9th International Workshop on Climate Informatics
October 02-04, 2019
Hosted by École Normale Supérieure, Paris, France

THE IMPORTANCE OF INDUCTIVE BIAS IN CONVOLUTIONAL MODELS FOR STATISTICAL DOWNSCALING

Jorge Baño-Medina[1], Jose M. Gutiérrez [1]

***Abstract*—Statistical downscaling is routinely used to produce regional climate change projections from coarse global model outputs. Along with the success of deep learning in multiple disciplines, recent studies outline the capability of deep neural models as a statistical downscaling technique. In this work, we analyze this problem from a multi-site perspective and highlight the benefits of a deep learning model. We argue that their merits are due to the existence of an inductive bias in multi-site architectures that prevents overfitting in over-parameterized models with no need for dimensionality reduction techniques. We frame the experiment in the largest to date downscaling intercomparison study, called VALUE. The result is a better local reproducibility of multi-site deep neural models in comparison with single-site neural models and VALUE's benchmark methods.**

I. MOTIVATION

In order to study the potential impacts of climate change, and to elaborate suitable adaptation measures, certain socio-economic sectors (e.g., agriculture, energy, health) require regional climate information for the next decades. The projections provided by the Global Climate Models (GCMs) are too coarse to be directly applied in many sectors. Statistical downscaling [1] bridges the gap between the low and the high-resolutions of GCMs and local observations, respectively, by learning empirical functions between large-scale atmospheric variables and a record of climate observations at a local scale. The perfect prognosis approach builds these relationships using reanalysis data.

Classical techniques including generalized linear models (GLMs) or analogs [2], and machine learning models, such as neural networks [3] or support vector machines [4], are among the most common statistical methods used by the downscaling community. Despite there have been numerous intercomparison studies, to date no method outstand against the others in terms of temporal reproducibility or spatial consistency.

However, in the last decade, the advances in the development of neural networks (e.g., stochastic gradient descent, new learning algorithms [5], ReLu activation function [6] and computing infrastructures) have made architectures with many layers tractable, becoming the state-of-the-art in other disciplines (e.g., image recognition [7]). Thus, despite convolutional architectures have existed since the 90s [8], it was not until recently when they first appeared as statistical downscaling models thanks to the deep learning machinery [9][10]. In particular, in [10] a variety of deep neural models were intercompared in order to shed light on the role of each element in the downscaling process. This was carried out applying the validation framework developed by the European COST-action VALUE [11], which represents the largest to date downscaling intercomparison study.

Due to the difficulties of standard models to downscale continental domains (e.g. Europe) several methods contributing to VALUE considered subdomains and performed dimensionality reduction on the predictors, mainly principal component analysis or nearest neighbour selection. Under this scenario, [10] showed that convolutional models can automatically treat high-dimensional spatial domains without any previous feature selection/extraction process. Furthermore, the non-linear spatial patterns learned by the convolutions resulted in a better reproduction of the local variability than the top-ranked methods in VALUE [12].

In this study we show that the multi-site character of neural models (i.e., downscaling to more than 1 site at a time) favours the simultaneous treatment of high-dimensional domains without leading to overfitting. The main idea behind multi-task learning in neural networks (see [13] for a review) is that the common properties are learned in a shared representation working as an

Corresponding author: J. Baño-Medina, bmedina@ifca.unican.es
[1]Meteorology Group. Instituto de Física de Cantabria, IFCA (CSIC - Univ. de Cantabria), Santander, 39005, Spain

inductive bias to the net and consequently improving generalization [14].

Therefore we extend the work done in [10] and investigate the influence of multi-site arquitectures over the implicit regularization of deep learning models in a statistical downscaling context.

