Deep Neural Networks for Statistical Downscaling of Climate Change Projections

1st Jorge Baño-Medina
Grupo de Meteorología
Instituto de Física de Cantabria
(CSIC-UC) Santander, Spain
bmedina@ifca.unican.es

2nd José Manuel Gutiérrez
Grupo de Meteorología
Instituto de Física de Cantabria
(CSIC-UC) Santander, Spain
gutierjm@ifca.unican.es

3rd Sixto Herrera
Grupo de Meteorología
Dept. Applied Mathematics
and Computing (UC) Santander, Spain
herreras@unican.es

***Abstract*—Global Climate Models (GCMs) are the main tool to predict the future evolution of the atmosphere at different time scales, from weather forecast (for the next few days) to climate change projections (where the forcing effect of greenhouse gases drives future climate trend projections for the next few decades). The main limitation of GCMs is their limited spatial resolution (hundreds of kilometers for climate change applications). A number of *statistical downscaling* techniques have been proposed in the last two decades to increase the resolution of these predictions taking into account the relationships between model outputs and local observations of the variables of interest. Besides the classical approaches based on (generalized) linear regression or analogs, a number of machine learning approaches have been applied to this problem. However, there is general consensus that only limited added value is obtained with these techniques when jointly considering model performance, interpretability and parsimony. In this Thesis we analyze the potential of deep learning in this field, which is yet unexplored. In particular, we analyze the promising properties of convolutional neural networks using as benchmark a recent intercomparison experiment of over 50 statistical downscaling methods over Europe (VALUE initiative, http://www.value-cost.eu). Some promising results are reported for a first illustrative example (precipitation occurrence), showing that these models automatically handle redundancy and perform geographical and variable selection/transformation of predictors in a robust and spatially consistent form. The relevance of this new approach is discussed in the context of a number of international initiatives where this Thesis will contribute.**

***Index Terms*—climate change, statistical downscaling, deep learning, convolutional neural networks**

I. Introduction

Global Climate Models (GCM) are key tools to simulate and predict the evolution of the climate system by numerically solving the physical equations governing its different components (atmosphere, hydrosphere, cryosphere, lithosphere) and the interactions among them [1], [2]. These models are solved over a 3D grid discretizing the globe (latitude, longitude, and height) and provide information for a large number of climatic variables, with typical spatial and temporal resolutions of hundreds of kilometers and days, respectively. GCMs are crucial for studying the future evolution of the climate and for assessing the impacts of climate change under different socio-economic emission scenarios (different plausible evolutions of concentrations of greenhouse gases in the atmosphere). For instance, politicians have recently adopted mitigation climate measures in the historical 2015 Paris agreement based on the assessment provided by the Intergubernamental Panel on Climate Change (IPCC) [3], which builds mainly on GCM projections for the present century [4].

One of the main practical limitations of GCMs is that they do not solve regional to local processes due to its coarse resolution, specially in areas where local phenomena are relevant (e.g., coastal areas and/or complex topography regions). Regional and local climate information is crucial to determine the effects of climate change in different impact areas, such as hydrology, agriculture and energy production. In order to bridge this gap, a number of statistical downscaling (SD) [1] techniques have been developed during the last two decades building on empirical relationships stablished between informative large-scale atmospheric variables provided by GCMs (predictors, e.g. humidity or temperature at different atmospheric height levels) and observational records of the variables of interest at regional/station scale (predictands). These relationships are learned using simultaneous daily records for predictors (normally obtained from a retrospective forecast dataset) and observational variables for a ($\sim$ 30 years) historical period (more details in [2], [5]). To generate future regional projections, the statistical models learned in a historical period are then applied to (the predictors from) future GCM projections. This poses a number of methodological issues that are analyzed in detail in Section II.

Besides the classical SD techiques based on linear approaches (e.g., generalized linear models) or non-parametric techniques (e.g., analogs) [6], a number of sophisticated machine learning techniques have been applied to this problem. For instance, the first applications of neural networks are dated back in the late 90s [7], [8]. The problems related to nonconvexity, time-consuming learning, and overfitting drove attention to alternative machine learning approaches, such as support vector machines (SVMs) [9] or random forests [10], [11]. However, there is general consensus that no method clearly outperforms the others and, in general, only limited added value is obtained with nonlinear techniques in the context of climate change when considering model performance, interpretability and parsimony [7], [12], [13], [14].

