in C. Chen et al. (Eds.), 2018: Proceedings of the 8th International Workshop on Climate Informatics (CI2018). NCAR Technical Note NCAR/TN-550+PROC, doi: 10.5065/D6BZ64XQ

8th International Workshop on Climate Informatics
September 19-21, 2018
CI 2018
Hosted by the National Center for Atmospheric Research in Boulder, CO

DEEP CONVOLUTIONAL NETWORKS FOR FEATURE SELECTION IN STATISTICAL DOWNSCALING

Jorge Baño-Medina[1] , José Manuel Gutiérrez[1]

***Abstract*—The potential of (deep) convolutional neural networks for automatic predictor selection in statistical downscaling over large continental domains is analyzed focusing on a simple illustrative example (precipitation occurrence). It is shown that these models automatically handle redundancy and perform geographical and variable selection/transformation of predictors in a robust and spatially consistent form, obtaining similar features for different predictor sets. Results are compared with best performing standard methods from the largest-to-date intercomparison of statistical downscaling methods (VALUE) using "perfect" reanalysis predictors.**

I. Motivation

Statistical downscaling (SD) is a methodology widely used to overcome the coarse resolution of Global Climate Models (GCM) and to provide climate information at local scales, which is crucial for decision-makers [1][2]. Under the perfect prognosis approach, SD bridges this gap by learning an empirical transfer function between large-scale atmospheric variables (predictors) and the local variable of interest (predictand) using "observed" data (reanalysis and observations, respectively). A variety of techniques have been applied to infer the transfer functions, including both linear (see [3] for a review) and nonlinear techniques such as SVMs [4] and neural networks [5][6]. Predictors must be carefully selected (both the particular variables and the geographical region of influence) in order to 1) reproduce the observed climatology in present climate conditions and 2) generalize to an out-sample dataset (cross-validation). Predictor selection is typically a time-consuming screening process where different combinations of variables and geographical domains are tested [7].

An extra difficulty for predictor selection is the high spatial correlation of large-scale variables, which may add unnecessary degrees of freedom to the model, making it prone to overfitting. To avoid this problem, SD methods normally build on feature compression —e.g. principal component analysis (PCs) [7][3]— and/or on feature selection —e.g. stepwise-based or regularization [6]—. The latter are straightforward and easy to apply (they typically work at a gridbox level), but provide no spatial consistency across the stations. On the other hand, PCs provide robust predictors (leading spatial variability modes) and introduce spatial consistency in the downscaled results. However, when dealing with large regions (e.g. Europe), they are typically computed separately over a number of predefined subregions (see, e.g. Fig. 1) in order to capture regional spatial variability, thus complicating the analysis and loosing spatial consistency.

As a result, predictor selection is one of the main limiting factors for the application of transfer-function SD techniques consistently over large (continental) domains. This is a timely topic due to the undergoing initiatives providing a framework for downscaling activities around the world (e.g. CORDEX [8]).

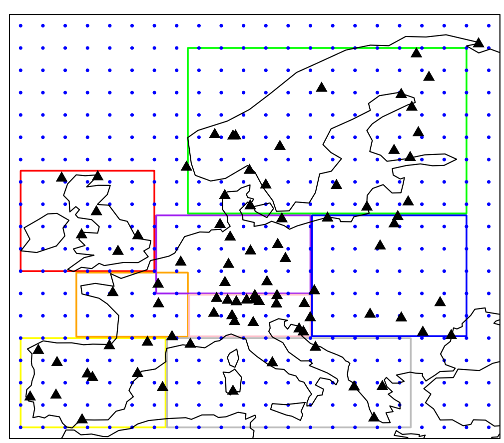

Fig. 1. Geographical domain indicating the renalysis' gridboxes (2° resolution, blue dots) and the 86 VALUE stations (black triangles). The coloured boxes show the eight PRUDENCE regions used to subdivide the domain by some VALUE downscaling methods [3].

Correspondence: Jorge Baño-Medina, bmedina@ifca.unican.es
[1]Santander Meteorology Group, Institute of Physics of Cantabria (CSIC-UC)

