https://doi.org/10.5194/gmd-2019-278
Preprint. Discussion started: 28 October 2019
© Author(s) 2019. CC BY 4.0 License.

Geoscientific Model Development Discussions

Configuration and Intercomparison of Deep Learning Neural Models for Statistical Downscaling

Jorge Baño-Medina [1], Rodrigo Manzanas [2], and José Manuel Gutiérrez [1]

[1]Santander Meteorology Group, Institute of Physics of Cantabria (CSIC-Univ. of Cantabria), Santander (Spain)
[2]Santander Meteorology Group, Dpto. de Matemática Aplicada y Ciencias de la Computación, Universidad de Cantabria, Santander (Spain)

Correspondence: Jorge Baño–Medina (bmedina@ifca.unican.es)

Abstract. Deep learning techniques (in particular convolutional neural networks, CNNs) have recently emerged as a promising approach for statistical downscaling due to their ability to learn spatial features from huge spatio-temporal datasets. However, existing studies are based on complex models, applied to particular case studies and using simple validation frameworks, which makes difficult a proper assessment of the (possible) added value offered by these techniques. As a result, these models are usually seen as black-boxes generating distrust among the climate community, particularly in climate change problems.

In this paper we undertake a comprehensive assessment of deep learning techniques for continental-scale statistical downscaling, building on the VALUE validation framework. In particular, different CNN models of increasing complexity are applied for downscaling temperature and precipitation over Europe, comparing them with a few standard benchmark methods from VALUE (linear and generalized linear models) which have been traditionally used for this purpose. Besides analyzing the adequacy of different components and topologies, we also focus on their extrapolation capability, a critical point for their possible application in climate change studies. To do this, we use a warm test period as surrogate of possible future climate conditions. Our results show that, whilst the added value of CNNs is mostly limited to the reproduction of extremes for temperature, these techniques do outperform the classic ones for the case of precipitation for most aspects considered. This overall good performance, together with the fact that they can be suitably applied to large regions (e.g. continents) without worrying about the spatial features being considered as predictors, can foster the use of statistical approaches in international initiatives such as CORDEX.

1 Introduction

The coarse spatial resolution of Global Climate Models (GCMs) is a major limitation for practical applications, since regional to local climate information is crucial for impact studies in many sectors. In order to bridge this gap, different *statistical downscaling* (SD, Maraun and Widmann, 2017) methods have been developed building on empirical relationships established between informative large-scale atmospheric variables (predictors) and local/regional variables of interest (predictands). Under the perfect prognosis approach, these relationships are learned from (daily) data using simultaneous observations for both the predictors (from a reanalysis) and predictands (historical local or gridded observations), and are subsequently applied to GCM

https://doi.org/10.5194/gmd-2019-278
Preprint. Discussion started: 28 October 2019
© Author(s) 2019. CC BY 4.0 License.

Geoscientific Model Development Discussions

simulated predictors (typically seasonal forecasts or multi-decadal climate change projections), to obtain locally downscaled values (see, e.g., Gutiérrez et al., 2013; Manzanas et al., 2018).

A number of standard perfect prognosis SD (hereafter just SD) techniques have been developed during the last two decades building on classic statistical techniques such as analogs or linear and generalized linear regression (see Gutiérrez et al., 2018, for an overview of methods). Moreover, several intercomparison studies have been conducted to understand their advantages and limitations taking into account a number of aspects such as temporal structure, extremes, or spatial consistency. In this regard, VALUE (Maraun et al., 2015) is a particularly relevant initiative which proposed an experimental validation framework for downscaling methods and conducted a comprehensive intercomparison study over Europe with over 50 contributing standard techniques (Gutiérrez et al., 2018).

Besides these classical SD methods, a number of machine learning techniques have been also adapted and applied for downscaling. For instance, the first applications of neural networks date back to the late 90s (Wilby et al., 1998; Schoof and Pryor, 2001). More recently, other alternative machine learning approaches have been applied, such as support vector machines (SVMs, Tripathi et al., 2006), random forests (Pour et al., 2016; He et al., 2016) or genetic programming (Sachindra and Kanae, 2019). There have been also a number of intercomparison studies analyzing classic and machine learning techniques (Wilby et al., 1998; Chen et al., 2010; Yang et al., 2016; Sachindra et al., 2018), with an overall consensus that no technique clearly outperforms the others and that limited added value —in terms of performance, interpretability and parsimony— is obtained with sophisticated machine learning options, particularly in the context of climate change studies.

In the last decade, machine learning has gained a renewed attention in several fields, boosted by major breakthroughs obtained with Deep Learning (DL) models (see Schmidhuber, 2015, for an overview). The advantage of DL resides in its ability to extract high-level feature representations in a hierarchical way due to its (deep) layered-structure. In particular, in spatiotemporal datasets, convolutional neural networks (CNN) have gained great attention due to its ability to learn spatial features from data (LeCun and Bengio, 1995). DL models allow to automatically treat high-dimensional problems avoiding the use of conventional feature extraction techniques (e.g. Principal Components, PCs), which are commonly used in more classic approaches (e.g., linear models and traditional fully-connected neural networks). Moreover, new efficient learning methods (e.g. batch, stochastic, and mini-batch gradient descent), regularization options (e.g. dropout), and computational frameworks (e.g. TensorFlow; see Wang et al., 2019, for an overview) have popularized the use of DL techniques, allowing to efficiently learn convolutional neural networks from (big) data avoiding overfitting. Different configurations of CNNs have proven successful in a variety of problems in several disciplines, particularly in image recognition (Schmidhuber, 2015). There are also a number of recent successful applications in climate science, including the detection of extreme weather events (Liu et al., 2016), the estimation of cyclone's intensity (Pradhan et al., 2018), the detection of atmospheric rivers (Chapman et al.), the emulation of model parameterizations (Gentine et al., 2018; Rasp et al., 2018; Larraondo et al., 2019) and full simplified models (Scher and Messori, 2019). The reader is referred to Reichstein et al. (2019) for a recent overview.

There have been some attempts to test the application of these techniques for SD, including simple illustrative examples of super-resolution approaches to recover high-resolution (precipitation) fields from low resolution counterparts with promising results (Vandal et al., 2017b; Rodrigues et al., 2018). In the context of perfect prognosis SD, deep learning applications have

https://doi.org/10.5194/gmd-2019-278
Preprint. Discussion started: 28 October 2019
© Author(s) 2019. CC BY 4.0 License.

Geoscientific Model Development Discussions

applied complex convolutional-based topologies (Vandal et al., 2017a; Pan et al., 2019), autoencoder arquitechtures (Vandal et al., 2019) and long-short term memory (LSTM) networks (Misra et al., 2018; Miao et al., 2019) over small case study areas and using simple validation frameworks, resulting in different conclusions about their performance, as compared to other standard approaches. Therefore, these complex (out-of-the-shelve in many cases) models are usually seen as black-boxes generating distrust among the climate community, particularly in climate change problems. Recently, Reichstein et al. (2019) outlined this problem and encouraged research towards the understanding of deep neural networks in climate science.

In this study we aim to shed light on this problem and perform a comprehensive evaluation of deep SD models of increasing complexity, assessing the particular role of the different elements conforming the deep neural network architecture (e.g., convolutional and fully-connected or dense layers). In particular, we use the VALUE validation framework over a continental region (Europe) and compare deep SD methods with a few standard benchmark methods best performing in the VALUE intercomparison (Gutiérrez et al., 2018). Besides this, we also focus on the extrapolation capability of the different methods, which is fundamental for climate change studies. Overall, our results show that simple deep CNNs outperform standard methods (particularly for precipitation) in most of the aspects analyzed.

The R code needed to fully replicate the experiments and results shown in this paper are freely available at GitHub (DOI: 10.5281/zenodo.3462428), together with a Jupyter notebook illustrating the use of the deep neural networks considered for climate downscaling in this work is also provided for interactive computing purposes (see the *code availability* section at the end).

2 Experimental Intercomparison Framework

2.1 Area of Study and Data

The VALUE COST Action (2012-2015) developed a framework to validate and intercompare downscaling techniques over Europe, focusing on different aspects such temporal and spatial structure and extremes (Maraun et al., 2015). The experimental framework for the first experiment (downscaling with 'perfect' reanalysis predictors) is publicly available at http://www.value-cost.eu/validation as well as the intercomparison results for over 50 different standard downscaling methods (Gutiérrez et al., 2018). Therefore, VALUE offers a unique opportunity for a rigorous and comprehensive intercomparison of different deep learning topologies for downscaling.

