step set to 1 and 5, respectively. EXP3_2, on the other hand, extends EXP1 by modifying the d_step from 1 to 5.

The weights assigned to loss functions in a GAN play a crucial role in the optimization process, balancing different objectives. In the case of a GAN developed for downscaling, such as our cGAN, the two main loss functions are typically the pixel loss and the adversarial loss, which need to be carefully balanced to ensure both accurate reconstruction and high realism in the generated data. The pixel loss (often represented as L1 or L2) encourages the generator to produce outputs that closely match real data at the pixel level, emphasizing detail precision. On the other hand, the adversarial loss pushes the generator to produce outputs that the discriminator cannot distinguish from real data, thus promoting realism in the generated images or data samples.

<i>Experiments</i>	<i>Pixel_loss weight</i>	<i>Adversarial_loss weight</i>	<i>d_step</i>
EXP2	0.7	0.3	1
EXP3_1	0.7	0.3	5
EXP3_2	0.5	0.2	5

Tab. 3.4: Sensitivity experiments concerning loss weights and *d_step* parameters.

Increasing the weight of the pixel loss from 0.5 to 0.7, as done in experiments EXP2 and EXP3_1, places greater emphasis on reconstruction quality. This adjustment ensures that the generator focuses more on minimizing pixel-level differences between the generated images and the real ones. As a result, the model tends to generate outputs that more closely resemble high-resolution targets, a crucial aspect in applications like downscaling, where precision is essential for reliable predictions. However, an excessive increase in the pixel loss weight may result in overly smooth or blurry outputs, as the adversarial component is downplayed, reducing the sharpness and natural variability in the generated data.

Similarly, increasing the adversarial loss weight from 0.2 to 0.3 shifts the balance toward improving the realism of the generated images. With a higher adversarial weight, the generator receives a stronger signal to deceive the discriminator, which leads to more realistic and higher-frequency details in the results. However, challenges also arise in this case: if the adversarial weight is too high relative to the pixel loss, the model might prioritize generating visually convincing images, at the expense of accurately reconstructing the original data, thus reducing overall prediction accuracy. This trade-off between realism and reconstruction fidelity is a well-known challenge in GAN training. Properly balancing these two losses is critical and often

requires domain-specific fine-tuning. In our context, where both accuracy and realism are needed to produce a new high-resolution dataset through AI, comparable to that produced by physically based models, increasing the adversarial loss weight should improve the model's ability to capture complex patterns in high-resolution data.

Another parameter investigated in the experiments discussed here is `d_step`, which defines the number of discriminator updates for each generator update. In traditional GAN training, the discriminator and generator are updated alternately, generally with a 1:1 ratio, i.e., `d_step = 1`. However, by varying the `d_step` value, greater flexibility can be introduced into the training process: increasing the number of discriminator updates allows it to more accurately capture real details compared to generated data before the generator is updated. Setting `d_step` to 5, as in experiments EXP3_1 and EXP3_2, allows the discriminator to learn more effectively before the generator attempts to catch up. Increasing `d_step` to 5 means that the generator will face a stronger discriminator, which could lead to more refined outputs, as the generator will be pushed to produce data that is increasingly indistinguishable from real data. However, this increased strength of the discriminator presents a significant challenge as it could also lead to instability during training. The generator may struggle to converge if the discriminator becomes too powerful, causing issues such as mode collapse, where the generator produces only a narrow range of examples. Conversely, keeping `d_step` at 1, as in EXP2 and EXP1, ensures that the generator and discriminator remain more aligned, promoting a more balanced learning process. However, this may slow down the improvement of high-quality details and realism in the generated data.

The combination of these two parameter adjustments in the experiments listed in Tab. 3.4, namely the weights assigned to the loss functions and the number of discriminator updates, defines different learning dynamics for the developed conditional GAN. By increasing both the pixel loss and adversarial loss weights, the model is pushed to produce outputs that are simultaneously more accurate and realistic. Additionally, adjusting `d_step` from 1 to 5 enhances the discriminator's learning capability, creating a more challenging learning environment for the generator, potentially leading to more detailed and high-quality outputs. However, the balance between these parameters is delicate, and the results space was cautiously investigated through various experimental setups to avoid training instability or poor convergence. To enable a meaningful comparison across the experiments, while balancing computational costs with the errors associated with both the generator and discriminator, we determined that limiting the training to 200 epochs was sufficient for an adequate evaluation.

To evaluate both the visual quality and the internal consistency of the generated dataset, we analyzed the performance of various cGAN configurations (Table 3.2) in terms of the evolution of the mean value, standard deviation, spatial and temporal correlation, and Fréchet Inception Distance (FID) across different training epochs, ranging from 20 to 200 steps in increments of 20. The analysis of the mean value revealed that the distribution of synthetic data progressively converges towards that of real data as training epochs increase, thereby reducing initial discrepancies attributed to the model's incremental stability. Additionally, the values of the standard deviation approach those of the real dataset, indicating an improvement in the GAN's ability to capture the variability inherent in the real dataset. Measures of spatial correlation suggest an enhancement in structural coherence, while temporal correlation assesses the model's effectiveness in maintaining temporal consistency; this aspect is particularly critical for sequential datasets, even though our model does not explicitly account for temporal elements.

Models EXP1 and EXP2, configured with $d_step = 1$ and loss weights of 0.5 and 0.2 (EXP1) and 0.7 and 0.3 (EXP2), exhibit (Fig. 3.22) similar values for mean, standard deviation, spatial correlation, and FID, although they demonstrate distinct trajectories during training. This observation suggests that, despite having analogous foundational configurations, the differences in loss weights induce variations in the learning dynamics. Conversely, models EXP3_1 and EXP3_2, configured with $d_step = 5$ and utilizing the same loss configurations as EXP2 and EXP1, respectively, exhibit mean values and standard deviations that deviate more significantly from the real dataset, indicating reduced accuracy in the generated data. Furthermore, the spatial correlation in EXP3 models is lower, implying a lack of structural coherence in the synthetic data, potentially arising from the increased interval between parameter updates for the discriminator.

When examining configurations with identical loss weights (EXP1 and EXP3_2, EXP2 and EXP3_1), a notable similarity in learning trajectories emerges as epochs increase, with a slight shift in results attributable to variations in the d_step parameter. Notably, although nearly overlapping, EXP1 and EXP3_2 display differences in mean and standard deviation values around the 140th epoch, suggesting that the configuration with weights 0.5 and 0.2 is likely more sensitive to the choice of d_step . This sensitivity may be due to the diminished influence of the adversarial loss in configurations EXP1 and EXP3_2, characterized by $pixel_loss = 0.5$ and $adversarial_loss = 0.2$, which renders the model's balance more dependent on the frequency of discriminator updates compared to configurations with $pixel_loss = 0.7$ and $adversarial_loss = 0.3$.

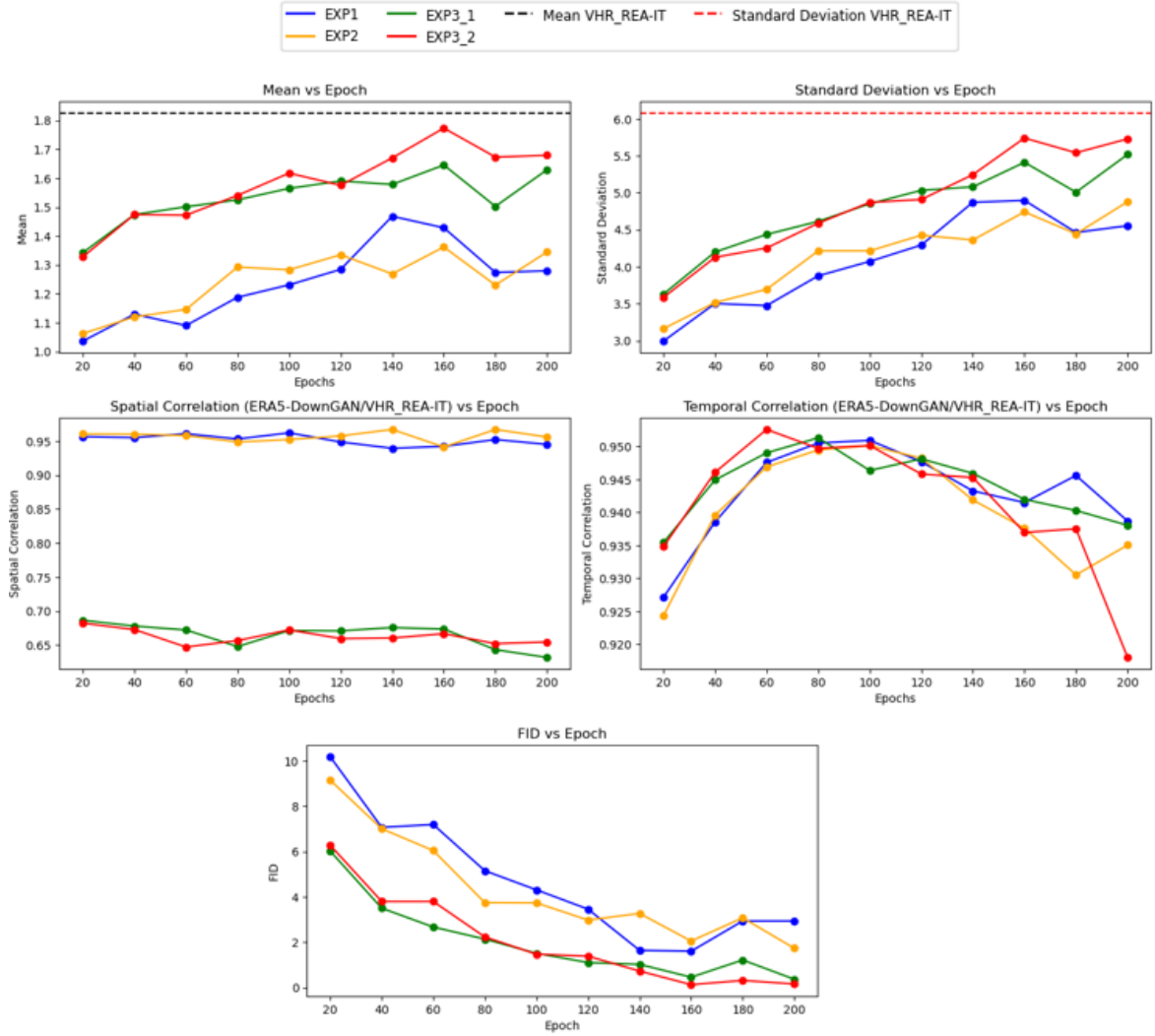


Fig. 3.22: Comparative analysis of performance metrics (mean, standard deviation, FID, spatial and temporal correlation) for experiments EXP1, EXP2, EXP3_1, and EXP3_2 as a function of epochs.

In Fig.3.23 we present the temporal evolution of the daily spatial average of total precipitation during the test period, comparing the benchmark configuration (EXP1) with the real high-resolution dataset (VHR_REA-IT) across various cGAN configurations. The data shown have been smoothed using a 30-day moving average. The results demonstrate that all cGAN configurations effectively capture the overarching trend of the real dataset (represented by the blue line), albeit with differing degrees of temporal shift. Notably, for the configurations with $d_step = 1$ (EXP1 and EXP2), the curves are nearly overlapping, indicating that the specific selection of loss weights has a minimal impact on the spatial average value in this context. However, both configurations utilizing $d_step = 1$ exhibit a tendency to underestimate precipitation when compared to the real dataset. Conversely, the experiments EXP3_1 and EXP3_2, which utilize $d_step = 5$, demonstrate a reduction in this underestimation, with certain

instances where the values produced by ERA5-DownGAN exceed those of the real dataset. The underestimation observed in the configurations with $d_step = 1$ can be attributed to the more frequent updates of the discriminator, which impose stricter constraints on the generator. This results in precipitation estimates that may be overly conservative. Conversely, configurations with $d_step = 5$ allow the generator to learn in a more flexible manner, thereby reducing the impact of immediate corrections from the discriminator. In this context, the generator is afforded the opportunity to explore a broader solution space, potentially leading to precipitation that, in certain instances, exceed the values of the real dataset.

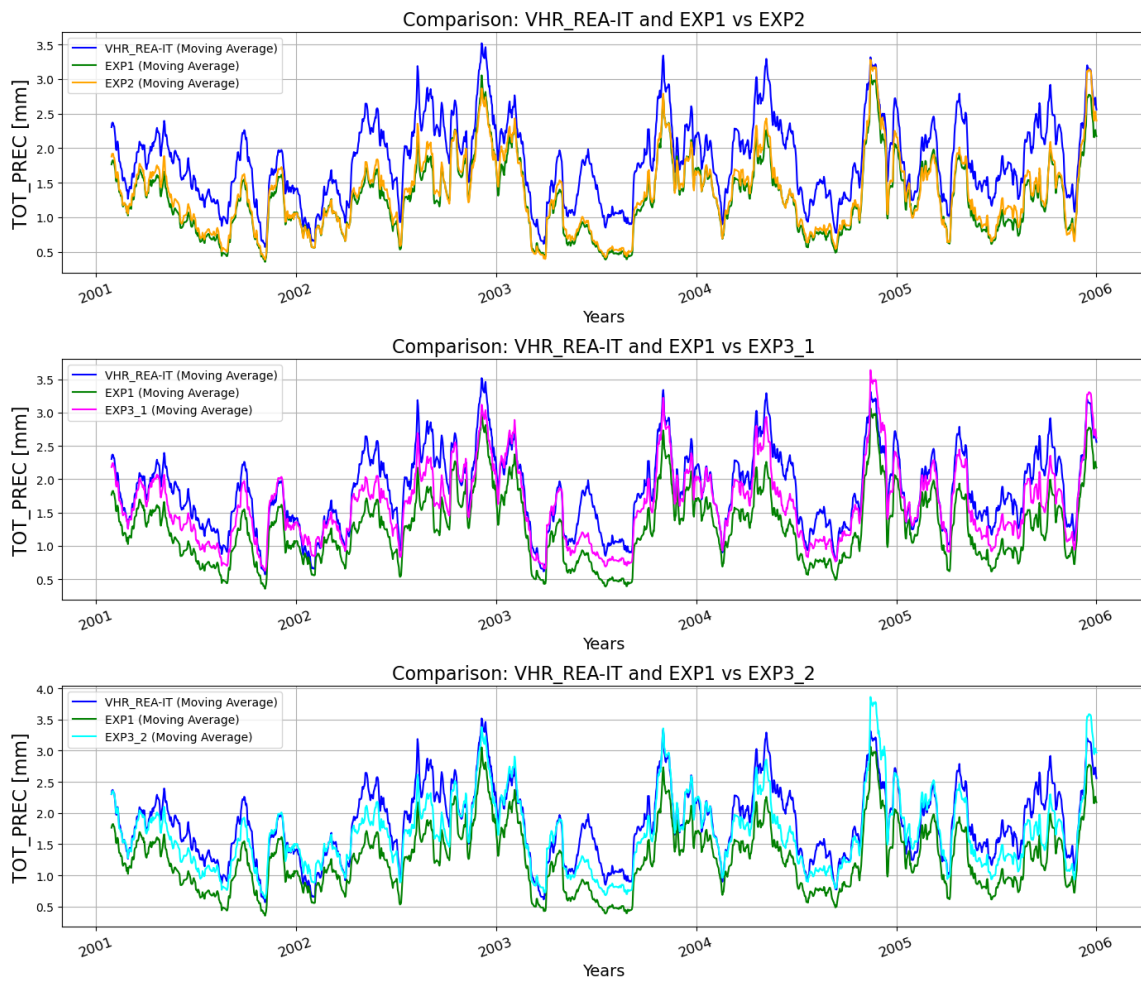


Fig. 3.23: Temporal evolution of the daily moving average of the total precipitation during the test period, calculated with a 30-day window. Comparison of the mean daily spatial average: benchmark (EXP1) and real high resolution dataset (VHR_REA-IT) with different cGAN configurations.

The analysis of boxplots across various cGAN configurations (Fig. 3.24) further corroborates the efficacy of the EXP3_1 and EXP3_2 experiments in generating mean values that are both higher and more closely aligned with those of the real dataset. Specifically, within the 90th to 99th percentile range, the boxplots for the EXP2, EXP3_1, and EXP3_2 configurations demonstrate a broader interquartile range (IQR), indicating enhanced dispersion in the central values. This finding underscores a more robust representation of extreme precipitation events in these configurations. Overall, the experiments employing a pixel loss of 0.7 and an adversarial loss of 0.3 reveal a distribution characterized by slightly elevated values, thereby achieving closer alignment with the real dataset compared to their counterparts utilizing a pixel loss of 0.5 and an adversarial loss of 0.2. These results suggest that the selected weightings in the loss functions significantly influence the model's capacity to effectively replicate extreme precipitation events.

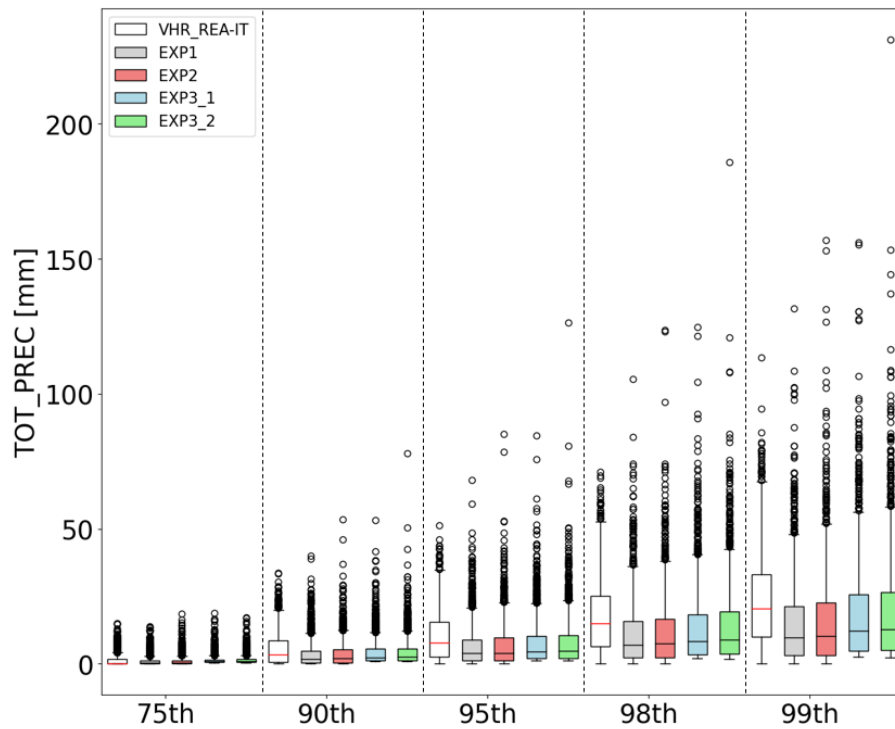


Fig. 3.24: Comparison of the total precipitation (TOT_PREC) distribution for real high-resolution data (VHR_REA-IT) with the white rectangle, and the downscaled artificial data (ERA5-DownGAN), created by the cGAN in EXP1, EXP2, EXP3_1 and EXP3_2 for each percentile from the 75th to the 99th.

To provide a comprehensive analysis of the performance of the various configurations implemented, Figure 3.25 presents a comparison of bias maps, mean absolute error (MAE), and root mean square error (RMSE) averaged over time, utilizing the VHR_REA-IT dataset as a

reference. The analysis indicates that EXP2, which adopted a pixel loss of 0.7 alongside an adversarial loss of 0.2, effectively reduced the general underestimation seen in EXP1, where the pixel loss was set at 0.5 and the adversarial loss remained at 0.2. This improvement was particularly evident in areas with complex orography. However, in regions where EXP1 exhibited a slight overestimation of precipitation values compared to the VHR_REA-IT dataset, EXP2 exacerbated this overestimation.

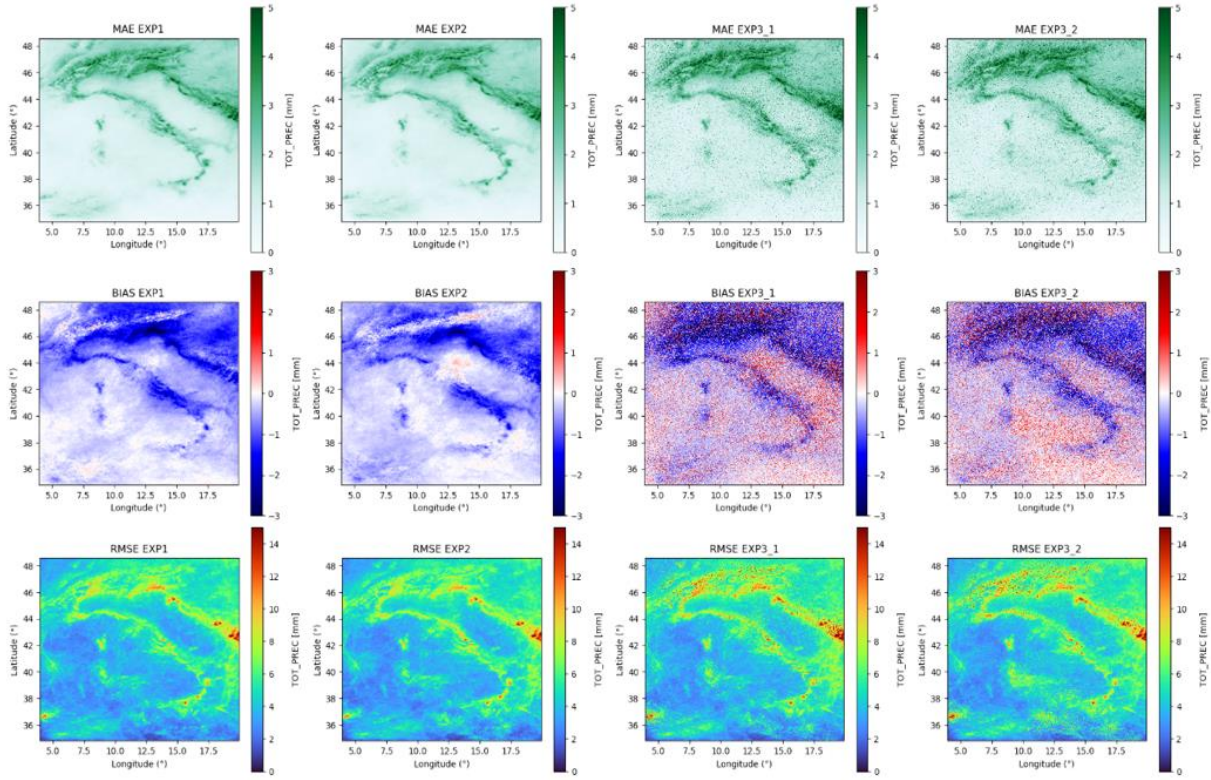


Fig. 3.25: Maps of time average. Mean absolute errors (MAEs), Mean bias (BIAS) and root-mean-square errors (RMSEs), comparing downscaled daily total precipitation for EXP1, EXP2, EXP3_1 and EXP3_2 (ERA5-DownGAN) with the high-resolution dataset (VHR_REA-IT) for the test period (2001-2005).