We acknowledge financial support from the project MULTI-SDM (CGL2015-66583-R, MINECO/FEDER).

Machine learning is currently a very active research area, boosted by the major deep learning (DL) breakthroughs in the field of neural networks (see [15] and references therein). Deep learning extracts high-level feature representations in a hierarchical way due to its (deep) layered-structure whose unit elements (neurons) are connected by adjustable weights. Different combinations of convolutions, auto-encoders and/or classic fully-connected layers allow to model a variety of problems in several disciplines. Moreover, new efficient learning methods (e.g. batch, stochastic, and mini-batch gradient descent), regularization options (e.g. dropout), and frameworks (e.g. TensorFlow) allow to efficiently learn these models from (big) data, avoiding overfitting. However, finding the adequate topology for a particular problem is still a challenging step. Particularly in computer vision, deep learning has outstanded against other machine learning techniques with a specific topology called deep convolutional neural network (CNN). CNN was first introduced in [16], appeared as a neural network specially designed for regular grid-structure inputs such as images (2D). The characteristic of CNN is that the parameters convolutes over the 2D map, exploiting the spatial structure and resulting into fewer parameters than traditional neural networks. Thus, the layers consists in a set of neurons spatially arranged called maps or filters that represents the spatial distribution of a particular feature learned by the net. The deeper the network, the more complex the features are in the filter maps.

Deep learning is a very active topic in many communities, such as bioinformatics [17]. In the case of meteorology and climate, there are only a few applications of deep learning such as the estimation of cyclone's intensity [18], the detection of extreme weather events [19] or a first approach for downscaling [20], among others. The latter establishes an analogy between images and atmospheric fields to generate super-resolution precipitation images and set the path for the application of deep CNN in statistical downscaling.

In this Thesis we will undertake a comprehensive analysis of deep learning for statistical downscaling of climate change projections, analyzing the adequacy of different components and topologies for this problem and assessing the underlying assumptions and methodological issues required for a robust applications of deep CNN in this field. We also analyze the replicability and explicability of results in order to gain confidence in these techniques, which are currently seen as black-box methods by the climate community. We will build and contribute to standard experimental frameworks and international initiatives focusing on statistical downscaling (e.g. IPCC, VALUE [21] and CORDEX [22]) and use as benchmarks the state-of-the-art methods developed therein. In particular, we will consider a recent intercomparison experiment of over 50 statistical downscaling methods over Europe developed in the framework of the VALUE initiative, which constitutes the largest to date intercomparison of statistical downscaling methods [6]. This approach will maximize the diffusion and impact of the results of this Thesis.

II. Hypothesis and Methodology

Statistical downscaling methods have to fulfill three assumptions in a climate change context in order to provide plausible results and to avoid statistical artifacts [1]:

1) *Explanation of local variability:* The methods should explain a large fraction of local variability in order to provide an informative link between the large scale (predictors) and the local scale (predictands). Besides the choice of the downscaling technique, the selection of informative predictors over suitable geographical domains plays a key role here in order to convey the appropriate large scale information to downscale the variable of interest (e.g. precipitation or temperature). This assumption is assessed using a variety of validation metrics, which measure (directly or indirectly) the percentage of local variability explained by the downscaling method. In this Thesis we will build on the previous work done in the VALUE initiative for the validation of statistical downscaling methods [21].
2) *Selection of robust predictors:* Since the downscaling methods are trained in a historical period using predictors from a retrospective GCM forecast (and simultaneous observations) and then applied to predictors from different GCM future projections, it is required that the predictors are realistically simulated by the different GCMs in present/historical climate. Therefore, suitable predictors for climate change studies typically restrict to large scale variables (such as pressure, wind components, temperature and humidity) at different heights (e.g. 850 and 500 mb, corresponding approximately to 1500 and 5000 meter above sea level, respectively). Surface variables (apart from sea level pressure or surface temperature) are commonly not used as predictors, since they strongly depend on the particular topography/resolution used by the GCM.
3) *Extrapolation capability:* This is a key requirement for SD methods, since the future climate change signal may be artificially biases otherwise. Therefore, the structure of the statistical downscaling method should provide extrapolation capabilities to future climates. This implies that the predictors used are credibly projected into the future by the GCM and that the statistical downscaling method can extrapolate out-of-sample records. Therefore, good cross-validation performance is a necessary but not sufficient requirement, since the values of predictors in future climates can be far away from the historical climate. Additional cross-validation experiments with pseudo-observations (using future model predictions as observations, since observations of the future climate are not available) have been suggested for this task and will be used in this Thesis [21].

