

POLYTECHNIQUE MONTRÉAL
affiliated with the Université de Montréal

Deep Learning for Precipitation Downscaling under Climate Change
THIS IS A TRANSLATED VERSION OF THE ORIGINAL THESIS (french)

PIERRE-LOUIS LEMAIRE
Department of mathematics and industrial engineering

Thesis submitted in partial fulfillment of the requirements for the degree of *Master of*
Applied Science
Mathematics

August 2025

POLYTECHNIQUE MONTRÉAL
affiliated with the Université de Montréal

This thesis entitled :

Deep Learning for Precipitation Downscaling under Climate Change
THIS IS A TRANSLATED VERSION OF THE ORIGINAL THESIS (french)

presented by **Pierre-Louis LEMAIRE**
in partial fulfillment of the requirements for the degree of *Master of Applied Science*
has been duly accepted by the examining committee consisting of:

Jonathan JALBERT, président
Julie CARREAU, membre et directrice de recherche
Youssef DIOUANE, membre et codirecteur de recherche
Richard LABIB, membre

DEDICATION

To automatic differentiation.

ACKNOWLEDGEMENTS

I would like to especially thank my research co-supervisors, Julie Carreau and Youssef Diouane, who guided and supported me through what felt like a bit of a blur at the beginning! Thank you for your supervision, your availability, and your advice. I have learned so much over the past two years, and you have helped shape a part of the person I will be in the future.

I formally thank Polytechnique Montréal and OURANOS for their trust and the financial support granted to this project.

This thesis would surely not have seen the light of day without the support and encouragement of those close to me. I want to deeply thank my family, and especially my parents, who were present throughout this work (despite the distance) and who were always by my side to listen, through moments of joy as well as fatigue.

I would also like to give a special thanks to Manille, who was by my side during almost the entire writing process, and whose purring greatly helped soften the workload!

Last but certainly not least, thank you my love for believing in me, for your unfailing support, and for your daily encouragement.

ABSTRACT

The acceleration of climate change is driving an intensification of extreme weather events, with major impacts on populations, infrastructure, and ecosystems. High-resolution climate projections are essential to anticipate these changes, but global climate models provide insufficient spatial resolution to capture fine-scale phenomena, leading to an underrepresentation of extremes, which are often highly localized. Regional climate models, while offering better resolution, remain costly and limit the production of large ensembles of high-resolution simulations.

Deep learning offers a promising alternative for the downscaling of climate variables. This thesis explores the use of a deep convolutional network (UNet) to perform downscaling of daily precipitation fields. Two main challenges are targeted: (1) the underestimation of high-intensity precipitation and (2) the degradation of performance under a non-stationary climate, as under climate change. These limitations hinder the practical application of such approaches. We hypothesize that they are partly due to the model's failure to learn the dependence of extreme precipitation on temperature changes.

Two main contributions are proposed. First, a soft constraint inspired by the Clausius-Clapeyron relations is added to the loss function to encourage the model to produce precipitation fields whose evolution with temperature remains consistent with the reference data. Second, two complementary approaches integrate priors from extreme value theory into the model. The non-stationarity linked to temperature evolution is thus encoded through the parametrization of extensions of the generalized Pareto distribution, improving the representation of extreme events under future climate.

Proposed approaches are evaluated on data originating from a large ensemble of regional climate simulations covering years 1955 to 2099. The results first highlight a clear dependence of precipitation extremes on mean temperature changes, while underscoring that deep learning models struggle to faithfully reproduce this relationship. They then demonstrate the potential of the proposed approaches to improve the prediction of extreme precipitation and open promising avenues for future research.

TABLE OF CONTENTS

DEDICATION	iii
ACKNOWLEDGEMENTS	iv
ABSTRACT	v
LIST OF TABLES	viii
LIST OF FIGURES	ix
LIST OF SYMBOLS AND ACRONYMS	xiii
LIST OF APPENDICES	xiv
CHAPTER 1 INTRODUCTION	1
CHAPTER 2 THEORETICAL FRAMEWORK	4
2.1 Deep Learning	4
2.1.1 Deep Convolutional Neural Networks	4
2.1.2 UNet	6
2.2 Extreme Value Theory	7
2.2.1 Block Maxima Approach	7
2.2.2 Peaks Over Threshold Approach	9
2.2.3 Extension of the Generalized Pareto Distribution	10
2.3 Relationship Between Precipitation and Temperature	12
2.3.1 Clausius-Clapeyron	12
2.3.2 Estimation by Binning Method	12
2.3.3 Estimation via Generalized Extreme Value Distribution	13
CHAPTER 3 LITERATURE REVIEW	15
3.1 Numerical Climate Modeling	15
3.2 Dynamic and Statistical Downscaling	16
3.3 Deep Learning for Downscaling	17
3.3.1 Deterministic Methods	18
3.3.2 Generative Methods	19
3.4 Physics-Informed Deep Learning	20

CHAPTER 4 DATA AND EXPLORATORY ANALYSIS	22
4.1 ClimEx: A Large Ensemble of Climate Simulations under Global Warming	22
4.2 Exploratory Statistical Analysis	23
CHAPTER 5 DEEP LEARNING AND PRECIPITATION DOWNSCALING UNDER CLIMATE CHANGE	27
5.1 UNet for Downscaling	27
5.1.1 Architecture Presentation	27
5.1.2 Strategy, Preprocessing, and Post-processing	29
5.2 UNet Improvements	32
5.3 Evaluation under Climate Change	35
5.3.1 TPSR Estimation via Generalized Extreme Value Distribution	37
CHAPTER 6 INTEGRATING THE TEMPERATURE-PRECIPITATION RELATIONSHIP FOR DOWNSCALING	42
6.1 Soft Clausius-Clapeyron Constraint via Binning Approach	42
6.1.1 Integrating the Precipitation-Temperature Relationship via Soft Constraints	42
6.1.2 Limitations of the Binning Approach	44
6.2 Integrating Statistical Priors from Extreme Value Theory	44
6.2.1 Modeling Precipitation via the Extended Generalized Pareto Distribution	45
6.2.2 Downscaling via Quantile Level Prediction	49
6.2.3 Learning via Quantile Level Alignment	53
CHAPTER 7 EVALUATION	54
7.1 Metrics	54
7.2 Results	56
7.2.1 Soft Constraint via Binning Approach	56
7.2.2 Integrating Statistical Priors from Extreme Value Theory	60
CHAPTER 8 CONCLUSION	65
8.1 Synthesis of Work	65
8.2 Research Perspectives	66
REFERENCES	67
APPENDICES	80

LIST OF TABLES

Table 5.1	MAE (in mm/day), proportions, and ℓ_2 norm of negative values for no transformation, T_1 , or T_2 . UNet trained on the 1960–2000 period and evaluated on 2000–2010.	31
Table 5.2	UNet trained on the 1955–2010 period and evaluated on the 2010–2020 period over 50 epochs. Comparison of performance with or without the post-processing constraint (5.8). Overall MAE (in mm/day) and MAE calculated in the vicinity of specific quantiles (0.5, 0.90, 0.95, and 0.99).	32
Table 5.3	MAE for different sequences of blocks per encoder level (symmetric decoder with one additional block per level). UNet trained for 50 epochs with 64 samples per mini-batch.	33
Table 5.4	MAE for different convolution types and initial channel numbers. UNet trained for 50 epochs.	34
Table 5.5	MAE for different convolution filter sizes. UNet trained for 50 epochs.	34
Table 5.6	MAE with certain residual block layers removed: (1) only SiLU, and (2) SiLU and Group Normalization (GN). UNet trained for 50 epochs.	34
Table 5.7	Mean negative log-likelihood of the GEV distributions over the domain for the 2020–2099 period for all combinations of $\mu(T)$ and $\sigma(T)$ parameter models.	39
Table 5.8	Mean TPSRs estimated at several quantile levels for reference data and UNet predictions for the 2020–2099 test period.	40
Table 7.1	Overall MAE and MAE evaluated in the vicinity of several quantile levels for bicubic interpolation over the test period.	56
Table 7.2	Overall MAE and MAE evaluated in the vicinity of several quantile levels, as a function of the coefficient γ of the loss function (6.1), for $s = 0.0$ mm/day (filtering threshold (6.4)).	57
Table 7.3	Overall MAE and MAE evaluated in the vicinity of several quantile levels, as a function of the coefficient γ of the loss function (6.1), for $s = 1.0$ mm/day (filtering threshold (6.4)).	57
Table 7.4	Overall MAE (in mm/day) and MAE evaluated in the vicinity of several quantile levels for the standard UNet $\mathcal{L}_{\text{reco}}$ and the ExtGPD quantile level prediction and alignment methods: $\mathcal{L}_{\text{ExtGPD}_1}$ (1) and $\mathcal{L}_{\text{ExtGPD}_2}$ (2), with or without temperature anomalies T	61

LIST OF FIGURES

Figure 1.1	Illustration of the theoretical framework for precipitation downscaling using deep learning.	2
Figure 2.1	Illustration of the receptive field of a cell (Layer 3 - orange), which corresponds to the portion of the input to which it is sensitive. As one progresses through the network layers, the receptive field expands (red in Layer 2, then orange in Layer 3), allowing neurons to capture higher-level features. In this example, the filters are of dimensions 3×3 , and the stride (the displacement of the convolution filter in each direction) is 1. The receptive field r_i of layer i , for a filter of dimension k_i and the stride of the previous layer s_{i-1} , is equal to: $r_i = r_{i-1} + (k_i - 1) \cdot s_{i-1}$	5
Figure 2.2	UNet composed of 3 levels, featuring transposed convolution or upsampling and convolution. The input and output dimensions are $64 \times 64 \times 3$, and the latent space is $16 \times 16 \times 256$	7
Figure 2.3	Probability densities for different values of ξ corresponding to the three families included in the GEV, with $\mu = 0$ and $\sigma = 1$	9
Figure 2.4	Densities of model (1) for "small" and "moderate" values (left) and "large" values (right), as a function of κ . The GPD shape is recovered when $\kappa = 1$. The Gamma distribution with parameters (1.4, 1.4) is shown for comparison. Figure inspired by [1].	11
Figure 4.1	Let Period A range from 1955 to 2030, and Period B be a five-year overlapping period from 2025 to 2099. (a) Absolute difference between Periods B and A of the median daily mean temperatures over the ClimEx domain; the black dotted line indicates the subdomain used in this thesis. (b) Median non-zero daily precipitation over Period A . (c) 95th quantile of non-zero daily precipitation over Period A . (d) Relative difference in median non-zero precipitation. (e) Relative difference in the 95th quantile of non-zero precipitation.	23
Figure 4.2	To simulate low-resolution (LR) observations, high-resolution (HR) daily precipitation from ClimEx is aggregated by averaging 16×16 cell blocks (256 cells). (Left) Histograms of daily precipitation, separated by intensity (0–5 mm/day top, and ≥ 5 mm/day bottom), comparing Periods A and B at high and low resolution. (Right) A randomly selected daily precipitation field (top) and its aggregated version (bottom).	25

Figure 5.1	Presentation of the UNet model architecture used, and modifications made to the original residual block [2].	28
Figure 5.2	Histograms of precipitation predicted by the UNet with or without the constraint (5.8) over the evaluation period (2010–2020).	33
Figure 5.3	(Top) Mean annual MAE (in mm/day) of the UNet over the test period. (Bottom) Relative evolution (in %) of the annual MAE compared to the 2020–2025 period, calculated in the vicinity of quantile levels 0.5, 0.75, 0.90, 0.95, and 0.99. Curves obtained by a 10-value moving average.	36
Figure 5.4	Comparison of the evolution of the Mean Absolute Error (MAE) of the UNet’s high-resolution precipitation predictions between the 2020–2059 and 2060–2099 periods relative to the temperature evolution over these same periods.	37
Figure 5.5	Comparison of TPSRs for the $q = 0.99$ quantile level estimated over the 2020–2099 test period from reference data and UNet predictions.	40
Figure 5.6	Box plots of TPSRs estimated at several quantiles for reference data and UNet predictions for the 2020–2099 test period.	41
Figure 6.1	\mathcal{L}_{CC} for the 1955–2020 period on a restricted domain as a function of increasing mini-batch sizes (256, 512, 1024, 2048, and 4096).	45
Figure 6.2	Negative log-likelihood obtained after optimizing the parameters defined in (6.22–6.24) for the ExtGPD distribution, fitted independently for each cell in the domain using reference data covering the 1955–2020 period. The numbered stars 1 to 4 identify the cells analyzed in Figures 6.5 and 6.6.	48
Figure 6.3	Values of the $\exp(\alpha_0)$, α_1 , $\exp(\beta_0)$, β_1 parameters for each cell in the domain after maximizing the log-likelihood (6.17).	50
Figure 6.4	Values of the ξ parameter after applying transformation (6.24).	51
Figure 6.5	(First row) Q-Q plots of data from the 2020–2099 period for cells 1 to 4 (see Figure 6.2) for the stationary models ($\alpha_1 = \beta_1 = 0$) fitted over the 1955–2020 period. (Second row) Histograms of the resulting quantiles.	51
Figure 6.6	(First row) Q-Q plots of data from the 2020–2099 period for cells 1 to 4 (see Figure 6.2) for the non-stationary models fitted over the 1955–2020 period. (Second row) Histograms of the resulting quantiles.	52

Figure 7.1	Annual MAE (in mm/day) for MSE loss ($\gamma = 0$), and for $s = 0$ or $s = 1$ mm/day with $\gamma = 10$. Bicubic interpolation error is not displayed to maintain clarity. Smoothing via a 10-value moving average.	58
Figure 7.2	Box plots of TPSRs (in %/ $^{\circ}C$) obtained via the method defined in Section 6.1 for each model. The test set is split into two periods: 2020–2050 and 2050–2099, to evaluate temporal evolution.	58
Figure 7.3	Precipitation histogram for each model, including all domain cells over the 2020–2099 test set period.	59
Figure 7.4	PSD_{rel} for each model relative to the reference data power spectral density, calculated according to the method detailed in Section 7.1. . .	60
Figure 7.5	Annual MAE (in mm/day) for the standard UNet (\mathcal{L}_{reco}) and the stationary and non-stationary variants of the quantile level prediction method (see Section 6.2.2) of the previously fitted ExtGPDs: \mathcal{L}_{ExtGPD_1} and $\mathcal{L}_{ExtGPD_1}(T)$. The curve is obtained via a 10-value moving average.	61
Figure 7.6	Annual MAE (in mm/day) for the standard UNet (\mathcal{L}_{reco}) and the stationary and non-stationary variants of the quantile level alignment method (see Section 6.2.3): \mathcal{L}_{ExtGPD_1} and $\mathcal{L}_{ExtGPD_1}(T)$. The curve is obtained via a 10-value moving average.	62
Figure 7.7	Precipitation histogram for each model, including all domain cells over the 2020–2099 test period.	63
Figure 7.8	PSD_{rel} for each model relative to the reference data PSD, calculated according to the method detailed in Section 7.1.	63
Figure 7.9	Low-resolution precipitation fields (1nn interpolation), interpolated by the bicubic method, predicted by each of the studied models, and high-resolution fields from ClimEx for two dates randomly selected from the test period.	64
Figure A.1	Negative log-likelihood following parameter optimization under the $M_{\mu,4} + M_{\sigma,2}$ combination for the reference data (left) and UNet predictions (right).	80
Figure A.2	μ_0 parameters of $M_{\mu,4}$ for the reference data (left) and UNet predictions (right).	80
Figure A.3	μ_1 parameters of $M_{\mu,4}$ for the reference data (left) and UNet predictions (right).	81
Figure A.4	σ_0 parameters of $M_{\sigma,2}$ for the reference data (left) and UNet predictions (right).	81

Figure A.5	σ_1 parameters of $M_{\sigma,2}$ for the reference data (left) and UNet predictions (right).	82
Figure A.6	ξ parameters for the reference data (left) and UNet predictions (right).	82

LIST OF SYMBOLS AND ACRONYMS

GCM	<i>Global Climate Model</i>
ESM	<i>Earth System Model</i>
RCM	<i>Regional Climate Model</i>
CMIP6	<i>Coupled Model Intercomparaison Project Phase 6</i>
CORDEX	<i>Coordinated Regional Climate Dowscaling Experiment</i>
GHG	<i>Greenhouse Gas</i>
TPSR	<i>Temperature-Precipitation Scaling Rates</i>
CNN	<i>Convolutional Neural Network</i>
GAN	<i>Generative Adversarial Network</i>
GEV	<i>Generalized Extreme Value distribution</i>
GPD	<i>Generalized Pareto Distribution</i>
ExtGPD	<i>Extended Generalized Pareto Distribution</i>

LIST OF APPENDICES

Appendix A Estimation of TPSRs via the Generalized Extreme Value Distribution 80

CHAPTER 1 INTRODUCTION

The acceleration of climate change is increasing the vulnerability of populations, ecosystems, and human infrastructure. It is accompanied by profound shifts in climate regimes, characterized notably by the intensification and increased frequency of extreme events such as heatwaves, floods, and intense precipitation. In particular, precipitation distribution could be significantly altered depending on the region, with some cases showing a decrease in mean annual precipitation alongside an increase in the frequency and intensity of extreme events.

To better anticipate these developments, it is essential to produce fine-scale climate projections. This allows for the modeling of strong disparities in certain atmospheric variables within "complex" zones, such as regions with rugged orography, coastal areas, or those subject to intense winds. Continuous improvements in climate models over the last fifty years have enabled the generation of global projections with a spatial resolution of approximately $\approx 50 - 100$ kilometers (km). However, this scale remains insufficient to simulate certain fine-scale processes that are nonetheless crucial for forecasting extreme events. Regional climate models (RCMs) provide a partial solution, offering a spatial resolution of about ≈ 10 km for the most advanced models, which significantly improves the representation of precipitation, particularly for high intensities. However, the high computational cost of these models limits the number of achievable simulations, especially when constructing large projection ensembles, which are indispensable for correctly representing natural climate variability.

The growing volume of atmospheric observations (from ground stations, balloons, buoys, and satellites) and numerical simulations, coupled with recent progress in deep learning, has opened new horizons in numerical climate modeling through the rise of data-driven approaches. Initially limited to relatively simple tasks, neural network-based methods are now capable of emulating global numerical climate models with comparable performance while drastically reducing computational costs. Furthermore, the use of deep learning for refining the spatial resolution of climate simulations has expanded significantly in recent years, offering a promising alternative to traditional downscaling approaches.

In this thesis, we focus on increasing the spatial resolution (also known as "downscaling") of precipitation from climate simulations using neural networks. This task, illustrated in Figure 1.1, consists of estimating the unknown downscaling function h , which links low-resolution precipitation fields to high-resolution ones, using a neural network f_θ parameterized by a set of weights θ . More specifically, we consider deep convolutional networks, which exploit the discrete convolution operation to efficiently process multidimensional data (*e.g.*, medical

imagery or atmospheric variable fields). Despite their potential, neural networks exhibit several well-known limitations that restrict their application in real-world contexts. On one hand, due to phenomena such as shortcut learning and overfitting, deep learning models can encode artificial or physically inconsistent relationships that do not hold outside their training domain and period—particularly in the context of climate change. On the other hand, these models struggle to correctly represent extreme values due to their bias toward low-frequency signals.

In the context of downscaling precipitation fields, the consequences of these limitations are well-documented and confirmed by our experiments:

1. A degradation in performance as the evaluation period moves further away from the training period, driven by the intensification of climate change;
2. High-resolution precipitation fields exhibiting blurry and poorly localized structures, as deep learning models tend to converge toward average solutions to minimize error at the expense of representing extreme events.

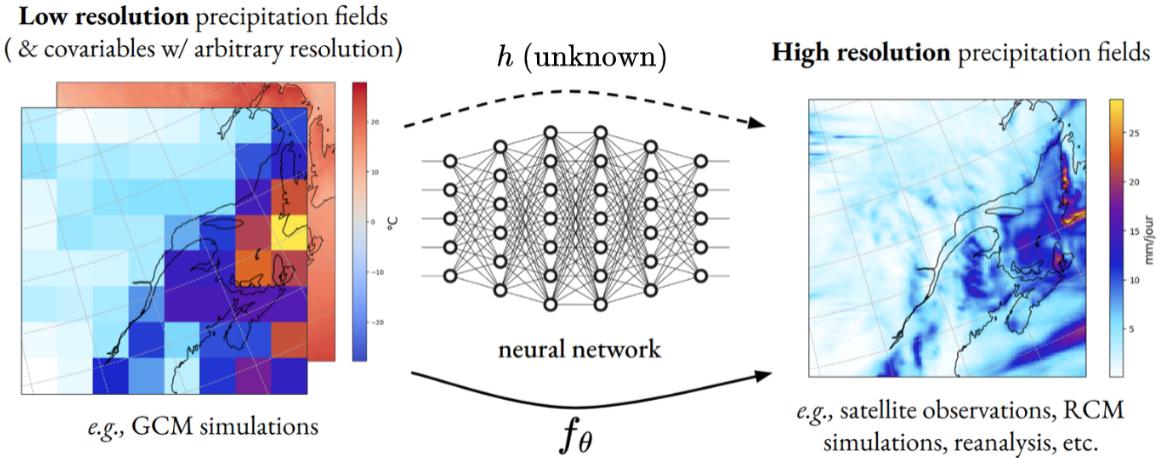


Figure 1.1 Illustration of the theoretical framework for precipitation downscaling using deep learning.

The central objective of this thesis is to mitigate these two primary limitations. We have chosen to limit the model size and fix a specific architecture *a priori* in order to focus the bulk of our efforts on developing methods that can be adapted to any type of deep learning model applied to precipitation field downscaling. To achieve this, we introduce concepts from atmospheric sciences, such as the Clausius-Clapeyron relation—which quantifies the change in saturation vapor pressure of a gas (*e.g.*, water vapor in the atmosphere) relative to

temperature—to integrate the dependence of precipitation extremes on temperature changes under climate change into the model. We also leverage methods from extreme value theory to incorporate statistical priors on precipitation distribution during model training, aiming to improve the representation of extremes while accounting for temperature-related non-stationarity.

This thesis is structured as follows: Section 2 presents the theoretical framework and the methods employed, followed by a literature review in Section 3. Section 4 describes the dataset used to evaluate the developed methods and provides an exploratory analysis to highlight specific characteristics. The deep learning model architecture, data processing, and performance evaluation under climate change are detailed in Section 5. Section 6 develops the approaches introduced to better represent extremes and integrate temperature dependence. The results of the proposed methods are presented and analyzed in Section 7. Finally, Section 8 summarizes the main contributions of this thesis, discusses their limitations, and proposes new research perspectives.

CHAPTER 2 THEORETICAL FRAMEWORK

This chapter aims to introduce the theoretical foundations of the methods and approaches used throughout this thesis. Since the methodological core of the presented work belongs to deep learning, we present the parent family of the model used: Deep Convolutional Neural Networks (CNNs), as well as the specific model chosen for our project. An overview of Extreme Value Theory (EVT) will then establish the theoretical basis for some of the downscaling strategies developed in this research. Finally, we introduce the Clausius-Clapeyron relation and the various methods used to estimate the scaling rates between temperature and precipitation. This will serve as a basis for evaluating the ability of deep learning models to learn physically consistent features to improve their robustness to the changes introduced by global warming.

2.1 Deep Learning

Deep learning is a branch of artificial intelligence that involves training networks with multiple layers of parameters using the backpropagation algorithm. These networks learn representations of a dataset and use these representations to solve tasks such as image recognition or text translation [3,4].

2.1.1 Deep Convolutional Neural Networks

Deep Convolutional Neural Networks (CNNs) constitute an important family of deep learning models designed primarily to process spatial data. Their structure consists of a succession of convolutional layers, each composed of one or more filters (or kernels) that extract characteristic spatial patterns from the data through the discrete convolution operation.

Consider a set $C \in \mathbb{N}_+^*$ of two-dimensional fields: $x \in \mathbb{R}^{C \times H \times W}$, with spatial dimensions $H \in \mathbb{N}_+^*$ and $W \in \mathbb{N}_+^*$, and a filter $k \in \mathbb{R}^{h \times w}$ of height $h \in \mathbb{N}_+^*$ and width $w \in \mathbb{N}_+^*$ where $h \ll H$ and $w \ll W$. The value of cell (i, j) in the two-dimensional field y , resulting from the application of the discrete convolution operation to the input fields x , can be expressed by the formula:

$$y_{i,j} = \sum_{c=1}^C \sum_{u=0}^h \sum_{v=0}^w k_{i+u,j+v} x_{c,i+u,j+v} \quad (2.1)$$

Since filter dimensions are generally much smaller than those of the data ($h \ll H$, and $w \ll W$), the number of parameters per layer remains low compared to a traditional dense (fully

connected) layer, where each input variable is connected to each output variable by a distinct parameter. This allows for the construction of deep models with many layers and multiple filters per layer. A CNN model, consisting of a succession of convolutional layers separated by pooling operators and non-linearities, possesses the capacity to learn complex structures representing its training dataset by exploiting both local and global connections within the data, as illustrated in Figure 2.1.

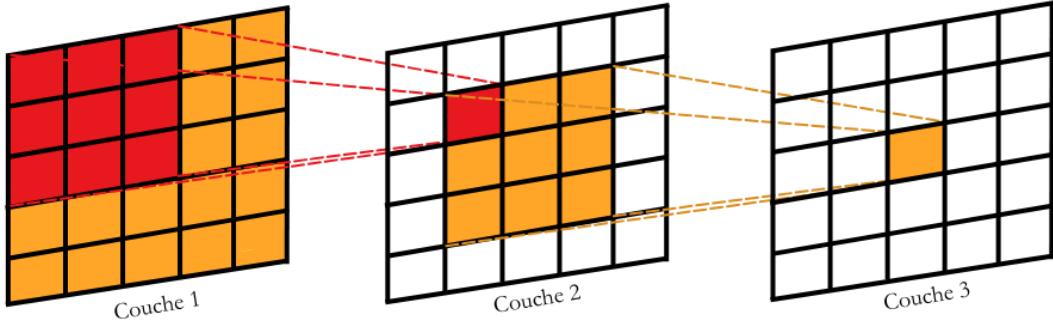


Figure 2.1 Illustration of the receptive field of a cell (Layer 3 - orange), which corresponds to the portion of the input to which it is sensitive. As one progresses through the network layers, the receptive field expands (red in Layer 2, then orange in Layer 3), allowing neurons to capture higher-level features. In this example, the filters are of dimensions 3×3 , and the stride (the displacement of the convolution filter in each direction) is 1. The receptive field r_i of layer i , for a filter of dimension k_i and the stride of the previous layer s_{i-1} , is equal to: $r_i = r_{i-1} + (k_i - 1) \cdot s_{i-1}$.

Non-linearity, inherent in complex data, is accounted for by adding activation functions between convolutional layers, such as $\text{ReLU}(z) = \max(0, z)$ or $\text{SiLU}(z) = z \cdot \text{sig}(z)$, where $\text{sig}(x) = 1/(1 + e^{-x})$ is the sigmoid function. Other operators, notably max-pooling, also help model the non-linearity of the data.

The parameters of the convolutional layers are estimated by stochastic gradient descent (SGD) via backpropagation. Let f_θ be the CNN parameterized by weights $\theta \in \mathbb{R}^d$, grouped into a vector of dimension d , intended to model $h : x \rightarrow y$, where x and y are real-valued tensors of arbitrary dimensions. The learning of parameters is performed by minimizing a cost function \mathcal{L} according to the rule:

$$\theta_{i+1} = \theta_i - \eta \cdot \nabla_{\theta_i} \mathcal{L}(f_{\theta_i}(x^{(ni:n(i+1))}), y^{(ni:n(i+1))}) \quad (2.2)$$

where η is the learning rate, and $x^{(ni:n(i+1))} \subset x$, $y^{(ni:n(i+1))} \subset y$ are the i -th mini-batches of data, each containing n samples. Since the parameters η and n are not learned, they are fixed manually or via hyperparameter optimization algorithms. The cost function plays a central

role in deep learning, as it defines the space in which the parameters θ are optimized, thus strongly influencing the performance and behavior of the model after training. The impact of the cost function for downscaling is explored in more detail in Section 3.3.1.

CNNs were first developed and used for image recognition, using pooling operators to reduce spatial dimensions while retaining the dominant features of the data [5]. They were quickly adapted to so-called *image-to-image* tasks, particularly in medical imaging, which require an output for every input pixel. It was in this context that the UNet architecture was introduced [6].

2.1.2 UNet

Initially designed for medical image segmentation, this architecture has established itself across multiple applications as a benchmark deep convolutional model. The UNet is composed of two symmetrical paths—contracting and expansive—functioning as an encoder/decoder:

- **The Encoder:** acts as a feature extractor $E : x \rightarrow z$. Containing L levels, E transforms a large-dimensional input $x \in \mathbb{R}^{C \times H \times W}$ into a contracted latent representation $z \in \mathbb{R}^{D \times H/2^L \times W/2^L}$, which is reduced in space but rich in features, also known as channels ($D \in \mathbb{R}_+^* \gg C$). Each level consists of at least one convolutional layer followed by an activation function and a pooling operation (average or max) to reduce spatial resolution.
- **The Decoder:** allows for the reconstruction of details in the output $y \in \mathbb{R}^{C \times H \times W}$ from the intermediate features $E_l(x) \in \mathbb{R}^{d_l \times H/2^l \times W/2^l}$ originating from level $l \in [|1, L|]$ of the encoder (with $C < d_l \in \mathbb{R}_+^* < D$ channels) and the final encoder output: $D : (z, E_{1:L-1}(x)) \rightarrow y$. Each level uses a transposed convolutional layer (or an upsampling operator followed by a convolutional layer) to increase spatial resolution, integrating the features learned at each level of the encoder via skip connections. As with the encoder, an activation function is applied after each convolutional layer.

