# "Regional Climate Model Emulator Based on Deep Learning: Concept and First Evaluation of a Novel Hybrid Downscaling Approach"

Lola Corre, Antoine Doury, Sébastien Gadat, Aurélien Ribes
and Samuel Somot

Toulouse
School of
Economics

# Regional Climate Model emulator based on deep learning: concept and first evaluation of a novel hybrid downscaling approach

Antoine Doury[1] · Samuel Somot[1] ·
Sebastien Gadat[2] · Aurélien Ribes[1] ·
Lola Corre[3]

**Abstract** Providing reliable information on climate change at local scale remains a challenge of first importance for impact studies and policymakers. Here, we propose a novel hybrid downscaling method combining the strengths of both empirical statistical downscaling methods and Regional Climate Models (RCMs). The aim of this tool is to enlarge the size of high-resolution RCM simulation ensembles at low cost.

We build a statistical RCM-emulator by estimating the downscaling function included in the RCM. This framework allows us to learn the relationship between large-scale predictors and a local surface variable of interest over the RCM domain in present and future climate. Furthermore, the emulator relies on a neural network architecture, which grants computational efficiency.

The RCM-emulator developed in this study is trained to produce daily maps of the near-surface temperature at the RCM resolution (12km). The emulator demonstrates an excellent ability to reproduce the complex spatial structure and daily variability simulated by the RCM and in particular the way the RCM refines locally the low-resolution climate patterns. Training in future climate appears to be a key feature of our emulator. Moreover, there is a huge computational benefit in running the emulator rather than the RCM, since training the emulator takes about 2 hours on GPU, and the prediction is nearly instantaneous.

However, further work is needed to improve the way the RCM-emulator reproduces some of the temperature extremes, the intensity of climate change, and to extend the proposed methodology to different regions, GCMs, RCMs, and variables of interest.

Antoine Doury
[1] CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France
[2] Toulouse School of Economics, Université Toulouse 1 Capitole, Institut Universitaire de France
[3] Météo-France, Toulouse, France
Tel.: +33561079104
E-mail: antoine.doury@meteo.fr

# Contents

## 1 Introduction

Climate models are an essential tool to study possible evolutions of the climate according to different scenarios of greenhouse gas emissions. These numerical models represent the physical and dynamical processes present in the atmosphere and their interactions with other components of the Earth System. The complexity of these models involves compromises between the computational costs, the horizontal resolution and, in some cases, the domain size.

Global Climate Models (GCMs) produce simulations covering the whole planet at reasonable cost thanks to a low spatial resolution (from 50 to 300 km). The large number of different GCMs developed worldwide allows to build large and coordinated ensembles of simulations, thanks to a strong international cooperation. These big ensembles (CMIP3/5/6, Meehl et al, 2007; Taylor et al, 2012; Eyring et al, 2016) are necessary to correctly explore the different sources of variability and uncertainties in order to deliver reliable information about future climate change at large spatial scales. However, the resolution of these models is too coarse to derive any fine scale information, which is of primary importance for impact studies and adaptation policies. Consequently, it is crucial to downscale the GCM outputs to a higher resolution. Two families of downscaling have emerged: empirical-statistical downscaling and dynamical downscaling. Both approaches have their own strengths and weaknesses.

Empirical Statistical Downscaling methods (ESD) estimate functions to link large scale atmosphere fields with local scale variables using observational data. Local implications of future climate changes are then obtained by applying these functions to GCM outputs. Gutiérrez et al (2019) present an overview of the different ESD methods and evaluate their ability to downscale historical GCM simulations. The great advantage of these statistical methods is their computational efficiency, which makes the downscaling of large GCM ensembles possible. On the other hand, they have two main limitations due to their dependency on observational data. First of all, they are applicable only for regions and variables for which local long-term observations are available. Secondly, they rely on the stationary assumption of the large-scale / local-scale relationship, which implies that a statistical model calibrated in the past and present climate remains reliable in the forthcoming climate. Studies tend to show that the calibration period has a non-negligible impact on the results (Wilby et al, 1998; Schmith, 2008; Dayon et al, 2015; Erlandsen et al, 2020).

Dynamical Downscaling (DD) is based on Regional Climate Models (RCMs). These models have higher resolution than GCMs (1 to 50 km) but are restricted to a limited area domain to keep their computational costs affordable. They are nested in a GCM, e.g., they receive at small and regular time intervals dynamical information from this GCM at their lateral boundaries. One key advantage of RCMs is to rely on the same physical hypotheses as the one

involved in GCMs. They provide a complete description of the state of the atmosphere over their domain through a large set of variables at high temporal and spatial resolution. The added value of RCMs has been demonstrated in several studies (Giorgi et al, 2016; Torma et al, 2015; Prein et al, 2016; Fantini et al, 2018; Kotlarski et al, 2015,  for examples). In order to deliver robust information about future local responses to climate change, it is necessary to explore the uncertainty associated with RCM simulations. Déqué et al (2007) and Evin et al (2021) assess that four sources of uncertainty are at play in a regional climate simulation: the choice of the driving GCM, the greenhouse gas scenario, the choice of the RCM itself and the internal variability. Their relative importance depends on the considered variables, spatial scale, and timeline. According to these results, it is important (Déqué et al, 2012; Evin et al, 2021; Fernández et al, 2019) to complete 4D matrices $[SCENARIO, GCM, RCM, MEMBERS]$ to deliver robust messages, where members are several simulations of each $(SCENARIO, GCM, RCM)$ triplet. However, the main limitation of RCM is their high computational costs, and completion of such matrices is impossible.

This study proposes a novel hybrid downscaling method to enlarge the size of RCM simulation ensembles. The idea is to combine the advantages of both dynamical and statistical downscaling to tackle their respective limits. This statistical RCM-emulator uses machine learning methods to learn the relationship between large scale fields and local scale variables inside regional climate simulations. It aims to estimate the downscaling function included in the RCM in order to apply it to new GCM simulations. This framework will allow to learn this function on the entire RCM domain, in past and future climate, under different scenarios. Besides, the emulator relies on Machine Learning algorithms with low computational costs, which will enable to increase RCM simulation ensembles and to better explore the uncertainties associated with these high resolution simulations.

Hybrid statistical-dynamical downscaling methods have already been proposed. They are methods which combine, in different ways, regional climate models and statistical approaches to obtain local climate information. Several studies such as Pryor and Barthelmie (2014), Vrac et al (2012) or Turco et al (2011) perform 2-step downscaling by applying ESD methods to RCM simulations. Colette et al (2012) apply bias correction methods to GCM outputs before downscaling with RCMs. Maraun and Widmann (2018) are among the first to mention the concept of emulators. Few studies have combined ESD and DD for the same purpose as in this study. For instance, Walton et al (2015) propose a statistical model which estimates from GCM outputs, the high resolution warming pattern for each month in California. It is calibrated using RCM simulations and relies on a simple linear combination of two predictors from the GCM (the monthly mean over the domain and an indicator for the land/sea contrast) plus an indicator for the spatial variance (obtained thanks to PCA). Berg et al (2015) adapt the same protocol for monthly changes in

precipitation over California. With respect to those pioneer studies, we propose here to further develop this approach by using a neural network based method and by emulating the full time series at the input time scale, allowing to explore daily climate extremes.

In recent years, climate science has taken advantage of the recent strides in performances of Deep Learning algorithms (see Lecun et al, 2015, for an overview). Indeed, thanks to their capacity to deal with large amounts of data and the strong ability of Convolutional Neural Network (CNN) (LeCun et al, 1998) to extract high level features from images, these algorithms are particularly adapted to climate and weather sciences. Reichstein et al (2019) present a complete overview of the climate studies applying Deep Learning and future possibilities. In particular, Vandal et al (2019) and Baño-Medina et al (2020, 2021) showed a good ability of convolutional neural network architecture to learn the transfer function between large scale fields and local variables in statistical downscaling applications. The concept of emulator is mentioned in Reichstein et al (2019) as surrogate models trained to replace a complex and computationally expensive physical model (entirely or only parts of it). Once trained, this emulator should be able to produce simulations much faster than the original model. In this context, the RCM-emulator proposed here is based on a different fully convolutional neural network architecture known as UNET (Ronneberger et al, 2015).

This study presents and validates the concept of statistical RCM-emulator. We will focus on emulating the near-surface temperature in a RCM over a specific domain, including high mountains, shore areas, and islands in Western Europe. This domain regroups areas where the RCM presents added value compared to GCM but remains small enough to perform quick sensitivity tests. This paper is organized as follows: Section 2 presents the whole framework to define, train and evaluate the emulator, while Section 3 shows the emulator results. Finally, Sections 4 and 5 discuss the results of the emulator and provide conclusions.

Table 1: Notations

| Notation | Description | Dimensions |
|---|---|---|
| $i, j$ | Spatial indexes over input grid | $\{i, j\} \in [\![1 , I]\!] \times [\![1 , J]\!]$ |
| $k, l$ | Spatial indexes over target grid | $\{k, l\} \in [\![1 , K]\!] \times [\![1 , L]\!]$ |
| $t$ | Temporal index, daily | $t \in \mathbb{N}$ |
| $x$ | 2-D variables index | List of 2D Variables : $V2D$ |
| $z$ | 1-D variables index | List of 1D Variables : $V1D$ |
| $X$ | Set of 2D input for the emulator | $X_{t,i,j,x} \in \{\mathbb{N} \times [\![1 , I]\!] \times [\![1 , J]\!] \times V2D\}$ |
| $Z$ | Set of 1D input for the emulator | $Z_{t,z}\{\mathbb{N} \times V1D\}$ |
| $Y$ | Target : Daily Near-Surface Temperature | $Y_{t,k,l} \in \{\mathbb{N} \times [\![1 , K]\!] \times [\![1 , L]\!]\}$ |
| $F$ | Downscaling function of the RCM | |
| $\hat{F}$ | Emulator : Estimation of F | |

## 2 Methodology

This section provides a complete description of the RCM-emulator used in this paper. The notations are summarised in Table 1. The RCM-emulator uses a neural network architecture to learn the relationship between large-scale fields and local-scale variables inside regional climate simulations. RCMs include a downscaling function (F) which transforms large scale information (X,Z) into high resolution surface variables (Y). The statistical RCM-emulator aims to solve a common Machine Learning problem

$$Y = F(X, Z)$$

which is to estimate $F$ by $\hat{F}$ in order to apply it to new GCM simulations. The following paragraphs describe the list of predictors used as inputs and their domain, the predictand (or target) and its domain, the neural network architecture, the framework used to train the emulator and the metrics used to evaluate its performances.

2.1 Models and simulations

This study focuses on the emulation of the daily near-surface temperature from EURO-CORDEX simulations based on the CNRM-ALADIN63 regional climate model (Nabat et al, 2020) driven by the CNRM-CM5 global climate model used in CMIP5 (Voldoire et al, 2013). The latter provides lateral boundary conditions, namely 3D atmospheric forcing at 6-hourly frequency, as well as sea surface temperature, sea ice cover and aerosol optical depth at monthly frequency. The simulations use a Lambert Conformal grid covering the European domain (EUR-11 CORDEX) at the 0,11° (about 12.5 km) scale (Jacob

Table 2: List of predictors

| 2D Variables | | | | | |
|---|---|---|---|---|---|
| **Field** | **Altitude Levels** | **Variables Notation** | **Units** | **Temporal Aggregation** | **Dimension** |
| Geopotential | 850 , 700, 500 hPa | zg500 zg700 zg850 | m | Daily mean | $[i,j]$ |
| Specific Humidity | 850 , 700, 500 hPa | hus500 hus700 hus850 | | Daily mean | $[i,j]$ |
| Temperature | 850 , 700, 500 hPa | ta500 ta700 ta850 | K | Daily mean | $[i,j]$ |
| Eastward Wind | 850 , 700, 500 hPa + Surface | ua500 ua700 ua850 uas | m/s | Daily mean | $[i,j]$ |
| Northward Wind | 850 , 700, 500 hPa + Surface | va500 va700 va850 vas | m/s | Daily mean | $[i,j]$ |
| Sea Level Pressure | Surface | psl | Pa | Daily mean | $[i,j]$ |
| Total Aerosol Optical Depth forcing | | TAOD | | Monthly mean | $[i,j]$ |
| 1-D Variables | | | | | |
| Daily spatial means of 2D variables | | $\overline{x}$ with $x \in V2D$ | | Daily | $[\#V2D]$ |
| Daily spatial standard deviation of 2D variables | | $\overset{\bullet}{x}$ with $x \in V2D$ | | Daily | $[\#V2D]$ |
| Total Anthropogenic greenhouses gas forcings | | ant_ghg | | Yearly | $[1]$ |
| Solar and Ozone forcings | | sol, oz | | Yearly | $[2]$ |
| Seasonal Indicators $Cos(\frac{2\pi t}{365})$ ; $Sin(\frac{2\pi t}{365})$ | | cos,sin | | Daily | $[2]$ |

et al, 2014). The historical period runs from 1951 to 2005. The scenarios (2006-2100) are based on two Representative Concentration Pathways from the fifth phase of the Coupled Model Intercomparison Project (CMIP5): RCP4.5 and RCP8.5 (Moss et al, 2010). The monthly aerosol forcing evolves according to the chosen RCP and the driving GCM.

## 2.2 Predictors

Neural Networks can deal with large datasets at low computational time. During their self optimisation, they are able to select the important variables and regions for the prediction. In this way, a large number of raw predictors can be given to the learning algorithm, with minimum prior selection (which could introduce some bias) or statistical pre-work (which might delete some of the information). Several ESD studies (Lemus-Canovas and Brands, 2020; Erlandsen et al, 2020) show that the right combination of predictors depends

strongly on the target region and the season. The RCM domains are often composed of very different regions in terms of orography, land types, distance to the sea, etc. For these reasons, we decided to give all potentially needed predictors to the emulator and leave the algorithm to determine the right combination to be used to predict each RCM grid point.

The set of predictors $(X, Z)$ used as input in the emulator is composed of 2 dimensional variables $X$, and 1D predictors $Z$ (Table 2). The set of 2D variables includes atmospheric fields commonly used in ESD (Baño-Medina et al, 2020; Gutiérrez et al, 2019) at different pressure levels. We also added the total aerosol optical depth present in the atmosphere since it constitutes a key regional driver of the regional climate change over Europe (Boé et al, 2020). It leads 20 2D predictors. These variables are normalised (see Equation 1) according to their daily spatial mean and standard deviation so that they all have the same importance for the neural network before the training.

The set of 1D variables includes external forcing also given to the RCM: the total concentration of greenhouse gases and the solar and ozone forcings. It also includes a cosinus, sinus vector to encode the information about the day of the year. Given that the 2D variables are normalised at each time step by their spatial mean, they don't carry any temporal information. For this reason, the daily spatial means and standard deviations time series for each 2D variable are included in the 1D input, bringing the size of this vector to 43. In order to always give normalised inputs to the emulator, $Z$ is normalized (see Equation 2 ) according to the means and standard deviations over a reference period (1971-2000 here) chosen during the emulator training. The same set of means and standard deviations will be used to normalise any low resolution data to be downscaled by the emulator.