As a consequence of these requirements, predictors must be carefully selected (both the particular variables and the geographical region of influence) in order to obtain credible results. Atmospheric predictors are very redundant and the same event can be driven by completely different physical processes depending on the region. As a consequence, identifying the adequate informative predictors —both the variables and the geographical regions of influence— for a certain task (e.g., downscaling precipitation) is a major challenge in statistical downscaling. For instance, this problem was reported in the largest to date intercomparison of standard SD methods performed in the VALUE initiative (http://www.value-cost.eu) as one of the most time-consuming tasks in most of the cases. Predictor selection was typically undertaken applying tedious feature selection (e.g., stepwise algorithm) and/or feature reduction (e.g., principal component analysis) techniques [6]. Moreover, when considering large continental domains (Europe in VALUE) most of the methods tackle the predictor selection task by subdividing Europe in 8 regions (see Figure 1), which hinders transferability to other domains.

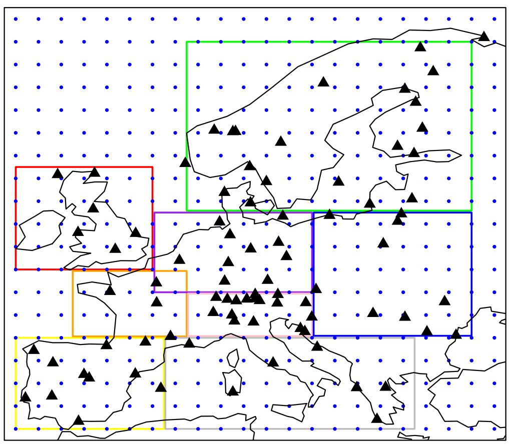

Fig. 1. Geographical domain indicating the renalysis' gridboxes or predictors (2° resolution, blue dots, aprox. 200 km) and the location of 86 stations or predictands (black triangles). The coloured boxes show 8 subregions, illustrating how traditional regions need to split big domains into subdomains due to dimensionality constraints; more details in [6].

In principle, machine learning techniques could help with these problems, allowing to develop automatic SD methods based on the available data, building on automatic feature selection approaches. In particular deep learning and in particular CNNs seem to be ideal candidates for this problem since convolutional layers could deal with the problem of selecting informative geographical regions for each predictand (e.g. location), whereas fully-connected layers could model potential nonlinearities connecting large and local scales. In Section IV we show some first results indicating the suitability of CNNs for this purpose.

Another disadvantage of statistical downscaling is that is very dependent on both the quality and quantity of data, in order to infer reliable models. This poses serious problems for downscaling in regions that lack of data, such as the Antarctica or Africa. Traditionally in these cases dynamical downscaling is preferred (see [22] for a review), due to the incapacity to infer the parameters of a statistical model. In this Thesis we will try to address this issue by using the concept of *transfer learning*, present in multi-task neural networks [23]. For downscaling problems, transfer learning can be interpreted as a way to answer the following question: can information of a certain region be useful to downscale climate in another place on the planet? Transfer learning has been sucesfully applied in other fields, such as natural language processing [24] and computer vision [25]. Furthermore, we believe that the ability to simultaneously downscale to various locations will result in more spatially coherent downscaling, very important in impact studies.

