WORKING PAPERS

N° 1497

January 2024

“On the suitability of a Convolutional Neural Network based RCM-Emulator for fine spatio-temporal precipitation”

Antoine Doury, Samuel Somot and Sébastien Gadat

Toulouse School of Economics

Noname manuscript No.
(will be inserted by the editor)

On the suitability of a Convolutional Neural Network based RCM-Emulator for fine spatio-temporal precipitation.

Antoine Doury[(1)] · Samuel Somot[(1)] · Sebastien Gadat[(2)]

Received: date / Accepted: date

Abstract High resolution climate models are necessary to capture local precipitation but are too expensive to explore the uncertainties associated with future projections. To solve this resolution-uncertainty dilemma, Doury et al (2022) proposed a neural network based RCM-emulator for the near-surface temperature, at a daily and 12km-resolution. It uses existing RCM simulations to learn the relationship between low-resolution predictors and high resolution surface variables. When trained the emulator can be applied to any GCM simulation to produce ensembles of high resolution emulated simulations. This study assess the suitability of applying the RCM-emulator for precipitation thanks to a novel asymmetric loss function targeting to reproduce the entire precipitation distribution over any grid point.

In perfect model evaluation, the resulting emulator shows striking ability to reproduce the RCM original series with an excellent spatio-temporal correlation. In particular, a very good behaviour is obtained for the two tails of the distribution, measured by the number of dry days and the 99th quantile. Moreover, it creates consistent precipitation objects with a slight lack of precision. The emulator quality holds for all simulations of the same RCM, with any driving GCM, ensuring transferability of the tool to GCMs never downscaled by the RCM.

A first showcase of downscaling GCM simulations showed that the RCM-emulator brings significant added-value with respect to the GCM as it produces adequate high resolution spatial structure and extremes' intensity. Nevertheless, further work is needed to understand the differences that occur with the RCM and establish a relevant evaluation framework for GCM applications.

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
E-mail: antoine.doury@meteo.fr

Keywords Emulator · Hybrid downscaling · Regional Climate Modeling · Statistical Downscaling · Precipitations · Deep Neural Network · Machine Learning · EURO-CORDEX · CORDEX

Declarations

Funding This work is part of IMPETUS4CHANGE, funded by the European Union's 503Horizon Europe research and innovation programme under grant agreement No 101081555. This study has received funding from Agence Nationale de la Recherche - France 2030 as part of the PEPR TRACCS programme under grand number ANR-22-EXTR-00XX

Conflicts of interest/Competing interests Nothing to mention

Availability of data and material (data transparency) Input data are available on the Earth System Grid Federation (ESGF).

Code availability (software application or custom code) The python code used to build and train the emulator and to pre-process the input data for the emulator is publicly available at https://github.com/antoinedoury/RCM-Emulator.

Contents

1 Introduction

Precipitation is the primary source of accessible freshwater on Earth. It plays a pivotal role in maintaining Earth's system equilibrium, supporting ecosystems, and crucially, sustaining human survival and activities (Masson-Delmotte et al, 2021). However, it also harbors the potential for catastrophic events. Intense rainfall can lead to devastating floods and adversely impact agricultural yields. Severe droughts inflict significant damage on ecosystems, agriculture, and access to potable water. Given the contemporary backdrop of global climate change, it is crucial to study potential changes in precipitation patterns and extremes.

The study of precipitation is inherently complex. It is a non-continuous variable, neither in temporal nor spatial terms. Precipitation occurrences are characterized by their frequency and intensity. Investigating precipitation series across diverse temporal and spatial scales is imperative for a comprehensive grasp of their inherent nature. While rainfall or snowfall may be influenced by extensive atmospheric circulations, they can also manifest as highly localized events due to small-scale physical processes (e.g., convective instability, cold pool.. Ducrocq et al (2008)), influenced by local topography or surface heterogeneity, among other factors. Fine spatial and temporal resolution is, therefore, imperative when modeling precipitation and studying its local changes in the context of global climate change.

Undeniably, Regional Climate Models (RCMs) stand out as one of the most widely employed modeling tools today, to fulfill the imperative for precise spatial and temporal resolution in projecting the future dynamics of precipitation. RCMs are a specific kind of climate models used to downscale at high-resolution and over a limited domain the low resolution simulations produced with Global Climate Models. Their high computational costs render unfeasible the production of large ensembles of high resolution simulations necessary to address the different sources of uncertainty associated with the local impacts of climate change (Hawkins and Sutton, 2009; Evin et al, 2019). To try to address this high-resolution versus large-ensemble dilemma, recent papers (Walton et al, 2015; Berg et al, 2015; Maraun and Widmann, 2018; Doury et al, 2022) introduced the concept of emulator for Regional Climate Model (RCM) as a solution to create large ensembles of high resolution climate projections blending the RCM approach with modern machine-learning techniques.

In this study, we propose testing whether the RCM-emulator introduced in Doury et al (2022) for near-surface temperature, is suitable for emulating daily precipitation for a RCM at its full resolution (12km) over Europe. The concept of the RCM-emulator involves using machine learning tools to learn the relationship between low-resolution altitude variables describing the atmospheric circulation on a specific day and a high-resolution local surface variable, such as daily precipitation. This downscaling function is learnt inside existing RCM simulations. The aim is to tackle the cost limitation of RCM by mimicking its downscaling function for a specific variable at a low computational cost and then by applying it to any global and low resolution simulation. RCM-emulators are categorized as hybrid downscaling methods because they incorporate both statistical and dynamical downscaling. Utilising historical and future RCM simulations in the training set enables the RCM-emulator to learn how this relationship may evolve under changing climate conditions. Moreover, it can also be applied over regions with no long series of good quality precipitation records.

Numerous studies have proposed statistical downscaling methods to estimate the relationship between large-scale and local-scale variables in observational records. Maraun et al (2010) or Gutiérrez et al (2019) provide an overview of available approaches for precipitations. Some very recent studies (Baño-Medina et al, 2020, 2021; Vandal et al, 2019; Wang et al, 2021) have successfully implemented convolutional neural networks for this purpose. The RCM-emulator employed in Doury et al (2022) and here is based on a fully convolutional neural network architecture called UNet (Ronneberger et al, 2015). It has exhibited an excellent ability to emulate the temperature, notably in reproducing the complex spatial structure and daily variability brought by the RCM. However, since precipitation is more challenging to model than temperature, this study proposes to explore the use of the loss function to help the neural network focusing on a specific task. Here the challenge will be the reproduce the entire distribution of precipitations. To address this, we devised a novel asymmetric loss function tailored for daily precipitation, which we will compare to two classical choices for regression problems.

After assessing the suitability of the RCM-Emulator for precipitation, we propose in this study to profit from the EURO-CORDEX simulations to evaluate the transferability of the tool. Indeed the emulator is trained using a given set of availabe RCM simulations (driven by a given GCM and RCP scenario) and it is crucial to study its behavior when downscaling other socio-economic scenarios or GCMs. Then, in a first step, we evaluate the emulator in a perfect model framework regarding all available simulations with the emulated RCM. Then in a final step, we propose a first showcase of application by downscaling GCM simulations.

This paper is organised into four main sections. In Section 2, we recall the concept of the RCM-emulator introduced in Doury et al (2022), define the technical aspects related to the neural network and the loss functions, and present the framework of the study, including the data, the target domain, and the associated predictors. Section 3 presents the detailed evaluation and comparison of the emulators within a perfect model framework, while Section 4 shows the results of applying the asymmetric emulator to GCM simulations. The concluding section summarizes the paper and initiates the discussion.

2 Methodology

In this section, we define the framework used to build and evaluate the RCM emulator for precipitation. Firstly, we recall the emulator concept and present the simulations and the chosen target domain and predictors for this study. We present the neural network architecture and the three loss functions used to train the three emulators for the inter-comparison. The perfect model framework approach used to train and evaluate the emulator is also recalled. Finally, we detail the metrics used to evaluate the emulator under different aspects.

2.1 RCM-Emulator concept and calibration process

Regional climate models (RCMs) are driven by global climate models (GCMs) as they continuously receive incoming data at their domain's borders from a specific GCM simulation at regular intervals. The resulting RCM simulation essentially represents a downscaling of the data from the driving GCM. Nevertheless, within the boundaries of its domain, the RCM develops its own narrative and may consequently deviate from the driving GCM. This can lead to significant differences, both on a daily scale and on a climatological scale, as discussed by Laprise et al (2008). This large scale transformation primarily arises from the chaotic nature of weather (Lucas-Picher et al, 2008), but it is also influenced by differences in how the models represent physical processes or their inherent complexity, as explored by Boé et al (2020) and Taranu et al (2022). Thanks to a lower computational cost, GCMs include generally more components than RCM such as ocean coupling or evolving aerosols. Consequently, Doury et al (2022) decided to develop an RCM emulator specifically to learn the downscaling process inside the RCM simulation while excluding the impact of large-scale transformations.

To isolate the downscaling function, the emulator is trained within a "perfect model" framework, where both the inputs and target data are sourced from the same RCM simulation. The methodology is detailed in Figure 1. The chosen predictors (described in Section 2.3) are upscaled to match the resolution of the GCM, typically around 150km, through a conservative interpolation method, which involves a straightforward average of all points encompassed within the low-resolution grid. A spatial moving average filter is then applied to eliminate any high-resolution features that might persist

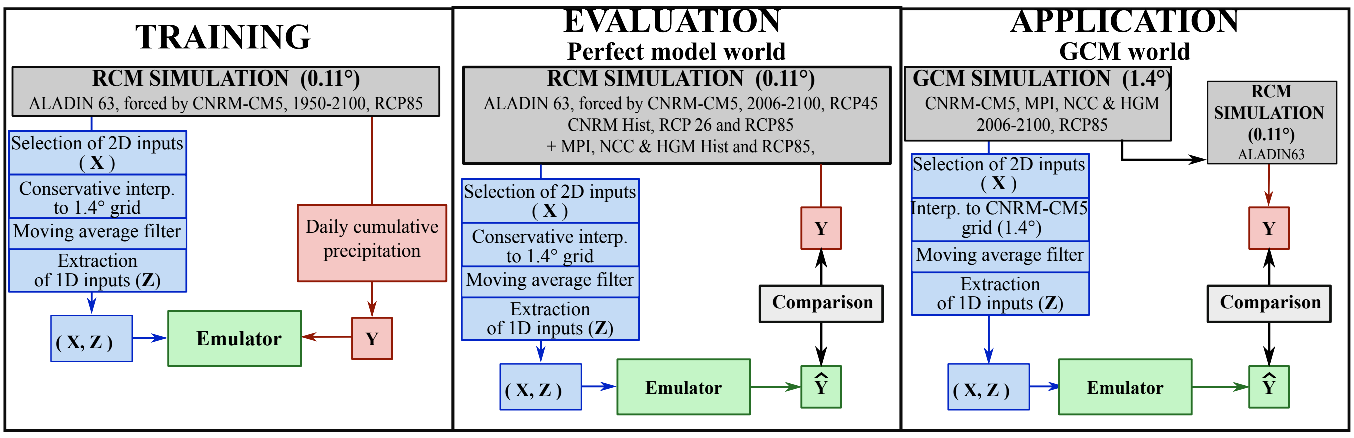

Fig. 1. Scheme of the training (left), perfect model evaluation (middle) and GCM world application (right) protocols. Redrawn from (Doury et al, 2022).

through the interpolation. Subsequently, the emulator is trained to accurately replicate the relationship between these "upscaled" inputs and the target variable, such as precipitation, at the resolution of the RCM.

This perfect model framework also facilitates a rigorous evaluation of the emulator, with the RCM series serving as an ideal reference that it should be capable of faithfully reproducing. In practical application, the emulator is directly applied to a GCM simulation, and the smoothing step is retained to consider the GCM at its effective resolution, as discussed by Klaver et al (2020).

2.2 Data: the RCM matrix

The emulator proposed in this study relies on the regional climate model ALADIN63 (Nabat et al, 2020). A total of ten simulations have been published with this RCM over the whole Europe in the EURO-CORDEX framework (Coppola et al, 2021). They downscale four different GCMs and three different scenarios of greenhouse gas emissions (cf Table 1). The CNRM-CM5 global climate model is developed in the same institute as ALADIN63, so they belong to the same family of models. CNRM-CM5 drove 4 ALADIN63 simulations, the historical (1951-2005) and three RCP scenarios (2.6, 4.5 and 8.5, on the period 2006-2100). MPI-ESM-LR, NorESM1-M and HadGEM2-ES are the three other GCMs used to drive ALADIN63 following the historical and RCP8.5 scenarios of greenhouses gases emissions. From now, CNRM-CM5 will be referred to as CNRM, MPI-ESM-LR as MPI, NorESM1-M as NCC and HadGEM2-ES as HGM.

Table 1: RCM x GCM x Scenario matrix

Driving Scenarios	Driving GCMs CNRM-CM5 (CNRM)	MPI-ESM-LR (MPI)	NorESM1-M (NCC)	HadGEM-ES2 (HGM)
Historical	x	x	x	x
RCP26	x			
RCP45	x			
RCP85	x	x	x	x

2.3 Predictands, predictors and neural network architecture.

This study focuses on the challenging task of emulating of daily precipitation from ALADIN63 at 0.11° horizontal resolution (about 12km). We selected a sub-domain of the EURO-CORDEX domain centred over the Alps, consisting of 128×128 grid points. The target domain is visible on the left side of Figure 2. It includes the entire Alps and goes from Sardinia until the north of France and from the Pyrenees until Croatia. This domain is of particular interest due to its diverse areas with distinct precipitation regimes. For example, the Cevennes (South-East of France) region is known for its very extreme events in autumn, similarly to other coastal areas of the Mediterranean region. The reliefs receive more precipitation than plane regions. They are known to be spots of RCM added value, especially regarding extremes (Torma et al, 2015). The flat regions of the north of the domain receive a lot of precipitation throughout the year but have less strong daily extremes than the southern regions. The Alps have also a specific precipitation regime with intense summer storms.The emulator is trained to replicate both land and ocean precipitations, although at times, we will concentrate our evaluation solely on land. Additionally, this domain is four time larger than the one in Doury et al (2022).

The emulator used in this paper for precipitation downscaling follows the principles developed in Doury et al (2022). It can be viewed as a conventional machine learning problem

$$Y_t = F(X_t, Z_t)$$

where (X_t, Z_t) are the low resolution predictors, Y_t the high resolution target variable (in this case daily amount of precipitation) at day t and F the downscaling function we aim to estimate using a neural network. The list of predictors and the standardization procedure remain consistent, encompassing both sets of 1D and 2D inputs, as detailed in Table 2. As we considered the daily precipitation we also provide daily inputs. For each day, we perform spatial normalization on each 2D input. The daily spatial mean and standard deviation are subsequently provided to the emulator through the set of 1D inputs, which also includes external forcings (yearly greenhouse gas concentrations, solar and ozone forcings) and the seasonal indicator (sinus-cosinus vector). More details can be found in Doury et al (2022). The input domain is adjusted to align with the new target domain. It is a 22*16 grid points on the CNRM-CM5 grid (1.4°) centred over the target domain, (the whole map on Figure 2, left).

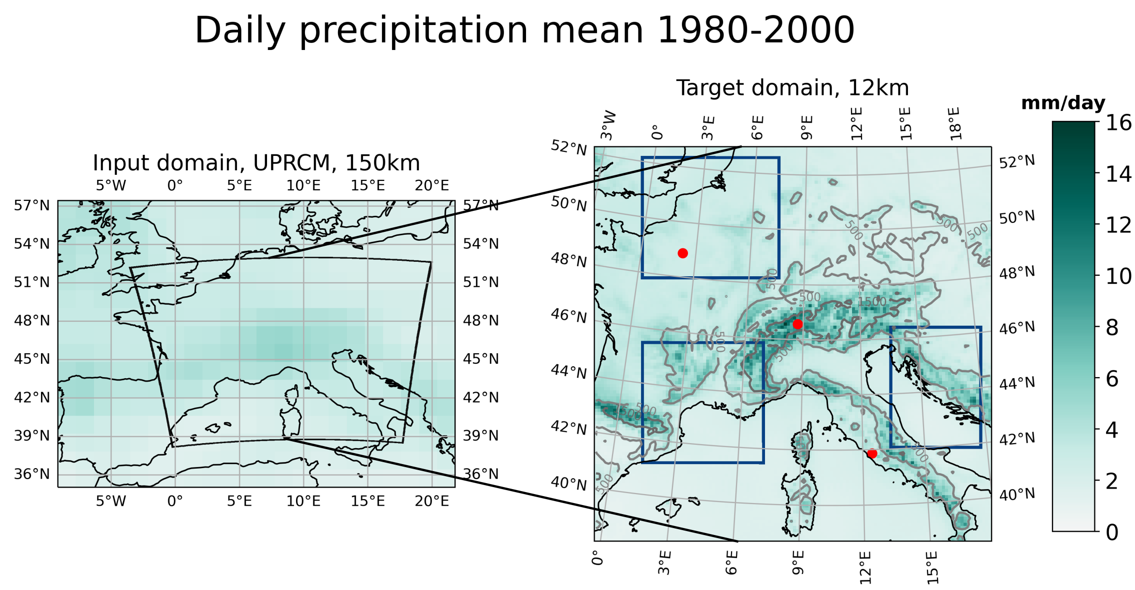

Fig. 2. Illustration of the input (left) and target (right) domains through the climatology of the daily rainfall over the 1980-2000 period. The black line on the left panel shows the target domain while the input domain is the entire map. On the target domain: the red points are the three illustrating points on Figure 6 and 10. From North to South, there is Paris, a high point (2247 meters) in the Swiss Alps and Roma. The three blue boxes are the three regions used for the SAL evaluation in section 3.2.1: The north region, centred over Belgium, the Cevennes region (south-east France) and the Dinaric Alps.