In particular, VALUE propose the use of twenty standard predictors from the ERA-Interim reanalysis, selected over a European domain (ranging from 36° to 72° in latitude and from -10° to 32° in longitude, with a 2° resolution) for the 30-year period 1979-2008. This predictor set is formed by five large-scale thermodynamic variables (geopotential height, zonal and meridional wind, temperature, and specific humidity) at four different vertical levels (1000,850,700 and 500 hPa) each. Daily standardized predictor values for the benchmarking linear and generalized linear techniques (see Section 2.3) are defined considering the closest ERA-Interim gridboxes to each E-OBS gridbox. However, the entire domain is used for the deep learning models, which allows to test their suitability to automatically handle high-dimensional input data, extracting relevant spatial

https://doi.org/10.5194/gmd-2019-278
Preprint. Discussion started: 28 October 2019
© Author(s) 2019. CC BY 4.0 License.

Geoscientific Model Development Discussions
Open Access
EGU

features (note that this is particularly important for continental wide applications). The left column of Figure 1 shows the climatology (and the grid) of two illustrative predictors used in this study.

The target predictands considered in this work are surface (daily) mean temperature and accumulated precipitation. Instead of the 86 representative local stations used in VALUE, we used the observational gridded dataset from E-OBS v14 (0.5° resolution). Note that this extended experiment allows for a better comparative with dynamical downscaling experiments carried out under the CORDEX initiative (Gutowski Jr. et al., 2016). The right column of Figure 1 shows the climatology of the two target predictands, temperature and precipitation.

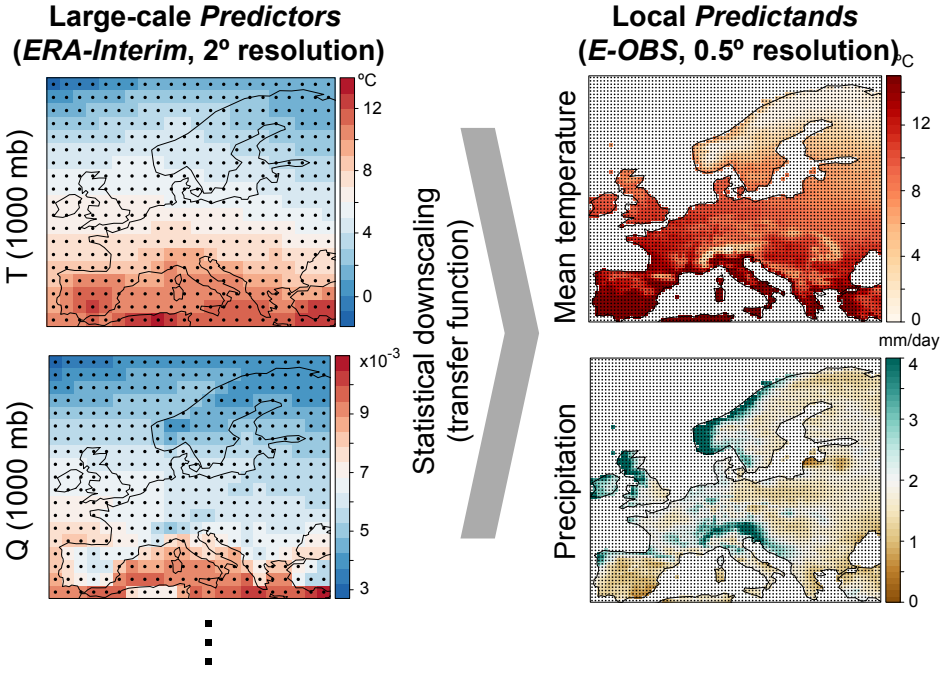

Figure 1. Climatology for (left) two typical predictors (air temperature, T, and specific humidity, Q, at 1000 mb), as given by the ERA-Interim reanalysis (2°) and (right), the observed target variables of this work, temperature and precipitation from E-OBS (0.5°). Dots indicate the center of each gridbox.

2.2 Evaluation Indices and Cross-Validation

The validation of downscaling methods is a multi-faceted problem with different aspects involved such as the representation of extremes (Hertig et al., 2019) or the temporal (Maraun et al., 2019) and spatial (Widmann et al., 2019) structure. VALUE developed a comprehensive list of indices and measures (available at the VALUE Validation Portal: http://www.value-cost.eu/validationportal) which allows to properly evaluate most of these aspects. Moreover, an implementation of these indices in an R package (VALUE, https://github.com/SantanderMetGroup/VALUE) is available for research reproducibility. In this work we consider the subset of VALUE metrics shown in Table 1 to assess the performance of the downscaling methods to reproduce

https://doi.org/10.5194/gmd-2019-278
Preprint. Discussion started: 28 October 2019
© Author(s) 2019. CC BY 4.0 License.

Geoscientific Model Development Discussions

105 the observations. Note that different metrics are considered for temperature and precipitation. The bias measures the average forecast error. For temperature, biases are given as absolute differences (in °C), whereas for precipitation they are expressed as relative differences with respect to the observed value (in %). Note that, beyond the bias in the mean, we also assess the bias in extreme percentiles, in particular the percentile 2 (P2, for temperature) and the 98 (P98, for both temperature and precipitation). We also consider the Root Mean Squared Error (RMSE), which measures the average magnitude of the forecast errors, weighted
110 according to the square of the error; in the case of precipitation, this metric is applied only for observation-prediction pairs for which the observed vale corresponded to a rainy (rainfall $>$ 1 mm) day. To evaluate how close the predictions follow the observations, we also assess correlation, in particular the Pearson coefficient for temperature and the Spearman rank one (adequate for non-gaussian variables) for precipitation; for the particular case of temperature, the seasonal cycle is removed from both observations and predictions in order to avoid its effect on the correlation. This is done by applying a 31-day width
115 moving window centered on each day in the time-series. For this variable we also consider the ratio of standard deviations, i.e., that of the predictions divided by that of the observations. Finally, to evaluate how well the probabilistic predictions of rain occurrence discriminate the binary event rain/no rain, we consider the ROC Skill Score (ROCSS) (see, e.g. Manzanas et al., 2014), which is based on the area under the ROC curve (see Kharin and Zwiers, 2003, for details).

Description	**Variable**	**Units**
Bias (for the mean)	temp., precip.	$^{\circ}C$, $\%$
Bias (for percentile 2)	temp.	$^{\circ}C$
Bias (for percentile 98)	temp., precip.	$^{\circ}C$, $\%$
Root Mean Square Error	temp., precip.	$^{\circ}C, mm/day$
Ratio of standard deviations	temp.	-
Pearson correlation	temp.	-
Spearman correlation	precip.	-
ROC Skill Score	precip.	-

Table 1. Subset of VALUE metrics used in this study to validate the different downscaling methods considered (see Table 2). The symbol '-' denotes adimensionality. For temperature and precipitation the biases are absolute (adimensional), respectively.

The VALUE framework builds on a cross-validation approach in which the 30-year period of study (1979-2008) is chrono-
120 logically split into five consecutive folds. We are particularly interested in analyzing the out-of-sample extrapolation capabilities of the deep SD models. Therefore, following the recommendations of Riley (2019, *"the question you want to answer should affect the way you split your data"*), we focus on the last fold, for which warmer conditions have been observed. Therefore, in this work we apply a simplified hold-out approach using for validation the period 2003-2008, and training the models using the remaining years (1979-2002). Figure 2 shows the climatology of the train period for both temperature and
125 precipitation (top and bottom panel, respectively), as well as the mean differences between the test and the train periods (taken the latter as reference). For temperature, warmer conditions are observed in the test period —over $0.7°$ for both mean values

https://doi.org/10.5194/gmd-2019-278
Preprint. Discussion started: 28 October 2019
© Author(s) 2019. CC BY 4.0 License.

Geoscientific Model Development Discussions

and extremes,— being especially significant for the 2nd percentile (cold days), for which temperatures increase up to 2° in northern Europe, compared with the training period. This allows us to estimate the extrapolation capabilities of the different methods, which is particularly relevant for climate change studies.