Despite these discrepancies, both MAE and RMSE values remained relatively stable between EXP1 and EXP2. These differences can be attributed to the increased weight assigned to pixel loss, which likely encouraged the model to prioritize the accurate reproduction of pixel values. While this approach aimed to enhance overall accuracy in regions where underestimation was prevalent, it may have inadvertently led to a “forcing” effect, elevating values in other areas. Notably, the bias values remain relatively low, ranging from -3 mm to +3 mm, while the MAE ranges from 0 to 5 mm. RMSE values, on the other hand, show a broader distribution, generally falling between 0 and 8 mm but peaking at 14 mm in specific locations, such as near Montenegro's Coastal Region.

Furthermore, in Experiments 3_1 and 3_2, where the discriminator update step (d_step) was increased from 1 to 5, the generated maps exhibited a marked decline in visual quality, characterized by undesirable artifacts. This phenomenon may arise from the excessive update frequency of the discriminator, which, updating more rapidly than the generator, can lead to a degree of overfitting to the real data, thereby limiting the generator’s ability to refine its representations. Consistent with Goodfellow et al. (2014) and Mescheder et al. (2017), a misalignment in the learning rates between the generator and discriminator may induce instability in image generation, ultimately compromising the quality of the final outputs.

In conclusion, the results suggest that EXP2, which utilizes a pixel loss of 0.7 and an adversarial loss of 0.3, represents a well-balanced compromise, demonstrating a slight mitigation of the underestimation observed in EXP1. In particular, EXP2 exhibits improved performance compared to EXP1 in extreme percentiles (Fig. 3.24). Although increasing the discriminator update steps (d_step) to 5 in EXP3_1 and EXP3_2 leads to better alignment of mean values and standard deviations with the real dataset compared to EXP1 and EXP2, this adjustment significantly compromises the structural coherence of the generated outputs. Consequently, at epoch 200, EXP2 emerges as the optimal configuration, providing the best trade-off between accuracy and stability among the evaluated experimental setups.

3.2.3 An Alternative Normalization Scheme

The complexity of precipitation fields motivated an investigation into alternative normalization schemes could enhance GAN output performance. We introduced Scheme 1 in Fig. 2.4, a method that scales all input data based on the minimum and maximum values obtained from a bilinear interpolation of the ERA5 dataset onto the VHR-REA-IT grid (ERA5_interp-VHR). Minimum and maximum values were computed separately for both training and testing periods’ temperature distributions, allowing ERA5-DownGAN to be denormalized according to ERA5_interp-VHR data from the testing period. Benchmark results from EXP1 were then compared with EXP4, which applies this new normalization strategy, to evaluate potential improvements in model accuracy and generalization.

Specifically, for both experiments (EXP1 and EXP4), the trends of the mean value, standard deviation, spatial and temporal correlations were analyzed in comparison to the real dataset (VHR_REA-IT), along with the Fréchet Inception Distance (FID) over the epochs ranging from 20 to 200 in increments of 20 (see Fig. 3.26). Additionally, frequency histograms for the various precipitation classes and maps of bias, mean absolute error (MAE), and root mean square error

(RMSE) were produced, averaging the results over the comparison period between the two experiments against the high-resolution dataset.

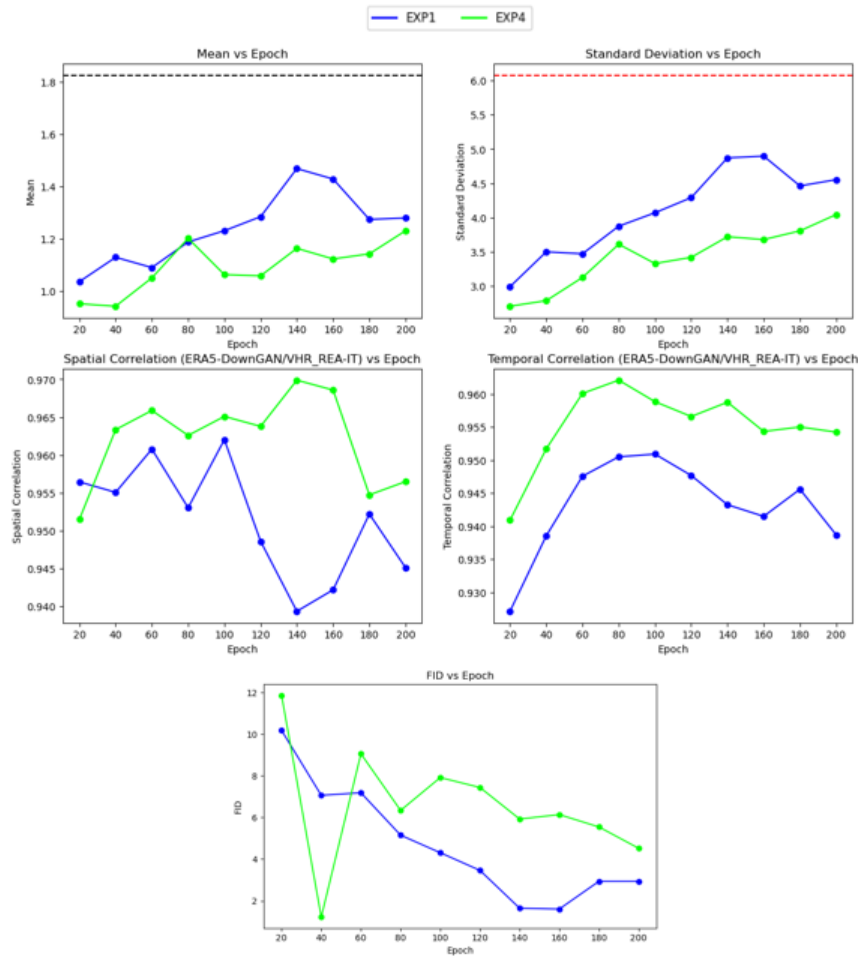


Fig. 3.26: Comparative analysis of performance metrics (mean, standard deviation, FID, spatial and temporal correlation) for experiments EXP1 and EXP4 as a function of epochs.

Although the spatial and temporal correlations in EXP4 are slightly higher than those in EXP1 (see Fig. 3.26), the mean and median values in EXP4 are further from the actual dataset compared to EXP1, and the Fréchet Inception Distance (FID) is also elevated. This suggests that the cGAN implemented in EXP4, despite its comparable performance to EXP1, encounters greater challenges in aligning with the mean values and distribution of the test period, which impacts its generalization capabilities. This discrepancy may be attributed to the training on data that are inherently smoother, resulting from the bilinear interpolation of ERA5 onto the VHR_REA-IT grid. In contrast, EXP1 was trained directly on the actual high-resolution dataset VHR_REA-IT. Consequently, the training values in EXP1, particularly those representing

frequency classes associated with high precipitation, are more representative of real-world conditions, facilitating improved generalization during the testing phase.

The observation that EXP4 fails to capture the higher values of the distribution as effectively as EXP1, resulting in a lower mean and reduced dispersion, initially suggested by Fig. 3.26, is confirmed by the frequency distribution analysis presented in Fig. 3.27. This analysis highlights the limitations of EXP4 in accurately representing the full spectrum of precipitation events, particularly those characterized by high intensity.

In particular, starting from 50 mm, EXP4 records a lower frequency not only compared to the actual dataset but also relative to EXP1. The most significant differences emerge from 100 mm, with a further deterioration at 200 mm, where only EXP1 presents values in this class, confirming its ability to represent the more intense event classes, a capability lacking in EXP4. An inversion of trend is observed exclusively for the weak event class, ranging from 1 to 5 mm, where both EXP1 and EXP4 show a higher frequency compared to the actual dataset. In contrast, for the intermediate precipitation classes (from 5 to 50 mm), both datasets exhibit a slightly lower frequency relative to VHR_REA-IT.

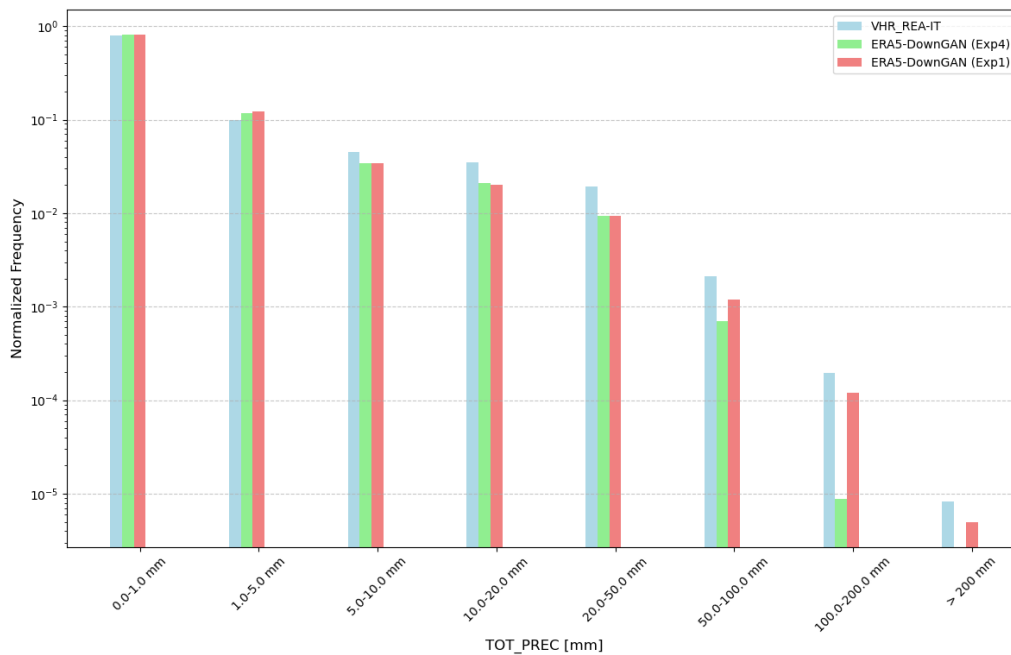


Fig. 3.27: Frequency distribution of precipitation events across different intensity classes for datasets: VHR_REA-IT (blue), ERA5-DownGAN EXP4 (green) and ERA5-DownGAN EXP1 (coral). The use of a logarithmic y-axis highlights the occurrence of higher-intensity precipitation events.

Figure 3.28 reveals that the two experiments exhibit no substantial differences in terms of mean bias, MAE, and RMSE. However, a notable reduction in bias is observed in EXP4 within the Friuli Venezia Giulia region and the Central Apennines, while a greater bias is evident in the mountainous areas of Croatia and Montenegro. This finding suggests that there is not a uniform mitigation of overestimation across all mountainous regions, contrasting with the results observed in EXP2 (Figure 3.25). Therefore, although EXP4 demonstrates a specific reduction in bias in certain locations, it does not ensure superior performance in mountainous contexts, where EXP2 provided a significant advantage over the benchmark EXP1.

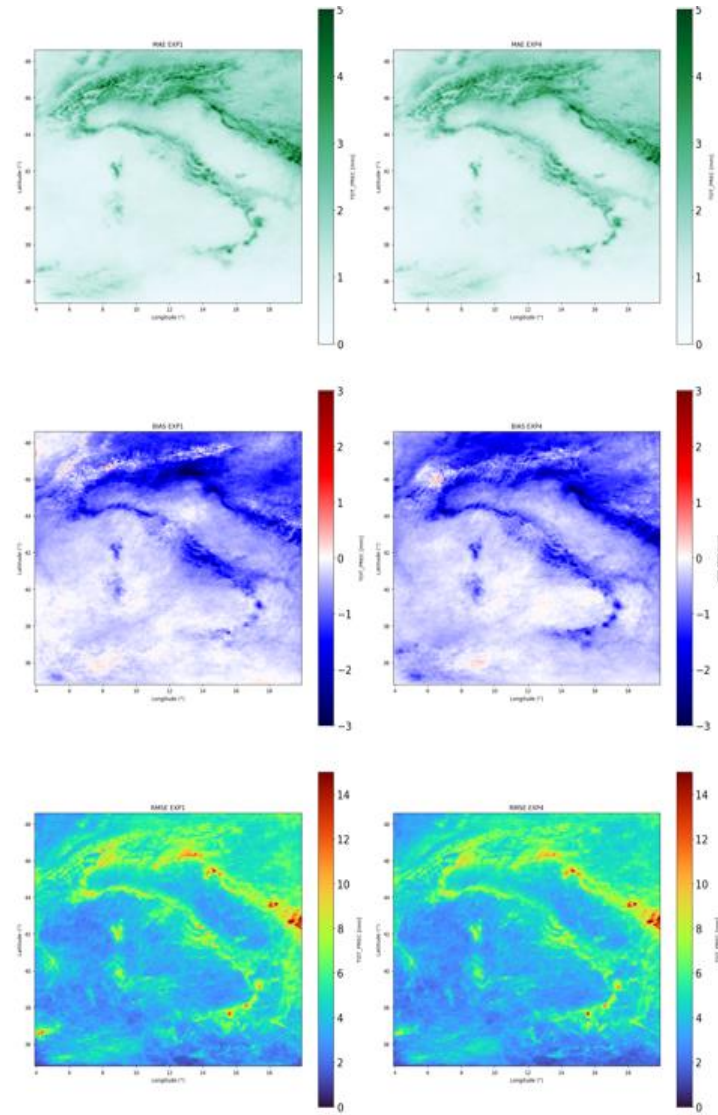


Fig. 3.28: Maps of time average. Mean absolute errors (MAEs), Mean bias (BIAS) and root-mean-square errors (RMSEs), comparing downscaled daily total precipitation for EXP1 and EXP4 (ERA5-DownGAN) with the high-resolution dataset (VHR_REA_IT) for the test period (2001-2005).

In conclusion, the results obtained from the comparison between EXP1 and EXP4, highlighting the differences in the use of Schema 3 and Schema 1 for normalization (Figure 2.4), further emphasize the importance of selecting an appropriate normalization scheme during the training of GANs. The findings confirm that Schema 3 proves superior in accurately representing the actual distribution of the atmospheric field under investigation, particularly at higher values. Therefore, to achieve more accurate and reliable precipitation forecasts, it is crucial to adopt more sophisticated normalization approaches, such as the one employed in EXP1. This supports the notion that effective normalization, directly based on high-resolution data and devoid of pre-processing interpolations, can enhance the model's generalization capabilities and overall performance.

3.2.4 Comparison of All Configurations

In this chapter, several experiments were conducted to evaluate the effectiveness of our cGAN architecture initially developed for the downscaling of 2-meter temperature (T_{2M}) and later adapted to handle precipitation fields (TOT_PREC), a more complex meteorological variable. The preliminary experiment (EXP0) indicated that while scaling between -1 and 1 was effective for T_{2M} , precipitation data required additional pre-processing to correct its highly skewed distribution. The introduction of a logarithmic transformation in EXP1 significantly improved the GAN's ability to replicate the high-resolution precipitation field (VHR_REA-IT), notably enhancing the model's precision in capturing spatial patterns across various intensities (Figs. 3.15, 3.19). This new pre-processing step enabled the model to generate precipitation distributions that closely resemble real high-resolution data, as demonstrated by its capacity to reconstruct daily and seasonal precipitation patterns. Percentile analysis (Fig. 3.18) highlighted that the GAN accurately represents the distributional characteristics of the real dataset, even at higher percentiles (from the 75th to the 99th), although it exhibits a more pronounced underestimation in the extreme classes (98th to 99th percentiles).

Nevertheless, the generated dataset maintains a realistic distribution and strong temporal correlation with high-resolution real data, showing substantial improvements over the low-resolution ERA5 dataset (Fig. 3.16). Using EXP1 as a benchmark, further parameter adjustments were made in EXP2 and EXP3 to optimize GAN performance. Increasing the weights of pixel and adversarial losses (from 0.5 to 0.7 and from 0.2 to 0.3, respectively) in EXP2 yielded marginal improvements in bias, particularly in complex orographic regions, with a closer alignment of mean values and standard deviations between ERA5-DownGAN and

VHR_REA-IT, though generating much higher outliers in the synthetic dataset than seen in the real dataset. In the configuration with `d_step` set to 5 (EXP3), aliasing phenomena emerged, indicating that a `d_step` value of 1 is crucial for maintaining spatial fidelity. The configuration with increased weight settings demonstrated a stronger dependency on the `d_step` setting.

Comparative analyses of seasonal and cumulative precipitation (Table 3.3, Figs. 3.19, Fig. 3.20) confirm the robustness of the GAN in reconstructing complex atmospheric patterns. The ERA5-DownGAN dataset successfully approximated the monthly cumulative precipitation of VHR_REA-IT, showing good correspondence across all months, with only slight underestimations during the summer months and good adherence to spatial structures in areas with intense events. Despite the presence of biases in orographically challenging areas, high correlation values over land areas confirm the GAN's ability to replicate trends in the VHR_REA-IT dataset (Fig. 3.20).

In summary, the results indicate that the developed GAN architecture, with the addition of specific pre-processing and targeted parameter optimization, performs downscaling of precipitation fields while maintaining good consistency with the high-resolution real dataset. It is important to note that the developed cGAN demonstrates sufficient flexibility to handle complex atmospheric variables such as precipitation, highlighting its potential applicability beyond T2M downscaling. The study further suggests that, for precipitation fields, the optimal number of training epochs (300) is higher than for T2M (100 epochs), without necessitating structural modifications to the GAN architecture. This adaptability to various meteorological variables, combined with the ability to replicate not only mean values but also specific patterns of the real high-resolution dataset while maintaining spatial and temporal coherence, underscores the potential of the developed methodology as a versatile tool for atmospheric downscaling applications.

3.2.5 Computational Details

All simulations were performed on the Juno supercomputer at the Euro-Mediterranean Center on Climate Change (CMCC), as the Zeus system was no longer available. Juno is equipped with advanced computational resources, including 12,240 processing cores and 20 NVIDIA A100 GPUs, which facilitate the tackling of ambitious challenges related to scalability and computational complexity. Juno's peak performance reaches 1134 TFlops, with each compute node featuring 72 cores and 512 GB of memory.

	years	lr	samples	batch size	epochs	time/training
EXP1	10	0.0001	4000	100	100	4 h 57 min
EXP2	10	0.0001	4000	100	100	5 h 28 min
EXP3_1	10	0.0001	4000	100	100	3 h 7 min
EXP3_2	10	0.0001	4000	100	100	1 h 9 min
EXP4	10	0.0001	4000	100	100	2 h 13 min

Tab. 3.5: Computational details for the training period on Juno Supercomputer.

In the context of this research, it is essential to highlight that the implementation of the Generative Adversarial Network (GAN) was carried out sequentially following a serial approach. The execution of the code efficiently utilized the available resources (1 compute node, with 72 cores), resulting in a significant increase in efficiency in the GAN-based downscaling approach. The average execution time for the training phase (for 100 epochs) was approximately ~3 h 22 min hours to simulate 10 years, considering the average time taken for the different configurations reported in the Tab. 3.5, and ~2.5 minutes to simulate 5 years during the testing phase (for all experiments). The execution of the precipitation downscaling experiments on Juno also represented an additional challenge within the work, allowing for the assessment of the portability of the developed architecture across different computational environments. This confers replicability to the methodology, enhancing its flexibility and applicability. This adaptability is crucial to ensure that the downscaling framework based on the developed cGAN can be effectively utilized in various research contexts and operational settings.

3.3 Potential Enhancements

Experiments related to the downscaling of precipitation fields to approximately 2.2 km have produced a dataset generated through statistical downscaling that is highly faithful to the high-resolution VHR_REA-IT dataset, produced through dynamic downscaling and representative of our reality. However, there are still significant margins for improvement. In particular, the use of convolutional layers could prove promising, either by integrating them into the current architecture or by replacing the linear layers currently in use. A convolutional

architecture enables faster training and broader model applicability (Park, S., & Shin, Y.-G., 2022), making the GAN suitable for downscaling over large geographic areas. Moreover, convolutional layers introduce translation invariance, allowing the model to recognize patterns regardless of their precise position in the data matrix. This characteristic is particularly advantageous for precipitation fields, where the specific locations of precipitation cores can vary rapidly.

ERA5-DownGAN has already demonstrated high fidelity in reproducing the high-resolution VHR_REA-IT dataset of total accumulated precipitation; however, there remains room for improvement, particularly in the representation of extreme precipitation events. The underestimation observed in the upper quantiles, between the 95th and 99th percentiles, highlights the need for further refinement of the learning process to ensure that the cGAN-generated field aligns more closely with the reference dataset, even during extreme events. One potential strategy involves incorporating a loss function based on Extreme Value Theory (EVT), such as the Extreme Value Loss, which has been shown to enhance deep learning models' ability to capture distribution tails by explicitly modelling extreme events (Ding, D. et al., 2019). Alternatively, approaches that separate the training of extreme values from the rest of the dataset may prove beneficial (Allouche, M. et al., 2022). The implementation of these developments would enable a more physically consistent representation of precipitation fields during extreme events, further strengthening the applicability of the proposed model.

Additionally, another interesting experiment could involve using a radar precipitation dataset as a high-resolution source for model training, instead of the current reanalysis. Radar datasets can incorporate information about localized meteorological phenomena that may not be adequately represented in reanalysis, thereby ensuring greater reliability in the downscaling process. This combination of advanced technologies and data could thus lead to significant progress in generating a high-resolution dataset through the developed cGAN, resulting in a deeper understanding of the relationships between global and local scales, particularly in cases of complex atmospheric dynamics.

4 | Expanding the Application of cGAN

Downscaling to U.S. Geographical Domain

This chapter presents the downscaling experiment conducted using a conditional Generative Adversarial Network (cGAN) developed for 2-meter temperature data across the United States (U.S.) during an internship at the National Center for Atmospheric Research (NCAR). We begin by discussing the significance of having a flexible downscaling method that can be applied to various geographical domains. Subsequently, we provide a detailed description of the cGAN architecture, emphasizing the preservation of the original design while increasing the number of neurons to accommodate the extensive spatial scale of the domain. Additionally, we explore the importance of the number of training epochs, highlighting its critical role in ensuring convergence of the results. The chapter details the datasets utilized, particularly the high-resolution Weather Research and Forecasting (WRF) data provided by NCAR. Overall, this chapter offers a comprehensive overview of how the cGAN-based approach was adapted to enhance downscaling for high-resolution atmospheric data in diverse contexts.

4.1 Context and Motivation for the Case Study

The ability of the developed downscaling method, which employs a conditional Generative Adversarial Network (cGAN), to perform downscaling across various geographical regions and to incorporate diverse input datasets is a key feature that highlights the flexibility of the approach. Having established the architecture's capability to downscale complex variables, this chapter focuses on its adaptability to different domains, particularly those significantly larger than our initial case study of the Italian Peninsula.

During an internship at the National Center for Atmospheric Research (NCAR), we directed our efforts toward applying statistical downscaling AI-based to the entire territory of the United States (Fig. 4.1; Tab. 4.1). We utilized data from the Weather Research and Forecasting (WRF)

model, configured by NCAR at a resolution of ~ 4 km. This high-resolution dataset served as the primary input during the training phase of our architecture, while the low-resolution counterpart came from ERA5 at ~ 31 km. This experiment presented several challenges for the developed method: adapting to various geographical contexts, maintaining flexibility in handling diverse datasets, and ensuring the model's portability across different supercomputing environments. Access to the Cheyenne supercomputing system enabled us to evaluate the model's performance on multiple platforms.