The neural network architecture is adapted from the UNet architecture (Ronneberger et al, 2015). The small differences with the one presented in Doury et al (2022) are due to the size of the input and target domains. As shown in Figure 3, the first layer of the network reshapes the 2D inputs from $[16, 22, 32]$ to $[16, 16, 64]$ in order to obtain squared images before the encoding path. On the other side, the expanding path is extended to reach the target domain size. This leads to a network of about 28 millions of parameters. The emulator presented in this paper is trained over the 150 years of the ALADIN63 simulations driven by the CNRM-CM5 historical and RCP85 runs. It takes about two hours and an half and 60 epochs to train the network on a GPU (Tesla V100 PCIe 16GB) using the keras environment (Chollet and others, 2015).

2D Variables		
Altitude (850, 700, 500 hPa)	Geopotential height, Humidity, Temperature, Northern & Eastern wind components	Daily
550hPa	Total Aerosols optical depth (ood550)	Monthly mean
Near-Surface	Sea level pressure, Northern & Eastern wind components	Daily
1D Variables		
Mean & Standard deviation for each 2D variables		Daily
Total anthropogenic greenhouses gases		Yearly
Solar and Ozone forcings		Yearly
Cosinus and Sinus seasonal indicators		Daily

Table 2: List of predictors, identical to Doury et al (2022)

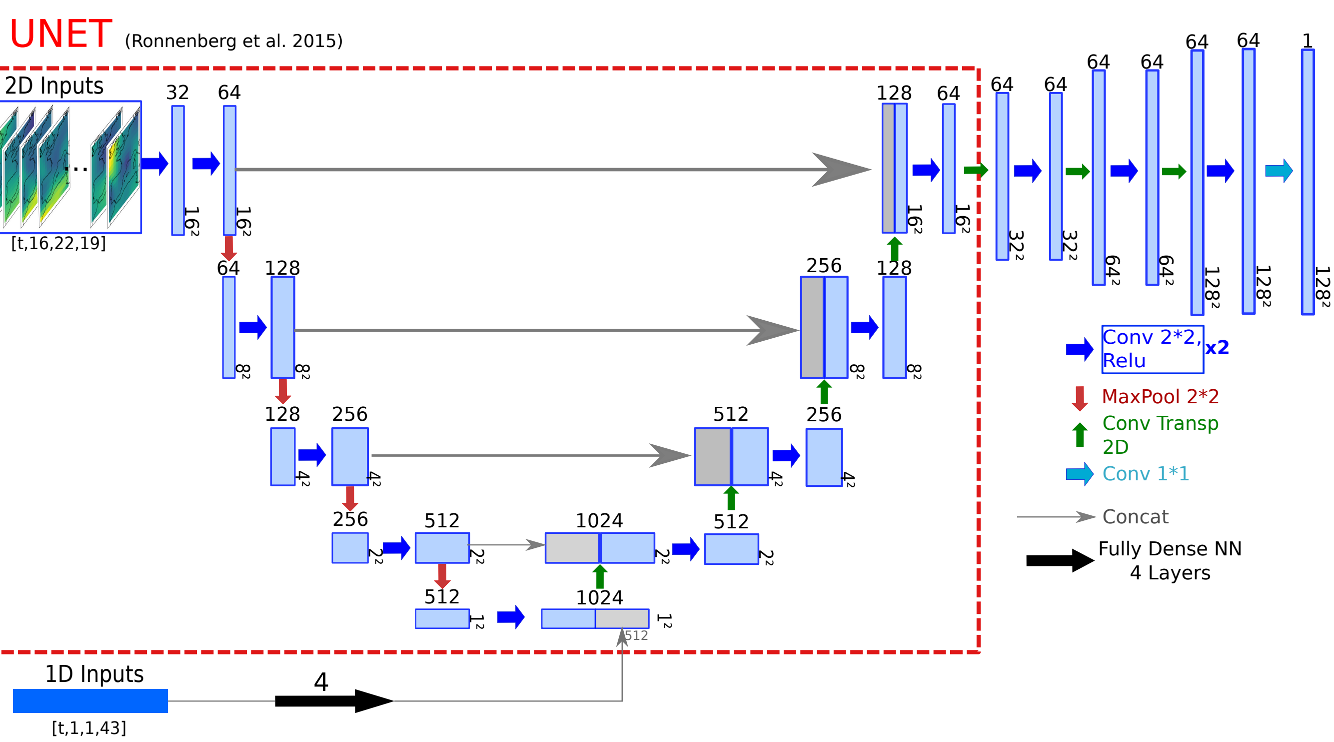

Fig. 3. Illustration of the neural network architecture, adapted from Doury et al (2022).

2.4 Loss function for the neural network training

Over this study, we propose a deeper look on the impact of the loss function on the emulator's performance. The loss function is an essential part of the neural network training. In the training phase, the network sees examples of inputs and target pairs. For each day of the training set, it makes a prediction and compares it with the truth. The loss function evaluates the network prediction against the expected outcome. The network parameters are then updated according to the loss function results. This operation is repeated until the cost (i.e. the loss mean on the training set) stabilises. The best combination of parameters has the lowest cost over a validation set, different from the training set. This is then a minimisation problem to find the best estimate $\hat{F}$ such that :

$$\hat{F} = \arg\min_{\theta \in \Theta} L(\mathcal{V}, \theta) \tag{1}$$

Where Θ is the ensemble of possible parameters, $\mathcal{V}$ the validation set and L the loss function.

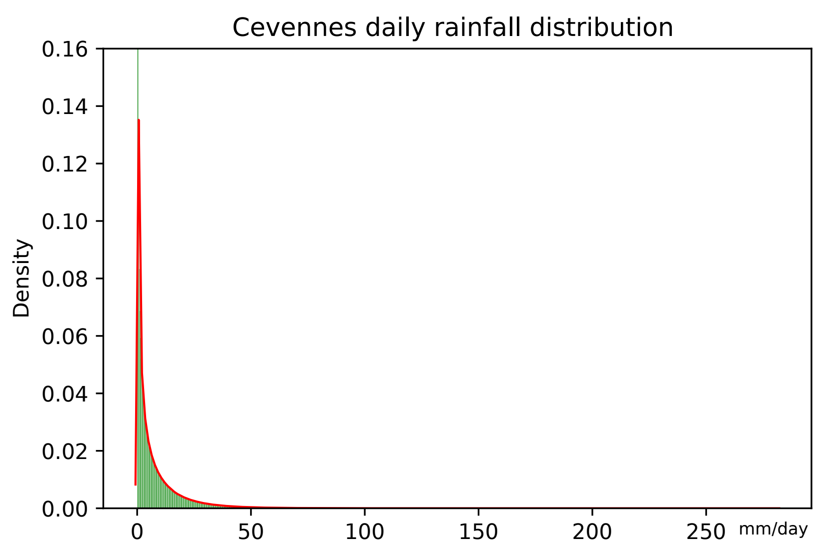

Fig. 4. Illustration of daily precipitation distribution (in mm/day), in the Cevennes box (cf Fig 2) all points and days are pooled.

Precipitations are particularly complicated to emulate with neural networks because of their distribution. Indeed, as illustrated in Figure 4, the distribution of precipitation looks like a highly left-skewed gamma distribution. There are many days with no precipitation and few ones with very high precipitation, which induces heavy tail distributions. These different events contribute non equally to the mean, with a few days having more impact than the other ones. It is of fundamental interest that the emulator reproduces well the entire distribution. The good reproduction of the frequency and intensity of rare extreme events constitutes a substantial added value of RCM, so the emulator should reproduce them accurately. The loss function is therefore a possible way to rebalance the data and to force the emulator to look more specifically into some specific part of the distribution (Ayzel et al, 2020).

We compare here three emulators, constructed with different loss functions:

– **Emul-MSE** uses the classical mean squared error for the loss function, as stated in Doury et al (2022). It corresponds to the L2 distance.

$$L(y, \hat{y}) = \frac{1}{N \times T} \sum_{t=0}^{T} \sum_{i \in \mathcal{D}} (y_{i,t} - \hat{y}_{i,t})^2 \quad (2)$$

With $\mathcal{D}$ the ensemble of grid points, N the number of grid points and T the number of days.

– **Emul-MAE** uses the mean absolute error. It corresponds to the L1 distance.

$$L(y, \hat{y}) = \frac{1}{N \times T} \sum_{t=0}^{T} \sum_{i \in \mathcal{D}} |y_{i,t} - \hat{y}_{i,t}| \quad (3)$$

– **Emul-ASYM** uses a specific loss function designed for the precipitation problem. It is based on the MAE loss function plus an asymmetric term which penalizes the emulator when it underestimates the true value while it was a raining day. The stronger the rain the stronger the penalty.

$$L(y, \hat{y}) = \frac{1}{N \times T} \sum_{t=0}^{T} \sum_{i \in \mathcal{D}} |y_{i,t} - \hat{y}_{i,t}| + \gamma_{i,t}^2 \times max(0, y_{i,t} - \hat{y}_{i,t}) \quad (4)$$

With $\gamma_{i,t} = G_i(y_{i,t})$ and G_i the cumulative distribution function of a random variable Y_i following a gamma distribution

$$Y_i \sim \mathit{\Gamma}_i : \mathit{\Gamma}(\alpha_i, \beta_i)$$

where the α_i and β_i parameters are fitted on the historical precipitation series at each grid point i.

The MAE and MSE losses are the most commonly used loss functions for regression problems. The MAE loss sums the absolute distance between an observation and its prediction. It gives the same weight to each observation. Knowing that daily rainfalls are strongly left-skewed, with a vast number of observations with a small amount of precipitation, the EMUL-MAE should be able to fit these days well. However, the rare cases with large precipitations could be less well reproduced.

The MSE loss function gives more weight to the significant errors than the small ones. The MSE generally shows the best results in regression problems and is equivalent to the maximum likelihood estimation in a Gaussian setting. It leads theoretically to the best estimate for normally distributed data knowing the inputs. In the case of precipitations, it is not likely to be the case because of their highly intermittent nature. So the MSE loss function might not be well suited. Note that Emul-MSE is the same emulator as

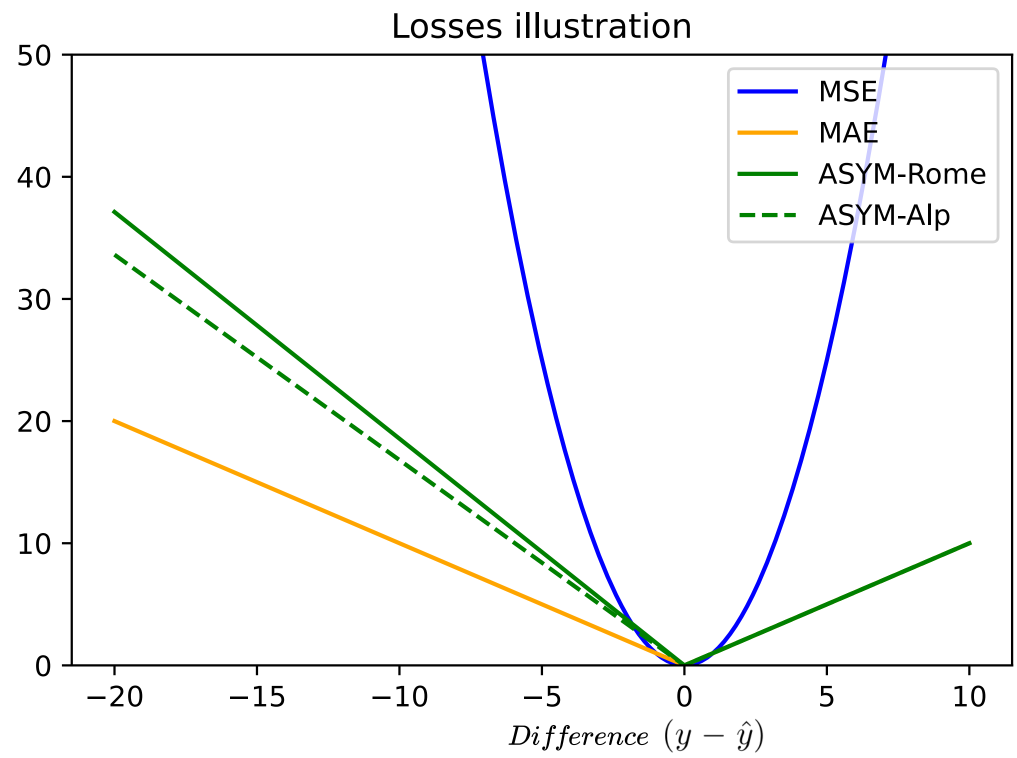

Fig. 5. Illustration of the three loss functions according to the error $(y - \hat{y})$. For the ASYM loss, as it depends on the true prediction and the location, we illustrate it with $y = 20mm/day$ and 2 locations: Roma and the Alps point already mentioned (Fig 2).

the one introduced in Doury et al (2022).

The choice of the asymmetric loss function comes from the results of both EMUL-MAE and EMUL-MSE presented in section 3. The idea is to add a penalty when the emulator underestimates strong precipitations. This is done by the asymmetric term: $max(0, y_{i,t} - \hat{y}_{i,t})$. Moreover it needs to depend on the rain intensity. The more extreme the precipitation, the rarest it is and so the higher the penalty should be. The $\gamma_{i,t}$ parameter determines how extreme is a given observation and defines the weight accordingly. At each grid point, we estimated the parameters of a gamma distribution on the rainy days (over 1mm) of the training set (using the scipy python package, Virtanen et al (2020)). The Gamma distribution has been widely used to described precipitation data (Katz, 1977; Vrac and Naveau, 2007) but other distribution could be considered. In order to make this parameter estimation more robust, we fit them yearly and then average these parameters over the years. It gives a map of the shape and scale parameters. The $\gamma_{i,t}$ parameter is then the evaluation of $y_{i,t}$ (the target value at point i and time t) by the Cumulative Distribution Function (CDF) associated to the gamma distribution Γ_i fitted for this point. It is an objective way to indicate the relative intensity of the precipitation for a given location.

2.5 Evaluation Metrics

In order to evaluate and compare the performances of the emulators we will evaluate their predictions with respect to the daily precipitation series from the corresponding RCM simulation (cf Fig.1). The evaluation relies on various metrics to compare the targeted (Y) and the predicted ($\hat{Y}$) series to have the most complete evaluation possible and understand the strengths and weaknesses of each emulators. The different metrics are detailed below.

2.5.1 Time series comparison

First of all we will evaluate in each grid point if the emulated time series matches the original RCM series through two metrics:

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

$$ACC(Y, \widehat{Y}) = \rho(Y_a, \widehat{Y_a}), \tag{5}$$

  with ρ 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:

$$RoV(Y, \widehat{Y}) = \frac{Var(\widehat{Y})}{Var(Y)} * 100 \tag{6}$$

Both metrics are computed at each grid point. Each map is summarised with its spatial mean and 5th and 95th super-quantiles. The super-quantile α is defined as the mean of all the values larger (resp. smaller) than the quantile of order α, when α is larger (resp. smaller) than 0.5.

2.5.2 Climatological scale metrics

It is necessary to evaluate the emulators at the climatological scale. We use three statistics over at least 20 years: the daily precipitation mean, the 99th quantile and the percentage of dry days (precipitations lower than 1 mm/day). These three metrics, often used in the climate community, are snapshots of the variable distribution from the mean and extreme sides. The biases maps are presented in percentage. When the biases are too strong, notably because of comparing very small values, we use the simple bias ($\hat{Y} - Y$), expressed in mm/days. Again, the statistics are computed point-wise, and each map is summarised by its spatial mean and super-quantiles.

$$\text{Relative bias } = \frac{(\hat{Y} - Y)}{Y} \times 100 \tag{7}$$

These three statistics will be looked at in present climate but also in climate change context. Each statistic will be computed in a future period and the climate change statistic is the relative difference with the past period. Then the simple bias is computed between RCM and emulator climate change statistics.