This decomposition of the large scale information consists in giving separately the spatial structure of the atmosphere $(X)$ and the temporal information $(Z)$ to the emulator. Thanks to the neural network architecture described in Section 2.4.3, we force the emulator to consider both items of information equally.

$$\widetilde{X}_{t,i,j,x} = \frac{X_{t,i,j,x} - \overline{X}_{t,x}}{\overset{\bullet}{X}_{t,x}} \tag{1}$$

$$\text{with } \overline{X}_{t,x} = \frac{\sum\limits_{i=1}^{I}\sum\limits_{j=1}^{J} X_{t,i,j,x}}{I \times J} \text{ and } \overset{\bullet}{X}_{t,x} = \sqrt{\frac{\sum\limits_{i=1}^{I}\sum\limits_{j=1}^{J}(X_{t,i,j,x} - \overline{X}_{t,x})^2}{I \times J}}$$

$$\widetilde{Z}_{t,z} = \frac{Z_{t,z} - \overline{Z}_{ref,z}}{\overset{\bullet}{Z}_{ref,z}} \qquad (2)$$

$$\text{with } \overline{Z}_{ref,z} = \frac{\sum\limits_{t \in R} Z_{t,z}}{nb\,days\,in\,R} \text{ and } \overset{\bullet}{Z}_{ref,z} = \sqrt{\frac{\sum\limits_{t \in R} (Z_{t,z} - \overline{Z}_{ref,z})^2}{nb\,days\,in\,R}}$$

where R is the reference period.

The aim of the emulator is to downscale GCM simulations. Klaver et al (2020) shows that the effective resolution of climate models is often larger (about 3 times) than its nominal resolution. For instance, CNRM-CM5 cannot be trusted at its own horizontal resolution ($\approx$ 150 km) but at a coarser resolution, probably about 450 to 600 km. For this reason, the set of 2D predictors are smoothed with a $3 \times 3$ moving average filter. The grid of the GCM is conserved, with each point containing a smoother information than the raw model outputs.

For this study the input domain is defined around the target domain (described in Section 2.3). It is a $16 \times 16$ ( $J = I = 16$ in Table 1) CNRM-CM5 grid box visible on Figure 1. Each observation given to the emulator (see Figure 1 for an illustration) is a day $t$ and it is composed of a 3D array ($X_{i,j,x}$), where the two first dimensions are the spatial coordinates and the third dimension lists the different variables chosen as predictors, and a 1D array ($Z_z$) regrouping all the 1 dimensional inputs.

### 2.3 Predictands

In this study, to assess the ability of the RCM-emulator to reproduce the RCM downscaling function, we focused on the emulation of the daily near surface temperature over a small but complex domain. The target domain for this study is a box of 64x64 RCM grid points at 12km resolution (about $600\,000\,km^2$) centred over the south of France (Figure 1). It gathers different areas of interest for the regional climate models. It includes three mountain massif (Pyrenees, Massif Central and French Alps) which are almost invisible at the GCM scale (specially the Pyrenees). The domain also includes coastlines on the Mediterranean side and on the Atlantic side. Thanks to a better representation of the coastline at the RCM resolution, it takes better into account the sea effect on the shore climate. It was also important for us to add small islands on our evaluation domain, such as the Baleares (Spain), since they are invisible on the GCM grid and the RCM brings important information. Finally, three major rivers (plotted in blue in Figure 1) are on the domain with interesting impacts on climate (commented in Section 3): the Ebro in Spain, the Garonne in southwest of France and the Rhone on the east of the domain. This domain should therefore illustrate the added value brought by a RCM at local scale and be a good test-bed on the feasibility of emulating
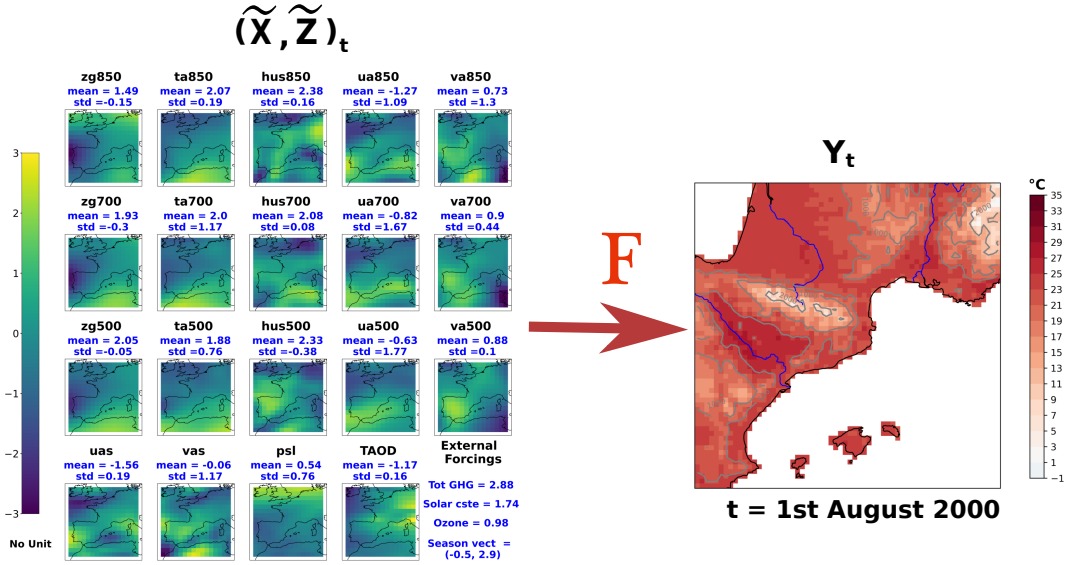
Fig. 1: Illustration of an observation. Left: each map represents a 2D input variables $(X)$, on the input domain, and the blue numbers correspond to the 1D variables $(Z)$. Right: an example of $Y$, the near surface temperature on the target domain.

high resolution models.

## 2.4 Deep learning with UNET

### 2.4.1 Neural network model as a black-box regression model

The problem of statistical downscaling and of emulation of daily near-surface variables may be seen as a statistical regression problem where we need to build the best relationship between the output response $Y$ and the input variables $(X, Z)$. When looking at the $L^2$ loss between the prediction $\hat{Y}$ and the true $Y$, the optimal link (denoted by $F$ below) is theoretically known as the conditional expectation:

$$F(X, Z) = E[Y \,|\, (X, Z)].$$

Unfortunately, since we will only have access to a limited amount of observations collected over a finite number of days, we shall work with a training set

formed by the collected observations $((X_t, Z_t), Y_t)_{1 \leq t \leq T}$ and try to build from the data an empirical estimation $\widehat{F}$ of the unknown $F$.

For this purpose, we consider a family of relationship between $(X, Z)$ and $Y$ generated by a parametric deep neural network, whose architecture and main parameters are described later on. We use the symbol $\theta$ to refer to the values that describe the mechanism of one deep neural network, and $\Theta$ as the set of all possible values of $\theta$. Hence, the family of possible relationships described by a collection of neural networks correspond to a large set $(F_\theta)_{\theta \in \Theta}$. Deep learning then corresponds to the minimization of a $L^2$ data-fidelity term associated to the collected observations:

$$\hat{F} = \arg \min_{F_\theta, \theta \in \Theta} \sum_{t=1}^{T} \|Y_t - F_\theta(X_t, Z_t)\|^2. \tag{3}$$

*2.4.2 Training a deep neural network*

To train our emulator between low resolution fields and one high resolution target variable, we used a neural network architecture called UNET whose architecture is described below. As usual in neural networks, the neurons of UNET are organised in layers. Given a set $E_{\mathbf{n}}$ of input variables denoted by $(x_i)_{i \in E_{\mathbf{n}}}$ of an individual neuron $\mathbf{n}$, the output of $\mathbf{n}$ corresponds to a non-linear transformation $\phi$ (called activation function) of a weighted sum of its inputs:

$$\mathbf{n}\left((x_i)_{i \in E_{\mathbf{n}}}\right) = \phi\left(\langle w, (x_i)_{i \in E_{\mathbf{n}}}\rangle\right).$$

The connection between the different layers and their neurons then depends on the architecture of the network. In a fully connected network (multilayer perceptron) all the neurons of a hidden layer are connected to all the neurons of the previous layer. The deepness of a network then depends on the number of layers.

As indicated in the previous paragraph, the machine learning procedure corresponds to the choice of a particular set of weights over each neuron to optimise a data fidelity term. Given a training set, a deep learning algorithm then solves a difficult multimodal minimisation problem as the one stated in (3) with the help of gradient descent strategies with stochastic algorithms. The weights associated with each neuron and each connection are then re-evaluated according to the evolution of the loss function, following the backpropagation algorithm Rumelhart and Hinton (1986). This operation is repeated over all the examples until the error is stabilized. Once the neural network is trained, it may be used for prediction, *i.e.*, to infer the value $Y$ from new inputs $(X, Z)$.

We emphasize that the bigger the dimension of the inputs and outputs, the larger the number of the parameters to be estimated and so the bigger the training set must be. Therefore, the quality of the training set is crucial: missing or wrong values will generate some additional fluctuation and errors in

the training process. Moreover, we also need to cover a sufficiently large variety of scenarios in the input variables to ensure that our training set covers a wide range of possible inputs. For all these reasons, climate simulation datasets are ideal to train deep learning networks.

### 2.4.3 UNET architecture

UNET is a specific network architecture that has been introduced in Ronneberger et al (2015) for its good abilities in image segmentation problems, which consists in identifying different objects and areas in an image. This incolves at gathering pixels that correspond to the same object. This key feature is of course naturally interesting for meteorological maps since the emulator needs to identify the different meteorological structures present in the low resolution predictors for a given day, in order to produce the corresponding high resolution near-surface temperature.

UNET is a fully convolutional neural network (LeCun et al, 1998), as it is composed only of convolutional layers. A convolutional layer applies different moving filters to the input images in order to decompose them in a set of features. The user choose the size of the filters and their number. This feature's extraction allows the network to decompose images in many features in order to identify the relevant part for the target prediction.

From a technical point of view, the original UNET (Ronneberger et al, 2015) is composed of two paths: the contracting one (also called encoder) and the expansive one (decoder). Figure 2 represents the UNET architecture scheme used in this study. It contains a cascade of standard steps of convolutional neural networks, that are described below. The different layers used in the network are:

– Each blue arrow $\Rightarrow$ corresponds to a convolution block composed of a layer built with a set of convolutional $2 \times 2$ filters. The number of filters increases all along the contracting part (the number of filters is respectively 64, 128, 256, 512 and 1024 and is given on the top of each block in Figure 2). The outputs of this layer are then normalised with a batch normalisation layer (see *e.g.* Ioffe and Szegedy (2015)) to improve the statistical robustness of the layer. Finally the ReLu activation function completes this block.

$$ReLu \; : \; R(z) = max(0, z)$$

– Each red arrow $\Downarrow$ is a Maxpooling layer. It performs $2 \times 2$ pooling on each feature map, which simply divides by 2 the spatial dimension by taking the maximum of each $2 \times 2$ block. It is applied through all convolution block outputs in the encoding path. The size of the images is indicated on the side of each block on Figure 2.

– Each green arrow $\Uparrow$ is a transpose convolution layer. It allows to perform up-sampling in the expansive part of the algorithm. It multiplies the spatial dimension by 2 by applying the same connection as the classical convolution but in the backward direction.

– The black arrow $\Longrightarrow$ represents a fully connected dense network of 4 layers which is applied on the 1-dimensional inputs ($Z$).
– The grey arrow $\Rightarrow$ represents a simple concatenation layer which recalls the layers from the encoding path in the decoding one.
– Finally, the light blue arrow $\Rightarrow$ is the output layer, which is a simple convolutional layer with a single filter and a linear activation function.

The Unet proposed here takes the same basis as the original one but enlarges the expansive path by a succession of up-sampling layers in order to reach target size (Figure 2). Moreover, the Unet architecture allowed us to add a 1-dimensional input (see Section 2.2 on Predictors) at the bottom of the "U" after a short fully dense network of 4 layers. At this point the neural network concatenates the encoded spatial information (from X) and the encoded temporal information (from Z). As illustrated in Figure 2 we force the Unet to give the same importance to these 2 inputs before starting the decoding path and recreating the high resolution temperature map.

We chose the Mean Square Error (MSE) as the loss function to train the network, as we are in a regression problem. Moreover, it is well adapted for variables following gaussian-like distribution such as temperature. The neural network was built and trained using the Keras Tensorflow environment in Python (*https://keras.io*). The network trained for this study has about 25 billion parameters to fit.

2.5 Training of the emulator : Perfect Model Framework

As any statistical downscaling and any machine learning method, the emulator needs to be calibrated on a training set. It consists in showing the emulator many examples of predictors and their corresponding target such that the parameters of the network can be fitted as mentioned in Section 2.4.2. The emulator is trained in a perfect model framework, with both predictors and predictands coming from the same RCM simulation. The intuitive path to train the emulator is to use GCM outputs as predictors and its driven RCM outputs as predictands, but there are many reasons for our choice. First of all, it guarantees a perfect temporal correlation between large scale predictors and a local scale predictand. Indeed, Figure 3 shows that GCM and RCM large scales are not always well correlated, with an average correlation of 0.9 and 10% of the days with a coefficient of correlation lower than 0.75. These mismatches are quite well known and often due to internal variability as explained by Sanchez-Gomez et al (2009); Sanchez-Gomez and Somot (2018). Moreover, there are more consistent biases (discussed in Section 4.1) between GCM and RCM large scales. It is of primary importance that the inputs and outputs used to calibrate the model are perfectly consistent, otherwise, the emulator will try to learn a non-existing or non-exact relationship. In this context, the perfect model framework allows us to focus on the downscaling function of the RCM, specifically.
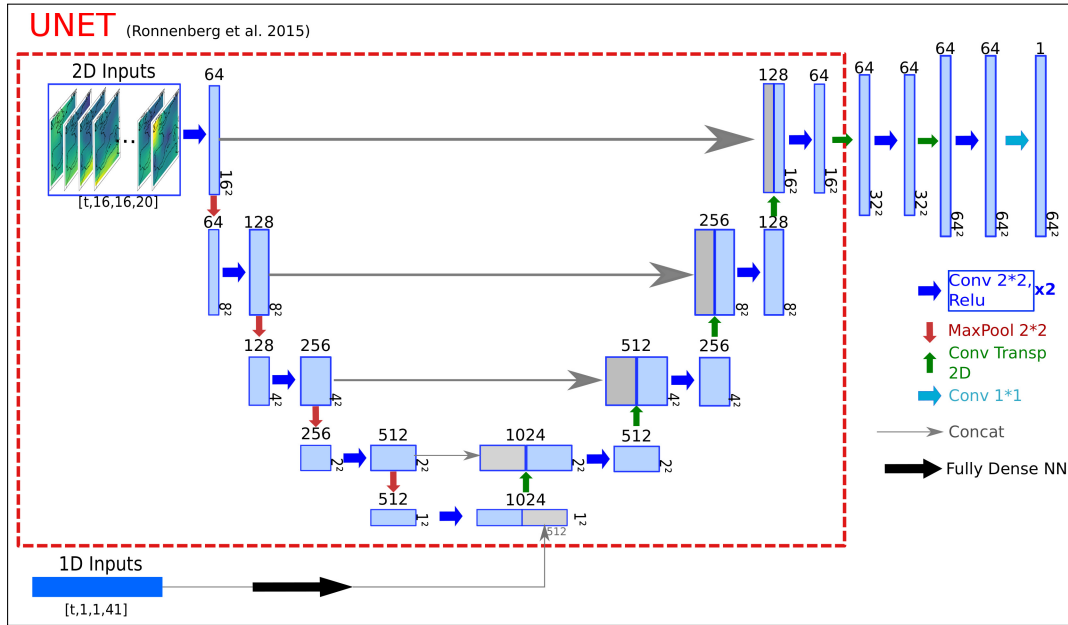
Fig. 2: Scheme of the neural network architecture used for this paper. The part of the network in the red frame corresponds to the original unet defined in Ronneberger et al (2015)

The training protocol is summarised in Figure 4a. In a first step, the RCM simulation outputs are upscaled to the GCM resolution (about 150km) thanks to a conservative interpolation. This first step transforms the RCM outputs into GCM-like outputs. This upscaled RCM is called UPRCM in the rest of this paper. In the second step, these UPRCM outputs are smoothed by a 3x3 moving average filter to respect the protocol described in Section 2.2. This smoothing also ensures to delete local scale information which might persist through the upscaling.