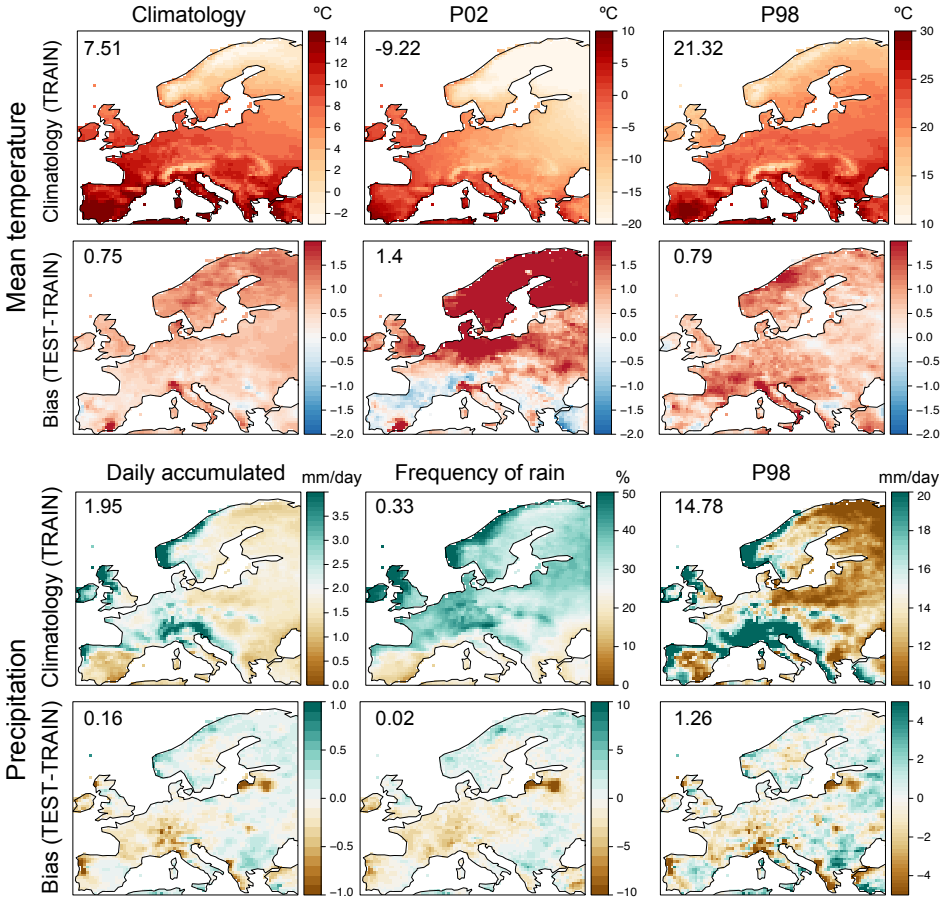

Figure 2. Top panel, top row: E-OBS climatology for the mean value, the P02 and the P98 of temperature in the train period (1979-2002). Top panel, bottom row: Mean difference between the test and train periods (the latter taken as reference) for the different quantities shown in the top row. Bottom panel: As the top panel, but for precipitation. In this case, the mean value, the frequency of rainy days and the P98 are shown. In all cases, the numbers within the panels indicate the spatial mean values.

130 Importantly, note that the differences between the test and train periods in Figure 2 reveal some inconsistencies in the dataset for both temperature (Southern Iberia and Alps) and precipitation (Northeastern Iberia and the Baltic states). This may be an artifact due to changes or interruptions in the national station networks used to construct E-OBS and may not correspond to a real change in the dataset. This will be taken into account when analyzing the results in Section 4.

https://doi.org/10.5194/gmd-2019-278
Preprint. Discussion started: 28 October 2019
© Author(s) 2019. CC BY 4.0 License.

Geoscientific Model Development Discussions

2.3 Classic Benchmark Models

We use as benchmark some state-of-the-art standard techniques which ranked among the top in the VALUE intercomparisson experiment. In particular, multiple linear and generalized linear regression models (hereafter referred to as GLM) exhibited good overall performance for temperature and precipitation, respectively (Gutiérrez et al., 2018). Here, we consider the version of these methods described in Bedia et al. (2019) which use the predictor values in the four gridboxes closest to the target location. This choice is a good compromise between feeding the model with full spatial information (all gridboxes, which is problematic due to the resulting high-dimensionality) and insufficient spatial representation when considering a single gridbox. For the sake of completeness we also illustrate the results obtained with a single gridbox, in order to provide an estimate of the added value of extending the spatial information considered for the different variables. These benchmark models are denoted GLM1 and GLM4 for one and four gridboxes, respectively (first two rows in Table 2).

In the case of temperature a single multiple regression model (i.e. GLM with Gaussian family) is used, whereas for precipitation two different GLMs are applied, one for the occurrence ($precipitation > 1mm$) and one for the amount of precipitation, using binomial and Gamma families with logarithmic link, respectively (see, e.g., Manzanas et al., 2015). In this case, the values from the two models are multiplied to obtain the final prediction or precipitation, although occurrence and amount are also evaluated separately.

Model	Architecture	Rationale
GLM1	20-1 ($\times$ 3258)	Simplest linear local model for benchmarking
GLM4	80-1 ($\times$ 3258)	Increasing the predictor's spatial domain
CNN-LM	20-**50**-**25**-**1**-3258	Using convolutions to automatically obtain meaningful spatial predictors
CNN1	20-**50**-**25**-**1**-3258	Testing the added value of CNN non-linearity
CNN10	20-**50**-**25**-**10**-3258	Increasing the complexity of last CNN features layer
CNN-PR	20-**10**-**25**-**50**-3258	Using standard topologies from pattern recognition
CNNdense	20-**50**-**25**-**10**-50-50-3258	Using complex dense CNN models

Table 2. Description of the deep learning architectures intercompared in this study, together with the two benchmark methods: GLM1 and GLM4 (these models are trained separately for each of the 3258 land-only gridboxes in E-OBS). Convolutional layers are indicated with boldfaced numbers. The numbers indicating the architecture correspond to the number of neurons in the different layers (in bold for convolutional layers).

3 Deep Neural Networks

Despite the success of deep learning in many fields, these models are still seen as black boxes generating distrust among the climate community, particularly in climate change problems, as their extrapolation capability has not been assessed yet. Recently, Reichstein et al. (2019) outlined this problem and encouraged research towards the understanding of deep neural

https://doi.org/10.5194/gmd-2019-278
Preprint. Discussion started: 28 October 2019
© Author(s) 2019. CC BY 4.0 License.

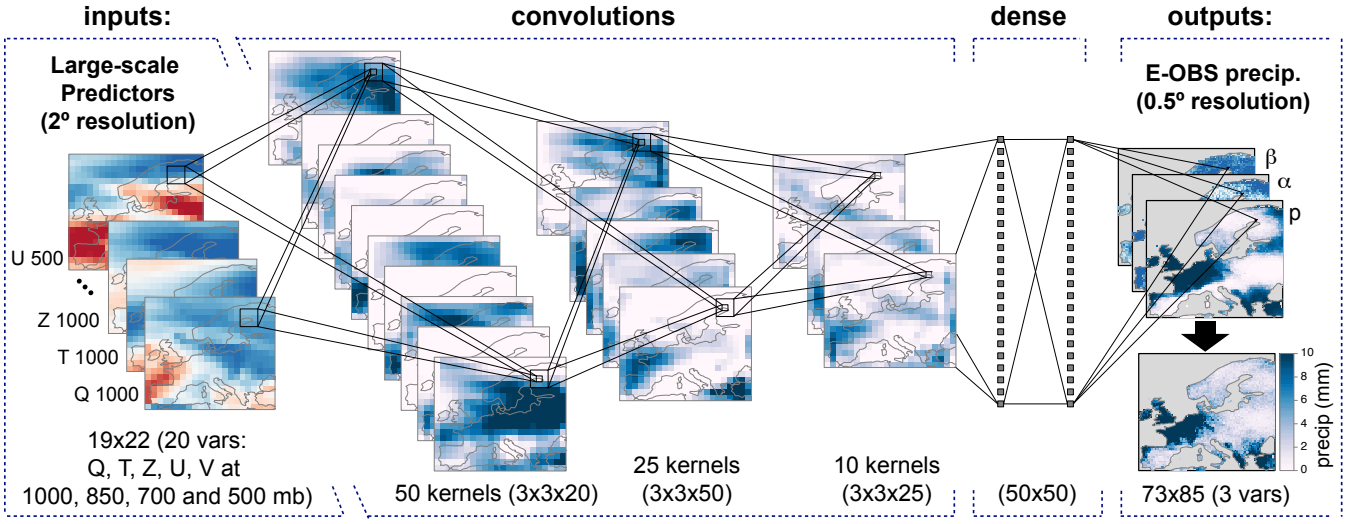

Figure 3. Scheme of the convoloutional neural network architecture used in this work to downscale European (E-OBS 0.5° grid) precipitation based on five coarse (2°) large-scale standard predictors (at four pressure levels). The network includes a first block of three convolutional layers with 50, 25 and 10 ($3 \times 3 \times \#inputs$) kernels, respectively, followed by two fully-connected (dense) layers with 50 neurons each. The output is modeled through a mixed binomial-lognormal distribution and the corresponding parameters are estimated by the network, obtaining precipitation as a final product, either deterministically (the expected value), or stochastically (generating a random value from the predicted distribution). The output layer is activated linearly except for the neurons associated to the parameter p which present sigmoidal activation functions.

networks in climate science. In this study we aim to shed light on the particular role of the different elements conforming the deep neural network architecture (e.g., convolutional and fully-connected or dense layers). To do this, we build and evaluate deep SD models of increasing complexity, starting with a simple benchmark linear model (GLM) and adding additional "deep" components, in particular convolution and dense layers, as shown schematically in Figure 3.