Fortunately, new sophisticated software and computation sources have been developed in order to easy the implementation of deep neural networks, permitting the construction of versatile and complex architectures. In particular, we will use the package TensorfFlow [ref], which facilitates the solid implementation of CNNs using the most relevant deep learning advances. Among them, we can construct deep neural networks with novel optimitation algorithms (e.g. Adam algorithm or Adagrad algorithm), different kinds of weight initializations (e.g. He's initialization or Xavier initialization), various activation functions (e.g. sigmoidal classification and ReLu activations), new regularization techniques (e.g. dropout), and pre-build complex hidden layers (e.g. convolutional layers). In this Thesis we will focus in deep CNNs, which builds on different parameters: kernel size, pooling, padding and number of filter maps. We will explore the effects of these parameters of convolutional layers as well as other deep learning advances in order to analyse the applicability of deep learning in statistical downscaling.

III. Objectives and Workplan

This Thesis is devoted to the study and application of deep neural networks for statistical downscaling in the context of climate change, building on the intercomparison framework developed in the VALUE initiative [21]. The main objetive is developing a CNN-based downscaling method which solves some of the outstanding open problems of statistical downscaling: 1) faces curse of dimensionality and automatically feature select/transform the predictors, 2) exploits the spatial structure, 3) operates over continental-sized domains, 4) is able to extrapolate the results to "unseen" regions by transfer learning and 5) quantifies the uncertainty of the predictions. In order to accomplish this objective we designed the following workplan.

A. *Complementary academic formation (Months 1 - 6)*

The first task to accomplish has been achieving and adequate complementary formation in machine learning and the new advances in deep learning. This has been done using standard texts (e.g. [26], [27]), and attending to special curses and workshops. Moreover, an extensive bibliographic search was conducted in order to get an up-to-date overview of the field, and the published references on machine learning applications to statistical downscaling. From this search we conclude that the topic is yet unexplored (there is a single general publication on this topic, [20]) and, therefore, the

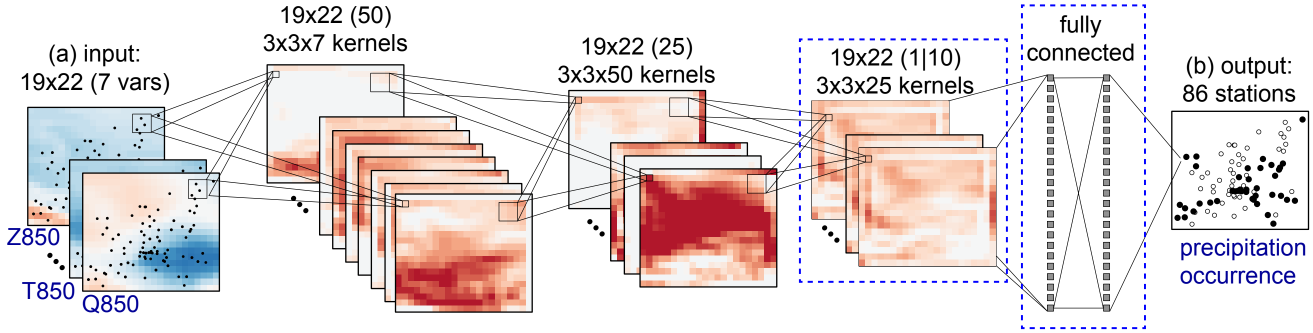

Fig. 2. Deep neural network architecture found in this Thesis to accomplish the objectives proposed. In particular, the net aims to downscale the precipitation's occurrence over 86 stations (output) based on several large-scale predictors of size 19×22 (input: 7 predictors), building on three convolutional layers with 50 kernels of size $3 \times 3 \times 7$, 25 kernels of size $3 \times 3 \times 50$ and one kernels of size $3 \times 3 \times 25$, respectively, followed by dense layers, with a total of over 50000 parameters. ReLU non-linear functions are used in all layers, except for the outputs, which are sigmoid for classification.

Thesis can be a significant and timely contribution to this field (see Section V).