2.5.3 PDF normalisation

Since the pdf for the rainfall are very heavy-tailed, it is difficult to compare them. We propose here to have a deeper look into the distributions thanks to the ASoP method introduced in Klingaman et al (2017) and used in multiple studies as Berthou et al (2020) or Vergara-Temprado et al (2020). It consists in computing the precipitation frequency following some well-chosen bins b_n defined in Eq 8. The bins are such that they contain a similar number of events for bins over 1mm and as long as the number of events is sufficient.

$$b_n = e\left(log(0.005) + \left[n\frac{(log(120) - log(0.005))^2}{59}\right]^{\frac{1}{2}}\right) \text{ with } n \in [\![0, 100]\!] \quad (8)$$

Then we can look at each bin's contribution C_n to the mean by multiplying each frequency by the corresponding bin's mean as described in Eq. 9. Both frequency and contribution are interesting in comparing the emulated series with the true RCM.

$$\begin{gathered} C_n = f_n m_n \\ \text{where } f_n \text{ and } m_n \text{ are the frequency and the mean of bin } b_n \end{gathered} \quad (9)$$

We use the skill score proposed in Berthou et al (2020) to evaluate the difference between the emulators and the RCM truth contributions curves. The fractional contributions are the actual contributions divided by the total mean precipitation of the series. They give information on the shape of the distribution independently from the mean. The Fractional Contribution Skill Score (FCSS) sums the absolute difference in each bin between the fractional contributions of an emulator and the targeted true series. The area under the FC curve is equal to 1, so the FCSS is equal to 0 when the two distributions are identical and to 2 when there is no overlap between them. It measures the differences between the two distribution shapes independently from the series mean. This score is illustrated on Figure 10 and further commented in the results section 3.1.3.

$$FCSS(Emul, RCM) = \sum_{n \in [\![0,100]\!]} |FC_n^{Emul} - FC_n^{RCM}|$$
$$\text{where } FC_n = \frac{C_n}{\sum_n C_n} = \frac{C_n}{mean} \tag{10}$$

2.5.4 SAL score

In order to further evaluate the performances of the emulator, we use an object-oriented score introduced in Wernli et al (2008). The SAL score aims to evaluate the spatial structure of precipitation objects from a predicted map versus a reference. It compares two maps of precipitation at a given time step. It accounts for the objects' structure (S-component), location (L-component) and the total amplitude of precipitation (A-component). In perfect model evaluation, the emulator should be able to reproduce the precipitation events accurately. This score indicates if the emulator recreates objects with the same characteristics than the RCM. Note that the days are dealt independently meaning that the life time of the objects is not considered.

The first step is to identify the precipitation objects. To do so, we used the *pysteps* (Pulkkinen et al, 2019) python library, which integrates a SAL implementation. On each daily map, the objects are define as the groups of at least 5 consecutive points with precipitation higher than a threshold equal to $R^* = \frac{1}{15}R^{(95)}$, $R^{(95)}$ being the 95^{th} quantile on the map. Multiple objects can be detected every day. Then, the three components are computed aiming to differentiate objectively different precipitation objects. The A- and S-components take values between -2 and 2 while the L-component takes values between 0 and 2. If all objects are similar on the maps the three components will be close to 0. A more detailed presentation of the score behavior can be find in Wernli et al (2008, 2009).

The results are then presented in a diagram where each day is represented by a point with the S and A components on the x and y axis respectively, and the L component given by the color of the point. SAL diagram are visible in Figure 12 and commented in Section 3.2.1. Following the recommendation of Wernli et al (2009), we apply this score on sub-domains of a maximum of 500km by side represented with blue squares on Figure 2.

3 Perfect model Evaluation

This section is divided in two parts. In a first evaluation step we evaluate and compare the three emulators in perfect model framework. We use the CNRM-ALADIN RCP45 simulation, from 2006 to 2100, which has not been seen during the training of the neural network (see Figure 1). After a first impression on the emulators' abilities through some examples, we extend the analysis with climatological and daily scores. This section also aims to understand the impact of the loss function on the trained emulator. A second step focuses the evaluation on the Emul-ASYM and comment the SAL results helping to objectively determine if the emulator is able to create precipitation objects. Finally the analysis is extended to all available ALADIN63 simulations (cf. Table 1) and study the emulator ability to reproduce their climate change projections.

3.1 Comparison of the three emulators

3.1.1 First look into the emulators' prediction

Before evaluating the emulators' performances with metrics, it seemed worthwhile to look into the raw series they produce. Figure 6 shows the times series at four grid points for the year 2022 in the evaluation simulation for the RCM truth and the three emulators. The three grid points show very different series. The Alps point series shows the strongest variability and intensities, with many days over 50 mm and almost no dry spell. The Paris series has minimal variability with numerous small precipitation days and low extremes compared with the other points. The Roma series shows dry spells during spring and summer 2022 in this simulation and has a very strong rainfall event in fall.

The emulators series are very encouraging. They reproduce the original series accurately, respecting each point's characteristics. They look like precipitation series as they appear to be able to produce periods with no precipitation and days with heavy rainfall. All emulators capture the extreme autumn rainfall in Roma and the dry spell between May and June. The very high variability over the Alpine point also appears to be well reproduced by the three emulators. On all points, the three emulators seem to miss some extremes simulated by the RCM, as it occurs several times that the red line comes higher than the others. However, it does not seem that Emul-MSE or Emul-MAE ever make stronger extremes than the RCM. At this point, it is impossible to decide if an emulator performs better than the others.

Figure 7 shows the precipitation field over the target for three days randomly picked along the simulation. It shows the RCM truth, the three emulators and the UPscaled precipitation field (UPRCM). The UPRCM helps to have an insight into the input resolution and shows how the RCM and the emulators refine it, even if precipitation is not part of the predictors. Several exciting points appear in this figure. First of all, the emulators' prediction on each panel is very coherent with the RCM. The precipitations are always well located with coherent intensity. It seems, however, that the emulators

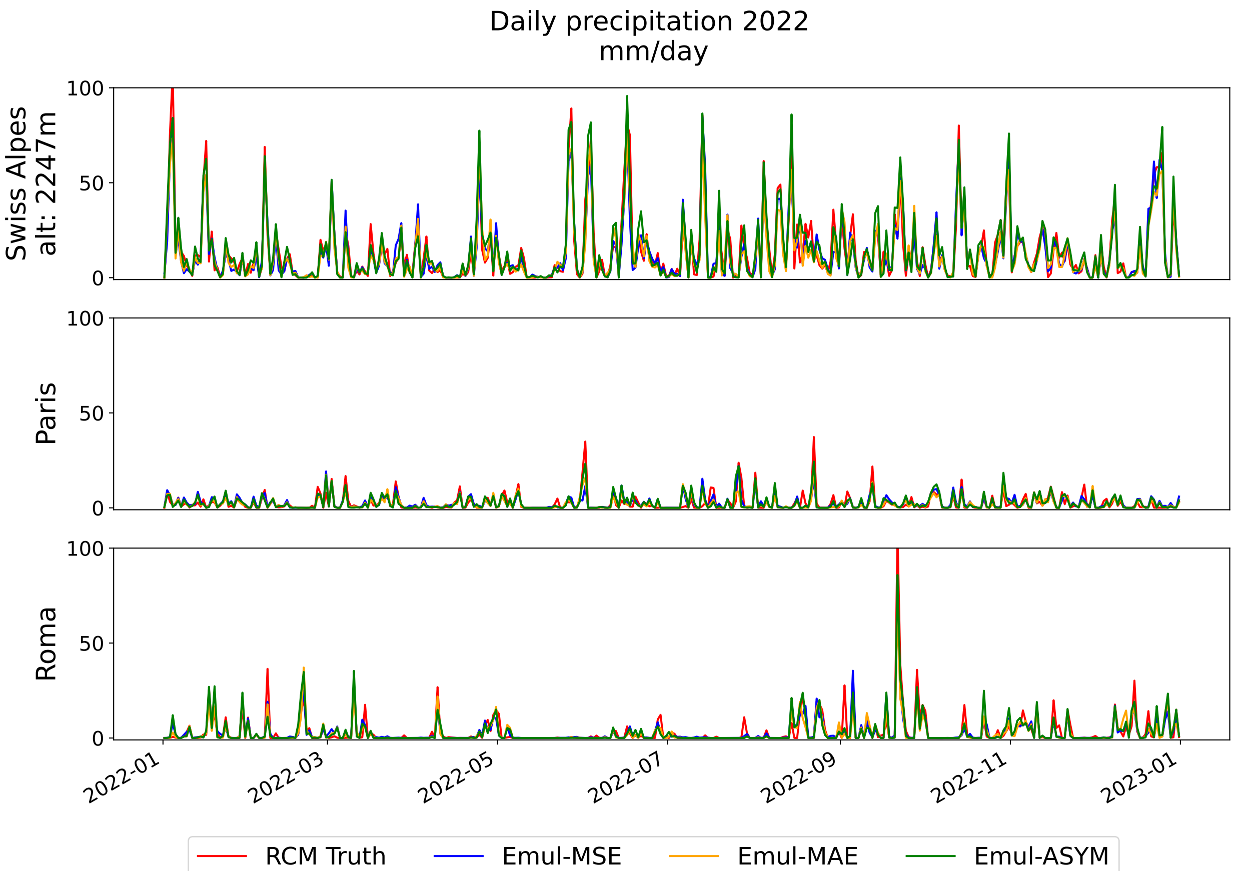

Fig. 6. Daily precipitation time series for four grid points. The RCM truth (in red) and the three emulators are plotted on each panel.

are producing too smooth objects. On the RCM maps, there are some very sharp and precise structures that the emulators fail to reproduce with the same precision. For example, on the lower panel, there is a hole with no rain over the southwest of France, which is missed by all emulators, even if Emul-MAE and Emul-ASYM make less intense precipitation over this area. The middle panel RCM map also shows very sharp structures that appear smoother in the emulators' maps. Nevertheless, the extreme points are well located for the three days.

In terms of intensities, the three emulators have mostly the correct spatial mean. Emul-ASYM reproduces better the spatial extremes as it has closer 95th superquantiles than Emul-MSE and Emul-MAE, which are both underestimating the spatial extremes on these three days. Emul-ASYM is overestimating the spatial SQ95 on the first panel, as it creates a more significant local extreme over the Alps than in the RCM map. It is, however, remarkable that this extreme is not inconsistent with the UPRCM map. Indeed it is interesting to notice the differences between the RCM and the UPRCM maps, which attest to the resolution's impact. The RCM is able to create sharp and well defined objects, with locally strong intensities. Regarding this aspect, the emulators seem to have an adequate capacity to refine the low-resolution maps and always recreate consistent high-resolution maps. Nevertheless, it

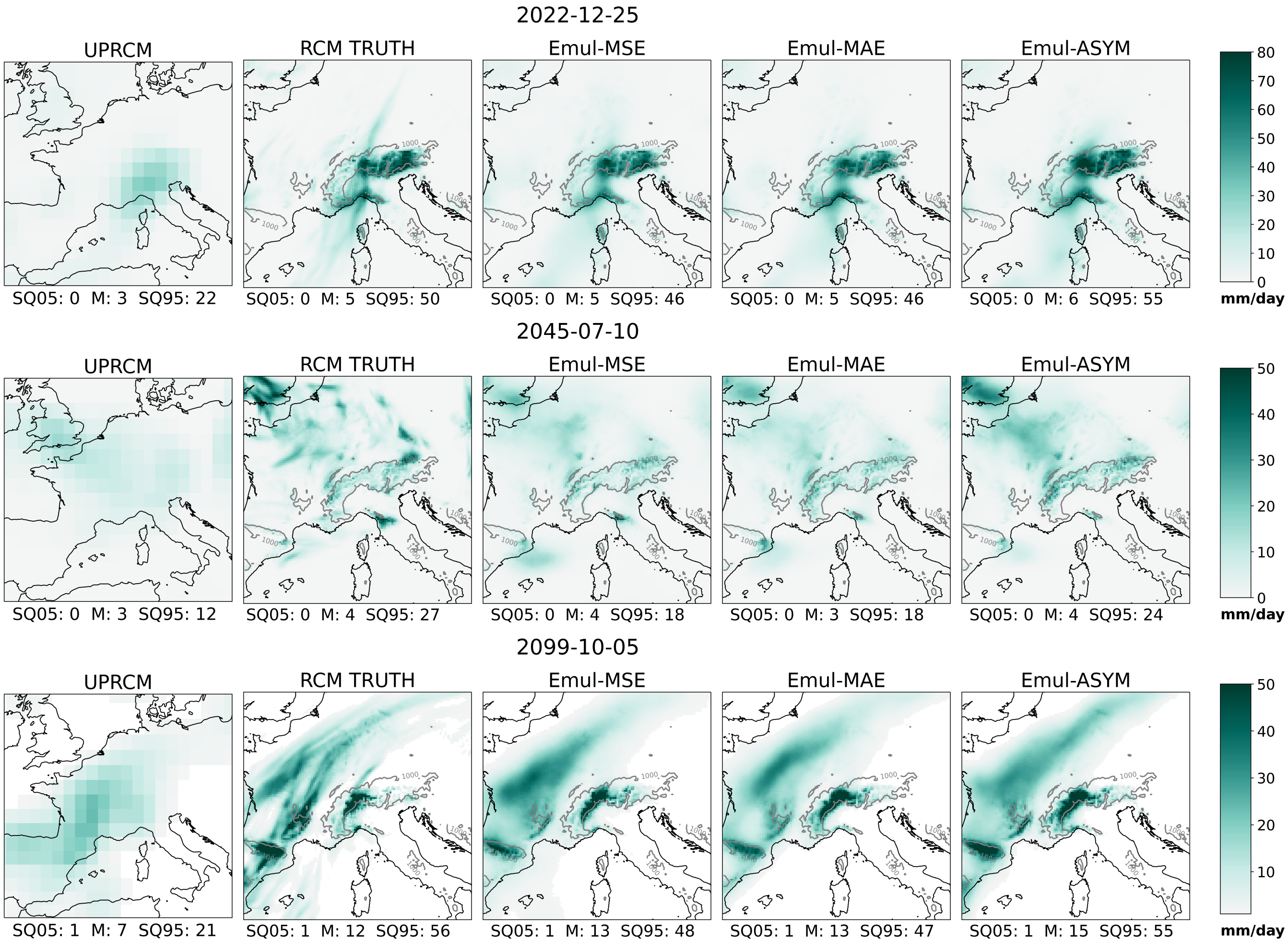

Fig. 7. 3 randomly chosen days illustrating the precipitation field of ALADIN63 at the Upscaled resolution (UPRCM), its native resolution (RCM truth). The three right-most plots show the precipitation field for each of the three emulators. The values corresponds to the spatial mean and 5th and 95th super-quantiles.

seems that the objects created by the emulator are smoother than the original RCM maps.

3.1.2 Daily scale analysis

In a second step, and to extend the first observations from the previous section, we can look at some scores over the time series. Firstly, the upper panel on figure 8 shows the Pearson correlation coefficients calculated between the RCM and the emulators' series in each grid point. The three emulators appear to have similar performances regarding this aspect, with a reasonable correlation (de-seasonalised and de-trended) with the true series over the whole domain. The best correlations are over the reliefs with Pearson coefficients larger 0.9. The lowest correlation appears over the driest area (cf Fig. 9), like the south of the Pyrenees or the North-East corner of the domain, but the correlations are still around 0.75.

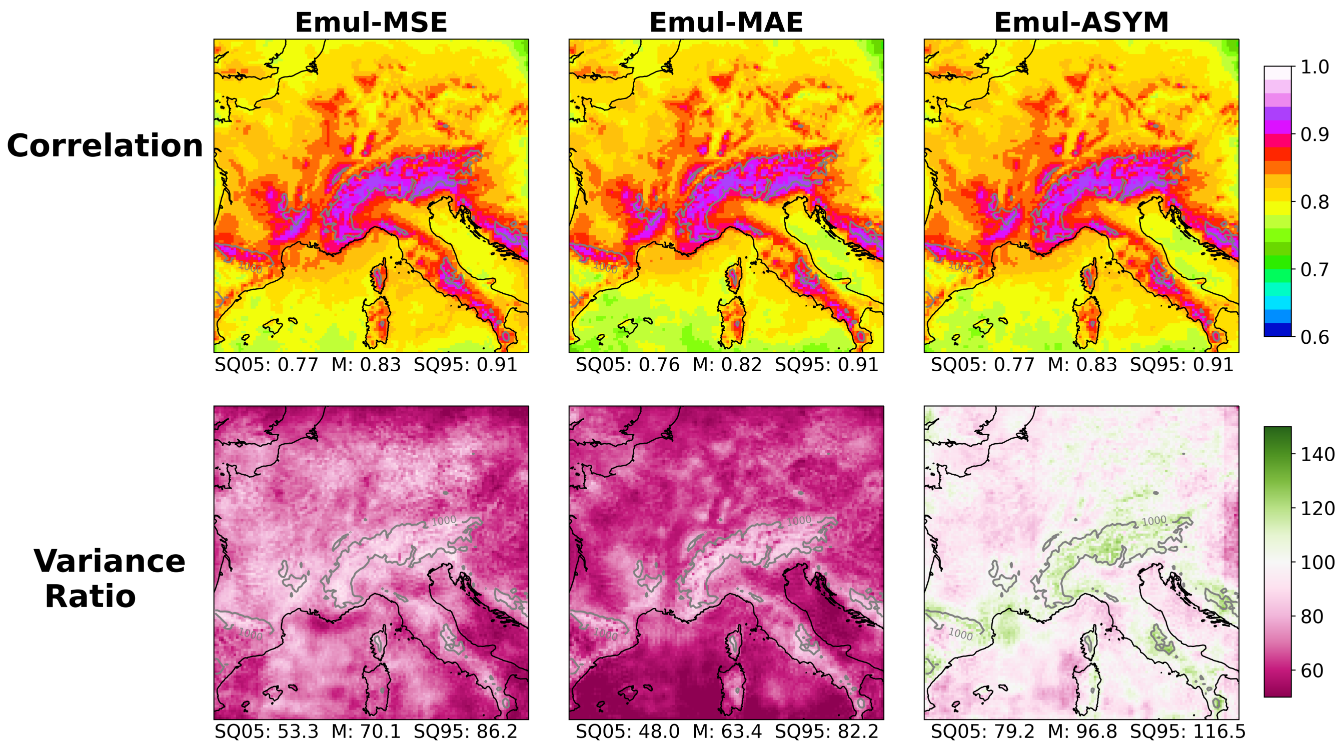

Fig. 8. Temporal Anomalies Correlation (up) and Ratio of variance (bottom) computed on the entire evaluation simulation (2006-2100) for the three emulators.