The basic neural network topology relies on feed-forward networks composed of several layers of non-linear neurons which are fully-connected between consecutive layers, from the input to the output (these are commonly referred to as "dense" networks; see Figure 3). Each of these connections is characterized by a weight which is learnt from data (e.g. the two layers of 50 neurons each in Figure 3 result in a total of 50×50 internal weights, besides the input and output connections). Differently to standard dense networks (whose input is directly the raw predictor data), convolutional networks generate data-driven spatial features to feed the dense network. These layers convolute the raw gridded predictors using 3D kernels (variable, latitude and longitude), considering a neighbourhood of the corresponding gridbox (3×3 in this work) in the previous layer (see Figure 3). Instead of fully-connecting the subsequent layers, kernel weights are shared across regions, resulting into a drastically reduction in the degrees of freedom of the network. Due to these convolutional operations, layers consists on filter maps, which can be interpreted as the spatial representation of the feature learned by the kernel. This is crucial when working with datasets with an underlying spatial structure. To maximize the performance of convolutional topologies, it is necessary to select an adequate

https://doi.org/10.5194/gmd-2019-278
Preprint. Discussion started: 28 October 2019
© Author(s) 2019. CC BY 4.0 License.

Geoscientific Model Development Discussions

number of layers, filter maps and kernel's size, which has been done here following an empirical screening procedure (not shown). Besides the different deep learning architectures, we also analyze the effect of basic elements such as the activation function or the layer configuration.

All the deep models used in this work have been trained using daily data for both predictors and predictand. For temperature, they estimate the mean of a gaussian distribution by minimizing the mean squared error. For precipitation, due to its mixed discrete-continuous nature, the network optimizes the negative log-likelihood of a Bernouilli-Gamma distribution following the approach previously introduced by Cannon (2008). In particular, the network estimates the parameter p (i.e., probability of rain) of the Bernouilli distribution for rain occurrence, and the parameters α (shape) and β (scale) of the Gamma rain amount model, as illustrated in the output layer of Figure 3. The final rainfall for a given day i, r_i, is then be inferred as the expected value of a gamma distribution, given by $r_i = \alpha_i * \beta_i$.

The first two methods analyzed in this work are the two benchmark GLM models (i.e. multiple linear regression for temperature and Bernoulli + Gamma GLM for precipitation) considering local predictors at the nearest (4 nearest) neighbouring gridboxes. They are labelled as GLM1 (GLM4) in Table 2. Selecting information only from the local gridboxes could be a limitation for the methods and, therefore, some GLM applications consider spatial features as predictors instead, such as Principal Components from the Empirical Orthogonal Functions (EOFs) (Gutiérrez et al., 2018). Convolutional networks are automatic feature extraction techniques which learn spatial features of increasing complexity from data in a hierarchical way, due to its (deep) layered-structure (LeCun and Bengio, 1995). Therefore, as third model we test the potential of convolutional layers for spatial feature extraction by considering a linear convolutional neural network with three layers (with 50, 25 and 1 features each) and linear activation functions (CNN-LM in Table 2). The benefits of non-linearity are tested considering the same convolutional network CNN-LM, but with non-linear (ReLu) activation functions in the hidden layers, making the model non-linear (CNN1 in Table 2). Moreover, the role of the number of convolutional features in the final layer is tested considering a non-linear convolutional model, but with 10 feature maps (coded as CNN10). Note that the previous models are built using a decreasing number of features in the subsequent convolutional layers. However, the approach usually used in computer vision for pattern recognition tasks is the contrary (i.e. the number of convolutional maps increases along the network). Therefore, we also tested this type of architecture considering a convolutional neural network with an increasing number of maps, (10, 25 and 50, labelled as CNN-PR). Finally, a general deep neural network is formed by including a dense (feed-forward) network as an additional block taking input from the convolutional layer (see Figure 3). This is the typical topology considered in practical applications, which combines both feature extraction and non-linear modeling capabilities (denoted as CNNdense in Table 2).

All deep learning models listed in Table 2 have been tested with and without padding (padding maintains the original resolution of the predictors throughout the convolutional layers, avoiding the loss of information that may occur near the borders of the domain), keeping in each case the best results for the final intercomparison. Padding was found to be useful only when the amount of feature maps in the last layer was small, so padding is only used for CNN1 model.

https://doi.org/10.5194/gmd-2019-278
Preprint. Discussion started: 28 October 2019
© Author(s) 2019. CC BY 4.0 License.

Geoscientific Model Development Discussions

200 # 4 Results

In this section we intercompare and discuss the performance of the different models shown in Table 2 for temperature (Section 4.1) and precipitation (Section 4.2).

4.1 Temperature

Figure 4 shows the validation results obtained for temperature in terms of the different metrics explained in Section 2.2. Each
205 panel contains 7 boxplots, one for each of the methods considered (Table 2), representing the spread of the results along the entire E-OBS grid. In particular, the gray boxes corresponds to the 25-75 percentile range, whereas the whiskers cover the 10-90 percentage range. The horizontal red line plots the median value obtained from the GLM4 method, which is considered as benchmark.

In general, all methods provide quite satisfactory results, with low biases and RMSE (panels a, c, e and f), a realistic variabil-
210 ity (d) and very high correlation values (after removing the annual cycle from the series). Among the classic linear methods, GLM4 clearly outperforms GLM1, which highlights the fact that including predictor information representative of a wider area around the target point helps to better describe the synoptic features determining the local temperature. However, most of the this local variability seems to be explained by linear predictor-predictand relationships, as both GLM4 and CNN-LM provide similar results to more sophisticated neural networks which account for non-linearity (regardless of their architecture).
215 Nevertheless, the biases provided by CNN1, CNN10, CNN-PR and CNNdense for P02 and P98 are lower than those obtained from the GLM1, GLM4 and CNN-LM (e, f), which suggests that non-linearity add some value for the prediction of extremes. Besides, CNN10 (identified with a darker gray) provides the lowest RMSE and the highest correlations, being overall the best method.

For a better spatial interpretation of these results, Figure 5 shows maps for each metric (in columns) for GLM1, GLM4
220 and CNN10 (in rows), representing the two initial benchmarking methods and the best-performing alternative found. It is important to highlight that the three methods present very little (mean) biases along the entire continent, which suggests their good extrapolation capability, and therefore, their potential suitability for climate change studies (recall that the anomalously warm test period that has been selected for this work may serve as a surrogate of the warmer conditions that are expected due to climate change).

225 Due to its strong local dependency, GLM1 leads to patchy (discontinuous) spatial patterns, something which is solved by GLM4 —including local predictor information representative of a wider area around the target point provides smother, continuous patterns.— Beyond this particular aspect, the improvement of GLM4 over GLM1 is evident for RMSE and correlation, and to a lesser extent also for the bias in P98. However, the best results are found for the CNN10 method, which improves all the validation metrics considered, and in particular, the bias for P2.

230 As already pointed out in Section 2.1, note that the anomalous results found for Southern Iberia could likely be related to issues in the E-OBS dataset.

https://doi.org/10.5194/gmd-2019-278
Preprint. Discussion started: 28 October 2019
© Author(s) 2019. CC BY 4.0 License.