The near-surface temperature on the target domain is extracted from the same RCM simulation. Following this procedure, the emulator is trained using the ALADIN63 simulation forced by CNRM-CM5, covering the 1950-2100 period with the RCP8.5 scenario from 2006. We chose the two most extreme simulations (historical and RCP8.5) for the training in order to most effectively cover the range of possible climate states since the emulator does not target to extrapolate any information. Future studies could explore the best combination of simulations to calibrate the emulator.
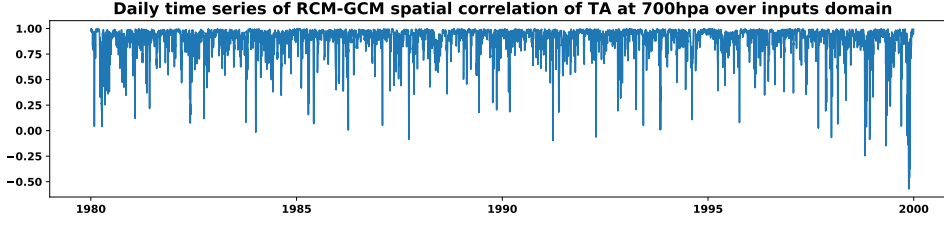
Fig. 3: Time series of spatial correlation of the atmospheric temperature at 700hpa between ALADIN63 and its driving GCM, CNRM-CM5, over the input domain.
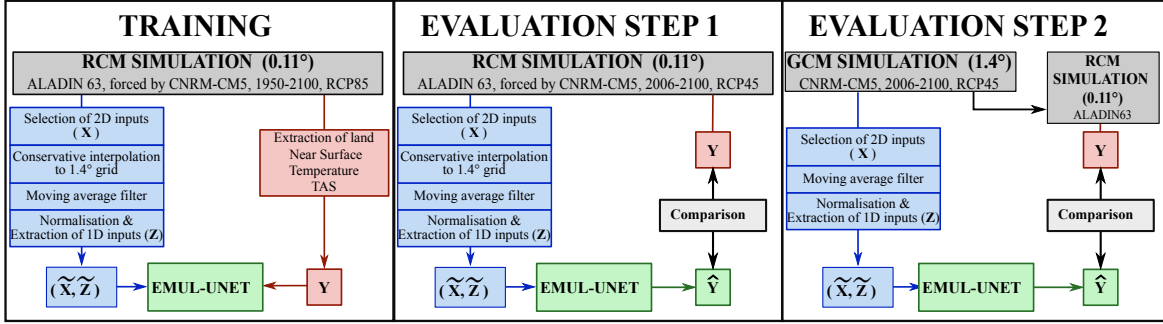


Fig. 4: Scheme of the protocols for the training (a) and the two steps of evaluations (b and c).

## 2.6 Evaluation metrics

The emulated temperature series ($\hat{Y}$) will be compared to the original RCM series ($Y$) (mentioned as "RCM truth" in the rest of this paper) through statistical scores described below. Each of these metrics will be computed in each point over the complete series:

– **RMSE**. The Root Mean Squared Error measures the standard deviation of the prediction error (in °C):

$$RMSE(Y,\widehat{Y}) = \sqrt{\frac{1}{T}\sum_t (Y_t - \widehat{Y}_t)^2}$$

– **Temporal Anomalies Correlation**.This is the Pearson correlation coefficient after removing the seasonal cycle:

$$ACC(Y,\widehat{Y}) = \rho(Y_a, \widehat{Y_a}),$$

with $\rho$ the Pearson correlation coefficient and $Y_a$ and $\widehat{Y_a}$ are the anomaly series after removing a seasonal cycle computed on the whole series.

– **Ratio of Variance**. It indicates the performance of the emulator in reproducing the local daily variability. We provide this score as a percentage.

$$RoV(Y, \widehat{Y}) = \frac{Var(\widehat{Y})}{Var(Y)} * 100$$

– **Wasserstein distance**. It measures the distance between two probability density functions $(P, Q)$. It relies on the optimal transport theory (Villani, 2009) and measures the minimum required "energy" to transform $P$ into $Q$. The energy here corresponds to the amount of distribution weight that is needed to be moved multiplied by the distance it has to be moved. In this study we use the 1-d Wasserstein distance, and its formulation between two samples becomes a rather simple function of ordered statistics:

$$W_1(f(Y), f(\widehat{Y})) = \sum_{i=1}^{T} |Y_{(i)} - \widehat{Y_{(i)}}|$$

with $f(\bullet)$ the probability density function associated with the sample $\bullet$.
– **Climatology**. We compare the climatology maps over present (2006-2025) and future (2081-2100, not shown in Section 3) climate. The RCM truth and emulator maps are shown with their spatial correlation and RMSE. The error (emulator minus RCM) map is also computed.
– **Number of days over 30°C**. Same as climatology for the maps showing the number of days over 30°C.
– **99th Percentile**. Same as climatology for the maps showing the 99th percentile of the daily distribution.
– **Climate Change**. Climate change maps for the climatology, the number of days over 30°C and the 99th percentile (delta between future (2080-2100) and present (2006-2100) period).

These metrics are at the grid point scale and are presented as maps. However, to summarise these maps with few numbers we can compute their means and their super-quantile of order 0.05 (SQ05) and 0.95 (SQ95). The super-quantile $\alpha$ is defined as the mean of all the values larger (resp. smaller) than the quantile of order $\alpha$, when $\alpha$ is larger (resp. smaller) than 0.5. These values are shown in the Figures of Section 3 and Tables 3 and 4 in supplementary material.

2.7 Benchmark

For this study, we propose as benchmark the near surface temperature from the input simulation (before the moving average smoothing), interpolated on the target grid by bilinear interpolation. As this study is the first to propose such an emulator, there is no already established benchmark. The one proposed here is a naive high-resolution prediction given available predictors (low-resolution). It allows the reader to locate the emulator somewhere in between the simplest possible downscaling (simple interpolation) and the most

complex one (RCM). All the metrics introduced in Section 2.6 will be applied to our benchmark.

## 3 Results

This section presents the emulator performances in terms of its computational costs (in Section 3.1) and its ability to reproduce the near surface temperature time series at high resolution. As illustrated in Figure 4, we will evaluate the emulator in two steps, (1) in the perfect model world (Section 3.2) and (2) when the emulator inputs come from a GCM simulation (Section 3.3). The RCM simulation used to evaluate the model (also called target simulation) is the ALADIN63, RCP45, 2006-2100 forced by the GCM simulation CNRM-CM5, RCP45, 2006-2100. Note that the emulator never saw the target simulation during the training phase. This evaluation exercise illustrates a potential use of the emulator: downscaling a scenario that the RCM has not previously downscaled.

### 3.1 Computational efficiency

The emulator is trained on a GPU (GeForce GTX 1080 Ti). About 60 epochs are necessary to train the network, and each epoch takes about 130 seconds with a batch size of 100 observations. The training of the emulator takes about 2 hours. Once the emulator is trained, the downscaling of a new low resolution simulation is almost instantaneous (less than a minute). It is a significant gain in time compared to RCM, even if these time lengths do not include the preparation of the inputs, which depends mainly on the size of the series to downscale and on the data accessibility. It would, when including the input preparation, take only a few hours on a simple CPU or GPU to produce a simulation with the trained emulator, while it takes several weeks to perform a RCM simulation on a super-computer.

### 3.2 Evaluation step 1 : Perfect model world

In a first step, the emulator is evaluated in the perfect model world, meaning that the inputs come from the UPRCM simulation. This first evaluation step is necessary to control the performances of the emulator in similar conditions as during its training. Moreover, the perfect model framework guarantees perfect temporal matches between the large scale low resolution fields (the inputs) and the local scale high resolution temperature (the target). The emulator should then be able to reproduce perfectly the temperature series that is simulated by the RCM. The benchmark for this first evaluation is the UPRCM near surface temperature re-interpolated on the RCM grid. It is referred as "I-UPRCM".
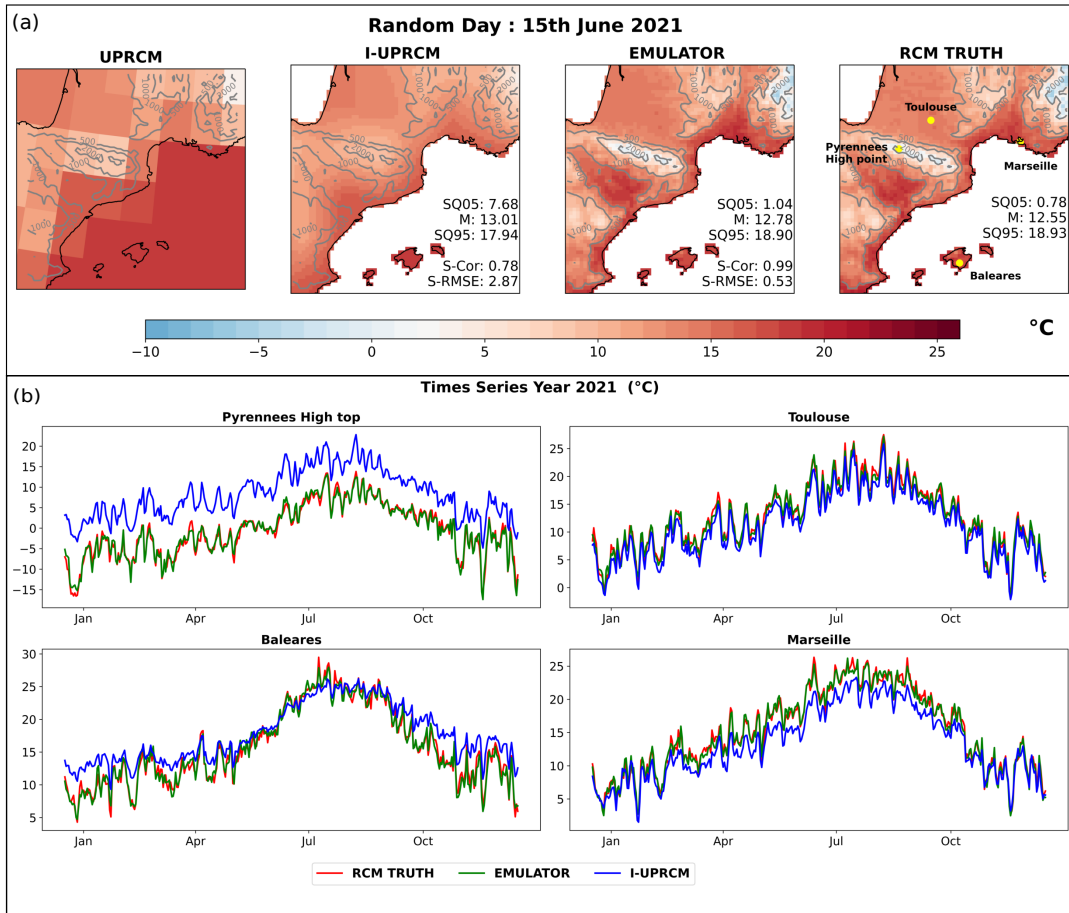
Fig. 5: Randomly chosen illustration of the production of the emulator with inputs coming from UPRCM: (a) temperature (°C) at a random day over the target domain for the RCM truth, emulator, interpolated UPRCM, and (b) random year time series (°C) for 4 particular grid points.

Figure 5a illustrates the production of the emulator for a random day regarding the target and the benchmark. The RCM truth map presents a refined and complex spatial structure largely missing in the UPRCM map. Moreover, it is evident on the I-UPRCM map that the simple bilinear interpolation does not recreate these high resolution spatial patterns. The emulator shows for this given day an excellent ability to reproduce the spatial structure of the RCM truth. It has very accurate spatial correlation and RMSE and estimates the right temperature range. On Figure 5b, we show the daily time series for four specific points shown on the RCM truth map (Marseille, Toulouse, a high Pyrenees grid point and a point in Majorca) during a random year. The