Skip connections are crucial for the UNet, as they preserve fine spatial information that would otherwise be lost during the spatial resolution reduction when contracting toward the latent space. Figure 2.2 represents a standard 3-level UNet architecture.

Numerous improvements to the initial architecture have been proposed, such as the integration of residual blocks, the addition of dense connections between levels, the use of attention mechanisms to weight image regions, or combinations of these enhancements [7–9]. In this

thesis, we will use a residual UNet model from the literature, described in more detail in Section 5.1.1.

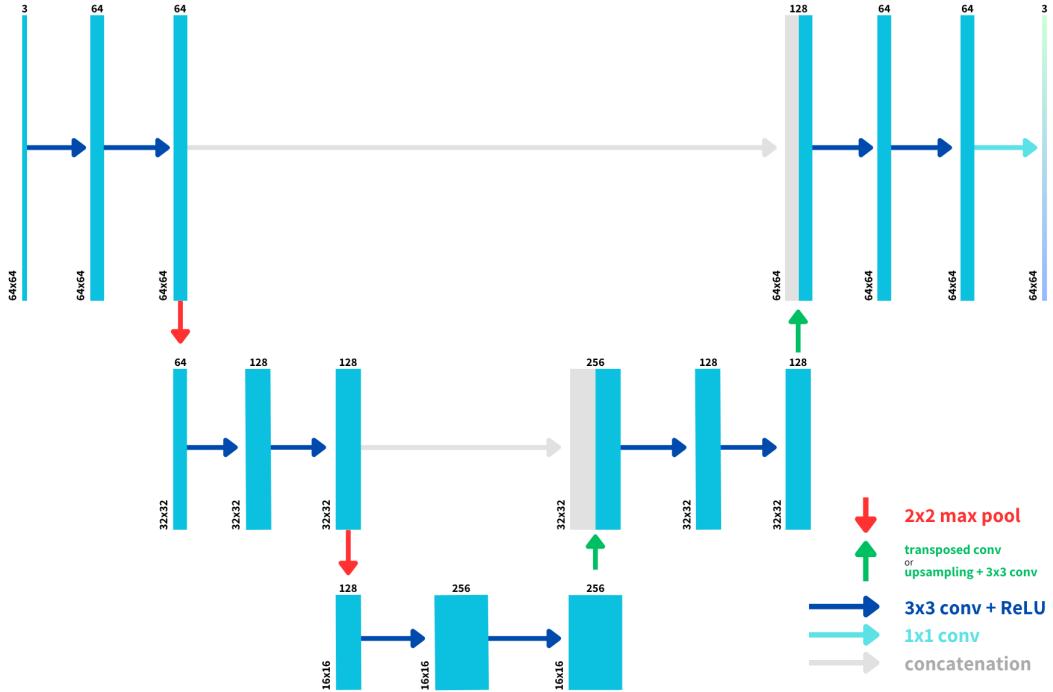


Figure 2.2 UNet composed of 3 levels, featuring transposed convolution or upsampling and convolution. The input and output dimensions are $64 \times 64 \times 3$, and the latent space is $16 \times 16 \times 256$.

2.2 Extreme Value Theory

Extreme Value Theory (EVT) is a branch of probability and statistics specifically concerned with rare events and the tails of distributions. This theory finds numerous applications in diverse fields such as meteorology and finance [10].

2.2.1 Block Maxima Approach

Consider a sequence of independent random variables X_1, \dots, X_n (corresponding to a "block") following the same cumulative distribution function F . The generalized extreme value distribution focuses on the behavior of the variable:

$$M_n = \max\{X_1, \dots, X_n\}. \quad (2.3)$$

Theoretically, it is possible to obtain the density of the random variable M_n using the independence of the variables $(X_i)_{[1,n]}$:

$$\mathbb{P}(M_n \leq z) = \mathbb{P}(X_1 \leq z, \dots, X_n \leq z) = \prod_{i=1}^n \mathbb{P}(X_i \leq z) = \{F(z)\}^n. \quad (2.4)$$

However, in practice, F is unknown, and tiny estimation errors (which are inevitable) can lead to large errors when n is large. Another approach is to directly approximate the distribution F^n as $n \rightarrow \infty$.

An important property to note is that for any $z < z_+ = \inf\{z \in \mathbb{R} \mid F(z) = 1\}$, we have $\{F(z)\}^n \rightarrow 0$ as $n \rightarrow \infty$. In other words, the probability density of the variable M_n asymptotically concentrates in the upper neighborhood of the bound z_+ . It is therefore necessary to perform a renormalization of M_n to stabilize its location and variance as n increases. Theorem 1 provides, under the renormalization hypothesis, a family of distributions for modeling F^n for sufficiently large n .

Theorem 1 (Fisher-Tippett-Gnedenko Theorem) *Let X_1, \dots, X_n be a sequence of independent and identically distributed random variables, and $M_n = \max\{X_1, \dots, X_n\}$. For two appropriate sequences $\{a_n > 0\}$ and $\{b_n\}$ such that:*

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq z\right) \rightarrow G(z) \quad \text{as } n \rightarrow \infty \quad (2.5)$$

for a non-degenerate distribution function G , then G is of the form:

$$G(z) = \begin{cases} \exp\{-[1 + \xi(\frac{z-\mu}{\sigma})]^{-1/\xi}\} & \text{if } \xi \neq 0 \\ \exp\{-\exp(\frac{z-\mu}{\sigma})\} & \text{if } \xi = 0 \end{cases} \quad (2.6)$$

defined for $\{z \in \mathbb{R} \mid 1 + \xi(z - \mu)/\sigma > 0\}$, with $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}_+^*$, and $\xi \in \mathbb{R}$.

In this form, G encompasses three families of probability distributions: Gumbel ($\xi = 0$), Fréchet ($\xi > 0$), and Weibull ($\xi < 0$), thus offering increased modeling flexibility. This family is known by the acronym GEV (Generalized Extreme Value).

Generally, the parameters μ , σ , and ξ are respectively called the location, scale, and shape parameters. Figure 2.3 shows the role of the shape parameter ξ , which controls the tail weight and behavior of G .

In practice, the unknown normalization constants $\{a_n > 0\}$ and $\{b_n\}$ do not pose difficulties

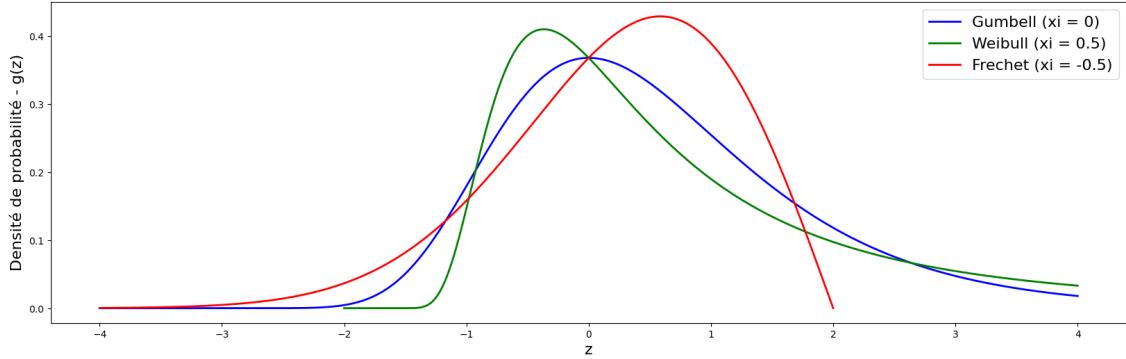


Figure 2.3 Probability densities for different values of ξ corresponding to the three families included in the GEV, with $\mu = 0$ and $\sigma = 1$.

because they can be absorbed by G . For sufficiently large n , and under the hypothesis that:

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq z\right) \approx G(z) \quad \text{then} \quad \mathbb{P}(M_n \leq z) \approx G\left(\frac{z - b_n}{a_n}\right) = G^*(z) \quad (2.7)$$

where G^* also belongs to the GEV family. Therefore, it is not necessary to consider $\{a_n > 0\}$ and $\{b_n\}$ in practice.

Finally, an important consideration should be noted regarding the values of ξ . The asymptotic properties necessary for maximum likelihood estimation are not satisfied for $\xi \leq -0.5$, notably because the Fisher information is no longer defined or tends toward infinity [11].

2.2.2 Peaks Over Threshold Approach

Modeling extreme values using the GEV has the disadvantage of severely restricting the number of data points used, while other extreme values may be available, without guaranteeing that the selected values are themselves extreme. Unlike block maxima modeling, the Peaks Over Threshold (POT) approach allows for the potential utilization of larger data series.

Theorem 2 (Pickands-Balkema-De Haan Theorem) *Let X_1, \dots, X_n be a sequence of independent and identically distributed random variables according to F . Under the same conditions as Theorem 1, i.e., $\lim_{n \rightarrow \infty} F^n(z) = G(z; \mu, \sigma, \xi)$, then for a sufficiently large threshold u , the distribution function of the variable $Y = X - u$, conditional on $X > u$, is given by:*

$$\lim_{u \rightarrow z+} \mathbb{P}(Y = X - u \leq y \mid X > u) = H(y) = \begin{cases} 1 - \left(1 + \xi \frac{y}{\sigma}\right)^{-\frac{1}{\xi}} & \text{if } \xi \neq 0 \\ 1 - \exp\left(-\frac{y}{\sigma}\right) & \text{if } \xi = 0 \end{cases} \quad (2.8)$$

defined for $\{y \in \mathbb{R} \mid 1 + \xi y / \tilde{\sigma} > 0\}$ and $\tilde{\sigma} = \sigma + \xi(u - \mu) > 0$, with $z+$ being the upper bound of the domain of X .

The family of distributions defined by Equation (2.8) is called the Generalized Pareto Distribution (GPD) family.

2.2.3 Extension of the Generalized Pareto Distribution

The extension of the Generalized Pareto Distribution aims to define a distribution capable of modeling the entire range of non-zero precipitation (low, moderate, and extreme) while preserving the asymptotic properties of extreme value theory in the tail of the distribution [1].

An interesting property of extreme value theory is its ability to also model low-intensity values. Take precipitation series as an example, which are lower-bounded by 0, and let us invert their sign to construct the variable $Y = -X$. The largest values of Y can then be modeled by a Generalized Pareto Distribution with a negative shape parameter ξ , corresponding to the Weibull case for the GEV (see Figure 2.3).

Let us fix a threshold v defining the boundary for low values of X . We denote $\xi = -1/\kappa$ with $\kappa > 0$. Following the previous hypothesis, we can find the form of the distribution for low values of X [1]:

$$\mathbb{P}(X \leq x \mid X \leq v) = \mathbb{P}(Y > -x \mid Y > -v) \approx \bar{H}_{-1/\kappa} \left(\frac{-x + v}{\sigma} \right) = \left(1 - \frac{1}{\kappa} \frac{-x + v}{\sigma} \right)^\kappa,$$

the condition $\bar{H}_{-1/\kappa}(0) = 0$, derived from the constraint imposed by the upper bound of Y on the survival function $\bar{H}_{-1/\kappa}$, also implies that $v = \kappa\sigma$, since $\bar{H}_{-1/\kappa}(0) = (1 - \frac{v}{\kappa\sigma})^\kappa = 0 \implies v = \kappa\sigma$. Consequently, low values of X can be approximately described by a power law, and it is possible to model the entire range of precipitation, excluding 0, by a law of the form:

$$\mathbb{P}(X \leq x) \approx \begin{cases} 1 - \text{cst} \times \bar{H}_\xi \left(\frac{x}{\sigma} \right) & \text{for "large" } x, \\ 1 - \text{cst} \times x^\kappa & \text{for "small" } x. \end{cases}$$

The distinction between "small" and "large" values implies the choice of a threshold, which considerably limits the flexibility of the method. However, the method aims to model the entire precipitation distribution in a simple and numerically efficient manner. To introduce the core of their approach, the authors use as a basis a method that extends the modeling capabilities of the Generalized Pareto Distribution [12]. Knowing that it is possible to sample a GPD by $\sigma H_\xi^{-1}(u)$ with $u \sim \mathcal{U}(0, 1)$, one can increase sampling flexibility by transforming

u by the inverse of a function K that shares the properties of a distribution function defined on $[0, 1]$. Thus, we have:

$$X = \sigma H_\xi^{-1}\{K^{-1}(u)\}. \quad (2.9)$$

To ensure that the tail of the distribution of X behaves in accordance with a GPD with ξ as the shape parameter, and that values near zero follow a power law, the authors impose three constraints that the function K must respect. Four families of parametric functions satisfying these constraints have been proposed:

1. $K(v) = v^\kappa$ with $\kappa > 0$;
2. $K(v) = pv^{\kappa_1} + (1-p)v^{\kappa_2}$ with $\kappa_2 \geq \kappa_1 > 0$ and $p \in [0, 1]$;
3. $K(v) = 1 - Q_\delta\{(1-v)^\delta\}$ with $\delta > 0$ and Q_δ the distribution function of a random variable $\sim \beta(1/\delta, 2)$;
4. $K(v) = [1 - Q_\delta\{(1-v)^\delta\}]^{\kappa/2}$, with $\kappa, \delta > 0$ and Q_δ defined previously.

For the sake of simplicity, and given its satisfactory performance in describing precipitation in several studies, we will limit ourselves to the use of model (1) in this thesis [1, 13, 14]. Figure 2.4 illustrates its behavior for several values of κ and provides a comparison with the Gamma distribution, which is particularly relevant for large values.

It is important to note that more complex methods extending the capabilities of the EGPD for precipitation have been proposed, but we leave their use for future work [15].

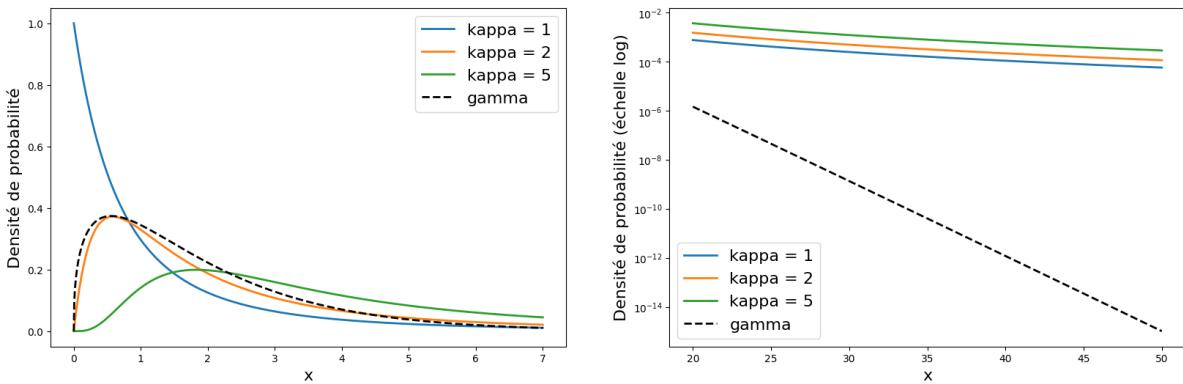


Figure 2.4 Densities of model (1) for "small" and "moderate" values (left) and "large" values (right), as a function of κ . The GPD shape is recovered when $\kappa = 1$. The Gamma distribution with parameters (1.4, 1.4) is shown for comparison. Figure inspired by [1].

2.3 Relationship Between Precipitation and Temperature

2.3.1 Clausius-Clapeyron

The Clausius-Clapeyron relation describes the link between the saturation vapor pressure of a substance and its temperature. Introduced between 1824 and 1850 through several works, this fundamental relation of thermodynamics plays an important role in atmospheric sciences, particularly in the study of precipitation, as it quantifies the air's capacity to hold water vapor according to its temperature. In the case of a phase change between liquid and vapor at saturation, the Clausius-Clapeyron relation is written as:

$$\frac{\partial e_s}{\partial T} = \frac{L_v e_s}{R_v T^2} \quad (2.10)$$

where e_s is the saturation vapor pressure (in hPa), T is the temperature (in Kelvin), L_v is the latent heat of vaporization (in $J \cdot K^{-1}$), and R_v is the specific gas constant for the substance under study (in $J \cdot kg^{-1} \cdot K^{-1}$).

Under the assumption that $L_v(T)$ is constant, the August-Roche-Magnus formula offers an approximate solution to the Clausius-Clapeyron equation for water vapor in the atmosphere:

$$e_s(T) \approx 6.1094 \exp \left(\frac{17.625T}{243.04 + T} \right) \quad (2.11)$$

where e_s is the saturation vapor pressure (in hPa), and T is the temperature (in °Celsius).

This solution has the characteristic of being in exponential form, with its derivative showing an exponential evolution of saturation pressure as a function of temperature, at a rate between 6 and 7%. This result represents the physical basis for the hypothesis of transferring the Clausius-Clapeyron formula to precipitation, assuming that precipitation extremes would increase as a function of temperature at a similar rate.

2.3.2 Estimation by Binning Method

The relationship between extreme precipitation and temperature has frequently been studied using the so-called "binning method." This approach was widely used in early attempts to empirically validate the hypothesis of extreme precipitation amplification according to the Clausius-Clapeyron relation [16, 17].

Low precipitation intensities (generally below $0.1\text{--}2 \text{ mm} \cdot \text{h}^{-1}$) are first excluded from the data series. The remaining values are then grouped into intervals (bins) defined according to the associated temperature. The size of these intervals can be determined either by a

fixed number of observations or by a given width on the temperature scale (*e.g.*, 2°C). For each interval, the mean temperature and a high quantile q (such as the 99th percentile level) of precipitation are calculated. A linear regression is then fitted to the logarithmically transformed data, modeling the relationship between the precipitation quantile P_q and the mean temperature \bar{T} [18]:

$$\log(P_q) = \alpha\bar{T} + \beta. \quad (2.12)$$

The percentage rate of precipitation increase per degree Celsius is obtained from the coefficient α :

$$\frac{\partial P_q}{\partial T} = 100(e^\alpha - 1) \quad (2.13)$$

This method has the advantage of being both simple to implement and easy to interpret. However, several limitations must be considered. On one hand, the choice of the initial exclusion threshold can significantly influence the resulting scaling rates. On the other hand, an inversion of the relationship between temperature and precipitation is frequently observed beyond a certain thermal threshold, typically between 20 and 30°C, which is explained by the decrease in available moisture at high temperatures. Two strategies are generally proposed to address this non-linearity: (1) substituting air temperature with dew point as the explanatory variable, or (2) explicitly integrating a hook/breakpoint into the linear regression [19].

Finally, a frequently used and conceptually similar alternative method consists of applying quantile regression directly to the entire dataset without prior binning [20, 21].

2.3.3 Estimation via Generalized Extreme Value Distribution

Several studies have nonetheless voiced criticisms of the binning approach [22, 23]. In particular, seasonal variations influencing both temperatures and precipitation simultaneously introduce a confounding factor that challenges the existence of a direct causal link between these two variables [24]. The relevance of the binning method for measuring the impact of climate change on extreme precipitation is therefore questioned, as the link between daily temperatures and precipitation (which can be explained by seasonal variations) is significantly different from the link associated with global warming.

To address this issue, the rate of precipitation increase as a function of temperature anomalies under climate change can be estimated using a non-stationary GEV distribution [24]. Annual precipitation maxima are selected and normalized by dividing them by the series median. A GEV distribution is then fitted to these data, using annual mean temperature anomalies (or anomalies limited to the season with the highest probability of extreme precipitation occurrence, *e.g.*, summer) as a covariate. The parameterization of the dependence of GEV

parameters on temperature anomalies varies by implementation. Some use [25]:

$$\mu(T) = \mu_0 + \mu_1 T, \quad \sigma(T) = \exp(\sigma_0 + \sigma_1 T), \quad \xi(T) = \xi. \quad (2.14)$$

The Temperature-Precipitation Scaling Rate (TPSR), denoted α , is obtained by calculating the quantile q for temperature anomalies T_1 and T_2 , leading to Q_1 and Q_2 . Any quantile of level q as a function of T_i can be obtained by the formula:

$$Q(q; \mu(T_i), \sigma(T_i), \xi(T_i)) = \begin{cases} \mu(T_i) - \sigma(T_i) \log[-\log(q)] & \text{if } \xi(T_i) = 0 \text{ and } q \in (0, 1); \\ \mu(T_i) + \frac{\sigma(T_i)}{\xi(T_i)} \{[-\log(q)]^{-\xi(T_i)} - 1\} & \text{si } \xi(T_i) > 0 \text{ and } q \in [0, 1], \\ & \text{or } \xi(T_i) < 0 \text{ and } q \in (0, 1]. \end{cases} \quad (2.15)$$

Finally, the TPSR α (in %/ $^{\circ}$ C) is given as the solution to the formula:

$$Q_2 = Q_1(1 + 0.001\alpha)^{T_2 - T_1} \quad (2.16)$$

In general, $\Delta T = T_2 - T_1 = 1$ is chosen, as the value of ΔT has little impact on the result [26].

CHAPTER 3 LITERATURE REVIEW

This chapter aims to provide a synthesis of the literature regarding the main themes and challenges addressed in this thesis. First, we will offer a brief summary of the history of numerical climate modeling, its major advancements, and its current challenges. We will then cover standard downscaling methods before focusing on the use of deep learning in this context, with particular attention to so-called "physics-informed" approaches.

3.1 Numerical Climate Modeling

The first numerical climate model was proposed in 1955, simulating the atmosphere across two vertical levels for a duration of 31 days [27]. These initial works, combined with progress in computing power and the growing volume of collected data, led to the emergence of the first Global Climate Models (GCMs) during the following decades [28].

GCMs are numerical models comprising, on one hand, a "dynamical core" that solves the fundamental equations of the atmosphere (*e.g.*, Navier-Stokes), and on the other hand, a physical component that models the exchanges between the planet's major "spheres": oceans, continents, and atmosphere. Certain physical processes occurring at scales finer than the resolution of GCMs are represented by empirical parameterizations; these are sources of uncertainty that motivate the push toward increasing the spatial and temporal resolution of models [29, 30]. By progressively integrating interactions between the climate and the biosphere and explicitly modeling more phenomena (marine biogeochemistry, aerosols, etc.), GCMs have evolved into what are now called Earth System Models (ESMs) [28, 31].

In parallel with this evolution, the applications of numerical climate modeling have expanded considerably. The advent of global climate simulations in recent decades has enabled major breakthroughs in understanding the causes and consequences of climate change [32, 33]. Climate projections established under various greenhouse gas (GHG) emission scenarios have significantly helped guide national and international policies, as well as anticipate the evolution of extreme weather events—such as floods associated with a 100-year return period—or estimate the future water needs of urban areas and the agricultural sector [29, 34–36]. Despite recent improvements, the current computational complexity of ESMs remains prohibitive, making the operational production of simulations at spatial resolutions lower than 50 kilometers (km) unfeasible [37].

Given that many physical processes can only be resolved at a finer spatial resolution, large-

scale simulations lack the physical details necessary to account for small-scale dynamics. However, there is an acute need for high-resolution climate simulations to assess the impacts of climate change within the framework of regional or local climate impact studies [38, 39], risk assessments related to extreme weather phenomena [40], hydrological modeling [41], or forest fire modeling [42]. To meet this need, downscaling techniques have been developed to increase the resolution of climate simulations, most often by restricting the coverage to a specific region. Two main approaches are distinguished: *dynamic* or *statistical* downscaling.

3.2 Dynamic and Statistical Downscaling

Dynamic downscaling consists of running a high-resolution simulation using a Regional Climate Model (RCM) over a restricted spatial domain, constrained at its boundaries by the outputs of a lower-resolution global model. Conversely, *statistical* approaches seek to establish empirical relationships between high-resolution data (*e.g.*, historical observations, satellite data, high-resolution simulations) and low-resolution climate simulations to estimate local climate variables from global fields.

Dynamic downscaling models integrate physical, dynamic, and biochemical processes specific to each simulated region. Lateral and upper boundary conditions are fixed by global model simulations at each time step, ensuring consistency with large-scale atmospheric circulation [43]. These models allow for simulations at spatial resolutions up to 10 km [44]. For reference, the 5th Canadian Regional Climate Model (CRCM5) provides climate simulations at a resolution of $0.11^\circ \approx 12$ km over the North American domain, including 685×668 cells across 56 vertical levels and 17 surface levels, with a 5-minute time step [45, 46].

Increasing the spatial and temporal resolution of RCMs makes it possible to model a larger number of fine-scale phenomena and interactions, thereby reducing the biases of global simulations [47]. However, their high numerical cost still limits the availability of long-term climate simulation ensembles, which are essential for modeling natural climate variability and evaluating a diversity of GHG emission scenarios [48].

Unlike the dynamic method, *statistical* downscaling does not explicitly solve physical processes. This family of methods aims to establish statistical relationships between low-resolution atmospheric variables from GCMs (or sometimes RCMs) and high-resolution data (either observed or from simulations in the case of emulators). Several methodological frameworks separate these methods [49]:

- **Perfect Prognosis (PP):** These methods involve calibrating a statistical model between large-scale predictors (ideally from reanalyses—simulations aligned with obser-

vations) and local observations. The resulting model, generally based on regression, is then applied to corrected or bias-free GCM outputs to generate local projections. The quality of predictors is crucial for the success of PP approaches.

- **Model Output Statistics (MOS):** This approach involves correcting the distribution of GCM outputs to align them with local observations by performing a statistical correction, typically via quantile mapping or scaling correction methods. MOS methods generally correspond to bias correction.
- **Weather Generators (WG):** These models rely on stochastic approaches that can fall under either the PP or MOS frameworks. WGs model the marginal properties and temporal structure of meteorological variables, and are sometimes capable of reproducing spatial dependence between predictors.

Statistical downscaling methods are commonly used in the literature, partly due to their simplicity of implementation and low associated computational costs. For example, some studies have evaluated the impact of global warming on flooding by applying methods from each of the aforementioned downscaling frameworks to refine the resolution of a GCM simulation ensemble [50].

Another methodological framework commonly used in deep learning, and adopted in this thesis, is called **Perfect Upscaling (PU)**. This framework, similar to PP, is based on the hypothesis that the relationship between predictors (low-resolution simulations and potential multi-resolution covariates) and predictands (high-resolution simulations or observations) is exactly the inverse of an aggregation process (*e.g.*, an average or an accumulation). It thus bypasses the GCM bias problem.

3.3 Deep Learning for Downscaling

Over the past ten years, deep learning has gradually gained popularity in the field of climate modeling [51]. These methods have demonstrated their effectiveness in various applications, including global climate simulation [52], weather forecasting with performance comparable to traditional numerical models [53], the parameterization of numerical climate models to improve fine-scale process modeling [54], and the downscaling of climate simulations [55].

Early applications of deep learning to climate downscaling were inspired by progress in super-resolution, a computer vision task aimed at reconstructing a high-resolution image from a degraded version. As an ill-posed inverse problem, super-resolution seeks to restore fine spatial details lost during degradation. These works laid the foundation for the first attempts

to adapt such models to climate simulation downscaling. Continuing from Section 2.1, we will present works using deterministic approaches separately from those based on generative methods.

3.3.1 Deterministic Methods

Deterministic models were the first to be used for climate variable downscaling. The potential of deep convolutional neural networks for this task was demonstrated by adapting a model from the super-resolution literature for the spatial downscaling of daily precipitation, increasing simulation resolution from 1° to 0.125° and achieving performance superior to Bias Correction Spatial Disaggregation (BCSD) or automated regression approaches [56–58]. The results also highlight the importance of adapting architectures from super-resolution (which primarily handle RGB images) to account for the specific characteristics of climate variables.

Subsequently, numerous works have implemented and compared CNNs to classic statistical methods. Some focused on modeling extreme precipitation [59], while others explored hybrid architectures combining CNNs and multilayer perceptrons (MLP) [60], or used CNNs to estimate the parameters of probability distributions intended to represent precipitation data [61]. Improvements to the model from [58] have also been proposed, significantly enhancing performance while reducing numerical cost [62]. UNet-type models have also been used for climate variable downscaling [63–65].

Other larger and more complex deterministic models (in terms of parameters and architecture) have also been successfully used, though these approaches are not explored in this thesis. A Transformer-type model pre-trained on very large heterogeneous datasets achieved remarkable performance on several tasks, including downscaling [66]. Other Transformer models (SwinIR and Uformer) were implemented for the downscaling and bias correction of temperature and wind speed [67]. The Mamba model, from the State Space Model family, has also been applied to precipitation downscaling constrained by topography [68]. Finally, other works have proposed an approach based on Fourier operators, allowing downscaling to an arbitrary resolution [69].

The loss function used for model training has also been the subject of several studies, particularly with the aim of improving the representation of extreme events. For example, it has been demonstrated that a UNet-type model, trained using a combination of an exponentially weighted quadratic loss and a structural similarity index measure (SSIM) from computer vision, is capable of modeling the entire high-resolution precipitation distribution, including the tail and extreme values [70]. An asymmetric loss function was also designed to penalize underestimations more heavily, proportional to the target value's intensity, enabling better

capture of extremes [65]. Finally, some works argue that the Charbonnier loss [71] $\mathcal{L}_{\text{charb}}$ offers a more robust alternative to standard \mathcal{L}_1 and \mathcal{L}_2 loss functions [67]:

$$\mathcal{L}_{\text{charb}}(\theta) = \frac{1}{N} \sum_{i=1}^N \sqrt{(\hat{y}_i - y_i)^2 + \epsilon^2} \quad (3.1)$$

$$\mathcal{L}_1(\theta) = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i| \quad (3.2)$$

$$\mathcal{L}_2(\theta) = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2 \quad (3.3)$$

where N is the number of samples, y_i is the target value, $\hat{y}_i = f_\theta(x_i)$ is the model prediction with parameters θ , and ϵ is a small positive term (*e.g.*, 10^{-3}) ensuring differentiability.