B. Deep learning for statistical downscaling (Months 6 - 30)

This task is the core of the Thesis and consists in exploring the deep learning developments in a statistical downscaling context building on TensorFlow. We focus in CNN due to their suitability to deal with spatial inputs (atmospheric fields in this case). However, many different kinds of CNN arquitechtures can be found in the literature, ranging from only convolutional layers to a combination of convolutional, autoencoders and dense layers. Exploring these configurations and understanding its influence on different validation metrics (e.g., temporal and spatial metrics, metrics related to extreme weather events) will be the main topic along the first part of the Thesis. To date we have explored the effects of CNN parameters such as the kernel size, whether to add padding or not (i.e., the filter map has the same resolution than the original map) and whether to incorporate pooling (i.e., a parameter highlighting the presence or not a certain parameter). Furthermore, as we are particularly interested in spatial metrics and the spatial consistency of the climate fields, we pay special attention to multi-task architectures rather than single-task. Thus, during this task we will also evaluate the benefits of multi-task architectures over single-task according to spatial metrics. In the end, the objective will be to come with a particular deep learning net that justifies its architecture based on its ability to handle predictor information and on the benefits obtained from transfer learning. Some first results have been already obtained in this task, corresponding to a simple illustrative classification example (precipitation occurrence, wether it has rained or not) used to analyze and understand the role of the different layers and elements in the downscaling process. This work is described in Section IV, which shows promising results to handle predictor redundancy and irrelevancy automatically.

C. Quantifying uncertainty (Months 18 - 30)

Bayesian neural networks have existed since the late 80's [28]. However their intractability in many neural network topologies along with a damage in the performance with respect to non-Bayesian neural networks [29], prevented from a widespread use of these models. Recently, Bayesian deep learning has simplified these problems by simply leaning on dropout [30]. Dropout consists in giving every neuron of the neural net a certain probability to exist in a particular step of the training process. Thus, at every new epoch a new subset of the original net is trained. This randomness generates distributions of predicted values that carry the uncertainty information. Dropout is easily implemented by TensorFlow and will basically consist in adding dropout to the deep learning architecture found to solve SD tasks.

D. Model Explanation (24 - 30 months)

The climate community is reluctant to black-box machine learning methods due to the inability to explain the results. Therefore, in this Thesis we analyze model explicability of deep learning in statistical downscaling, trying to understand what elements are key for the different components and layers of the model, and which factors influence the performance of the model when compared to benchmark methods. We will examine the physical interpretation (relative to the problem under study) of the different hidden layers. For instance, in Section IV we analyze a simple example that allows for a interpretability of the convolutional layers. Thus, we will try to understand the physics underlying the coupling of large and local scales, as learned by the deep neural network model. For instance, precipitation is driven by different processes in the Mediterranean and in North Europe, but are these physical processes captured as patterns responsible of the improvement of the downscaling with respect to benchmark models?

E. Divulgation and contribution to international initiatives (Months 24 - 36)

The results of this Thesis will be published in artificial intelligence and climate journals and conferences, such as Conferencia de la Asociación Española para la Inteligencia Artificial (CAEPIA), Neural Information Processing Systems (NIPS) and Climate Informatics (CI). In fact, the results described in Section IV have been already submitted to the Climate Informatics congress that will take place in September in Colorado. Moreover, in order to maximize the diffusion and visibility of this work, (see Section V). Finally, the submission of the Thesis is planned for mid 2020.

IV. First Results

In this section we describe the first results obtained during the first year of the realization of the Thesis. We have explored different deep learning topologies and have obtained a particular CNN configuration able to shed light and respond to the objectives described in the previous section for a simple statistical downscaling case-study corresponding to a classification problem: downscaling the occurrence of precipitation. In particular we consider the experimental framework of the first VALUE intercomparison experiment, consisting on downscaling over 86 stations located over Europe (Figure 1). As benchmark we use one of the best performing models participating in this experiment, based on Generalized Linar Models (GLM), in particular logistic regression for this case [6].