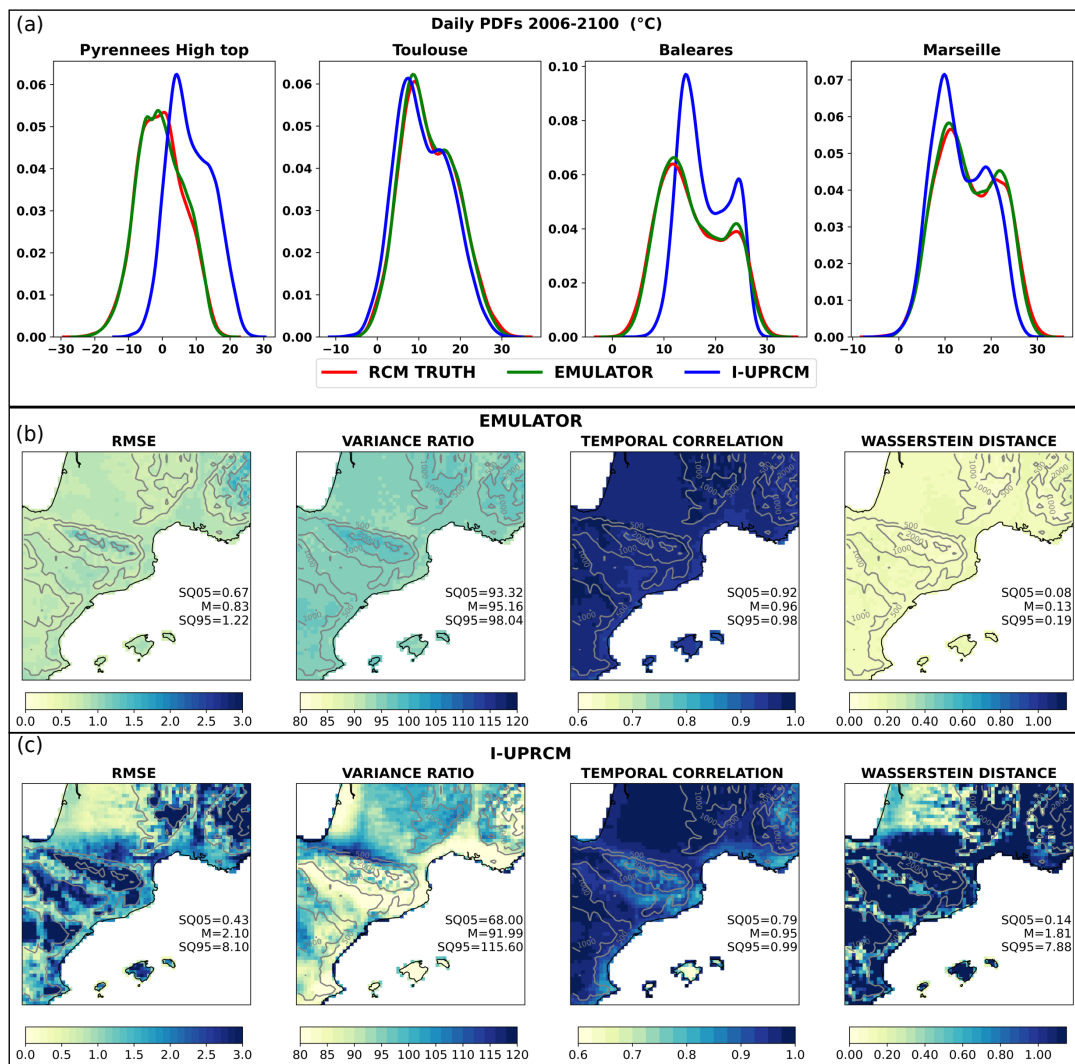Fig. 6: (a) Daily probability density functions from the RCM truth, the emulator and the I-UPRCM at 4 particular grid points over the whole simulation period. (b) (resp. (c)) Maps of performance scores of the emulator (resp. of the I-UPRCM) with respect to the RCM truth computed over the whole simulation period. For each map, the values of the spatial mean and super-quantiles (SQ05 and SQ95) are added.
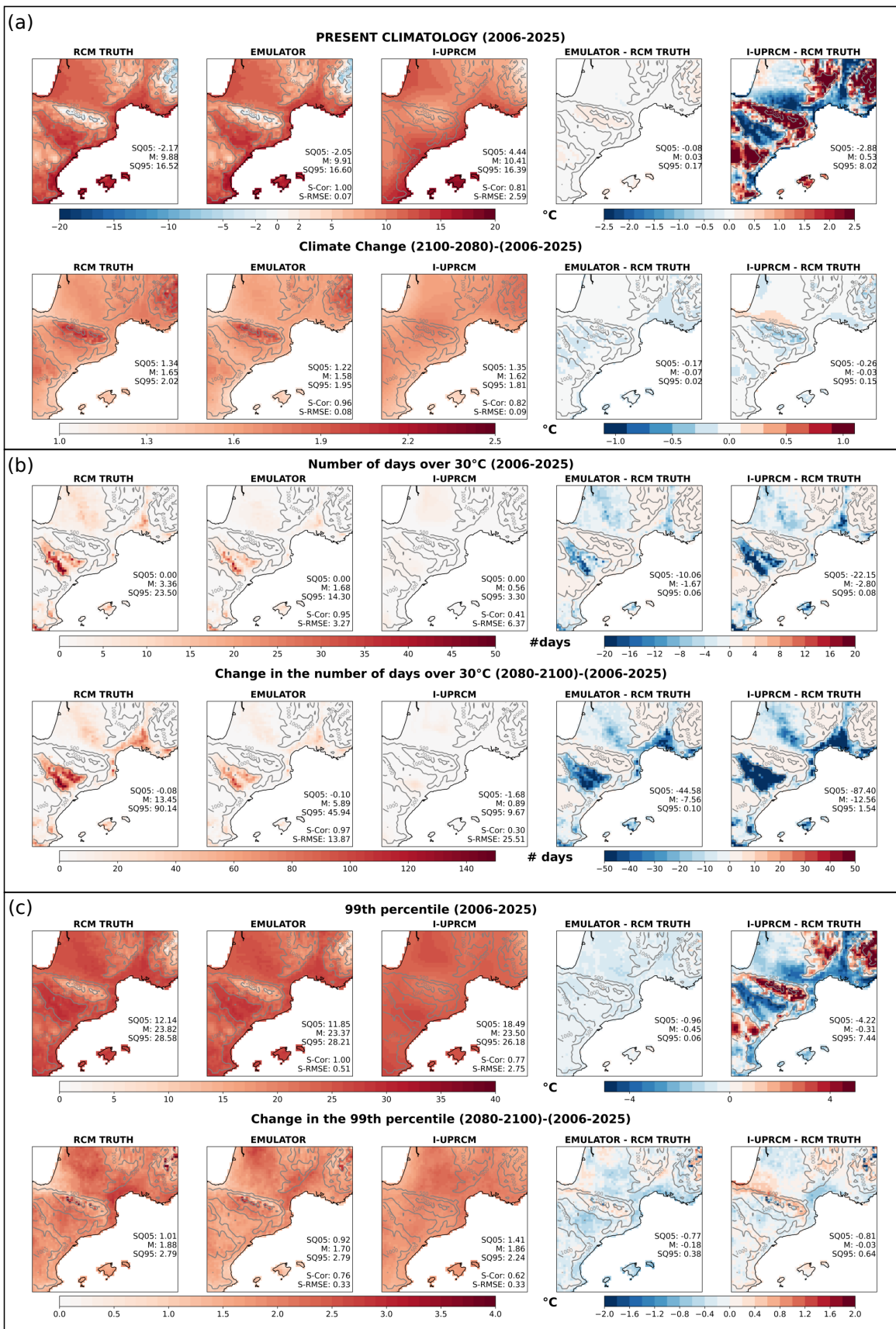
Fig. 7: (a) Maps of long-term mean climatologies, (b) number of days over 30°C and (c) 99th percentile of daily near-surface temperature for a present-climate period (2006-2025) and for the climate change signal (2080-2100 minus 2006-2025), for the RCM truth, the emulator and the I-UPRCM. On each line, the two last maps show the error map of the emulator and the I-UPRCM. For each map, the spatial mean and super-quantiles (SQ05 and SQ95) are added, as well as the spatial correlation and spatial RMSE for the emulator and I-UPRCM maps.

RCM transforms the large scale temperature (visible on I-UPRCM) differently over the four points. In the Pyrenees, the RCM shifts the series and seems to increase the variance. In Marseille, it appears to produce a warmer summer without strongly impacting winter characteristics, and it seems to be the exact opposite in Majorca. On the contrary, in Toulouse, I-UPRCM and RCM are close. For each of these 4 cases, the emulator reproduces the sophistication of the RCM series almost perfectly.

Figure 5a gives the impression that the emulator has a good ability to reproduce the complex spatial structure brought by the RCM, and we can generalise this result with the other figures. First of all, the spatial correlation (equal to 1) and the very low spatial RMSE ($0.07°$C) of the climatology maps in Figure 7 in the present climate confirm that this good representation of the temperature's spatial structure is robust when averaging over long periods. In particular, it is worth noting that the altitude dependency is well reproduced as well as the warmer patterns in the Ebro and Rhone valley or along the coastlines. The performance scores (Figure 6b) of the emulator support this result. Their spatial homogeneity tends to show that the emulator does not have particular difficulties over complex areas. Moreover, the comparison with the interpolated UPRCM shows the added-value of the emulator and in particular its ability to reinvent the fine-scale spatial pattern of the RCM truth from the large-scale field. Indeed, the score maps of the I-UPRCM (Figure 6c) or the present climatology error map shows strong spatial structures, highlighting the regions where the RCM brings added value and that the emulator reproduces successfully.

The RCM resolution also allows to have a better representation of the daily variability at the local scales over critical regions. The difference in variance between the I-UPRCM and the RCM is visible on Figure 6c. The I-UPRCM underestimates the variability and is poorly correlated over the higher reliefs, the coastlines and the river valley. In contrast, the emulator reproduces more than 90% of the RCM variance over the whole domain. The RMSE and temporal correlation maps of the emulator confirm the impression given by Figure 5b that it sticks almost perfectly to the RCM truth series. Moreover, the RCM daily variability is strongly dependent on the region. Indeed, the RCM transforms the "I-UPRCM" pdfs in different ways across the domain (visible on Figures 6ac). Figures 6ab show that the emulator succeeds particularly well in filling these gaps.

The emulator's good representation of the daily variability and temporal correlation involves a good representation of the extreme values. The probability density functions of the four specific points on Figure 6a show that the entire pdfs are fully recreated, including the tails. The Wasserstein distance map extends this result to the whole target domain. The two extreme scores computed for the present climate on Figure 7bc confirm these results. The 99th percentile emulated map is almost identical to the target one verified

by the difference map, with a maximum difference of less than 1°C for values over 35°C. The spatial pattern of the 99th percentile map is here again correctly captured by the emulator, particularly along the Garonne river that concentrates high extremes. The number of days over 30°C is a relatively more complicated score to reproduce since it involves an arbitrary threshold. The emulator keeps performing well with a high spatial correlation between the emulator and the RCM truth. However, it appears that the emulator misses some extreme days, involving a lack in the intensity of some extremes.

Finally, the high-resolution RCM produces relevant small-scale structures in the climate change maps. In particular, RCMs simulate an elevation-dependent warming (see the Pyrenées and Alps areas Kotlarski et al, 2015), a weaker warming near the coasts (see the Spanish or Atlantic coast) and a specific signal over the islands as shown in the second lines of Figures 7ac. It can be asked if the emulator can reproduce these local specificities for the climate change signal. The emulator is able to capture this spatial structure of the warming but with a slight lack of intensity which is general over the whole domain. The reproduction of the climate change in the extremes suffers the same underestimation of the warming but also offers the same good ability to reproduce the spatial structure, with high spatial correlation.

This first evaluation step shows that if the emulator is still perfectible, in particular when looking at extremes or climate change intensity, it is able to almost perfectly reproduce the spatial structure and daily variability of the near surface temperature in the perfect model world.

3.3 Evaluation step 2 : GCM world

In this second evaluation step, we directly donwscale a GCM simulation that has not been downscaled by the RCM. The benchmark for this evaluation is the near surface temperature from the GCM, interpolated on the target grid. It will be referred to as I-GCM. The emulated series and the benchmark are compared to the RCM simulation driven by the same GCM simulation.

Figure 8a illustrates the production of the emulator regarding the benchmark and RCM truth for the same day as Figure 5a. First of all, as for the I-UPRCM, the I-GCM map does not show any of the complex RCM spatial structures. The I-GCM is less correlated with the RCM and warmer than the I-UPRCM. In contrast, the emulator reproduces the complex spatial structure of the RCM very well with a spatial correlation of 0.98 but appears to have a warm bias with respect to the RCM truth. The four time series are consistent with the previous section, with fundamental differences between I-GCM and RCM, which the emulator captured very well. However, the correlation between the emulator and the RCM seems to be not as good as in the perfect

model framework.

Figure 8b is a very good illustration of the RCM-GCM large scale de-correlation issue presented in Section 2.4.2. Indeed the less good correlation of the emulator with the RCM is probably due to mismatches between GCM and RCM large scales. For instance,in the beginning of November, on the time series shown on Figure 8b, the RCM seems to simulate a cold extreme on the whole domain, which appears neither in the interpolated GCM nor in the emulator. The same kind of phenomenon occurs regularly along the series and is confirmed by lower temporal correlations between the RCM truth and the I-GCM (Figure 9c) than with the I-UPRCM (Figure 6c). According to this, the emulated series can not present a good temporal correlation with the RCM truth since it is a daily downscaling of the GCM large scale. Keeping in mind these inconsistencies, it is still possible to analyse the performances of the emulator if we leave aside these scores which are influenced by the poor temporal correlation (RMSE, ACC).

As in the first step of the evaluation (Section 3.2), the spatial structure of the RCM truth is well reproduced by the emulator. The present climatology map (Figure 10a) has a perfect spatial correlation with the RCM. The added value from the emulator is clear if compared to the interpolated GCM. The spatial temperature gradients simulated by the GCM seem to be mainly driven by the distance to the sea. On the other hand, the emulator manages to recreate the complex structures created by the RCM, related to relief and coastlines. The emulator capacity to reproduce the RCM spatial structure seems as good as in Section 3.2. The scores in Figure 9b present a good spatial homogeneity, exactly like in the previous section 6b.

The error map in Figure 10(a) shows that the emulator is warmer in the present climate than the RCM truth ($+0.96°C$). This bias presents a North-South gradient with greater differences over the North of the target domain, which is consistent with the Wasserstein distance map on Figure 9b. The Wasserstein metric shows that the density probability functions from the emulated series are further away from the RCM truth with GCM-inputs than in UPRCM mode. The similarities between the Wasserstein scores and the present climatology difference map indicate that the emulator shifts the mean.

The daily variability is well reproduced by the emulator. As mentioned before, the weaker RMSE (Figure 9b) is mainly due to the lower correlation between GCM and RCM. But the ratio of variance demonstrates that the emulator manages to reproduce the daily variability over the whole domain. The RCM brings a complex structure of this variability (higher variability in the mountains than in plains, for example), and the emulator, as in the first evaluation step, recreates this fine scale. Moreover, the daily pdfs of the emulator (Figure 9a) are very consistent with the RCM ones, and the same range of

values is covered for each of the four particular points.

This good representation of daily variability tends to suggest that the emulator can reproduce the local extremes. Figures 10bc confirm these results, with a very high spatial correlation between the emulator and the RCM truth maps in present climate. The warmer extremes along the three river valleys are present in both RCM and emulator maps, while they are absent from the I-GCM maps. The warm bias observed in the present climatology map also impacts these scores. The emulator map of the number of days over $30°C$ in the present climate shows more hot days than the RCM but the same spatial structure. The map of the 99th percentile over the 2006-2025 period shows the same observation, with a warm bias $(+0.82)$ slightly lower than the climatology bias.

Finally, the climate change signal is also well captured by the emulator. The different spatial patterns that bring the high resolution of the RCM in the Figures 10abc are also visible in the emulator climate change Figures. The emulator represents a weaker warming than the RCM, observable in average warming but also on the map of extremes. This underestimated warming is mainly due to the warm bias between GCM and RCM, which is less intense in the future. For instance, the warming from the emulator is $0.27°C$ weaker on average over the domain (with almost no spatial variation) than in the RCM. This number corresponds approximately to the cold bias from the I-GCM $(0.19°C)$ plus the missed warming by the emulator in the perfect model framework $(0.07)$. This tends to show that the emulator performs well in the GCM world but reproduces the GCM-RCM biases.