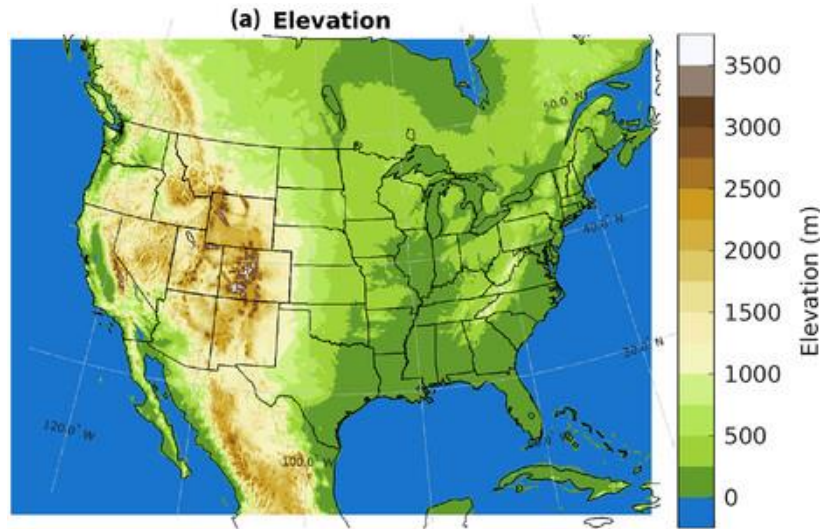


Fig. 4.I: Domain of the 4-km CONUS404 simulation with land surface elevation height (m). Derived by Rassussen M. L. et al., 2023.

In this context, our experiment not only demonstrates the effectiveness of the developed cGAN within the specific setting of the United States, but also emphasizes its potential for application across additional geographical domains and datasets. This flexibility represents a significant advancement in downscaling technologies, as models that can adapt to different input variables and geographical contexts greatly enhance their applicability, making them essential tools for assessing climate risks and informing planning decisions.

Data	longitude and latitude	horizontal resolution	n° grid points
ERA5	(Lon = : 55.50 W – 138.90°; Lat = 18.3° N–57.50° N)	0.28° (≈ 31 km)	157x333
WRF	(Lon = : 56.95 W - 139.05°; Lat = 18.1° N–57.91° N)	0.02° (≈ 4 km)	1359x1015

Tab. 4.I: Domain details for LW and HR dataset: geographical boundaries, horizontal resolution, number of grid points.

4.2 Datasets and *Pre-processing*

The datasets employed for the application of our cGAN for the U.S. case study include ERA5 for the low-resolution data at 31 km, and WRF at 4 km for the high-resolution counterpart. The specifics of the domain coverage and the number of grid points for each dataset are summarized in Table 4.I. As the ERA5 dataset has already been introduced in Chapter 2, the following section will focus on the description of the high-resolution WRF dataset.

4.2.1 High-Resolution Dataset: WRF-NCAR (Weather Research and Forecasting)

The dataset utilized for high resolution in our cGAN model derives from the 4 km simulations of the Weather Research and Forecasting (WRF) model, which has produced a new dataset known as CONUS404. This dataset represents a significant advancement in the availability of high-resolution hydroclimatic data for the contiguous United States, covering a period of over 40 years, from October 1979 to September 2021, with a spatial resolution of 4 km. Generated through the WRF model, the dataset results from a dynamic downscaling process applied to global atmospheric reanalysis data from ERA5 of the European Centre for Medium-Range Weather Forecasts (ECMWF). The configuration of the WRF model used includes 51 vertical levels and employs several physical parameterization schemes, such as the Morrison Microphysics Scheme for precipitation simulation, the Mellor-Yamada-Janjic (MYJ) Planetary Boundary Layer Scheme for modeling atmospheric stability, and the RRTMG Radiation Scheme for solar and terrestrial radiation. These schemes were selected to enhance the representation of atmospheric processes, especially in complex orographic contexts. The simulations were executed on the Cheyenne supercomputer at the National Center for Atmospheric Research (NCAR), utilizing modern architectures with Graphics Processing Units (GPUs) to accelerate numerical calculations, thereby enabling more detailed and long-term simulations. The results discussed in Rasmussen R. M. et al. (2023) demonstrate that the dataset effectively captures the dynamics of temperature and precipitation, highlighting an annual thermal bias ranging from -1 °C to +1 °C, with a tendency to underestimate winter temperatures, particularly in northern areas. Furthermore, the dataset exhibits a remarkable capacity to represent extreme precipitation events, allowing for the distinction of spatial variability at a mesoscale. However, the authors note limitations such as the underestimation of snowpack and a positive bias in surface solar radiation, indicating the need for further optimizations. Despite these challenges, CONUS404 emerges as a valuable resource for the availability of high-

resolution atmospheric fields over an extended temporal range and a vast domain. The CONUS404 data are archived and publicly accessible at (<https://doi.org/10.5066/P9PHPK4F>) (Rasmussen R. M. et al. 2023), promoting the sharing of scientific data for research and innovation in the field.

4.2.2 Assessing the Portability of the Developed cGAN Downscaling Model on Different Supercomputer Architectures

The portability of code is a fundamental aspect of scientific programming, referring to the ability of an application to run on different hardware and software platforms without requiring substantial modifications. This concept is crucial for ensuring the replicability of scientific results and facilitating the sharing of discoveries among researchers utilizing diverse infrastructures. Portability also enables researchers to leverage the specific capabilities of various supercomputers, optimizing computational efficiency and reducing execution times. In the context of our study focused on downscaling for the United States, we utilized both the Casper and Cheyenne supercomputers (<https://ncar-hpc-docs-arc-iframe.readthedocs.io/compute-systems/casper/>), both provided by NCAR. While Cheyenne featured a node-based architecture with Intel Xeon processors and 36 cores per node, making it highly capable of handling complex and long-term simulations, our work primarily relied on Casper. This choice was driven by the fact that Cheyenne was being phased out and was eventually decommissioned in 2024. Cheyenne system was designed to support a wide range of workloads, from climate simulations to artificial intelligence applications. In contrast, Casper is particularly suited for operations that require the use of Graphics Processing Units (GPUs), which can significantly accelerate the model training processes, especially for deep learning algorithms and other highly parallelizable applications. The GPU architecture of Casper allows for more efficient execution of intensive calculations compared to traditional processors. The infrastructure specifications of the Casper supercomputer are detailed below (Tab. 4.2).

Category	Nodes	Memory	Processors	Storage	Networking	GPUs
Data Analysis & Visualization	22 Supermicro 7049GP-TRT SuperWorkstation	Up to 384 GB DDR4-2666 per node	2 x 18-core 2.3 GHz Intel Xeon Gold 6140 per node	2 TB NVMe SSD local	1 Mellanox ConnectX-4 100Gb Ethernet, 1 Mellanox ConnectX-6 HDR100 InfiniBand	9 nodes: 16GB NVIDIA Quadro GP100 GPU >3 nodes: 40GB NVIDIA Ampere A100 GPU
GPU Visualization	6 Supermicro L40 GPU Visualization	768 GB DDR5 per node	1 x 48-core AMD EPYC per node	1.6 TB NVMe SSD local	1 Mellanox ConnectX-6 100Gb Ethernet VPI adapter	None

Machine Learning / Deep Learning	4 Supermicro SuperServer with 4 V100 GPUs	768 GB DDR4-2666 per node	2 x 18-core 2.6 GHz Intel Xeon Gold 6240 per node	2 TB NVMe SSD local	1 Mellanox ConnectX-4 100Gb Ethernet, 2 Mellanox ConnectX-6 HDR200 InfiniBand adapters	4 V100 32GB SXM2 GPUs with NVLink
	6 Supermicro SuperServer with 8 V100 GPUs	1152 GB DDR4-2666 per node	2 x 18-core 2.6 GHz Intel Xeon Gold 6240 per node	2 TB NVMe SSD local	1 Mellanox ConnectX-4 100Gb Ethernet, 2 Mellanox ConnectX-6 HDR200 InfiniBand adapters	8 V100 32GB SXM2 GPUs with NVLink
	8 Supermicro with 4 A100 GPUs	1024 GB memory per node	2 x 64-core 2.45 GHz AMD EPYC Milan 7763 per node	1.5 TB NVMe SSD local	4 Mellanox ConnectX-6 network adapters	4 A100 80GB SXM4 GPUs with NVLink
	2 Supermicro with 4 H100 GPUs	1024 GB DDR5 per node	2 x 32-core Intel Xeon Gold processors per node	1.5 TB NVMe SSD local	4 Mellanox ConnectX-6 network adapters	4 H100 80GB SXM4 GPUs with NVLink
High-Throughput Computing	62 small-memory workstation nodes	384 GB DDR4-2666 per node	2 x 18-core 2.6 GHz Intel Xeon Gold 6240 per node	1.6 TB NVMe SSD local	1 Mellanox ConnectX-5 100Gb Ethernet, 1 Mellanox ConnectX-6 HDR200 InfiniBand	None
	2 large-memory 36-core workstation nodes	1.5 TB DDR4-2666 per node	2 x 18-core 2.3 GHz Intel Xeon Gold 6240 per node	1.6 TB NVMe SSD local	1 Mellanox ConnectX-5 100Gb Ethernet, 1 Mellanox ConnectX-6 HDR200 InfiniBand	None
	6 large-memory 64-core workstation nodes	1.5 TB DDR5 per node	1 x 64-core AMD EPYC per node	1.6 TB NVMe SSD local	1 Mellanox ConnectX-6 100Gb Ethernet VPI adapter	None
Research Data Archive	4 Supermicro Workstation nodes	94 GB DDR4-2666 per node	2 x 16-core 2.3 GHz Intel Xeon Gold 5218 per node	1.92 TB SSD local	1 Mellanox ConnectX-6 VPI 100Gb Ethernet	None

Table 4.2: Infrastructure of the Casper Supercomputer (NCAR)

The portability of our code has involved the creation of conda environments tailored to each supercomputer, facilitating the management of libraries and dependencies essential for the application’s proper functioning. Furthermore, we carefully managed how the code is executed across the nodes, optimizing operational efficiency. It is noteworthy that previous experiments conducted in Italy had already anticipated the portability of the architecture developed on other supercomputers; as previously mentioned, these experiments were initially carried out on the Zeus supercomputer and later transitioned to Juno, both operated by CMCC, when the former was decommissioned. The integration of portability within our code allows for a seamless transition between Cheyenne and Casper, effectively leveraging the different hardware architectures to optimize performance and efficiency in our experiments. Such practices expand

the scope of simulations and provide greater flexibility in scientific research, fostering innovation and collaboration among diverse institutions.

4.3 Results and Analysis: Comparison with Dynamical Downscaling

In this section, we discuss the results obtained from the downscaling of ERA5 data using the developed Conditional Generative Adversarial Network (cGAN) approach, comparing them with those from the Weather Research and Forecasting (WRF) model configured at 4 km, achieved through dynamic downscaling. The data provided by NCAR, used as input for the cGAN model, spans the period from January 2001 to December 2014. Consistent with the methodology applied in the experiment conducted over Italy, 10 years (January 2001 to December 2011) were allocated for training, while the remaining two years (January 2012 to December 2014) were designated for testing. The focus of this experiment, centered on the United States, is on two fundamental aspects: the importance of setting an adequate number of neurons for the considered geographical domain and the impact of the number of training epochs on the obtained results. The decision to concentrate on these two key points for this experiment was driven by the desire to highlight other essential aspects of training a GAN, which become particularly important when working with large geographical domains.

The choice of the number of neurons in neural networks is crucial for representing the complexity of atmospheric dynamics. An insufficient number of neurons can lead to overly simplified modeling, resulting in inaccurate outcomes, while an excessive number may cause overfitting, where the network learns to memorize the training data rather than generalize to new data. Therefore, for the domain of the United States, it is essential to find a balance that allows the cGAN to effectively learn the local climatic peculiarities, especially considering orographic variations, which are particularly relevant in this region.

The second aspect analyzed concerns the impact of the number of training epochs. Previous researches suggest that the number of epochs significantly influences the network's ability to learn temporal dynamics. An excessive number of epochs can lead to a model that, while performing well on the training data, fails to generalize to the test data. Conversely, too few epochs may prevent the model from fully assimilating the information, resulting in suboptimal predictions. This experiment demonstrates that the accurate configuration of the number of neurons and the training epochs is of fundamental importance in the context of GAN-based downscaling, with direct effects on the quality of the generated atmospheric fields.

4.3.I Overall Architecture: Upgrade of the Number of Neurons

Similar to the configuration of the number of neurons employed in the experiments conducted in Italy, which was derived from an empirical approach, the selection of the architecture illustrated in Fig. 4.2 for the application in the United States is also grounded in empirical experimentation and theoretical considerations regarding the significance of an appropriate number of neurons. Notably, it was essential to increase the number of neurons while maintaining the same architectural framework as the Italian case study, owing to the substantially larger dimensions of the geographical domain in the United States. This adjustment is warranted, as an expansive domain encompasses a wider array of climatic and geographical features that must be effectively captured to ensure accurate modeling. The augmentation of the neuron count enables the model to learn and represent these complexities more, thereby enhancing its generalization capabilities across the diverse conditions inherent to the study area. Consequently, such optimization of the architecture is imperative for effectively addressing the intricacies of the data and for improving the model's super-resolution capabilities.

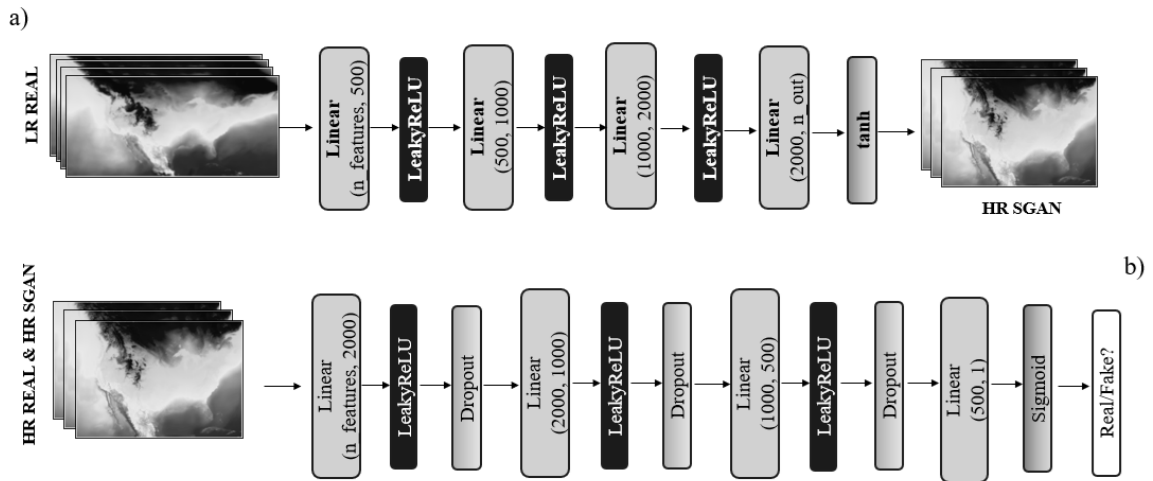


Fig. 4.2: Architectures of (a) generator and (b) discriminator. Size input generator 157×333 , size input discriminator 1359×1015 , the number of frames per sequence 100 (batch size = 100) for both data sets considered here. After training, the best generator is used for the test period. The architecture displayed in this figure shows an increased number of neurons in each layer compared to the configuration presented in Fig. 2.6.

The analysis of the results for the testing period begins by comparing the configuration employed for the cGAN application over Italy, referred to as ERA5-DownGAN_Italy for clarity, with the revised setup for this new domain, referred to as ERA5-DownGAN_U.S.,

where the number of neurons was increased. The temporal evolution of the dataset generated by the cGAN, in comparison with the low-resolution ERA5 dataset and the high-resolution real-world counterpart from WRF, as shown in Fig. 4.3, reveals clear distinctions between the two configurations.

The temporal evolution of the results comparing the two configurations, ERA5-DownGAN_Italy and ERA5-DownGAN_U.S., reveals interesting differences between them. Notably, with the increase in the number of neurons (Fig. 4.3b), the model demonstrated an improved ability to capture the minima, which are now represented by lower temperatures compared to the previous configuration (Fig. 4.3a). However, this adjustment also resulted in a shift towards higher values for the maxima, reflecting a certain distortion in the model's behavior. Nevertheless, it is important to emphasize that both configurations, despite the positive or negative shift, show a good temporal agreement with the real high-resolution dataset (WRF). However, this trend is obtained by averaging over the entire geographic area of interest, making it susceptible to compensation effects.

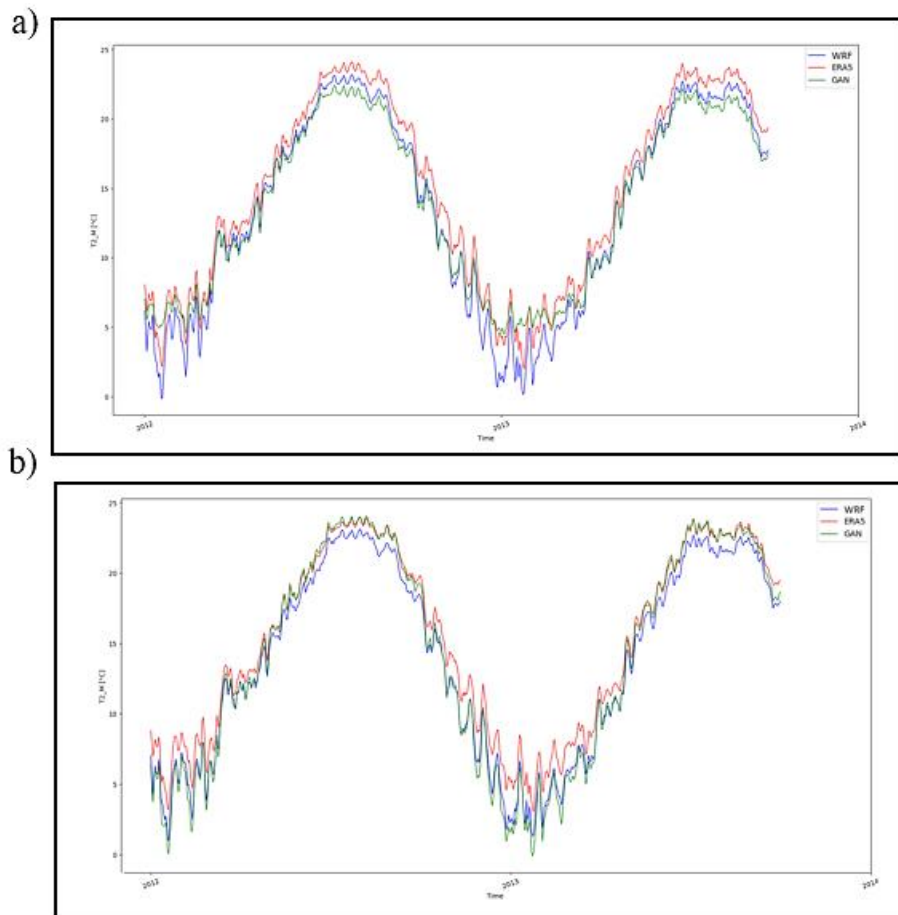


Fig. 4.3: Temporal evolution of the daily spatial average of 2m-temperature during the test period. The real low-resolution data (ERA5) is shown in red, the real high-resolution data (WRF) in blue, and the downscaled

artificial data generated by the GAN (ERA5-DownGAN) in green. Panel (a) compares ERA5-DownGAN_Italy, while panel (b) compares ERA5-DownGAN_U.S.

Consequently, based on this preliminary exploratory analysis, we opted to compare the spatial correlation maps (Fig. 4.5) derived from averaging over the entire testing period. This comparison aims to highlight the differences in reconstruction between the two employed configurations, which are also evident in Fig. 4.4.

The ERA5-DownGAN_U.S. configuration demonstrates a significantly enhanced capacity for reconstructing the 2-meter temperature field compared to the ERA5-DownGAN_Italy configuration (Fig. 4.4). Specifically, the latter is characterized by artifacts and a sparse representation of the temperature field, which can be attributed to the insufficient number of neurons in the model architecture tailored for Italy. This limitation restricts the model's ability to capture the underlying spatial patterns and variations within the dataset, thereby underscoring the importance of employing the ERA5-DownGAN_U.S. configuration with an increased number of neurons for the domain under investigation.

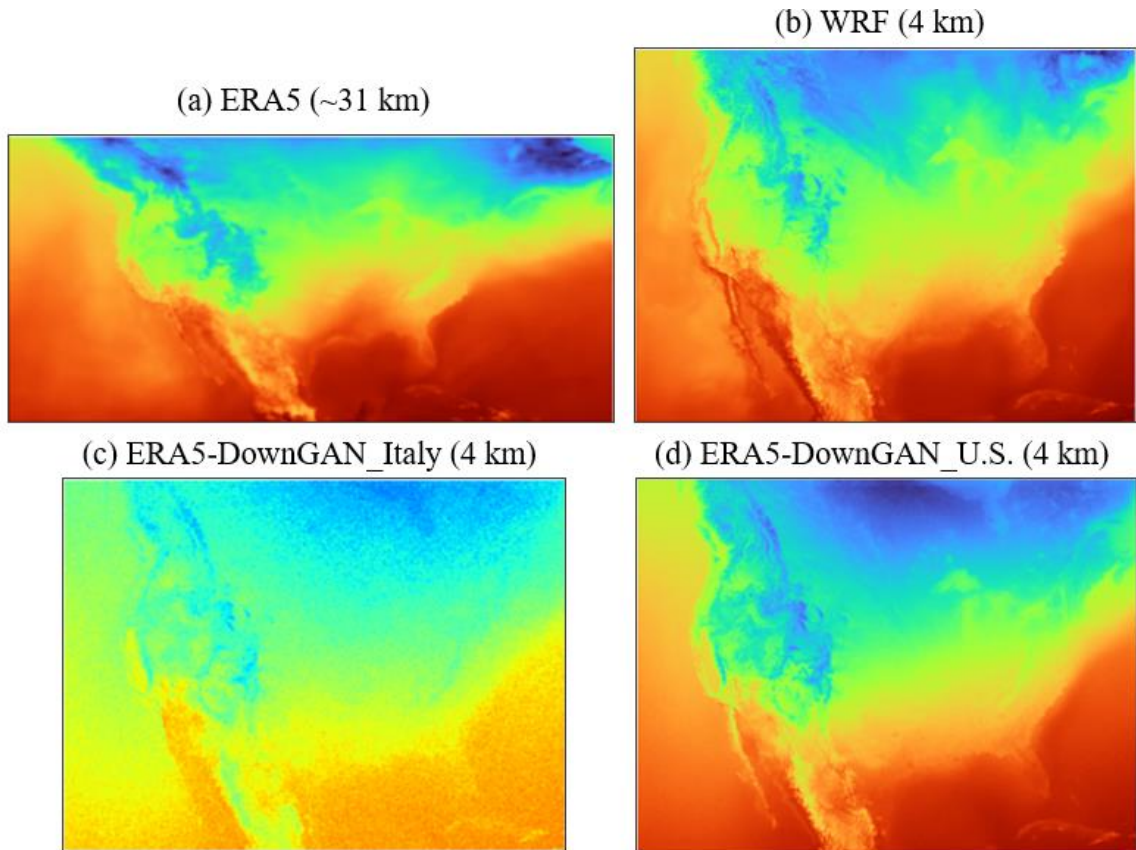


Fig. 4.4: Examples of one day from the test data set. (a)–(d) Real low-resolution data (ERA5), the real high-resolution data (WRF), the downscaled artificial data ERA5-DownGAN_Italy, and the downscaled artificial data ERA5-DownGAN_U.S.