The lower panel on Figure 8 shows the variance ratio for the three emulators against the RCM truth. Emul-ASYM manages to reproduce in each point the RCM variance much better than the two others. Its variance ratio ranges from 80 to 120 percent, with a big part of the map being very light, showing about 100% variance reproduction. It slightly overestimates the relief's variance and slightly underestimates it over the regions with low rain average (cf Fig. 9). On the other hand, both Emul-MSE and Emul-MAE vastly underestimate the variance over the whole domain, even if Emul-MSE is slightly better.

It seems that the three emulators follow the large scale the same way, as they can recreate the chronology of the original RCM series very accurately. They can identify where and when the precipitations occur at the grid point scale, as shown by the good correlation maps. However, the loss choice seems to substantially impact the reproduction of the events' intensity as the emulators have different variance ratio maps. Let us see if this is confirmed when we look at aggregated statistics.

3.1.3 Climatological scale analysis

In this section, we look at some aggregated statistics to evaluate if the series produced by the emulator are statistically similar to the RCM one and how they differ. Figure 9 shows three climatological metrics over 20 years in the present period for the RCP4.5 simulation which is not in the training set. The upper panel shows the average daily precipitation over 2006-2025, the middle one is the 99th quantile, and the lower one shows the proportion of dry days. This figure illustrates well the impact of each loss function on the emulator.

The Emul-MSE mean is very similar to the RCM map. The spatial mean and superquantiles are the same. The bias map shows that it slightly underestimates the RCM values, but at maximum by 15% and over regions with low precipitations. However, it presents much poorer results on the other part of the distribution: it largely underestimates the 99th quantile (-15% on average) and the number of dry days (-10% on average). It is due to the nature of the mean squared error loss, mainly concentrating around the mean.

The Emul-MAE is, meanwhile, very accurate for the representation of dry days, very slightly overestimating them. However, it fails to reproduce the mean and the 99th quantile maps, broadly underestimating them. The MAE loss gives the same weight to all errors. Since the number of dry days is the most represented (between 35 and 85% of the days are between 0 and 1 mm) they weigh much more in the emulator training, so it mainly focuses on them.

The Emul-ASYM aims to correct the EMUL-MAE by giving more weight to the rainy days, proportionally to the amount of rain. It has similar performances to Emul-MAE over the dry days' map, which is expected since both emulators have the same loss function on this part of the distribution. However, the Emul-ASYM mean and 99th quantile maps are also very accurate. It shows in both cases less than 15% bias over the worst points and almost no bias on average over the maps. Regarding both climatologic maps, it seems to slightly overestimate the precipitation over the reliefs where it is raining the most and under-estimates at the driest points. Nevertheless, these errors are small, and the Emul-ASYM is clearly the best option if we aggregate the performances for the three metrics.

On all maps in Figure 9, it is striking to see how well the emulators reproduce the complex spatial structures. Emul-MAE and Emul-MSE have strong biases that are uniform over the domain. All three statistics present locally different patterns, and the emulators reproduce that. For instance, on the 99th quantile maps, there is a strong pattern in the Cevennes, just south of the Massif Central (France), which is much less intense in the daily mean map. It is the same for the emulators' maps. The spatial structure over Italy is also very complex; there is a thin line over the reliefs with more rainy days and higher extremes, which is also almost perfectly reproduced by the emulators. Similar examples exist for the entire domain.

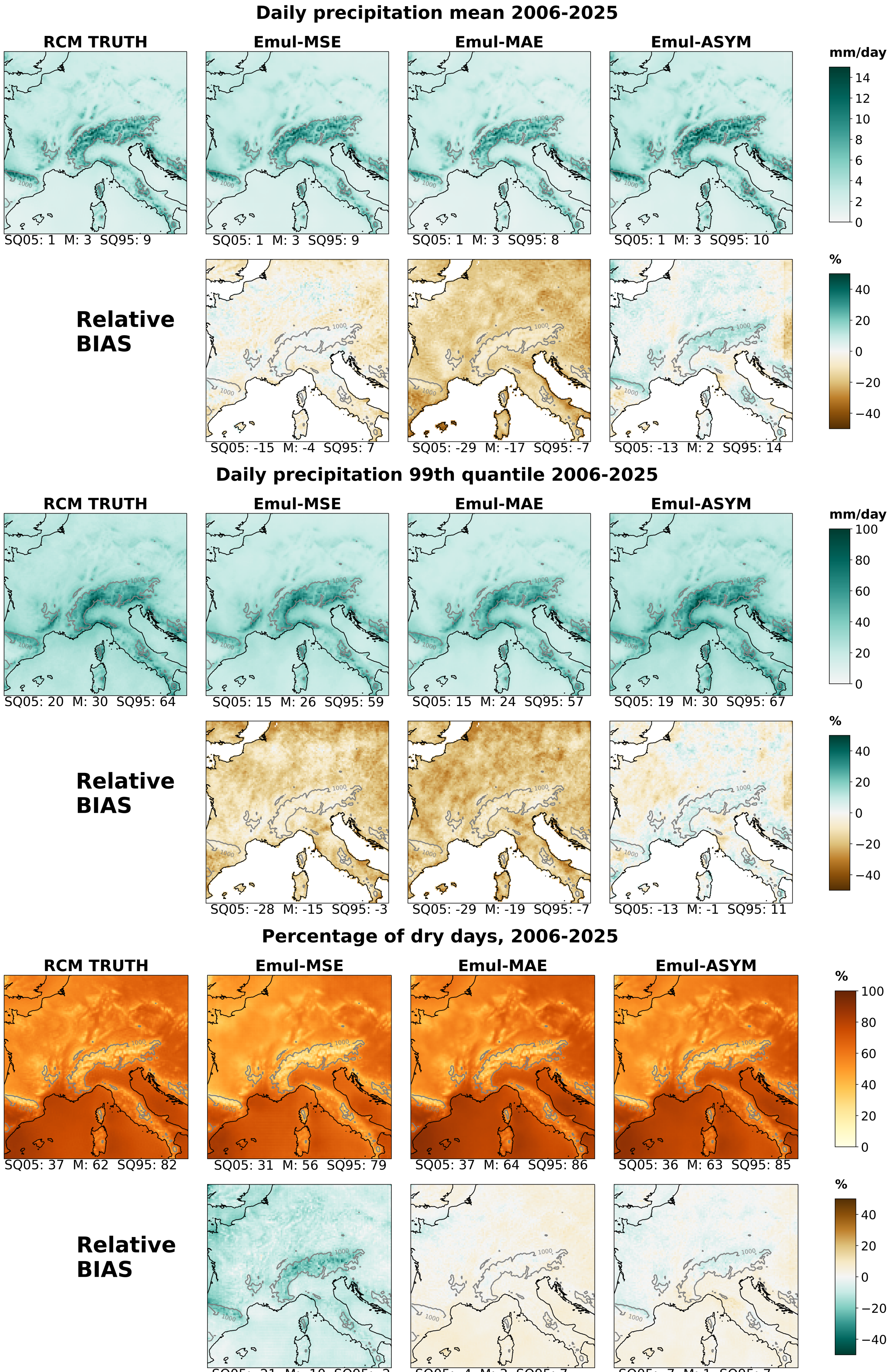

Fig. 9. (Upper) the mean map of daily precipitation accumulations over the 2006-2025 period, (middle) the 99th quantile map over the same period and (lower) the percentage of dry days. These three statistics are shown for the RCM and the three emulators. For each emulator and each metric, the relative bias maps are shown. The spatial mean and 95th and 5th superquantiles are given for each map.

In order to extend this result, we can look at the entire distribution using the ASoP method described in section 2.5.3. In Figure 10, the pdf analysis is detailed for the three grid points previously used: Paris, Roma and a high point in the Swiss Alps. The first column shows the events frequencies for each bin defined in section 2.5.3. Most days fall in bins under 0.1mm/day as the red curve comes from high on the left part of the plots. The Emul-ASYM and the Emul-MAE reproduce this part well, while the Emul-MSE underestimates the very low precipitations ($\leq 0.1mm/day$) and overestimates the ones between 0.01 and 10mm/day. It is less pronounced for the Alps point, where the event distribution is more uniform across the bins than the other three points. Emul-ASYM reproduces the frequency of these stronger events better than the two other emulators.

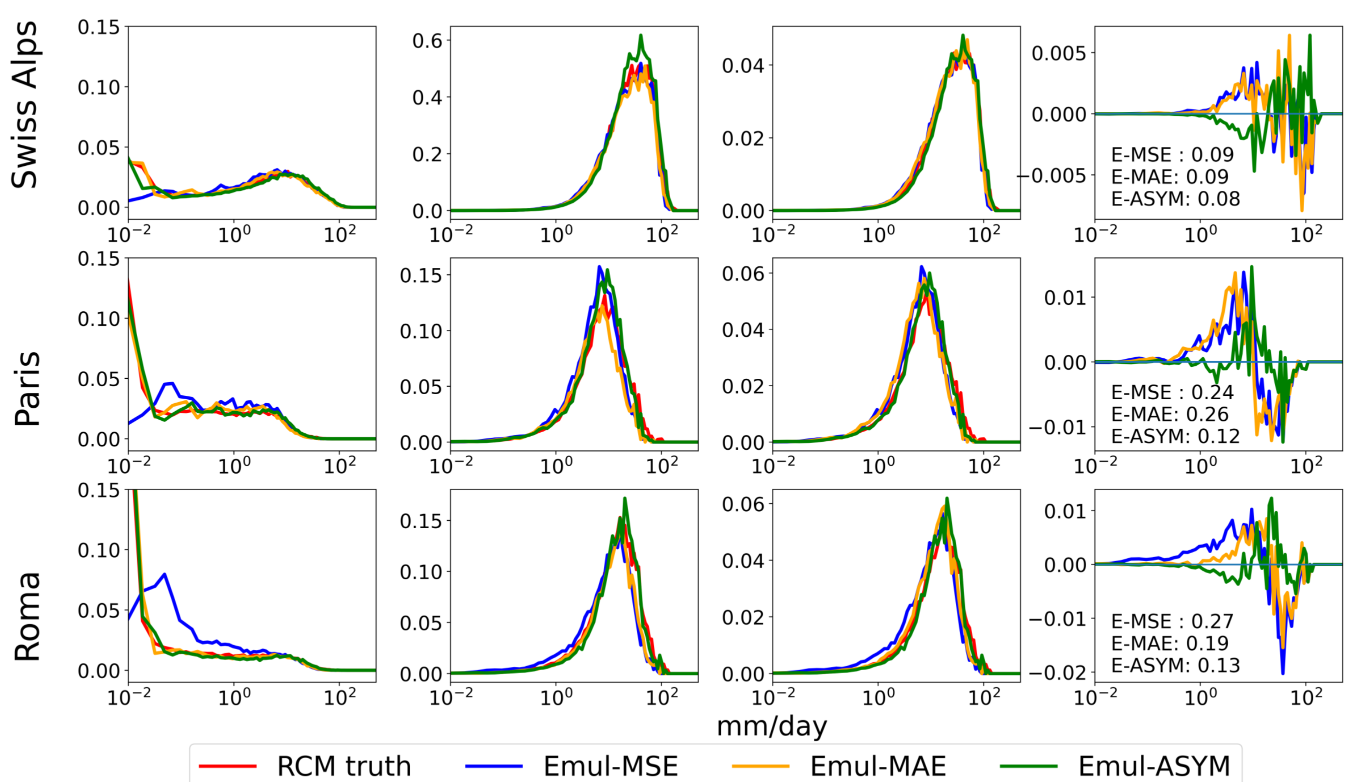

Fig. 10. Illustration of the probability density function analysis following the ASoP method (Klingaman et al, 2017) on three example grid points. Each line is a point and each column is a different step of the method. The first column shows the frequency of events in each bins, the second and the third the actual and the fractional contribution and the last column illustrates the skill score. The number in the last column plots are the scores for each emulator at the corresponding point.

The second column shows the actual contributions to the mean, which are the frequencies multiplied by the bins' mean. The first remark is that Emul-ASYM slightly overestimates the contribution of the precipitations around 10mm, which probably led to the wet bias on the mean map of figure 9. Emul-MAE produces insufficient rainfall over $\sim$ 8mm as the right part of the distribution is shifted to the left. The same remark applies to the Emul-MSE to a minor extent, which has a better reproduction of the mean, confirming what we saw in figure 9. Nevertheless, the Emul-ASYM matches better the right tail of the curve.

The last column illustrates the fractional contributions skill score by plotting the difference between the emulators and the RCM distributions on the third column. The fractional contributions are the actual contributions normalized by the mean of the series, allowing us to compare only the shape of the distribution across the bins. It helps to see that the Emul-MSE and Emul-MAE distribution are generally left-shifted, with too many small precipitations and not enough strong events. The Emul-ASYM curve generally looks better even if it tends to produce slightly too many precipitations between the mean and the 75th quantile. The regularization term in the Emul-ASYM loss appears to play its role pretty well as the distribution of precipitation is closer to the real one but might sometimes be too strong.

The skill score measures the area between the emulators' fractional contribution and the RCM one, and we can see that the Emul-ASYM outperforms the others over these three points. It is interesting to notice that Emul-MSE and Emul-MAE perform better over the Alps point, where the precipitations are more uniformly distributed across the bins. Finally, Figure 11 shows that the Emul-ASYM skill score is better over the whole domain. It generalizes the distribution analysis and confirms that the specifically designed loss function is more adapted than the two others to reproduce the highly skewed distribution of precipitation.

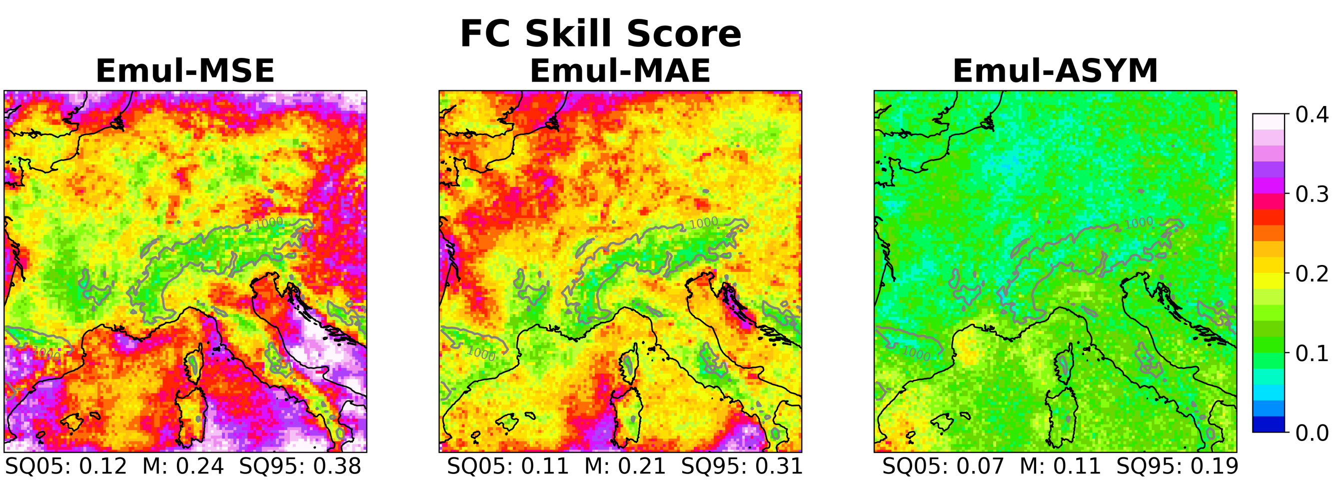

Fig. 11. Fractional Contribution Skill Score maps for the three emulators.