This section shows that the emulator remains robust when applied to GCM inputs since it provides a realistic high-resolution simulation. As in the first step, the emulator exhibits several desirable features with an outstanding ability to reproduce the complex spatial structure of the daily variability and climatology of the RCM. We also showed that the emulator remains consistent with its driving large scale, which leads to inconsistencies with the RCM. In the next section, we will develop this discussion further.
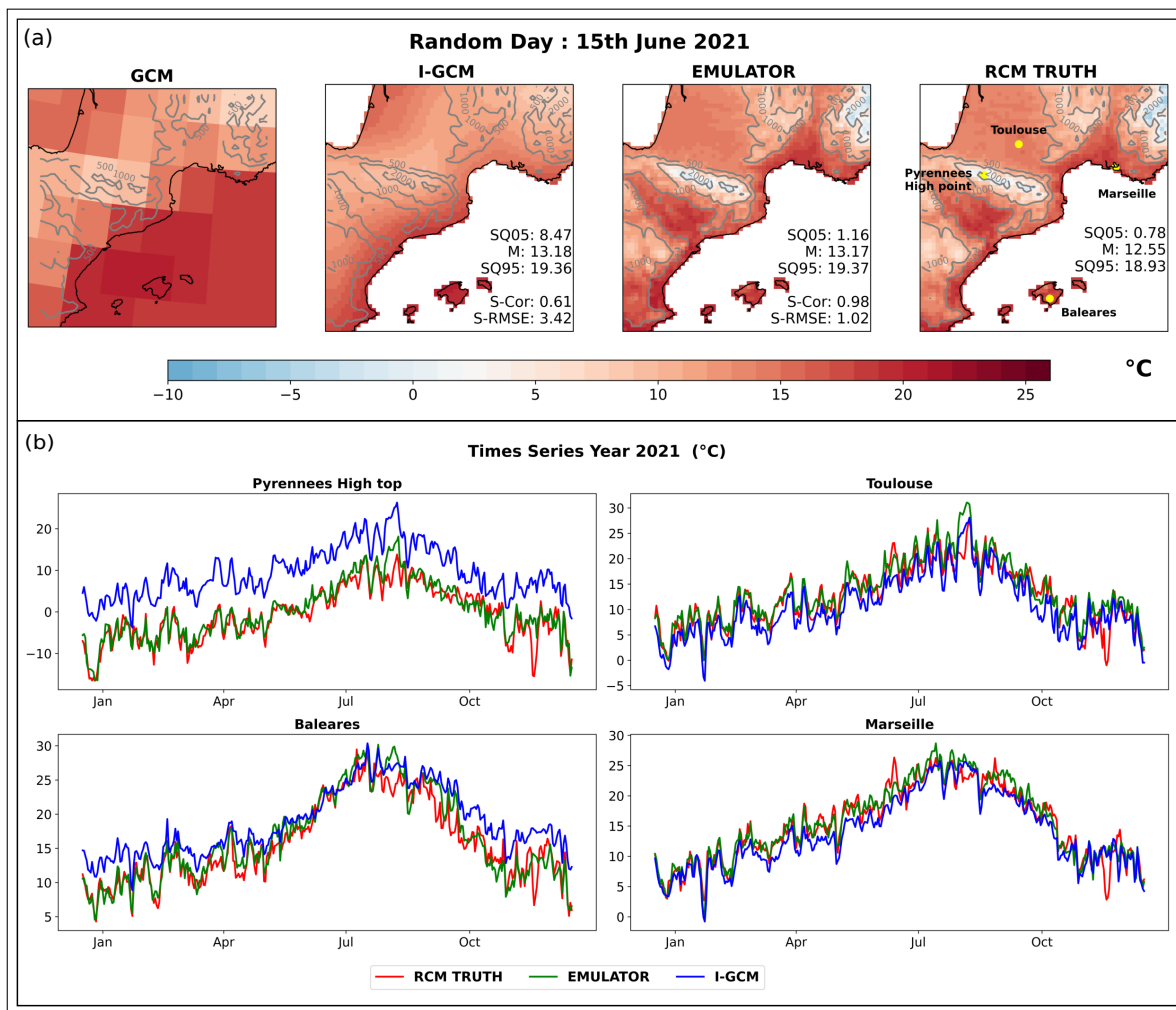
Fig. 8: Randomly chosen illustration of the production of the emulator with inputs coming from the GCM: (a) near-surface temperature (°C) at a random day over the target domain for the RCM truth, the emulator, the interpolated GCM, and the GCM and (b) random year time series (°C) for 4 particular grid points.

## 4 Discussion

### 4.1 On the inconsistencies between GCM and RCM

Several recent studies (Sørland et al, 2018; Bartók et al, 2017; Boé et al, 2020) have highlighted the existence of large scale biases for various variables between RCMs and their driving GCM, and have discussed the reasons behind
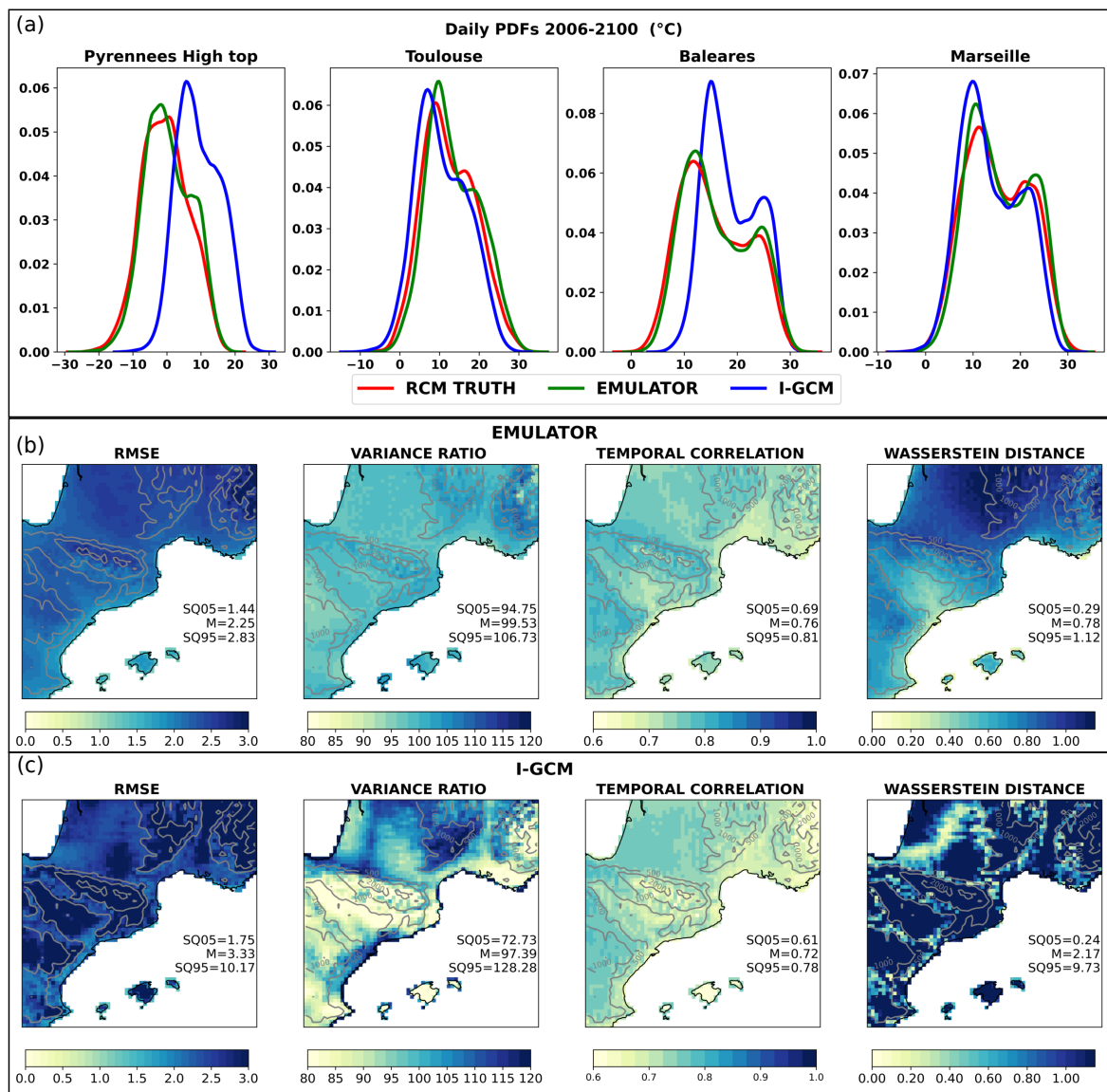
Fig. 9: (a) Daily probability density functions from the RCM truth, the emulator and the I-GCM at 4 particular grid points over the whole simulation period. (b) (resp. (c)) Maps of performance scores of the emulator (resp. of the I-GCM) with respect to the RCM truth computed over the whole simulation period. For each map, the spatial mean and super-quantiles (SQ05 and SQ95) are added.
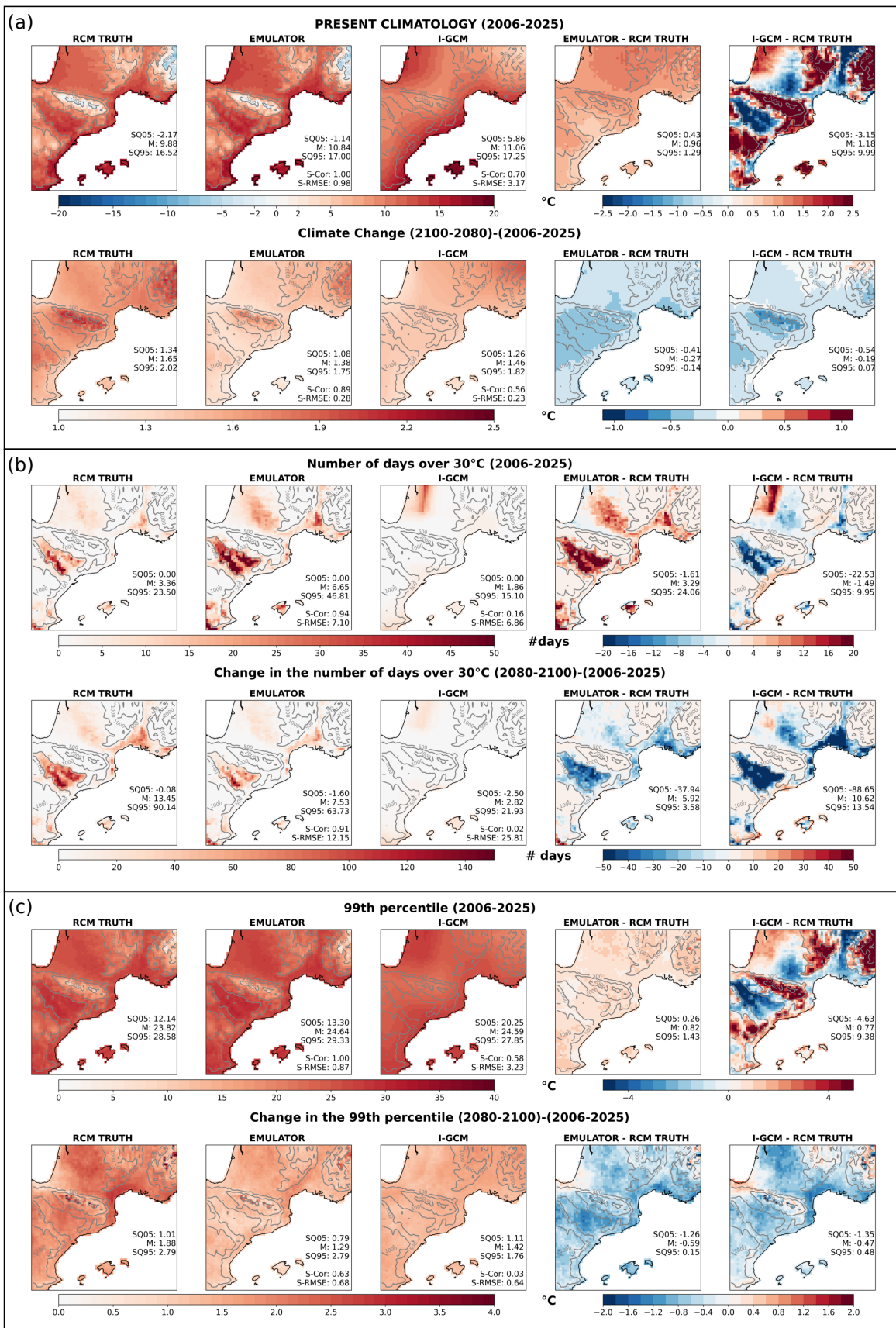
Fig. 10: (a)Maps of climatologies, (b) Number of days over 30°C and (c) the 99th percentile of daiy temperature, for the present-climate period (2006-2025) and for the climate change signal (2080-2100 minus 2006-2025), for the RCM truth, the emulator and the I-GCM. On each line, the two last maps show the error maps of the emulator and the I-GCM. For each map, the spatial mean and super-quantiles (SQ05 and SQ95) are added, as well as the spatial correlation and spatial RMSE for the emulator and I-GCM maps.

these inconsistencies. From a theoretical point of view, it is still controversial as to whether these inconsistencies are for good or bad reasons (Laprise et al, 2008) and therefore if the emulator should or should not reproduce them. In our study, the emulator is trained in such a way that it focuses only on learning the downscaling function of the RCM, i.e., from the RCM large scale to the RCM small scale. Within this learning framework, the emulator can not learn GCM-RCM large-scale inconsistencies, if there should be any. Therefore, when GCM inputs are given to the emulator, the estimated RCM downscaling function is applied to the GCM large scales fields, and any GCM-RCM bias is conserved between the emulated serie and the RCM one. Figure 11 shows the biases for the present-climate climatology between the GCM and the UPRCM over the input domain for TA700 and ZG700, at the GCM resolution. The GCM seems generally warmer than the UPRCM, which could partly explain the warm bias observed between the emulator results and the RCM truth in present climate (e.g., Figure 10). These large scale biases between GCM and RCM raise the question of using the RCM to evaluate the emulator when applied to GCM data. Indeed, if these inconsistencies are for bad reasons (e.g., inconsistent atmospheric physics or inconsistent forcings), the emulator somehow corrects the GCM-RCM bias for the emulated variable. In this case, the RCM simulation cannot be considered as the targeted truth. However, if the RCM revises the large scale signal for good reasons (e.g., upscaling of the local added-value due to refined representation of physical processes), then the design of the emulator should probably be adapted.

In future studies, we plan to use RCM runs with spectral nudging (Colin et al, 2010), two-way nested GCM-RCM runs or global high-resolution simulations for testing other modelling frameworks to further develop and evaluate the emulator.

4.2 On the stationary assumption

In the introduction of this paper, we state that the stationary assumption is one of the main limitations of empirical statistical downscaling. The emulator proposed here is similar in many ways to a classical ESD method, the main difference being that the downscaling function is learnt in a RCM simulation. The framework used to train the emulator is a good opportunity to test the stationary assumption for the RCM-emulator. We train the same emulator, with the same neural network architecture and same predictor set, but on the historical period (1951-2005) only. Results are reported in Tables 3 and 4 in the supplementary material, this version being named 'Emul-Hist'.
The perfect model (Table 3) evaluation constitutes the best way to evaluate the validity of this assumption properly. Emul-Hist has a cold bias over the whole simulation regarding the RCM truth and the range of emulators described in subsection 4.4. Moreover, this bias is much stronger for the future period (from 0.3°C in 2006-2025 to 1.3°C in 2080-2100). Emul-Hist manages
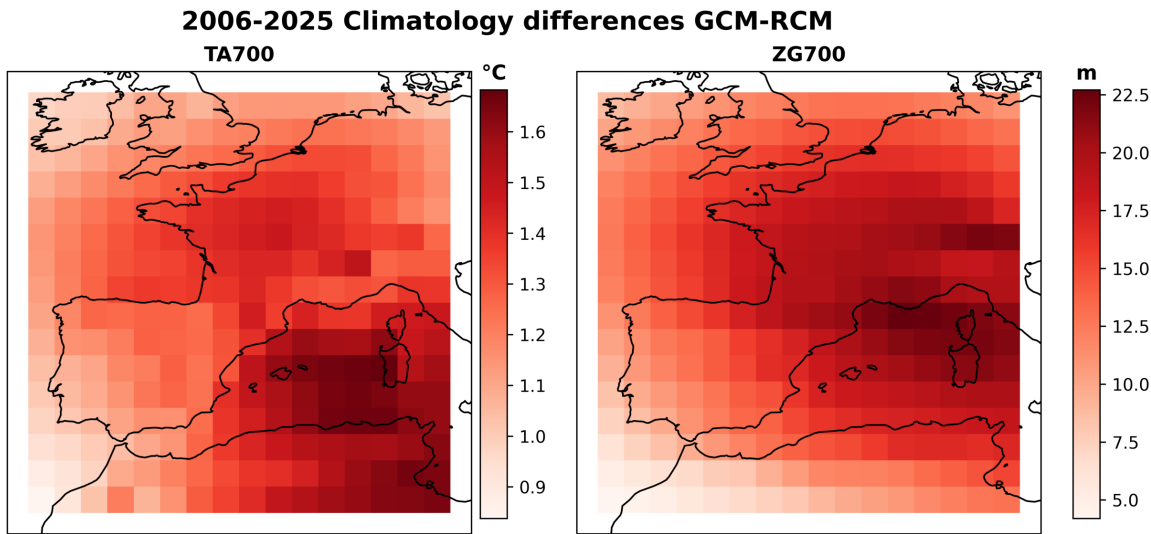
Fig. 11: Present (2006-2025) climatology differences for the atmospheric temperature and geopotential at 700 hpa: CNRM-CM5 RCP45 minus ALADIN63 driven by CNRM-CM5 RCP45 upscaled on the GCM grid.

to reproduce only 30% of the climate change simulated by the RCM. It also fails to capture most of the spatial structure of the warming since the spatial correlation between the Emul-hist and RCM climate change maps (0.86) is close to the I-UPRCM (0.82) and largely weaker that for the main emulator (0.95) (see Figure 12). The Emul-hist average RMSE (1.35°C) over the whole series is also out of emulator range ([0.8; 0.86]). Results in GCM evaluation are also presented (Table 4), but due to the lack of proper reference, it is difficult to use them to assess the stationary assumption. However, it presents the same cold bias regarding the ensemble of emulators. These results demonstrate the importance of training the emulator in the wider range of possible climate states.