The topology of deep learning architectures vary depending on the task to be solved. According to the objectives described in Section III, the configuration for this problem should automatically handle the selection of predictors, dealing with the typical redundancy and irrelevancy properties of atmospheric predictors. In particular to this study we have used the following redundant set: the specific humidity at 850 and 700 hPa, the temperature at 850 and 700 hPa, and the geopotential height at 1000, 850 and 500 hPa. In order to address the latter and find an adequate configuration, we have tested a combination of convolutional layers with dense layers (see Figure 2). The output layer has 86 neurons, one for each station of Figure 1. For this case-study, an only-convolutional configuration (i.e., excluding the two last fully connected layers) achieves higher validation scores and provides more interpretability with respect to unraveling the implicit geographical feature selection for this simple example. However, the search has not been limited to define wether the net should be partially or totally convolutional, and the number of filter maps in the last layer has been found to be crucial for the performance. Performance suffers from the curse of dimensionality and thus the addition of unnecessary filter maps to the neural architecture harms the statistical significance of the model. This is specially relevant if they are added in the last layer directly having an impact in the space's size where there is the boundary layer separating rainy from non-rainy days. As a consequence, reducing the number of filter maps to 1 in the last layer has shown better performance than with a bigger number of filter maps (e.g., 10), demonstrating that only 1 filter map is necessary to this simple case. In this only 1 filter map architecture, the CNN filters the information coming from the original predictors retaining a pattern that best downscales the precipitation and representing it spatially in its third layer, generating a new novel super-predictor. By this way, the CNN feature selects/transforms the original predictors into only 1. But does CNN also performs a domain feature-selection?. In Figure 3 we observe the coefficient's values linking the third layer with the output layer for a) Madrid and b) Helsinki. It is interesting that the CNN automatically finds the area influencing the local climate for a particular station, from its surrounding geographical region, ignoring the rest of the continental domain.

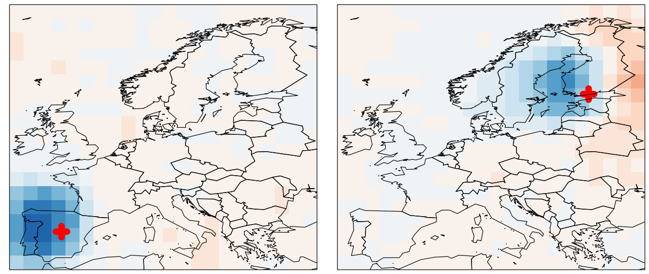

Fig. 3. Weights from the last convolution layer to the outputs (the stations) for Madrid (left) and Helsinki (right). A 5×5 spatial moving average is applied to represent the effect of kernels. Blue/red colors indicate positive/negative weights.

The resulting CNN shows better performance than the considered state-of-the-art benchmark method (GLM), in terms of a standard validation measure: Relative Operating Characteristics Skill Score (ROCSS), which is a standard accuracy measure for probabilistic predictions of binary variables. In Figure 4 we observe the ROCSS as a function of the a) training epochs and b) stations for three different configurations of deep learning models and for the benchmark (GLM). We observe how the deep CNN with only 1 feature map in its last layer and with no fully connected layers after the convolutions achieves the best results. In particular, it is remarkable how deep learning models have achieved considerably higher results than the benchmark model, which additionally required a tedious pre-analysis of the predictors.

V. Relevance

Statistical downscaling is an important problem in the context of climate change, since it allows to transfer the global information of GCMs to the regional and local scales needed in impact studies. There are a number of important international initiatives which focus on this problem, including the IPCC (Intergubernamental Panel on Climate Change, [4]) and CORDEX (COordinated Regional climate Downscaling EXperiment, [22]). Also, at a national level, the Spanish National Adaptation Plan (PNACC) provides regional climate change information for Spain building, among others, on statistical downscaling methods (see http://escenarios.adaptecca.es). At an international level, one of the main limitations of statistical downscaling method is that they need to be applied at a global level, considering continental domains. As we have seen in the previous sections, state-of-the-art statistical downscaling methods cannot operate automatically and require human intervention to define suitable geographical regions, predictors, etc. CNNs may provide an alternative to circumvent these problems and may contribute to these initiatives providing downscaled results with global coverage. The Thesis will be performed in the framework of these initiatives, which will sever as ideal platforms for the dissemination of results. At