Geoscientific Model Development
Discussions
Open Access
EGU

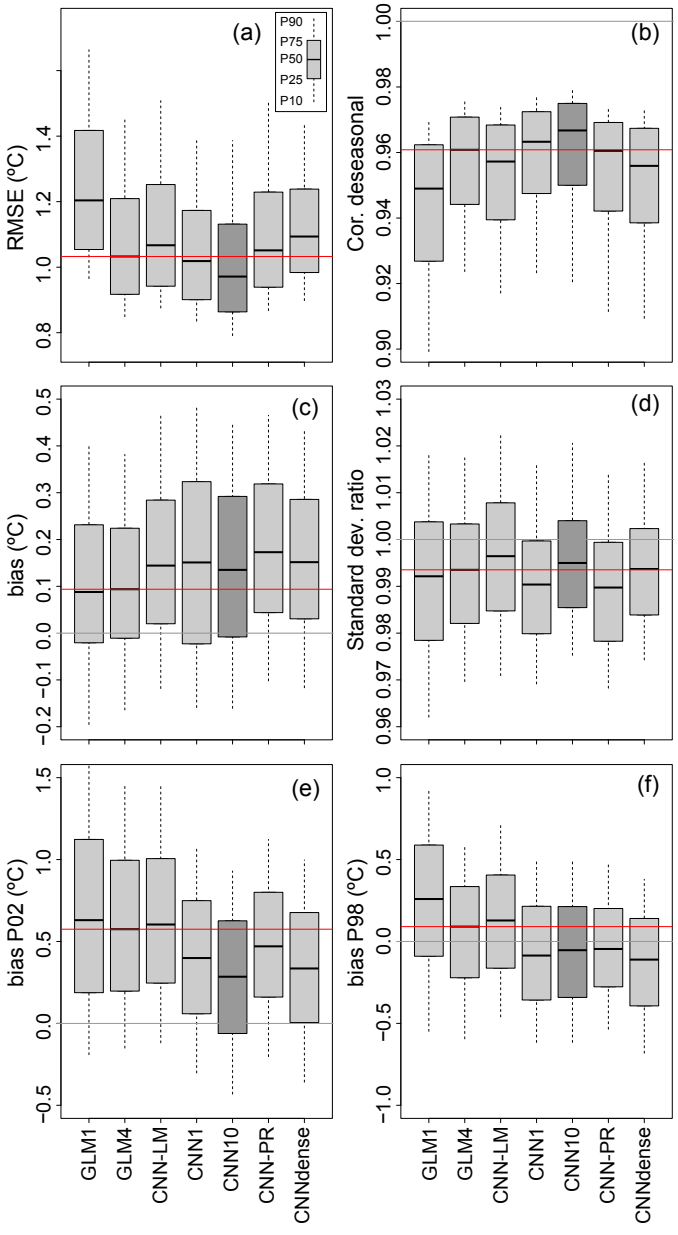

Figure 4. Validation results obtained for temperature. Each panel (corresponding to a particular metric) contains 7 boxplots, one for each of the methods tested, which represents the spread of the results along the entire E-OBS grid (the gray boxes corresponds to the 25-75 percentile range, whereas the whiskers cover the 10-90 percentage range). The horizontal red line plots the median value obtained from the GLM4 method, which is considered as benchmark, whereas the gray one indicates the 'perfect' value for each metric. The dark shaded box indicates the best performing method (CNN10 in this case).

https://doi.org/10.5194/gmd-2019-278
Preprint. Discussion started: 28 October 2019
© Author(s) 2019. CC BY 4.0 License.

Geoscientific Model Development Discussions

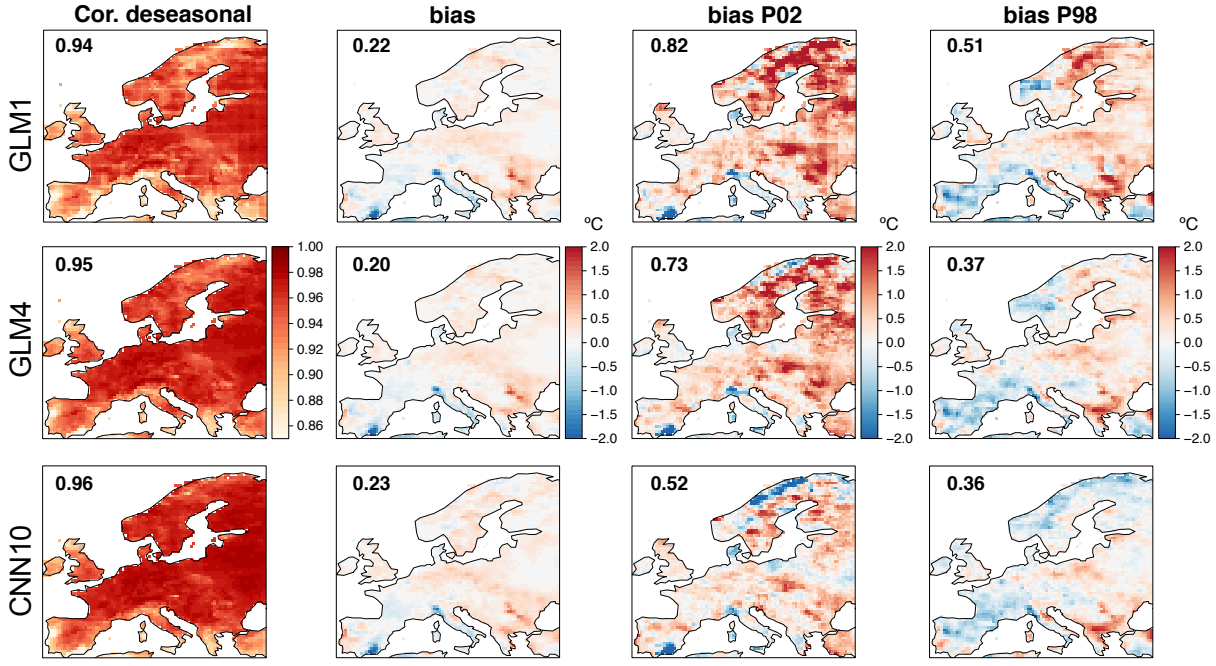

Figure 5. Maps showing the spatial results obtained in terms of the different metrics considered for temperature (in columns) for the two benchmarking versions of GLM (top and middle row) and the best-performing method, the CNN10 (bottom row). The numbers within the panels show the spatial mean absolute values (to avoid error compensation).

4.2 Precipitation

Figure 6 is similar to Figure 4, but for the case of precipitation (note that the validation metrics considered for this variable differ). Similarly to the case of temperature, GLM4 performs notably better than GLM1, in particular for the ROCSS (panel a), the RMSE (b), and the correlation (c). Nevertheless, with the exception of CNN-LM and CNN-PR, convolutional networks yield in general better results than GLM4. Differently to the case of temperature, this indicates that accounting for non-linear predictor-predictand relationships is key to better describe precipitation, especially in terms of ROCSS and correlation. Moreover, the standard architecture for pattern recognition (CNN-PR), is not suitable for this prediction problem. In terms of errors (RMSE and the different biases considered), all convolutional networks perform similarly, exhibiting very little, centered around zero biases for the mean. With respect to the P98, the slight underestimation shown by deterministic configurations (e) can be solved by stochastically sampling from the predicted Gamma distribution (f), but at the cost of losing part of the temporal and spatial correlation achieved by deterministic set-ups (not shown). Note that, as usual, the correlations found for all methods are much lower than those obtained for temperature, with the CNN-LM method yielding similar values to those obtained with GLM4. This suggests that choosing the 4 nearest gridboxes as predictors allows to capture the key spatial features that affect the downscaling of precipitation with linear models (at least over Europe). Differently to the case of temperature, note also that there is not a significant change in the climatological mean between the train and test periods for precipitation (see Figure 2),

https://doi.org/10.5194/gmd-2019-278
Preprint. Discussion started: 28 October 2019
© Author(s) 2019. CC BY 4.0 License.

Geoscientific Model Development Discussions

so the particular train/test partition considered in this work does not allow to carry out a proper assessment of the extrapolation capability of the different methods.

Overall, the best results are obtained for the CNN1 (marked with a darker gray) and CNNdense, which differ from CNN10 in the amount of neurons placed in the last layer. This suggests that whilst 1 feature map was a little restrictive for the case of temperature, 10 maps oversized the network for precipitation, worsening its generalization capability for this variable.

Figure 7 is the equivalent to Figure 5 but for precipitation. Again, the best-performing method (CNN1 in this case; bottom row) is shown, together with the two benchmarking versions of GLM (top and middle rows). In all cases, the deterministic implementation is considered. As for temperature, GLM4 provides better results than GLM1 for all metrics, being the spatial pattern of improvement rather uniform in all cases. Likewise, CNN1 outperforms GLM4 for all metrics and regions, especially over Central and Northern Europe. These results suggest the suitability of convolutional neural networks to downscale precipitation, which may be a consequence of their ability to automatically extract the important spatial features determining the local climate, as well as to efficiently model the non-linearity established between local precipitation and the large-scale atmospheric circulation.