3.3.2 Generative Methods

Due to their probabilistic nature, generative methods offer two main advantages over deterministic approaches: (1) by modeling the conditional distribution of a variable, models in this family better capture the natural variability associated with the climate, and (2) it is possible to quantify prediction uncertainties, which is crucial for downscaling extreme weather events.

To leverage the advantages of generative methods while retaining the capabilities of the UNet architecture, a method was developed to adapt this model to a probabilistic context for medical image segmentation [72]. The authors propose learning a latent space representing the parameters of a Gaussian distribution and sampling from this distribution to enrich UNet predictions. This approach models the variability inherent in solutions due to the ill-posed nature of the inverse problem. In a collaborative project not mentioned in this thesis, we successfully apply this method to precipitation downscaling, improving the preservation of high-frequency data and the prediction of extreme events.

Although not addressed in detail in this thesis, several generative approaches have been successfully applied to climate downscaling. In particular, several recent studies have highlighted the superior performance of diffusion models [73] for high-resolution field reconstruction [74–76]. A Generative Adversarial Network (GAN) of the CycleGAN type [77, 78], capable of faithfully reproducing fine spatial structures of precipitation while correcting systematic biases present in ESM outputs, was also recently proposed [79].

3.4 Physics-Informed Deep Learning

The nature of the downscaling task, combined with the inherent physical constraints of climate modeling, constitutes a dual motivation for the development of "physics-informed" methods.

First, downscaling (in the climate domain) constitutes an ill-posed inverse problem for two reasons: (1) the solution is not necessarily unique due to information loss between high and low resolution, and (2) the solution does not depend continuously on the input data, as small perturbations can cause significant variations in the solution [80]. To address these limitations, it is necessary to introduce priors on the system to reduce sensitivity to perturbations and constrain the space of admissible solutions. Second, predictions from deep learning methods can exhibit physical inconsistencies [81]. The introduction of priors—whether simple or sophisticated—aims to improve the realism of results, strengthen robustness against out-of-distribution (OOD) situations, and increase user confidence in the models.

Among "physics-informed" approaches, two main categories are generally distinguished. The first includes methods that integrate constraints (hard or soft) into the cost function to enforce total or partial compliance with physical laws or conservation principles. The second category relies on incorporating priors directly within the model architecture, using system knowledge (also called "domain knowledge") to guide its design. A literature review lists various approaches aimed at integrating physical constraints and climate science priors into deep learning methods applied to climate modeling [82].

Within constraint-based approaches, a distinction is generally made between hard and soft constraints. A constraint is considered "hard" when it must be strictly satisfied by any admissible solution, whereas a "soft" constraint is introduced as a penalty in the loss function, weighted by a coefficient. This offers more flexibility but fewer guarantees. In this vein, authors have proposed several hard constraint formulations to enforce mass conservation in a downscaling framework applied to water vapor columns, comparing the performance of unconstrained deep learning models with those using hard or soft constraints [83]. Their conclusions indicate that the introduction of constraints reduces errors in regions with high spatial complexity, such as coastal or mountainous zones, also contributing to a notable improvement in the visual quality of predictions. Other works exploit the Clausius-Clapeyron relation to integrate hard or soft constraints into post-processing and atmospheric variable downscaling contexts, respectively [84, 85]. Finally, another hard constraint approach consists of imposing, in post-processing, an equilibrium constraint between high-resolution GAN predictions and the aggregated values of input fields to reinforce physical consistency and

improve the robustness of predictions under climate change [70].

The second category encompasses a broader family of methods that propose integrating priors through the design of deep learning architectures or data processing steps. For example, one approach aimed at improving the downscaling of precipitation events at very high resolution (250 meters) adapts a video super-resolution method by replacing the "optical flow" field used to transmit information between frames with the wind advection field [86]. This substitution exploits knowledge of the strong dependence between precipitation and wind dynamics, thereby generating spatially consistent fine-scale fields. Another approach proposes integrating an orographic model (enabling better modeling of precipitation structures in complex terrain) within a GAN and conditioning the generative model's latent space using a Gaussian process ensemble to obtain a statistically and physically enriched starting point [87]. Some propose modeling climate processes as an advection partial differential equation (PDE) and predicting the evolution of the velocity field using neural networks [88]. This approach explicitly integrates the physical dynamics of the climate system into predictions. Finally, other authors introduce a series of transformations inspired by atmospheric sciences, allowing various climate modeling tasks to be performed in a climate-invariant space [89].

CHAPTER 4 DATA AND EXPLORATORY ANALYSIS

The objective of this chapter is to present the dataset used in this thesis in detail and to provide an exploratory statistical analysis of the variables of interest.

4.1 ClimEx: A Large Ensemble of Climate Simulations under Global Warming

The ClimEx project (also known as CRCM5-LE) provides an ensemble of 50 independent regional climate simulations at a spatial resolution of $0.11^\circ \approx 12$ km. These simulations cover two geographic domains—Northeastern North America and Europe—over a period ranging from 1950 to 2099, with a temporal resolution of 5 minutes [90]. To obtain these high-resolution simulations, dynamical downscaling was performed using an ensemble of 50 independent global simulations generated by the Canadian ESM CanESM2 at a spatial resolution of $2.8^\circ \approx 310$ km. These served as the forcing for the Canadian RCM CRCM5 (version 5), developed at the Université du Québec à Montréal (UQAM). The independence of the ensemble members is based on the introduction of initial perturbations applied to the CanESM2 global simulation: 5 perturbations in 1850, followed by 10 in 1950. These perturbations allow for the generation of 50 independent global trajectories starting from 1955, capturing natural climate variability. The radiative forcing applied from 2006 onward corresponds to the RCP8.5 greenhouse gas emission scenario. As a result, approximately fifty variables per simulation were archived, with time steps ranging from hourly to daily. Some variables are available via open access at a daily time step through the PAVICS platform.

The work presented in this thesis relies on a subset of the ClimEx dataset. Specifically, we use the outputs from a single ensemble member (member *kdf*), leaving the study of inter-member learning and generalization for future work. Daily precipitation accumulations (**pr**), expressed in $\text{kg} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, are extracted, along with daily temperature statistics: mean (**tas**), minimum (**tasmin**), and maximum (**tasmax**), expressed in Kelvins (K). The integration of additional climate covariates for precipitation downscaling is left for future research. Furthermore, due to numerical constraints, the initial spatial domain—comprising 280×280 cells—is restricted to a region of 128×128 cells covering southern Quebec, part of the Canadian Maritime provinces, and the northeastern United States (see Figure 4.1 (a)).

4.2 Exploratory Statistical Analysis

To identify trends potentially linked to climate change, we distinguish between two periods: **A** and **B** (1955–2030 and 2025–2099), each covering 75 years of daily data with a 5-year overlap. Period **A** can be considered a period of moderate climate change, while Period **B** represents significant climate change under the RCP8.5 scenario. The map of the absolute difference in the median of mean daily temperatures, presented in Figure 4.1 (a), highlights general warming across the entire ClimEx domain, particularly in the North. It also underscores the heterogeneity of the response within the chosen domain regarding temperature evolution under climate change. The largest difference in the subdomain is 5.07°C, the smallest is 2.40°C, and the median is 3.52°C—compared to 5.51°C, 2.11°C, and 3.37°C for the entire domain. This suggests that the chosen subdomain is representative of the ClimEx domain in terms of climate change response.

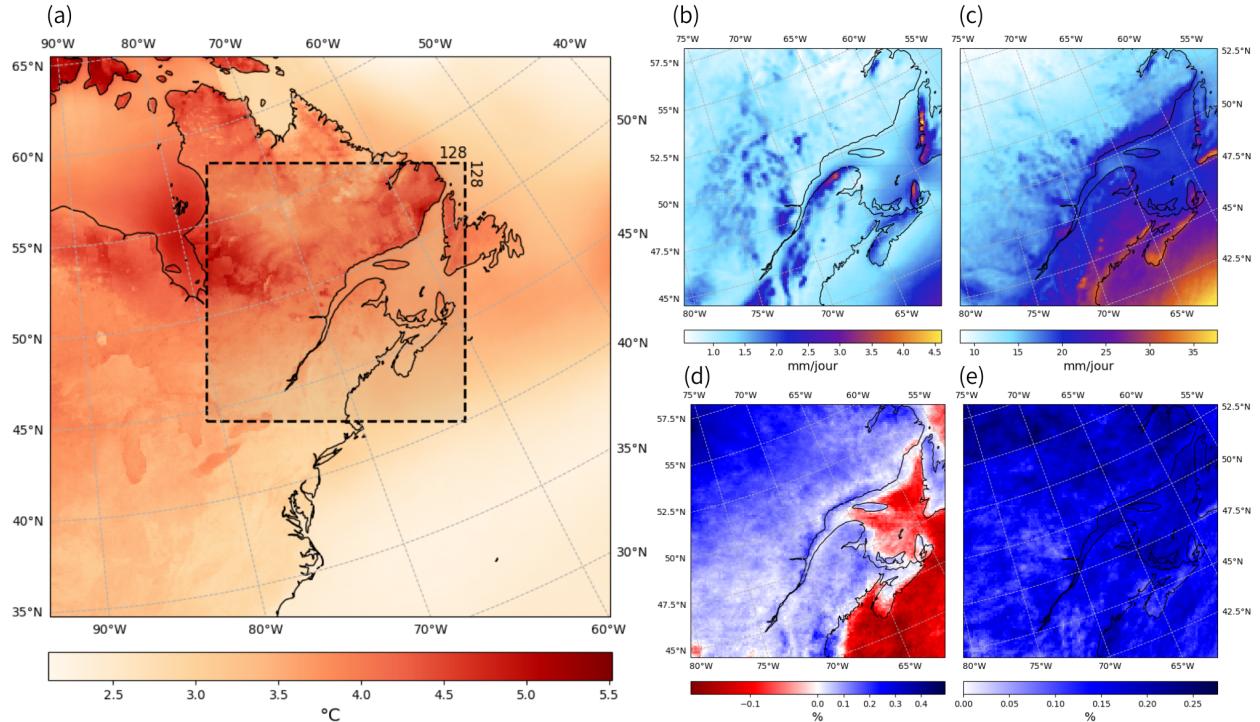


Figure 4.1 Let Period **A** range from 1955 to 2030, and Period **B** be a five-year overlapping period from 2025 to 2099. (a) Absolute difference between Periods **B** and **A** of the median daily mean temperatures over the ClimEx domain; the black dotted line indicates the subdomain used in this thesis. (b) Median non-zero daily precipitation over Period **A**. (c) 95th quantile of non-zero daily precipitation over Period **A**. (d) Relative difference in median non-zero precipitation. (e) Relative difference in the 95th quantile of non-zero precipitation.

Maps (b:e) in Figure 4.1 are restricted to the 128×128 cell subdomain. Maps (b) and (c)

present the median and the 95th quantile of non-zero precipitation for Period **A**, respectively. It is observed that regions with a high median correspond to mountainous coastal regions: the Chic-Choc Mountains in Gaspésie, the Cape Breton Highlands in Nova Scotia, and the Long Range Mountains in Newfoundland and Labrador. Due to their complex orography, these regions generally pose greater challenges during downscaling, as the fine dynamics associated with this terrain are not captured by GCMs. The heavy precipitation observed in coastal regions (map (c)) is explained by the frequent convergence of low-pressure systems originating from the North Atlantic, whose intensity is amplified by moisture from the Gulf Stream.

Maps (d) and (e) illustrate the relative variations between Periods **B** and **A** for the median and 95th quantile of daily non-zero precipitation. Marked spatial patterns are observed for the median, with a strong increase over continental areas (up to 49.2%), relative stability along the coasts, and a notable decrease in median precipitation over the Atlantic Ocean and the Great Lakes (bottom left of map (d)), reaching up to -18.7%. This heterogeneity suggests that the evolution of light-to-moderate precipitation depends heavily on external factors, such as orography or regional atmospheric conditions. In contrast, extreme precipitation (represented by the 95th quantile) is increasing across the entire domain without a specific spatial pattern. The increase ranges from 5.14% to 27.6%, with an average of 15.99%. With all due caution, this generalized upward trend, parallel to the rise in mean temperatures (cf. map (a)), appears consistent with the theoretical Clausius-Clapeyron relation presented in Section 2.3.

As stated in Section 3.2, the work in this thesis relies on the **Perfect Upscaling** methodological framework. We assume that the high-resolution precipitation fields correspond to the inverse of an aggregation by averaging. Data from ClimEx are considered high-resolution (HR), and spatial averaging over blocks of 256 cells (16×16) is performed to obtain low-resolution (LR) data. A random example of the result of this operation is shown on the right side of Figure 4.2. Logically, a significant loss of information occurs due to spatial aggregation. In this case, the fine patterns associated with precipitation largely disappear, making it difficult to identify precise meteorological structures or to locate extreme precipitation events, which are often concentrated in small areas or bands.

The analysis of precipitation histograms (see Figure 4.2), calculated over the aggregation of all cells in the domain, highlights the effects of spatial aggregation: at low resolution, the density of light precipitation increases due to the "dilution" of rain-free cells into neighboring zones. Conversely, moderate precipitation becomes clearly under-represented—a direct consequence of spatial smoothing. Finally, most of the distribution tail, corresponding to heavy

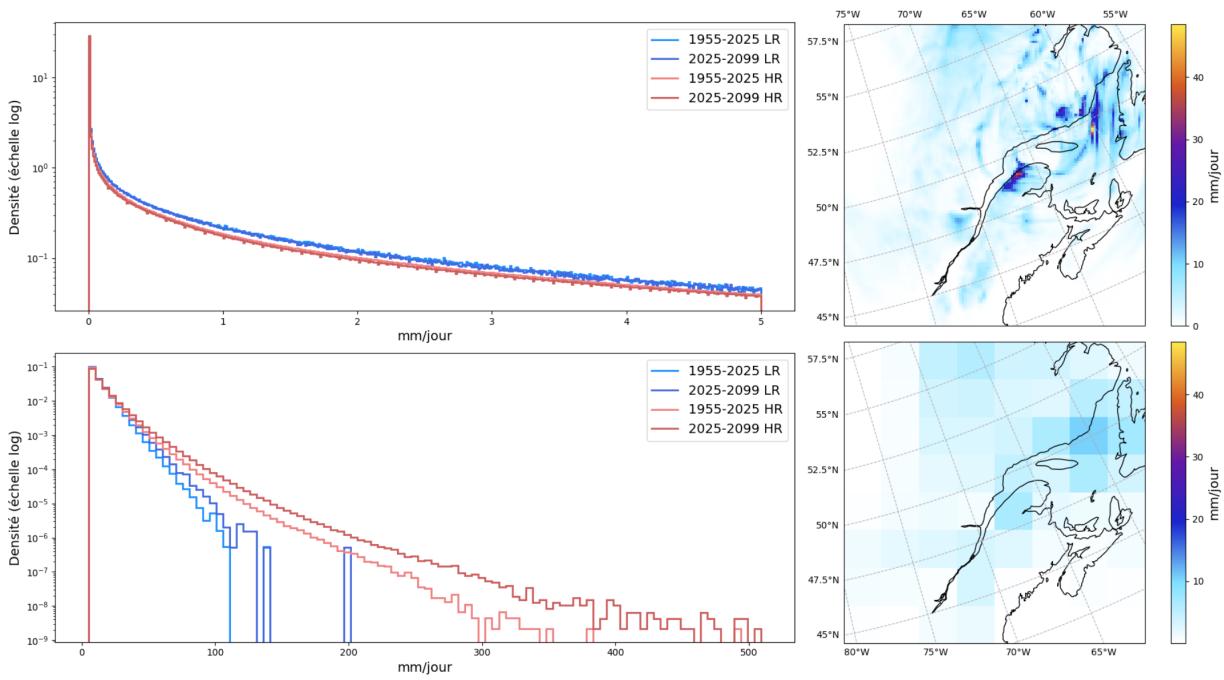


Figure 4.2 To simulate low-resolution (LR) observations, high-resolution (HR) daily precipitation from ClimEx is aggregated by averaging 16×16 cell blocks (256 cells). **(Left)** Histograms of daily precipitation, separated by intensity (0–5 mm/day top, and ≥ 5 mm/day bottom), comparing Periods A and B at high and low resolution. **(Right)** A randomly selected daily precipitation field (top) and its aggregated version (bottom).

precipitation at high resolution, is strongly attenuated or entirely absent at low resolution. When separating the data by period, it is clearly observed that the precipitation distribution tail for Period **B** is generally heavier, evidencing a marked increasing trend in extreme precipitation under climate change.

The ability to restore fine spatial details specific to precipitation events, to faithfully reconstruct their distribution tail, and to integrate the amplification of extreme values linked to climate change constitute the main challenges that the work presented in this thesis seeks to address.

CHAPTER 5 DEEP LEARNING AND PRECIPITATION DOWNSCALING UNDER CLIMATE CHANGE

In this chapter, we detail the architecture of the deep learning model used to perform the downscaling of precipitation fields. We then present an initial performance evaluation and analyze its limitations to motivate the methodological improvements proposed in the following chapter.

5.1 UNet for Downscaling

This section aims to present the original architecture of the UNet model used in this thesis and describes the strategy employed to perform precipitation field downscaling with this model.

5.1.1 Architecture Presentation

As previously indicated, we use an architecture derived from prior work as a starting point [2]. The architecture follows a classic encoder/decoder structure, where each level is composed of several residual blocks with dedicated skip connections and the integration of temporal information to guide the UNet. We present the details of this architecture here, as illustrated in Figure 5.1.

The encoder is composed of four levels, each doubling the number of channels (the dimension of "features," analogous to the three RGB channels in a color image) of the previous one and halving their spatial dimensions. Thus, an input of 128×128 cells is transformed into feature maps of 16×16 cells. Each encoder level includes three residual blocks: the first performs spatial average pooling, the second doubles the number of channels, and the third preserves dimensions. The decoder features four residual blocks per level: the first performs upsampling, and the others reduce the number of channels while integrating information from the corresponding encoder skip connections.

The original structure of the residual blocks is illustrated in Figure 5.1 and may include, depending on the context, a pooling operator or nearest-neighbor upsampling. Since no significant performance gain was observed during tests with bilinear or cubic interpolation, nearest-neighbor upsampling was retained for its simplicity and the absence of introduced bias. The activation function used is $\text{SiLU}(x) = x \text{sig}(x)$, frequently employed due to its differentiability at all points, its smooth gradient comparable to the sigmoid, and its non-

linear behavior close to that of the $\text{ReLU}(x) = \max(0, x)$ function [91]. Group normalization is used to stabilize training, as it is more robust than batch normalization for small batch sizes [92]. Group normalization centers the values within a group (fixed at 32 here) of feature maps and reduces their variance to unity. These maps correspond to the different outputs (equivalent to channels) produced by the filters of a convolutional layer.

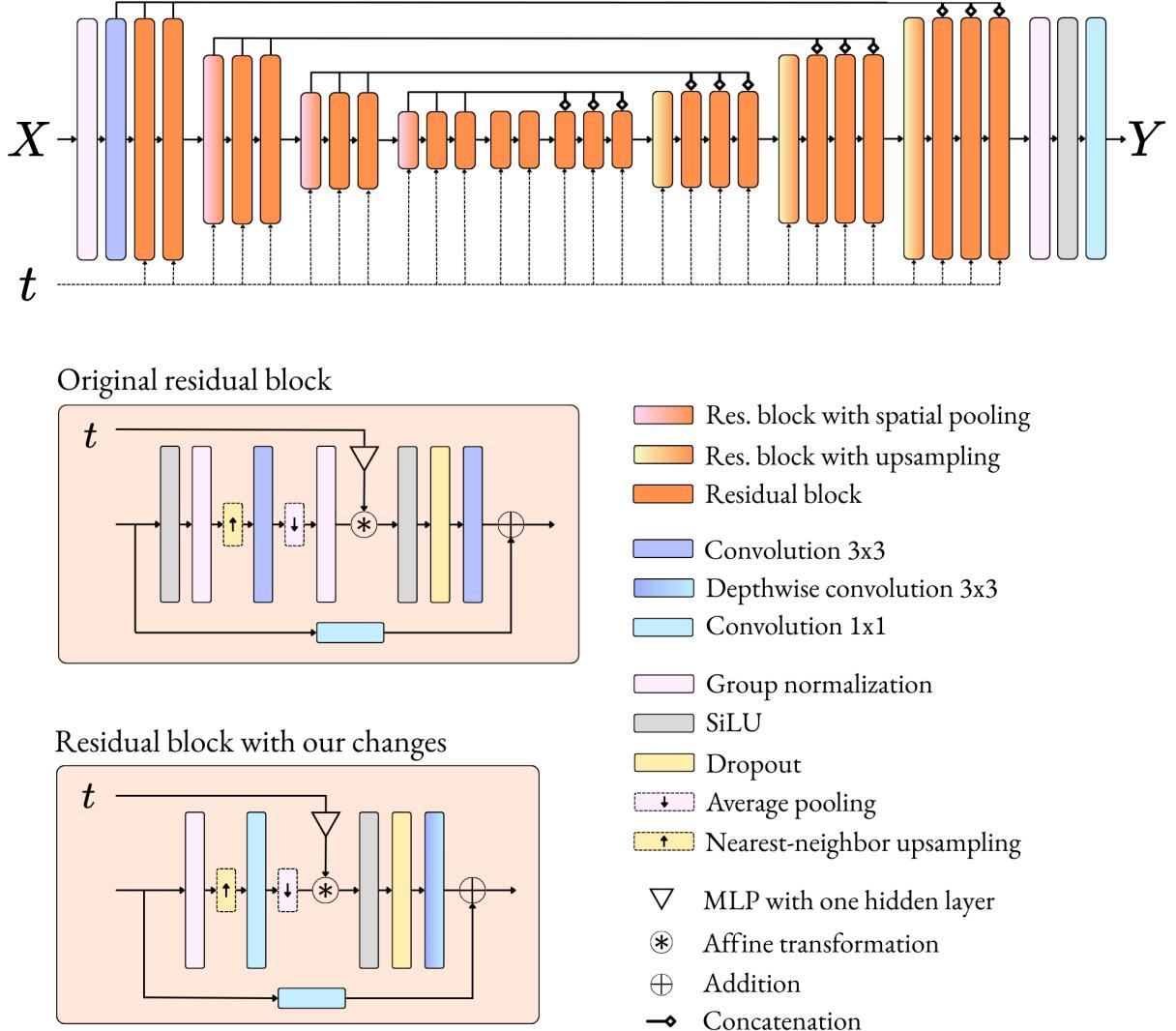


Figure 5.1 Presentation of the UNet model architecture used, and modifications made to the original residual block [2].

A dropout layer is introduced in each residual block to randomly set $100 \times p\%$ of the feature map values to zero during training. This mechanism limits overfitting by preventing the

model from "memorizing" a specific solution [93]. The probability p is a hyperparameter that we keep fixed at $p = 0.1$, as in the original architecture.

Temporal information related to the input X is integrated multiple times into the UNet to account for seasonal differences and future evolution. We use a representation that models the cycle of days within a month, months within a year, and the progression of years. For each date (m/d/y), we calculate the variable:

$$t = \left(\frac{a - a_{\min}}{a_{\max} - a_{\min}} \right) \cos \left(\frac{2\pi j}{31} \right) \sin \left(\frac{2\pi m}{12} \right), \quad (5.1)$$

where $j \in [|1, 31|]$, $m \in [|1, 12|]$, and $a \in [|1955, 2099|]$ are the day, month, and year of the precipitation field, respectively. a_{\max} and a_{\min} correspond to the maximum and minimum years of the training period. The variable t is then transformed by a Multilayer Perceptron (MLP) containing one hidden layer of dimension 128, producing two outputs: a scale parameter $\gamma \in \mathbb{R}$ and a phase parameter $\beta \in \mathbb{R}$. These two parameters are used to apply an affine transformation to the feature maps x , as illustrated in Figure 5.1:

$$y = \gamma x + (1 + \beta). \quad (5.2)$$

Finally, the model parameters θ are initialized randomly by sampling from a "Kaiming" uniform distribution [94]:

$$\theta_l \sim \mathcal{U}(-b, b) \quad \text{with} \quad b = \sqrt{\frac{9}{n_{l-1}}}, \quad (5.3)$$

where n_{l-1} corresponds to the number of input channels for layer l .

5.1.2 Strategy, Preprocessing, and Post-processing

Within the *Perfect Upscaling* framework, low-resolution precipitation fields are artificially generated by spatial averaging of the high-resolution daily data from ClimEx (see Section 4). To practically evaluate deep learning capabilities for downscaling, we adopt an aggregation factor of 16—significantly higher than in most existing works—while maintaining a sufficient level of information across our restricted domain. Each low-resolution field thus comprises 8×8 cells, resulting from the aggregation of 256 high-resolution cells.

The UNet architecture requires the input and output to have similar spatial dimensions to ensure the continuity of skip connections. Thus, it is necessary to upsample the low-resolution precipitation fields before feeding them into the model. An alternative approach

could involve extending the decoder to achieve a higher output resolution than the input, but as our experiments were inconclusive, we decided to focus on the first approach [95]. Many studies use bilinear or bicubic interpolation for this upsampling. We chose to perform nearest-neighbor upsampling to avoid introducing bias at this stage, noting no significant impact on performance in preliminary experiments.

Let $x, y \in \mathbb{R}^{H \times W}$ be two high-resolution precipitation fields totaling $H \times W$ cells, obtained respectively by interpolation from low resolution and directly from the high-resolution data. One approach is to train a UNet f_θ to predict y from x and any potential covariates v , i.e., $f_\theta(x, v) = y$. An alternative is to model only the residual $r = y - x$, by learning $f_\theta(x, v) = r$. This strategy allows the learning to focus on details absent from the interpolation rather than the entire field, some components of which have already been lost by the encoder. We adopt this residual approach hereafter, having observed consistently superior performance, in agreement with prior studies [2, 74].

Daily precipitation and mean temperature fields between 1955 and 2020 are used to constitute the training set. To optimize the UNet architecture, this set is split into a training subset covering 1955–2010 and a validation set ranging from 2010 to 2020. The test set extends from 2020 to 2099 to evaluate the model’s performance in a context of marked climate change and to analyze the evolution of its capabilities as climate effects intensify.

Parameter optimization is performed using the **Adam** algorithm [96], with a mini-batch size fixed at $N = 32$ unless otherwise stated. The UNet model is trained by minimizing the Mean Squared Error (MSE) between its predictions and target values. Let $R \in \mathbb{R}^{N \times H \times W}$ be the target residual fields, $X \in \mathbb{R}^{N \times H \times W}$ the precipitation fields interpolated from low-resolution data, $V \in \mathbb{R}^{N \times H \times W}$ the covariate fields, and f_θ the UNet parameterized by θ . The loss function used for training is then given by:

$$\mathcal{L}(\theta) = \frac{1}{N H W} \sum_{i=1}^N \|f_\theta(X_i, V_i) - R_i\|_2^2. \quad (5.4)$$

Preprocessing

As illustrated by the histograms in Figure 4.2, precipitation is heavily concentrated around zero with a marked distribution tail. This structure motivates applying a transformation to precipitation values for two main reasons: (1) to expand the range of light precipitation, which is often under-represented, and (2) to reduce the proportion of negative predictions, which indicate physical inconsistency and generate an overabundance of zeros after truncation. To meet these objectives, we compare two transformations, denoted T_1 and T_2 , designed to relax

the positivity constraint in the transformed space, given by the following equations [70]:

$$T_1(x, \epsilon) = \log(e^{x+\epsilon} - 1), \quad (5.5)$$

$$T_2(x, \epsilon) = \log(1 + x) - \log(\epsilon), \quad (5.6)$$

with $\epsilon = 10^{-5}$ (the value yielding the least inconsistency in our preliminary experiments). The results for precipitation downscaling presented in Table 5.1 clearly demonstrate their effectiveness in reducing both the proportion of negative predictions and the magnitude of these inconsistencies. Both transformations also improve the Mean Absolute Error (MAE) in a comparable manner.

Table 5.1 MAE (in mm/day), proportions, and ℓ_2 norm of negative values for no transformation, T_1 , or T_2 . UNet trained on the 1960–2000 period and evaluated on 2000–2010.

	MAE	Neg. %	ℓ_2 Neg.
Without transformation	1.1236	12.88	228.73
With T_1	1.0866	2.20	18.3
With T_2	1.0856	9.47	27.92

For the remainder of this work, we retain the T_1 transformation. Data are normalized using the `min/max` method along the temporal axis. Let $\tilde{X} \in \mathbb{R}^{N \times H \times W}$ be a sequence of N transformed precipitation fields (using T_1); the normalization result is obtained by: $\tilde{X}_{\text{norm}} = \frac{\tilde{X} - \min_n \tilde{X}_n}{\max_n \tilde{X}_n - \min_n \tilde{X}_n}$. To remain consistent with the climate simulation downscaling framework, we use the minima and maxima of the low-resolution precipitation fields to normalize the high-resolution fields: $\tilde{Y}_{\text{norm}} = \frac{\tilde{Y} - \min_n \tilde{X}_n}{\max_n \tilde{X}_n - \min_n \tilde{X}_n}$. The normalized residuals are therefore equal to:

$$R = \frac{\tilde{Y}_{\text{norm}} - \tilde{X}_{\text{norm}}}{\max_n \tilde{X}_n - \min_n \tilde{X}_n}. \quad (5.7)$$

Post-processing

Once the preprocessing operations (residuals, normalization, transformation) are reversed, an additional post-processing step is introduced. Indeed, the model is not subject to any explicit constraint ensuring consistency between the value of a low-resolution cell and the distribution of predicted values in the corresponding high-resolution cells. This imbalance, while sometimes subtle during training, can become problematic out-of-distribution. To address this, we impose a constraint on the UNet predictions after training, ensuring this

cross-scale consistency [79]:

$$\hat{y}_i^* = \left(\frac{\sum_i^{N_{\text{grid}}} x_i}{\sum_i^{N_{\text{grid}}} \hat{y}_i} \right) \hat{y}_i, \quad (5.8)$$

where \hat{y}_i denotes a high-resolution precipitation cell predicted by the UNet, x_i the corresponding low-resolution value, and N_{grid} the number of high-resolution cells aggregated into each low-resolution cell.