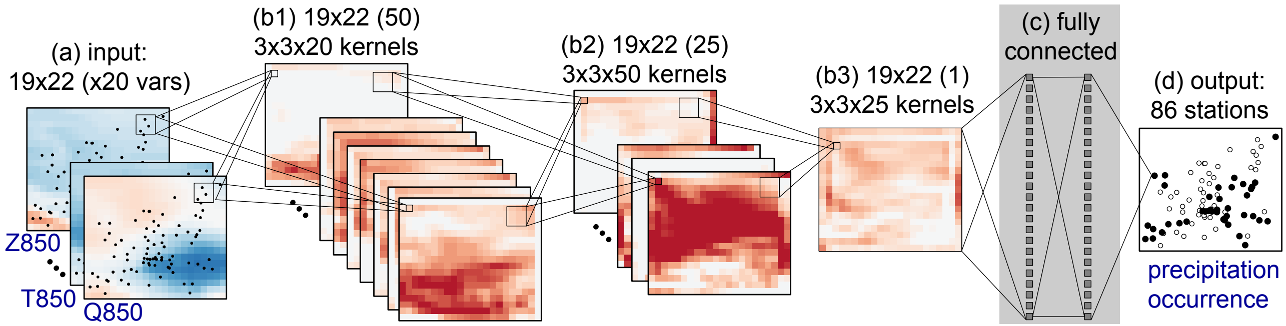

Fig. 2. Optimum neural network architecture for downscaling precipitation occurrence over 86 stations in Europe (output, panel d) based on twenty large-scale European-wide predictors (a). Three convolutional layers (b1-3) with 50, 25 and one 3×3 kernels, respectively, were considered, together with a fully connected layer (c). ReLU non-linear functions are used in all layers, except for the outputs, which are sigmoid for classification. Note that a padding of three zero-value gridboxes was used in the convolutional layers in order to keep the original 19×22 size of the predictors in the different layers.

In this paper we analyze the potential of convolutional neural networks for this problem, and show that they perform automatic selection/transformation of predictors —both the (geographical) domain and the variables,— in a robust and spatially consistent form. We consider the experimental validation framework recently developed by the VALUE European initiative [9][3], where over fifty SD methods were intercompared (cross-validation for the period 1979-2008) over 86 stations across Europe (see Figure 1), considering twenty standard predictors (geopotential height Z, U and V wind components, temperature T and humidity Q, all at 1000, 850, 700 and 500 mb) from the ERA-Interim reanalysis over a $19{\times}22$ $2°$ resolution gridboxes domain covering Europe.

Here we illustrate the simple case of downscaling precipitation occurrence (using a threshold value of 1mm to separate wet and dry days) and focus on the problem of automatic feature selection at a continental domain level.

II. METHODS

Convolutional Neural Networks [10] were introduced as a particular type of neural network specific for grid-structured inputs, such as images. These models were later refined with the new deep learning methodological developments (stochastic gradient descend, dropout, etc.), frameworks (here we use TensorFlow), and technologies (e.g. GPUs), boosting the development of groundbreaking models in different areas [11]. In the case of meteorology and climate, deep learning has been used to estimate cyclone's intensity [12], to detect extreme weather events [13], to emulate subgrid parameterizations [14], and to downscale climate information [15], among others. The latter work established 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 work we use different neural network architectures combining both convolutional and fully connected layers (see Fig.2). The successive convolutional layers provide regional spatial features optimized to jointly downscale precipitation at the different stations (transfer learning). The dense layer builds on these features to learn a downscaling function for each of the stations. We tested different configurations for the three convolutional layers scheme shown in Fig.2, varying the size of the kernel from 3×3 to 9×9 (with a stride of one) and the number of kernels in each layer from 250 to 25 (denoted as $250 - 25$), $50 - 5$ and $10 - 1$, respectively. Different configurations of the dense layer were also tested, with $100 - 10$ and $50 - 5$ neurons, respectively.

We used ReLu activation functions (with He's initialization [16]) in the convolutional layers and a sigmoidal function (with Xavier normal initialization [17]) in the output layer. In order to preserve the original resolution of the atmospheric fields (19×22) a padding was applied to the network. The Adam optimizer [18] learning algorithm was used with a batch size of 100. The learning process was performed splitting the data in different datasets: the training/validation set with 90%/10% of randomly selected days from the period 1985-2008 (resulting into 6567 and 730 samples for training and validation, respectively) and the test set (2192 days that belong to the period 1979-1984). We have used the validation dataset to perform early-stopping according to the Relative Operating Characteristics Skill Score (2 * ROCArea - 1) or ROCSS [19], which ranges between -1 (perfectly wrong prediction) to 1 (prefect prediction), with zero indicating no skill (the same as the climatology).

As benchmark we use one of the good performing SD methods contributing to VALUE (method 'GLM' in [3]), based on Generalized Linear Models (logistic regression for precipitation occurrence) using combined PCs for a subset of seven predictors (Z1000, Z850, Z700, T850, T700, Q850, Q700) over 8 subregions covering Europe (see Figure 1).