3.1.4 Conclusion on the comparison

Until here, we have analysed the role the loss function can play in the calibration of the emulator. Table 3 summarise the results obtained on the three emulator. They all demonstrated an excellent capacity to reproduce

Table 3: Summary of the emulators' comparison results

Emulators	Temporal correlation	Average precipitation	Low precipitation	Heavy precipitation	Variance	PDF Shape
MSE	++	++	−	−	−	−
MAE	++	−	++	−	−	-
ASYM	++	++	++	++	+	+

the daily precipitation time series with a good temporal correlation with the original RCM series. Nevertheless the loss function impacts strongly the intensity of the events. The MSE loss function penalizes strongly the large error which centers the prediction around the mean because of the chaotic nature of precipitations. Thus, if the mean daily precipitations is well represented the extremes are underestimated. On the other hand the MAE reproduces well the low precipitations but underestimate the intensity of larger events. Finally, the Emul-ASYM, thanks to a regularization term added in the loss function, managed to reproduce better the entire rainfall distribution at each grid point of the domain with notably a better reproduction of the extremes. Therefore, the loss function plays here as a cursor to set the event intensities, while the chronology of the series is captured from the predictors. From now on, we will consider only the EMUL-ASYM.

3.2 Deeper evaluation of Emul-ASYM

3.2.1 Object oriented analysis

Figure 7 seems to illustrate that the precipitation objects created by the emulator are smoother than in the RCM. The SAL method presented in section 2.5.4 is an objected-oriented evaluation approach which compares on two maps the object similarities.

Following the recommendation of Wernli et al (2009), we limited the evaluation to three subdomains of about 500km by the side. The blue boxes represent them on figure 2. The first subdomain focuses on the Cevennes regions. This part of South France is well known for its extreme autumn precipitation events. These events are the object of multiple studies (Ribes et al, 2019; Caillaud et al, 2021) because of their strong socio-economic impacts. It is then important to assess whether the emulator is able or not to reproduce such events. The second domain is another hotspot for Mediterranean extreme precipitation events (Ivušić et al, 2021) located in Croatia, over the Dinaric Alps and the North of the Adriatic Sea. The last subdomain is centred around Belgium, including the South-East of England, the North-East of France and West of Germany. This region presents a different climatology with extreme events of smaller intensities occurring more in winter.

Figure 12 presents the SAL scores' results. For each region, there are five SAL diagrams. The left most diagram represents the results for all rainy days. Then going to the right we consider only days where the spatial 99th percentile of the RCM truth series is above an increasing threshold. The threshold and the number of considered days are indicated on each diagram. Thus, from left to right we consider only more and more extreme events. The first general comment is that over all these diagrams, the emulator reproduces accurately the large majority of the events. Indeed the red boxes regroup 90% of the days and they are always centred around 0 with most points in deep blue, showing good Location score.

On the first column representing all rainy days, the emulator underestimates the global amount of precipitation over the domain, with the red box being pulled down left. As it gets more centered when we look only at stronger events we can conclude that the emulator misses some small precipitation objects. Knowing the chaotic nature of rainfall, we assume that it is perfectly fine if the emulator misses or add some small events. Moreover, the SAL metrics are one-sided: they evaluate how the predicted map matches the reference one. As we fix the threshold according to the RCM true series, it is logical that events, especially small ones, are missed or underestimated by the emulator. Besides, when we fix the threshold according to the emulated series, then the emulator overestimates the amplitude of some small RCM events and the red box is pushed up-right. It shows that the emulator sometimes misses small objects and sometimes creates some.

On the right of the figure, when we look at days with heavier precipitation, the amplitude gets centred around zero or slightly positive on the right-most column of the two Mediterranean regions. In addition, the emulator tends to produce larger objects with a positive S-component. However, the centre of the object is most of the time well located. It tends to generalize that the emulator produces smoother objects than the RCM, especially on significant intensities events.

There is a correlation between the amplitude and the structure metrics. It can attest that the emulator always creates objects consistent with the RCM. They are either smaller or bigger in terms of both shape and amplitude. On all diagrams, we can see some days with lousy location and structure scores but the correct amplitude. They are typical of days where the emulator produced too smooth objects and did not peak like the RCM. The emulator produces one large object with medium intensity, while the RCM produces multiple peaked objects with high intensities. It implies bad locations and structure scores but good amplitude.

Generally speaking, the emulator manages to reproduce the precipitation objects simulated by the RCM, even if they do not always have the perfect characteristics. The emulator captures most of the extreme events with the most suitable characteristics. The emulator seems nevertheless to produce smoother objects. A further analysis, with an application to a hydrological impact study, should be conducted to determine whether it is a fundamental limitation and how we could maybe adapt the emulator.

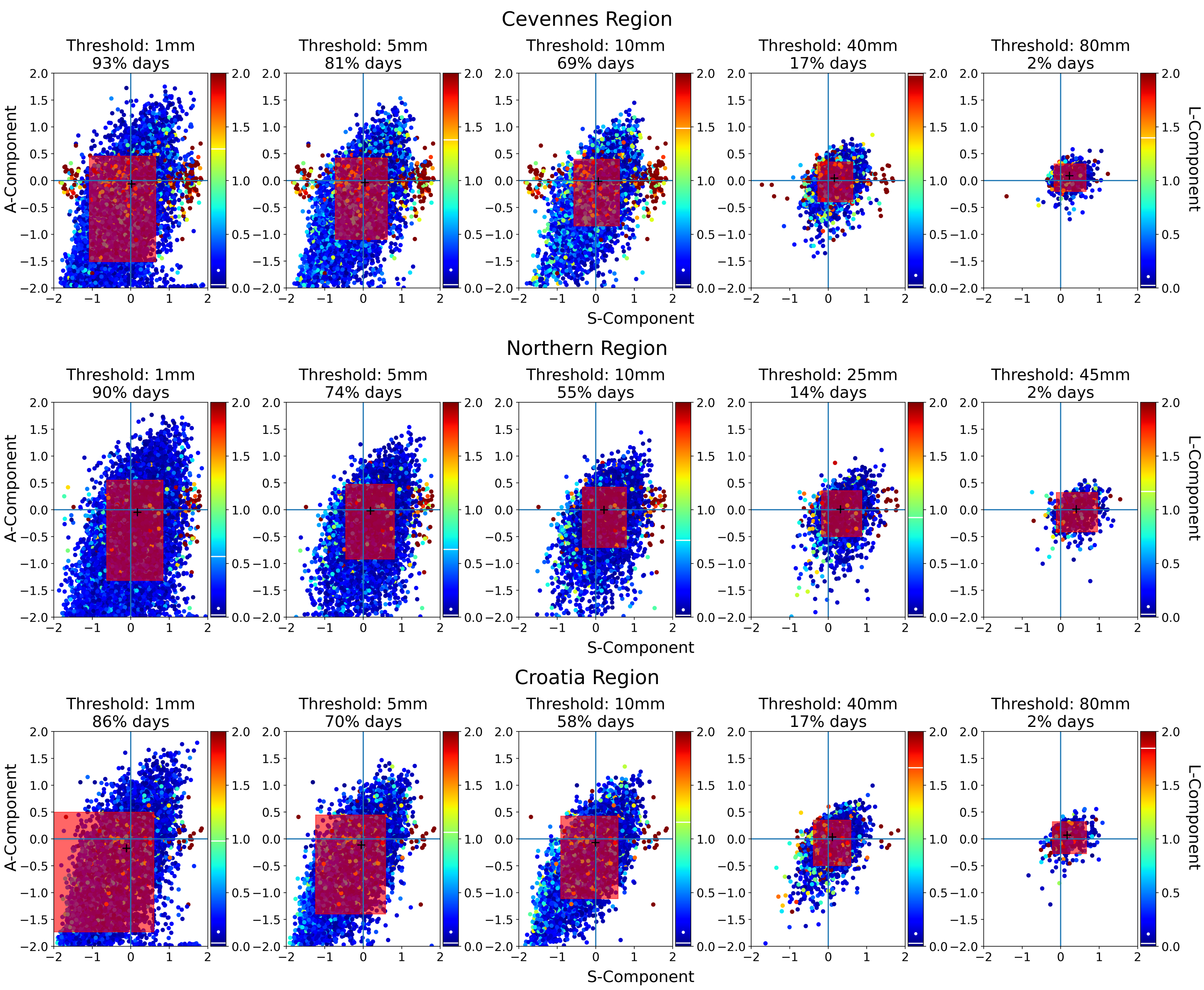

Fig. 12. SAL diagram for the three regions: Cevennes (up), North of the domain centred around Belgium (middle), and a region over Croatia and the North of the Adriatic sea. From left to right, the panel show the SAL results for days with maximum events intensities above an increasing threshold. Each point on the diagram represents a day with the Amplitude component on the y-axis, the Structure on the x-axis and the color give the Location score. The red box includes 90% of the points, and the black cross indicates the A and S median. The 5th, 50th and 95th quantiles are given in white on the colormap for the Location component.

3.2.2 ALADIN63 matrix extension

In order to give more robustness to the good performances of the Emul-ASYM, we can extend the evaluation to all ALADIN63 simulations available for our target domain. Indeed, up to now we focused the evaluation on the ALADIN simulation driven by CNRM-CM5 RCP4.5, which share the same driving GCM. The EURO-CORDEX matrix gives us the opportunity to evaluate the emulator on simulation driven a by different GCMs. This question of transferability to different GCMs, is an important challenge as it is a necessary condition for the application of the emulator for the downscaling of large ensemble of simulation.

Figure 13 summarizes climatological maps as the ones shown on Figure 9. The three panels (from left to right) correspond to the three statistics we looked at in Section 3.1.3: the mean amount of daily precipitation, the 99th quantile and the percentage of dry days over the 2006-2025 period. On each panel, the upper part shows the summary statistics for the raw maps of the RCM and the emulator, and the lower part summarises the relative bias maps of the emulator with respect to the RCM truth. On each panel, the columns correspond to a simulation. Each bar shows the spatial mean of the map, the upper bound shows the 95th super-quantile and the lower bound shows the 05th super-quantile. The first column shows the results for the CNRM RCP85 simulation, which has been used to train the emulator. The results on this simulation are given here as an indicator and cannot be taken alone to evaluate the emulator's performances. On each panel, the second column is the summary of the evaluation on the CNRM-RCP45 simulation presented on Figure 9. The bars illustrate well the main conclusions with for example a slight over-estimation over the wettest point (as the green bar goes higher) or the low biases on the lower panel.

The results are encouraging as the performances of the emulator are very similar across simulations, even if those are different in several aspects. For instance, the 3 CNRM simulations have higher daily means than the three others since the spatial mean and superquantiles are higher. It is less evident on the 99th quantile maps, where only the NCC simulation produces "eye-visible" less intense extremes. The emulator's bars reproduce the diversity of behavior of the various GCM-RCM pairs in terms of spatial patterns on the statistics maps. As observed previously, the emulator overestimates the average daily precipitation on the wettest points and underestimates it over the driest points whatever the statistics, which stays valid for all simulations. The biases on land points are similar to the ones observed for the CNRM RCP45 simulation, showing that the emulator reproduces each simulation with the same accuracy. In all these simulations, the emulator reproduces the three parts of the distribution well over the whole domain.

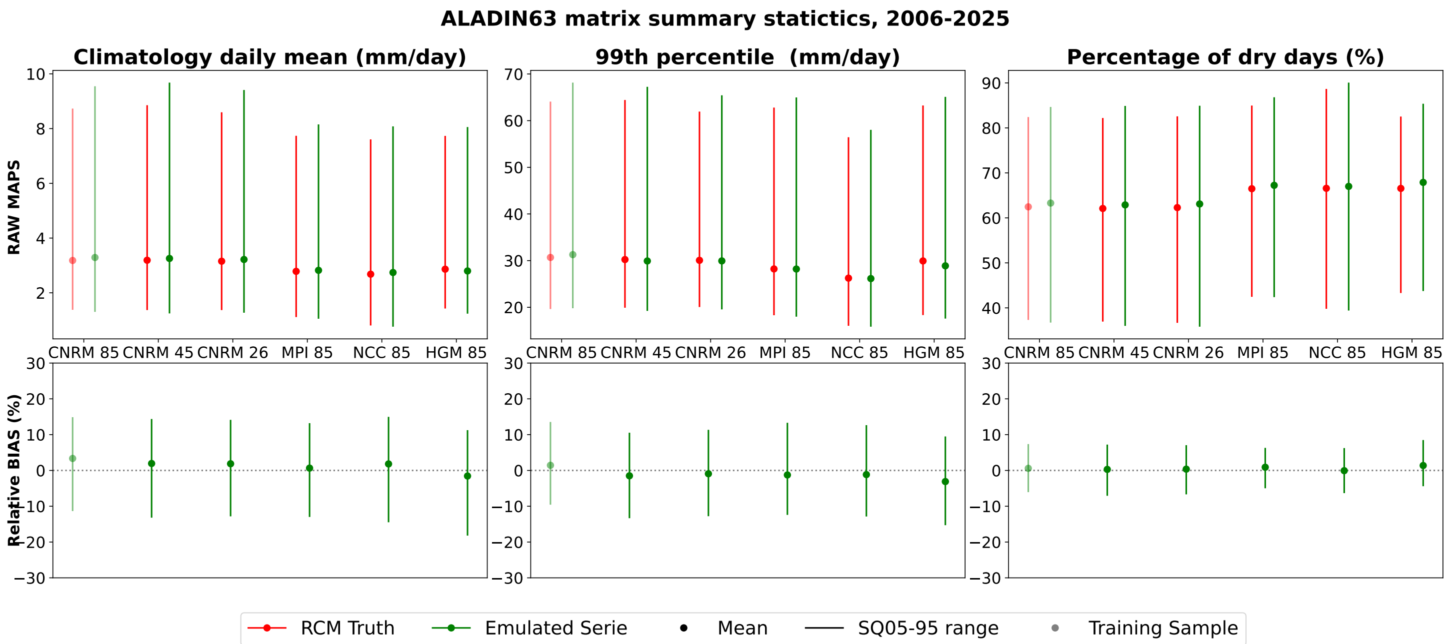

Fig. 13. Summary plots of the three climatological statistics regrouping the results on all ALADIN63 simulations. On each error bar, the lower (resp. upper) bound is the spatial 5th (resp. 95th) superquantile and the spatial mean is represented by the dot. The upper panels show the raw maps summary statistics for the RCM (in red) and the Emul-ASYM (in green), and the lower panels show them for the relative bias maps.

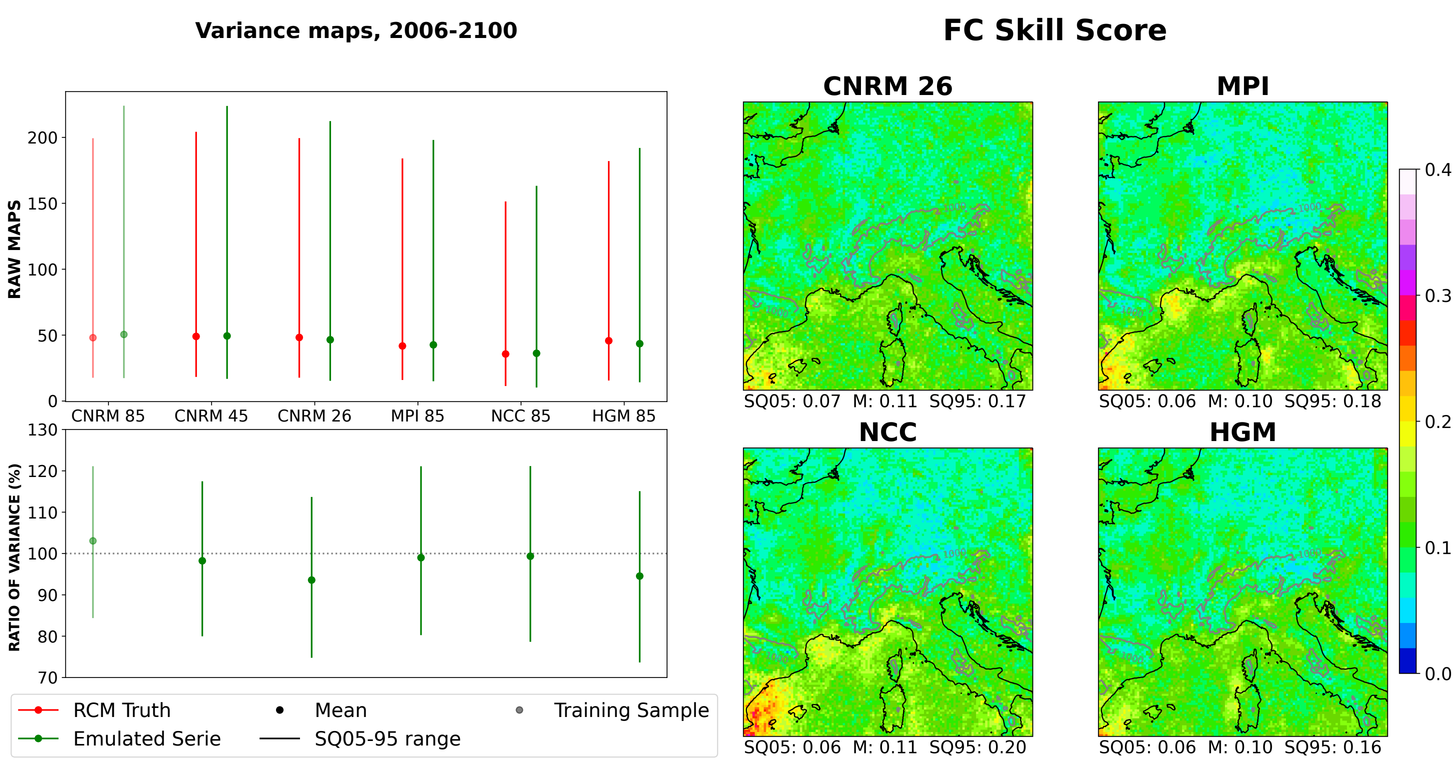

Fig. 14. The left panel show the summary plots for the temporal variance, the upper panel summarizes the raw variance maps, and the lower one shows the ratio of variances. The right panel shows the Fractional Contribution Skill Score (FCSS) for the 2006-2100 period for the four simulations that have not been shown yet (CNRM 26, MPI, NCC, HGM).