Table 5.2 presents the MAE of the UNet with and without the application of the post-processing constraint (5.8), as well as the MAE calculated in the vicinity of specific quantile levels (0.5, 0.9, 0.95, and 0.99). It is observed that despite a slight increase in overall error primarily due to performance degradation in average precipitation (i.e., around the median), applying the constraint allows for a clear improvement in predictions for extreme values (beyond the 90th quantile). Furthermore, the prediction histogram (see Figure 5.2) highlights an expansion in the range of predicted precipitation, reflecting a better representation of high intensities.

Table 5.2 UNet trained on the 1955–2010 period and evaluated on the 2010–2020 period over 50 epochs. Comparison of performance with or without the post-processing constraint (5.8). Overall MAE (in mm/day) and MAE calculated in the vicinity of specific quantiles (0.5, 0.90, 0.95, and 0.99).

	MAE	MAE-q (0.5)	MAE-q (0.9)	MAE-q (0.95)	MAE-q (0.99)
Without constraint	1.1754	0.7824	3.6363	6.3347	16.1048
With constraint	1.1928	0.8594	3.6044	6.0203	14.6546

5.2 UNet Improvements

To improve the UNet’s performance, we integrate several recommendations from the literature aimed at strengthening deep convolutional models [97]. In accordance with the authors of these recommendations, we perform the modifications sequentially and retain only those that improve the MAE. It is important to emphasize that our main goal here is not to identify the optimal architectural choices for the UNet—which would involve comparing multiple combinations of modifications—but rather to leverage the literature to attempt to improve its performance.

The first modification involves adjusting the number of residual blocks per level to concentrate the model’s capacity on potentially more influential levels (see Figure 5.1). While the original encoder follows a (3–3–3–3) sequence, we experiment with two distributions: (2–2–4–4) and (4–4–2–2). Table 5.3 shows a marginal improvement in performance despite an

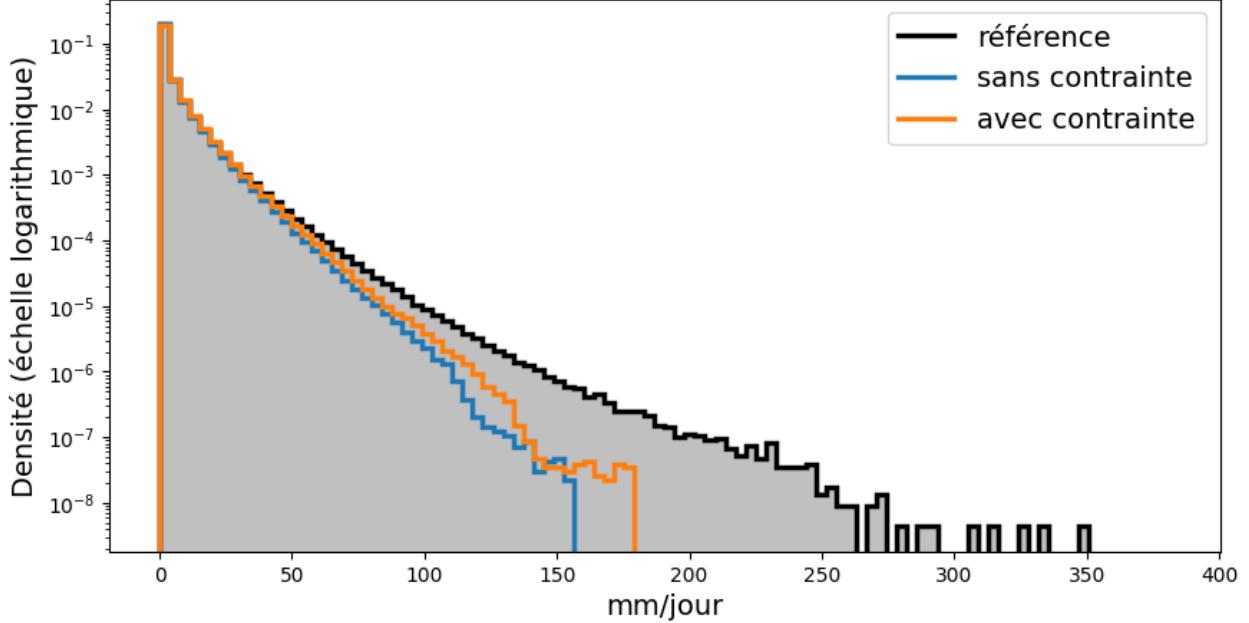


Figure 5.2 Histograms of precipitation predicted by the UNet with or without the constraint (5.8) over the evaluation period (2010–2020).

increase in the number of parameters for one of the sequences. **For the remainder of this work, we retain the initial number of blocks per level.**

Table 5.3 MAE for different sequences of blocks per encoder level (symmetric decoder with one additional block per level). UNet trained for 50 epochs with 64 samples per mini-batch.

Residual block sequence	MAE	Rel. MAE change	No. of parameters
(3 – 3 – 3 – 3)	1.1975	NA	5.709×10^6
(2 – 2 – 4 – 4)	1.1947	-0.23%	6.740×10^6
(4 – 4 – 2 – 2)	1.1953	-0.18%	4.679×10^6

The number of channels at the first encoder level is initially fixed at 32 to respect our GPU RAM constraints (≈ 24 GB). To optimize parameter usage, we modify the first two convolutional layers of each level. The first, responsible for varying the number of channels (in the encoder or decoder), uses 1×1 filters instead of 3×3 , concentrating learning solely on the channel dimension. The second, which preserves the number of channels, is replaced by a depthwise convolution—i.e., a convolution applied independently to each channel without cross-channel interaction. These adjustments significantly reduce the number of parameters, allowing the initial number of channels to be doubled while maintaining a constant memory budget. Table 5.4 highlights the clear improvement in MAE obtained through these modifi-

cations.

For the remainder of this work, we retain this modification.

Table 5.4 MAE for different convolution types and initial channel numbers. UNet trained for 50 epochs.

Convolution types	Init. channels	MAE	Rel. MAE change	Parameters
$(3 \times 3), (3 \times 3)$	32	1.1928	NA	5.709×10^6
$(1 \times 1), (3 \times 3 \text{ depthwise})$	64	1.1782	-1.22%	4.609×10^6

Given the reduction in model parameters, it has been suggested that increasing the filter size of the convolutional layers from (3×3) to (7×7) could improve performance by expanding the receptive field. Table 5.5 indicates a significant degradation in our case.

For the remainder of this work, we retain the (3×3) filter size.

Table 5.5 MAE for different convolution filter sizes. UNet trained for 50 epochs.

Filter size	MAE	Rel. MAE change	No. of parameters
(3×3)	1.1782	NA	4.609×10^6
(7×7)	1.1883	+0.86%	4.796×10^6

The final modifications consist of: (1) removing the first SiLU layer from the residual block, and (2) removing the second normalization layer from the residual block. These recommendations are inspired by Transformer model architectures. Table 5.6 indicates that these two modifications slightly improve the UNet’s performance. Notably, the error decreases for all evaluated quantiles, indicating a real performance gain.

For the remainder of this work, we remove these layers from the residual blocks.

Table 5.6 MAE with certain residual block layers removed: (1) only SiLU, and (2) SiLU and Group Normalization (GN). UNet trained for 50 epochs.

Layers removed	MAE	Rel. MAE change	No. of parameters
None	1.1782	NA	4.609×10^6
SiLU	1.1741	-0.35%	4.609×10^6
SiLU + GN	1.1732	-0.42%	4.600×10^6

The modifications made to the original architecture restore a MAE similar to that obtained without the post-processing constraint (5.8)—a total MAE reduction of 1.64%—while preserving the improvements for high precipitation quantiles.

5.3 Evaluation under Climate Change

To motivate the relevance of this project’s objective—to strengthen the climate change robustness of deep convolutional models for precipitation downscaling—it is first necessary to establish the sensitivity of the UNet to climate change.

We train the UNet for 50 epochs on the 1955–2020 period, totaling 23,725 daily precipitation fields. Its performance is then evaluated on a test set covering the 2020–2099 period. It should be noted that during this future period, the domain’s average temperature increases significantly, doubling between the 2020–2029 decade (5.37°C) and the 2090–2099 decade (11.04°C). This evaluation period is thus representative of a climate heavily impacted by climate change.

Figure 5.3 presents (top) the UNet’s annual MAE over the evaluation period (in mm/day) and (bottom) the relative evolution of the MAE compared to the average error over 2020–2025, evaluated in the vicinity of several quantiles. We first notice a clear increase in error as climate change intensifies, rising from ≈ 1.2 mm/day between 2020 and 2030 to over ≈ 1.4 mm/day in the final decade, corresponding to a 16.67% increase. If we fit a linear relationship to the annual errors, we obtain a linear error trend of 0.0031—i.e., an additional +0.15 mm/day of average error every 50 years. This is significant, as this magnitude corresponds roughly to the first quartile of precipitation on the test domain.

Evaluating the MAE in the vicinity of different quantiles allows for the identification of the precipitation intensities that contribute most to the model’s performance degradation. The second graph in Figure 5.3 shows that this degradation is more pronounced as the quantile level increases: the relative evolution of the MAE reaches approximately 25% for the $q = 0.99$ quantile level, compared to only 5 to 8% for the $q = 0.5$ level by the end of the period. This observation supports the hypothesis that the UNet fails to satisfactorily model the dependency relationship between rising temperatures and the intensification of extreme precipitation.

We split the test period into two sub-periods: 2020–2059 and 2060–2099, each covering half of the test period. The relative evolution of the MAE between the first and second sub-periods, presented in Figure 5.4a, is positive across almost the entire geographic domain. A correlation can be observed in the northwest of the domain between the marked increases in MAE and mean temperature (Figure 5.4b). Note that the portion of the domain covered by the ocean cannot be analyzed from the same perspective, as precipitation there is subject to different phenomena (e.g., oceanic depressions, Gulf Stream). This finding reinforces our hypothesis that the UNet does not adequately integrate the mechanism of precipitation increase under

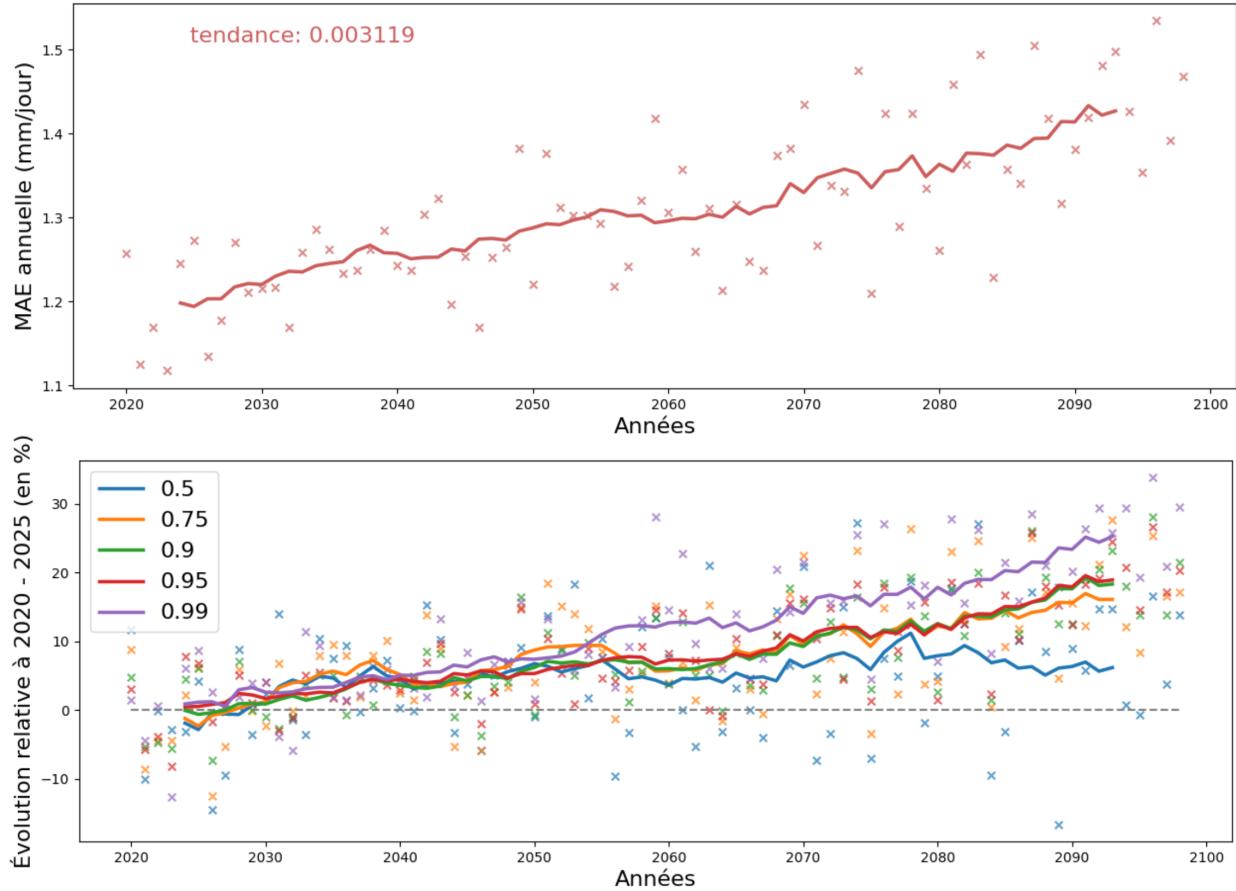


Figure 5.3 (**Top**) Mean annual MAE (in mm/day) of the UNet over the test period. (**Bottom**) Relative evolution (in %) of the annual MAE compared to the 2020–2025 period, calculated in the vicinity of quantile levels 0.5, 0.75, 0.90, 0.95, and 0.99. Curves obtained by a 10-value moving average.

rising temperatures.

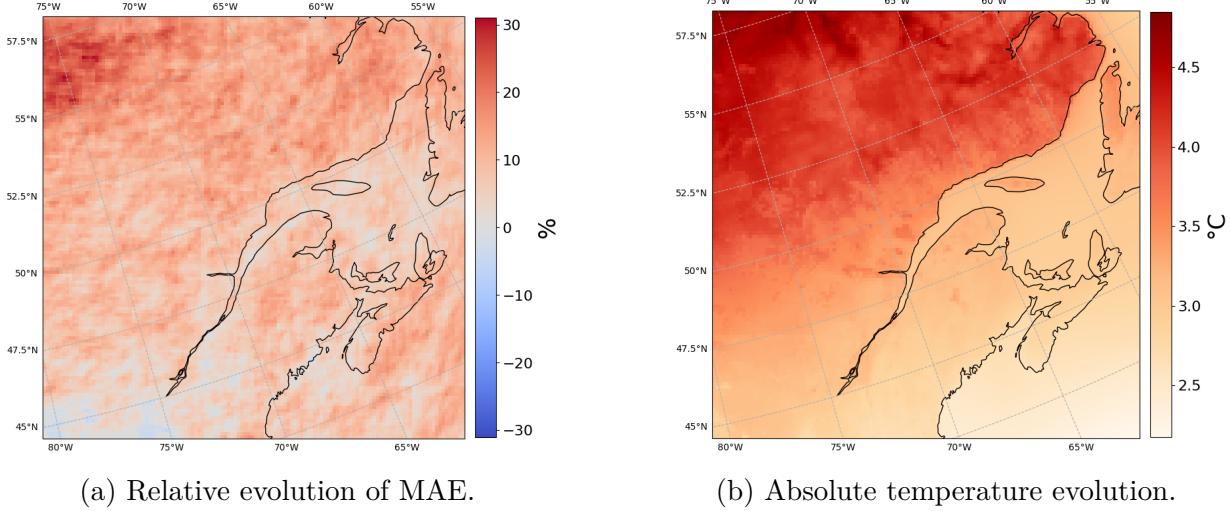


Figure 5.4 Comparison of the evolution of the Mean Absolute Error (MAE) of the UNet’s high-resolution precipitation predictions between the 2020–2059 and 2060–2099 periods relative to the temperature evolution over these same periods.

5.3.1 TPSR Estimation via Generalized Extreme Value Distribution

To better understand the relationship linking UNet performance under climate change to temperature evolution, we apply the method described in Section 2.3.3 to estimate the scaling rates of annual precipitation maxima as a function of temperature anomalies (TPSRs), both for the reference data and the UNet predictions. We adopt the following methodology, detailed in several steps:

We select the annual precipitation maxima per cell over the estimation period. The maxima series are then normalized (division by the median over the period). Simultaneously, seasonal temperature anomalies, calculated for the May-to-September period of each year, are extracted to analyze the sensitivity of extreme precipitation to temperature evolution [25]. For each cell, a GEV distribution is fitted to the precipitation maxima, with parameters modeled as functions of temperature anomalies.

To reduce uncertainties related to sampling when estimating GEV parameters, we apply spatial grouping, which involves aggregating data from neighboring cells in a 3×3 grid around each analyzed cell [26]. Parameters for each cell will therefore be estimated using data from its nine neighboring cells. To avoid inconsistent parameters and rates, we do not consider cells on the edges of the domain.

To determine the functional form of the dependence of GEV parameters on temperature anomalies T , we consider the following models:

- **Location parameter** $\mu(T)$: three models:

$$\begin{aligned} M_{\mu,1} : \mu(T) &= \mu_0, \\ M_{\mu,2} : \mu(T) &= \mu_0 + \mu_1 T, \\ M_{\mu,3} : \mu(T) &= \exp(\mu_0 + \mu_1 T). \end{aligned}$$

- **Scale parameter** $\sigma(T)$: two models:

$$\begin{aligned} M_{\sigma,1} : \sigma(T) &= \sigma_0, \\ M_{\sigma,2} : \sigma(T) &= \exp(\sigma_0 + \sigma_1 T). \end{aligned}$$

- **Shape parameter**: independent of T as it is complex to estimate:

$$\xi(T) = \xi.$$

Parameters are estimated by maximum likelihood using the L-BFGS-B optimization algorithm (Limited-memory BFGS with Bounds constraints) to control the shape parameter limits: $\xi \in (-0.5, 0.5)$ and ensure the positivity of the scale parameter: $\sigma(T) > 0$ [98]. Let y_1, \dots, y_n be the normalized precipitation maxima for a given cell, t_1, \dots, t_n the temperature anomalies, and $\mu(T), \sigma(T)$, and ξ the parameters; the GEV log-likelihood is given by:

$$\ell(\mu, \sigma, \xi) = \begin{cases} -\sum_{i=1}^n \log \sigma(t_i) - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^n \log \left(1 + \xi \frac{y_i - \mu(t_i)}{\sigma(t_i)}\right) \\ -\sum_{i=1}^n \left(1 + \xi \frac{y_i - \mu(t_i)}{\sigma(t_i)}\right)^{-1/\xi} & \text{if } \xi \neq 0, \\ -\sum_{i=1}^n \log \sigma(t_i) - \sum_{i=1}^n \frac{y_i - \mu(t_i)}{\sigma(t_i)} - \sum_{i=1}^n \exp\left(\frac{y_i - \mu(t_i)}{\sigma(t_i)}\right) & \text{si } \xi = 0. \end{cases} \quad (5.9)$$

defined on the support $\forall i : 1 + \xi \frac{y_i - \mu(t_i)}{\sigma(t_i)} > 0$. We use the JAX library to automatically calculate the Jacobian of $\ell(\mu, \sigma, \xi)$ and parallelize the cell-wise optimization. Table 5.7 indicates the mean negative log-likelihood (NLL) for each combination of $\mu(T)$ and $\sigma(T)$ models. The results confirm the relevance of integrating temperature anomalies, with the $M_{\mu,1} + M_{\sigma,1}$ combination yielding the highest NLL. We retain the $M_{\mu,3} + M_{\sigma,2}$ combination.

Once parameters are estimated for each cell, we calculate the TPSRs from the fitted GEV distribution. To do this, we calculate the mean temperature anomalies $T^{(1)} = (1/n) \sum_{i=1}^n t_i$, $T^{(2)} = T_1 + 1$, and the associated level q quantiles, $z_q^{(1)}$ and $z_q^{(2)}$, using the GEV quantile

Table 5.7 Mean negative log-likelihood of the GEV distributions over the domain for the 2020–2099 period for all combinations of $\mu(T)$ and $\sigma(T)$ parameter models.

Model combination	Mean NLL
$M_{\mu,1} + M_{\sigma,1}$	289.91
$M_{\mu,2} + M_{\sigma,1}$	281.67
$M_{\mu,3} + M_{\sigma,1}$	281.56
$M_{\mu,1} + M_{\sigma,2}$	285.99
$M_{\mu,2} + M_{\sigma,2}$	275.67
$M_{\mu,3} + M_{\sigma,2}$	275.53

function:

$$z_q(T) = \begin{cases} \mu(T) + \frac{\sigma(T)}{\xi} [(-\log q)^{-\xi} - 1] & \text{if } \xi \neq 0, \\ \mu(T) - \sigma(T) \log(-\log q) & \text{if } \xi = 0. \end{cases} \quad (5.10)$$

Finally, for each cell, we calculate the TPSR α (in %/ $^{\circ}\text{C}$) using the formula:

$$z_q^{(2)} = z_q^{(1)} (1 + 0.01\alpha)^{T^{(2)} - T^{(1)}}. \quad (5.11)$$

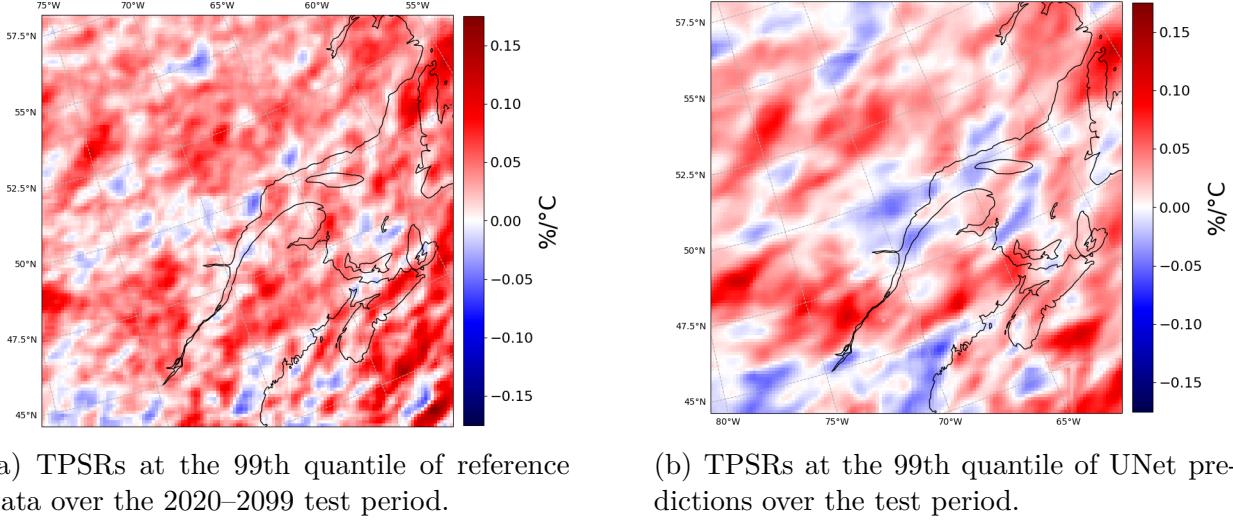
The per-cell negative log-likelihoods for the $M_{\mu,3} + M_{\sigma,2}$ combination for the reference data and the UNet predictions are available in Appendix A. Per-cell values for each parameter for both the reference data and UNet predictions are also included.

Results

Figure 5.5 presents the TPSRs evaluated at the $q = 0.99$ quantile level for the reference data (Figure 5.5a) and the UNet predictions (Figure 5.5b) for the 2020–2099 test period. It is important to note that TPSR evaluation carries high uncertainties between climate simulation members. The mean rate of 3.2%/ $^{\circ}\text{C}$ observed in the reference data here falls within the range of rates observed in other studies across all ClimEx members [25].

It is observed that the proportion of negative rates appears higher when estimated from the UNet predictions and is of greater magnitude. This observation holds for the rates of all estimated quantiles, as indicated by Table 5.8 and Figure 5.6, which present the mean and quartiles of the rates for each quantile, respectively. Nonetheless, it is noted that the spatial distribution of TPSRs is roughly similar between the reference data and the UNet predictions.

Since the rates estimated from UNet predictions are consistently lower than those obtained from the reference data, this suggests that the predicted extreme precipitation progresses



(a) TPSRs at the 99th quantile of reference data over the 2020–2099 test period.

(b) TPSRs at the 99th quantile of UNet predictions over the test period.

Figure 5.5 Comparison of TPSRs for the $q = 0.99$ quantile level estimated over the 2020–2099 test period from reference data and UNet predictions.

more slowly with rising temperatures than in reality. We conclude that the UNet’s difficulties in correctly representing extremes are amplified in a climate change context.

Table 5.8 Mean TPSRs estimated at several quantile levels for reference data and UNet predictions for the 2020–2099 test period.

Quantiles	Reference	UNet
0.5	2.24%	1.49%
0.75	2.61%	1.69%
0.90	2.87%	1.83%
0.95	3.00%	1.91%
0.99	3.20%	1.92%

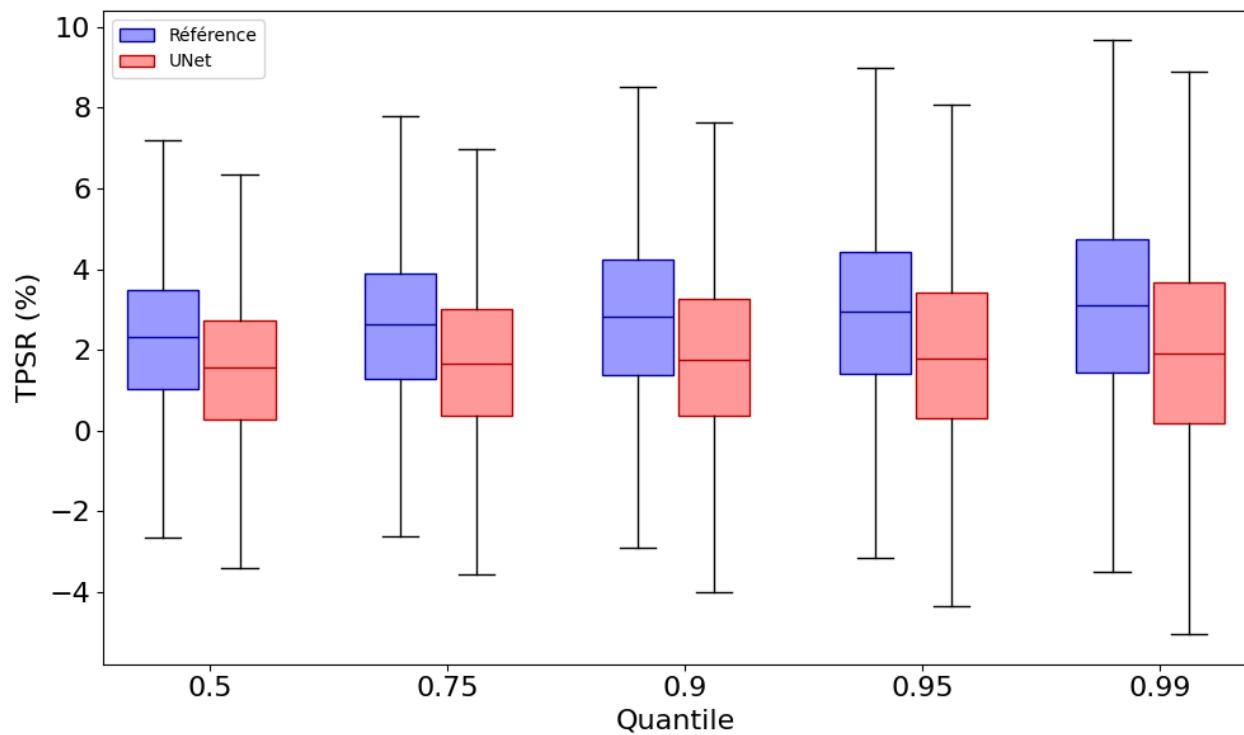


Figure 5.6 Box plots of TPSRs estimated at several quantiles for reference data and UNet predictions for the 2020–2099 test period.

CHAPTER 6 INTEGRATING THE TEMPERATURE-PRECIPITATION RELATIONSHIP FOR DOWNSCALING

In the previous chapter, we hypothesized that the degradation of UNet performance under climate change results, in part, from its inability to correctly reproduce the scaling rates of extreme precipitation relative to temperature evolution. In this chapter, we explore several approaches aimed at better integrating this relationship during the training of the deep learning model.