Finally, notice that the anomalous results found over north-eastern Iberia and the Baltic states might be due to issues in the E-OBS dataset. Nonetheless, particularly bad results are also found over the Greek peninsula (especially for the mean bias), for which we do not envisage a clear explanation.

5 Conclusions

Deep learning techniques have gained increasing attention due to the promising results obtained in various disciplines. In particular, convolutional neural networks (CNN) have recently emerged as a promising approach for statistical downscaling in climate due to their ability to learn spatial features from huge spatio-temporal datasets, which would allow for an efficient application of statistical downscaling to large domains (e.g. continents). Within this context, there have been a number of intercomparison studies analyzing classic and machine learning (including CNN) techniques. However, these studies are based on different case studies and use different validation frameworks, which makes difficult a proper assessment of the (possible) added value offered by CNNs and, in some cases, offer contradictory results (e.g. Vandal et al., 2019; Sachindra et al., 2018).

In this paper we build on a comprehensive framework for validating statistical downscaling techniques (the VALUE validation framework) and evaluate the performance of different CNN models of increasing complexity for downscaling temperature and precipitation over Europe, comparing them with a few standard benchmark methods from VALUE (linear and generalized linear models). Besides analyzing the adequacy of different network architectures, we also focus on their extrapolation capability, a critical point for their possible application in climate change studies, and use a warm test period as surrogate of possible future climate conditions.

Regarding the classic (generalized) linear methods, our results show that using predictor data in several gridboxes helps to better describe the synoptic features determining the local climate, yielding thus better predictions both for temperature and precipitation. Besides, for the case of temperature, we find that the added value of non-linear CNNs (regardless of the architec-

https://doi.org/10.5194/gmd-2019-278
Preprint. Discussion started: 28 October 2019
© Author(s) 2019. CC BY 4.0 License.

Geoscientific Model Development Discussions

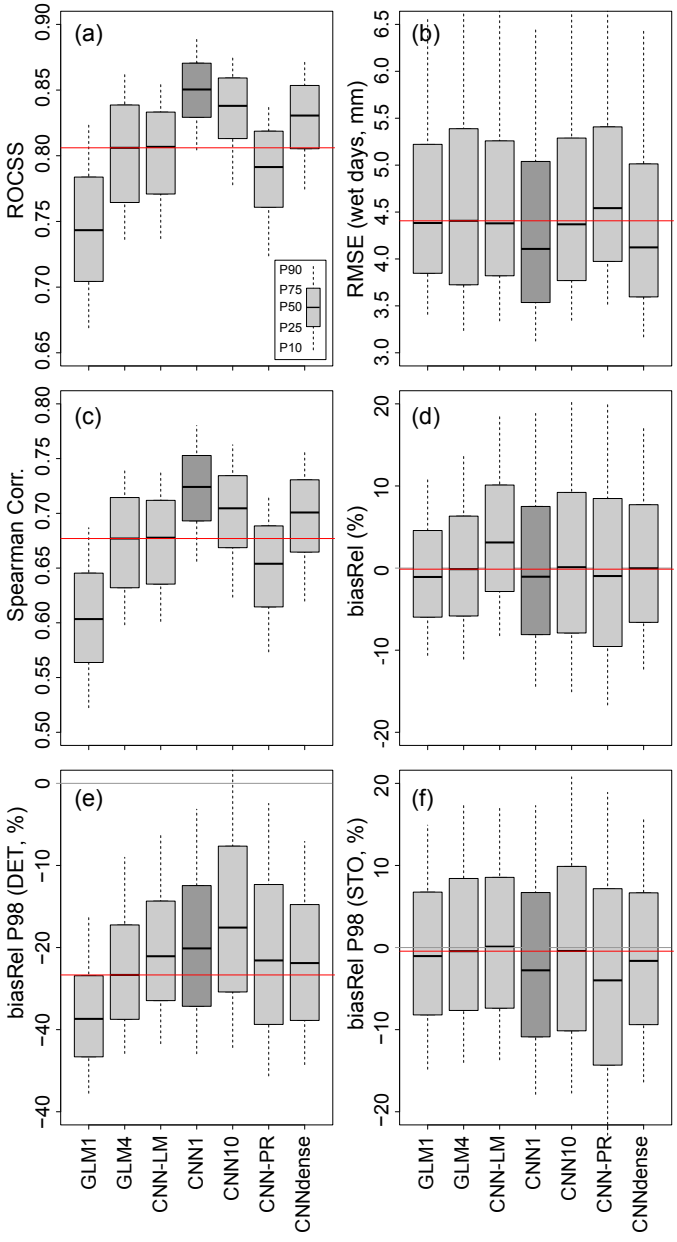

Figure 6. As Figure 4, but for precipitation.

ture considered) is limited to the reproduction of extremes, as most of the local variability of this variable is well captured with classic linear methods. Moreover, for precipitation, CNNs yield in general better results than standard generalized linear methods, which may reflect the ability of these techniques to automatically extract the important spatial features determining the

https://doi.org/10.5194/gmd-2019-278
Preprint. Discussion started: 28 October 2019
© Author(s) 2019. CC BY 4.0 License.

Geoscientific Model Development Discussions
Open Access
EGU

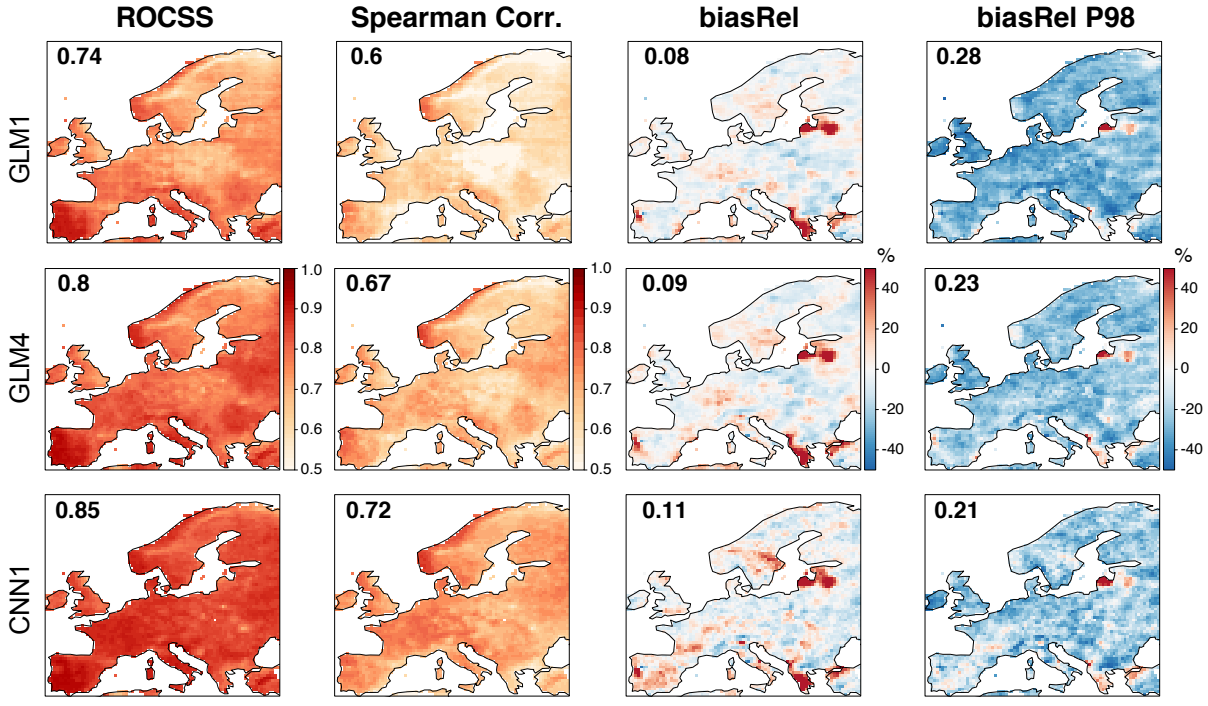

Figure 7. As Figure 5 but for precipitation. In this case, CNN1 is taken as best-performing method (bottom row). The numbers within the panels show the spatial mean absolute values (to avoid error compensation).

local climate, as well as to efficiently model the non-linearity established between this variable and the large-scale atmospheric circulation.