II. DATA

Here we build on the research done in [10] and frame the experiment in the VALUE intercomparison project [12]. Some VALUE methods divided Europe in eight regions for a better characterization of regional climates and a simplification of the input space. Though the eight regions were treated simultaneously as a whole in [10], for simplicity in this study we focus on the domain covering the Iberian peninsula (see Figure 1). According to VALUE, we use the following predictor variables from the ERA-Interim reanalysis [15]: temperature, geopotential, specific humidity and zonal and meridional wind at 250,500 750 and 1000 hPa, resulting into 20 variables per gridpoint. On the other hand we use precipitation of the E-OBS dataset [16] as the response variable (i.e., predictand) on a daily scale. Thus, the objective is to learn empirical functions linking the resolution of ERA-Interim (i.e., 2°) to the E-OBS resolution (0.5° in this example). The cross-validation experiment defined in VALUE indicates the splitting of the data in 5 chronological folds: 1979-1984, 1985-1990, 1991-1996, 1997-2002, 2003-2008. In contrast to [10], where only the 5^{th} fold was used as test set, in this study we reconstruct the 5 folds permitting the analysis of the performance of the methods in the same experimental framework proposed in VALUE.

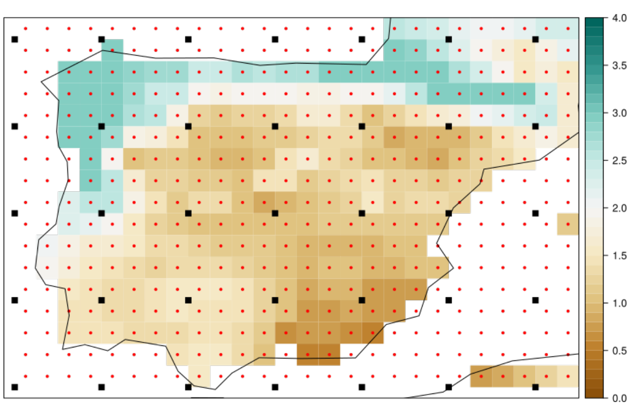

Fig. 1. Climatology of the precipitation in mm/day for the period 1978-2008 over the VALUE Iberian Peninsula subdomain. The black squares represent the predictor's gridpoints (i.e., a total of 35) whereas the red dots indicate the predictand's resolution (i.e., a total of 296 over land).

III. METHODS

In order to evaluate the implicit regularization of neural-based multi-site models in statistical downscaling we start from the intercomparison study done in [10]. According to the latter, an architecture made of 3 convolutional hidden layers with 50,25 and 1 filters, respectively, outperforms the VALUE benchmark methods and other deep learning models in terms of local reproducibility. This success was achieved partly due to the spatial features learned by the convolutions but, in order to study the role played by the inductive bias, we build single-site versions of the architecture described (i.e., everything is identical except from the output layer which reduces to a single gridbox).

Thus, the net uses as input layer 20 feature maps, corresponding to the 20 variables indicated in Section II and fully connect the last convolutional layer with the output layer. Due to the discrete-continuous nature of precipitation, the net minimizes the negative-loglikelihood of a Bernouilli-Gamma distribution as done in [3] and [17]. Therefore, the output layer consists of 3 feature maps (i.e., 3 neurons in single-site mode) matching the parameters p (i.e., probability of rain in a Bernouilli distribution), α (i.e., shape parameter of a Gamma distribution) and β (i.e., scale parameter of a Gamma distribution). The mean daily rainfall for a given day i, can be recovered as the expected value of the conditional Gamma distribution where $\mu_i = \alpha\beta$.

Moreover, we compare both single-site and multi-site models with 2 methods that ranked among the best in a recent contribution to the VALUE experiment (see [18]): a generalized linear model and analogs with a moving window of 4 and 25 neighbours, respectively (e.g., the predictor's gridpoints selected are the 4 closest to the predictand's localization in the case of the GLM).

In terms of reproducibility, we rely on climate4R [19], which is a set of R packages designed to handle climate data and promotes transparent climate data access. Moreover, the downscaling phase can be done with the climate4R library downscaleR [18] for the analogs and the GLM methods. On the other hand, the deep learning models rely on the R version of Keras. The latter facilitates the implementation of the deep learning's new technological developments. To train the models we perform early-stopping as a Keras's callback with a patience of 15 epochs and a learning rate of 5e-4 using the Adam optimizer.