6.1 Soft Clausius-Clapeyron Constraint via Binning Approach

As indicated in Section 2.3.2, two methods allow for evaluating whether precipitation extremes are consistent with the temperature dependence described by the Clausius-Clapeyron relation. In this section, we focus on the first approach, known as the "binning method," and present an attempt to integrate this method into the training of the UNet model to better incorporate the relationship between temperature and precipitation.

6.1.1 Integrating the Precipitation-Temperature Relationship via Soft Constraints

Unlike saturation vapor pressure, whose temperature dependence is exactly described by the Clausius-Clapeyron relation, extreme precipitation does not strictly obey this law. The Clausius-Clapeyron relation for precipitation constitutes a statistical approximation observed under certain conditions rather than a deterministic physical constraint. Therefore, it is not appropriate to impose an exact or "hard" physical constraint on the UNet. Nonetheless, to better integrate the existing dependency between precipitation extremes and temperature, one can seek to "guide" the UNet toward equilibrium via a penalty in its loss function—i.e., a "soft" constraint.

Let $\mathcal{C} : Y \in \mathbb{R}^{N \times 2 \times H \times W} \rightarrow c \in \mathbb{R}^{H \times W}$ be the function that calculates a field of rates c (in $\%/\text{ }^{\circ}\text{C}$) for the evolution of precipitation extremes as a function of temperature for a sequence of daily precipitation and temperature fields Y . In the context of mini-batch UNet training, we consider a mini-batch of N precipitation fields interpolated from low-resolution data $x \in \mathbb{R}^{N \times h \times w}$, the high-resolution target residual fields $r \in \mathbb{R}^{N \times H \times W}$, the high-resolution temperature fields $T \in \mathbb{R}^{N \times H \times W}$, the UNet f_{θ} parameterized by θ , and $\gamma \geq 0$ as the penalty

weight. The loss function with the soft binning-based constraint is given by:

$$\mathcal{L}_\theta(x, T, r) = \mathcal{L}_{\text{reco}}(x, T, r, \theta) + \gamma \mathcal{L}_{CC}(x, T, r, \theta), \quad (6.1)$$

$$\text{where } \mathcal{L}_{\text{reco}}(x, T, r, \theta) = \frac{1}{N H W} \sum_{n=1}^N \sum_{i=1}^H \sum_{j=1}^W (f_\theta(x_n, T_n)_{i,j} - y_{n,i,j})^2, \quad (6.2)$$

$$\text{and } \mathcal{L}_{CC}(x, T, r, \theta) = \frac{1}{H W} \sum_{i=1}^H \sum_{j=1}^W \{\mathcal{C}(f_\theta(x, T)_{.,i,j} + r_{.,i,j}, T_{.,i,j}) - \mathcal{C}(x_{.,i,j} + r_{.,i,j}, T_{.,i,j})\}^2. \quad (6.3)$$

The function \mathcal{C} encompasses the various steps of the "binning" approach. Given a series of N daily precipitation values $y_{1:N}$ (in mm/day) and temperatures $t_{1:N}$ (in $^\circ\text{C}$):

1. **Filtering low intensities:** Retain only the days where precipitation exceeds a threshold s :

$$\mathcal{S} = \{n^* \in \{1, \dots, N\} \mid y_{n^*} > s\}. \quad (6.4)$$

2. **Sorting and Binning:** Order the indices $n^* \in \mathcal{S}$ by increasing temperature t_l , then divide the resulting list into D intervals $\{G_1, \dots, G_D\}$ of constant capacity m :

$$G_d = \{n_d^{*(1)}, \dots, n_d^{*(m)}\}, \quad \text{where } m \approx \frac{|\mathcal{S}|}{D}. \quad (6.5)$$

3. **Calculating group statistics:** For each group $d \in \{1, \dots, D\}$:

$$\bar{t}_d = \frac{1}{m} \sum_{n^* \in G_d} t_{n^*}, \quad y_d^q = \text{quantile}_q(\{y_{n^*} : n^* \in G_d\}). \quad (6.6)$$

4. **Regression to estimate dependency:** A log-linear relationship is assumed:

$$\log y_d^q = \alpha + \beta \bar{t}_d + \varepsilon_d. \quad (6.7)$$

The least-squares estimator for β is given by:

$$\hat{\beta} = (T_D^T T_D)^{-1} T_D^T Q_y, \quad (6.8)$$

where

$$T_D = [\bar{t}_1, \dots, \bar{t}_D], \quad Q_y = [\log y_1^q, \dots, \log y_D^q]. \quad (6.9)$$

5. Rate Calculation: Finally, the rate c is calculated using the following formula:

$$c = \exp(\hat{\beta}) - 1. \quad (6.10)$$

Having observed no strong sensitivity to the number of intervals D , we fix $D = 10$ to obtain sufficiently high interval capacities. Likewise, observing no strong sensitivity to quantile levels among $q \in \{0.90, 0.95, 0.99\}$, we fix $q = 0.90$ to reduce estimation errors when m (the number of data points per interval) is small. To ensure that each UNet prediction mini-batch contains enough values above the threshold s to fill $D = 10$ bins, the model is first trained using $\mathcal{L}_{\text{reco}}$ for two epochs. Results and the impact of hyperparameters s and γ are presented in Section 7.2.1.

6.1.2 Limitations of the Binning Approach

The function \mathcal{C} , used to estimate the dependency between temperature and extreme precipitation, is defined based on binning that requires a sufficient number of observations to produce stable rates. In our context, the input to \mathcal{C} consists of the (y_n, t_n) pairs belonging to the training mini-batch of N pairs $\{(y_n, t_n)\}_{n=1}^N$.

However, to obtain a robust estimation of the rates from this procedure, the mini-batch size N must be relatively large. This constraint implies bulky mini-batches, significantly increasing the RAM required during network training and limiting the method's application for networks like the UNet.

To illustrate this limitation, we restricted the spatial domain to 32×32 cells to free up memory and increase the training mini-batch size. Figure 6.1 presents the evolution of \mathcal{L}_{CC} as a function of mini-batch size, clearly showing the necessity of setting $N > 512$ for the errors to converge toward zero. Unfortunately, the resources at our disposal (though significant) only allow us to increase the mini-batch size up to $N = 150$ for the full domain.

An unexplored avenue in this thesis would be to train the UNet model on overlapping spatial "patches" of the domain. This approach would reduce the weight of the spatial dimensions of the mini-batches, thus allowing for an increased number of samples [99].

6.2 Integrating Statistical Priors from Extreme Value Theory

As explained in Section 2.3.3, the binning approach has the disadvantage of not being able to truly measure the causal link between temperature evolution and precipitation extremes, as seasonal variations introduce a significant confounding factor [24].

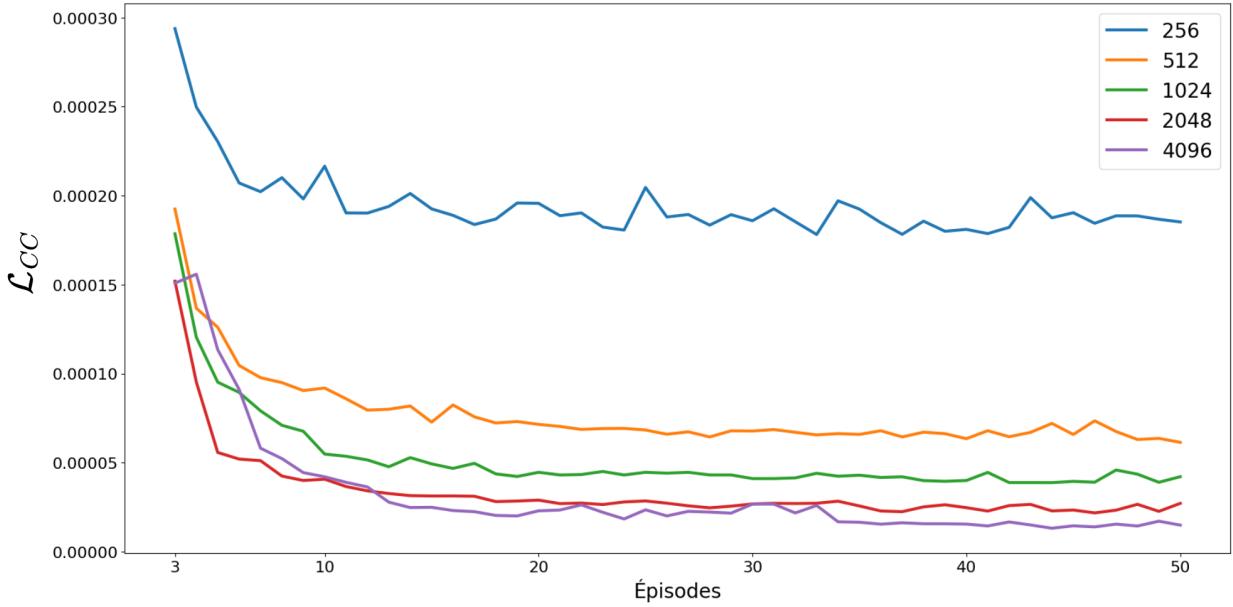


Figure 6.1 \mathcal{L}_{CC} for the 1955–2020 period on a restricted domain as a function of increasing mini-batch sizes (256, 512, 1024, 2048, and 4096).

Based on extreme value theory, alternative approaches model extreme precipitation by capturing non-stationarity via parameters that depend on annual or seasonal temperature anomalies. In this section, we study how to leverage these distributions as statistical priors to strengthen the robustness of the deep learning model in a climate change context.

6.2.1 Modeling Precipitation via the Extended Generalized Pareto Distribution

The Extended Generalized Pareto Distribution (ExtGPD), presented in Section 2.2.3, allows for modeling the entire range of precipitation (excluding zero values) while preserving the asymptotic properties of extreme value theory [1]. Our objective is to obtain a precise precipitation distribution (dependent on temperature anomalies) for each cell in our domain using our training data, and then to use this distribution during deep learning-based downscaling.

The cumulative distribution function (CDF) of a random variable following an ExtGPD is given by:

$$F(x) = K \left\{ H_\xi \left(\frac{x}{\sigma} \right) \right\}, \quad (6.11)$$

where K is a parametric model subject to certain conditions, H_ξ is the CDF of the generalized Pareto distribution (see eq. 2.8) with $\xi \geq 0$, and $\sigma > 0$ is the scale parameter. We use the model proposed by the authors: $K(v) = v^\kappa$ with $\kappa > 0$. Consequently, the probability density

function (PDF) is written as:

$$f(x) = \frac{\partial}{\partial x} F(x) = \frac{\partial}{\partial x} H_\xi \left(\frac{x}{\sigma} \right)^\kappa = \kappa H_\xi \left(\frac{x}{\sigma} \right)^{\kappa-1} h_\xi \left(\frac{x}{\sigma} \right) \sigma^{-1}, \quad (6.12)$$

where H_ξ and h_ξ are the CDF and PDF of the generalized Pareto distribution, respectively:

$$H_\xi(z) = \begin{cases} 1 - (1 + \xi z)^{-1/\xi} & \text{if } \xi \neq 0, \\ 1 - \exp(-z) & \text{if } \xi = 0, \end{cases} \quad (6.13)$$

$$h_\xi(z) = \begin{cases} (1 + \xi z)^{-(\xi+1)/\xi} & \text{if } \xi \neq 0, \\ \exp(-z) & \text{if } \xi = 0. \end{cases} \quad (6.14)$$

We must estimate three parameters for each cell in the domain: $\kappa > 0$, $\sigma > 0$, and $\xi \geq 0$ to fit the ExtGPD to the precipitation data. We perform parameter estimation by maximizing the left-censored log-likelihood, as suggested by the authors. The "drizzle effect," responsible for an overrepresentation of low precipitation values, is accounted for by applying a threshold of 1 mm/day [100]. Given our choice of $K(v) = v^\kappa$, with F as the ExtGPD CDF and f as the PDF, the log-likelihood of the ExtGPD parameters for threshold $C = 1$ mm/day and $y_1, \dots, y_n \stackrel{\text{i.i.d.}}{\sim} F$ (a series of non-zero precipitation) is written as:

$$\ell_C(\kappa, \sigma, \xi) = \log \left\{ \prod_{i:y_i < C} F(C) \prod_{i:y_i \geq C} f(y_i) \right\} \quad (6.15)$$

$$= \kappa \sum_{i:y_i < C} \log \left\{ H_\xi \left(\frac{C}{\sigma} \right) \right\} + \sum_{i:y_i \geq C} \log \kappa + (\kappa - 1) \sum_{i:y_i \geq C} \log \left\{ H_\xi \left(\frac{y_i}{\sigma} \right) \right\} \quad (6.16)$$

$$+ \sum_{i:y_i \geq C} \log \left\{ h_\xi \left(\frac{y_i}{\sigma} \right) \right\} - \sum_{i:y_i \geq C} \log \sigma. \quad (6.17)$$

Thus, for $\xi = 0$, we obtain:

$$\ell_C(\kappa, \sigma, \xi = 0) = \kappa \sum_{i:y_i < C} \log \left\{ 1 - \exp \left(-\frac{C}{\sigma} \right) \right\} + (\kappa - 1) \sum_{i:y_i \geq C} \log \left\{ 1 - \exp \left(-\frac{y_i}{\sigma} \right) \right\} \quad (6.18)$$

$$+ \sum_{i:y_i \geq C} \log \kappa - \sum_{i:y_i \geq C} \frac{y_i}{\sigma} - \sum_{i:y_i \geq C} \log \sigma, \quad (6.19)$$

and for $\xi > 0$:

$$\ell_C(\kappa, \sigma, \xi > 0) = \kappa \sum_{i:y_i < C} \log \left\{ 1 - \left(1 + \xi \frac{C}{\sigma} \right)^{-1/\xi} \right\} + (\kappa - 1) \sum_{i:y_i \geq C} \log \left\{ 1 - \left(1 + \xi \frac{y_i}{\sigma} \right)^{-1/\xi} \right\} \quad (6.20)$$

$$+ \sum_{i:y_i \geq C} \log \kappa - (\xi + 1)/\xi \sum_{i:y_i \geq C} \log \left\{ 1 + \xi \frac{y_i}{\sigma} \right\} - \sum_{i:y_i \geq C} \log \sigma. \quad (6.21)$$

To estimate the parameters of the distribution by maximizing the log-likelihood defined by Equation (6.17), we use the L-BFGS optimization algorithm. To ensure compliance with parameter constraints, we apply appropriate transformations to the parameters. This approach proved more stable than the direct use of the L-BFGS-B algorithm, which imposes bounds but showed instabilities when fitting this distribution.

To include non-stationarity linked to temperature evolution, we choose to make the parameters σ and κ dependent on annual temperature anomalies. The parameters κ , σ , and ξ for temperature anomalies T are defined during optimization by the following transformations:

$$\kappa(T) = \exp(\alpha_0 + \alpha_1 T) \quad (6.22)$$

$$\sigma(T) = \exp(\beta_0 + \beta_1 T) \quad (6.23)$$

$$\xi(T) = 0.5/\{1 + \exp(-\nu)\} \quad (6.24)$$

where $\alpha_0, \alpha_1, \beta_0, \beta_1, \nu \in \mathbb{R}$. To prevent parameters κ and σ from diverging toward unrealistic values during optimization, we add a penalty term of 10^6 to the negative log-likelihood for cases where $\kappa(T) > 2$ or $\sigma(T) > 30$. These thresholds were defined empirically based on the maximum bounds observed during previous estimations with explicit constraints. As a reminder, we fix $C = 1$ mm/day.

The mean negative log-likelihood (NLL) obtained in the stationary case ($\alpha_1 = \beta_1 = 0$) is 43,784.86, compared to 43,778.69 in the non-stationary case. Figure 6.2 presents the spatial distribution of the negative log-likelihood in the non-stationary case. High values are observed in regions with significant orography, notably the Laurentians, the Chic-Choc Mountains, the Cape Breton Plateau, and the Long Range Mountains. The NLL is also high in the southeastern part of the domain—an oceanic zone characterized by heavy precipitation, as shown in Figure 4.1(c).

Figures 6.3 and 6.4 present the spatial distribution of the parameters obtained in the non-stationary case. Figure 6.3a shows that the stationary parameter $\kappa(T)$ is higher over continental areas, particularly in mountainous regions. This suggests more spread-out distributions

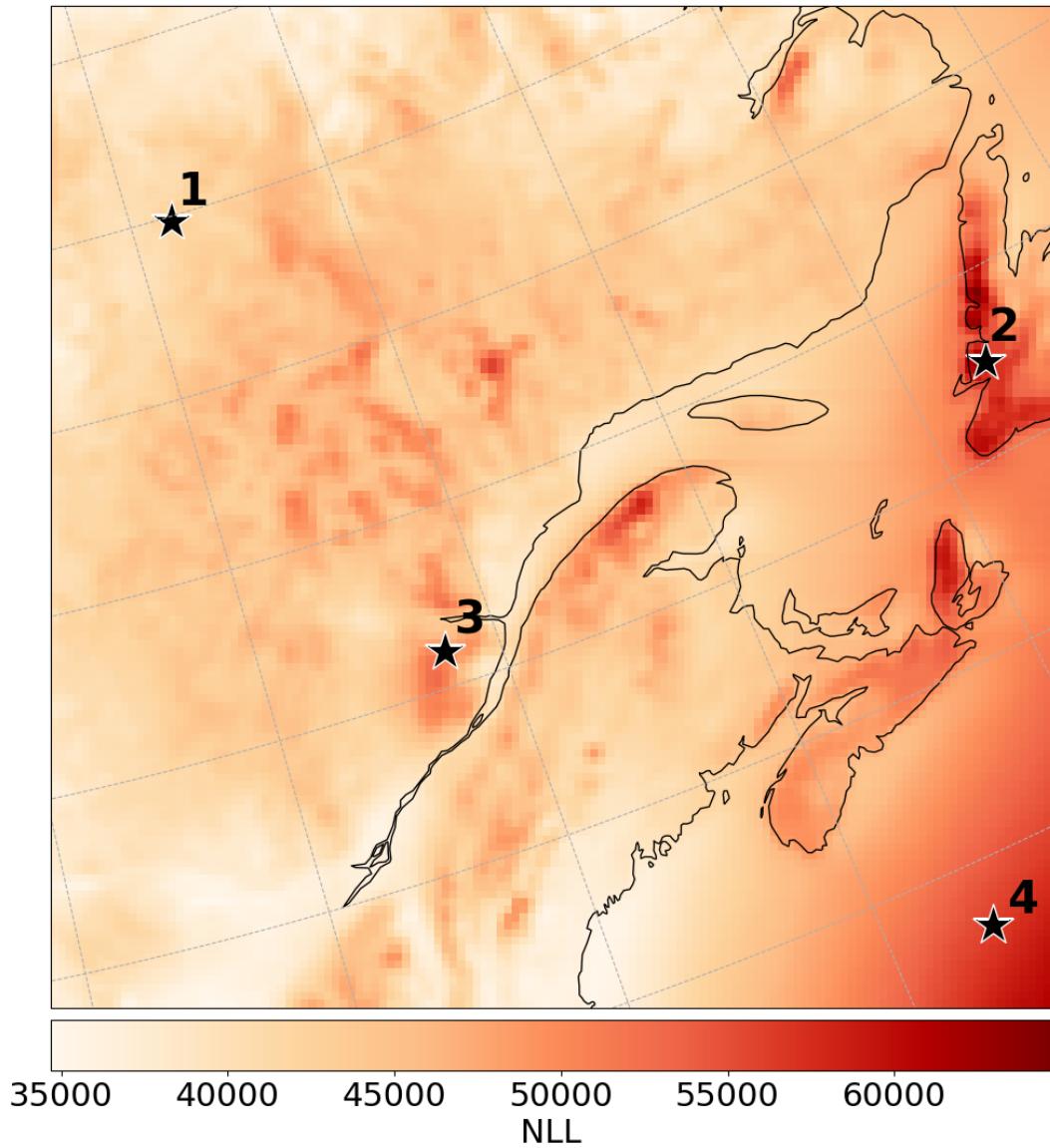


Figure 6.2 Negative log-likelihood obtained after optimizing the parameters defined in (6.22–6.24) for the ExtGPD distribution, fitted independently for each cell in the domain using reference data covering the 1955–2020 period. The numbered stars 1 to 4 identify the cells analyzed in Figures 6.5 and 6.6.

with reduced density around low precipitation intensities. Figure 6.3b shows that the stationary parameter $\sigma(T)$ clearly distinguishes between continental and oceanic zones, with the latter associated with higher scale parameter values.

Regarding the non-stationary components shown in Figures 6.3c and 6.3d, we observe an opposite dynamic over maritime zones, with an increase in the scale parameter $\sigma(T)$ and a decrease in $\kappa(T)$. Furthermore, the coefficient β_1 is generally positive across the entire domain, indicating an increasing response of extreme precipitation to rising temperatures. Finally, Figure 6.4 reveals that the shape parameter ξ is higher over continental areas, reflecting heavier distribution tails.

Four cells, representative of the diversity of the geographic domain, are identified by stars 1 to 4 in Figure

These Q-Q plots and corresponding histograms illustrate the importance of accounting for temperature anomaly evolution. In the stationary case ($\alpha_1 = \beta_1 = 0$) shown in Figure 6.5, we observe a systematic underestimation of high quantiles (over the 2020–2099 period) by the distribution fitted over 1955–2020. Conversely, Figure 6.6 shows nearly perfect Q-Q plots and well-aligned histograms. Thus, including temperature anomalies allows for a more faithful adaptation of non-zero precipitation distributions in a climate change context.

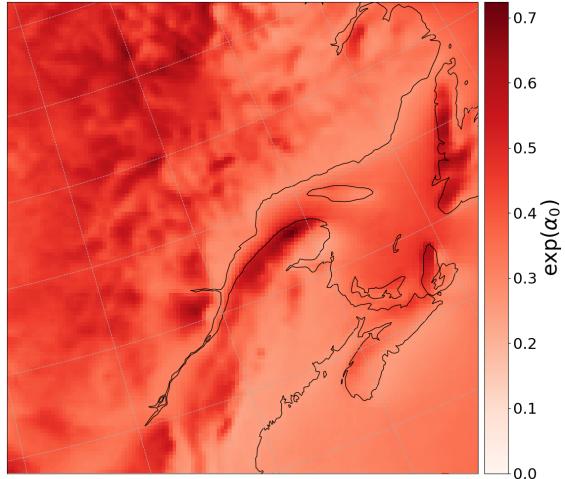
6.2.2 Downscaling via Quantile Level Prediction

Unlike the classic downscaling approach—which involves training a deep learning model to directly predict the high-resolution precipitation value (or, in our case, the residual relative to the low resolution)—we propose a method based on the statistical structure of high-resolution precipitation. Specifically, for each cell, we leverage the previously fitted ExtGPD.

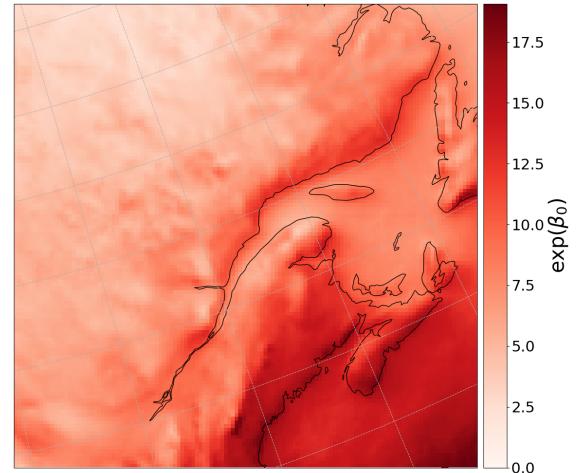
Let x and y be two precipitation fields at low and high resolution, respectively, associated with a field of seasonal temperature anomalies tp . For each cell (i, j) of the high-resolution grid, the downscaling model aims to predict the quantile level $\hat{p}_{ij}(x) \in (0, 1)$ such that \hat{y}_{ij} (the corresponding precipitation value) is obtained by inverting the CDF F_{ij} , which is dependent on temperature anomalies and estimated from the fitted ExtGPD for that cell:

$$\hat{y}_{ij} = F_{ij}^{-1}(\hat{p}_{ij}(x), tp_{i,j}). \quad (6.25)$$

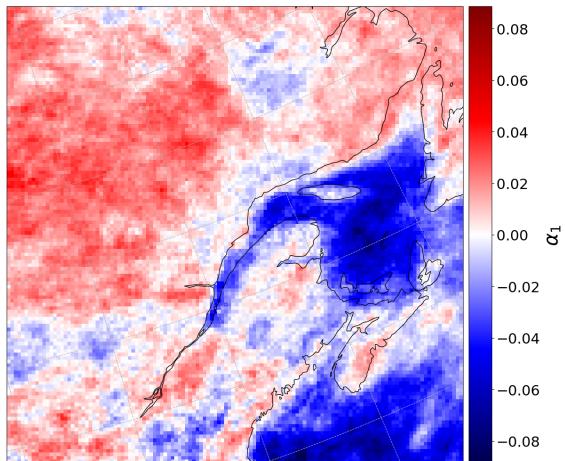
This approach integrates extreme value theory properties to improve the representation of high-intensity precipitation. Furthermore, by making the CDF $F_{i,j}$ dependent on annual temperature anomalies, we strengthen the model's robustness under climate change. The



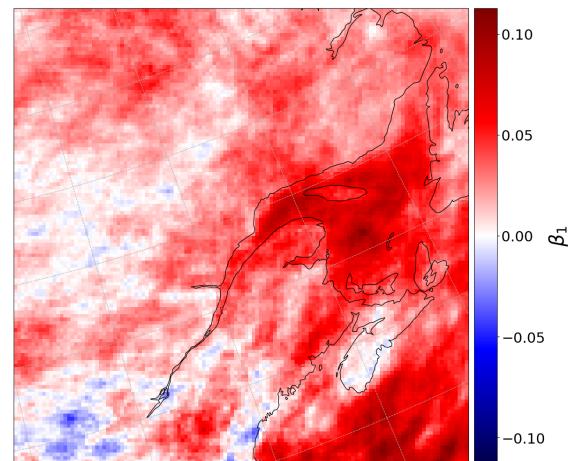
(a) Values of the $\exp(\alpha_0)$ parameter corresponding to the stationary part of the $\kappa(T)$ parameter.



(b) Values of $\exp(\beta_0)$ corresponding to the stationary part of the $\sigma(T)$ parameter.



(c) Values of α_1 corresponding to the non-stationary part of the $\kappa(T)$ parameter.



(d) Values of β_1 corresponding to the non-stationary part of the $\sigma(T)$ parameter.

Figure 6.3 Values of the $\exp(\alpha_0)$, α_1 , $\exp(\beta_0)$, β_1 parameters for each cell in the domain after maximizing the log-likelihood (6.17).

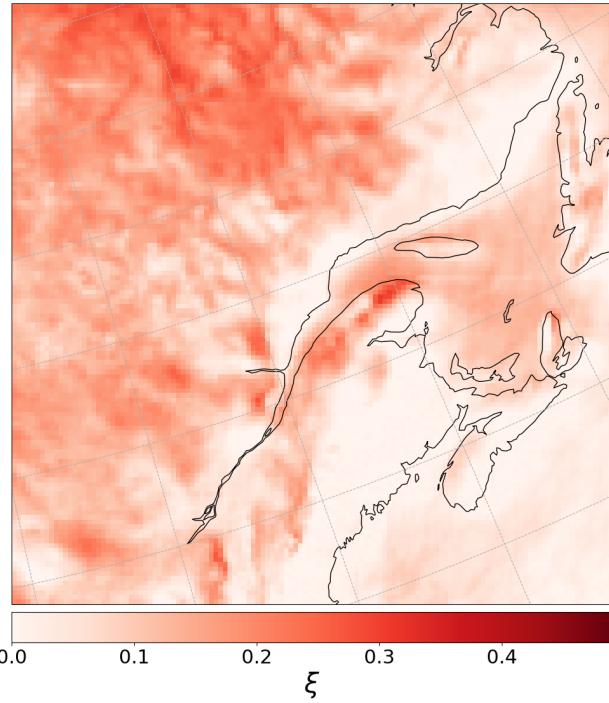


Figure 6.4 Values of the ξ parameter after applying transformation (6.24).