The analysis is the same regarding the variance maps summarized in Figure 14. The daily variance differs according to the simulation. For example, the RCM simulation driven by NCC has a smaller variance than the CNRM simulations or the HGM. The emulator reproduces in each case the variance maps quite accurately. However, in every simulation, it strengthens the variance where it is the strongest. The variance ratio summary plot confirms that the analysis made for the CNRM RCP45 in section 3.1.2 extends to all other simulations. The emulator can reproduce the daily time series with globally acceptable variance at every grid point. The temporal correlations (not shown) are also similar to what we observed on the CNRM-RCP45 simulation across all simulations. In the worse cases, it misestimates the variance by about 20%. Figure 14 also shows the FC skill score maps for the four missing evaluation simulations (the RCP45 simulation in in 11). Here again, we can observe that the emulators reproduce the shape of the precipitation distribution correctly at each grid point in all simulations. It is impressive to see how similar are these four maps. The emulator has similar performances across all simulations at the grid point scale.

3.2.3 Climate change reproduction

In order to finalise the evaluation of the emulator in the perfect model framework, we can look at the climate change maps. To do so, we will look at the three statistics used in the previous sections: the mean daily precipitation, the 99th quantile and the percentage of dry days. In each simulation, we compute the relative changes in a future period (2070-2100) versus a past period (1950-1980). The changes in precipitation are likely to be different according to the seasons over western Europe so we will look at the seasonal climate change here. The different studies about changes in precipitation amount over the region project a decrease in summer precipitations, notably around the Mediterranean sea, and an increase of winter precipitation on the North. Besides, a possible increase in extreme precipitation, especially over northern Europe, is expected. The results for the four seasons and the three statistics on all simulations are summarised through summary plots in Figure 15 while the results for the MPI and HGM simulations are illustrated in Figure 16. We chose those two maps as they show very contrasted climate change signal.

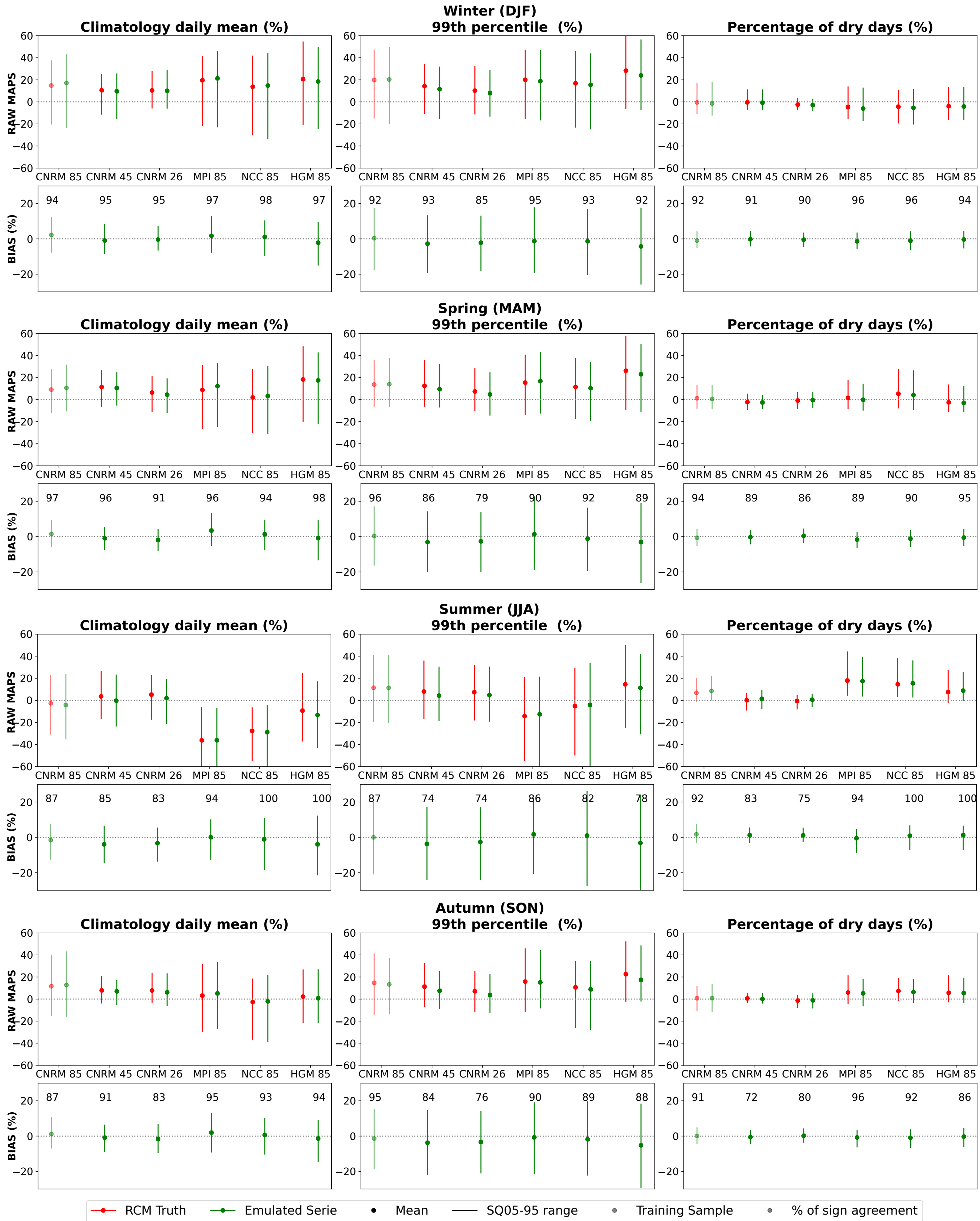

Fig. 15. Same as Fig 13 for the seasonal climate change (2070-2100 vs 1950-80) summary plots for the three statistics of interest: the daily precipitation mean, the 99th quantile and the percentage of dry days. The changes are the relative difference between the future period and the past one. The biases are simple bias between the emulator and RCM relative change maps. On each bias summary plot the number indicates the % of points where RCM and emulator agree on the sign.

The first remark is that on all plots summarising the raw maps, the green bar sticks very well to the red one, implying that the emulator correctly reproduces the maps and the intensity of the local changes. It is particularly notable on the summer plot, where the differences between the projections are the strongest. The MPI and NCC simulations show a substantial decrease in the mean daily precipitation over the entire map, associated with a global increase in the percentage of dry days. On the other hand, the HGM simulation projects an increase in average daily rainfall over some regions in summer. The emulator reproduces each simulation specificity with mainly the right intensity. Figure 16 shows summer and winter changes for the MPI and HGM simulations. It illustrates well that the emulator correctly captures the big spatial pattern. Still, in summer, we can observe that the emulator precisely places the regions where the HGM simulation produces an increase in average rainfall. This increase matches an increase of the 99th quantile in the same regions, and the emulator produces the same relationship. Similar analysis can exist on the winter maps, concluding that the emulator reproduces the ALADIN63 simulation with excellent accuracy.

Nevertheless, the emulator's maps are more continuous than the RCM maps, especially for the 99th quantile maps, which are patchy. It results in significant local biases between the emulator and the RCM maps. It partly explains the large biases on the bias maps summary plots in Figure 15. Generally, the emulator tends to overestimate some changes as we can see that the green bar is often longer than the red one. The number given on top of the bias maps summary plots shows the percentage of sign agreement between RCM and emulator over the grid points. It shows that the emulator identifies well the changes as these numbers are very high (always above 75%, very often above 90%). Moreover, on the bias maps of Figure 16, the hatching shows the points where RCM and emulator disagree on the signs. It is visible that they mostly correspond to points with minor changes.

To conclude, the emulator can reproduce high-resolution climate change maps with the same strong spatial pattern and intensities. Another relevant remark, not shown here, is that Emul-MSE and Emul-MAE have the same ability as Emul-ASYM to reproduce the climate change maps. It means that each emulator keeps the same biases along the simulation, and the changes are mainly driven by the large scale, which the emulators captures well.

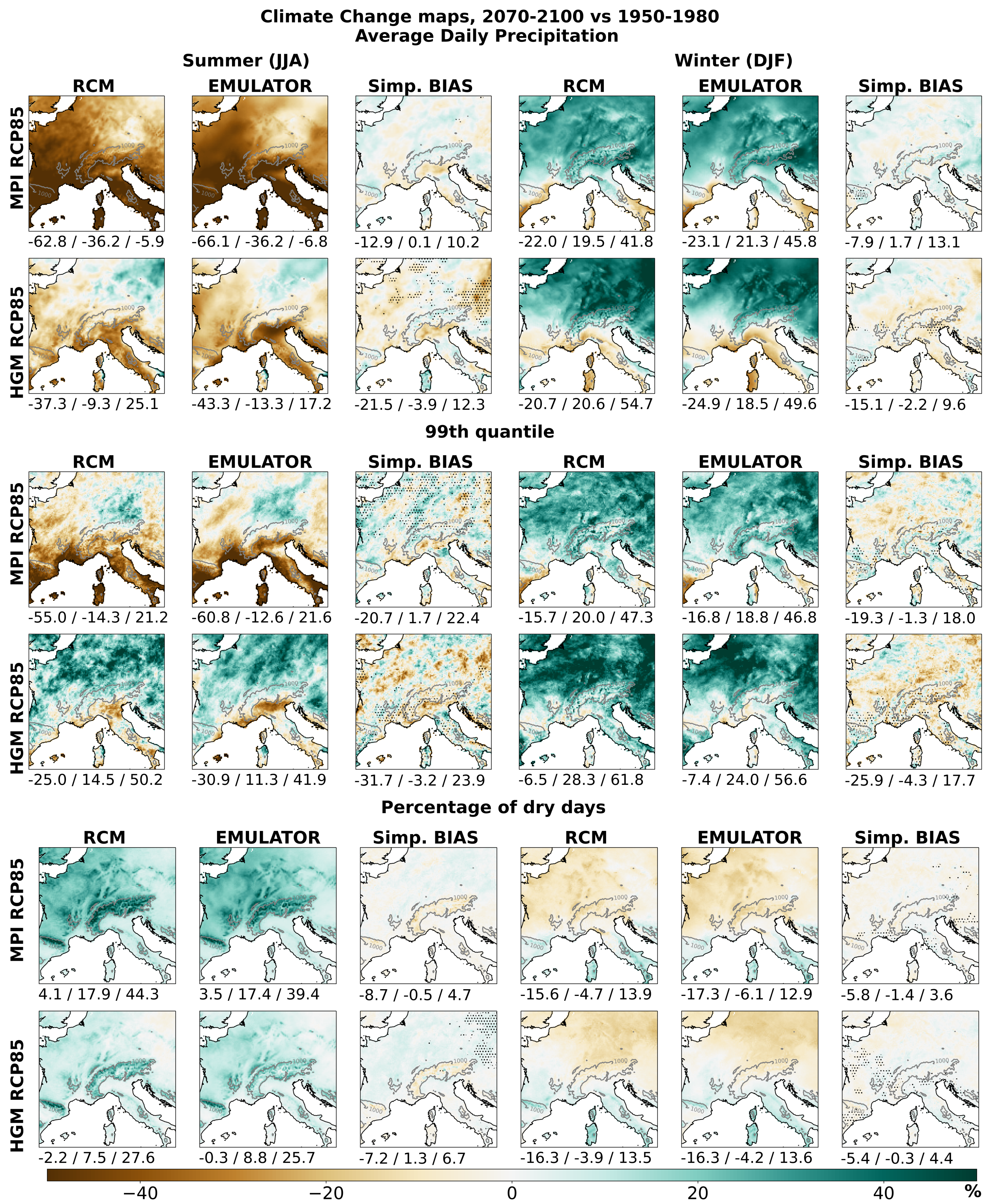

Fig. 16. Relative changes (in %) between 2070-2100 and 1950-1980 for the MPI and HGM driven simulations regarding (up) the mean map of daily precipitation accumulations, (middle) the 99th quantile map and (low) the percentage of dry days. These three statistics are shown for the RCM and the emulator, plus the simple bias map between the two. For each map, the spatial mean and 95th and 5th superquantiles are given. The hatching indicates the point where RCM and emulator disagree on the sign.

3.2.4 Conclusions on Emul-ASYM

Through sections 3.1 and 3.2 we have analysed the ability of the emulator trained with the asymetric loss function to reproduce the precipitation field simulated by the RCM. The conclusion on the emulator performances are summarised here:

- The emulator is able to produce realistic precipitation time series well correlated to the RCM ones and with the right spatio-temporal variability.
- The grid-point regularization term in the asymetric loss function helps to respect and reproduce the entire complex distribution of precipitation everywhere on the target domain.
- The emulator tend to underestimate the precipitation in generally dry regions and overestimate it in the wettest parts of the domain.
- The emulator creates coherent objects of precipitation, with generally the right characteristics even if they tend to be too smooth (i.e. less sharp and precise than the RCM objects).
- Those conclusions are the same for any RCM simulations available to evaluate the emulator in perfect model, including the ones driven by different GCMs than the one used during the training. It notably showed good ability to reproduce the diversity across simulations which attests for the good transferability of the learnt function and so gives some confidence on its applicability to various GCMs simulations. This is a key results for future applications.
- Finally the climate change maps obtained from the emulated series are almost identical to the RCM ones. It gives a lot of confidence to use the emulator in climate change context.

The emulator present therefore satisfactory results in perfect model evaluation and, even is there is space for improvements. The proposed loss function allowed to reproduce correctly the entire precipitation distribution at the grid point scale validating so far the use of the RCM emulator for precipitation downscaling.

4 GCM data application

This section aims to assess the emulator's applicability to GCM simulations. The ultimate objective of the emulator is to downscale large ensembles of GCM simulations to generate high-resolution simulations, allowing the study of local precipitation evolution and the associated uncertainty. Hence, it is crucial to evaluate if the emulator is indeed applicable to GCM simulations while maintaining similar performance levels than in perfect model. The application protocol is illustrated in the right panel of Figure 1, where the emulator processes GCM data after interpolating them onto a common grid. In this evaluation, we utilized the emulator to downscale four RCP85 GCM simulations—CNRM-CM5, MPI-ESM-LR, HadGEM2-ES, and

NorESM1 (refer to Table 1), which were employed to drive ALADIN63. The corresponding RCM simulations serve as a comparison basis, yet they cannot be deemed as the reference truth for the emulated series. Indeed, as elucidated in Doury et al (2022) and in Section 2.1, differences between an RCM simulation and its driving GCM entail low day-to-day correlation and long-term statistical disparities. The challenge of this section therefore lies in evaluating whether the emulator generates a series that aligns with the large-scale characteristics of the GCM while incorporating high-resolution features from the RCM. Another way to frame the objective of this section is that we try to identify if the Emulator in GCM application mode is able to reproduce an added-value with respect to its driving GCM similar to the one proposed by the original RCM. Consequently, we will compare the emulator's output with both the RCM and GCM series. Our expectation is that the emulator produces a series consistent with the GCM's large scale while integrating high-resolution features akin to those introduced by the RCM.