We underline that not all ESD methods are expected to behave that poorly with respect to projected warming. However learning in the future is one of the main differences between our RCM emulator approach and the standard ESD approach that relies on past observations.

4.3 On the selection of the predictors

For this study, we chose to use a large number of inputs with almost no prior selection, leaving the emulator to select the right combination of inputs for each grid point. However, we are aware that it involves a lot of data, which is not always available, and leads to several computations due to the different
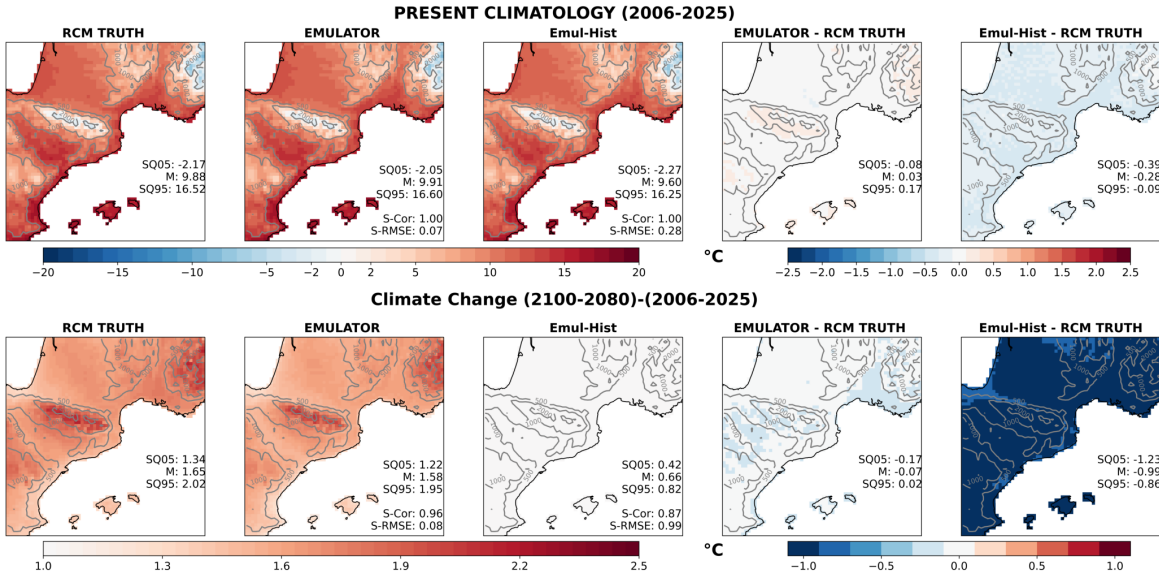
Fig. 12: Maps of climatologies over present period (2006-2025) and climate change signal (2080-2100 versus 2006-2025), for the RCM truth, the Emulator presented in section 3 and the Emul-Hist. The two last columns correspond to the error maps of the Emulator and Emul-Hist with respect to the RCM truth. For each map, the spatial mean and super-quantiles (SQ05 and SQ95) are added, as well as the spatial correlation and spatial RMSE for the Emul-Hist and I-UPRCM maps.

preprocessing steps described in Section 2. For this reason, we tried to build an emulator with fewer inputs in $X$ including only the variables from Table 2 at 700hpa and removing the solar and ozone forcings from $Z$. The results are reported in tables 3 and 4. They show that having more inputs seems to increase the quality of the emulated series, but the "cheap" emulator presents satisfying results and can be considered as a serious option. For instance, in the first step of evaluation (Table 3), the cheap emulator provides a good representation of the spatial structure (S-Cor=0.94 vs 0.96 for the main emulator) but is more biased than the main emulator (-0.3 versus -0.18 °C). Having a specific selection of inputs for specific areas would probably increase the performance of the Cheap emulator. However, this is not in the spirit of the tool proposed here, which aims to be as simple and as general as possible.

4.4 On the non reproducibility of neural network training

While neural networks have experienced considerable success over the last decades and the number of applications is constantly increasing, they have also been largely criticised for their lack of transparency due to an excessive

number of parameters. Several studies (see Guidotti et al, 2018, for review) have tried to provide the keys to "open the black box". However, users of deep neural networks should also be aware of their non-reproducibility. Indeed the training of deep neural networks with GPUs involves several sources of randomness (initialisation, operation ordering, etc.). A few recent studies have raised the issue for medical applications with a Unet (Marrone et al, 2019; Bhojanapalli et al, 2021), but to the best of our knowledge, no study has addressed it.

In order to document this issue for the RCM emulator developed in this paper and assess its robustness, we propose a Monte Carlo experiment where the same configuration of the emulator (as described in Section 2) is trained 31 times, resulting in 31 emulators. The results of this experiment are illustrated in Figure 13 and summarised in tables 3 and 4. The results of Section 3 are based on a randomly picked emulator, mentioned as the main emulator in appendix and plotted in darker green in Figure 13. All the emulators from the Monte Carlo experiment show consistent results with the ones presented in Section 3. The RMSEs (and correlated scores such as variance) are really close to each other ($[0.8; 0.86]^{\circ}C$), which is expected since it is the loss function used to fit the neural network. However, the path taken to minimise the loss during training might vary from one emulator to another, leading to bigger differences in climatological metrics (see tables). For instance, in the UPRCM evaluation step, the average error in the future climatology varies from .003 to .139°C. The results from the GCM world evaluation step (Table 4) are similar to those in the perfect model framework. This stability shows again the robustness of the emulator when using GCM large scale fields.

We believe that the readers of this study and any potential emulator users should be aware of this characteristic of deep learning neural networks. However, it is worth mentioning that this randomness in the training of the network does not impact the key conclusions of the results (Section 3), which prove their robustness. Training many times the emulator as in this test may be a way to reach probabilistic RCM-emulators and will be investigated in future studies.
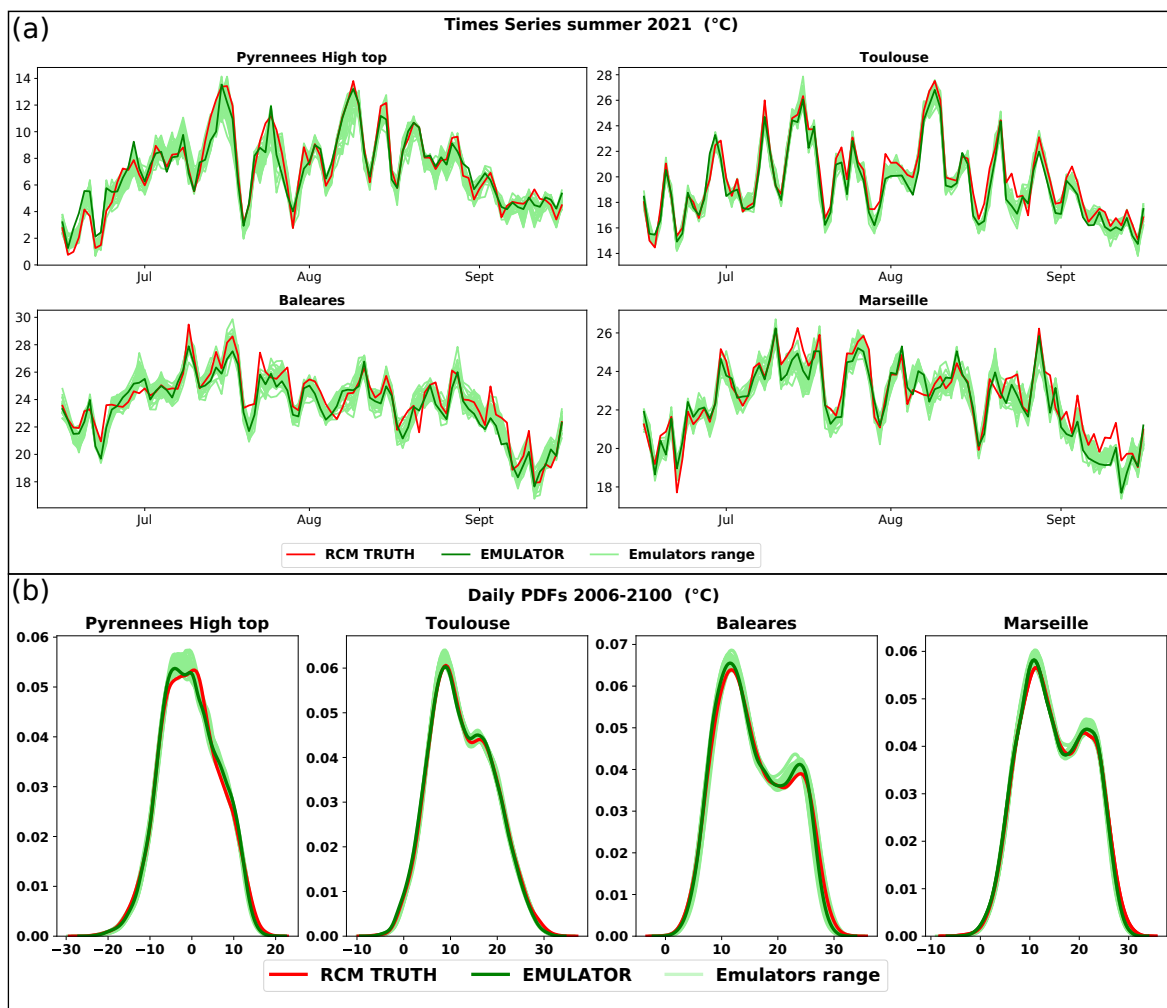
Fig. 13: Illustration of the results for the Monte Carlo experiment. (a) present a year time series for 4 given points and (b) the pdfs on the whole serie. The red line refers to the RCM truth, the dark green line is the main emulator, and the light green lines are the 30 emulators from the Monte Carlo experiment.

## 5 Conclusion

This study aims to explore a novel hybrid downscaling approach that emulates the downscaling function of a RCM. That is to say learning the transformation of the large-scale climate information into a local climate information performed by a regional climate model. Here, we develop this approach for the near surface temperature and a Southwest European domain. This new method, called RCM-emulator, is designed to help increase the size of the

high-resolution regional simulation ensembles at a lower cost.

To achieve this overall goal, we develop a specific conceptual framework. The emulator is trained using existing RCM simulations which allows it to learn the large scale/local scale relationship in different climates and in particular in future climate. Simply speaking, the general functioning of a RCM can be broken down into a large scale transformation and a downscaling function. To focus on the downscaling function, the emulator is trained in a perfect model framework, where both predictors and predictand come from the same RCM simulation. The emulator takes daily large-scale and low-resolution information as input and produces daily maps of the near surface temperature at the RCM resolution. It is worth noting that the downscaling function likely depends on the RCM choice. So the emulator developed in this study is RCM-dependent.

Technically speaking, the RCM-emulator is based on a fully convolutional neural network algorithm, called Unet. A key point of the emulator is the substantial computational gain regarding RCM computational cost. Training the emulator used in this study took two hours on a GPU. While this time depends on the target domain size it will never exceed some hours even for much bigger domains. Once the emulator is trained, the downscaling of a new low resolution simulation is instantaneous. These time lengths do not include the preparation of the inputs; nevertheless the gain remains evident when a RCM simulation involves weeks of computation on a super-computer.

The emulator is evaluated in both perfect model and GCM worlds. The results show that the emulator generally fulfils its mission by capturing very well the transformation from low resolution information to the high resolution near surface temperature. Firstly, the emulator is robust to different sources of input, which validates our conceptual framework. Secondly, the emulator succeeds very well in reproducing the high resolution spatial structure and daily variability of the RCM. The perfect model evaluation shows that the emulator is able to reproduce the original series almost perfectly. Moreover, it appears clearly that training the emulator in future climate improves its ability to reproduce warmer climate. Nevertheless the emulator shows some limitations in accurately simulating extreme events and of the complete climate change magnitude. Future work should focus on these two aspects to further improve the emulator ability. It is worth noting that similar emulators could easily be built over different domains, for different RCMs and for surface variables other than near-surface temperature.

Finally, and even if it was not our original goal, this study highlights, as others before, the RCM-GCM inconsistencies at large scales. As the emulator focuses on the downscaling function of the RCM, it does not learn to reproduce these large scale transformations. This raises the question of how to evaluate the emulator when downscaling GCM simulation. An RCM run might not

be the correct reference in such a case. Secondly, it also puts into question the final role of a RCM-emulator. Under the strong hypothesis that the large scale transformation carried out by the RCM results from physics or forcings inconsistencies – which will require further investigation – the emulator provide a corrected high-resolution series since it is consistent with the GCM large scale.

## 6 Acknowledgement

# 7 Supplementary material

Table 3: Summary of the results obtained in the 2nd step of evaluation (GCM world). The first line (RCM truth) is the RCM absolute values. The following lines are the scores according to the metrics defined in 2.6, for the different emulators described in this study. The three first lines give the mean and the range of the ensemble of emulators described in 4.4. For the ensemble minimums and maximums the name ( E1 to E30 ) of the emulator corresponding to the value is given in brackets. "Main Emulator" is the one used for the results section (3). "Cheap" emulator is the one including only the variables at 700hpa described in Section 4.3. "Emul-Hist" is the emulator trained only on the historical period.