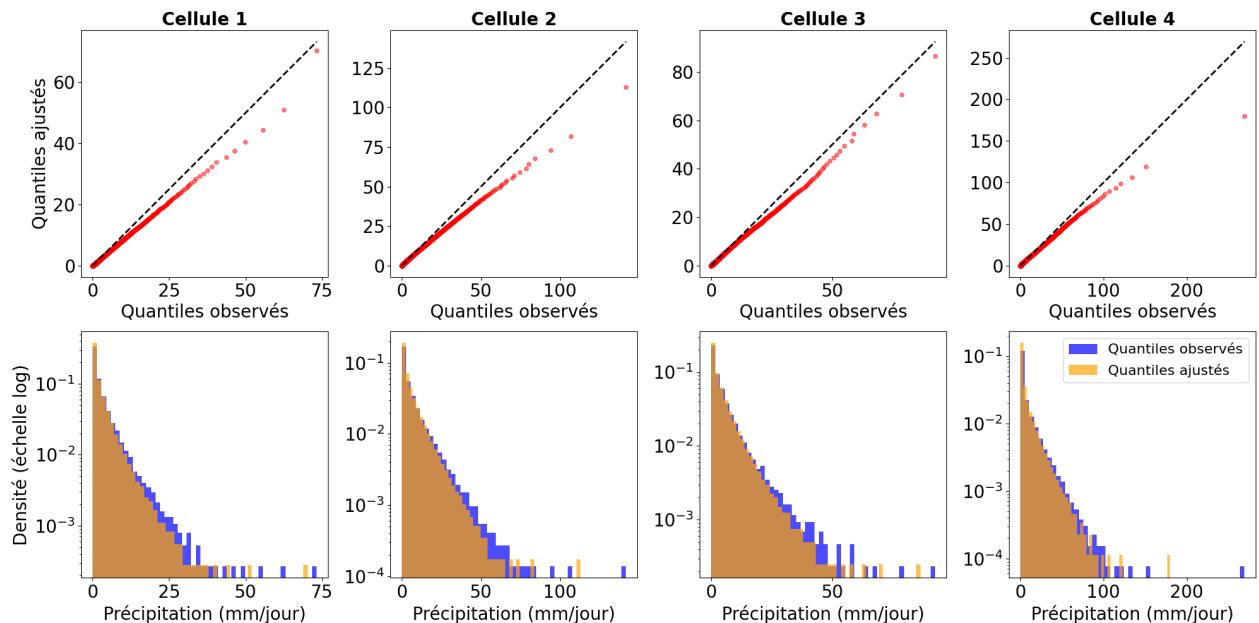


Figure 6.5 **(First row)** Q-Q plots of data from the 2020–2099 period for cells 1 to 4 (see Figure 6.2) for the stationary models ($\alpha_1 = \beta_1 = 0$) fitted over the 1955–2020 period. **(Second row)** Histograms of the resulting quantiles.

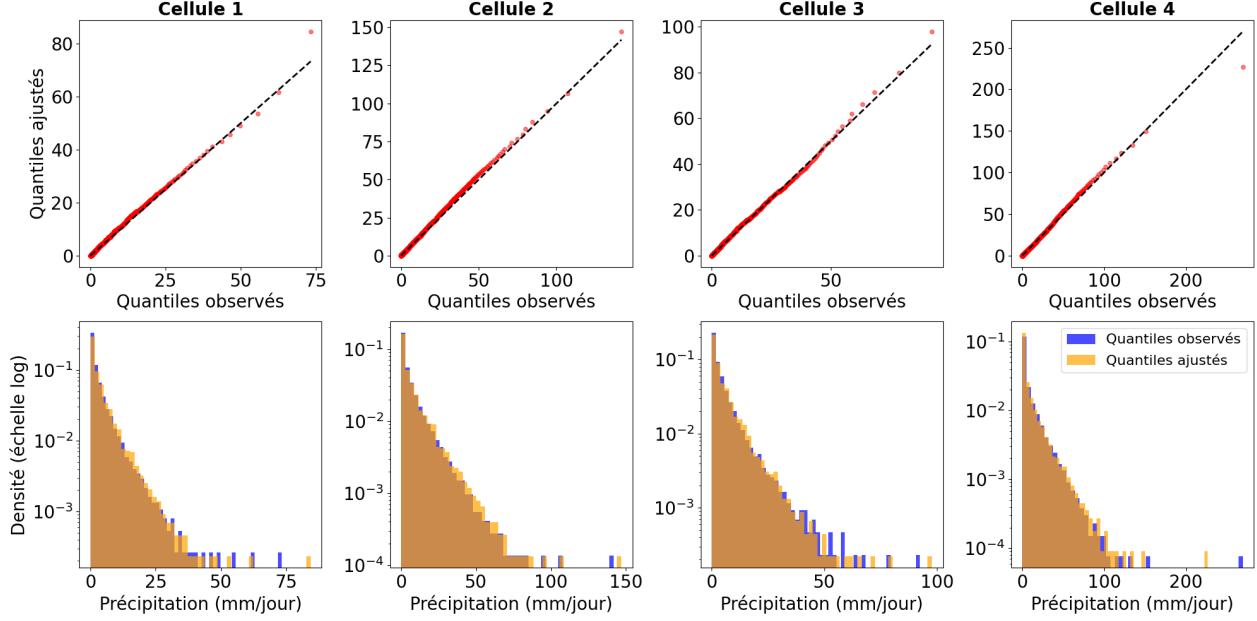


Figure 6.6 (**First row**) Q-Q plots of data from the 2020–2099 period for cells 1 to 4 (see Figure 6.2) for the non-stationary models fitted over the 1955–2020 period. (**Second row**) Histograms of the resulting quantiles.

inverse CDF of the ExtGPD can be calculated using the formula:

$$F^{-1}(p) = \begin{cases} \frac{\sigma}{\xi} \left[\left(1 - p^{1/\kappa}\right)^{-\xi} - 1 \right] & \text{if } \xi > 0, \\ -\sigma \log(1 - p^{1/\kappa}) & \text{if } \xi = 0. \end{cases} \quad (6.26)$$

The entire predicted high-resolution field \hat{y} is then reconstructed by applying this operation to each cell of the \mathcal{G}_{HR} grid:

$$\hat{y} = \left\{ \hat{y}_{ij} = F_{ij}^{-1}(\hat{p}_{ij}(x), tp_{i,j}) \mid (i, j) \in \mathcal{G}_{HR} \right\}. \quad (6.27)$$

Since the quantile levels associated with the fitted ExtGPDs are not uniformly distributed (they are largely concentrated toward low values near zero), we choose to train the f_θ model to predict a logarithmic transformation of the quantile level. Specifically, if $p_{i,j} = F(y_{i,j}, tp_{i,j})$ denotes the quantile level corresponding to the target value $y_{i,j}$ according to the temperature-dependent ExtGPD CDF F , then the model is trained to approximate:

$$f_\theta(x)_{i,j} = \log(1 - p_{i,j}) \implies \hat{p}_{i,j} = 1 - \exp(-f_\theta(x)_{i,j}) \quad (6.28)$$

where x denotes the interpolated low-resolution input, $f_\theta(x)_{i,j}$ the model output for cell (i, j) , and $\hat{p}_{i,j}$ the predicted quantile level.

Let $Y \in \mathbb{R}^{N \times H \times W}$ be a mini-batch of N high-resolution daily reference precipitation fields, $X \in \mathbb{R}^{N \times H \times W}$ a corresponding mini-batch interpolated from low-resolution data, and $T \in \mathbb{R}^{N \times H \times W}$ the mini-batch of high-resolution temperature anomalies. Let $\kappa, \sigma_0, \sigma_1, \xi \in \mathbb{R}^{H \times W}$ be the parameter fields pre-fitted for each grid cell, and $F(\cdot, T_{i,j})$ the temperature-dependent ExtGPD CDF for cell (i, j) .

We then train the UNet by minimizing the following loss function:

$$\mathcal{L}_{\text{ExtGPD}_1}(\theta) = \frac{1}{NHW} \sum_{n=1}^N \sum_{h=1}^H \sum_{w=1}^W \left(f_\theta(X_n)_{h,w} - \log(1 - F(Y_{n,h,w}, T_{n,h,w})) \right)^2, \quad (6.29)$$

6.2.3 Learning via Quantile Level Alignment

Another approach involves using the precipitation prediction $\hat{Y}_{n,i,j} = f_\theta(X_n)_{i,j}$ produced by the UNet, then projecting this value into the quantile level space by applying the CDF $F_{i,j}(\cdot, T_{n,i,j})$ associated with cell (i, j) . The predicted quantile level is then compared to the target quantile level $p_{n,i,j} = F(Y_{n,i,j}, T_{n,i,j})$ without applying the previously detailed logarithmic transformation, since the predicted precipitation value is not directly dependent on the quantile level:

$$\mathcal{L}_{\text{ExtGPD}_2}(\theta) = \frac{1}{NHW} \sum_{n=1}^N \sum_{h=1}^H \sum_{w=1}^W \left(F(f_\theta(X_n)_{h,w}, T_{n,h,w}) - F(Y_{n,h,w}, T_{n,h,w}) \right)^2. \quad (6.30)$$

CHAPTER 7 EVALUATION

This chapter aims to present and analyze the results of each method proposed in the previous chapter, in comparison with those of the UNet presented in Chapter 5, which was trained on the residual mean squared error loss (5.4).

7.1 Metrics

The primary goal of downscaling is to reproduce the fine structures of high-resolution precipitation fields from degraded fields that only capture large-scale structures. The downscaling model must also be capable of generalizing within a climate change context as mean temperatures rise.

Achieving this objective requires the simultaneous satisfaction of several performance criteria:

1. **Reproduction of local intensities:** The model must be able to faithfully reproduce high-resolution precipitation values as well as the extent of precipitation events.
2. **Reconstruction of statistical distribution:** It is essential that the model can reconstruct the high-resolution precipitation distribution, particularly by accurately representing extreme values. This implies a correct estimation of the distribution tails.
3. **Frequency fidelity:** The reconstructed fields must contain the correct amount of information at each spatial scale. This assumes an accurate reconstruction of precipitation structures across a range of spatial frequencies with minimal distortion.

To compare the models' ability to satisfy criterion (1), we measure the Mean Absolute Error (MAE) (expressed in mm/day) between the downscaled precipitation fields and the reference data. Predicting high-resolution extreme precipitation represents a particular challenge for models. To more finely evaluate performance in the extreme regions of the distribution, we also calculate the MAE in the vicinity of several quantile levels, following this methodology:

1. Quantiles of levels $q \pm 0.025$ are estimated on the reference dataset.
2. We identify data points from the reference set whose precipitation values fall between these two quantiles. This defines a local range around the quantile of level q .
3. The MAE is then calculated only on this restricted subset of points.

We also calculate the MAE for each year of the test period. This evaluation allows us to analyze the evolution of errors over time and compare the generalization capacity to climate change across models.

To evaluate compliance with criterion (2), we aggregate all cells in the domain and compare precipitation histograms using a logarithmic scale to focus on the distribution tail.

Finally, to evaluate compliance with criterion (3), we analyze the frequency fidelity of the reconstructed precipitation fields using the Radially Averaged Power Spectral Density (RA-PSD). This metric quantifies the distribution of the spatial variance of the field as a function of spatial frequencies, thus providing a direct measure of the models' ability to restore structures present at different scales.

For a precipitation field $Y \in \mathbb{R}_+^{H \times W}$, the RA-PSD is obtained from the two-dimensional discrete Fourier transform:

$$\hat{Y}(k_h, k_w) = \sum_{h=0}^{H-1} \sum_{w=0}^{W-1} Y_{h,w} e^{-2\pi i \left(\frac{k_h h}{H} + \frac{k_w w}{W} \right)} \quad (7.1)$$

$$\text{PSD}(k_h, k_w) = \frac{1}{H W} |\hat{Y}(k_h, k_w)|^2 \quad (7.2)$$

The two-dimensional power spectral density (PSD) is then radially averaged—i.e., by intervals of successive rings of radius $k = \sqrt{k_h^2 + k_w^2}$ —to obtain a one-dimensional PSD, denoted $\text{PSD}(k)$. The resulting one-dimensional PSD vectors for each precipitation field are then averaged, frequency by frequency, to produce a mean PSD representative of the dataset for each model.

To evaluate the frequency reconstruction quality of the models, we analyze the relative RA-PSD between the generated fields and the reference data, defined by:

$$\text{PSD}_{\text{rel}}(k) = \frac{\text{PSD}_{\text{model}}(k)}{\text{PSD}_{\text{reference}}(k)} \quad (7.3)$$

A value of $\text{PSD}_{\text{rel}}(k) = 1$ indicates a perfect restoration of variance at the corresponding spatial scale. Values below 1 reflect a loss of variance (under-representation of structures), while values above 1 reveal an overestimation.

7.2 Results

Each compared model is trained according to the same procedure and with the same training hyperparameters. The random seed is fixed at 351 for all training sessions. The optimization algorithm used is `Adam`, with its standard parameters and a learning rate of 10^{-3} . Daily precipitation fields are randomly shuffled, and the mini-batch size is fixed at 32 days, except for the binning approach variants, for which it is fixed at 150. Each model is trained for 50 epochs.

As a reminder, the training set covers the 1955–2020 period, corresponding to 23,725 data points per cell, and the test set covers the period from 2020 to 2099, totaling 28,835 data points per cell.

Bicubic interpolation is used as the baseline method, providing a lower bound for comparison to evaluate model performance. The mean absolute error of this method is presented in Table 7.1.

Table 7.1 Overall MAE and MAE evaluated in the vicinity of several quantile levels for bicubic interpolation over the test period.

MAE	MAE-q (0.5)	MAE-q (0.75)	MAE-q (0.9)	MAE-q (0.95)	MAE-q (0.99)
1.570	1.181	2.446	4.050	6.983	20.949

7.2.1 Soft Constraint via Binning Approach

We analyze the performance of the approach consisting of adding a soft constraint to the model’s loss function, intended to align the model’s predictions with the relationship between temperatures and precipitation extremes. This constraint, defined by the binning approach, is detailed in Section 6.1.

To better understand the model’s behavior when applying this constraint, we compare the results for several values of the parameter $\gamma \in \{0, 1, 10, 50, 100\}$, which modulates the intensity of the penalty applied to the loss function ($\gamma = 0$ corresponds to the quadratic loss, and $\gamma = 100$ implies a penalty of roughly the same order of magnitude as the initial loss function). We also compare the performance of the models as a function of the filtering threshold s (6.4), fixed at 0 or 1 mm/day.

Tables 7.2 and 7.3 present the MAE values over the test period for the different configurations studied. No significant performance gain is observed following the introduction of the penalty, regardless of the threshold s . On the contrary, for coefficients $\gamma > 10$, a marked degradation

in performance is noted. However, the results obtained for moderate coefficients ($\gamma \in \{1, 10\}$) are systematically better when a precipitation filtering of $s = 1$ mm/day is applied—with the exception of the median. This improvement can be explained by a concentration of learning on high intensities, as low precipitation values are excluded. Nevertheless, as this effect remains marginal, these results must be interpreted with caution and require further investigation to be consolidated.

Table 7.2 Overall MAE and MAE evaluated in the vicinity of several quantile levels, as a function of the coefficient γ of the loss function (6.1), for $s = 0.0$ mm/day (filtering threshold (6.4)).

γ	MAE	MAE-q (0.5)	MAE-q (0.75)	MAE-q (0.9)	MAE-q (0.95)	MAE-q (0.99)
0	1.329	0.918	2.273	4.045	6.744	17.151
1	1.331	0.931	2.290	4.044	6.710	17.123
10	1.328	0.944	2.255	3.939	6.624	17.306
50	1.375	0.987	2.314	4.032	6.813	17.773
100	1.393	1.010	2.315	4.028	6.810	18.073

Table 7.3 Overall MAE and MAE evaluated in the vicinity of several quantile levels, as a function of the coefficient γ of the loss function (6.1), for $s = 1.0$ mm/day (filtering threshold (6.4)).

γ	MAE	MAE-q (0.5)	MAE-q (0.75)	MAE-q (0.9)	MAE-q (0.95)	MAE-q (0.99)
0	1.329	0.918	2.273	4.045	6.744	17.151
1	1.323	0.933	2.268	3.989	6.599	17.083
10	1.321	0.937	2.260	3.908	6.503	17.376
50	1.365	1.010	2.272	3.879	6.594	18.128
100	1.391	1.037	2.297	3.948	6.730	18.308

For subsequent evaluations, we retain the configuration with $\gamma = 10$ for each threshold s , as this coefficient generally offers the best performance. The evolution of the MAE over the test period, presented in Figure 7.1, shows that the penalized models (orange and purple curves) fail to mitigate the increasing error trend over time. Furthermore, Figure 7.2 confirms the limitations mentioned in Section 6.1.2: models trained with a penalty ($\gamma > 0$) do not improve the alignment of TPSRs with those of the reference data, compared to training based solely on $\mathcal{L}_{\text{reco}}$. Finally, the division into two sub-periods suggests that these evolution rates do not exhibit dynamics sensitive to the acceleration of climate change.

The extent of precipitation predicted by the models is not improved by adding the constraint, as shown in Figure 7.3, which presents precipitation histograms for each model over the test period (with all domain cells aggregated). However, a marginal improvement in relative PSD

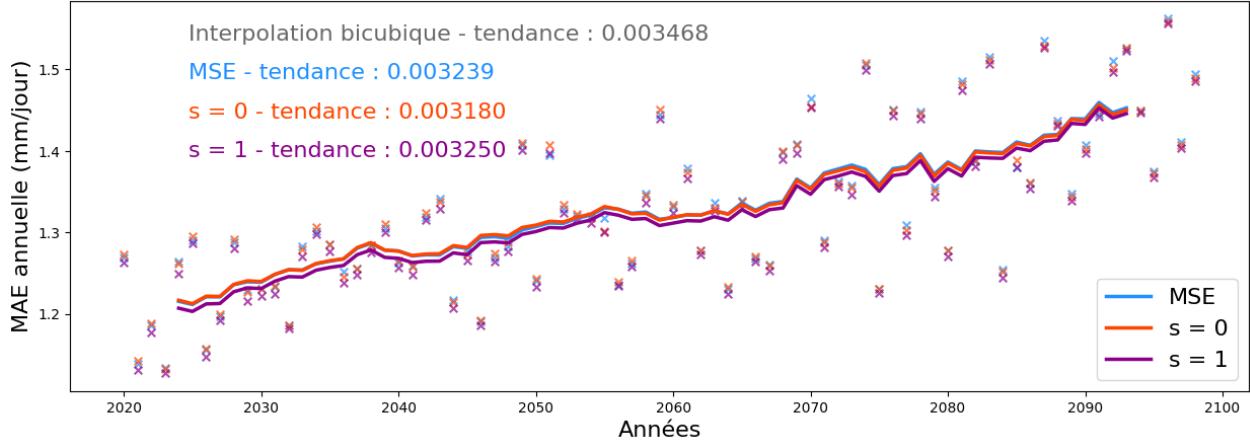


Figure 7.1 Annual MAE (in mm/day) for MSE loss ($\gamma = 0$), and for $s = 0$ or $s = 1$ mm/day with $\gamma = 10$. Bicubic interpolation error is not displayed to maintain clarity. Smoothing via a 10-value moving average.

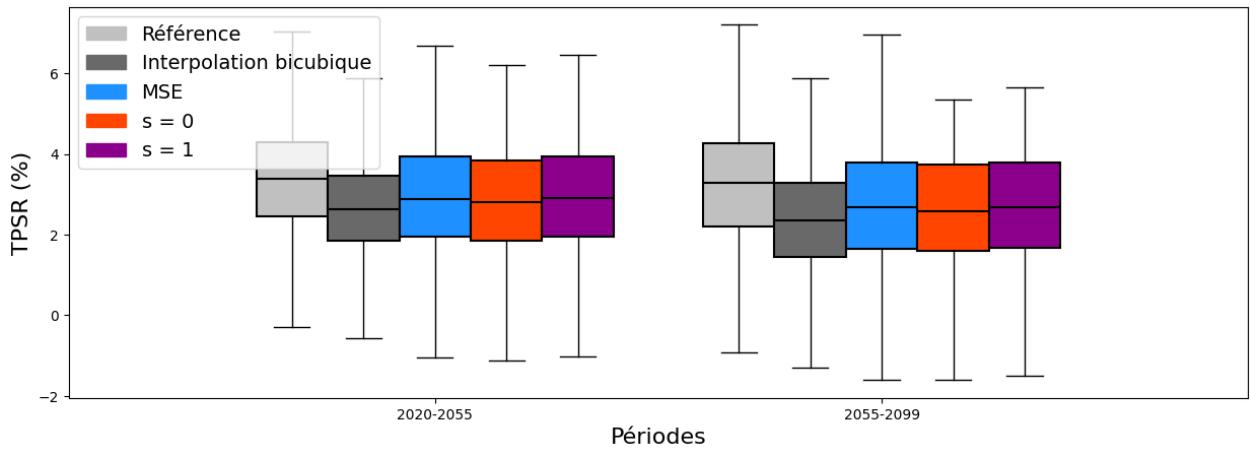


Figure 7.2 Box plots of TPSRs (in $\%/\text{ }^{\circ}\text{C}$) obtained via the method defined in Section 6.1 for each model. The test set is split into two periods: 2020–2050 and 2050–2099, to evaluate temporal evolution.

is observed for frequencies above approximately 0.047 km^{-1} , corresponding to a spatial scale of less than four cells. The drop in PSD curves at these frequencies is also observed in the reference data. It may reflect a distortion inherent to the RCM model at this scale, possibly linked to a transition between two physical schemes or a numerical filtering effect.

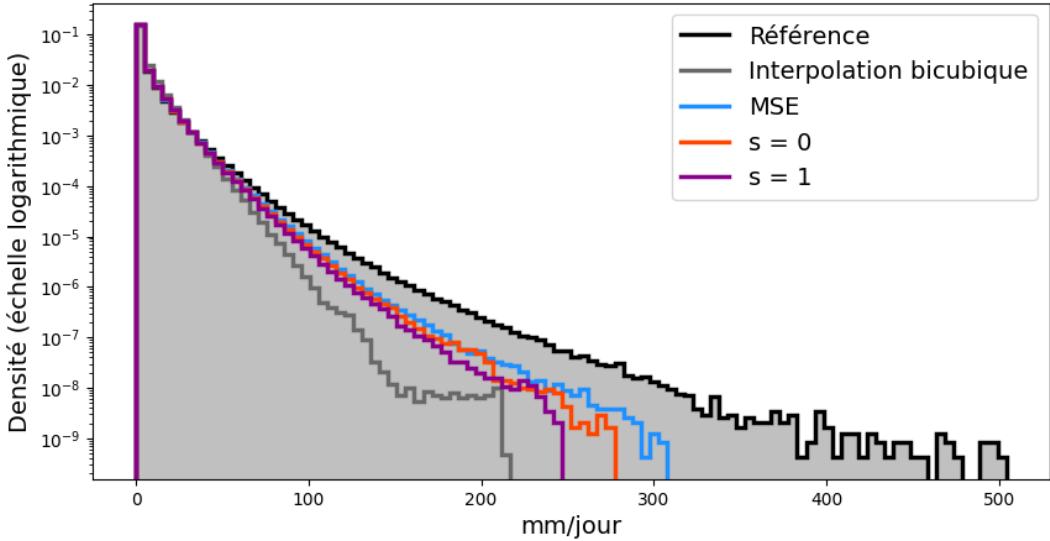


Figure 7.3 Precipitation histogram for each model, including all domain cells over the 2020–2099 test set period.

As discussed in Section 6.1.2, a batch size much larger than $N = 150$ samples would be necessary to obtain a robust TPSR estimation using the binning method. Such an estimation is indispensable for the penalty derived from this method to define an optimization landscape consistent with the stated goal: aligning the progression rates of precipitation extremes as a function of temperature with those observed in the reference data.

However, RAM limitations make it impossible to use batch sizes larger than 256 while maintaining a domain large enough to contain a diversity of information.

One solution would be to train the model on a set of overlapping sub-domains (patches) extracted from the initial domain. This approach would significantly reduce memory requirements, making it possible to train the UNet with the proposed loss function for batch sizes exceeding 512. We therefore encourage further work in this direction to truly evaluate the relevance of this approach for improving extreme prediction and climate change robustness.

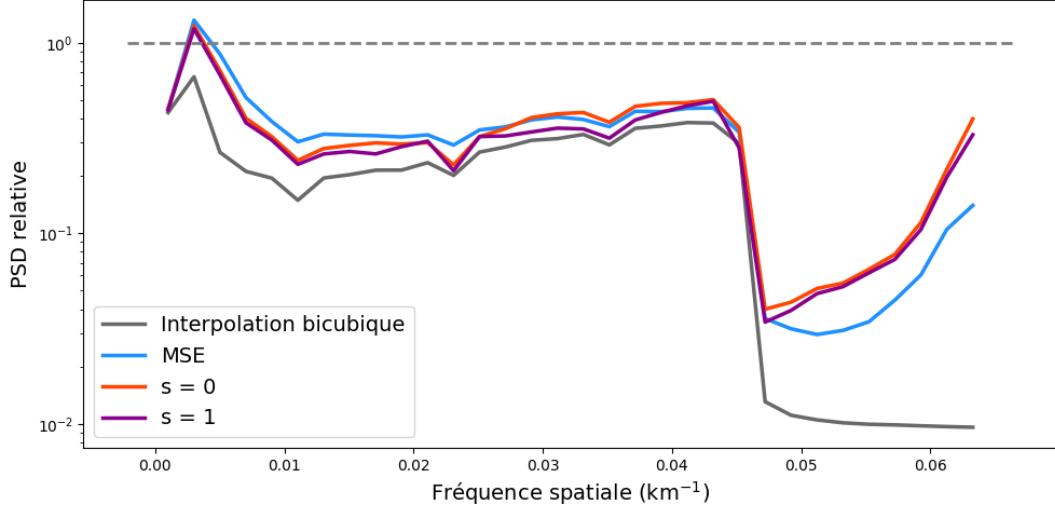


Figure 7.4 PSD_{rel} for each model relative to the reference data power spectral density, calculated according to the method detailed in Section 7.1.

7.2.2 Integrating Statistical Priors from Extreme Value Theory

We evaluate each of the proposed methods in two configurations: stationary and non-stationary, corresponding to the exclusion or inclusion of temperature anomalies in the loss function and quantile calculation, respectively. However, these anomalies are never used as explicit explanatory covariates at the model input. Indeed, their inclusion alongside low-resolution precipitation and mean daily temperature systematically led to a performance degradation during our experiments. Only one of the evaluated models has access to annual temperature anomalies during the test period evaluation: $\mathcal{L}_{\text{ExtGPD}_1}(T)$, as this variable is necessary for calculating the precipitation quantile in the non-stationary ExtGPD case.

Table 7.4 highlights the value of using the statistical prior offered by the ExtGPD for quantile levels $q \in [0.75, 0.95]$. Nevertheless, note that the median of high-resolution precipitation is systematically subject to larger errors following these approaches. Despite improvements, progress remains marginal; therefore, no definitive conclusion can be drawn regarding the benefit of these methods for improving precipitation extreme prediction.

Similarly, the inclusion of temperature anomalies—while beneficial for modeling precipitation distribution under climate change, as noted in Section 6.2.1—appears to have the opposite effect of what was expected, with a faster increase in errors compared to other models, as seen in Figure 7.5. However, this should not necessarily be considered a performance degradation. As indicated by the precipitation histogram of the $\mathcal{L}_{\text{ExtGPD}_1}(T)$ model in Figure 7.7, predicting quantile levels and including temperature anomalies allows for broadening the

Table 7.4 Overall MAE (in mm/day) and MAE evaluated in the vicinity of several quantile levels for the standard UNet $\mathcal{L}_{\text{reco}}$ and the ExtGPD quantile level prediction and alignment methods: $\mathcal{L}_{\text{ExtGPD}_1}$ (1) and $\mathcal{L}_{\text{ExtGPD}_2}$ (2), with or without temperature anomalies T .

Model	MAE	MAE (0.5)	MAE (0.75)	MAE (0.9)	MAE (0.95)	MAE (0.99)
$\mathcal{L}_{\text{reco}}$	1.329	0.918	2.273	4.045	6.744	17.151
1	1.330	0.942	2.223	3.920	6.662	17.729
$1 + T$	1.336	0.943	2.257	3.993	6.734	17.629
2	1.318	0.928	2.251	3.956	6.590	17.221
$2 + T$	1.321	0.926	2.258	3.986	6.631	17.175

range of intensities obtained under climate change. Yet, if a precipitation event containing one of these extreme intensities is not precisely located by the UNet, the model will be doubly penalized compared to another model that largely underestimated the precipitation intensities in the field.

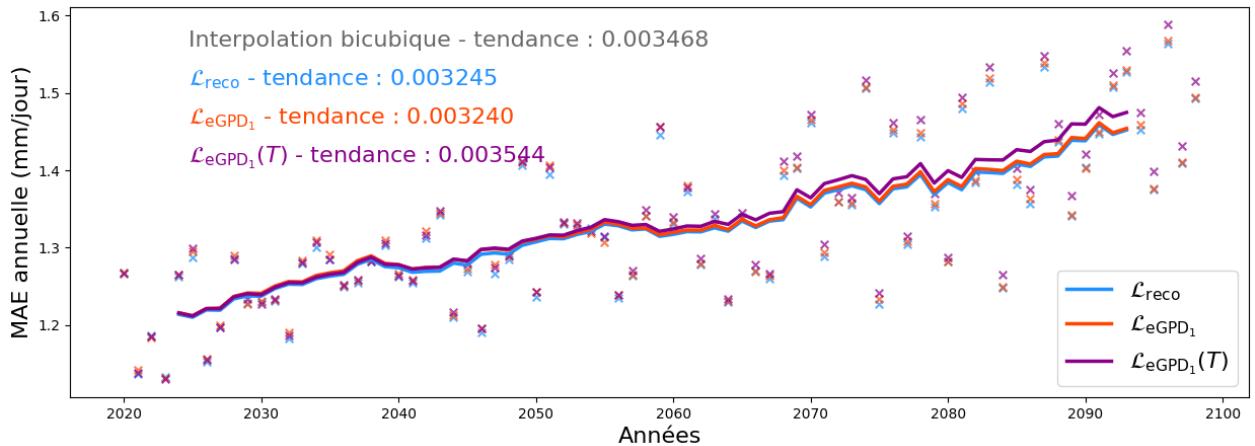


Figure 7.5 Annual MAE (in mm/day) for the standard UNet ($\mathcal{L}_{\text{reco}}$) and the stationary and non-stationary variants of the quantile level prediction method (see Section 6.2.2) of the previously fitted ExtGPDs: $\mathcal{L}_{\text{ExtGPD}_1}$ and $\mathcal{L}_{\text{ExtGPD}_1}(T)$. The curve is obtained via a 10-value moving average.