IV. EVALUATION

In Figure 2, we can observe the downscaling done with the CNN-MS (Convolutional Neural Network Multi-Site) for a particular day, using as reference the ERA-Interim's total precipitation. Whereas reanalysis' rainfall pattern lacks from enough resolution, neural-based downscaling is able to capture local structures in the precipitation pattern.

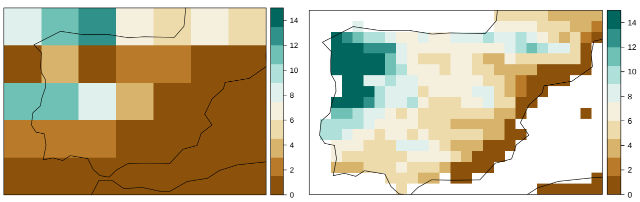

Fig. 2. ERA-Interim's (left) and downscaled (right) total precipitation (mm/day) over Iberia for 01-01-1979.

Multi-site neural models achieve optimum scores in comparison with benchmark methods (i.e., GLM and analogs) and their equivalent single-site networks, in terms of Roc Skill Score (ROCSS), Root Mean Squared Error (RMSE) and spearman correlation (see Table I). On the other hand, the climatological relative bias is better adjusted by the GLM and analogs than by convolutional models, being very close to 0. Though multi-task architectures are very little biased this is especially relevant for single-site models whose relative values overpass 0.5 for most of the folds.

The CNN-MS reproduces slightly better the occurrence of precipitation than the linear model and far better than the single-site model (CNN-SS), according to the ROCSS index. Thus, when downscaling is wanted for single sites, convolutional models suffer from the same high-dimensional issues than classical methods (e.g. GLM). In order to take advantage of the ability of CNNs to extract spatial information and to add nonlinearity to the task, multi-site models are needed, introducing an inductive bias to the model which plays a major role to prevent overfitting. The same result can be observed for the RMSE, where CNN-MS obtains lower values than classical methods and especially than CNN-SS, which again suffers from overfitting. The success in the reproducibility of the occurrence of precipitation, measured by the ROCSS, and in the amount of precipitation, measured by the RMSE, traduces also into better correlation values of multi-site models than the rest of the methods tested.

In order to visualize a more detailed description of the advantages of multi-site models over single-site ones, we plot their differences in the validation indices

ROCSS	**fold 1**	**fold 2**	**fold 3**	**fold 4**	**fold 5**
analogs	0.61	0.58	0.59	0.58	0.57
GLM	0.87	0.86	0.87	0.87	0.86
CNN-SS	0.77	0.75	0.74	0.75	0.78
CNN-MS	**0.90**	**0.88**	**0.89**	**0.89**	**0.89**

RMSE	**fold 1**	**fold 2**	**fold 3**	**fold 4**	**fold 5**
analogs	4.18	4.17	4.1	4.19	4.14
GLM	4.62	4.25	4.03	4.13	4.16
CNN-SS	4.75	4.75	11.84	5.22	5.18
CNN-MS	**3.83**	**3.64**	**3.69**	**3.79**	**3.48**

Sp. Cor	**fold 1**	**fold 2**	**fold 3**	**fold 4**	**fold 5**
analogs	0.58	0.57	0.58	0.57	0.56
GLM	0.72	0.70	0.71	0.71	0.70
CNN-SS	0.72	0.72	0.73	0.73	0.71
CNN-MS	**0.76**	**0.73**	**0.76**	**0.75**	**0.74**

Rel. bias	**fold 1**	**fold 2**	**fold 3**	**fold 4**	**fold 5**
analogs	**-0.04**	**0.05**	-0.11	-0.07	-0.06
GLM	0.11	0.09	**-0.02**	**0.05**	**0.05**
CNN-SS	0.65	0.68	0.17	0.67	0.55
CNN-MS	0.32	0.16	0.15	0.2	0.16