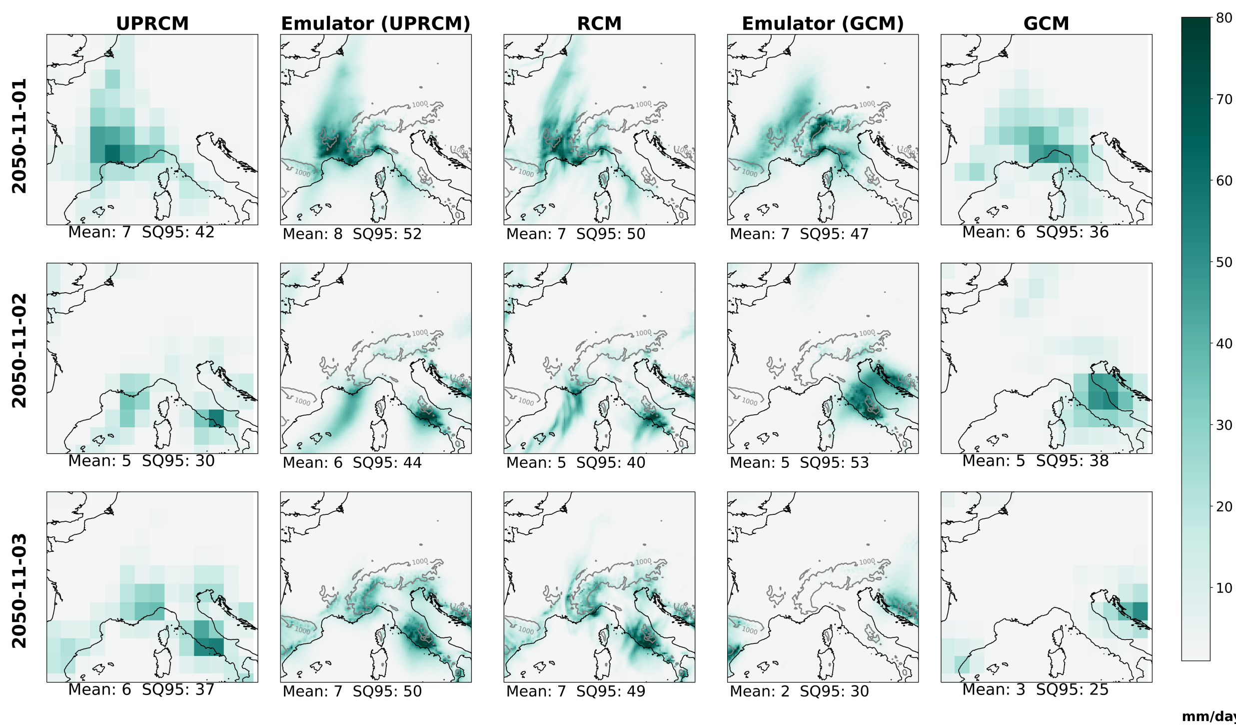

Fig. 17. Illustration of three consecutive days for the UPRCM, the emulator downscaling the UPRCM, the RCM, the emulator downscaling the GCM, and the GCM precipitation fields.

4.1 Illustration of the daily GCM/RCM differences

Figure 17 showcases the precipitation field for three consecutive autumn days in the CNRM RCP85 simulation. Each day includes the RCM truth simulation alongside the emulated maps in perfect model (UPRCM) and application (GCM) mode, complemented the UPRCM and GCM precipitation maps for the respective days. It is important to remember here that the low-resolution precipitation field is not a predictor. The UPRCM precipitation is simply the RCM map interpolated on the GCM grid, and we use it to compare with the GCM precipitation map.

These three days vividly illustrate the daily low correlation between the RCM and its driving GCM. Comparing the low-resolution maps reveals distinct chronologies. For instance, on day 1, the RCM depicts a significant Mediterranean event in southern France, later moving toward the Alps and Italy. In contrast, the GCM on day 1 exhibits a heavily localized precipitation event more eastward, over the southern Alps. These disparities result in very different extremes between the simulations at the daily scale.

However, the three high-resolution maps offer assurance regarding the emulator's ability to downscale GCM simulations. It generates a series consistent with the GCM, depicting precipitation objects that align with the story presented by the GCM. Moreover, the emulator refines the high resolution in a manner similar to the RCM. For instance, on day one, it precisely localizes extremes in the Alps and along the northern Italian coast. On day two, the GCM's situation over Italy closely resembles the RCM's depiction on day three, with the emulator producing similar events in mid-Italy in both cases. The emulator also adjusts the intensity of extremes, generating stronger extremes compared to the GCM as captured by the SQ95. However, it exhibits similar limitations in both UPRCM and GCM applications, with objects appearing overly blurred and lacking sharpness, as discussed in section 3.2.1. This consistency underscores the emulator's stability when downscaling GCM data. These three days exemplify the challenge of evaluating the emulator in application mode without a proper reference, given the day-to-day mismatches that hinder distinguishing potential emulator issues from large-scale-induced divergences.

4.2 Present climate analysis

In this section we analyse the series downscaled by the emulator in present climate. As in the perfect model evaluation, we compute the annual average daily rainfall, the 99th quantile and the percentage of dry days in the present climate (2006-2025) in the four simulations. We compare the emulator's maps with the RCM ones and the GCM ones.

The most striking observation lies in the added value brought by both the RCM and the emulator when compared to the GCM maps. CNRM, among the GCMs, exhibits some spatial structure across all three statistical measures, while the remaining three show notably flat maps, especially concerning extremes. The emulator's maps exhibit a high spatial correlation with the RCM ones, effectively replicating the fine-scale spatial structure across mean climate conditions and within dry or wet extremes. It successfully captures topography-driven spatial patterns, portraying areas like the central Alps experiencing more precipitation compared to the rest of the range across all RCM and emulator simulations. Additionally, intricate structures over Italy and the Mediterranean coastline are faithfully reproduced by the emulator. Another point of validation is the spatial super-quantile that are comparable with the RCM, confirming the emulator's high-resolution consistency with the RCM.

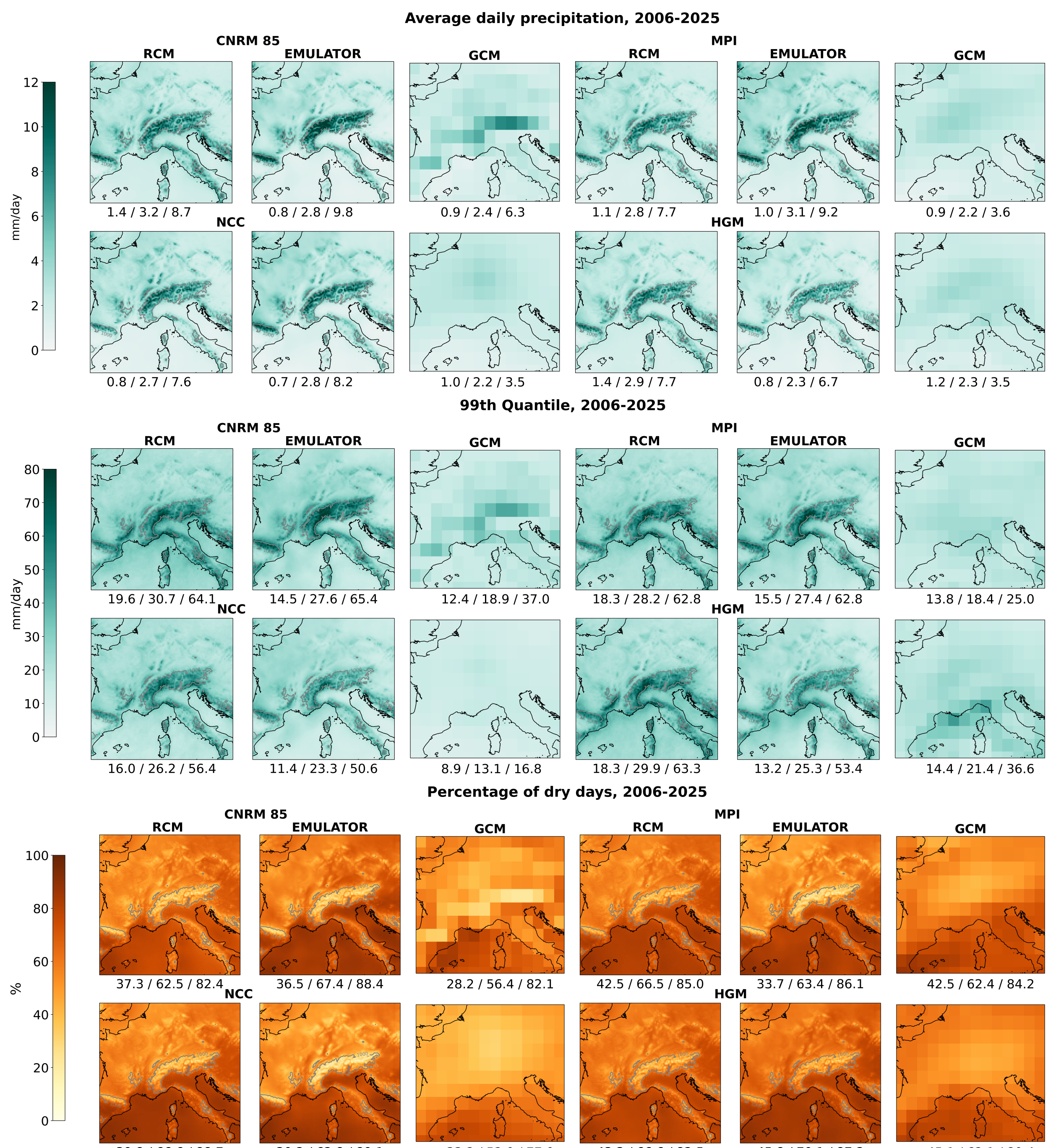

Fig. 18. Present (2006-2025) climate statistics of 4 simulations (CNRM RCP85, MPI, NCC and HGM) for (Upper) the mean map of daily precipitation accumulations, (middle) the 99th quantile map and (lower) the percentage of dry days. For each simulation, we see the RCM, the emulated one and the corresponding GCM map. The spatial mean and 95th and 5th superquantiles are given for each map.

In all four simulations and across the three statistical measures, significant disparities exist between the emulator and the RCM maps. As explained in sections 2.1, the daily inconsistencies between GCM and RCM large scales can lead to climatological differences. For instance, the emulator driven by CNRM generates more intense precipitation over the Alps than the RCM simulation, resulting in a higher 99th quantile and fewer dry days in the region. Conversely, the HGM-driven emulator simulation reflects a drier tendency, characterized by a lower 99th quantile and a larger number of dry days across the entire domain. The consistency between the three statistics and the fact that the differences vary accross simulations tend to support the hypothesis of real large scale differences rather than a problem in the emulator downscaling.

However, some biases in the emulator's outputs warrant attention. For instance, all emulated simulations underestimate the 99th quantile over the Cevennes in southern France. This region is recognized for its extreme events, an area where the RCMs usually bring a proven added-value at daily scale. While the emulator generates significant extreme events here, they appear comparatively less intense than those over the Alps in contrast to the RCM maps, where they exhibit a similar intensity. Dedicated studies specifically investigating the added value of emulators compared to RCMs and GCMs by analyzing particular events could certainly be conducted. However, such studies are beyond the scope of our current investigation.

4.3 Climate change analysis

In order to complete the study of the emulator ability to downscale GCM simulations, we propose to look at climate change maps. Given the inherent challenges in assessing the emulator's performance when downscaling GCMs, we will emphasize specific examples in this section. While the emulator is not expected to precisely replicate the changes simulated by the RCM, it should align with those produced by the GCM while integrating small-scale features consistent with the RCM. We compare the changes in autumn precipitation presented in Figures 19 and 20 produced by the emulator maps for the four simulations with the RCM and the driving GCM simulations.

These figures affirm the emulator's capability to incorporate high-resolution features into GCM simulations. In terms of both extremes and mean changes, the emulator generally aligns with the patterns observed in the GCM maps. For instance, the CNRM simulation exhibits an intensification of autumn precipitation over the northern domain, particularly noticeable in the 99th quantiles. The emulator echoes this trend, demonstrating a consistent signal with a more refined localization of pronounced changes, notably over northern and western France. The Emulator also clarifies the North-South contrast

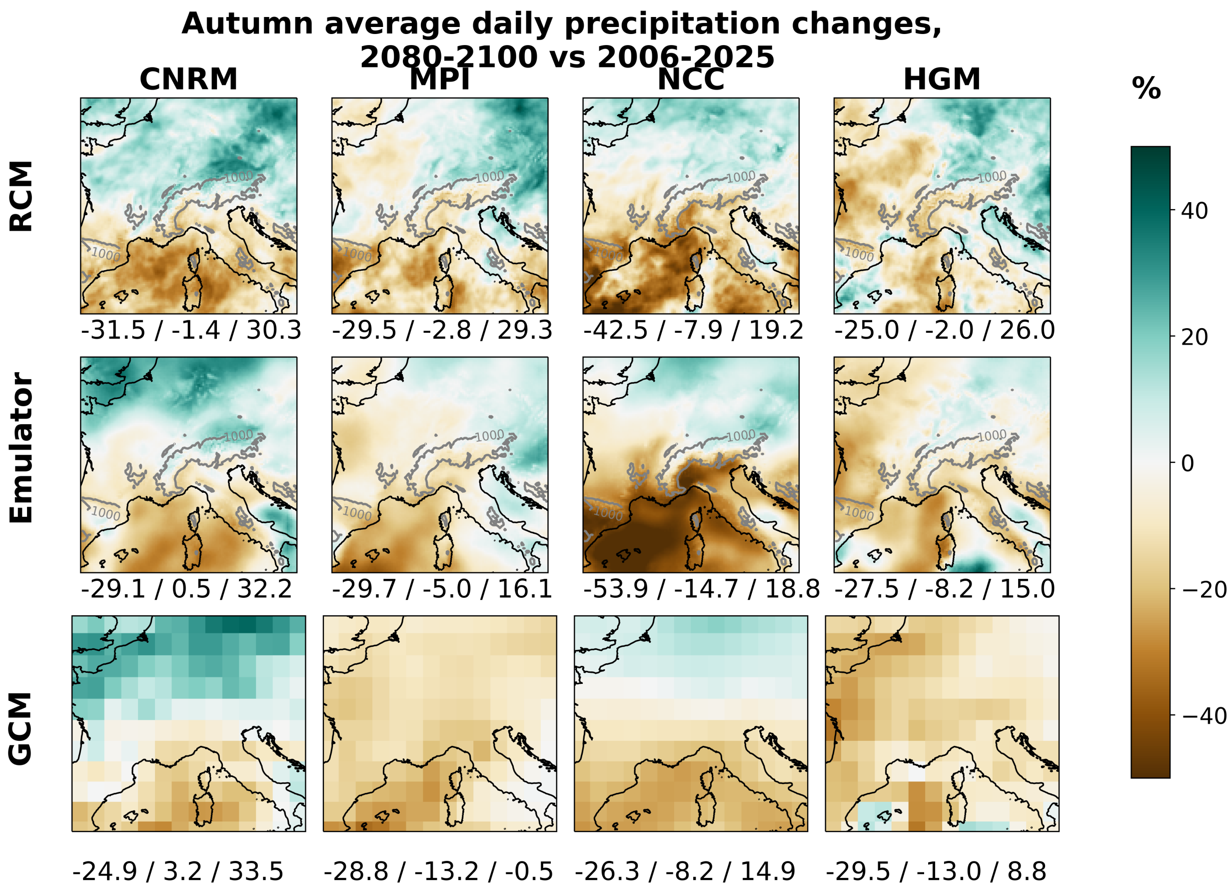

Fig. 19. Autumn relative changes of average daily precipitation between future (2080-2100) and present (2006-2025) period for the 4 GCM simulations downscaled with the emulator: CNRM, MPI, NCC and HGM under RCP85 scenario. From up to down, the rows show: the RCM, the emulator downscaling GCM, and the GCM maps. The spatial mean and 95th and 5th superquantiles are given for each map.

in precipitation change in the Alps with respect to the GCM low-resolution map, mimicing well the RCM pattern.

Moreover, the emulator appears capable of modifying the signal produced by the GCM. For instance, both the MPI and HGM simulations indicate a decrease in average precipitation across the entire domain, despite showing an intensification in the 99th quantile. In contrast, the emulator portrays an increase in autumn precipitations over the eastern domain, propelled by a more substantial intensification of extreme events in those regions. If it is difficult to assess for the validity of the modification, it is in agreement with the two other emulated simulations and the four RCM maps.

Even if some spatial structures are consistent between the RCM and the emulator maps, they remain fundamentally distinct. The emulator's structures are generally smoother than the RCM ones. However, the maps produced by the emulator include realistic high resolution features influenced by topography or coastline for example. Setting aside the differences in smoothness, distinguishing between the RCM and emulator maps becomes a challenging task.

4.4 Conclusion on GCM applications

To conclude on the emulator suitability to downscale GCM simulations, we have seen across different time horizon that the emulator behaves as expected. It applies the downscaling to the GCM large scale as it produces realistic high resolution fields. The consistency between the GCM patterns and the emulator ones plus the presence of high resolution features coherent with the RCM simulations give confidence in emulator downscaling. However, the emulator does not learn how to reproduce the large-scale transformations carried out by the RCM, resulting in differences between the precipitation simulations produced by the emulator and the RCM. In this context, it is difficult to give full confidence to the emulator when downscaling GCM simulations and further studies must be conducted in this purpose. In particular, it seems important to look for a proper evaluation framework of the Emulators in application mode.