The "double penalty" thus challenges the relevance of using MAE alone to measure UNet performance under climate change, especially regarding precipitation extremes.

During the experiments conducted with the ExtGPD quantile level prediction models, we observed a systematic tendency to underestimate high quantiles, leading to simulated precipitation significantly lower than that observed in the reference data. To mitigate this effect, the logarithmic transformation was introduced in the loss function (6.29), which partially corrected this bias. The improvement in extreme precipitation representation and the better

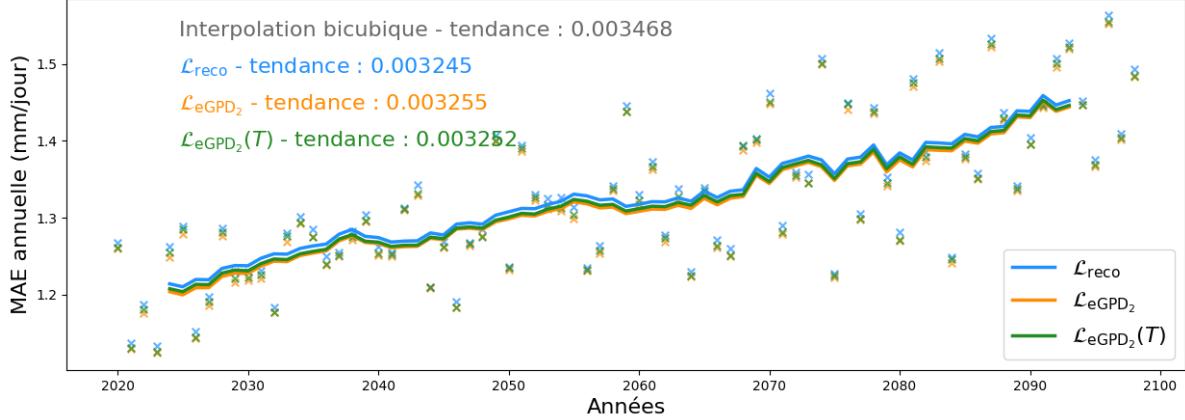


Figure 7.6 Annual MAE (in mm/day) for the standard UNet ($\mathcal{L}_{\text{reco}}$) and the stationary and non-stationary variants of the quantile level alignment method (see Section 6.2.3): $\mathcal{L}_{\text{ExtGPD}_1}$ and $\mathcal{L}_{\text{ExtGPD}_1(T)}$. The curve is obtained via a 10-value moving average.

preservation of spatial variance (see Figure 7.8) suggest that integrating ExtGPD-derived information into deep learning models for precipitation downscaling retains significant potential.

Figure 7.9 illustrates simulated precipitation fields for two distinct days of the test period, comparing the outputs of the various studied models with the high-resolution reference data from ClimEx. This visualization highlights the complexity of the precipitation downscaling task, especially without resorting to numerous covariates.

The proposed methods do not fully achieve the set objectives: (1) improving extreme precipitation downscaling, and (2) strengthening the UNet’s robustness against climate change, particularly by integrating the dependency relationship between precipitation and temperature evolution. Despite this, the results obtained remain promising and encourage further research to better exploit the potential of these approaches.

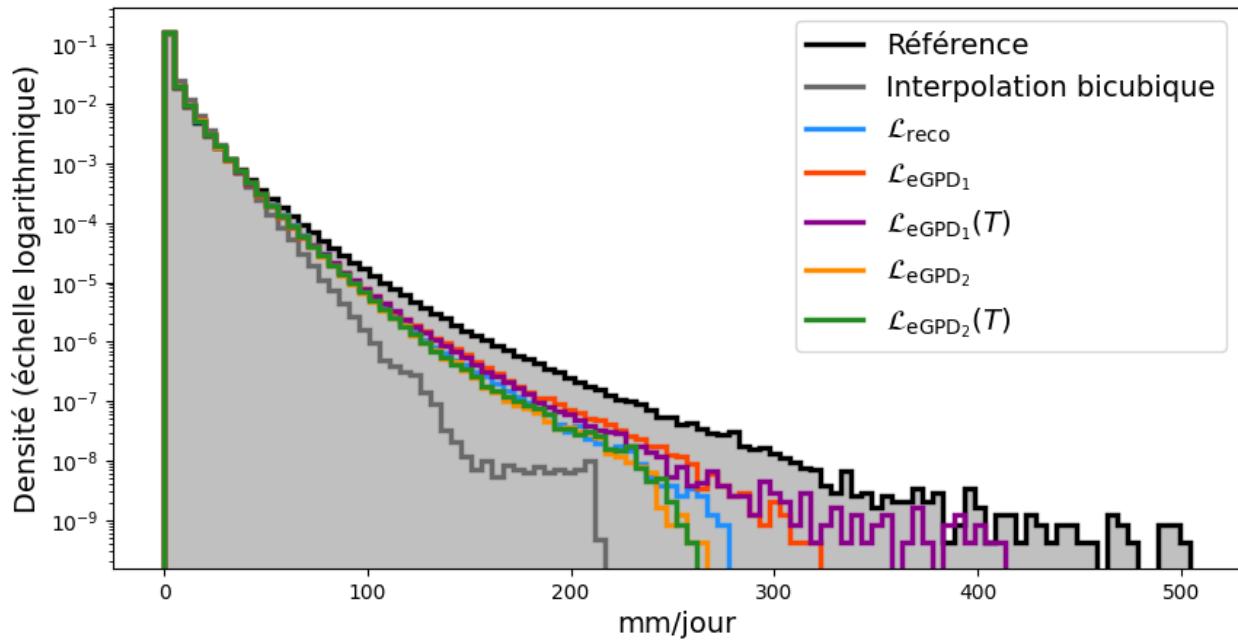


Figure 7.7 Precipitation histogram for each model, including all domain cells over the 2020–2099 test period.

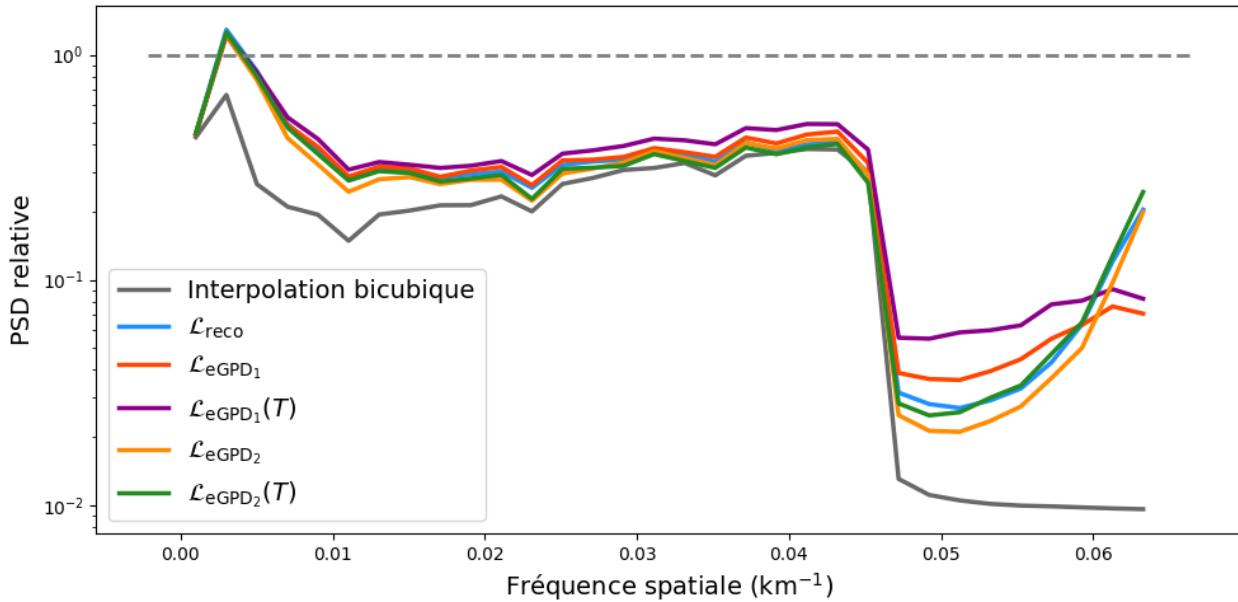


Figure 7.8 PSD_{rel} for each model relative to the reference data PSD, calculated according to the method detailed in Section 7.1.

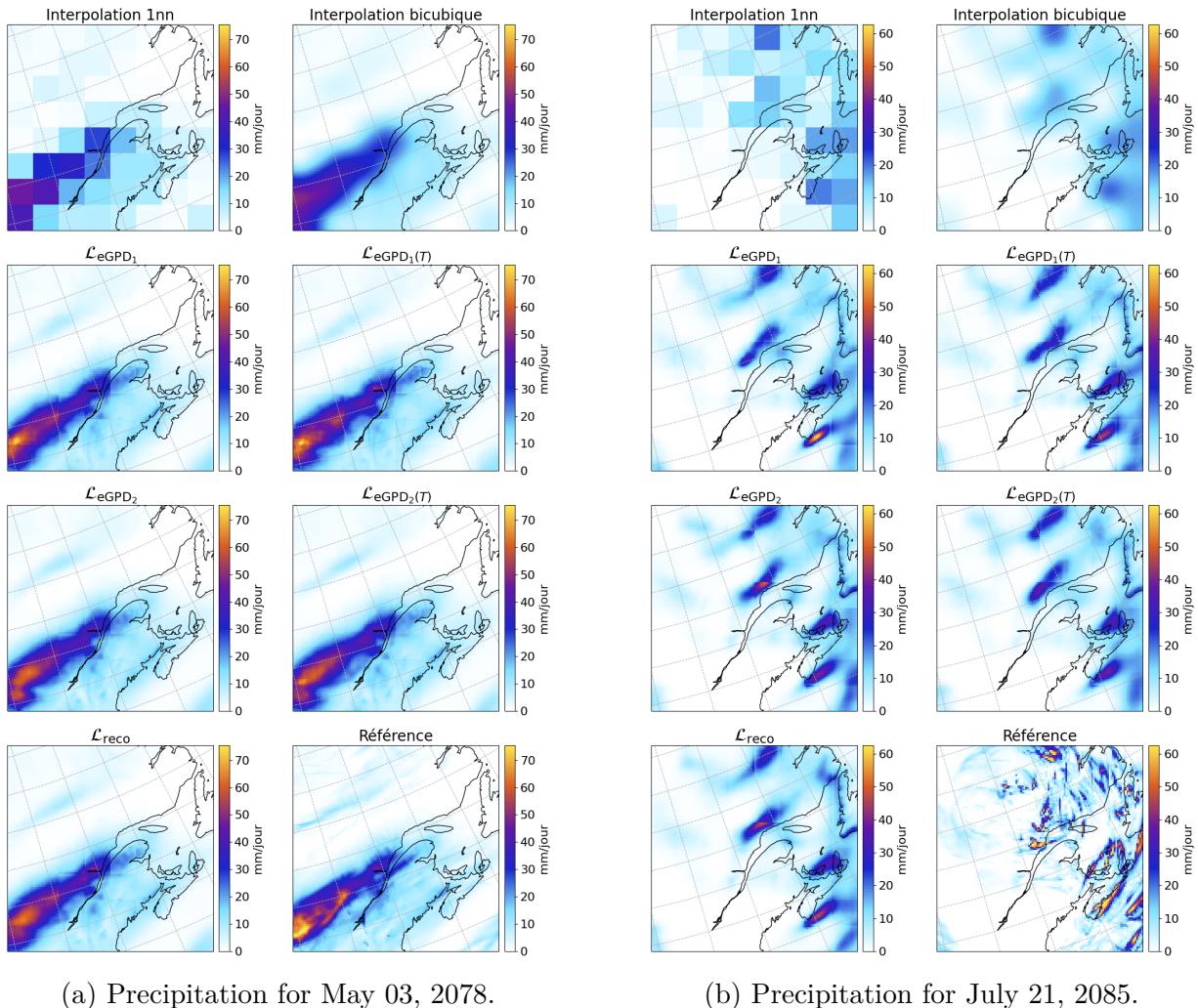


Figure 7.9 Low-resolution precipitation fields (1nn interpolation), interpolated by the bicubic method, predicted by each of the studied models, and high-resolution fields from ClimEx for two dates randomly selected from the test period.

CHAPTER 8 CONCLUSION

8.1 Synthesis of Work

Climate change is intensifying extreme events and profoundly altering the distribution of atmospheric variables such as precipitation, with major impacts on populations and ecosystems. High-resolution climate projections are essential for anticipating and adapting to these developments. Deep learning methods represent a significant opportunity to produce high-resolution simulations at a low computational cost. This thesis focused on precipitation downscaling using deep learning, highlighting certain vulnerabilities of neural networks, notably their difficulty in predicting extremes and their lack of robustness in the face of climate change.

After implementing a UNet-type deep convolutional network for the precipitation field downscaling task—including the definition of a data pre- and post-processing strategy—several structural improvements from the literature were integrated to enhance its performance.

The evaluation of the UNet under climate change, accompanied by the implementation of TPSR estimation methods partially inspired by extreme value theory, led to the formulation of the hypothesis that the degradation of precipitation downscaling performance under climate change is linked to the lack of integration of the Clausius-Clapeyron relationship within the learned features.

This observation led to the proposal of incorporating TPSR information into UNet training using the binning approach. Despite a functional implementation, the potential of this method could not be fully explored, prompting a rethink of the integration strategy for the information obtained through this approach. Consequently, two alternative approaches were developed, consisting of conditioning the learning process with a cumulative distribution function (CDF) derived from extreme value theory. This function represents the precipitation distribution for each cell as a function of the evolution of temperature anomalies. The parameters of this distribution were estimated for each cell in the domain, and we explicitly modeled the dependence on temperature anomalies, allowing for a better representation of precipitation under climate change.

At this stage, the results do not allow for a definitive positive conclusion regarding the effectiveness of the proposed methods in overcoming the identified limitations of the UNet for precipitation downscaling. However, certain elements suggest significant potential and indicate that further work is needed to better exploit the developed approaches.

8.2 Research Perspectives

As discussed in this thesis, one promising strategy would be to partition the spatial domain into overlapping sub-domains (patches) and train the UNet on this new structure. Such an approach would significantly increase mini-batch sizes, an essential condition for ensuring a robust estimation of TPSRs via the binning approach. As suggested by the results in Section 6.1.2, this method would also facilitate the convergence of the \mathcal{L}_{CC} loss function defined in (6.1).

Regarding methods based on the extension of the generalized Pareto distribution, a more in-depth analysis of the influence of temperature anomalies on precipitation modeling would be necessary. It would also be relevant to consider a combination of the proposed loss functions to merge the strengths of each approach. Finally, drawing inspiration from deep learning work applied to inverse problems, the introduction of sampling layers directly derived from the ExtGPD distribution within the UNet could constitute a promising avenue. This architectural integration would allow for the explicit incorporation of priors on the shape of precipitation distributions, as well as their dependence on temperatures.

REFERENCES

- [1] P. Naveau, R. Huser, P. Ribereau, and A. Hannart, “Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection,” *Water Resources Research*, vol. 52, no. 4, pp. 2753–2769, 2016. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/2015WR018552>
- [2] R. A. Watt and L. A. Mansfield, “Generative Diffusion-based Downscaling for Climate,” Apr. 2024, arXiv:2404.17752 [physics]. [Online]. Available: <http://arxiv.org/abs/2404.17752>
- [3] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, publisher: Nature Publishing Group. [Online]. Available: <https://www.nature.com/articles/nature14539>
- [4] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, <http://www.deeplearningbook.org>.
- [5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” in *Advances in Neural Information Processing Systems*, vol. 25. Curran Associates, Inc., 2012. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html
- [6] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241.
- [7] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, “Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation,” *CoRR*, vol. abs/1802.06955, 2018. [Online]. Available: <http://arxiv.org/abs/1802.06955>
- [8] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, “Unet++: A nested u-net architecture for medical image segmentation,” in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, 2018, pp. 3–11.

- [9] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, “Attention u-net: Learning where to look for the pancreas,” in *Medical Imaging with Deep Learning*, 2018. [Online]. Available: <https://openreview.net/forum?id=Skft7cijM>
- [10] S. Coles, *An Introduction to Statistical Modeling of Extreme Values*. London: Springer London, 2001. [Online]. Available: <http://dx.doi.org/10.1007/978-1-4471-3675-0>
- [11] R. L. Smith, “Maximum likelihood estimation in a class of nonregular cases,” *Biometrika*, vol. 72, no. 1, pp. 67–90, 1985. [Online]. Available: <http://www.jstor.org/stable/2336336>
- [12] I. Papastathopoulos and J. Tawn, “Extended generalised pareto models for tail estimation,” *Journal of Statistical Planning and Inference*, vol. 143, no. 1, pp. 131–143, 2013.
- [13] A. Haruna, J. Blanchet, and A.-C. Favre, “Performance-based comparison of regionalization methods to improve the at-site estimates of daily precipitation,” *Hydrology and Earth System Sciences*, vol. 26, no. 10, pp. 2797–2811, 2022. [Online]. Available: <https://hess.copernicus.org/articles/26/2797/2022/>
- [14] J. S. Nanditha, G. Villarini, and P. Naveau, “Assessing future changes in daily precipitation extremes across the contiguous United States with the extended Generalized Pareto distribution,” *Journal of Hydrology*, vol. 659, p. 133212, 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0022169425005505>
- [15] P. Gamet and J. Jalbert, “A flexible extended generalized pareto distribution for tail estimation,” *Environmetrics*, vol. 33, no. 6, p. e2744, 2022. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/env.2744>
- [16] G. Lentink and E. van Meijgaard, “Increase in hourly precipitation extremes beyond expectations from temperature changes,” *Nature Geoscience*, vol. 1, no. 8, pp. 511–514, Aug. 2008. [Online]. Available: <https://doi.org/10.1038/ngeo262>
- [17] R. Hardwick Jones, S. Westra, and A. Sharma, “Observed relationships between extreme sub-daily precipitation, surface temperature, and relative humidity,” *Geophysical Research Letters*, vol. 37, no. 22, 2010, _eprint: <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2010GL045081>. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1029/2010GL045081>

- [18] H. Ali, H. J. Fowler, G. Lenderink, E. Lewis, and D. Pritchard, “Consistent Large-Scale Response of Hourly Extreme Precipitation to Temperature Variation Over Land,” *Geophysical Research Letters*, vol. 48, no. 4, p. e2020GL090317, 2021, _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1029/2020GL090317>. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1029/2020GL090317>
- [19] A. Papadopoulos-Zachos and C. Anagnostopoulou, “A methodological approach to improving extreme precipitation reanalysis data using the clausius-clapeyron relationship: A case study in a mediterranean city,” *Atmosphere*, vol. 15, no. 10, 2024. [Online]. Available: <https://www.mdpi.com/2073-4433/15/10/1195>
- [20] Y. Moustakis, C. J. Onof, and A. Paschalis, “Atmospheric convection, dynamics and topography shape the scaling pattern of hourly rainfall extremes with temperature globally,” *Communications Earth & Environment*, vol. 1, no. 1, pp. 1–9, Aug. 2020, publisher: Nature Publishing Group. [Online]. Available: <https://www.nature.com/articles/s43247-020-0003-0>
- [21] H. Van de Vyver, B. Van Schaeybroeck, R. De Troch, R. Hamdi, and P. Termonia, “Modeling the Scaling of Short-Duration Precipitation Extremes With Temperature,” *Earth and Space Science*, vol. 6, no. 10, pp. 2031–2041, 2019, _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1029/2019EA000665>. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1029/2019EA000665>
- [22] J. Bao, S. C. Sherwood, L. V. Alexander, and J. P. Evans, “Future increases in extreme precipitation exceed observed scaling rates,” *Nature Climate Change*, vol. 7, no. 2, pp. 128–132, Feb. 2017. [Online]. Available: <https://www.nature.com/articles/nclimate3201>
- [23] Q. Sun, F. Zwiers, X. Zhang, and G. Li, “A comparison of intra-annual and long-term trend scaling of extreme precipitation with temperature in a large-ensemble regional climate simulation,” *Journal of Climate*, vol. 33, no. 21, pp. 9233 – 9245, 2020. [Online]. Available: <https://journals.ametsoc.org/view/journals/clim/33/21/jcliD190920.xml>
- [24] X. Zhang, F. W. Zwiers, G. Li, H. Wan, and A. J. Cannon, “Complexity in estimating past and future extreme short-duration rainfall,” *Nature Geoscience*, vol. 10, no. 4, pp. 255–259, Apr. 2017. [Online]. Available: <https://www.nature.com/articles/ngeo2911>
- [25] A. Pérez Bello, A. Mailhot, and D. Paquin, “The Response of Daily and Sub-Daily Extreme Precipitations to Changes in Surface and Dew-Point Temperatures,” *Journal of Geophysical Research: Atmospheres*, vol. 126, no. 16, p. e2021JD034972, 2021,

- _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1029/2021JD034972>. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1029/2021JD034972>
- [26] C. Li, F. Zwiers, X. Zhang, and G. Li, “How Much Information Is Required to Well Constrain Local Estimates of Future Precipitation Extremes?” *Earth’s Future*, vol. 7, no. 1, pp. 11–24, 2019, _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1029/2018EF001001>. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1029/2018EF001001>
- [27] N. A. Phillips, “The general circulation of the atmosphere: A numerical experiment,” *Quarterly Journal of the Royal Meteorological Society*, vol. 82, no. 352, 1956. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/qj.49708235202>
- [28] A. Dahan, “Putting the Earth System in a numerical box? The evolution from climate modeling toward global change,” *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, vol. 41, no. 3, pp. 282–292, Sep. 2010. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1355219810000511>
- [29] T. Schneider, J. Teixeira, C. S. Bretherton, F. Bréint, K. G. Pressel, C. Schär, and A. P. Siebesma, “Climate goals and computing the future of clouds,” *Nature Climate Change*, vol. 7, no. 1, pp. 3–5, Jan. 2017, publisher: Nature Publishing Group. [Online]. Available: <https://www.nature.com/articles/nclimate3190>
- [30] V. Balaji, F. Couvreux, J. Deshayes, J. Gautrais, F. Hourdin, and C. Rio, “Are general circulation models obsolete?” *Proceedings of the National Academy of Sciences*, vol. 119, no. 47, p. e2202075119, Nov. 2022, publisher: Proceedings of the National Academy of Sciences. [Online]. Available: <https://www.pnas.org/doi/abs/10.1073/pnas.2202075119>
- [31] G. M. Flato, “Earth system models: an overview,” *WIREs Climate Change*, vol. 2, no. 6, pp. 783–800, 2011. [Online]. Available: <https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wcc.148>
- [32] T. J. Crowley, “Causes of climate change over the past 1000 years,” *Science*, vol. 289, no. 5477, pp. 270–277, 2000. [Online]. Available: <https://www.science.org/doi/abs/10.1126/science.289.5477.270>
- [33] IPCC, *Climate Change 2022: Impacts, Adaptation and Vulnerability: Working Group II Contribution to the Sixth Assessment Report of the Intergovernmental Panel on*

- Climate Change*, 1st ed. Cambridge University Press, Jun. 2023. [Online]. Available: <https://www.cambridge.org/core/product/identifier/9781009325844/type/book>
- [34] IPCC, *Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, P. R. Shukla, J. Skea, A. R. Reisinger, and IPCC, Eds. Geneva: Cambridge University Press, 2022. [Online]. Available: <https://www.cambridge.org/core/product/identifier/9781009157926/type/book>
- [35] S. Seneviratne, X. Zhang, M. Adnan, W. Badi, C. Dereczynski, A. Di Luca, S. Ghosh, I. Iskandar, J. Kossin, S. Lewis, F. Otto, I. Pinto, M. Satoh, S. Vicente-Serrano, M. Wehner, and B. Zhou, *Weather and Climate Extreme Events in a Changing Climate*. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press, 2021, p. 1513–1766.
- [36] P. C. D. Milly, R. T. Wetherald, K. A. Dunne, and T. L. Delworth, “Increasing risk of great floods in a changing climate,” *Nature*, vol. 415, no. 6871, pp. 514–517, Jan. 2002, publisher: Nature Publishing Group. [Online]. Available: <https://www.nature.com/articles/415514a>
- [37] P. Neumann, P. Düben, P. Adamidis, P. Bauer, M. Brück, L. Kornblueh, D. Klocke, B. Stevens, N. Wedi, and J. Biercamp, “Assessing the scales in numerical weather and climate predictions: will exascale be the rescue?” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 377, no. 2142, p. 20180148, Feb. 2019, publisher: Royal Society. [Online]. Available: <https://royalsocietypublishing.org/doi/10.1098/rsta.2018.0148>
- [38] C. L. Walsh, D. Roberts, R. J. Dawson, J. W. Hall, A. Nickson, and R. Hounsome, “Experiences of integrated assessment of climate impacts, adaptation and mitigation modelling in London and Durban,” *Environment & Urbanization*, vol. 25, no. 2, pp. 361–380, Oct. 2013, publisher: SAGE Publications Ltd. [Online]. Available: <https://doi.org/10.1177/0956247813501121>
- [39] A. T. Ogunrinde, O. , Phillip G., A. , Akinola S., , and J. T. Fasimmirin, “Evaluation of the impact of climate change on the characteristics of drought in Sahel Region of Nigeria: 1971–2060,” *African Geographical Review*, vol. 40, no. 2, pp. 192–210, Apr. 2021, publisher: Routledge. [Online]. Available: <https://www.tandfonline.com/doi/full/10.1080/19376812.2020.1814826>

- [40] N. Pilgj, M. Taszarek, M. Kryza, and H. E. Brooks, “Reconstruction of Violent Tornado Environments in Europe: High-Resolution Dynamical Downscaling of ERA5,” *Geophysical Research Letters*, vol. 49, no. 11, p. e2022GL098242, 2022, _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1029/2022GL098242>. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1029/2022GL098242>
- [41] A. Foughali, Y. Tramblay, Z. Bargaoui, J. Carreau, and D. Ruelland, “Hydrological Modeling in Northern Tunisia with Regional Climate Model Outputs: Performance Evaluation and Bias-Correction in Present Climate Conditions,” *Climate*, vol. 3, no. 3, pp. 459–473, Sep. 2015, number: 3 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: <https://www.mdpi.com/2225-1154/3/3/459>
- [42] J. Gaudreau, L. Perez, and P. Drapeau, “BorealFireSim: A GIS-based cellular automata model of wildfires for the boreal forest of Quebec in a climate change paradigm,” *Ecological Informatics*, vol. 32, pp. 12–27, Mar. 2016. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1574954115002022>
- [43] F. Giorgi and W. J. G. Jr, “Regional Dynamical Downscaling and the CORDEX Initiative,” *Annual Review of Environment and Resources*, vol. 40, no. Volume 40, 2015, pp. 467–490, Nov. 2015, publisher: Annual Reviews. [Online]. Available: <https://www.annualreviews.org/content/journals/10.1146/annurev-environ-102014-021217>
- [44] F. Giorgi, “Thirty years of regional climate modeling: Where are we and where are we going next?” *Journal of Geophysical Research: Atmospheres*, vol. 124, no. 11, pp. 5696–5723, 2019. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018JD030094>
- [45] A. Martynov, R. Laprise, L. Sushama, K. Winger, L. Šeparović, and B. Dugas, “Reanalysis-driven climate simulation over CORDEX North America domain using the Canadian Regional Climate Model, version 5: model performance evaluation,” *Climate Dynamics*, vol. 41, no. 11, pp. 2973–3005, Dec. 2013. [Online]. Available: <https://doi.org/10.1007/s00382-013-1778-9>
- [46] D. Paquin, M. Giguère, C. McCray, O. Asselin, C. Gauthier, D. Matte, M.-P. Labonté, and P. Bourgault, “CRCM5-CMIP6 : A dynamically-downscaled ensemble of CMIP6 simulations.” Apr. 2025. [Online]. Available: <https://zenodo.org/records/15196952>
- [47] C. E. Iles, R. Vautard, J. Strachan, S. Joussaume, B. R. Eggen, and C. D. Hewitt, “The benefits of increasing resolution in global and regional climate simulations for European climate extremes,” *Geoscientific Model Development*, vol. 13, no. 11,