**Distribution scores**

| | RMSE SQ05 | RMSE MEAN | RMSE SQ95 | ACC SQ05 | ACC MEAN | ACC SQ95 | Variance ratio SQ05 | Variance ratio MEAN | Variance ratio SQ95 | Wasserstein SQ05 | Wasserstein MEAN | Wasserstein SQ95 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ens MAX | 0,699 (E6) | 0,86 (E6) | 1,256 (E30) | 0,927 (E11) | 0,965 (E11) | 0,981 (E8) | 0,95 (E9) | 0,975 (E9) | 1,001 (E8) | 0,144 (E6) | 0,24 (E6) | 0,356 (E13) |
| Ens MEAN | 0,665 | 0,817 | 1,199 | 0,923 | 0,963 | 0,98 | 0,928 | 0,957 | 0,978 | 0,081 | 0,14 | 0,263 |
| Ens MIN | 0,642 (E19) | 0,8 (E11) | 1,167 (E19) | 0,913 (E13) | 0,96 (E6) | 0,978 (E6) | 0,893 (E6) | 0,921 (E6) | 0,954 (E20) | 0,04 (E9) | 0,105 (E8) | 0,18 (E28) |
| Main Emulator | 0,67 | 0,826 | 1,221 | 0,92 | 0,962 | 0,98 | 0,933 | 0,952 | 0,98 | 0,084 | 0,133 | 0,186 |
| I-UPRCM | 0,43 | 2,104 | 8,097 | 0,791 | 0,945 | 0,993 | 0,68 | 0,92 | 1,156 | 0,137 | 1,815 | 7,884 |
| Cheap | 0,92 | 1,105 | 1,343 | 0,871 | 0,932 | 0,96 | 0,937 | 0,977 | 1,018 | 0,075 | 0,152 | 0,275 |
| Emul-Hist | 1,157 | 1,348 | 1,702 | 0,897 | 0,948 | 0,971 | 0,884 | 0,926 | 0,963 | 0,735 | 0,869 | 1,006 |

**PRESENT PERIOD 2006-2025**

| | CLIMATOLOGY SpaCor | CLIM SQ05 | CLIM Mean | CLIM SQ95 | Days SpaCor | Days SQ05 | Days Mean | Days SQ95 | 99th SpaCor | 99th SQ05 | 99th Mean | 99th SQ95 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| RCM truth values | | -2,166 | 9,879 | 12,138 | | 0,000 | 3,356 | 23,500 | | 16,518 | 23,817 | 28,580 |
| Ens MAX | 1,0 (E1) | -0,06 (E7) | 0,063 (E16) | 0,308 (E22) | 0,963 (E11) | -5,269 (E22) | -0,57 (E22) | 0,179 (E22) | 0,999 (E3) | -0,611 (E9) | -0,193 (E22) | 0,323 (E9) |
| Ens MEAN | 1 | -0,169 | 0,001 | 0,172 | 0,949 | -8,351 | -1,338 | 0,073 | 0,998 | -0,869 | -0,379 | 0,071 |
| Ens MIN | 1,0 (E1) | -0,298 (E1) | -0,069 (E12) | 0,071 (E4) | 0,927 (E6) | -12,217 (E6) | -1,92 (E6) | -0,025 (E4) | 0,997 (E9) | -1,108 (E2) | -0,6 (E6) | -0,193 (E20) |
| Main Emulator | 1 | -0,076 | 0,033 | 0,171 | 0,949 | -10,064 | -1,675 | 0,093 | 0,998 | -0,963 | -0,449 | 0,059 |
| I-UPRCM | 0,805 | -2,884 | 0,529 | 8,016 | 0,407 | -22,153 | -2,796 | 7,793 | 0,767 | -4,22 | -0,315 | 7,443 |
| Cheap | 1 | -0,163 | 0,031 | 0,271 | 0,918 | -7,765 | -1,589 | 0,088 | 0,998 | -0,837 | -0,303 | 0,19 |
| Emul-Hist | 1 | -0,393 | -0,275 | -0,088 | 0,833 | -16,931 | -2,787 | -1,095 | 0,997 | -1,661 | -0,981 | -0,486 |

**FUTURE PERIOD 2080-2100**

| | CLIMATOLOGY SpaCor | CLIM SQ05 | CLIM Mean | CLIM SQ95 | Days SpaCor | Days SQ05 | Days Mean | Days SQ95 | 99th SpaCor | 99th SQ05 | 99th Mean | 99th SQ95 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| RCM truth values | | | 11,530 | 14,040 | | | 16,802 | 112,293 | | 17,888 | 25,700 | 30,550 |
| Ens MAX | 1,0 (E1) | -0,158 (E7) | -0,003 (E16) | | 0,984 (E3) | -30,672 (E11) | -5,748 (E11) | 0,361 (E22) | 0,999 (E26) | -0,889 (E22) | -0,345 (E22) | 0,168 (E22) |
| Ens MEAN | 1 | -0,271 | -0,078 | | 0,974 | -45,22 | -8,06 | 0,035 | 0,998 | -1,148 | -0,567 | -0,058 |
| Ens MIN | 1,0 (E1) | -0,388 (E13) | -0,139 (E12) | | 0,948 (E13) | -61,912 (E6) | -10,625 (E6) | 0,0 (E30) | 0,997 (E1) | -1,439 (E2) | -0,827 (E6) | -0,238 (E6) |
| Main Emulator | 1 | -0,163 | -0,036 | | 0,975 | -52,413 | -9,234 | 0,009 | 0,998 | -1,247 | -0,632 | -0,029 |
| I-UPRCM | 0,803 | -2,868 | 0,5 | | 0,364 | -107,142 | -15,352 | 1,658 | 0,775 | -4,421 | -0,341 | 7,294 |
| Cheap | 1 | -0,369 | -0,143 | | 0,958 | -46,35 | -9,075 | 0,014 | 0,997 | -1,447 | -0,726 | -0,109 |
| Emul-Hist | 1 | -1,458 | -1,266 | | 0,796 | -100,009 | -16,018 | 0 | 0,992 | -3,741 | -2,479 | -1,672 |

**CLIMATE CHANGE 2080-2100 vs 2006-2025**

| | CLIMATOLOGY SpaCor | CLIM SQ05 | CLIM Mean | CLIM SQ95 | Days SpaCor | Days SQ05 | Days Mean | Days SQ95 | 99th SpaCor | 99th SQ05 | 99th Mean | 99th SQ95 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| RCM truth values | | 1,336 | 1,651 | 2,019 | | | 13,447 | 90,138 | | 1,009 | 1,882 | 2,794 |
| Ens MAX | 0,964 (E26) | -0,155 (E28) | -0,057 (E9) | 0,051 (E1) | 0,976 (E3) | -28,173 (E11) | -4,893 (E11) | 0,311 (E22) | 0,836 (E26) | -0,565 (E11) | -0,055 (E11) | 0,48 (E11) |
| Ens MEAN | 0,955 | -0,187 | -0,079 | 0,012 | 0,965 | -39,669 | -6,722 | 0,097 | 0,804 | -0,733 | -0,188 | 0,312 |
| Ens MIN | 0,945 (E1) | -0,212 (E22) | -0,113 (E6) | -0,018 (E6) | 0,938 (E13) | -52,364 (E6) | -8,705 (E6) | 0,048 (E29) | 0,722 (E1) | -0,839 (E16) | -0,277 (E10) | 0,224 (E16) |
| Main Emulator | 0,956 | -0,175 | -0,069 | 0,017 | 0,966 | -44,583 | -7,559 | 0,095 | 0,761 | -0,772 | -0,183 | 0,382 |
| I-UPRCM | 0,821 | -0,26 | -0,029 | 0,148 | 0,305 | -87,397 | -12,556 | 1,536 | 0,615 | -0,808 | -0,026 | 0,639 |
| Cheap | 0,942 | -0,299 | -0,174 | -0,069 | 0,956 | -41,427 | -7,486 | 0,107 | 0,725 | -1,027 | -0,423 | 0,206 |
| Emul-Hist | 0,867 | -1,233 | -0,99 | -0,858 | 0,661 | -84,118 | -13,23 | 0,074 | 0,545 | -2,338 | -1,498 | -0,917 |

Table 4: Summary of the results obtain in the 2nd step of evaluation (GCM world). The first line (RCM truth) are the RCM absolute values. The following lines are the scores according to the metrics defined in 2.6, for the different emulators described in this study. The three first lines summarize gives the mean and the range of the ensemble of emulators described in 4.4. For the ensemble minimums and maximums the name ( E1 to E30 ) of the emulator corresponding to the value is given in bracket. "Main Emulator" is the one used for the results section (3). "Cheap" emulator is the one including only the variables at 700hpa described in Section 4.3. "Emul-Hist" is the emulator trained only on the historical period.

**PRESENT PERIOD 2006-2025 — distribution metrics**

| | RMSE | | | ACC | | | Variance ratio | | | Wasserstein distance | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | SQ05 | MEAN | SQ95 | SQ05 | MEAN | SQ95 | SQ05 | MEAN | SQ95 | SQ05 | MEAN | SQ95 |
| Ens MAX | 1,438 (Em30) | 2,272 (Em7) | 2,854 (Em22) | 0,698 (Em8) | 0,759 (Em8) | 0,809 (Em2) | 0,96 (Em8) | 1,013 (Em22) | 1,088 (Em26) | 0,294 (Em2) | 0,82 (Em7) | 1,226 (Em27) |
| Ens MEAN | 1,404 | 2,238 | 2,804 | 0,694 | 0,757 | 0,807 | 0,934 | 0,99 | 1,061 | 0,209 | 0,748 | 1,108 |
| Ens MIN | 1,377 (Em14) | 2,204 (Em12) | 2,76 (Em4) | 0,689 (Em7) | 0,754 (Em10) | 0,806 (Em1) | 0,902 (Em6) | 0,956 (Em6) | 1,027 (Em20) | 0,134 (Em12) | 0,658 (Em12) | 1,003 (Em30) |
| Main Emulator | 1,435 | 2,246 | 2,826 | 0,691 | 0,755 | 0,808 | 0,947 | 0,995 | 1,067 | 0,291 | 0,782 | 1,118 |
| GCM | 1,755 | 3,335 | 10,169 | 0,61 | 0,724 | 0,776 | 0,727 | 0,974 | 1,283 | 0,244 | 2,173 | 9,728 |
| Cheap | 0,92 | 1,105 | 1,343 | 0,871 | 0,932 | 0,96 | 0,937 | 0,977 | 1,018 | 0,075 | 0,152 | 0,275 |
| Emul-Hist | 1,541 | 2,16 | 2,752 | 0,683 | 0,749 | 0,8 | 0,886 | 0,95 | 1,017 | 0,155 | 0,278 | 0,552 |

**PRESENT PERIOD 2006-2025 — CLIMATOLOGY / Number of days over 30°C / 99th Percentile**

| | CLIMATOLOGY | | | | Number of days over 30°C | | | | 99th Percentile | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | SpaCor | SQ05 | Mean | SQ95 | SpaCor | SQ05 | Mean | SQ95 | SpaCor | SQ05 | Mean | SQ95 |
| RCM truth values | | -2,166 | 9,879 | 16,518 | | 0,000 | 3,356 | 23,500 | | 12,138 | 23,817 | 28,580 |
| Ens MAX | 0,999 (Em3) | 0,408 (Em19) | 1,002 (Em7) | 1,412 (Em27) | 0,958 (Em1) | -1,35 (Em22) | 5,101 (Em22) | 34,218 (Em22) | 0,998 (Em7) | 0,3 (Em22) | 0,96 (Em22) | 1,633 (Em5) |
| Ens MEAN | 0,998 | 0,272 | 0,927 | 1,307 | 0,949 | -2,1 | 3,078 | 23,779 | 0,997 | -0,005 | 0,732 | 1,335 |
| Ens MIN | 0,997 (Em1) | 0,128 (Em27) | 0,831 (Em12) | 1,198 (Em12) | 0,937 (Em21) | -3,103 (Em21) | 1,639 (Em6) | 13,308 (Em20) | 0,995 (Em27) | -0,354 (Em27) | 0,497 (Em20) | 1,053 (Em20) |
| Main Emulator | 0,999 | 0,428 | 0,959 | 1,291 | 0,944 | -1,609 | 3,295 | 24,059 | 0,998 | 0,258 | 0,825 | 1,427 |
| I-GCM | 0,702 | -3,155 | 1,185 | 9,989 | 0,157 | -22,525 | -1,491 | 9,947 | 0,584 | -4,631 | 0,771 | 9,383 |
| Cheap | 1 | -0,163 | 0,031 | 0,271 | 0,918 | -7,765 | -1,589 | 0,088 | 0,998 | -0,837 | -0,303 | 0,19 |
| Emul-Hist | 0,999 | 0,099 | 0,657 | 0,983 | 0,858 | -10,493 | -2,045 | 0,12 | 0,996 | -0,979 | -0,168 | 0,461 |

(*Sub-columns for these blocks correspond to the "Emul vs RCM truth" header: SpaCor, SQ 05, Mean, SQ95.*)

**FUTURE PERIOD 2080-2100 — CLIMATOLOGY / Number of days over 30°C / 99th Percentile**

| | CLIMATOLOGY | | | | Number of days over 30°C | | | | 99th Percentile | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | SpaCor | SQ05 | Mean | SQ95 | SpaCor | SQ05 | Mean | SQ95 | SpaCor | SQ05 | Mean | SQ95 |
| RCM truth values | | -0,188 | 11,530 | 17,888 | | 0,000 | 16,802 | 112,293 | | 14,040 | 25,700 | 30,550 |
| Ens MAX | 0,999 (Em22) | 0,17 (Em11) | 0,733 (Em7) | 1,181 (Em16) | 0,971 (Em8) | -13,169 (Em22) | 0,863 (Em22) | 32,274 (Em22) | 0,996 (Em4) | -0,327 (Em22) | 0,409 (Em22) | 1,249 (Em5) |
| Ens MEAN | 0,998 | 0,009 | 0,647 | 1,069 | 0,956 | -26,271 | -3,053 | 12,458 | 0,996 | -0,703 | 0,144 | 0,986 |
| Ens MIN | 0,997 (Em1) | -0,152 (Em27) | 0,56 (Em12) | 0,97 (Em30) | 0,936 (Em2) | -39,843 (Em20) | -6,688 (Em20) | 0,42 (Em20) | 0,993 (Em27) | -0,996 (Em2) | -0,14 (Em20) | 0,736 (Em20) |
| Main Emulator | 0,998 | 0,158 | 0,692 | 1,106 | 0,955 | -25,328 | -2,621 | 12,246 | 0,996 | -0,583 | 0,233 | 1,083 |
| I-GCM | 0,695 | -3,28 | 0,999 | 9,593 | 0,069 | -109,274 | -12,113 | 21,336 | 0,564 | -5,188 | 0,304 | 9,041 |
| Cheap | 1 | -0,369 | -0,143 | 0,088 | 0,958 | -46,35 | -9,075 | 0,014 | 0,997 | -1,447 | -0,726 | -0,109 |
| Emul-Hist | 0,998 | -0,96 | -0,503 | -0,104 | 0,764 | -99,231 | -15,871 | 0,025 | 0,989 | -3,263 | -1,933 | -0,774 |

**CLIMATE CHANGE 2080-2100 vs 2006-2025 — CLIMATOLOGY / Number of days over 30°C / 99th Percentile**

| | CLIMATOLOGY | | | | Number of days over 30°C | | | | 99th Percentile | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | SpaCor | SQ05 | Mean | SQ95 | SpaCor | SQ05 | Mean | SQ95 | SpaCor | SQ05 | Mean | SQ95 |
| RCM truth values | | 1,336 | 1,651 | 2,019 | | -0,077 | 13,447 | 90,138 | | 1,009 | 1,882 | 2,794 |
| Ens MAX | 0,912 (Em28) | -0,388 (Em28) | -0,255 (Em28) | -0,134 (Em11) | 0,943 (Em28) | -28,153 (Em22) | -4,239 (Em22) | 9,981 (Em22) | 0,756 (Em12) | -1,056 (Em11) | -0,509 (Em11) | 0,218 (Em4) |
| Ens MEAN | 0,888 | -0,44 | -0,28 | -0,155 | 0,917 | -37,716 | -6,131 | 3,56 | 0,666 | -1,192 | -0,588 | 0,135 |
| Ens MIN | 0,869 (Em22) | -0,474 (Em6) | -0,308 (Em6) | -0,178 (Em10) | 0,879 (Em6) | -49,897 (Em20) | -8,332 (Em20) | 0,272 (Em20) | 0,47 (Em13) | -1,383 (Em13) | -0,659 (Em30) | 0,019 (Em12) |
| Main Emulator | 0,885 | -0,411 | -0,268 | -0,139 | 0,909 | -37,944 | -5,916 | 3,576 | 0,625 | -1,257 | -0,592 | 0,153 |
| I-GCM | 0,556 | -0,539 | -0,186 | 0,073 | 0,017 | -88,647 | -10,622 | 13,536 | 0,035 | -1,355 | -0,467 | 0,48 |
| Cheap | 0,942 | -0,299 | -0,174 | -0,069 | 0,956 | -41,427 | -7,486 | 0,107 | 0,725 | -1,027 | -0,423 | 0,206 |
| Emul-Hist | 0,733 | -1,44 | -1,16 | -0,985 | -0,59 | -94,595 | -13,826 | 0,108 | 0,358 | -2,593 | -1,765 | -0,959 |