The significant improvement in the results obtained from the ERA5-DownGAN_U.S. configuration compared to those from the ERA5-DownGAN_Italy configuration is also reflected in the correlation map (Fig. 4.5) illustrating the relationship between the real high-resolution WRF dataset and the high-resolution outputs generated by the two examined GAN configurations. The optimized configuration for the U.S. consistently exhibits high average correlation values, ranging from 0.85 to 1, with the lowest values observed in certain oceanic regions, dropping to 0.7. In contrast, the correlation map derived from the comparison between the WRF dataset and the results from the configuration optimized for Italy shows significantly lower values, ranging from 0.7 across the entire oceanic areas to 0.85. Furthermore, this map highlights a sparse distribution similar to that observed in the 2-meter temperature field maps, underscoring the limitations of the Italian configuration.

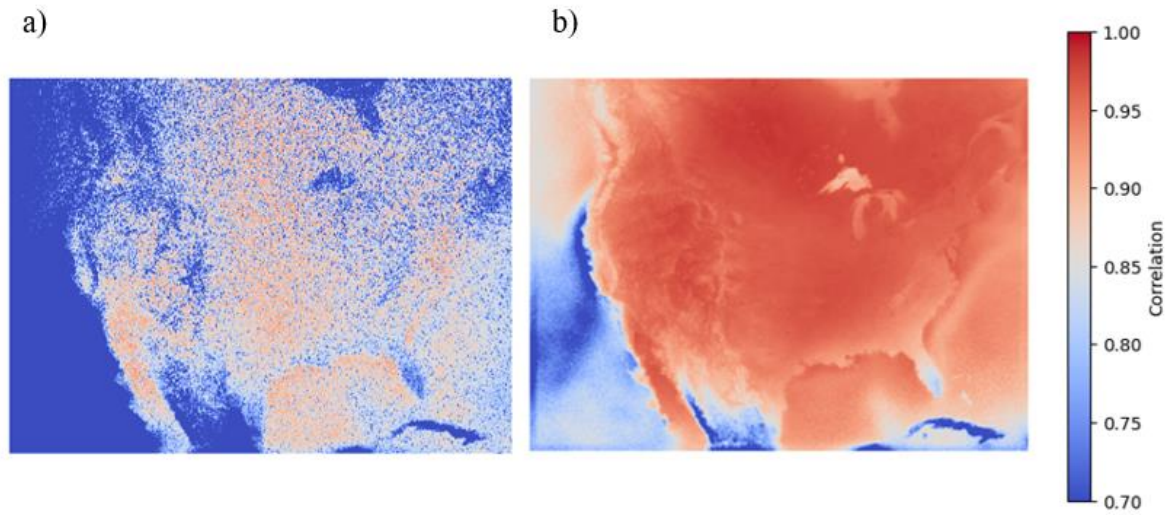


Fig. 4.5: Correlation maps showing the relationship between the WRF dataset and the downscaled artificial data from the GAN configurations. (a) ERA5-DownGAN_Italy; (b) ERA5-DownGAN_U.S.

Figure 4.6 provides insight into the temporal alignment between the two datasets (real and generated by the GAN). The scatter plots were obtained by considering the spatial mean value for each time step, both for the high-resolution real dataset (x-axis) and for the values generated by the GAN in the two configurations (y-axis). In comparing the two configurations of the developed cGAN, represented by the scatter plots in panels a and b, notable differences emerge in the model's ability to reproduce the high-resolution real dataset values. Ideally, if the values generated by the GAN correspond exactly to the real ones, the points should align along the 45° line, indicating a perfect match between real and generated values.

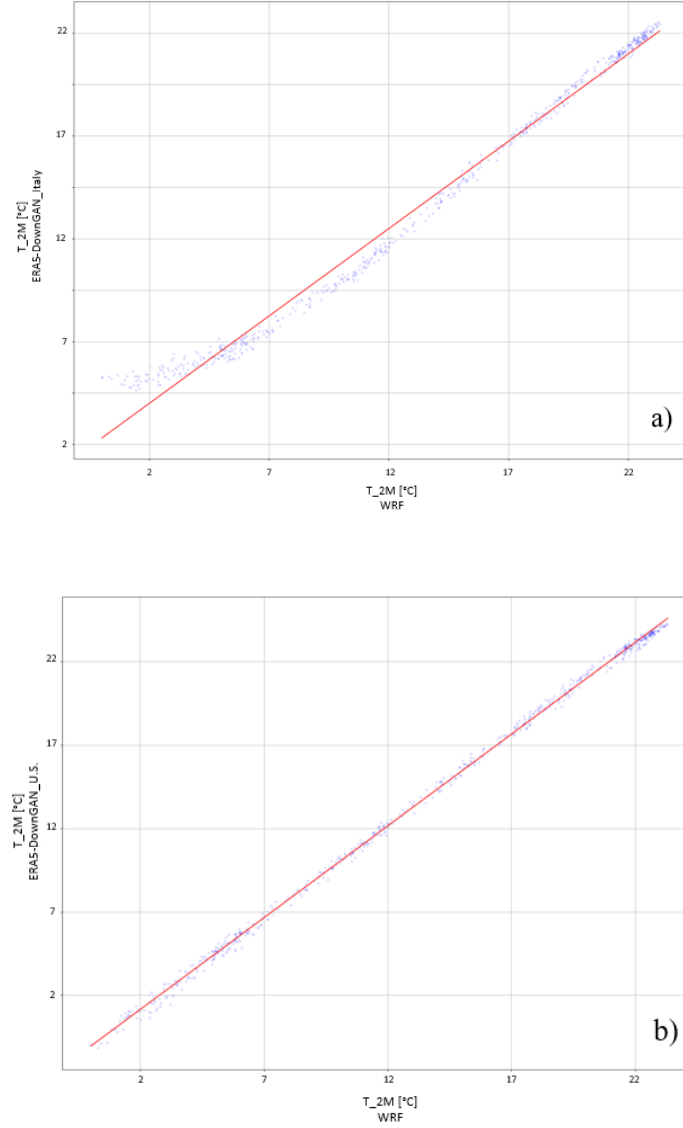


Fig. 4.6: Scatter diagram for evaluating the correlation between WRF and the downscaled artificial data from the GAN configurations. (a) ERA5-DownGAN_IItaly; (b) ERA5-DownGAN_US.

In panel a), representing the ERA5-DownGAN_IItaly configuration, the points deviate from the 45° line: for lower and higher spatial mean values, the points are positioned above the line, while for intermediate values, they fall below it. This pattern suggests that the GAN in this configuration tends to overestimate the generated values at the extremes of the dataset (low and high real values) and underestimate them for intermediate values. These discrepancies indicate that the model struggles to accurately capture the dynamics present in the real dataset, particularly in the extreme ranges, likely due to its limited capacity to represent complex patterns.

Conversely, in panel b), corresponding to the ERA5-DownGAN_U.S. configuration, which uses a larger number of neurons, the points are more closely aligned with the 45° line, with a slight overestimation for higher values. This significant improvement suggests that the GAN in this configuration is more accurate in reproducing the spatial mean values of the real dataset, even in situations where more “extreme” values are present. The increased number of neurons likely allowed the model to learn the complex relationships between the data better, enhancing the agreement with the high-resolution dataset, particularly in the extreme ranges, and reducing the overestimation and underestimation errors observed in the previous configuration.

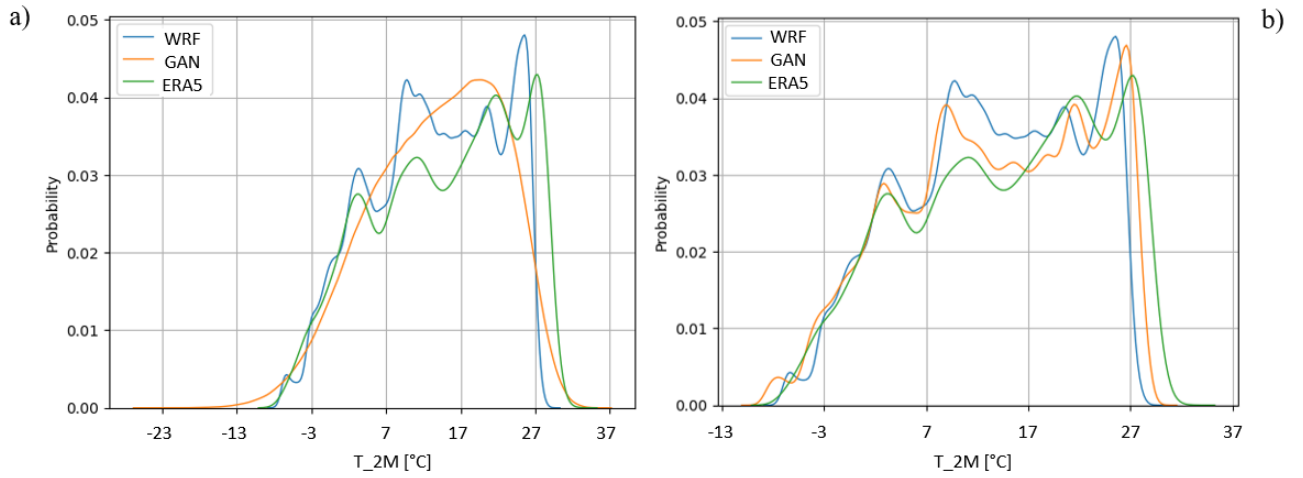


Fig. 4.7: Comparison of the probability density functions (PDF) of the 2m-temperature distribution for the ERA5 (green), WRF (blue), and the downscaled artificial data from the GAN configurations (orange). (a) ERA5-DownGAN_Italy; (b) ERA5-DownGAN_U.S.

The analysis of the Probability Density Functions (PDFs) for the two configurations of the GAN, illustrated in Figure 4.7 with panels a and b, reveals notable differences in their ability to replicate the high-resolution temperature distribution of the WRF dataset. In panel a, representing the ERA5-DownGAN_Italy configuration, the PDF exhibits a flattened shape that approximates a Gaussian distribution, indicating an inability to capture the peaks present in the high-resolution WRF dataset. This configuration shows lower values, extending down to -25°C, whereas both the high-resolution and low-resolution real datasets are limited to a minimum of -12°C. In contrast, the second configuration depicted in panel b, ERA5-DownGAN_U.S., which utilizes a greater number of neurons, successfully captures not only the range of the WRF dataset (-12°C to 30°C) but also all the peaks within that range, resulting in a shape that closely resembles the original distribution. However, it is important to note that this configuration displays a slight rightward shift of approximately 1°C in the results produced by the GAN for

values above 22°C. This shift aligns with previous observations made in the scatter plots, where spatially averaged values were considered, while the current analysis of the PDFs takes into account the average over time across all points in space.

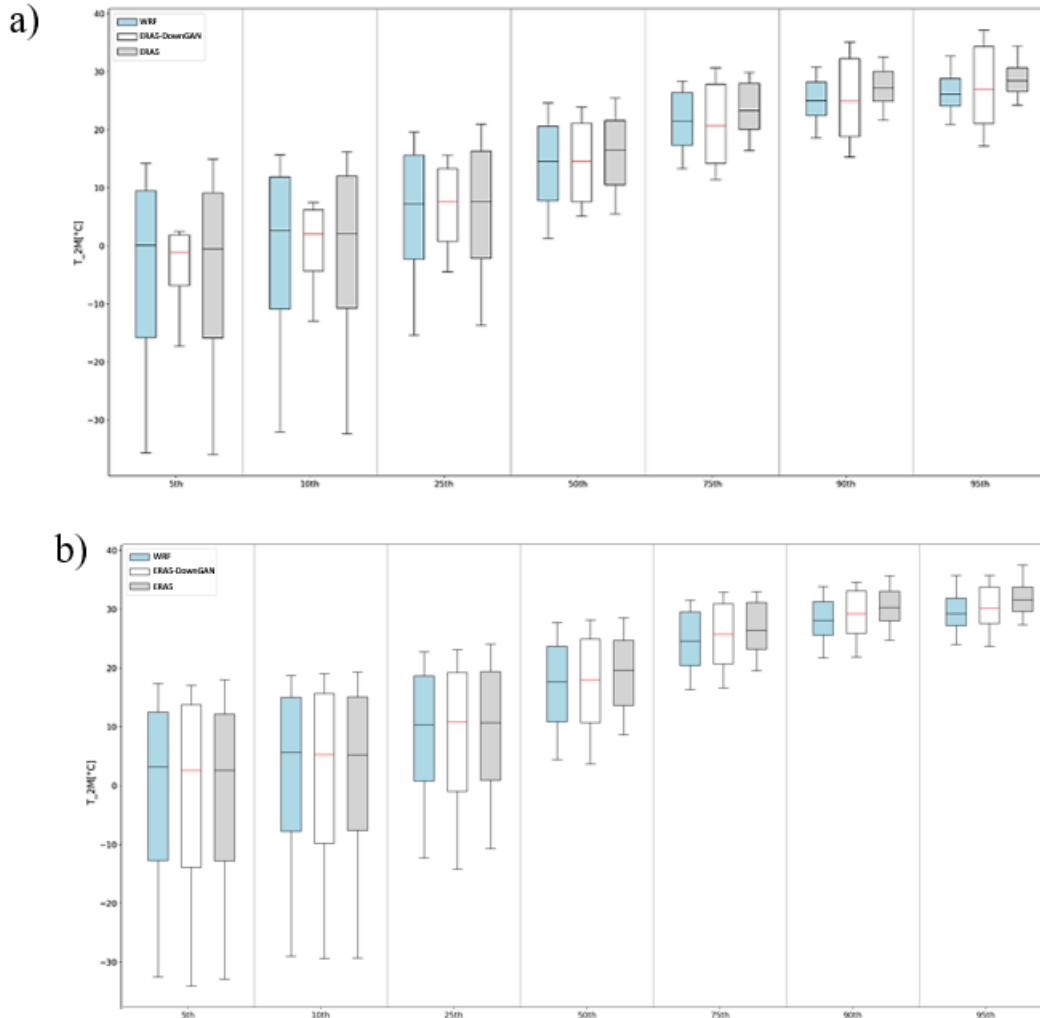


Fig. 4.8: Comparison of the 2m-temperature (T_{2M}) distribution for real low-resolution data (ERA5) with the gray rectangle, the real high-resolution data (WRF) with the blue rectangle, and the downscaled artificial data from the GAN configurations (white). (a) ERA5-DownGAN_Italy; (b) ERA5-DownGAN_U.S.

Previous analyses primarily concentrated on spatial or temporal averages, which can mask underlying variability within the datasets. In contrast, Figure 4.8 presents a more comprehensive assessment by considering all available values and evaluating the percentiles from the 1st to the 95th. This approach enables a deeper understanding of the distributional characteristics of the temperature data, facilitating the identification of subtle variations and extreme values that may be critical for accurately capturing the dynamics of the underlying processes. By analyzing the complete dataset in this manner, we aim to enhance the robustness

of our evaluation, providing insights that are essential for assessing the performance of the GAN configurations tested. Fig. 4.8, is organized into two panels (a and b), which illustrate the performance of the two configurations of the Generative Adversarial Network (GAN) tested: ERA5-DownGAN_Italy and ERA5-DownGAN_U.S. In panel a, the dataset generated by the GAN for the low percentiles, ranging from the 1st to the 25th, shows a median aligned with that of the real datasets (WRF and ERA5). However, the interquartile range (IQR) of the boxplots is significantly reduced, with the upper whisker shifted towards notably lower values (approximately 2 °C) compared to WRF (approximately 15 °C). For the central percentile (50), there is good agreement in terms of both median and variability, although the upper whisker is shifted towards lower values by about 1 °C, while the lower whisker extends to higher values by approximately 3 °C compared to WRF. For the higher percentiles, from the 79th to the 95th, the situation reverses, revealing reduced variability, with upper whiskers directed towards lower values and lower whiskers extending to higher values. This behavior suggests that the ERA5-DownGAN_Italy configuration may lack an adequate number of neurons to capture the complexities of the domain, resulting in an underestimation of extreme values in the lower percentiles and a poor ability to represent distribution dynamics in the higher percentiles. In contrast, panel b demonstrates excellent agreement in terms of median, variability, and whisker length. Generally, for all percentiles, there is a slight tendency towards greater variability, attributed to an increase in lower values for the lower percentiles and higher values for the upper percentiles. This improvement is attributable to the increased number of neurons in the ERA5-DownGAN_U.S. configuration, which has enabled the model to more effectively learn the complex relationships present in the dataset, leading to a more accurate reproduction of thermal distributions.

Finally, when comparing the high-resolution real dataset (WRF) with the low-resolution dataset (ERA5), there is good agreement in terms of median, variability, and whisker length for the low percentiles up to the 25th. However, from the 50th percentile onward, an increase in variability and higher upper whiskers is observed, with values greater than those in WRF, while the lower whisker shows lower values. This behavior is closely related to the difference in resolution: ERA5, being a coarser dataset, tends to average values over a less fine grid, resulting in a loss of detail and an increase in discrepancies in extreme values.

This comprehensive evaluation highlights the enhanced performance of the ERA5-DownGAN_U.S. configuration in accurately reproducing the temperature distribution characteristics found in the real datasets, further reconfirming the importance of utilizing an

appropriate number of neurons tailored to the number of points that define the domain under examination.

4.3.2 The Role of Epochs in Model Training

The number of epochs constitutes a critical parameter in the training regimen of Generative Adversarial Networks (GANs), as it substantially influences the quality and stability of the generated outputs. In the present case study focusing on downscaling within the United States, the number of epochs employed for comparative analyses was established at 15. This decision was predicated upon the observation that an increase in the number of neurons within the network correlates with heightened model complexity. Although such complexity can potentially enhance the capacity of the conditional GAN (cGAN) to capture intricate features of the domain, it concurrently necessitates a restart of the training process for a domain as expansive and multifaceted as the one under consideration. This aspect has been addressed and integrated into our architectural framework in subsequent phases, yet it remains beyond the scope of this discussion.

It is noteworthy that the GAN exhibited preliminary signs of convergence (assessed through the evaluation of generator and discriminator losses, denoted as d_loss and g_loss , respectively) at the 15-epoch mark, indicating that the model was effectively assimilating the characteristics of the dataset. Thus, the selection of this specific epoch count was justified. Nevertheless, the importance of epoch selection cannot be underestimated, as it plays a pivotal role in balancing the trade-off between underfitting and overfitting. Insufficient epochs may result in a model that fails to adequately capture the underlying data structures, while excessive epochs may lead to overfitting, wherein the network memorizes the training data rather than generalizing effectively to unseen data.

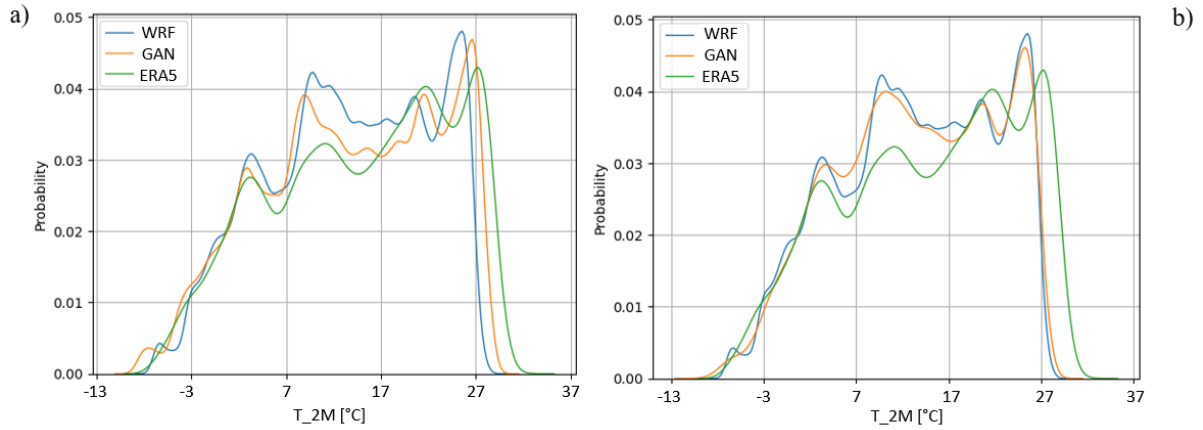


Fig. 4.9: Comparison of the 2m-temperature (T_{2M}) distribution for real low-resolution data (ERA5) with the gray rectangle, the real high-resolution data (WRF) with the blue rectangle, and the downscaled artificial data from the GAN configurations (white). (a) ERA5-DownGAN_Italy; (b) ERA5-DownGAN_U.S.

To illustrate this point, Fig. 4.9 provides a comparative analysis of the probability distributions (PDFs) generated with 10 and 15 epochs. In particular, in panel (a), which represents the results obtained with 15 epochs, the GAN clearly demonstrates its ability to capture all the peaks present in the real WRF dataset. This indicates that the model is effectively learning the underlying data distribution and can represent the complexities inherent in the dataset. Conversely, in panel (b), which corresponds to a lower epoch count, the peaks are notably smoothed out, although the overall shift remains smaller. This suggests that while the model is able to approximate the general trend of the data, it struggles to capture the finer details and variations present in the real dataset. The diminished representation of the peaks indicates that a lower number of epochs may lead to an oversimplification of the data, compromising the GAN's ability to generate outputs that truly reflect the complexities of the underlying distribution. Such findings underscore the importance of selecting an appropriate epoch count during training to ensure the model can adequately learn and represent the intricacies of the target domain.

4.4 Summary Evaluation

In conclusion, the developed Conditional Generative Adversarial Network (cGAN) has demonstrated a remarkable capacity for adaptation to downscaling across various geographic domains, including those characterized by larger spatial extents. This flexibility is a key strength of architecture, as it allows for replication and application in diverse contexts, thereby significantly enhancing the model's potential for broader climate data applications. The comparative analysis of the ERA5-DownGAN_Italy and ERA5-DownGAN_U.S.

configurations has elucidated the considerable impact of architectural choices, particularly the number of neurons, on the model's performance in downscaling tasks.