Note that the overall good results found for the CNNs tested here, together with the fact that they can be suitably applied to large domains without worrying for the spatial features being considered as predictors, can foster the use of statistical approaches in the framework of international initiatives for downscaling such as CORDEX, which has traditionally relied on dynamical simulations to-date.

Code availability. For the purpose of research transparency, we provide the full code needed to reproduce the experiments presented in this paper, which can be found in the Santander Meteorology Group GitHub (https://github.com/SantanderMetGroup/DeepDownscaling) and in Zenodo (Baño Medina et al., 2019). The code builds on the open-source `climate4R` (Iturbide et al., 2019) and `keras` (Chollet et al., 2015) `R` frameworks, for the benchmark and the CNN models, respectively. The former is an open `R` framework for climate data access, processing (e.g. collocation, binding, and subsetting), visualization, and downscaling, allowing for a straightforward application of wide range of downscaling methods (Bedia et al., 2019). The latter is a popular `R` framework for deep learning which builds on `TensorFlow`.

Moreover, the validation of the methods has been carried out with the package `R_VALUE` and its climate4R wrapper `climate4R.value` (https://github.com/SantanderMetGroup/climate4R.value), which enables a direct application of the VALUE validation metrics.

https://doi.org/10.5194/gmd-2019-278
Preprint. Discussion started: 28 October 2019
© Author(s) 2019. CC BY 4.0 License.

Geoscientific Model Development Discussions

Author contributions. Baño-Medina, J. and Gutiérrez, J.M. conceived the study; Baño-Medina, J. implemented the code to develop the convolutional neural networks, and generated the results of the paper; all authors analyzed the results and wrote the manuscript; Baño-Medina, J. and Manzanas, R. prepared the code and notebooks for reproducibility.

Competing interests. No competing interests are present.

Acknowledgements. The authors acknowledge the funding provided by the project MULTI-SDM (CGL2015-66583-R, MINECO/FEDER). We also acknowledge the E-OBS dataset from the EU-FP6 project UERRA (http://www.uerra.eu) and the Copernicus Climate Change Service, and the data providers in the ECA&D project (https://www.ecad.eu).

https://doi.org/10.5194/gmd-2019-278
Preprint. Discussion started: 28 October 2019
© Author(s) 2019. CC BY 4.0 License.

Geoscientific Model Development Discussions

References

Baño Medina, J., Manzanas, R., and Gutiérrez, J. M.: SantanderMetGroup/DeepDownscaling: 2019_deepDownscaling_GMD, https://doi.org/10.5281/zenodo.3462428, 2019.

Bedia, J., Baño Medina, J., Legasa, M. N., Iturbide, M., Manzanas, R., Herrera, S., Casanueva, A., San-Martín, D., Cofiño, A. S., and Gutiérrez, J. M.: Statistical downscaling with the downscaleR package: Contribution to the VALUE intercomparison experiment, Geoscientific Model Development Discussions, 2019, 1–33, https://doi.org/10.5194/gmd-2019-224, 2019.

Cannon, A. J.: Probabilistic Multisite Precipitation Downscaling by an Expanded Bernoulli-Gamma Density Network, Journal of Hydrometeorology, 9, 1284–1300, https://doi.org/10.1175/2008JHM960.1, https://journals.ametsoc.org/doi/full/10.1175/2008JHM960.1, 2008.

Chapman, W. E., Subramanian, A. C., Monache, L. D., Xie, S. P., and Ralph, F. M.: Improving Atmospheric River Forecasts With Machine Learning, Geophysical Research Letters, 0, https://doi.org/10.1029/2019GL083662.

Chen, S.-T., Yu, P.-S., and Tang, Y.-H.: Statistical downscaling of daily precipitation using support vector machines and multivariate analysis, Journal of Hydrology, 385, 13–22, https://doi.org/10.1016/j.jhydrol.2010.01.021, http://www.sciencedirect.com/science/article/pii/S0022169410000533, 2010.

Chollet, F. et al.: Keras, https://keras.io, 2015.

Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G., and Yacalis, G.: Could Machine Learning Break the Convection Parameterization Deadlock?, Geophysical Research Letters, 45, 5742–5751, https://doi.org/10.1029/2018GL078202, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018GL078202, 2018.

Gutiérrez, J. M., San-Martín, D., Brands, S., Manzanas, R., and Herrera, S.: Reassessing Statistical Downscaling Techniques for Their Robust Application under Climate Change Conditions, Journal of Climate, 26, 171–188, https://doi.org/10.1175/JCLI-D-11-00687.1, 2013.

Gutiérrez, J. M., 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, M. J., Ramos, P., Zerenner, T., Turco, M., Bosshard, T., Bartholy, J., Pongracz, R., Keller, D. E., Fischer, A. M., Cardoso, R. M., Soares, P. M. M., Czernecki, B., and Pagé, C.: 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, 0, https://doi.org/10.1002/joc.5462, https://rmets.onlinelibrary.wiley.com/doi/10.1002/joc.5462, 2018.

Gutowski Jr., W. J., Giorgi, F., Timbal, B., Frigon, A., Jacob, D., Kang, H.-S., Raghavan, K., Lee, B., Lennard, C., Nikulin, G., O'Rourke, E., Rixen, M., Solman, S., Stephenson, T., and Tangang, F.: WCRP COordinated Regional Downscaling EXperiment (CORDEX): A diagnostic MIP for CMIP6, Geosci. Model Dev., 9, 4087–4095, https://doi.org/10.5194/gmd-9-4087-2016, https://www.geosci-model-dev.net/9/4087/2016/, 2016.

He, X., Chaney, N. W., Schleiss, M., and Sheffield, J.: Spatial downscaling of precipitation using adaptable random forests, Water Resources Research, 52, 8217–8237, https://doi.org/10.1002/2016WR019034, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016WR019034, 2016.

Hertig, E., Maraun, D., Bartholy, J., Pongracz, R., Vrac, M., Mares, I., Gutiérrez, J. M., Wibig, J., Casanueva, A., and Soares, P. M. M.: Comparison of statistical downscaling methods with respect to extreme events over Europe: Validation results from the perfect predictor experiment of the COST Action VALUE, International Journal of Climatology, 39, 3846–3867, https://doi.org/10.1002/joc.5469, https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.5469, 2019.

https://doi.org/10.5194/gmd-2019-278
Preprint. Discussion started: 28 October 2019
© Author(s) 2019. CC BY 4.0 License.

Iturbide, M., Bedia, J., Herrera, S., Baño-Medina, J., Fernández, J., Frías, M. D., Manzanas, R., San-Martín, D., Cimadevilla, E., Cofiño, A. S., and Gutiérrez, J. M.: The R-based climate4R open framework for reproducible climate data access and post-processing, Environmental Modelling & Software, 111, 42–54, https://doi.org/10.1016/j.envsoft.2018.09.009, http://www.sciencedirect.com/science/article/pii/S1364815218303049, 2019.

Kharin, V. V. and Zwiers, F. W.: On the ROC score of probability forecasts, Journal of Climate, 16, 4145–4150, https://doi.org/10.1175/1520-0442(2003)016<4145:OTRSOP>2.0.CO;2, 2003.

Larraondo, P. R., Renzullo, L. J., Inza, I., and Lozano, J. A.: A data-driven approach to precipitation parameterizations using convolutional encoder-decoder neural networks, arXiv [physics], http://arxiv.org/abs/1903.10274, arXiv: 1903.10274, 2019.

LeCun, Y. and Bengio, Y.: Convolutional networks for images, speech, and time series, The handbook of brain theory and neural networks, 3361, 1995, 1995.

Liu, Y., Racah, E., Prabhat, Correa, J., Khosrowshahi, A., Lavers, D., Kunkel, K., Wehner, M., and Collins, W.: Application of Deep Convolutional Neural Networks for Detecting Extreme Weather in Climate Datasets, arXiv:1605.01156 [cs], http://arxiv.org/abs/1605.01156, arXiv: 1605.01156, 2016.

Manzanas, R., Frías, M. D., Cofiño, A. S., and Gutiérrez, J. M.: Validation of 40 year multimodel seasonal precipitation forecasts: The role of ENSO on the global skill, Journal of Geophysical Research: Atmospheres, 119, 1708–1719, https://doi.org/10.1002/2013JD020680, 2014.

Manzanas, R., Brands, S., San-Martín, D., Lucero, A., Limbo, C., and Gutiérrez, J. M.: Statistical downscaling in the tropics can be sensitive to reanalysis choice: A case study for precipitation in the Philippines, Journal of Climate, 28, 4171–4184, https://doi.org/10.1175/JCLI-D-14-00331.1, 2015.