# References

Baño-Medina J, Manzanas R, Gutierrez JM, Gutiérrez JM (2020) Configuration and intercomparison of deep learning neural models for statistical downscaling. Geoscientific Model Development 13(4):2109–2124, DOI 10.5194/gmd-13-2109-2020, URL https://gmd.copernicus.org/articles/13/2109/2020/

Baño-Medina J, Manzanas R, Gutiérrez JM (2021) On the suitability of deep convolutional neural networks for continental-wide downscaling of climate change projections. Climate Dynamics (0123456789), DOI 10.1007/s00382-021-05847-0, URL https://doi.org/10.1007/s00382-021-05847-0

Bartók B, Wild M, Folini D, Lüthi D, Kotlarski S, Schär C, Vautard R, Jerez S, Imecs Z (2017) Projected changes in surface solar radiation in CMIP5 global climate models and in EURO-CORDEX regional climate models for Europe. Climate Dynamics 49(7-8):2665–2683, DOI 10.1007/s00382-016-3471-2

Berg N, Hall A, Sun F, Capps S, Walton D, Langenbrunner B, Neelin D (2015) Twenty-first-century precipitation changes over the los angeles region. Journal of Climate 28(2):401–421, DOI 10.1175/JCLI-D-14-00316.1, URL http://dx.doi.org/10.1175/JCLI-D-14-

Bhojanapalli S, Wilber K, Veit A, Rawat AS, Kim S, Menon A, Kumar S (2021) On the Reproducibility of Neural Network Predictions. arXiv preprint arXiv:210203349 pp 1–19, URL http://arxiv.org/abs/2102.03349

Boé J, Somot S, Corre L, Nabat P (2020) Large discrepancies in summer climate change over Europe as projected by global and regional climate models: causes and consequences. Climate Dynamics DOI 10.1007/s00382-020-05153-1

Colette A, Vautard R, Vrac M (2012) Regional climate downscaling with prior statistical correction of the global climate forcing. Geophysical Research Letters 39(13):1–5, DOI 10.1029/2012GL052258

Colin J, DéQué M, Radu R, Somot S (2010) Sensitivity study of heavy precipitation in Limited Area Model climate simulations: Influence of the size of the domain and the use of the spectral nudging technique. Tellus, Series A: Dynamic Meteorology and Oceanography 62(5):591–604, DOI 10.1111/j.1600-0870.2010.00467.x

Dayon G, Boé J, Martin E (2015) Transferability in the future climate of a statistical downscaling method for precipitation in France. J Geophys Res Atmos 120:1023–1043, DOI 10.1002/2014JD022236

Déqué M, Rowell DP, Lüthi D, Giorgi F, Christensen JH, Rockel B, Jacob D, Kjellström E, De Castro M, Van Den Hurk B (2007) An intercomparison of regional climate simulations for Europe: Assessing uncertainties in model projections. Climatic Change 81(SUPPL. 1):53–70, DOI 10.1007/s10584-006-9228-x

Déqué M, Somot S, Sanchez-Gomez E, Goodess CM, Jacob D, Lenderink G, Christensen OB (2012) The spread amongst ENSEMBLES regional scenarios: Regional climate models, driving general circulation models and inter-annual variability. Climate Dynamics 38(5-6):951–964, DOI 10.1007/s00382-

011-1053-x

Erlandsen HB, Parding KM, Benestad R, Mezghani A, Pontoppidan M (2020) A hybrid downscaling approach for future temperature and precipitation change. Journal of Applied Meteorology and Climatology 59(11):1–46, DOI 10.1175/jamc-d-20-0013.1, URL https://doi.org/10.1175/JAMC-D-20-

Evin G, Somot S, Hingray B (2021) Balanced estimate and uncertainty assessment of European climate change using the large EURO-CORDEX regional climate model ensemble. Earth System Dynamics Discussions (March):1–40, DOI 10.5194/esd-2021-8

Eyring V, Bony S, Meehl GA, Senior CA, Stevens B, Stouffer RJ, Taylor KE (2016) Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. Geoscientific Model Development 9(5):1937–1958, DOI 10.5194/gmd-9-1937-2016, URL https://gmd.copernicus.org/articles/9/1937/2016/

Fantini A, Raffaele F, Torma C, Bacer S, Coppola E, Giorgi F, Ahrens B, Dubois C, Sanchez E, Verdecchia M (2018) Assessment of multiple daily precipitation statistics in ERA-Interim driven Med-CORDEX and EURO-CORDEX experiments against high resolution observations. Climate Dynamics 51(3):877–900, DOI 10.1007/s00382-016-3453-4

Fernández J, Frías MD, Cabos WD, Cofiño AS, Domínguez M, Fita L, Gaertner MA, García-Díez M, Gutiérrez JM, Jiménez-Guerrero P, Liguori G, Montávez JP, Romera R, Sánchez E (2019) Consistency of climate change projections from multiple global and regional model intercomparison projects. Climate Dynamics 52(1-2):1139–1156, DOI 10.1007/s00382-018-4181-8, URL https://doi.org/10.1007/s00382-018-4181-8

Giorgi F, Torma C, Coppola E, Ban N, Schär C, Somot S (2016) Enhanced summer convective rainfall at Alpine high elevations in response to climate warming. Nature Geoscience 9(8):584–589, DOI 10.1038/ngeo2761

Guidotti R, Monreale A, Ruggieri S, Turini F, Giannotti F, Pedreschi D (2018) A survey of methods for explaining black box models. ACM Computing Surveys 51(5), DOI 10.1145/3236009

Gutiérrez JM, Maraun D, Widmann M, Huth R, Hertig E, Benestad R, Roessler O, Wibig J, Wilcke R, Kotlarski S, San Martín D, Herrera S, Bedia J, Casanueva A, Manzanas R, Iturbide M, Vrac M, Dubrovsky M, Ribalaygua J, Pórtoles J, Räty O, Räisänen J, Hingray B, Raynaud D, Casado MJ, Ramos P, Zerenner T, Turco M, Bosshard T, Štěpánek P, Bartholy J, Pongracz R, Keller DE, Fischer AM, Cardoso RM, Soares PM, Czernecki B, Pagé C (2019) An intercomparison of a large ensemble of statistical downscaling methods over Europe: Results from the VALUE perfect predictor cross-validation experiment. International Journal of Climatology 39(9):3750–3785, DOI 10.1002/joc.5462

Ioffe S, Szegedy C (2015) Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. Proceedings of the 32nd International Conference on Machine Learning, PMLR 1(37):448–456

Jacob D, Petersen J, Eggert B, Alias A, Christensen OB, Bouwer LM, Braun A, Colette A, Déqué M, Georgievski G, Georgopoulou E, Gobiet A, Menut L,

Nikulin G, Haensler A, Hempelmann N, Jones C, Keuler K, Kovats S, Kröner N, Kotlarski S, Kriegsmann A, Martin E, van Meijgaard E, Moseley C, Pfeifer S, Preuschmann S, Radermacher C, Radtke K, Rechid D, Rounsevell M, Samuelsson P, Somot S, Soussana JF, Teichmann C, Valentini R, Vautard R, Weber B, Yiou P (2014) EURO-CORDEX: New high-resolution climate change projections for European impact research. Regional Environmental Change 14(2):563–578, DOI 10.1007/s10113-013-0499-2

Klaver R, Haarsma R, Vidale PL, Hazeleger W (2020) Effective resolution in high resolution global atmospheric models for climate studies. Atmospheric Science Letters 21(4):1–8, DOI 10.1002/asl.952, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/asl.952

Kotlarski S, Lüthi D, Schär C (2015) The elevation dependency of 21st century European climate change: An RCM ensemble perspective. International Journal of Climatology 35(13):3902–3920, DOI 10.1002/joc.4254

Laprise R, de Elía R, Caya D, Biner S, Lucas-Picher P, Diaconescu E, Leduc M, Alexandru A, Separovic L (2008) Challenging some tenets of Regional Climate Modelling Network for Regional Climate Modelling and Diagnostics. Meteorol Atmos Phys 100:3–22, DOI 10.1007/s00703-008-0292-9

LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. Proceedings of the IEEE 86(11):2278–2323, DOI 10.1109/5.726791

Lecun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444, DOI 10.1038/nature14539

Lemus-Canovas M, Brands S (2020) Assessing several downscaling methods for daily minimum and maximum temperature in a mountainous area. Are we able to statistically simulate a warmer climate in the Pyrenees? In: EGU General Assembly Conference Abstracts, EGU General Assembly Conference Abstracts, p 11389

Maraun D, Widmann M (2018) Statistical Downscaling and Bias Correction for Climate Research. Cambridge University Press, DOI 10.1017/9781107588783, URL https://www.cambridge.org/core/product/identifier/9781107588783/type/book

Marrone S, Piantadosi G, Olivieri S, Sansone C (2019) Reproducibility of deep CNN for biomedical image processing across frameworks and architectures. European Signal Processing Conference 2019-Septe:0–4, DOI 10.23919/EUSIPCO.2019.8902690

Meehl GA, Covey C, Delworth T, Latif M, McAvaney B, Mitchell JF, Stouffer RJ, Taylor KE (2007) The WCRP CMIP3 multimodel dataset: A new era in climatic change research. Bulletin of the American Meteorological Society 88(9):1383–1394, DOI 10.1175/BAMS-88-9-1383

Moss RH, Edmonds JA, Hibbard KA, Manning MR, Rose SK, Van Vuuren DP, Carter TR, Emori S, Kainuma M, Kram T, Meehl GA, Mitchell JF, Nakicenovic N, Riahi K, Smith SJ, Stouffer RJ, Thomson AM, Weyant JP, Wilbanks TJ (2010) The next generation of scenarios for climate change research and assessment. Nature 463(7282):747–756, DOI 10.1038/nature08823

Nabat P, Somot S, Cassou C, Mallet M, Michou M, Bouniol D, Decharme B, Drugé T, Roehrig R, Saint-Martin D (2020) Modulation of radiative aerosols effects by atmospheric circulation over the Euro-Mediterranean region. Atmospheric Chemistry and Physics 20(14):8315–8349, DOI 10.5194/acp-20-8315-2020, URL https://doi.org/10.5194/acp-20-8315-2020

Prein AF, Gobiet A, Truhetz H, Keuler K, Goergen K, Teichmann C, Fox Maule C, van Meijgaard E, Déqué M, Nikulin G, Vautard R, Colette A, Kjellström E, Jacob D (2016) Precipitation in the EURO-CORDEX 0.11 and 0.44 simulations: high resolution, high benefits? Climate Dynamics 46(1-2):383–412, DOI 10.1007/s00382-015-2589-y

Pryor SC, Barthelmie RJ (2014) Hybrid downscaling of wind climates over the eastern USA. Environmental Research Letters 9(2), DOI 10.1088/1748-9326/9/2/024013

Reichstein M, Camps-Valls G, Stevens B, Jung M, Denzler J, Carvalhais N, Prabhat (2019) Deep learning and process understanding for data-driven Earth system science. Nature 566(7743):195–204, DOI 10.1038/s41586-019-0912-1, URL https://doi.org/10.1038/s41586-019-0912-1 http://dx.doi.org/10.1038/s41586-019-0912-1

Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol 9351, pp 234–241, DOI 10.1007/978-3-319-24574-4_28

Rumelhart DE, Hinton GE (1986) Learning Representations by Back-Propagating Errors. Cognitive Modeling (2):3–6, DOI 10.7551/mitpress/1888.003.0013

Sanchez-Gomez E, Somot S (2018) Impact of the internal variability on the cyclone tracks simulated by a regional climate model over the Med-CORDEX domain. Climate Dynamics 51(3):1005–1021, DOI 10.1007/s00382-016-3394-y

Sanchez-Gomez E, Somot S, Déqué M (2009) Ability of an ensemble of regional climate models to reproduce weather regimes over Europe-Atlantic during the period 1961-2000. Climate Dynamics 33(5):723–736, DOI 10.1007/s00382-008-0502-7

Schmith T (2008) Stationarity of regression relationships: Application to empirical downscaling. Journal of Climate 21(17):4529–4537, DOI 10.1175/2008JCLI1910.1

Sørland SL, Schär C, Lüthi D, Kjellström E (2018) Bias patterns and climate change signals in GCM-RCM model chains. Environmental Research Letters 13(7), DOI 10.1088/1748-9326/aacc77

Taylor KE, Stouffer RJ, Meehl GA (2012) An overview of CMIP5 and the experiment design. Bulletin of the American Meteorological Society 93(4):485–498, DOI 10.1175/BAMS-D-11-00094.1

Torma C, Giorgi F, Coppola E (2015) Added value of regional climate modeling over areas characterized by complex terrain-precipitation over the Alps. Journal of Geophysical Research 120(9):3957–3972, DOI 10.1002/2014JD022781, URL http://doi.wiley.com/10.1002/2014JD022781

Turco M, Quintana-Seguí P, Llasat MC, Herrera S, Gutiérrez JM (2011) Testing MOS precipitation downscaling for ENSEMBLES regional climate models over Spain. Journal of Geophysical Research Atmospheres 116(18):1–14, DOI 10.1029/2011JD016166

Vandal T, Kodra E, Ganguly AR (2019) Intercomparison of machine learning methods for statistical downscaling: the case of daily and extreme precipitation. Theoretical and Applied Climatology 137(1-2):557–570, DOI 10.1007/s00704-018-2613-3, URL https://arxiv.org/abs/1702.04018

Villani C (2009) Optimal Transport, vol 338. DOI 10.1007/978-3-540-71050-9

Voldoire A, Sanchez-Gomez E, Salas y Mélia D, Decharme B, Cassou C, Sénési S, Valcke S, Beau I, Alias A, Chevallier M, Déqué M, Deshayes J, Douville H, Fernandez E, Madec G, Maisonnave E, Moine MP, Planton S, Saint-Martin D, Szopa S, Tyteca S, Alkama R, Belamari S, Braun A, Coquart L, Chauvin F (2013) The CNRM-CM5.1 global climate model: Description and basic evaluation. Climate Dynamics 40(9-10):2091–2121, DOI 10.1007/s00382-011-1259-y

Vrac M, Drobinski P, Merlo A, Herrmann M, Lavaysse C, Li L, Somot S (2012) Dynamical and statistical downscaling of the French Mediterranean climate: Uncertainty assessment. Natural Hazards and Earth System Science 12(9):2769–2784, DOI 10.5194/nhess-12-2769-2012

Walton DB, Sun F, Hall A, Capps S (2015) A hybrid dynamical-statistical downscaling technique. Part I: Development and validation of the technique. Journal of Climate 28(12):4597–4617, DOI 10.1175/JCLI-D-14-00196.1

Wilby RL, Hassan H, Hanaki K (1998) Statistical downscaling of hydrometeorological variables using general circulation model output. Journal of Hydrology 205(1-2):1–19, DOI 10.1016/S0022-1694(97)00130-3