Specifically, the results indicate that the increase in the number of neurons within the U.S. configuration led to substantial improvements in the cGAN's ability to accurately reconstruct the 2-meter temperature field. This advancement is reflected in the model's enhanced capacity to capture intricate spatial patterns and mitigate artifacts that were prevalent in the Italy-optimized configuration. Such improvements are further evidenced by the elevated correlation values between the GAN-generated fields and the high-resolution WRF dataset, as illustrated in the correlation maps and scatter plots. The strengthened relationship between the generated outputs and the high-resolution data underscores the effectiveness of the cGAN in producing realistic and coherent representations of temperature distributions.

Moreover, the analysis of the Probability Density Functions (PDFs) highlights a critical disparity between the two configurations. The ERA5-DownGAN_Italy configuration struggled to accurately represent the temperature value distribution, resulting in a flattened PDF that failed to capture the key peaks observed in the high-resolution WRF dataset. In contrast, the ERA5-DownGAN_U.S. configuration successfully mirrored the distribution of the WRF dataset, encompassing the full range of temperature values and accurately depicting the relevant peaks. While this configuration displayed a slight rightward shift in the results, which could potentially be minimized with extended training, the overall alignment with the high-resolution dataset indicates a significant advancement in the model's generative capabilities. The importance of epoch selection has been underscored throughout this study, with findings demonstrating that an insufficient number of training epochs can lead to oversimplified representations of the underlying data. The notable discrepancies between outputs generated with 10 and 15 epochs further emphasize the necessity for thorough training to ensure that the cGAN can adequately capture the full complexity of the dataset. Insufficient epochs may result in a model that approximates the general trend of the data without adequately reflecting the finer details and variations inherent in the real dataset.

Although a direct comparison in terms of training hours between the Italian and U.S. domains is not provided, due to the differing architectures used, driven by the dataset storage systems on their respective supercomputers, some preliminary observations can be made regarding the scalability challenges when transitioning from smaller to significantly larger geographical domains. For instance, training over 100 epochs takes about 4 hours for the Italian domain,

whereas the U.S. domain requires approximately 25 hours for 15 epochs. In contrast, the testing phase incurs considerably lower costs, with the Italian domain requiring only 2 minutes and the U.S. domain less than 7 minutes. While the computational costs associated with the architecture developed in this study are highly competitive within the landscape of ML techniques for downscaling and in comparison to dynamic downscaling, it is important to note that porting the architecture to GPUs could further enhance the competitiveness of computational costs, enabling downscaling of even larger geographical domains.

Overall, this investigation underscores the critical interplay between network architecture and training parameters in the effective application of GANs for climate data downscaling. With the increasing demand for high-resolution climate data, the findings of this study highlight the potential of the developed conditional Generative Adversarial Network (cGAN) as a powerful and flexible tool for advancing the understanding of climate dynamics at high and very high spatial resolutions across various geographical contexts.

5 | Discussion

The integration of artificial intelligence (AI) into climate and meteorological sciences has rapidly established a prominent role, demonstrating significant potential for solving complex challenges and enhancing predictive capabilities. The exponential growth in the adoption of advanced machine learning (ML) models and deep learning (DL) architectures has led to the development of specific platforms and collaborative programs, such as AI2ES (AI for Environmental Sciences), supported by NOAA for analyzing extreme environmental phenomena; ML4ESS (Machine Learning for Earth System Science), developed by NCAR to integrate ML into Earth system modeling; and ECMWF's AI4Weather, which aims to refine long-term climate and weather forecasting. A pioneering result of these initiatives is ECMWF's use of AI-based global forecasting models, such as Google DeepMind's GraphCast, which accelerates computational processes and enables deeper investigation of unresolved questions by identifying latent relationships in atmospheric data. These advancements signal a profound shift, with AI emerging as a crucial tool not only for improving forecast efficiency but also for advancing the understanding of complex climate phenomena, thereby transforming the approach to climate science.

A critical challenge in climate science is downscaling, given the high computational costs associated with traditional dynamic downscaling based on physical models. The growing adoption of ML/DL models and architectures by major international research centers, such as ECMWF, NCAR, and NOAA, has motivated the exploration and development, in this work, of an innovative downscaling model fully based on DL architecture, specifically a conditional generative adversarial network (cGAN) trained on physically grounded datasets. The objective of this architecture is to maintain spatial coherence during the downscaling process, transitioning from the original low-resolution field (~ 31 km) to the new high-resolution dataset, ERA5-DownGAN (~ 2.2 km), comparable to high-resolution reanalysis (VHR_REA-IT) achieved through dynamic downscaling based on the COSMO-CLM model. The main goal of

this study was to demonstrate the comparability between these two high-resolution datasets by evaluating not only the cGAN's capacity to reconstruct spatio-temporal patterns but also its fidelity in replicating the statistics of the original field and its accuracy in capturing extreme values, an area where many ML/DL models typically face performance limitations. To further test the robustness and flexibility of the proposed architecture, experiments were conducted on a different geographic domain (centered on the United States) and on complex climate variables, such as precipitation, characterized by highly skewed value distributions. Moreover, to highlight the model's adaptability, optimizations in individual case studies were achieved by maintaining the base architecture fixed and adjusting only a few configurable parameters, such as the weights associated with the loss function, the normalization scheme, and the number of neurons. Finally, an analysis was conducted on the behavior of the cGAN in the presence of extreme weather events, such as a significant heatwave, to evaluate its robustness in anomalous climatic conditions relative to climatological averages.

5.1 General Conclusions

The AI-assisted downscaling model, developed in this study, demonstrates versatility and ease of application across diverse climatological contexts and complex atmospheric variables. Specifically, the dataset produced by ERA5-DownGAN (~2.2 km), downscaled from ERA5 (~31 km), has shown the capability to replicate patterns and median values that are nearly identical to those simulated by the dataset (VHR_REA-IT) produced with a dynamical model for 2m-temperature, with a slight tendency to produce a cooler temperature field compared to its dynamical counterpart.

Even in anomalous situations, ERA5-DownGAN faithfully replicated the patterns of its dynamic counterpart; for example, during an intense heatwave, it exhibited a tendency to either amplify or slightly simplify the anomalous conditions in terms of value ranges. This simplification of the meteorological conditions was particularly evident in the Po Valley, where the model assimilated the bias inherent in the convection-permitting dynamical model that produced VHR_REA-IT, which overestimates the mean temperature field.

A thorough analysis of extremes over the entire period, averaged over the spatial domain, demonstrates that the ERA5-DownGAN can capture even the most extreme percentiles (from the 95th to the 99th), reflecting a distribution that is approximately identical in terms of median and dispersion compared to that produced by the dynamic model.

For the total cumulative precipitation field, the model successfully replicated the observed patterns in the dataset produced at the same resolution by its dynamical counterpart, exhibiting some anomalies only at the extreme percentiles (98th and 99th). These results were obtained without significant changes in the architecture, aside from appropriate settings for certain hyperparameters and the normalization used.

The application of the developed model to the U.S. domain for downscaling 2m-temperature also yielded extremely encouraging results, with no architectural alterations other than an increase in the number of neurons to accommodate a domain, that is n-times larger than that used in previous experiments focused on the Italian peninsula.

Based on the promising results, the AI-assisted downscaling approach developed proves to be a viable option for achieving high-performance downscaling, yielding results comparable to those of its dynamic counterpart while simultaneously reducing the computational resources involved in the process. An important innovation of this study lies in the use of a conditional Generative Adversarial Network (cGAN), with the stipulation that both the generative and discriminative models utilize real datasets during the training phase, for both high-resolution and low-resolution counterparts. This approach conditions the generation process, enabling the establishment of empirical connections between the two datasets while maintaining correlation between the two fields.

5.2 Future Developments

Despite the promising results achieved, the complexity of the proposed architecture opens the door to multiple opportunities for improvement. These enhancements can be implemented both on the datasets utilized, such as through the adoption of cross-validation training windows, and on the architecture itself. Among the potential future directions outlined by this study is the exploration of the model on a broader set of atmospheric variables, including wind fields and air humidity fields. This would enable the generation of a high-resolution dataset containing the primary atmospheric variables investigated, based on the proposed statistical downscaling using cGAN. Moreover, it is crucial to recognize the significance of seasonal factors and temporal variations arising from the selection of distinct training and testing phases. A promising approach could be the implementation of a spatiotemporal cross-validation framework that carefully considers the training and testing periods. This strategy could significantly enhance the generative model's ability to address seasonal fluctuations and temporal disparities.

Another improvement could arise from the adoption of convolutional layers in place of the current linear layers. While linear layers have been employed to provide a solid architectural foundation and greater control over the parameters, the use of convolutional layers could offer significant advantages, especially concerning complex variables such as precipitation. Convolutional architectures allow for faster training and improved model generalization, thereby mitigating the risk of overfitting. Furthermore, the translational invariance introduced by convolutional layers enables the model to recognize patterns irrespective of their precise position within the data matrix, which is a crucial aspect in meteorological forecasting.

Finally, to further enhance the performance of the developed downscaling model, it is advisable to consider integrating advanced strategies inspired by two GAN frameworks: the Progressive Growing Generative Adversarial Network (ProGAN) and the Latent Adversarial Generative (LAG). ProGAN is recognized for its ability to progressively generate high-resolution images, optimizing image quality as the training process advances (Li G. and Cao G., 2024). By implementing a progressive growth strategy within the layers of our cGAN, we can improve the quality of the generated images, focusing on finer details and superior spatial coherence.

On the other hand, the integration of LAG proves essential for modeling uncertainty in the data, allowing for the generation of a greater number of plausible samples rather than deterministic predictions (Li G. and Cao G., 2024). This approach effectively represents the intrinsic variability within climatic datasets. By incorporating LAG into our cGAN, we anticipate producing outputs that are not only realistic but also representative of data variability, thereby further enhancing the quality and reliability of the generated high-resolution dataset. These developments could not only expand the capabilities of the developed downscaling model but also contribute to a better understanding of atmospheric dynamics through the provision of high-resolution datasets across multiple geographical domains in significantly shorter timescales compared to dynamic downscaling, while ensuring consistency with similarly resolved datasets produced by physically based models.

A | 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 introduced in Section 2.2. 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°.

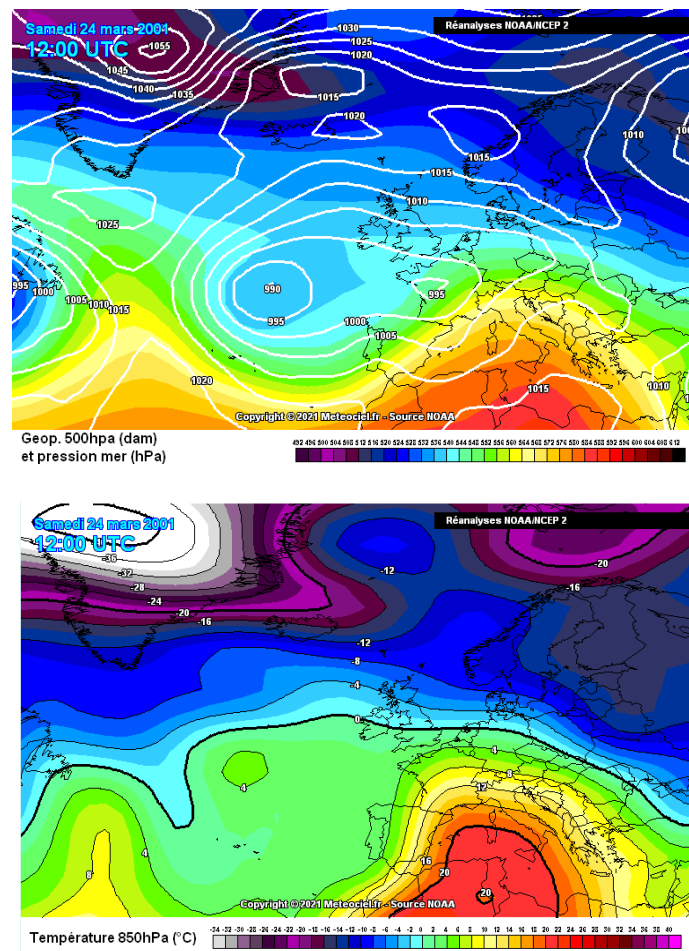


Fig. A.I: Synoptic analysis at 12:00 UTC from NOAA/NCEP reanalysis on 24th March 2001. Maps of geopotential at 500 hPa (upper) and temperature (°C) at 850 hPa (lower). Reproduced from (www.meteociel.it).

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. A.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 Chapter 3, 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. A.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. A.2, Panel b). This finding confirms the effectiveness of the developed cGAN model in the downscaling application for the temperature field, as discussed in Chapter 3.

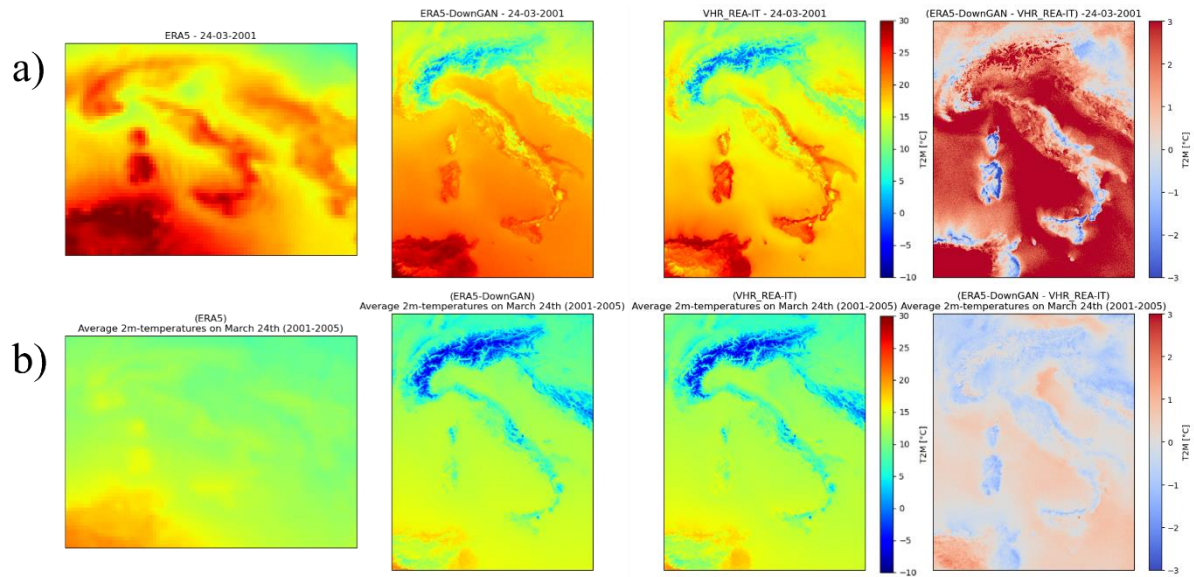


Fig. A.2.: Panel (a): Maps for March 24, 2021, for ERA5 (LR), ERA5-DownGAN (HR), and VHR_REA-IT (HR), along with the difference between ERA5-DownGAN and VHR_REA-IT. Panel (b): Average of all March 24 observations from 2001 to 2005 for ERA5 (LR), ERA5-DownGAN (HR), and VHR_REA-IT (HR), as well as the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

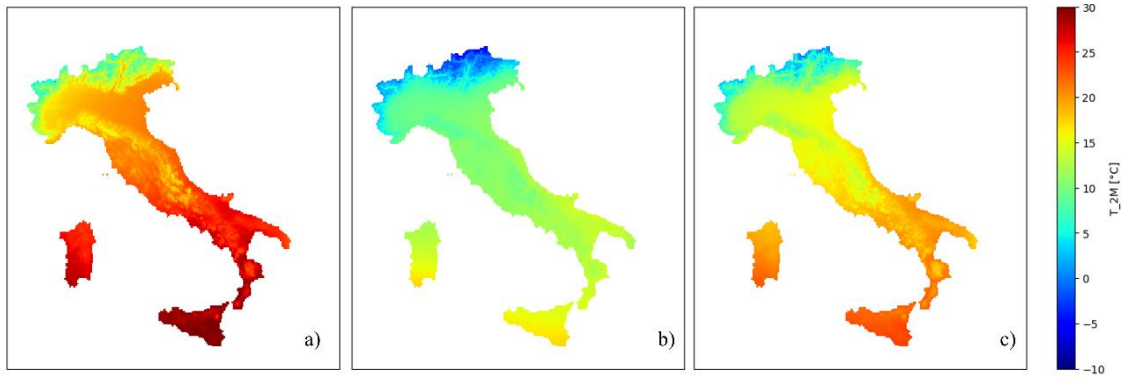


Fig. A.4: Maps of SCIA observed-gridded temperature dataset. (a) Maximum temperature (tmax); (b) Minimum temperature (tmin) and the derived mean temperature (c).

<i>Dataset</i>	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. A.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. A.5, along with the descriptive statistics presented in Tab. A.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.

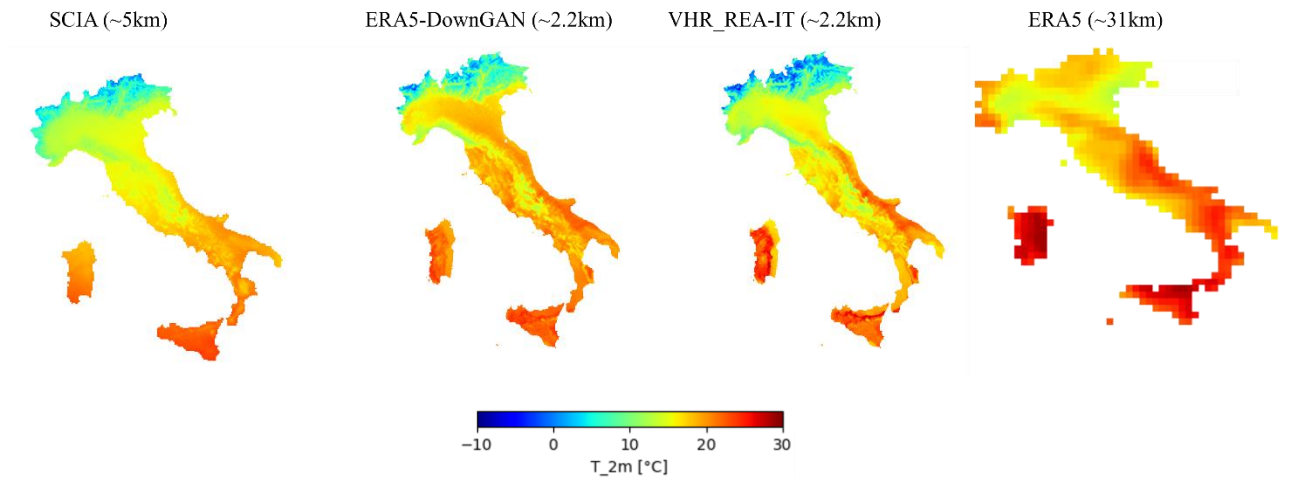


Fig. A.5: Comparison of 2m-temperature field only over Italian Peninsula on 24th March 2001. SCIA observed-gridded temperature dataset; ERA5-DownGAN; VHR_REA-IT and ERA5.

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 A.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 A.7, Panel a).

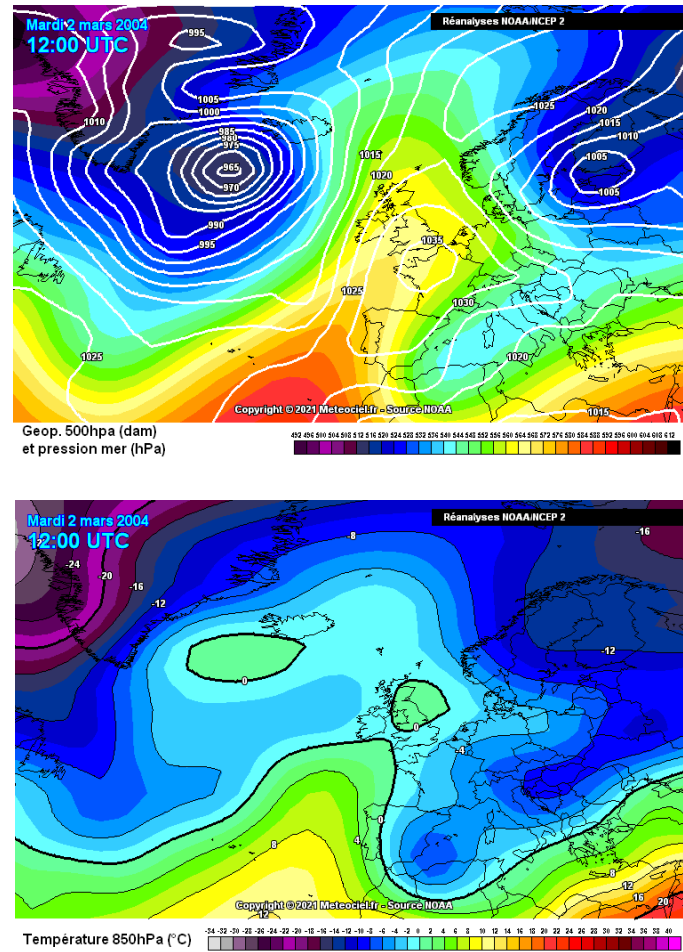


Fig. A.6: Synoptic analysis at 12:00 UTC from NOAA/NCEP reanalysis on 2nd March 2004. Maps of geopotential at 500 hPa (upper) and temperature (  C) at 850 hPa (lower). Reproduced from (www.meteociel.it).

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. A.7, Panel d). This result further confirms the effectiveness of the developed cGAN model in downscaling the 2m-temperature field.

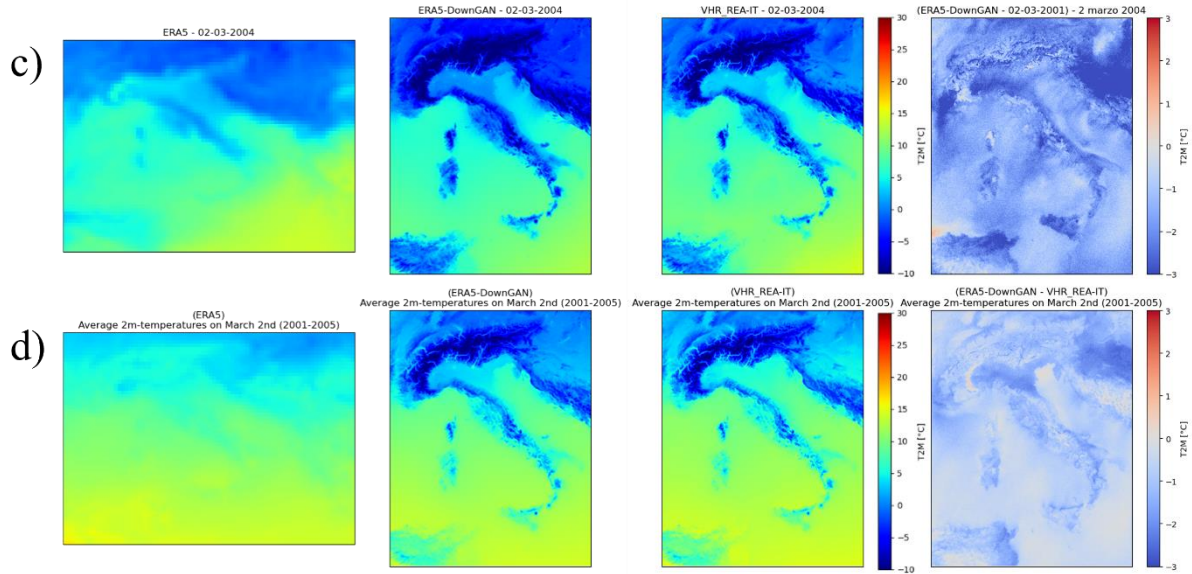


Fig. A.7: Panel (a): Maps for March 2, 2024, for ERA5 (LR), ERA5-DownGAN (HR), and VHR_REA-IT (HR), along with the difference between ERA5-DownGAN and VHR_REA-IT. Panel (b): Average of all March 2 observations from 2001 to 2005 for ERA5 (LR), ERA5-DownGAN (HR), and VHR_REA-IT (HR), as well as the difference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

<i>Dataset</i>	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. A.2: Comparison on 2nd 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.