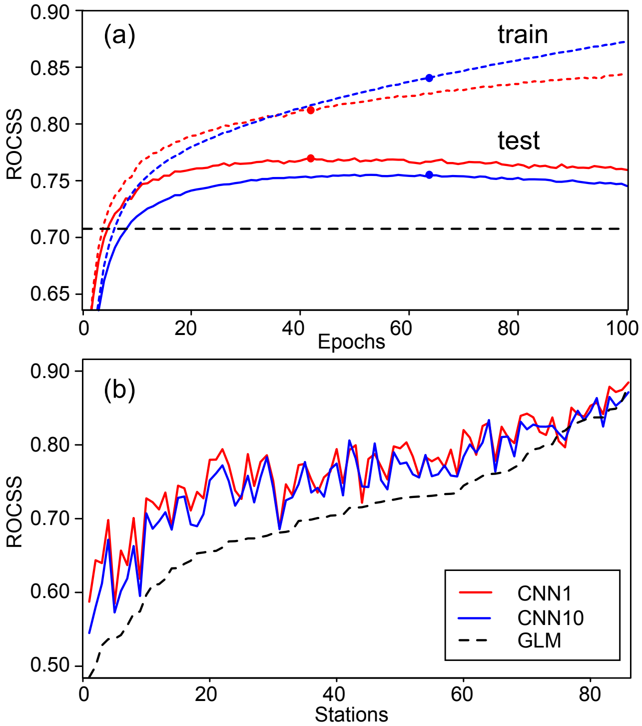

Fig. 4. Results (ROCSS) of the different downscaling models as a function of (a) the epochs (train/test results are indicated by solid/dashed lines; the dots indicate early stopping) and (b) the stations for the trained models, with stations sorted according to the GLM results. The deep learning models have 1 filter map in the third hidden layer (CNN1) or 10 filter maps (CNN10).

a national level, one of the main limitations of statistical downscaling methods is that they do not provide spatially consistent results. CNNs could also circumvent this problem, thus contributing to a better provision of regional climate change information for impact studies in Spain.

Acknowledgments

Funding was provided by the project MULTI-SDM (CGL2015-66583-R, MINECO/FEDER).

References

[1] D. Maraun and M. Widmann, *Statistical Downscaling and Bias Correction for Climate Research*. Cambridge University Press, 2018.

[2] M. . Gaertner, J. M. Gutirrez, and M. Castro, "Escenarios regionales de cambio climtico.," vol. 26, no. 2.

[3] IPCC, *Summary for Policymakers*, book section SPM, p. 130. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press, 2013.

[4] M. Collins, R. Knutti, J. Arblaster, J.-L. Dufresne, T. Fichefet, P. Friedlingstein, X. Gao, W. Gutowski, T. Johns, G. Krinner, M. Shongwe, C. Tebaldi, A. Weaver, and M. Wehner, *Long-term Climate Change: Projections, Commitments and Irreversibility*, book section 12, p. 10291136. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press, 2013.

[5] J. T. Schoof, "Statistical downscaling in climatology," vol. 7, no. 4, pp. 249–265.

[6] J. M. Gutirrez, D. Maraun, M. Widmann, R. Huth, E. Hertig, R. Benestad, O. Roessler, J. Wibig, R. Wilcke, S. Kotlarski, D. San Martn, S. Herrera, J. Bedia, A. Casanueva, R. Manzanas, M. Iturbide, M. Vrac, M. Dubrovsky, J. Ribalaygua, J. Prtoles, O. Rty, J. Risnen, B. Hingray, D. Raynaud, M. J. Casado, P. Ramos, T. Zerenner, M. Turco, T. Bosshard, P. tpnek, J. Bartholy, R. Pongracz, D. E. Keller, A. M. Fischer, R. M. Cardoso, P. M. M. Soares, B. Czernecki, and C. Pag, "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*, Mar. 2018.