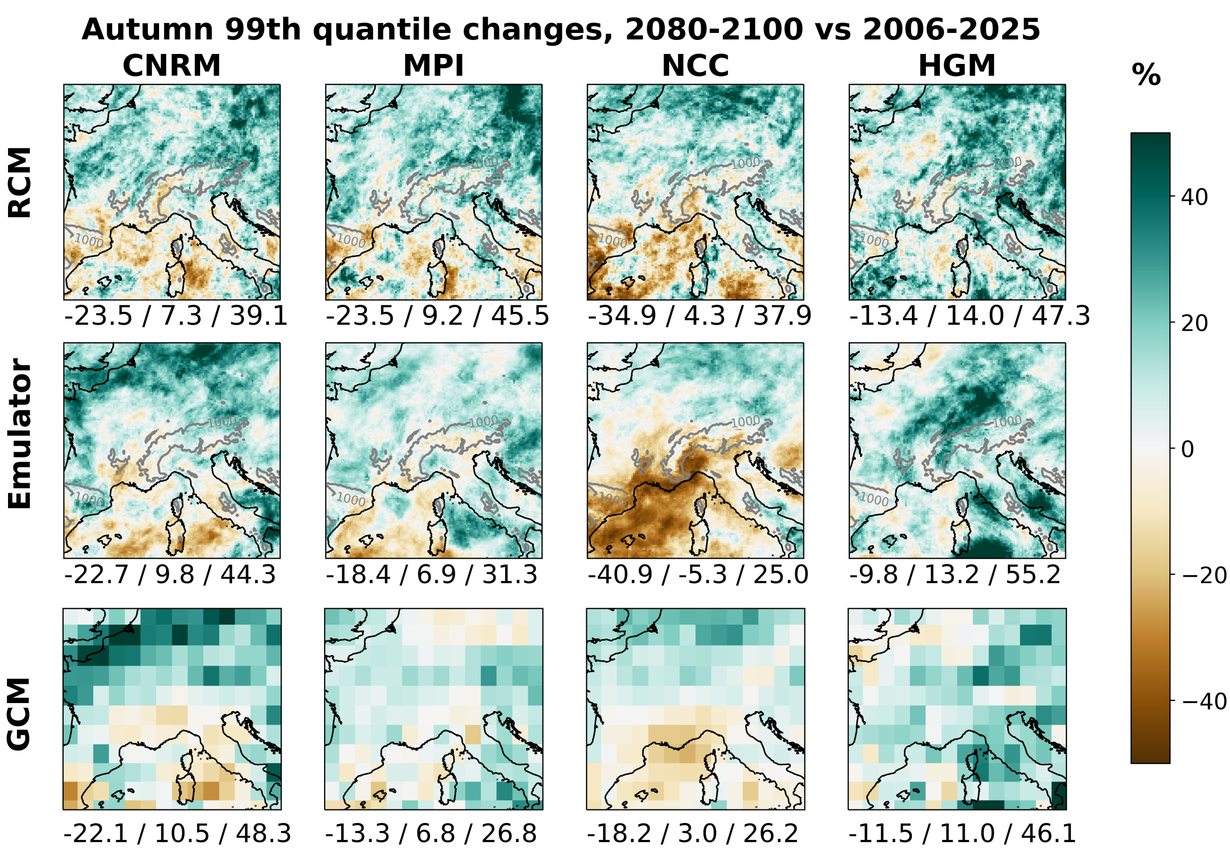

Fig. 20. Same as Figure 19 for the 99th quantile changes

5 Conclusion

This study aims to propose a credible solution to the high computational costs of Regional Climate Models to build large ensembles of high-resolution precipitation projections at daily scale. It extends the RCM-emulator introduced in Doury et al (2022) for the case of temperature downscaling. RCM-emulators belong to the family of hybrid downscaling methods. They use RCM simulations to estimate the downscaling relationship between low-resolution and large-scale variables and a high-resolution surface variable. It is important to recall here that the present study propose an emulator of a given RCM, CNRM-ALADIN63, in its EURO-CORDEX configuration. This manuscript has three main objectives:

1. Addressing the suitability of the emulator for the complex variable of precipitation, including the extreme parts of its distribution.
2. Studying the transferability of the trained emulator to different sources of inputs.
3. Evaluating the emulator behavior when applied to GCM simulations.

To address these objectives we extended the Doury et al (2022)'s work with some developments while keeping as most the same basis. Indeed a strength of the RCM-emulator should be its universality across domain or variables. Thus the emulator presented here relies on the same perfect model framework as in Doury et al (2022), it takes the same list of predictors and the neural network architecture is simply adapted to match the new input and target domains. The target domain considered here is four times bigger which also implied increasing the size of the input domain. Because of the non-gaussian nature of precipitation we proposed an asymmetric loss function and put those results in perspective with two classical functions for regression problems (MSE and MAE). Finally we also extended the evaluation of the emulator to a larger test set including simulations driven by various GCMs allowing to study its transferability. A first result is the good stability of the methodology set in Doury et al (2022) with a bigger domain even regarding to computational efficiency.

Regarding the first main objective we have shown that RCM-emulators are a credible strategy to downscale precipitation fields. The perfect model evaluation ensures a perfect reference against which we can precisely evaluate and compared the three emulators. All of them managed to capture the relationship between the daily large scale circulation and the associated high resolution precipitation accumulation as they all showed very good temporal correlation. It validates the concept of the emulator as it is possible to identify and learn the RCM downscaling function associated to precipitation. Nevertheless, only the asymmetric loss function ensured the emulator to reproduce the full high resolution daily variability that the RCM creates as well as the entire precipitation distribution including strong and rare events. Indeed, we have seen that a dedicated loss function to re-balance the data is necessary to deal with precipitation, and the one introduced here is a credible strategy. We also evaluated the accuracy of precipitation object created by the emulator. We found that they are quite realistic and coherent even if they tend too be smoother and less precise than the RCM ones. An other defaults of the asymmetric loss function we designed is that it leads to an over-estimation of the precipitation where it rains the most and under-estimation where it rains the less. Therefore, the loss function is a critical aspect to ensure that emulators suit well a given variable. The asymmetric loss function is a proposition that showed some success, but other loss functions or different strategy could be used in the same purpose in future studies.

The EURO-CORDEX matrix allowed us to study the emulator's behavior when we move out from the world corresponding to the Scenarios/GCM/RCM triplet used for training. We highlighted the robustness of the learnt function as it presents similar performances across all available simulations. The emulator notably managed to reproduce the specificity of each simulation in present climate but also in climate change signal. Indeed each simulation showed different climate change signals with different spatial patterns and variability over the domain and the emulator showed an excellent ability to reproduce this diversity. This question of transferability is essential for the potential applications it opens to the emulator. Our result tends to show that the emulator can be used to downscale various GCMs and various scenarios.

A critical point in the emulator evaluation is to ensure its good applicability to GCM simulations as it is its purpose. Because the emulator is trained in perfect model framework (i.e. with both target and input coming from the same RCM simulation), it learns only the downscaling function. Thus in GCM application it applies this function to the large scale provided by the GCM which is very likely to differ from the RCM one and so the run produced by the emulator is expected to differ from the RCM simulation driven by the same GCM. We expect the emulator to be coherent with the GCM large scale but also to include high resolution features brought by the RCM. We analysed the emulator performance over 4 GCMs and under different time horizons: we looked at some daily maps and at climatological statistics in present climate and in climate change. The conclusions are robust over all those aspects, the emulator brings a strong added-value with respect to its driving GCM that is consistent with the original RCM added-value. However, there are substantial differences between RCM and emulator maps, and it is difficult to assess if they results from large scale discrepancies between the RCM and its driving GCM, or from a misconception of the emulator. Further studies focused on given phenomenon or including other, specifically designed, simulations are probably necessary to assess if we can have a complete trust in the current version of the emulator when it is used to downscale GCM simulations.

References

Ayzel G, Scheffer T, Heistermann M (2020) RainNet v1.0: A convolutional neural network for radar-based precipitation nowcasting. Geoscientific Model Development 13(6):2631–2644, DOI 10.5194/GMD-13-2631-2020

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

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-

Berthou S, Kendon EJ, Chan SC, Ban N, Leutwyler D, Schär C, Fosser G (2020) Pan-European climate at convection-permitting scale: a model intercomparison study. Climate Dynamics 55(1-2):35–59, DOI 10.1007/s00382-018-4114-6

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

Caillaud C, Somot S, Alias A, Bernard-Bouissières I, Fumière Q, Laurantin O, Seity Y, Ducrocq V (2021) Modelling Mediterranean heavy precipitation events at climate scale: an object-oriented evaluation of the CNRM-AROME convection-permitting regional climate model. Climate Dynamics 2021 56:5 56(5):1717–1752, DOI 10.1007/S00382-020-05558-Y, URL https://link.springer.com/article/10.1007/s00382-020-05558-y

Chollet F, others (2015) Keras. \url{https://keras.io}

Coppola E, Nogherotto R, Ciarlo' JM, Giorgi F, van Meijgaard E, Kadygrov N, Iles C, Corre L, Sandstad M, Somot S, Nabat P, Vautard R, Levavasseur G, Schwingshackl C, Sillmann J, Kjellström E, Nikulin G, Aalbers E, Lenderink G, Christensen OB, Boberg F, Sørland SL, Demory ME, Bülow K, Teichmann C, Warrach-Sagi K, Wulfmeyer V (2021) Assessment of the European Climate Projections as Simulated by the Large EURO-CORDEX Regional and Global Climate Model Ensemble. Journal of Geophysical Research: Atmospheres 126(4):e2019JD032,356, DOI 10.1029/2019JD032356, URL https://onlinelibrary.wiley.com/doi/full/10.1029/2019JD032356 https://onlinelibrary.wiley.com/doi/abs/10.1029/2019JD032356 https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2019JD032356

Doury A, Somot S, Gadat S, Ribes A, Corre L (2022) Regional climate model emulator based on deep learning: concept and first evaluation of a novel hybrid downscaling approach. Climate Dynamics 1:1–29, DOI 10.1007/S00382-022-06343-9/FIGURES/16, URL https://link.springer.com/article/10.1007/s00382-022-06343-9

Ducrocq V, Nuissier O, Ricard D, Lebeaupin C, Thouvenin T (2008) A numerical study of three catastrophic precipitating events over southern France. II: Mesoscale triggering and stationarity factors. Quarterly Journal of the Royal Meteorological Society 134(630):131–145, DOI 10.1002/QJ.199, URL https://onlinelibrary.wiley.com/doi/full/10.1002/qj.199 https://onlinelibrary.wiley.com/doi/abs/10.1002/qj.199 https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.199

Evin G, Favre ACC, Hingray B (2019) Stochastic generators of multi-site daily temperature: comparison of performances in various applications. Theoretical and Applied Climatology 135(3-4):811–824, DOI 10.1007/s00704-018-2404-x, URL http://link.springer.com/10.1007/s00704-018-2404-x

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

Hawkins E, Sutton R (2009) The potential to narrow uncertainty in regional climate predictions. Bulletin of the American Meteorological Society 90(8):1095–1107, DOI 10.1175/2009BAMS2607.1

Ivušić S, Güttler I, Somot S, Guérémy JF, Horvath K, Alias A (2021) Modelling extreme precipitation over the Dinaric Alps: An evaluation of the CNRM-ALADIN regional climate model. Quarterly Journal of the Royal Meteorological Society 147(741):4425–4453, DOI 10.1002/qj.4187

Katz RW (1977) Precipitation as a Chain-Dependent Process. JApMe 16(7):671–676, DOI 10.1175/1520-0450(1977)016, URL https://ui.adsabs.harvard.edu/abs/1977JApMe..16..671K/abstract

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

Klingaman NP, Martin GM, Moise A (2017) ASoP (v1.0): A set of methods for analyzing scales of precipitation in general circulation models. Geoscientific Model Development 10(1):57–83, DOI 10.5194/gmd-10-57-2017

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. Meteorology and Atmospheric Physics 100(1-4):3–22, DOI 10.1007/s00703-008-0292-9, URL http://link.springer.com/10.1007/s00703-008-0292-9

Lucas-Picher P, Caya D, Elía R, Laprise R (2008) Investigation of regional climate models' internal variability with a ten-member ensemble of 10-year simulations over a large domain. Climate Dynamics 31(7-8):927–940, DOI 10.1007/S00382-008-0384-8/FIGURES/8, URL https://link.springer.com/article/10.1007/s00382-008-0384-8

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

Maraun D, Wetterhall F, Ireson AM, Chandler RE, Kendon EJ, Widmann M, Brienen S, Rust HW, Sauter T, Themel M, Venema VK, Chun KP, Goodess CM, Jones RG, Onof C, Vrac M, Thiele-Eich I (2010) Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. Reviews of Geophysics 48(3), DOI 10.1029/2009RG000314

Masson-Delmotte V, Zhai P, Chen Y, Goldfarb L, Gomis MI, Matthews JBR, Berger S, Huang M, Yelekçi O, Yu R, Zhou B, Lonnoy E, Maycock TK, Waterfield T, Leitzell K, Caud N (2021) Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change Edited by. DOI 10.1017/9781009157896, URL www.ipcc.ch

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

Pulkkinen S, Nerini D, Pérez Hortal AA, Velasco-Forero C, Seed A, Germann U, Foresti L (2019) Pysteps: An open-source Python library for probabilistic precipitation nowcasting (v1.0). Geoscientific Model Development 12(10):4185–4219, DOI 10.5194/gmd-12-4185-2019

Ribes A, Thao S, Vautard R, Dubuisson B, Somot S, Colin J, Planton S, Soubeyroux JM (2019) Observed increase in extreme daily rainfall in the French Mediterranean. Climate Dynamics 52(1-2):1095–1114, DOI 10.1007/S00382-018-4179-2/FIGURES/11, URL https://link.springer.com/article/10.1007/s00382-018-4179-2

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

Taranu IS, Delire C, Somot S, Alias A, Julien Boé (2022) Mechanisms behind large-scale inconsistencies between regional and global climate model-based projections over Europe. Clim Dyn [preprint] DOI 10.21203/RS.3.RS-1915811/V1, URL https://www.researchsquare.com https://www.researchsquare.com/article/rs-1915811/v1

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

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

Vergara-Temprado J, Ban N, Panosetti D, Schlemmer L, Schär C (2020) Climate models permit convection at much coarser resolutions than previously considered. Journal of Climate 33(5):1915–1933, DOI 10.1175/JCLI-D-19-0286.1

Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson J, Millman KJ, Mayorov N, Nelson AR, Jones E, Kern R, Larson E, Carey CJ, Polat I, Feng Y, Moore EW, VanderPlas J, Laxalde D, Perktold J, Cimrman R, Henriksen I, Quintero EA, Harris CR, Archibald AM, Ribeiro AH, Pedregosa F, van Mulbregt P, Vijaykumar A, Bardelli AP, Rothberg A, Hilboll A, Kloeckner A, Scopatz A, Lee A, Rokem A, Woods CN, Fulton C, Masson C, Häggström C, Fitzgerald C, Nicholson DA, Hagen DR, Pasechnik DV, Olivetti E, Martin E, Wieser E, Silva F, Lenders F, Wilhelm F, Young G, Price GA, Ingold GL, Allen GE, Lee GR, Audren H, Probst I, Dietrich JP, Silterra J, Webber JT, Slavič J, Nothman J, Buchner J, Kulick J, Schönberger JL, de Miranda Cardoso JV, Reimer J, Harrington J, Rodríguez JLC, Nunez-Iglesias J, Kuczynski J, Tritz K, Thoma M, Newville M, Kümmerer M, Bolingbroke M, Tartre M, Pak M, Smith NJ, Nowaczyk N, Shebanov N, Pavlyk O, Brodtkorb PA, Lee P, McGibbon RT, Feldbauer R, Lewis S, Tygier S, Sievert S, Vigna S, Peterson S, More S, Pudlik T, Oshima T, Pingel TJ, Robitaille TP, Spura T, Jones TR, Cera T, Leslie T, Zito T, Krauss T, Upadhyay U, Halchenko YO, Vázquez-Baeza Y (2020) SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature Methods 17(3):261–272, DOI 10.1038/s41592-019-0686-2

Vrac M, Naveau P (2007) Stochastic downscaling of precipitation: From dry events to heavy rainfalls. Water Resources Research 43(7):7402, DOI 10.1029/2006WR005308, URL https://onlinelibrary.wiley.com/doi/full/10.1029/2006WR005308 https://onlinelibrary.wiley.com/doi/abs/10.1029/2006WR005308 https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2006WR005308

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

Wang J, Liu Z, Foster I, Chang W, Kettimuthu R, Kotamarthi VR (2021) Fast and accurate learned multiresolution dynamical downscaling for precipitation. Geoscientific Model Development 14(10):6355–6372, DOI 10.5194/gmd-14-6355-2021

Wernli H, Paulat M, Hagen M, Frei C (2008) SAL - A novel quality measure for the verification of quantitative precipitation forecasts. Monthly Weather Review 136(11):4470–4487, DOI 10.1175/2008MWR2415.1

Wernli H, Hofmann C, Zimmer M (2009) Spatial forecast verification methods intercomparison project: Application of the SAL technique. Weather and Forecasting 24(6):1472–1484, DOI 10.1175/2009WAF2222271.1