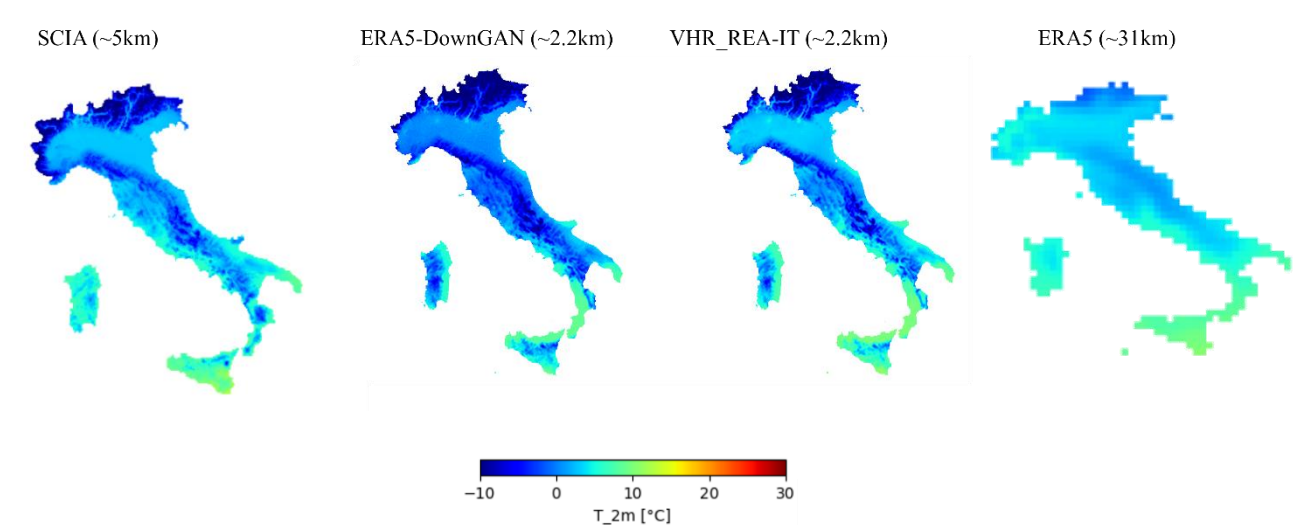


Fig. A.8:. Comparison of 2m-temperature field only over Italian Peninsula on 2nd March 2004. SCIA observed-gridded temperature dataset; ERA5-DownGAN; VHR_REA-IT and ERA5.

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 | Lateral work: Dynamical Downscaling over the Italian Peninsula Using WRF, COSMO, and ICON Atmospheric Models

In the context of enhancing climate modeling efforts, it is essential to investigate the intrinsic differences between statistical approaches and physically-based traditional climate models. To this end, a complementary study was conducted focusing on the configuration of the Weather Research and Forecasting (WRF) model in its version 4.2.1. This work aimed to develop an optimized configuration for the Italian peninsula, achieving a horizontal resolution of 2 km. Through extensive sensitivity testing, various physical parameterizations were

examined, including the boundary layer (PBL), microphysics, and short- and long-wave radiation schemes. This analysis facilitated an evaluation of the developed WRF configuration's performance compared to other atmospheric models, such as COSMO and ICON, at the same resolution, highlighting differences in predictive capabilities. The findings of this research were published in the paper by Manco I. et al. (2023), titled "A Comparative Performance Study of WRF, COSMO, and ICON Atmospheric Models for the Italian Peninsula at Very High Resolution." (DOI:10.3369/tethys.2023.20.01) The results demonstrate strong agreement and excellent performance of the optimized WRF configuration over the Italian Peninsula. Understanding these dynamics is essential for informing the statistical downscaling process using AI-based methods, as a solid foundation in physically-based models enhances the accuracy and relevance of local climate simulations.

Sensitivity Group	Planetary Boundary Layer Scheme	Deep Convection Scheme	Microphysics Scheme	Land Surface Scheme	Map Projections
ref	Mellor	no cumulus parameterization	Thompson graupel	unified Noah land-surface model	Mercator
Sim 1	Mellor	no cumulus parameterization	Thompson graupel	unified Noah land-surface model	Conformal Lambert
Sim 2	Mellor	Tiedtke	Thompson graupel	unified Noah land-surface model	Conformal Lambert
Sim 3	Mellor	no cumulus parameterization	Thompson graupel	thermal diffusion	Conformal Lambert
Sim 4	Mellor	no cumulus parameterization	WSM 6-class graupel	thermal diffusion	Conformal Lambert
Sim 5	Mellor	no cumulus parameterization	Morrison 2-moment	thermal diffusion	Conformal Lambert
Sim 6	Mellor	no cumulus parameterization	NSSL 2-moment 4-ice	thermal diffusion	Conformal Lambert
Sim 7	YSU	no cumulus parameterization	Thompson graupel	thermal diffusion	Conformal Lambert
Sim 8	YSU	no cumulus parameterization	WSM 6-class graupel	thermal diffusion	Conformal Lambert
Sim 9	YSU	no cumulus parameterization	Morrison 2-moment	thermal diffusion	Conformal Lambert
Sim 10	YSU	no cumulus parameterization	NSSL 2-moment 4-ice	thermal diffusion	Conformal Lambert

Tab. B.I: The sensitivity analysis about: PBL, microphysics, surface physics schemes and geographic projection has allowed to determinate the best configuration of the WRF model, that minimizes the error metrics for the precipitation, wind speed and temperature fields (Sim 9). Reproduced by Manco I. et al 2023.

The atmospheric forcing used was the Integrated Forecasting System (IFS) with analyses provided by the European Centre for Medium-Range Weather Forecasts (ECMWF), with a grid-point distance between neighboring points of approximately 9 km. The simulations were conducted in a single run of 15 days and the lateral boundary conditions for each model were refreshed every 6 hours.

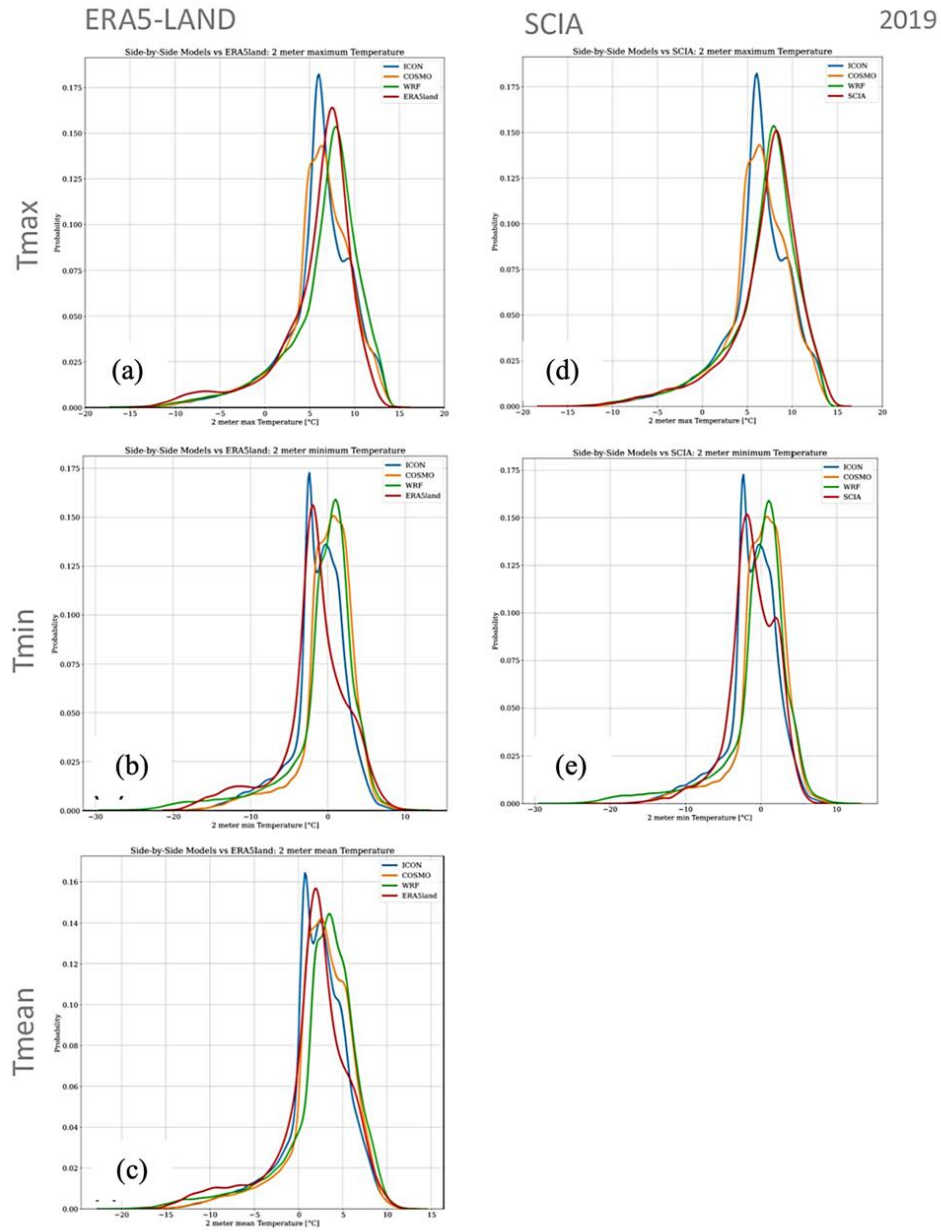


Fig. B.I: PDFs for 2019 of T_{max} (a,d), T_{min}(b,e) and T_{mean} (c). Panels (a,b,c) identify the evaluation against ERA5-Land and panels (d,e) identify the evaluation against SCIA. Reproduced by Manco I. et al 2023.

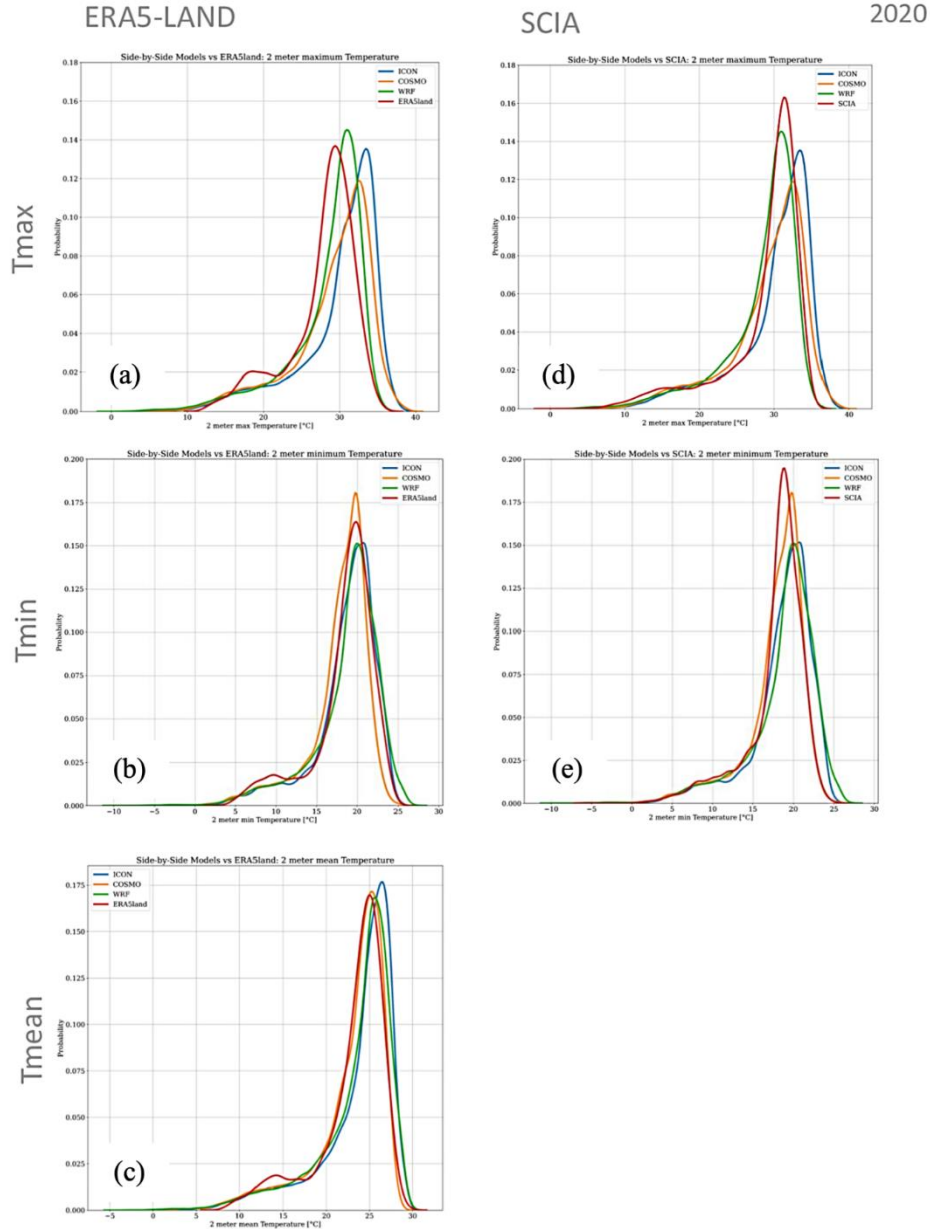


Fig. B.2: PDFs for 2020 of $T_{\max}$ (a,d), $T_{\min}$ (b,e) and T_{mean} (c). Panels (a,b,c) identify the evaluation against ERA5-Land and panels (d,e) identify the evaluation against SCIA. Reproduced by Manco I. et al 2023.

Through this sensitivity analysis, is possible to identify Sim9 as the most effective model configuration by assessing the simulated temperature and precipitation fields against observational data (details not shown here). A set of statistical metrics was employed for validation, such as MAE (mean absolute error), RMSE (root mean square error), MBIAS (mean bias), IoA (Index of Agreement), and Taylor diagrams. Notably, the application of the Planetary Boundary Layer "YSU" scheme led to significant performance improvements, providing a more

accurate representation of both temperature and precipitation ranges. Among the various microphysics options, the "Morrison 2-moment scheme" proved to be the best for accurately modeling precipitation. Additionally, employing a Lambert projection minimized distortions compared to the Mercator projection. For all simulations, the Monin-Obukhov scheme (Janjic Eta) was utilized to effectively parameterize the surface layer.

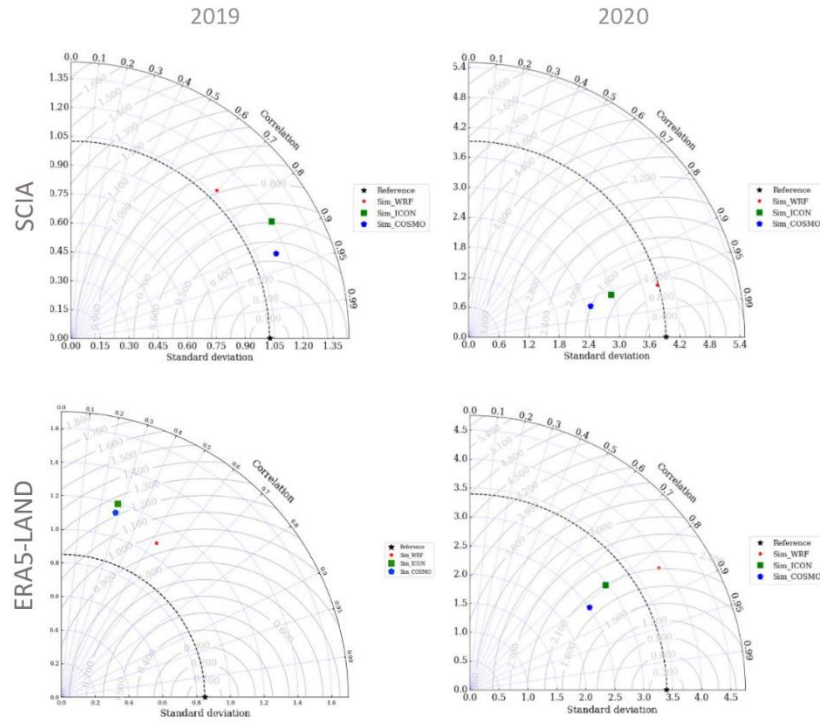


Fig. B.3: Taylor Diagrams of the daily cumulative precipitation at 2m in WRF, COSMO and ICON evaluated against ERA5-Land (a,c) and SCIA (b,d) in 2019 (left column) and 2020 (right column). Reproduced by Manco I. et al 2023.

The best-model configurations for the Weather Research and Forecasting (WRF), COSMO, and ICON models were rigorously evaluated against two observational datasets, ERA5-Land and SCIA. This assessment aimed to quantify their efficacy in simulating key atmospheric variables, namely minimum temperature (t_{min}), mean temperature (t_{mean}), maximum temperature (t_{max}), shown in Fig. B.I and Fig. B.2, total precipitation (Tot prec), in Fig. B.3, and wind speed at 10 meters (W10m), in Fig. B. 4, during two distinct periods: winter 2020 and summer 2019. All models demonstrated robust performance in reproducing temperature, precipitation, and wind speed, achieving high correlation coefficients ranging from 80% to 99% for temperature variables. Additionally, low root mean square error (RMSE) values were observed, particularly in the simulation of temperature fields. In terms of precipitation, the models achieved an

accuracy of 70% to 96% when evaluated against SCIA data. However, discrepancies in performance were noted when comparing the models with the coarser resolution of ERA5-Land, which may inhibit the ability to resolve fine-scale dynamic processes effectively.

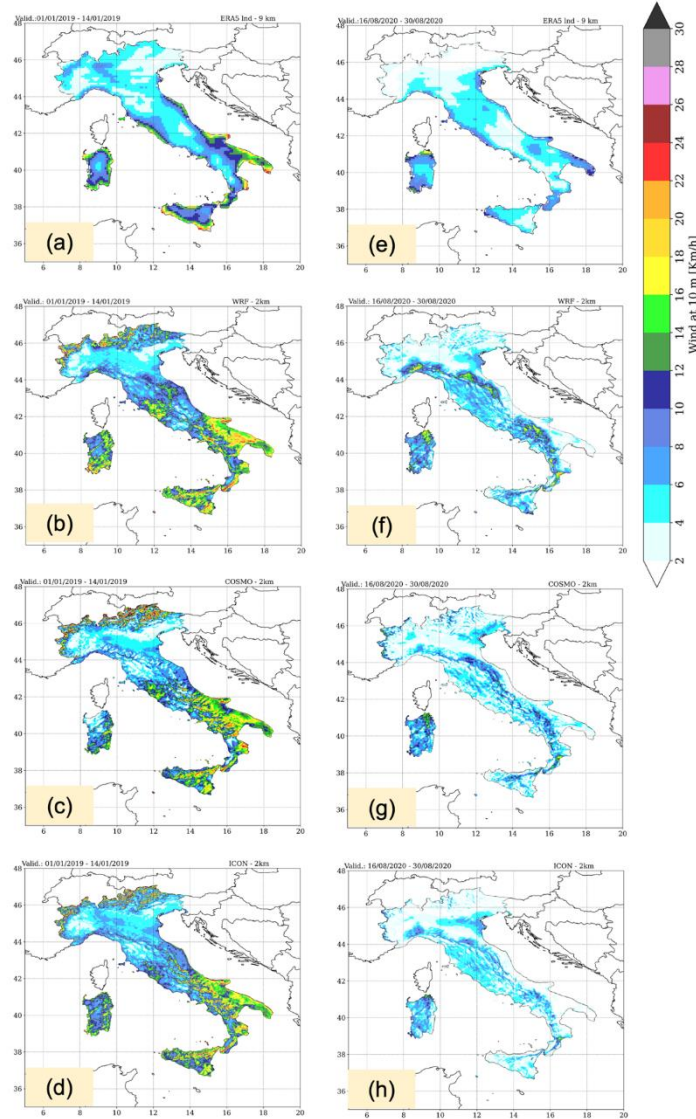


Fig. B.4: Map of time-mean wind speed at 10 m for ERA5-Land (a,f), SCIA (b,g), WRF (c,h), COSMO (d,i) and ICON (e,l). Reproduced by Manco I. et al 2023.

In particular, WRF accurately represented specific meteorological conditions, including cold temperatures and strong winds in winter, and intense convection in summer. The configuration of WRF, optimized for the Italian domain, performed on par with or exceeded ICON configurations from De Lucia et al. (2022) and COSMO configurations described by Baldauf et al. (2011), with all models tested at a similar horizontal resolution of approximately 2 km.

C | Lateral work: AI-Based Method for Identifying Historical Recurrent Precipitation Patterns and Evaluating Changes in Future Projections

The high-resolution dataset generated through the statistical downscaling methodology based on a cGAN developed in this study is intended to serve as a robust and versatile product, comparable to dynamically downscaled datasets. Its applicability has been assessed across various scientific questions commonly encountered in meteorology and climatology. Among these applications, a key focus was placed on identifying geographically recurring patterns during extreme precipitation events. For this purpose, a supplementary methodology (see Fig. C.I) was developed, utilizing the k-means clustering algorithm (alongside trials with alternative algorithms, including hierarchical clustering). This investigation was carried out in parallel as part of the PNRR-HAMMON project, in collaboration with the CMCC.

Starting from a high-resolution dataset of daily cumulative precipitation and the creation of an extreme-event catalog, this approach identifies statistically recurring extreme regimes over a selected region and delineates their geographic extent. This process enables the classification of areas affected by similar precipitation values during extreme events. Figure C.I presents the average precipitation values recorded within the catalog of extreme events, defined as shown in Fig. C.3.