[7] R. L. Wilby, T. M. L. Wigley, D. Conway, P. D. Jones, B. C. Hewitson, J. Main, and D. S. Wilks, "Statistical downscaling of general circulation model output: A comparison of methods," vol. 34, no. 11, pp. 2995–3008.

[8] J. Schoof and S. Pryor, "Downscaling temperature and precipitation: a comparison of regression-based methods and artificial neural networks," vol. 21, no. 7, pp. 773–790.

[9] S. Tripathi, V. V. Srinivas, and R. S. Nanjundiah, "Downscaling of precipitation for climate change scenarios: A support vector machine approach," vol. 330, no. 3, pp. 621–640.

[10] S. H. Pour, S. Shahid, and E.-S. Chung, "A hybrid model for statistical downscaling of daily rainfall," vol. 154, pp. 1424–1430.

[11] X. He, N. W. Chaney, M. Schleiss, and J. Sheffield, "Spatial downscaling of precipitation using adaptable random forests," vol. 52, no. 10, pp. 8217–8237.

[12] S.-T. Chen, P.-S. Yu, and Y.-H. Tang, "Statistical downscaling of daily precipitation using support vector machines and multivariate analysis," vol. 385, no. 1, pp. 13–22.

[13] T. Vandal, E. Kodra, and A. R. Ganguly, "Intercomparison of machine learning methods for statistical downscaling: The case of daily and extreme precipitation,"

[14] C. Yang, N. Wang, S. Wang, and L. Zhou, "Performance comparison of three predictor selection methods for statistical downscaling of daily precipitation," pp. 1–12.

[15] J. Schmidhuber, "Deep learning in neural networks: An overview," vol. 61, pp. 85–117.

[16] Y. LeCun and Y. Bengio, "Convolutional networks for images, speech, and time series," *The handbook of brain theory and neural networks*, vol. 3361, no. 10, p. 1995, 1995.

[17] S. Min, B. Lee, and S. Yoon, "Deep Learning in Bioinformatics," p. 46.

[18] R. Pradhan, R. S. Aygun, M. Maskey, R. Ramachandran, and D. J. Cecil, "Tropical Cyclone Intensity Estimation Using a Deep Convolutional Neural Network," *IEEE Transactions on Image Processing*, vol. 27, pp. 692–702, Feb. 2018.

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

[20] T. Vandal, E. Kodra, S. Ganguly, A. Michaelis, R. Nemani, and A. R. Ganguly, "DeepSD: Generating High Resolution Climate Change Projections through Single Image Super-Resolution," Mar. 2017.

[21] D. Maraun, M. Widmann, J. M. Gutirrez, S. Kotlarski, R. E. Chandler, E. Hertig, J. Wibig, R. Huth, and R. A. Wilcke, "VALUE: A framework to validate downscaling approaches for climate change studies," vol. 3, no. 1, p. 2014EF000259.

[22] W. J. Gutowski Jr., F. Giorgi, B. Timbal, A. Frigon, D. Jacob, H.-S. Kang, K. Raghavan, B. Lee, C. Lennard, G. Nikulin, E. O'Rourke, M. Rixen, S. Solman, T. Stephenson, and F. Tangang, "WCRP COordinated regional downscaling EXperiment (CORDEX): a diagnostic MIP for CMIP6," vol. 9, no. 11, pp. 4087–4095.

[23] S. Ruder, "An Overview of Multi-Task Learning in Deep Neural Networks," *arXiv:1706.05098 [cs, stat]*, June 2017. arXiv: 1706.05098.

[24] R. Collobert and J. Weston, "A Unied Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning," p. 8.

[25] R. Girshick, "Fast R-CNN," pp. 1440–1448, 2015.

[26] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2006.

[27] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[28] N. Tishby, E. Levin, and S. A. Solla, "Consistent inference of probabilities in layered networks: predictions and generalizations," in *International 1989 Joint Conference on Neural Networks*, pp. 403–409 vol.2, 1989.

[29] G. E. Hinton, "Keeping Neural Networks Simple by Minimizing the Description Length of the Weights," p. 9.

[30] Y. Gal, "Uncertainty in Deep Learning," p. 166.