- pp. 5583–5607, Nov. 2020, publisher: Copernicus GmbH. [Online]. Available: <https://gmd.copernicus.org/articles/13/5583/2020/>
- [48] F. von Trentini, M. Leduc, and R. Ludwig, “Assessing natural variability in RCM signals: comparison of a multi model EURO-CORDEX ensemble with a 50-member single model large ensemble,” *Climate Dynamics*, vol. 53, no. 3, pp. 1963–1979, Aug. 2019. [Online]. Available: <https://doi.org/10.1007/s00382-019-04755-8>
- [49] D. Maraun, M. Widmann, and J. M. Gutiérrez, “Statistical downscaling skill under present climate conditions: A synthesis of the VALUE perfect predictor experiment,” *International Journal of Climatology*, vol. 39, no. 9, pp. 3692–3703, 2019, _eprint: <https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/joc.5877>. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/joc.5877>
- [50] H. Tabari, S. M. Paz, D. Buekenhout, and P. Willems, “Comparison of statistical downscaling methods for climate change impact analysis on precipitation-driven drought,” *Hydrology and Earth System Sciences*, vol. 25, no. 6, pp. 3493–3517, 2021. [Online]. Available: <https://hess.copernicus.org/articles/25/3493/2021/>
- [51] C. O. de Burgh-Day and T. Leeuwenburg, “Machine learning for numerical weather and climate modelling: a review,” *Geoscientific Model Development*, vol. 16, no. 22, pp. 6433–6477, Nov. 2023, publisher: Copernicus GmbH. [Online]. Available: <https://gmd.copernicus.org/articles/16/6433/2023/>
- [52] O. Watt-Meyer, G. Dresdner, J. McGibbon, S. K. Clark, B. Henn, J. Duncan, N. D. Brenowitz, K. Kashinath, M. S. Pritchard, B. Bonev, M. E. Peters, and C. S. Bretherton, “ACE: A fast, skillful learned global atmospheric model for climate prediction,” Dec. 2023, arXiv:2310.02074 [physics]. [Online]. Available: <http://arxiv.org/abs/2310.02074>
- [53] I. Price, A. Sanchez-Gonzalez, F. Alet, T. R. Andersson, A. El-Kadi, D. Masters, T. Ewalds, J. Stott, S. Mohamed, P. Battaglia, R. Lam, and M. Willson, “Probabilistic weather forecasting with machine learning,” *Nature*, vol. 637, no. 8044, pp. 84–90, Jan. 2025. [Online]. Available: <https://doi.org/10.1038/s41586-024-08252-9>
- [54] S. Rasp, M. S. Pritchard, and P. Gentine, “Deep learning to represent subgrid processes in climate models,” *Proceedings of the National Academy of Sciences*, vol. 115, no. 39, pp. 9684–9689, 2018. [Online]. Available: <https://www.pnas.org/doi/abs/10.1073/pnas.1810286115>

- [55] N. Rampal, S. Hobeichi, P. B. Gibson, J. Baño-Medina, G. Abramowitz, T. Beucler, J. González-Abad, W. Chapman, P. Harder, and J. M. Gutiérrez, “Enhancing regional climate downscaling through advances in machine learning,” *Artificial Intelligence for the Earth Systems*, vol. 3, no. 2, p. 230066, 2024. [Online]. Available: <https://journals.ametsoc.org/view/journals/aies/3/2/AIES-D-23-0066.1.xml>
- [56] A. W. Wood, E. P. Maurer, A. Kumar, and D. P. Lettenmaier, “Long-range experimental hydrologic forecasting for the eastern united states,” *Journal of Geophysical Research: Atmospheres*, vol. 107, no. D20, pp. ACL–6, 2002.
- [57] M. Hessami, P. Gachon, T. Ouarda, and A. St-Hilaire, “Automated regression-based statistical downscaling tool,” *Environmental Modelling & Software*, vol. 23, pp. 813–834, 06 2008.
- [58] T. Vandal, E. Kodra, S. Ganguly, A. Michaelis, R. Nemani, and A. R. Ganguly, “DeepSD: Generating High Resolution Climate Change Projections through Single Image Super-Resolution,” in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD ’17. New York, NY, USA: Association for Computing Machinery, 2017, pp. 1663–1672. [Online]. Available: <https://dl.acm.org/doi/10.1145/3097983.3098004>
- [59] T. Vandal, E. Kodra, and A. R. Ganguly, “Intercomparison of machine learning methods for statistical downscaling: the case of daily and extreme precipitation,” *Theoretical and Applied Climatology*, vol. 137, no. 1, pp. 557–570, Jul. 2019. [Online]. Available: <https://doi.org/10.1007/s00704-018-2613-3>
- [60] B. Pan, K. Hsu, A. AghaKouchak, and S. Sorooshian, “Improving Precipitation Estimation Using Convolutional Neural Network,” *Water Resources Research*, vol. 55, no. 3, pp. 2301–2321, 2019, _eprint: <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2018WR024090>. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1029/2018WR024090>
- [61] J. Baño-Medina, R. Manzanas, and J. M. Gutiérrez, “Configuration and intercomparison of deep learning neural models for statistical downscaling,” *Geoscientific Model Development*, vol. 13, no. 4, pp. 2109–2124, 2020. [Online]. Available: <https://gmd.copernicus.org/articles/13/2109/2020/>
- [62] L. S. Passarella, S. Mahajan, A. Pal, and M. R. Norman, “Reconstructing High Resolution ESM Data Through a Novel Fast Super Resolution Convolutional Neural Network (FSRCNN),” *Geophysical Research Letters*, vol. 49, no. 4, p. e2021GL097571, 2022,

- _eprint: <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2021GL097571>. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1029/2021GL097571>
- [63] K. Höhlein, M. Kern, T. Hewson, and R. Westermann, “A comparative study of convolutional neural network models for wind field downscaling,” *Meteorological Applications*, vol. 27, no. 6, p. e1961, Nov. 2020. [Online]. Available: <https://rmets.onlinelibrary.wiley.com/doi/10.1002/met.1961>
- [64] M. van der Meer, S. de Roda Husman, and S. Lhermitte, “Deep learning regional climate model emulators: A comparison of two downscaling training frameworks,” *Journal of Advances in Modeling Earth Systems*, vol. 15, no. 6, p. e2022MS003593, 2023, e2022MS003593 2022MS003593. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2022MS003593>
- [65] A. Doury, S. Somot, and S. Gadat, “On the suitability of a convolutional neural network based RCM-emulator for fine spatio-temporal precipitation,” *Climate Dynamics*, Jul. 2024. [Online]. Available: <https://doi.org/10.1007/s00382-024-07350-8>
- [66] T. Nguyen, J. Brandstetter, A. Kapoor, J. K. Gupta, and A. Grover, “ClimaX: A foundation model for weather and climate,” in *Proceedings of the 40th International Conference on Machine Learning*. PMLR, Jul. 2023, pp. 25 904–25 938, iSSN: 2640-3498. [Online]. Available: <https://proceedings.mlr.press/v202/nguyen23a.html>
- [67] X. Zhong, F. Du, L. Chen, Z. Wang, and H. Li, “Investigating transformer-based models for spatial downscaling and correcting biases of near-surface temperature and wind-speed forecasts,” *Quarterly Journal of the Royal Meteorological Society*, vol. 150, no. 758, pp. 275–289, 2024, _eprint: <https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.4596>. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/qj.4596>
- [68] Z. Liu, H. Chen, L. Bai, W. Li, W. Ouyang, Z. Zou, and Z. Shi, “MambaDS: Near-Surface Meteorological Field Downscaling With Topography Constrained Selective State-Space Modeling,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–15, 2024. [Online]. Available: <https://ieeexplore.ieee.org/document/10752514/>
- [69] Q. Yang, A. Hernandez-Garcia, P. Harder, V. Ramesh, P. Sattigeri, D. Szwarcman, C. D. Watson, and D. Rolnick, “Fourier Neural Operators for Arbitrary Resolution Climate Data Downscaling,” *Journal Machine Learning Research*, no. 25, 2024.

- [70] P. Hess and N. Boers, “Deep Learning for Improving Numerical Weather Prediction of Heavy Rainfall,” *Journal of Advances in Modeling Earth Systems*, vol. 14, no. 3, p. e2021MS002765, Mar. 2022. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021MS002765>
- [71] P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud, “Two deterministic half-quadratic regularization algorithms for computed imaging,” in *Proceedings of 1st International Conference on Image Processing*, vol. 2, Nov. 1994, pp. 168–172 vol.2. [Online]. Available: <https://ieeexplore.ieee.org/document/413553/>
- [72] S. Kohl, B. Romera-Paredes, C. Meyer, J. De Fauw, J. R. Ledsam, K. Maier-Hein, S. M. A. Eslami, D. Jimenez Rezende, and O. Ronneberger, “A Probabilistic U-Net for Segmentation of Ambiguous Images,” in *Advances in Neural Information Processing Systems*, vol. 31. Curran Associates, Inc., 2018. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2018/hash/473447ac58e1cd7e96172575f48dca3b-Abstract.html
- [73] J. Ho, A. Jain, and P. Abbeel, “Denoising Diffusion Probabilistic Models,” in *Advances in Neural Information Processing Systems*, vol. 33. Curran Associates, Inc., 2020, pp. 6840–6851. [Online]. Available: <https://proceedings.neurips.cc/paper/2020/hash/4c5bcfec8584af0d967f1ab10179ca4b-Abstract.html>
- [74] M. Mardani, N. Brenowitz, Y. Cohen, J. Pathak, C.-Y. Chen, C.-C. Liu, A. Vahdat, M. A. Nabian, T. Ge, A. Subramaniam, K. Kashinath, J. Kautz, and M. Pritchard, “Residual corrective diffusion modeling for km-scale atmospheric downscaling,” *Communications Earth & Environment*, vol. 6, no. 1, pp. 1–10, Feb. 2025, publisher: Nature Publishing Group. [Online]. Available: <https://www.nature.com/articles/s43247-025-02042-5>
- [75] Z. Y. Wan, R. Baptista, Y.-f. Chen, J. Anderson, A. Boral, F. Sha, and L. Zepeda-Núñez, “Debias Coarsely, Sample Conditionally: Statistical Downscaling through Optimal Transport and Probabilistic Diffusion Models,” Oct. 2023, arXiv:2305.15618 [cs]. [Online]. Available: <http://arxiv.org/abs/2305.15618>
- [76] H. Addison, E. Kendon, S. Ravuri, L. Aitchison, and P. A. Watson, “Machine learning emulation of precipitation from km-scale regional climate simulations using a diffusion model,” Jul. 2024, arXiv:2407.14158 [physics]. [Online]. Available: <http://arxiv.org/abs/2407.14158>

- [77] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative Adversarial Nets,” in *Advances in Neural Information Processing Systems*, vol. 27. Curran Associates, Inc., 2014. [Online]. Available: https://papers.nips.cc/paper_files/paper/2014/hash/f033ed80deb0234979a61f95710dbe25-Abstract.html
- [78] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks,” in *2017 IEEE International Conference on Computer Vision (ICCV)*. Venice: IEEE, Oct. 2017, pp. 2242–2251. [Online]. Available: <http://ieeexplore.ieee.org/document/8237506/>
- [79] P. Hess, M. Drüke, S. Petri, F. M. Strnad, and N. Boers, “Physically constrained generative adversarial networks for improving precipitation fields from Earth system models,” *Nature Machine Intelligence*, vol. 4, no. 10, pp. 828–839, Oct. 2022, publisher: Nature Publishing Group. [Online]. Available: <https://www.nature.com/articles/s42256-022-00540-1>
- [80] C. Charles, “Introduction aux problèmes inverses,” Jan. 2014, note de statistique et d’informatique - Université de Liège.
- [81] J. Gonzalez-Abad, A. Hernandez-Garcia, P. Harder, D. Rolnick, and J. M. Gutierrez, “Multi-Variable Hard Physical Constraints for Climate Model Downscaling,” *Proceedings of the AAAI Symposium Series*, vol. 2, no. 1, pp. 62–67, Jan. 2024. [Online]. Available: <https://ojs.aaai.org/index.php/AAAI-SS/article/view/27650>
- [82] K. Kashinath, M. Mustafa, A. Albert, J.-L. Wu, C. Jiang, S. Esmaeilzadeh, K. Azizzadenesheli, R. Wang, A. Chattopadhyay, A. Singh, A. Manepalli, D. Chirila, R. Yu, R. Walters, B. White, H. Xiao, H. A. Tchelepi, P. Marcus, A. Anandkumar, P. Hassanzadeh, and n. Prabhat, “Physics-informed machine learning: case studies for weather and climate modelling,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 379, no. 2194, p. 20200093, Feb. 2021, publisher: Royal Society. [Online]. Available: <https://royalsocietypublishing.org/doi/full/10.1098/rsta.2020.0093>
- [83] P. Harder, A. Hernandez-Garcia, V. Ramesh, Q. Yang, P. Sattegeri, D. Szwarcman, C. D. Watson, and D. Rolnick, “Hard-constrained deep learning for climate downscaling,” *J. Mach. Learn. Res.*, vol. 24, no. 1, pp. 365:17 534–365:17 573, 2023.
- [84] F. Zanetta, D. Nerini, T. Beucler, and M. A. Liniger, “Physics-Constrained Deep Learning Postprocessing of Temperature and Humidity,” Dec. 2023,

- section: Artificial Intelligence for the Earth Systems. [Online]. Available: <https://journals.ametsoc.org/view/journals/aies/2/4/AIES-D-22-0089.1.xml>
- [85] P. Jiang, Z. Yang, J. Wang, C. Huang, P. Xue, T. C. Chakraborty, X. Chen, and Y. Qian, “Efficient Super-Resolution of Near-Surface Climate Modeling Using the Fourier Neural Operator,” *Journal of Advances in Modeling Earth Systems*, vol. 15, no. 7, p. e2023MS003800, 2023. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1029/2023MS003800>
- [86] B. Teufel, F. Carmo, L. Sushama, L. Sun, M. N. Khaliq, S. Bélair, A. Shamseldin, D. N. Kumar, and J. Vaze, “Physics-informed deep learning framework to model intense precipitation events at super resolution,” *Geoscience Letters*, vol. 10, no. 1, p. 19, Apr. 2023. [Online]. Available: <https://doi.org/10.1186/s40562-023-00272-z>
- [87] A. Saha and S. Ravela, “Statistical-physical adversarial learning from data and models for downscaling rainfall extremes,” *Journal of Advances in Modeling Earth Systems*, vol. 16, no. 6, p. e2023MS003860, 2024, e2023MS003860 2023MS003860. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2023MS003860>
- [88] Y. Verma, M. Heinonen, and V. Garg, “ClimODE: Climate forecasting with physics-informed neural ODEs,” in *The Twelfth International Conference on Learning Representations*, 2024. [Online]. Available: <https://openreview.net/forum?id=xuY33XhEGR>
- [89] T. Beucler, P. Gentine, J. Yuval, A. Gupta, L. Peng, J. Lin, S. Yu, S. Rasp, F. Ahmed, P. A. O’Gorman, J. D. Neelin, N. J. Lutsko, and M. Pritchard, “Climate-invariant machine learning,” *Science Advances*, vol. 10, no. 6, p. eadj7250, Feb. 2024, publisher: American Association for the Advancement of Science. [Online]. Available: <https://www.science.org/doi/10.1126/sciadv.adj7250>
- [90] M. Leduc, A. Mailhot, A. Frigon, J.-L. Martel, R. Ludwig, G. B. Brietzke, M. Giguère, F. Brissette, R. Turcotte, M. Braun, and J. Scinocca, “The climex project: A 50-member ensemble of climate change projections at 12-km resolution over europe and northeastern north america with the canadian regional climate model (crcm5),” *Journal of Applied Meteorology and Climatology*, vol. 58, no. 4, pp. 663 – 693, 2019. [Online]. Available: <https://journals.ametsoc.org/view/journals/apme/58/4/jamc-d-18-0021.1.xml>
- [91] S. Elfwing, E. Uchibe, and K. Doya, “Sigmoid-weighted linear units for neural network function approximation in reinforcement learning,” *Neural Networks*, vol. 107, pp.

- 3–11, 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0893608017302976>
- [92] Y. Wu and K. He, “Group Normalization,” *International Journal of Computer Vision*, vol. 128, no. 3, pp. 742–755, Mar. 2020. [Online]. Available: <https://doi.org/10.1007/s11263-019-01198-w>
- [93] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting,” *Journal of Machine Learning Research*, vol. 15, no. 1, p. 1929–1958, Jan. 2014.
- [94] K. He, X. Zhang, S. Ren, and J. Sun, “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification,” in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1026–1034.
- [95] J.-S. Giroux, S.-P. Breton, and J. Carreau, “Interpolation-free deep learning for meteorological downscaling on unaligned grids across multiple domains with application to wind power,” 2024. [Online]. Available: <https://arxiv.org/abs/2410.03945>
- [96] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” 2017. [Online]. Available: <https://arxiv.org/abs/1412.6980>
- [97] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, “A convnet for the 2020s,” *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.
- [98] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, “A limited memory algorithm for bound constrained optimization,” *SIAM Journal on Scientific Computing*, vol. 16, no. 5, pp. 1190–1208, 1995. [Online]. Available: <https://doi.org/10.1137/0916069>
- [99] A. Pérez, M. S. Cruz, D. S. Martín, and J. M. Gutiérrez, “Transformer based super-resolution downscaling for regional reanalysis: Full domain vs tiling approaches,” Oct. 2024. [Online]. Available: <http://arxiv.org/abs/2410.12728>
- [100] E. Kjellström, F. Boberg, M. Castro, J. H. Christensen, G. Nikulin, and E. Sánchez, “Daily and monthly temperature and precipitation statistics as performance indicators for regional climate models,” *Climate Research*, vol. 44, no. 2-3, pp. 135–150, Dec. 2010. [Online]. Available: <https://www.int-res.com/abstracts/cr/v44/n2-3/p135-150/>

APPENDIX A ESTIMATION OF TPSRS VIA THE GENERALIZED EXTREME VALUE DISTRIBUTION

Figures A.1, A.2, A.3, A.4, A.5, and A.6 present the NLLs and the parameters μ_0 , μ_1 , σ_0 , σ_1 , and ξ estimated from the reference data and UNet predictions for the 2020–2099 test period, respectively.

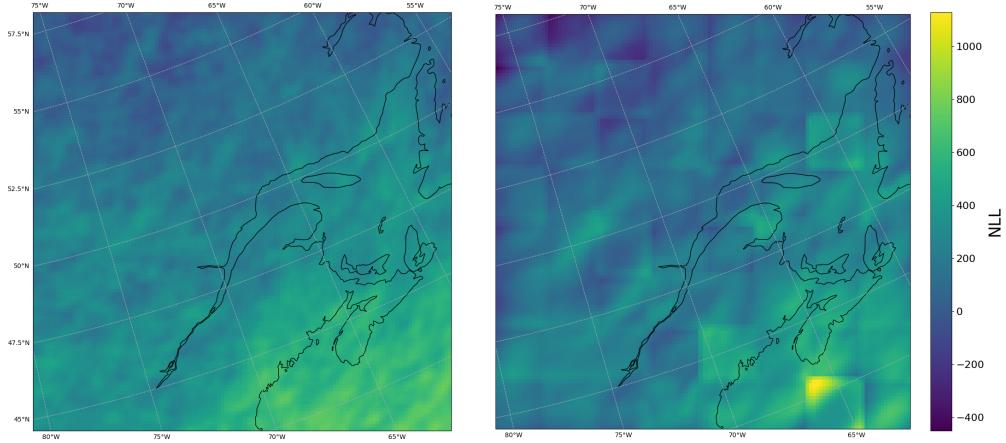


Figure A.1 Negative log-likelihood following parameter optimization under the $M_{\mu,4} + M_{\sigma,2}$ combination for the reference data (**left**) and UNet predictions (**right**).

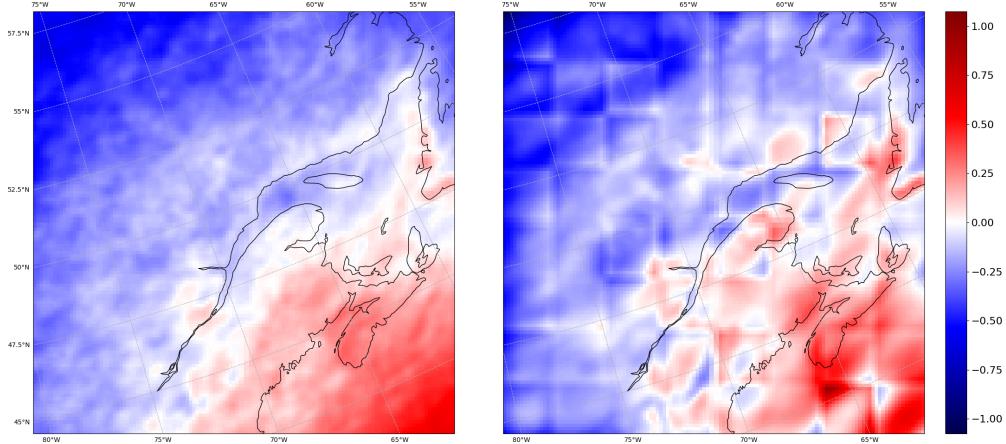


Figure A.2 μ_0 parameters of $M_{\mu,4}$ for the reference data (**left**) and UNet predictions (**right**).

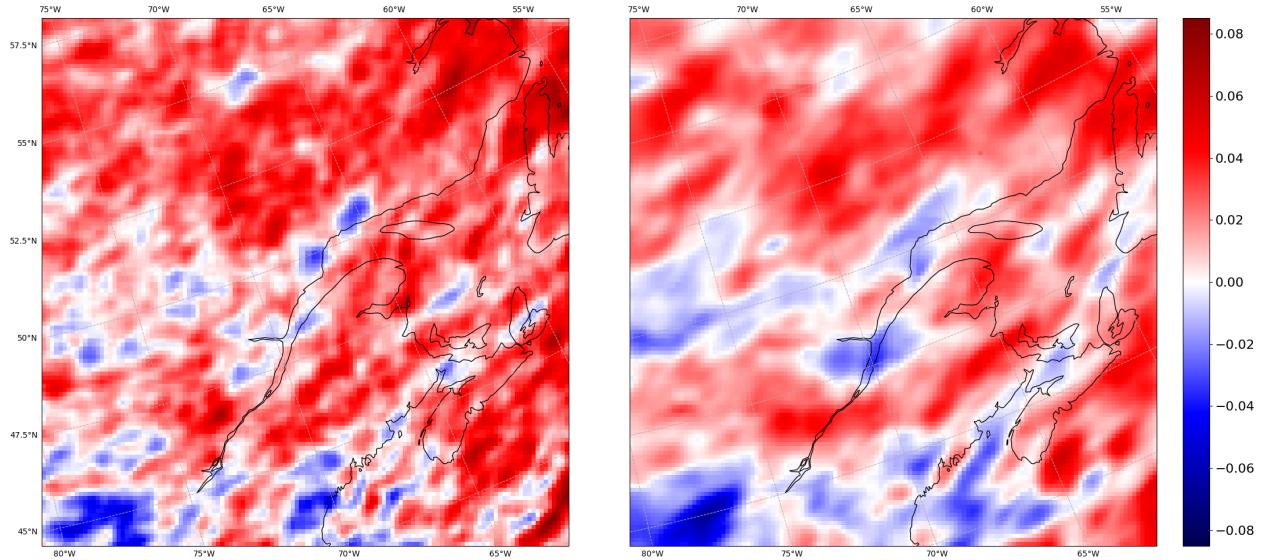


Figure A.3 μ_1 parameters of $M_{\mu,4}$ for the reference data (**left**) and UNet predictions (**right**).

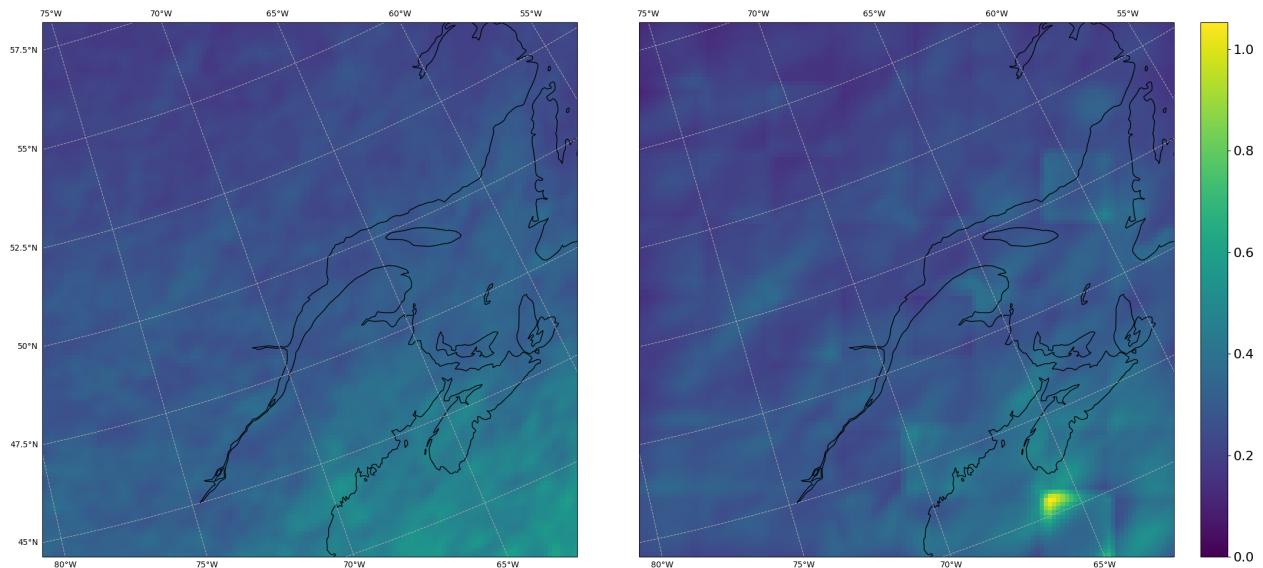


Figure A.4 σ_0 parameters of $M_{\sigma,2}$ for the reference data (**left**) and UNet predictions (**right**).

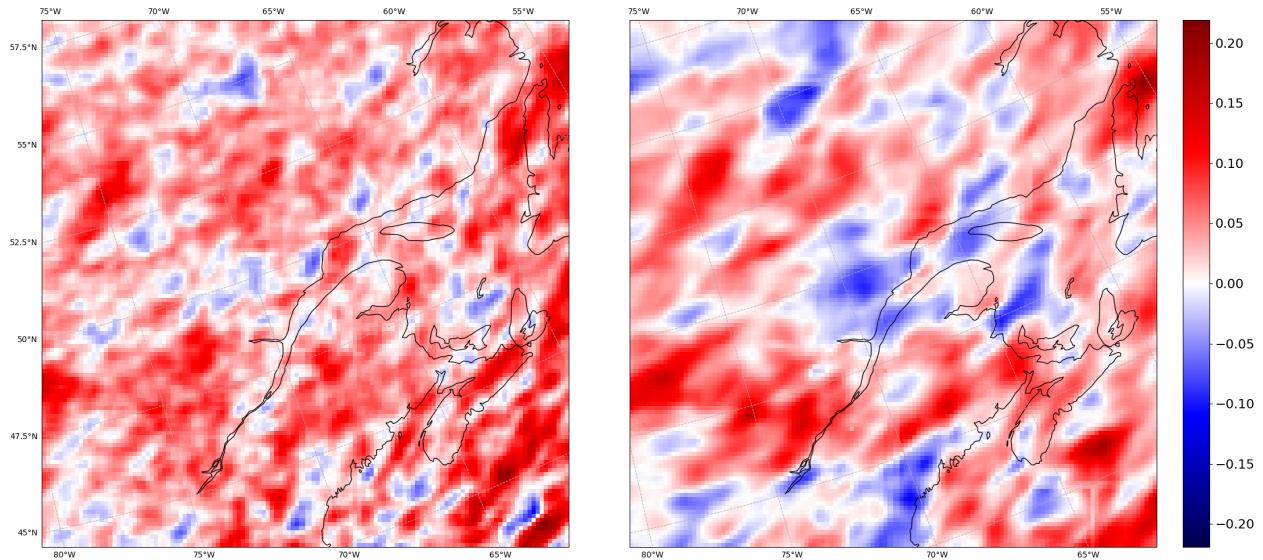


Figure A.5 σ_1 parameters of $M_{\sigma,2}$ for the reference data (**left**) and UNet predictions (**right**).

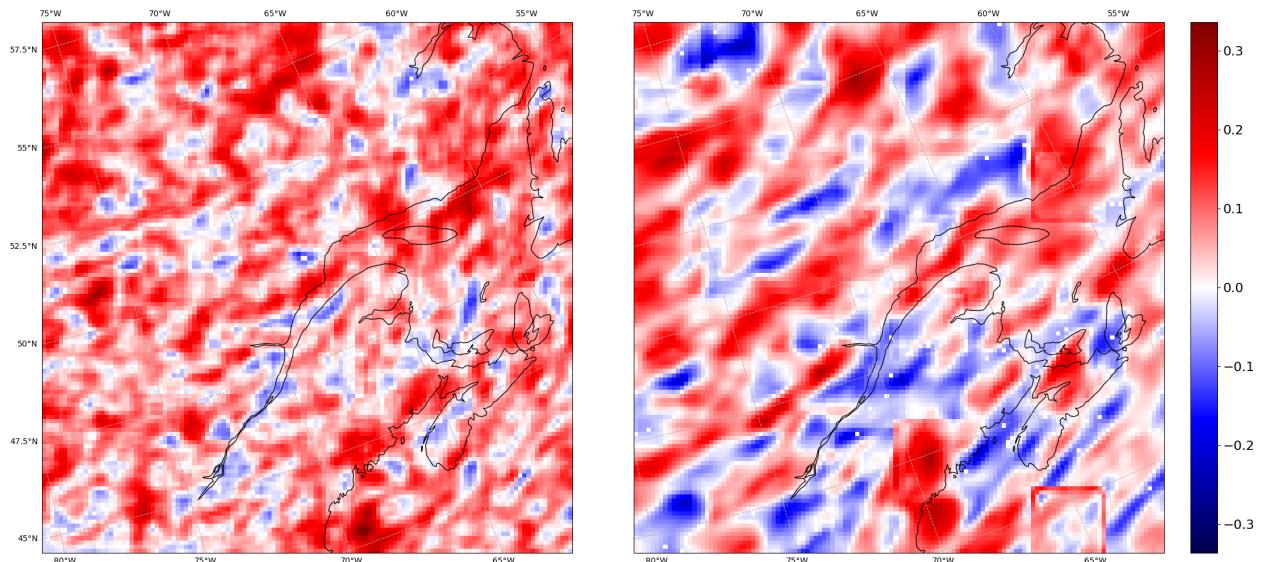


Figure A.6 ξ parameters for the reference data (**left**) and UNet predictions (**right**).