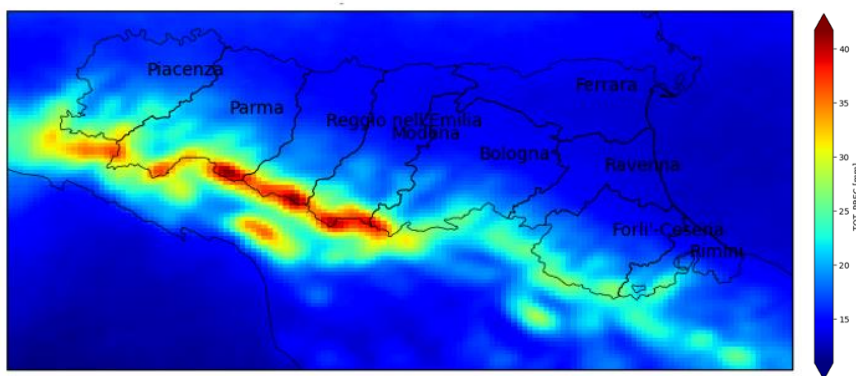


Fig. C.I: Map of Mean Precipitation on Extreme Event Days Exceeding the 90th Percentile.

The k-means algorithm is an unsupervised clustering technique widely used to partition datasets into k groups, or clusters, based on data proximity (Lloyd, 1982; MacQueen, 1967). The algorithm begins with the random selection of k centroids, around which data points are iteratively grouped. At each iteration, each data point is assigned to the nearest centroid, followed by recalculating each centroid's position based on the mean of its assigned points. This process repeats until the centroids stabilize or a predefined number of iterations is reached. However, a known limitation of the k-means algorithm is the requirement to set the number of clusters k beforehand, which directly influences the clustering result (Jain A. K., 2010). Determining the optimal k value is crucial for obtaining meaningful clusters within the dataset. To address this, this study applied the inertia plot, or "elbow method" (Thorndike, 1953), to identify the best k . The plot represents the sum of squared distances between data points and their respective centroids for various k values. The optimal number of clusters, in our case equal to 10, is typically indicated by the "elbow" in the plot, where increasing the number of clusters yields only marginal improvements, suggesting an appropriate level of differentiation among groups.

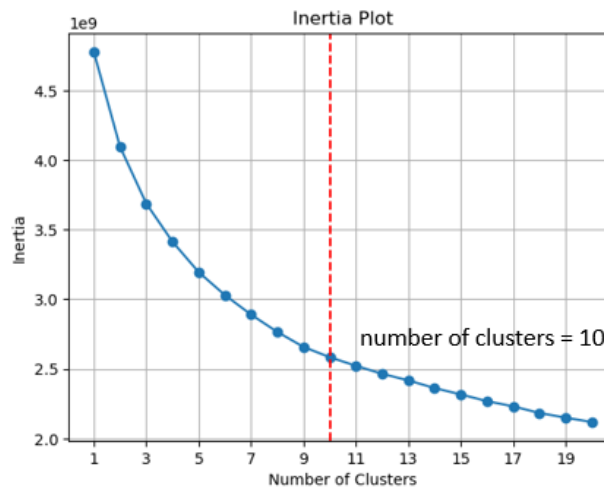


Fig. C.2: Inertia Plot to identify the optimal numbers of clusters.

The k-means algorithm is particularly effective at identifying natural clusters or recurring patterns within large datasets, including meteorological data, making it highly valuable in climate studies (Jain A. K., 2010; Xu R. and Wunsch D. I. I., 2005). Its application allows researchers to extract meaningful information about precipitation regimes, especially when examining extreme events. By isolating areas with homogeneous behaviors, k-means facilitates

the analysis of climate phenomena at regional scales, aiding in the understanding of spatial patterns associated with extreme weather conditions (Deka, P., and Saha, U., 2023).

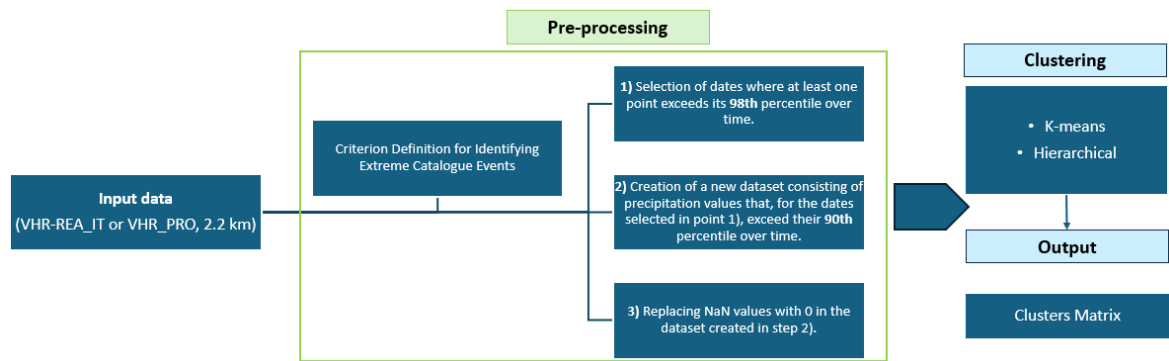


Fig. C.3: Work-flow of clustering methodology.

Figure C.4 presents a case study of the Emilia-Romagna region, which is notably affected by very intense precipitation events throughout the entire historical period 1981-2005).

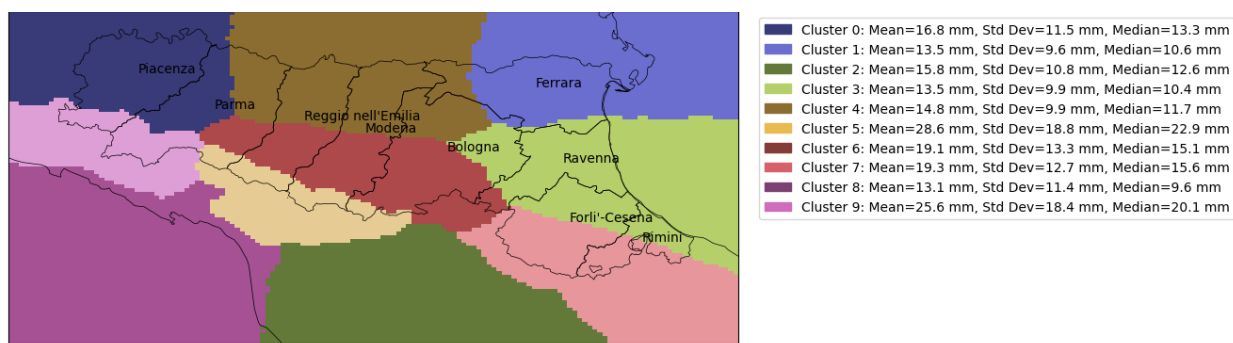


Fig. C.4: K-means Clustering 1981-2020.

One limitation of the k-means algorithm is its difficulty in accurately representing non-spherical patterns within data. However, through the developed methodology, which includes a preliminary filtering of extreme events, the algorithm successfully identifies the core of the most intense precipitation, as also shown in Fig. C.I following the clustering process (highlighted in orange in Fig. C.5). This outcome confirms k-means’ ability, once optimized for extreme events, to precisely delineate clusters corresponding to areas of higher precipitation intensity.

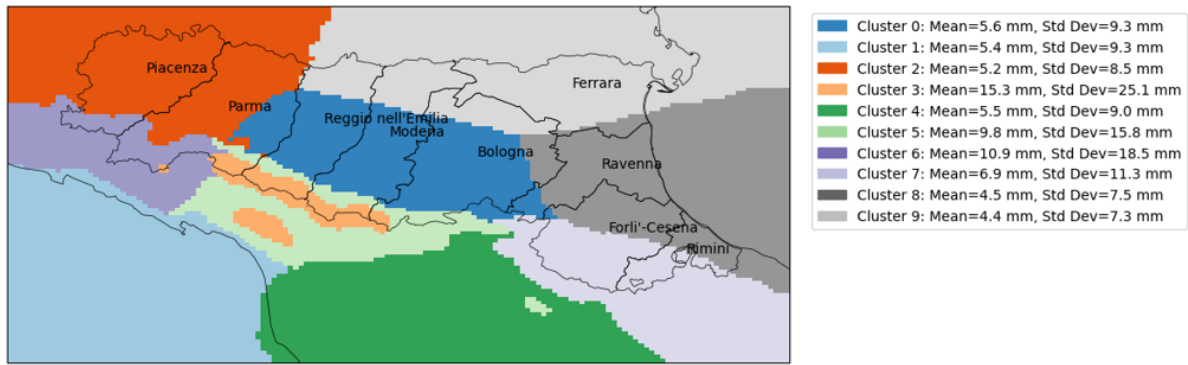


Fig. C.5: K-means Clustering 1981-2020 (winter).

The investigation initially examined the historical period from 1981 to 2005, both as a whole and seasonally. Subsequently, in line with the findings of Puente A. et al. (2015), which emphasize the importance of extending cluster-based analyses to future scenarios, we aimed to understand how these clusters may evolve in terms of localization and spatial extent under the RCP8.5 scenario. To this end, the VHR_PRO dataset (Raffa M. et al., 2023) was utilized to compare historical data with future projections (Fig. C.6).

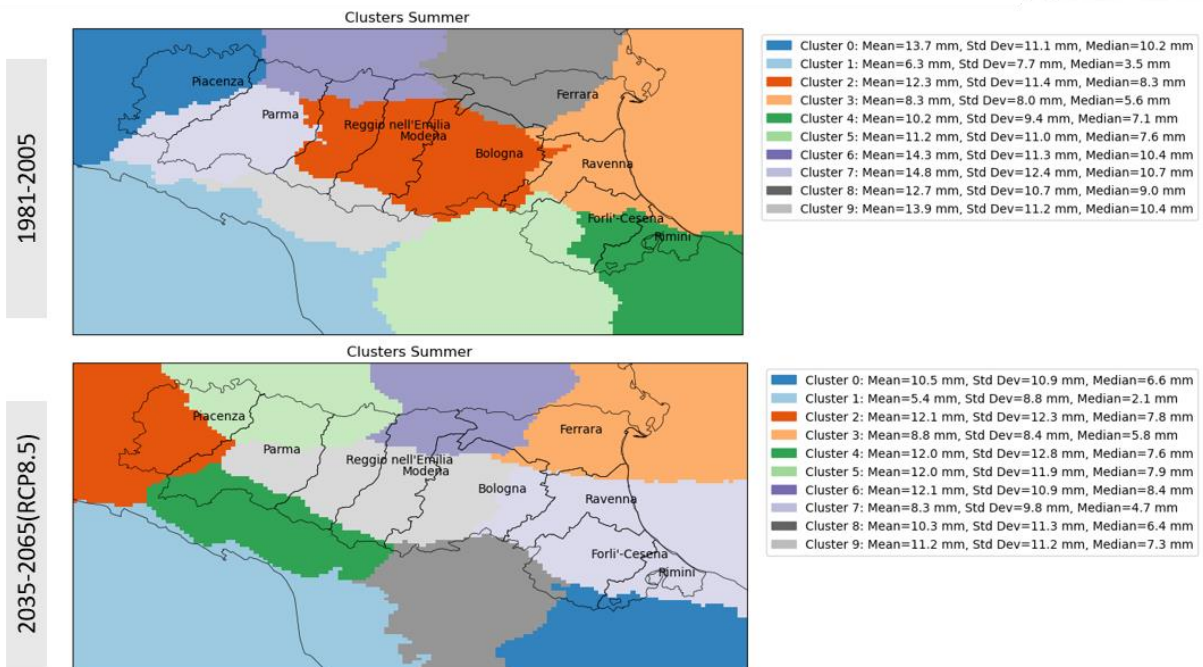


Fig. C.6: Comparison of cluster distribution for the historical (1981-2005) and future (2035-2065) RCP8.5 during the summer.

In conclusion, the investigation revealed no significant differences in the extent and localization of clusters when comparing historical data with projections under the RCP8.5 scenario. Key statistical measures such as mean, median, and standard deviation remained relatively consistent across the datasets.

However, when examining seasonal variations, distinct differences emerged. During the winter months, the RCP8.5 scenario exhibited significantly higher precipitation values for the cluster representing the most intense precipitation events compared to historical data. In contrast, summer showed notable changes in the localization and extent of the clusters. Specifically, in the provinces of Parma and Piacenza, lower average precipitation values and higher standard deviations were observed, indicating increased variability in precipitation patterns during this season. This highlights the importance of conducting seasonal analyses to capture the nuanced changes that may occur under future climate scenarios. The results of this preliminary investigation using the dynamic dataset lay the groundwork as a benchmark for future experiments involving the dataset generated through statistical downscaling using the GAN model developed in this study.

Code and data availability

The cGAN developed in this study will be made publicly available upon publication. The code will be accessible through Zenodo and will be fully open-source. The high-resolution dataset produced by cGAN-downscaling will be available on the DDS (Data Delivery System) of CMCC (<https://dds.cmcc.it/>). The implementation is written in Python and operates within a Miniconda environment. The machine learning components require the following libraries: PyTorch, Torchvision, Scipy, and Numpy. For visualization, the required libraries include Matplotlib and Torchvision.utils. A list of dependencies, including the specific versions of these libraries, is provided below to ensure proper environment setup.

Name	Version	Build	Channel
_libgcc_mutex	0.1	conda_forge	conda-forge
_openmp_mutex	4.5	2_gnu	conda-forge
absl-py	1.4.0	pyhd8ed1ab_0	conda-forge
affine	2.4.0	pypi_0	pypi
aiohttp	3.8.4	py39h72bdee0_0	conda-forge
aiosignal	1.3.1	pyhd8ed1ab_0	conda-forge
anyio	3.6.2	pyhd8ed1ab_0	conda-forge
argon2-cffi	21.3.0	pyhd8ed1ab_0	conda-forge
argon2-cffi-bindings	21.2.0	py39h7f8727e_0	anaconda
asttokens	2.2.1	pyhd8ed1ab_0	conda-forge
async-timeout	4.0.2	py39h06a4308_0	anaconda

Name	Version	Build	Channel
attrs	22.1.0	pyh71513ae_1	conda-forge
autopep8	2.0.4	pypi_0	pypi
babel	2.11.0	pyhd8ed1ab_0	conda-forge
backcall	0.2.0	pyh9f0ad1d_0	conda-forge
backports	1.1	pyhd3eb1b0_0	anaconda
backports.functools_lru_cache	1.6.4	pyhd8ed1ab_0	conda-forge
beautifulsoup4	4.11.1	pyha770c72_0	conda-forge
black	23.10.1	pypi_0	pypi
blas	1.0	mk1	anaconda
bleach	5.0.1	pyhd8ed1ab_0	conda-forge
blinker	1.5	pyhd8ed1ab_0	conda-forge
blosc	1.21.0	h4ff587b_1	anaconda
brotli	1.0.9	h166bdaf_8	conda-forge
brotli-bin	1.0.9	h166bdaf_8	conda-forge
brotlipy	0.7.0	py39h27cfd23_1003	anaconda
brunslis	0.1	h2531618_0	anaconda
bzip2	1.0.8	h7f98852_4	conda-forge
c-ares	1.18.1	h7f98852_0	conda-forge

Name	Version	Build	Channel
ca-certificates	2024.11.26	h06a4308_0	
cachetools	5.3.0	pyhd8ed1ab_0	conda-forge
cartopy	0.22.0	py39hddac248_1	conda-forge
cdo	1.5.7	pypi_0	pypi
certifi	2024.8.30	py39h06a4308_0	
cf_xarray	0.7.6	pyhd8ed1ab_0	conda-forge
cffi	1.15.1	py39h5eee18b_3	anaconda
cfitsio	3.470	h5893167_7	anaconda
cftime	1.6.2	py39h2ae25f5_1	conda-forge
charls	2.2.0	h2531618_0	anaconda
charset-normalizer	2.1.1	pyhd8ed1ab_0	conda-forge
click	8.1.3	unix_pyhd8ed1ab_2	conda-forge
click-plugins	1.1.1	pypi_0	pypi
cligj	0.7.2	pypi_0	pypi
cloudpickle	2.2.1	py39h06a4308_0	anaconda
colorama	0.4.6	py39h06a4308_0	
comm	0.1.2	pyhd8ed1ab_0	conda-forge
contourpy	1.0.6	py39hf939315_0	conda-forge

Name	Version	Build	Channel
cryptography	38.0.1	py39h9ce1e76_0	anaconda
curl	7.86.0	h5eee18b_0	anaconda
cycler	0.11.0	pyhd8ed1ab_0	conda-forge
cytoolz	0.12.0	py39h5eee18b_0	anaconda
dask-core	2023.6.0	py39h06a4308_0	anaconda
dbus	1.13.18	hb2f20db_0	anaconda
debugpy	1.5.1	py39h295c915_0	anaconda
decorator	5.1.1	pyhd8ed1ab_0	conda-forge
defusedxml	0.7.1	pyhd8ed1ab_0	conda-forge
distributed	2023.6.0	py39h06a4308_0	anaconda
docutils	0.18.1	pyhd8ed1ab_0	conda-forge
entrypoints	0.4	pyhd8ed1ab_0	conda-forge
esmpy	8.1.0	py39h635207f_0	conda-forge
flask	2.2.2	py39h06a4308_0	anaconda
fonttools	4.39.0	pyhd8ed1ab_0	conda-forge
fsspec	2023.5.0	pyhd8ed1ab_0	conda-forge
future	0.18.3	pyhd8ed1ab_0	conda-forge
geospatial-python	0.11.0	pyhd8ed1ab_0	conda-forge

Name	Version	Build	Channel
gdal	3.7.2	h2ee2c1b_0	conda-forge
geoviews	1.10.4	pyhd8ed1ab_0	conda-forge
git	2.39.1	h9ed2024_0	conda-forge
greenlet	2.0.2	py39h5eee18b_0	conda-forge
h5py	3.8.0	py39h0fd1996_0	anaconda
icu	70.1	h2f8d375_0	conda-forge
importlib-metadata	6.0.0	pyhd8ed1ab_0	conda-forge
ipykernel	6.22.0	py39h06a4308_0	anaconda
ipython	8.11.0	py39h06a4308_0	anaconda
ipython_genutils	0.2.0	pyhd8ed1ab_0	conda-forge
jinja2	3.1.2	pyhd8ed1ab_0	conda-forge
joblib	1.2.0	pyhd8ed1ab_0	conda-forge
json5	0.9.6	pyhd8ed1ab_0	conda-forge
jsonschema	4.17.3	pyhd8ed1ab_0	conda-forge
jupyter_client	8.2.0	py39h06a4308_0	anaconda
jupyter_core	5.3.0	pyhd8ed1ab_0	conda-forge
jupyterhub	3.2.0	py39h06a4308_0	anaconda

Acknowledgments

I would like to express my heartfelt gratitude to all the individuals and institutions that have contributed to the research work developed throughout this journey. In particular, sincere recognition is due to my supervisor, Prof. Antonio Navarra, who, with his pioneering vision, has believed in this work and in me from the very beginning, encouraging me to believe in the tangible possibility of contributing a piece to the vast scientific landscape.

A warm thank you is also extended to Dr. Paola Mercogliano, whose guidance over the years has been essential. Thanks to her support, I have been able to enrich my knowledge base and tackle increasingly stimulating challenges aimed at shaping me into a researcher: not always with ready answers, but always ready to find them at all costs.

I would like to extend special thanks to Dr. Marco Briscolini for believing in the potential of this work, and to E4 Computer Engineering for financing it. A deserved acknowledgment goes to Dr. Walter Riviera and Dr. Andrea Zanetti, and in general to Intel Corporation, for their invaluable support in the field of Artificial Intelligence, especially as I began to transition from physically-based models to the realm of AI models.

I would also like to thank Dr. Ethan Gutmann for accepting me as a visitor at the Research Application Laboratory (RAL) of the National Center for Atmospheric Research (NCAR) in Boulder, Colorado. The engaging meetings, accompanied by a whiteboard and markers, allowed me to push my research toward new frontiers. A heartfelt thank you is also due to his entire research group, with whom I had productive and stimulating scientific discussions, as well as to NCAR for hosting me and providing the necessary computational resources. I wish to express my gratitude to the Euro-Mediterranean Center for Climate Change (CMCC) for supplying the computational resources essential to my work.

Finally, I would like to extend a heartfelt thank you to my PhD colleagues and to my colleagues at the REMHI division of CMCC, who have contributed to enriching this journey even further.

Bibliography

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146, 1999-2049. <https://doi.org/10.1002/qj.3803>.

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(88). <https://doi.org/10.3390/data6080088>.

Kondrup, C., Mercogliano, P., Bosello, F., Mysiak, J., Scoccimarro, E., Rizzo, A., Ebrey, R., Ruiter, M. de, Jeuken, A., & Watkiss, P. (2022). Climate adaptation modelling (p. 238). Springer Nature. <https://doi.org/10.1007/978-3-030-86211-4>.

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. *Irrigation and Drainage*, <https://doi.org/10.1002/ird.2653>.

Sun, Q., Fang, J., Dang, X., Xu, K., Fang, Y., Li, X., and 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>.

Tabor, K., & Williams, J. W. (2010). Globally downscaled climate projections for assessing the conservation impacts of climate change. *Ecological Applications*, 20(2), 554–565. <http://www.jstor.org/stable/27797827>.

Zhang, F., & Georgakakos, A. P. (2012). Climate change impacts on the hydrology of the North American Great Lakes. *Climatic Change*, 111(3), 945–972. <https://doi.org/10.1007/s10584-011-0167-9>.

- Benestad, R., Hanssen-Bauer, I., & Chen, D. (2008). Empirical-Statistical Downscaling. World Scientific Editions. <https://doi.org/10.1142/6908>.
- Yin, S., & Chen, D. (2020). Weather generators. In Oxford Research Encyclopedia of Climate Science. <https://doi.org/10.1093/acrefore/9780190228620.013.768>
- Panofsky, H. W. and Brier, G. W. (1968), Some applications of statistics to meteorology. The Pennsylvania State University Press.
- Liu, Jiaming & Yuan, Di & Zhang, Liping & Zou, Xia & Song, Xingyuan. (2015). Comparison of Three Statistical Downscaling Methods and Ensemble Downscaling Method Based on Bayesian Model Averaging in Upper Hanjiang River Basin, China. *Advances in Meteorology*. 2016. <https://doi.org/10.1155/2016/7463963>.
- Leinonen, J., Grazioli, J., & Berne, A. (2021). Reconstruction of the mass and geometry of snowfall particles from multi-angle snowflake camera (MASC) images. *Atmospheric Measurement Techniques*, 14(10), 6851–6866. <https://doi.org/10.5194/amt-14-6851-2021>.
- Kumar, B., Atey, K., Singh, B.B. et al. On the modern deep learning approaches for precipitation downscaling. *Earth Sci Inform* 16, 1459–1472 (2023). <https://doi.org/10.1007/s12145-023-00970-4>.
- Mamalakis, A., Barnes, E. A., & Ebert-Uphoff, I. (2022). Investigating the Fidelity of Explainable Artificial Intelligence Methods for Applications of Convolutional Neural Networks in Geoscience. *Artificial Intelligence for the Earth Systems*, 1(4), e220012. <https://doi.org/10.1175/AIES-D-22-0012.1>
- Russell, S., & Norvig, P. (1995). Artificial intelligence: A modern approach. Prentice Hall.
- Newell, A., & Simon, H. A. (1972). Human problem solving. Prentice Hall.
- Hinton, G. E., & Bengio, Y. (1986). Learning representations by back-propagating errors. In *Proceedings of the IEEE 1986 Workshop on Neural Networks for Signal Processing* (pp. 1–7).
- de Burgh-Day, C. O., & Leeuwenburg, T. (2023). Machine learning for numerical weather and climate modelling: A review. *Geoscientific Model Development*, 16(22), 6433–6477. <https://doi.org/10.5194/gmd-16-6433-2023>.