Manzanas, R., Lucero, A., Weisheimer, A., and Gutiérrez, J. M.: Can bias correction and statistical downscaling methods improve the skill of seasonal precipitation forecasts?, Climate Dynamics, 50, 1161–1176, https://doi.org/10.1007/s00382-017-3668-z, 2018.

Maraun, D. and Widmann, M.: Statistical Downscaling and Bias Correction for Climate Research by Douglas Maraun, Cambridge University Press, 2017.

Maraun, D., Widmann, M., Gutiérrez, J. M., Kotlarski, S., Chandler, R. E., Hertig, E., Wibig, J., Huth, R., and Wilcke, R. A.: VALUE: A framework to validate downscaling approaches for climate change studies, Earth's Future, 3, 2014EF000259, https://doi.org/10.1002/2014EF000259, http://onlinelibrary.wiley.com/doi/10.1002/2014EF000259/abstract, 2015.

Maraun, D., Huth, R., Gutiérrez, J. M., Martín, D. S., Dubrovsky, M., Fischer, A., Hertig, E., Soares, P. M. M., Bartholy, J., Pongracz, R., Widmann, M., Casado, M. J., Ramos, P., and Bedia, J.: The VALUE perfect predictor experiment: Evaluation of temporal variability, International Journal of Climatology, 39, 3786–3818, https://doi.org/10.1002/joc.5222, 2019.

Miao, Q., Pan, B., Wang, H., Hsu, K., and Sorooshian, S.: Improving Monsoon Precipitation Prediction Using Combined Convolutional and Long Short Term Memory Neural Network, Water, 11, 977, https://doi.org/10.3390/w11050977, https://www.mdpi.com/2073-4441/11/5/977, 2019.

Misra, S., Sarkar, S., and Mitra, P.: Statistical downscaling of precipitation using long short-term memory recurrent neural networks, Theoretical and Applied Climatology, 134, 1179–1196, https://doi.org/10.1007/s00704-017-2307-2, http://link.springer.com/10.1007/s00704-017-2307-2, 2018.

Pan, B., Hsu, K., AghaKouchak, A., and Sorooshian, S.: Improving Precipitation Estimation Using Convolutional Neural Network, Water Resources Research, 55, 2301–2321, https://doi.org/10.1029/2018WR024090, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018WR024090, 2019.

https://doi.org/10.5194/gmd-2019-278
Preprint. Discussion started: 28 October 2019
© Author(s) 2019. CC BY 4.0 License.

Pour, S. H., Shahid, S., and Chung, E.-S.: A Hybrid Model for Statistical Downscaling of Daily Rainfall, Procedia Engineering, 154, 1424–1430, https://doi.org/10.1016/j.proeng.2016.07.514, http://www.sciencedirect.com/science/article/pii/S1877705816319038, 2016.

Pradhan, R., Aygun, R. S., Maskey, M., Ramachandran, R., and Cecil, D. J.: Tropical Cyclone Intensity Estimation Using a Deep Convolutional Neural Network, IEEE Transactions on Image Processing, 27, 692–702, https://doi.org/10.1109/TIP.2017.2766358, 2018.

Rasp, S., Pritchard, M. S., and Gentine, P.: Deep learning to represent subgrid processes in climate models, Proceedings of the National Academy of Sciences, 115, 9684–9689, https://doi.org/10.1073/pnas.1810286115, http://www.pnas.org/content/115/39/9684, 2018.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., and Prabhat: Deep learning and process understanding for data-driven Earth system science, Nature, 566, 195–204, https://doi.org/10.1038/s41586-019-0912-1, https://www.nature.com/articles/s41586-019-0912-1, 2019.

Riley, P.: Three pitfalls to avoid in machine learning, Nature, 572, 27, https://doi.org/10.1038/d41586-019-02307-y, 2019.

Rodrigues, E. R., Oliveira, I., Cunha, R. L. F., and Netto, M. A. S.: DeepDownscale: a Deep Learning Strategy for High-Resolution Weather Forecast, arXiv:1808.05264 [cs, stat], http://arxiv.org/abs/1808.05264, arXiv: 1808.05264, 2018.

Sachindra, D. A. and Kanae, S.: Machine learning for downscaling: the use of parallel multiple populations in genetic programming, Stochastic Environmental Research and Risk Assessment, https://doi.org/10.1007/s00477-019-01721-y, 2019.

Sachindra, D. A., Ahmed, K., Rashid, M. M., Shahid, S., and Perera, B. J. C.: Statistical downscaling of precipitation using machine learning techniques, Atmospheric Research, 212, 240–258, https://doi.org/10.1016/j.atmosres.2018.05.022, http://www.sciencedirect.com/science/article/pii/S0169809517310141, 2018.

Scher, S. and Messori, G.: Weather and climate forecasting with neural networks: using GCMs with different complexity as study-ground, Geoscientific Model Development Discussions, pp. 1–15, https://doi.org/https://doi.org/10.5194/gmd-2019-53, https://www.geosci-model-dev-discuss.net/gmd-2019-53/, 2019.

Schmidhuber, J.: Deep learning in neural networks: An overview, Neural Networks, 61, 85–117, https://doi.org/10.1016/j.neunet.2014.09.003, http://www.sciencedirect.com/science/article/pii/S0893608014002135, 2015.

Schoof, J. and Pryor, S.: Downscaling temperature and precipitation: a comparison of regression-based methods and artificial neural networks, International Journal of Climatology, 21, 773–790, https://doi.org/10.1002/joc.655, http://onlinelibrary.wiley.com/doi/10.1002/joc.655/abstract, 2001.

Tripathi, S., Srinivas, V. V., and Nanjundiah, R. S.: Downscaling of precipitation for climate change scenarios: A support vector machine approach, Journal of Hydrology, 330, 621–640, https://doi.org/10.1016/j.jhydrol.2006.04.030, http://www.sciencedirect.com/science/article/pii/S0022169406002368, 2006.

Vandal, T., Kodra, E., and Ganguly, A. R.: Intercomparison of Machine Learning Methods for Statistical Downscaling: The Case of Daily and Extreme Precipitation, arXiv:1702.04018 [stat], http://arxiv.org/abs/1702.04018, arXiv: 1702.04018, 2017a.

Vandal, T., Kodra, E., Ganguly, S., Michaelis, A., Nemani, R., and Ganguly, A. R.: DeepSD: Generating High Resolution Climate Change Projections through Single Image Super-Resolution, https://arxiv.org/abs/1703.03126, 2017b.

Vandal, T., Kodra, E., and Ganguly, A. R.: Intercomparison of machine learning methods for statistical downscaling: the case of daily and extreme precipitation, Theoretical and Applied Climatology, 137, 557–570, https://doi.org/10.1007/s00704-018-2613-3, https://doi.org/10.1007/s00704-018-2613-3, 2019.

Wang, Z., Liu, K., Li, J., Zhu, Y., and Zhang, Y.: Various Frameworks and Libraries of Machine Learning and Deep Learning: A Survey, Archives of Computational Methods in Engineering, https://doi.org/10.1007/s11831-018-09312-w, https://doi.org/10.1007/s11831-018-09312-w, 2019.

https://doi.org/10.5194/gmd-2019-278
Preprint. Discussion started: 28 October 2019
© Author(s) 2019. CC BY 4.0 License.

Widmann, M., Bedia, J., Gutiérrez, J. M., Bosshard, T., Hertig, E., Maraun, D., Casado, M. J., Ramos, P., Cardoso, R. M., Soares, P. M. M., Ribalaygua, J., Pagé, C., Fischer, A. M., Herrera, S., and Huth, R.: Validation of spatial variability in downscaling results from the VALUE perfect predictor experiment, International Journal of Climatology, 39, 3819–3845, https://doi.org/10.1002/joc.6024, https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.6024, 2019.

Wilby, R. L., Wigley, T. M. L., Conway, D., Jones, P. D., Hewitson, B. C., Main, J., and Wilks, D. S.: Statistical downscaling of general circulation model output: A comparison of methods, Water Resources Research, 34, 2995–3008, https://doi.org/10.1029/98WR02577, http://onlinelibrary.wiley.com/doi/10.1029/98WR02577/abstract, 1998.

Yang, C., Wang, N., Wang, S., and Zhou, L.: Performance comparison of three predictor selection methods for statistical downscaling of daily precipitation, Theoretical and Applied Climatology, pp. 1–12, https://doi.org/10.1007/s00704-016-1956-x, https://link.springer.com/article/10.1007/s00704-016-1956-x, 2016.