TABLE I
RESULTS FOR THE VALIDATION INDICES. THE BEST RESULTS PER INDEX AND FOLD ARE IN BOLD.

per gridpoint (see Figure 3). Therefore, we observe how the improvement in the ROCSS (Figure 3a) and in the correlation (Figure 3c) of CNN-MS with respect to CNN-SS is a generalized situation, attaining higher values over the majority of the Iberia region. There are very few isolated gridpoints who show the opposite behaviour such as the one located in the Balearic Islands. We hypothesize that this may be due to its very particular local climatology what prevents it from benefiting from multi-site mode. For the amount of precipitation, we observe negative values (Figure 3b), indicating lower errors for the CNN-MS over CNN-SS, especially in southern and northeastern Iberia.

V. CONCLUSIONS

According to the results described in Section IV, we can conclude that:

1) The deep learning machinery is suitable to benefit from high-dimensional large-scale information avoiding the implicit loss of information of dimensionality reduction techniques.
2) If a convolutional model is designed to downscale only at 1 site, then it suffers from the same issues than traditional approaches. Under this scenario, deep learning adds little value to the task.
3) When the interest is to downscale to multiple sites, then multi-task convolutional architectures outperform classical and single-site models in terms of local reproducibility, thanks to the existence of an inductive bias.

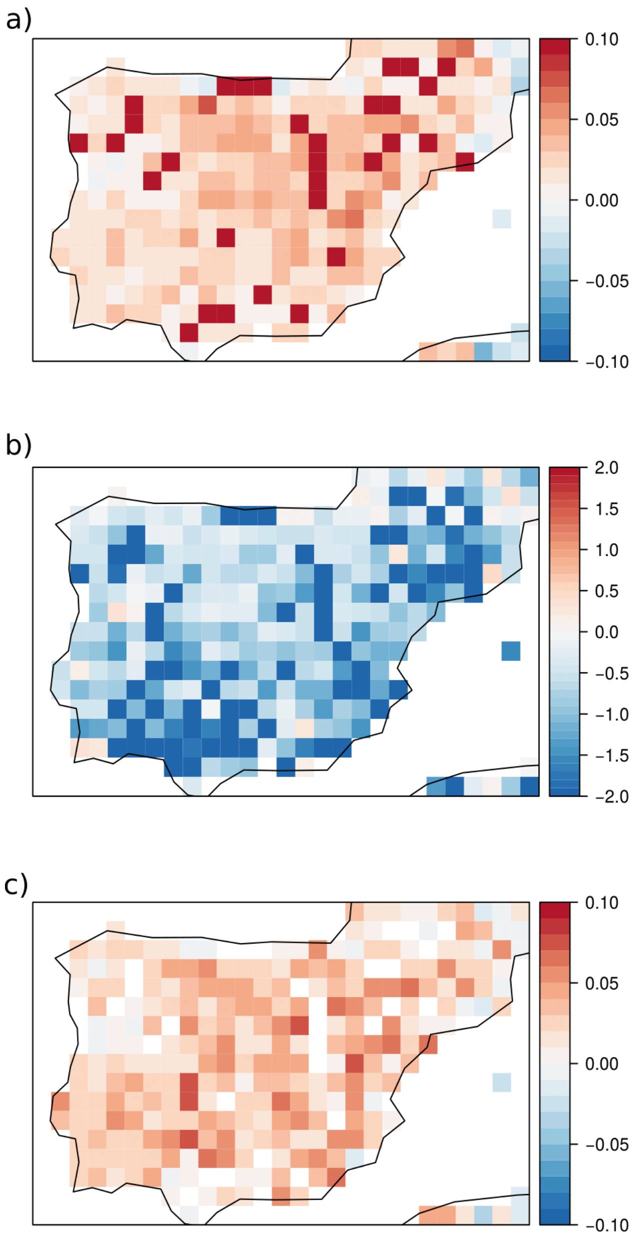

Fig. 3. Differences in the a) ROCSS, b) RMSE and c) Spearman correlation over the Iberian region between the CNN-MS and the CNN-SS models.