Wang J., Wang X., Guan J., Zhang L., Zhang F., Chang T. STPF-Net: Short-Term Precipitation Forecast Based on a Recurrent Neural Network. *Remote Sensing*. 2024; 16(1):52. <https://doi.org/10.3390/rs16010052>.

Haupt, S. E., Chapman, W., Adams, S. V., Kirkwood, C., Hosking, J. S., Robinson, N. H., Lerch, S., & Subramanian, A. C. (2021). Towards implementing artificial intelligence post-processing in weather and climate: proposed actions from the Oxford 2019 workshop. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 379(2194). <https://doi.org/10.1098/rsta.2020.0091>.

Baño-Medina, J., Manzananas, R., & Gutiérrez, J. M. (2020). Configuration and intercomparison of deep learning neural models for statistical downscaling. *Geoscientific Model Development*, 13(4), 2109–2124. <https://doi.org/10.5194/gmd-13-2109-2020>.

Sadai, S., Condrón, A., DeConto, R., & Pollard, D. (2020). Future climate response to Antarctic Ice Sheet melt caused by anthropogenic warming. *Science Advances*, 6(39), eaat1169. <https://doi.org/10.1126/sciadv.aaz1169>.

Barnes, E. A., Hurrell, J. W., Ebert-Uphoff, I., Anderson, D., & others. (2019). Viewing forced climate patterns through an AI lens. *Geophysical Research Letters*, 46(4), e2019GL084944. <https://doi.org/10.1029/2019GL084944>.

Reichstein, M., Camps-Valls, G., Stevens, B., Prabhat, & others. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, 566(7743), 195. <https://doi.org/10.1038/s41586-019-0912-1>.

Maraun, D., & Widmann, M. (2018). Statistical downscaling and bias correction for climate research. Cambridge University Press. <https://doi.org/10.1017/9781107588783>.

Klein, W., & Glahn, H. (1974). Forecasting local weather by means of model output statistics. *Bulletin of the American Meteorological Society*, 55, 1217–1217. [https://doi.org/10.1175/1520-0477\(1974\)055<1217:FLWBMO>2.0.CO;2](https://doi.org/10.1175/1520-0477(1974)055<1217:FLWBMO>2.0.CO;2).

Wilby, R. L., & Wigley, T. M. L. (1997). Downscaling general circulation model output: a review of methods and limitations. *Progress in Physical Geography: Earth and Environment*, 21(4), 530-548. <https://doi.org/10.1177/030913339702100403>.

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>.

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. *Climate Dynamics*, 60(5–6), 1751–1779. <https://doi.org/10.1007/s00382-022-06343-9>.

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 Resources Research*, 34(11), 2995–3008. <https://doi.org/10.1029/98WR02577>.

Dayon, G., Boé, J., & Martin, E. (2015). Transferability in the future climate of a statistical downscaling method for precipitation in France. *Journal of Geophysical Research: Atmospheres*, 120(3), 1023–1043. <https://doi.org/10.1002/2014JD022236>.

Erlandsen, H. B., Parding, K. M., Benestad, R., Mezghani, A., & Pontoppidan, M. (2020). A hybrid downscaling approach for future temperature and precipitation change. *Journal of Applied Meteorology and Climatology*, 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. *WMO Bulletin*, 58, 175–183.

Boé, J., A. Mass, and J. Deman, 2023: A simple hybrid statistical–dynamical downscaling method for emulating regional climate models over Western Europe. Evaluation, application, and role of added value? *Climate Dyn.*, 61, 271–294, <https://doi.org/10.1007/s00382-022-06552-2>.

Sharifi, E., B. Saghafian, and R. Steinacker, 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>.

Holden, P. B., N. R. Edwards, P. H. Garthwaite, and R. D. Wilkinson, 2015: Emulation and interpretation of high-dimensional climate model outputs. *J. Appl. Stat.*, 42, 2038–2055, <https://doi.org/10.1080/02664763.2015.1016412>.

Feng, D., H. Beck, K. Lawson, and C. Shen, 2023: The suitability of differentiable, physics-informed machine learning hydrologic models for ungauged regions and climate change impact

assessment. *Hydrol. Earth Syst. Sci.*, 27, 2357–2373, <https://doi.org/10.5194/hess-27-2357-2023>.

Hutengs, C., and M. Vohland, 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>.

Limon, G., & Jablonowski, C. (2023). Probing the skill of random forest emulators for physical parameterizations via a hierarchy of simple CAM6 configurations. *Journal of Advances in Modeling Earth Systems*, 15, e2022MS003395. <https://doi.org/10.1029/2022MS003395>.

Baño-Medina, J., R. Manzananas, and J. M. Gutiérrez, 2020: Configuration and intercomparison of deep learning neural models for statistical downscaling. *Geosci. Model Dev.*, 13, 2109–2124, <https://doi.org/10.5194/gmd-13-2109-2020>.

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. *Nature Climate Change*, 7, 362–370. <https://doi.org/10.1038/nclimate3418>.

Nishant, N., S. Hobeichi, S. Sherwood, G. Abramowitz, Y. Shao, C. Bishop, and A. Pitman, 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>.

Hobeichi, S., N. Nishant, Y. Shao, G. Abramowitz, A. Pitman, S. Sherwood, C. Bishop, and S. Green, 2023: Using machine learning to cut the cost of dynamical downscaling. *Earth's Future*, 11, e2022EF003291, <https://doi.org/10.1029/2022EF003291>.

Bittner, M., S. Hobeichi, M. Zawish, S. Diatta, R. Ozioko, S. Xu, and A. Jantsch, 2023: An LSTM-based downscaling framework for Australian precipitation projections. *NeurIPS 2023 Workshop on Tackling Climate Change with Machine Learning*, New Orleans, LA, NeurIPS, <https://www.climatechange.ai/papers/neurips2023/46>.

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>.

Leinonen, J., D. Nerini, and A. Berne, 2021: Stochastic superresolution for downscaling time-evolving atmospheric fields with a generative adversarial network. *IEEE Trans. Geosci. Remote Sens.*, 59, 7211–7223, <https://doi.org/10.1109/TGRS.2020.3032790>.

Doury, Antoine & Somot, Samuel & Gadat, Sébastien & Ribes, Aurélien & Corre, Lola. (2022). Regional climate model emulator based on deep learning: concept and first evaluation of a novel hybrid downscaling approach. *Climate Dynamics*. 60. 1-29. [10.1007/s00382-022-06343-9](https://doi.org/10.1007/s00382-022-06343-9).

Lu, L., Jin, P., & Karniadakis, G. E. (2019). DeepONet: Learning nonlinear operators for identifying differential equations based on the universal approximation theorem of operators. *arXiv preprint*. <https://doi.org/10.48550/arXiv.1910.03193>.

Haykin, S. (1998). *Neural Networks: A Comprehensive Foundation* (2nd ed.). Prentice Hall. <https://doi.org/10.5555/12345678>.

Wood, J., & Shawe-Taylor, J. (1996). Representation theory and invariant neural networks. *Discrete Applied Mathematics*, 69(1-2), 33-60. [https://doi.org/10.1016/0166-218X\(95\)00075-3](https://doi.org/10.1016/0166-218X(95)00075-3).

Fischler, M. A., & Firschein, O. (1990). Intelligence: The eye, the brain and the computer. *Journal of Social and Biological Systems*, 13(3), 284-287. [https://doi.org/10.1016/0140-1750\(90\)90016-Y](https://doi.org/10.1016/0140-1750(90)90016-Y).

Rumelhart, D. & Hinton, G. & McClelland, James. (1986). A General Framework for Parallel Distributed Processing. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. 1.

Child, J. M. (1916). The manuscripts of Leibniz on his discovery of the differential calculus. *The Monist*, 26(4), 577-629. <https://doi.org/10.1086/283286>.

Churchland, P. S., & Sejnowski, T. J. (1992). *The computational brain*. The MIT Press.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324. <https://doi.org/10.1109/5.726791>.

Widrow, B., & Stearns, S. D. (1985). *Adaptive signal processing*. Prentice-Hall.

Haykin, S. (2002) *Adaptive Filter Theory*. 4th Edition, Prentice Hall, Upper Saddle River.

Miralles, O., D. Steinfeld, O. Martius, and A. C. Davison, 2022: Downscaling of historical wind fields over Switzerland using generative adversarial networks. *Artif. Intell. Earth Syst.*, 1, e220018, <https://doi.org/10.1175/AIES-D-22-0018.1>.

Wang, J., Z. Liu, I. Foster, W. Chang, R. Kettimuthu, and V. R. Kotamarthi, 2021: Fast and accurate learned multiresolution dynamical downscaling for precipitation. *Geosci. Model Dev.*, 14, 6355–6372, <https://doi.org/10.5194/gmd-14-6355-2021>.

Goodfellow, I., Y. Bengio, and A. Courville, 2016: *Deep Learning*. MIT Press, 800 pp.

Kingma, D.P. and Welling, M. (2014) Auto-Encoding Variational Bayes. *ICLR 2014*, 14-16 April 2014, Banff, Canada.

Rezende, D. J., & Mohamed, S. (2015). Variational inference with normalizing flows. *Proceedings of the International Conference on Learning Representations (ICLR)*. <https://doi.org/10.48550/arXiv.1505.05770>.

Oord, A. v. d., Kalchbrenner, N., & Kavukcuoglu, K. (2016). Pixel recurrent neural networks. *arXiv preprint arXiv:1601.06759*. <https://doi.org/10.48550/arXiv.1601.06759>.

Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., & Chen, X. (2016). Improved techniques for training GANs. *Advances in Neural Information Processing Systems*, 29. <https://doi.org/10.5555/6125>.

Mirza, M., & Osindero, S. (2014). Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*. <https://doi.org/10.48550/arXiv.1411.1784>.

Vandal, T., E. Kodra, S. Ganguly, A. Michaelis, R. Nemani, and A. R. Ganguly, 2017: DeepSD: Generating high resolution climate change projections through single image super-resolution. *arXiv*, 1703.03126v1, <https://doi.org/10.48550/arXiv.1703.03126>.

Stengel, K., A. Glaws, D. Hettinger, and R. N. King, 2020: Adversarial super-resolution of climatological wind and solar data. *Proc. Natl. Acad. Sci. USA*, 117, 16 805–16 815, <https://doi.org/10.1073/pnas.1918964117>.

Liu, Y., A. R. Ganguly, and J. Dy, 2020: Climate downscaling using YNet: A deep convolutional network with skip connections and fusion. *Proc. 26th ACM SIGKDD Int. Conf. on Knowledge Discovery & Data Mining*, Online, Association for Computing Machinery, 3145–3153, <https://doi.org/10.1145/3394486.3403366>.

Maraun, D., Wetterhall, F., Ireson, A. M., Thiele-Eich, I., & et al. (2010). Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Reviews of Geophysics*, 48(3), RG3003. <https://doi.org/10.1029/2009RG000314>.

Wilby, R. L., C. W. Dawson, and E. M. Barrow, 2002: SDSM} A decision support tool for the assessment of regional climate change impacts. *Environ. Modell. Software*, 17, 145–157, [https://doi.org/10.1016/S1364-8152\(01\)00060-3](https://doi.org/10.1016/S1364-8152(01)00060-3).

Hay, L. E., Clark, M. P., Wilby, R. L., Gutowski, W. J. Jr., Leavesley, G. H., Pan, Z., Arritt, R. W., & Takle, E. S. (2002). Use of regional climate model output for hydrologic simulations. *Journal of Hydrometeorology*, 3(5), 571–590. [https://doi.org/10.1175/1525-7541\(2002\)003<0571>2.0.CO;2](https://doi.org/10.1175/1525-7541(2002)003<0571>2.0.CO;2)

Eastman, J.R. and Fulk, M. (1993) Long Sequence Time Series Evaluation Using Standardised Principal Components. *Photogrammetric Engineering and Remote Sensing*, 59, 991-996.

Wilks, D.S. (1992) Adapting Stochastic Weather Generation Algorithms for Climate Change Studies. *Climatic Change*, 22, 67-84. <https://doi.org/10.1007/BF00143344>.

Piani, C., G. Weedon, M. Best, S. Gomes, P. Viterbo, S. Hagemann, and J. Haerter, Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models, *Journal of Hydrology*, 395 (3), 199–215, 2010b.

Xie, L., Lin, K., Wang, S., Wang, F., & Zhou, J. (2018). Differentially private generative adversarial network. *arXiv preprint arXiv:1802.06739*. <https://doi.org/10.48550/arXiv.1802.06739>.

Wang, Y., G. Sivandran, and J. M. Bielicki, 2018: The stationarity of two statistical downscaling methods for precipitation under different choices of cross-validation periods. *Int. J. Climatol.*, 38, e330–e348, <https://doi.org/10.1002/joc.5375>.

Rummukainen, M. (2010) State-of-the-Art with Regional Climate Models. *Climate Change*, 1, 96-82. <http://dx.doi.org/10.1002/wcc.8>.

Vandal, T., Kodra, E., Ganguly, S., Michaelis, A., Nemani, R., & Ganguly, A. R. (2020). DeepSD: Generating high resolution climate change projections through single image super-resolution. *arXiv preprint arXiv:1703.03126*. <https://doi.org/10.48550/arXiv.1703.03126>.

- 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. *International Journal of Climatology*, 39(1), 240–258. <https://doi.org/10.1002/joc.5897>.
- Adinolfi, M., Raffa, M., Reder, A., & Mercogliano, P. (2023). Investigation on potential and limitations of ERA5 Reanalysis downscaled on Italy by a convection-permitting model. *Climate Dynamics*, 61, 1-24. <https://doi.org/10.1007/s00382-023-06803-w>.
- Dubey A K, Jain V. A comparative study of relu and leaky-relu activation functions for convolutional neural networks [M]. *Applications of computing, automation and wireless systems in electrical engineering*. Springer, Singapore, 2019: 873-880.
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 1026-1034. <https://doi.org/10.1109/ICCV.2015.123>.
- Zhang, H., Goodfellow, I., Metaxas, D., & Odena, A. (2019). Self-attention generative adversarial networks. *Proceedings of the 36th International Conference on Machine Learning*, 97, 7354-7363. <https://doi.org/10.48550/arXiv.1906.06972>.
- 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*. <https://doi.org/10.48550/arXiv.1207.0580>.
- Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. <https://doi.org/10.48550/arXiv.1511.06434>.
- Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, Santamaría J, Fadhel MA, Al-Amidie M, Farhan L. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J Big Data*. 2021;8(1):53. doi: 10.1186/s40537-021-00444-8. Epub 2021 Mar 31. PMID: 33816053; PMCID: PMC8010506.
- Yu, X.; Liu, T.; Gong, M.; and Tao, D. 2018. Learning with biased complementary labels. In *ECCV (1)*, volume 11205 of *Lecture Notes in Computer Science*, 69–85. Springer. Zhou, Z.-H.; Zhang, M.-L.; Huang, S.-J.; and Li, Y.-F. 2012. Multi-instance multi-label learning. *Artificial Intelligence* 176(1):2291–2320.

- Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 4396-4405. <https://doi.org/10.1109/CVPR.2019.00453>.
- Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., & Chen, X. (2016). Improved techniques for training GANs. *arXiv preprint arXiv:1606.03498*. <https://arxiv.org/abs/1606.03498>.
- Bottou, L., & Bousquet, O. (2007). The tradeoffs of large scale learning. In *Advances in Neural Information Processing Systems 20* (pp. 161-168). Vancouver, British Columbia, Canada: Neural Information Processing Systems Foundation.
- Bejani, M., & Ghatee, M. (2021). A systematic review on overfitting control in shallow and deep neural networks. *Artificial Intelligence Review*, 54(9), 1-48. <https://doi.org/10.1007/s10462-021-09975-1>.
- Hoffer, E., Hubara, I., & Soudry, D. (2017). Train longer, generalize better: Closing the generalization gap in large batch training of neural networks. In *Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS 2017)*. Retrieved from <https://arxiv.org/abs/1705.08741>.
- Keskar, N. S., Mudigere, D., Nocedal, J., & Tang, P. (2016). On large-batch training for deep learning: Generalization gap and sharp minima. *arXiv preprint arXiv:1609.04836*. <https://doi.org/10.48550/arXiv.1609.04836>.
- Kornblith, S., Norouzi, M., Lee, H., & Hinton, G. (2019). Similarity of neural network representations revisited. In *Proceedings of the 36th International Conference on Machine Learning (ICML 2019)* (Vol. 97, pp. 3519-3528). PMLR. Retrieved from <https://proceedings.mlr.press/v97/kornblith19a.html>.
- Smith, L. N. (2018). A disciplined approach to neural network hyper-parameters: Part 1 — learning rate, batch size, momentum, and weight decay. *arXiv preprint arXiv:1803.09820v2*. <https://doi.org/10.48550/arXiv.1803.09820>.
- Hestness, J., Narang, S., Ardalani, N., Diamos, G., Jun, H., Kianinejad, H., Patwary, M. M. A., Yang, Y., & Zhou, Y. (2017). Deep learning scaling is predictable, empirically. *arXiv preprint arXiv:1712.00409*. <https://doi.org/10.48550/arXiv.1712.00409>.

Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. In Proceedings of the International Conference on Learning Representations (ICLR). Retrieved from <https://arxiv.org/abs/1412.6980>

Loshchilov, I., & Hutter, F. (2016). SGDR: Stochastic gradient descent with warm restarts. arXiv preprint arXiv:1608.03983. <https://api.semanticscholar.org/CorpusID:14337532>.

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/10.1145/3422622>.

PyTorch. (2023). PyTorch: An open source machine learning framework. Retrieved from <https://pytorch.org>.

Arjovsky, M., Chintala, S., & Bottou, L. (2017). Wasserstein GAN. In Proceedings of the 34th International Conference on Machine Learning (ICML 2017) (Vol. 70, pp. 214-223). PMLR. Retrieved from <http://proceedings.mlr.press/v70/arjovsky17a.html>.

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., & Courville, A. (2017). Improved training of Wasserstein GANs. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS 2017) (pp. 5769–5779). Curran Associates Inc.

Mescheder, L., Geiger, A., & Nowozin, S. (2018). Which training methods for GANs do actually converge? arXiv preprint arXiv:1801.04406v4. <https://doi.org/10.48550/arXiv.1801.04406>.

Obaid, H. S., Dheyab, S. A., & Sabry, S. S. (2019). The impact of data pre-processing techniques and dimensionality reduction on the accuracy of machine learning. In 2019 9th Annual Information Technology, Electromechanical Engineering and Microelectronics Conference (IEMECON) (pp. 279-283). IEEE. <https://doi.org/10.1109/IEMECONX.2019.8877011>.

Singh, D., & Singh, B. (2019). Investigating the impact of data normalization on classification performance. Applied Soft Computing, 82, Article 105524. <https://doi.org/10.1016/j.asoc.2019.105524>.

Leinonen, J., Nerini, D., & Berne, A. (2020). Stochastic super-resolution for downscaling time-evolving atmospheric fields with a generative adversarial network. IEEE Transactions on Geoscience and Remote Sensing, PP(99), 1-11. <https://doi.org/10.1109/TGRS.2020.3032790>.

Daizong Ding, Mi Zhang, Xudong Pan, Min Yang, and Xiangnan He. Modeling extreme events in time series prediction. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '19, pp. 1114–1122, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450362016. doi: 10.1145/3292500.3330896. URL <https://doi.org/10.1145/3292500.3330896>.

Allouche, M., Girard, S., & Gobet, E. (2022). EV-GAN: Simulation of extreme events with ReLU neural networks. *Journal of Machine Learning Research*, 23, 1-39.

Rasmussen, R. M., Chen, F., Liu, C. H., Ikeda, K., Prein, A., Kim, J., Schneider, T., Dai, A., Gochis, D., Dugger, A., Zhang, Y., Jaye, A., Dudhia, J., He, C., Harrold, M., Xue, L., Chen, S., Newman, A., Dougherty, E., Abolafia-Rosenzweig, R., Lybarger, N. D., Viger, R., Lesmes, D., Skalak, K., Brakebill, J., Cline, D., Dunne, K., Rasmussen, K., & Miguez-Macho, G. (2023). CONUS404: The NCAR–USGS 4-km long-term regional hydroclimate reanalysis over the CONUS. *Bulletin of the American Meteorological Society*, 104(8), E1382–E1408. <https://doi.org/10.1175/BAMS-D-21-0326.1>.

Li, G., & Cao, G. (2024). Generative adversarial models for extreme geospatial downscaling. arXiv preprint arXiv:2402.14049v2. <https://doi.org/10.48550/arXiv.2402.14049>.

Desiato, F., Lena, F., & Toreti, A. (2007). SCIA: A system for a better knowledge of the Italian climate. *Bollettino di Geofisica Teorica ed Applicata*, 48(3), 351-358. <https://doi.org/10.4430/bgta0024>.

Manco, I., De Lucia, C., Repola, F., Fedele, G., & Mercogliano, P. (2024). A comparative performance study of WRF, COSMO, and ICON atmospheric models for the Italian Peninsula at very high resolution. *Tethys*, 20, 1-20. <https://doi.org/10.3369/tethys.2023.20.01>.

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651-666. <https://doi.org/10.1016/j.patrec.2009.09.011>.

Xu, R., & Wunsch, D. I. I. (2005). Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, 16(5), 645-678. <https://doi.org/10.1109/TNN.2005.845141>.

Deka, P., & Saha, U. (2023). Introduction of k-means clustering into random cascade model for disaggregation of rainfall from daily to 1-hour resolution with improved preservation of extreme rainfall. *Journal of Hydrology*, [Volume], 129478. <https://doi.org/10.1016/j.jhydrol.2023.129478>.

- Arroyo Puente, Á., Tricio, V., Corchado Rodríguez, E., & Herrero Cosío, Á. (2015). A comparison of clustering techniques for meteorological analysis. In 10th International Conference on Soft Computing Models in Industrial and Environmental Applications. Springer. https://doi.org/10.1007/978-3-319-19719-7_11.
- Raffa, M., Adinolfi, M., Reder, A., Marras, G., Mancini, M., Scipione, G., Santini, M., & Mercogliano, P. (2023). Very high resolution projections over Italy under different CMIP5 IPCC scenarios. *Scientific Data*, 10, Article 21. <https://doi.org/10.1038/s41597-023-02144-9>.
- Hewitson, B. C. and Crane, R. G. (1996), 'Climate downscaling: techniques and application', *Clim. Res.* 7, 85–95.
- Giorgi, F. and Mearns, L. O. (1999), 'Introduction to special section: Regional Climate Modeling Revisited', *J. Geophys. Res.* D6, 6335–6352.
- Kingma, D. P., and J. Ba, 2014: Adam: A method for stochastic optimization. *arXiv*, 1412.6980v9, <https://doi.org/10.48550/arXiv.1412.6980>.