Further work will consist in investigating the limits of the predictand's domain or how the addition of completely different climatological sites affects the downscaling. This will be done under the international initiative CORDEX (Coordinated Regional Climate Downscaling Experiment), where one of the objectives is to provide downscaled climate change projections over the entire globe.

ACKNOWLEDGMENTS

We 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). Funding for the authors 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, Jan. 2018. Google-Books-ID: AMhJDwAAQBAJ.

[2] E. Zorita and H. von Storch, "The Analog Method as a Simple Statistical Downscaling Technique: Comparison with More Complicated Methods," *Journal of Climate*, vol. 12, pp. 2474–2489, Aug. 1999.

[3] A. J. Cannon, "Probabilistic Multisite Precipitation Downscaling by an Expanded BernoulliGamma Density Network," *Journal of Hydrometeorology*, vol. 9, pp. 1284–1300, Dec. 2008.

[4] S. Tripathi, V. V. Srinivas, and R. S. Nanjundiah, "Downscaling of precipitation for climate change scenarios: A support vector machine approach," *Journal of Hydrology*, vol. 330, pp. 621–640, Nov. 2006.

[5] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[6] A. F. Agarap, "Deep Learning using Rectified Linear Units (ReLU)," *arXiv:1803.08375 [cs, stat]*, Mar. 2018. arXiv: 1803.08375.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, pp. 84–90, May 2017.

[8] 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.

[9] T. Vandal, E. Kodra, S. Ganguly, A. Michaelis, R. Nemani, and A. R. Ganguly, "DeepSD: Generating High Resolution Climate Change Projections through Single Image Super-Resolution," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '17*, (Halifax, NS, Canada), pp. 1663–1672, ACM Press, 2017.

[10] J. Baño-Medina, R. Manzanas, and J. M. Gutiérrez, "Configuration and Intercomparison of Deep Learning Neural Models for Statistical Downscaling," *Submitted to Geoscientific Model Development*, 2019.

[11] D. Maraun, M. Widmann, J. M. Gutiérrez, 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," *Earth's Future*, vol. 3, pp. 1–14, Jan. 2015.

[12] J. M. Gutiérrez, 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. Pagé, "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*, vol. 39, pp. 3750–3785, 2019.

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

[14] R. Caruana, “Multitask Learning,” p. 35.

[15] “The ERAInterim reanalysis: configuration and performance of the data assimilation system - Dee - 2011 - Quarterly Journal of the Royal Meteorological Society - Wiley Online Library.”

[16] R. C. Cornes, G. van der Schrier, E. J. M. van den Besselaar, and P. D. Jones, “An Ensemble Version of the E-OBS Temperature and Precipitation Data Sets,” *Journal of Geophysical Research: Atmospheres*, vol. 123, pp. 9391–9409, Sept. 2018.

[17] P. M. Williams, “Modelling Seasonality and Trends in Daily Rainfall Data,” p. 7.

[18] J. Bedia, J. Baño-Medina, M. N. Legasa, M. Iturbide, R. Manzanas, S. Herrera, D. San Martín, A. S. Cofiño, and J. M. Gutiérrez, “Statistical downscaling with climate4R: Contribution to the VALUE intercomparison experiment,” *Submitted to Geoscientific Model Development*, 2019.

[19] M. Iturbide, J. Bedia, S. Herrera, J. Baño-Medina, J. Fernndez, M. D. Fras, R. Manzanas, D. San-Martn, E. Cimadevilla, A. S. Cofio, and J. M. Gutiérrez, “The R-based climate4r open framework for reproducible climate data access and post-processing,” *Environmental Modelling & Software*, vol. 111, pp. 42–54, Jan. 2019.