III. RESULTS

Optimum ROCSS results were found with 50, 25 and 1 3×3 kernels, as shown in Fig.2. The different dense layers exhibited similar results, with slightly smaller test and larger train errors than the case with no dense layer (e.g. connecting the output directly to the convolutional spatial features). This could be the consequence of mixing the regional features in the dense neurons, thus impeding site focalization which is crucial for this task. Therefore, we chose a pure convolutional architecture in this example, thus yielding better interpretability. Fig. 3 shows the ROCSS as a function of (a) the epochs (to analyze the training process, with train/validation results indicated by solid/dashed lines) and (b) the stations for the following models: 1) the benchmarking seven-predictor GLM, 2) the optimum CNN models with 1 and 10 kernels in the last layer (CNN1 and CNN10, respectively) both using the same seven predictors, and 3) a CNN1 model trained with the full predictor dataset (twenty variables, instead of the seven used for the GLM method). Results indicate that CNN models outperform the GLM, particularly in bad (GLM) performing stations. Moreover, a single final kernel (CNN1) achieves the best results, whereas CNN10 exhibits fast growing training errors and smaller test results. It is also remarkable that the model can manage redundancy, achieving the best results with the full set of predictors.

One of the nice features of CNNs is that they automatically optimize the geographical area of influence for the different stations. In our simple architecture this information is encoded in the weights from the last convolution layer to the outputs (the 86 stations). As an example, Fig. 4 shows these weights for two illustrative stations: Madrid and Helsinki. We observe that the relevant geographical information is automatically determined as a Wester pattern for Madrid, and an East to West gradient for Helsinki.

Finally, we can also analyze the robustness of CNN downscaling models by comparing the convolutional predictor resulting from different predictor sets —for instance, the two cases with seven (partial information) and twenty (full information) input variables analyzed

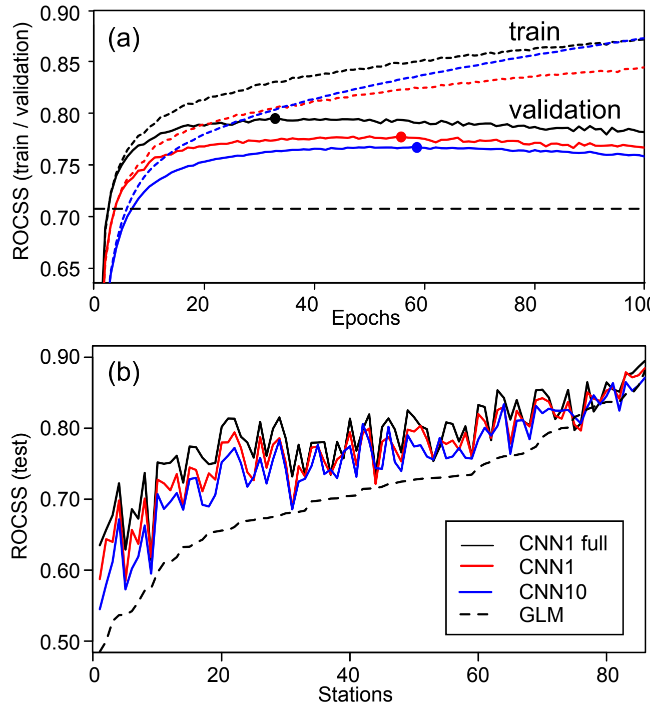

Fig. 3. Results (ROCSS) of different downscaling models as a function of (a) the epochs (train/validation results are indicated by solid/dashed lines; the dots indicate early stopping) and (b) the stations (sorted according to the GLM test results).

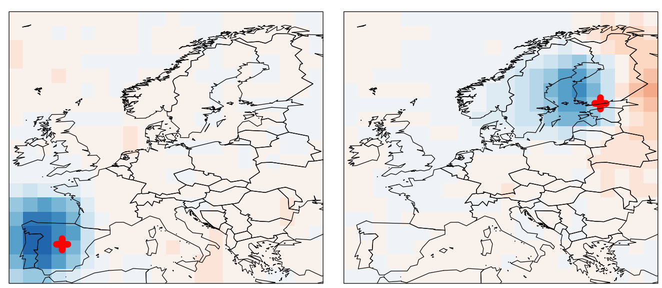

Fig. 4. Weights from the last convolution layer to the outputs (the stations) for Madrid (left) and Helsinki (right). A 5×5 spatial moving average is applied to represent the effect of kernels. Blue/red colors indicate positive/negative weights.

in Fig. 3.— Fig. 5 shows the spatial patterns of the original predictor variables and the features (at the bottom) resulting from these two CNN models for a particular illustrative date. It can be easily seen that similar patterns are learned in both cases, resembling mainly the humidity (Q) and pressure (Z), which are the most influential variables in this task [3]. Moreover, the features of the full model carry also information from the rest of explanatory variables, yielding a slightly different spatial pattern and attaining better results, as shown in Fig. 3.

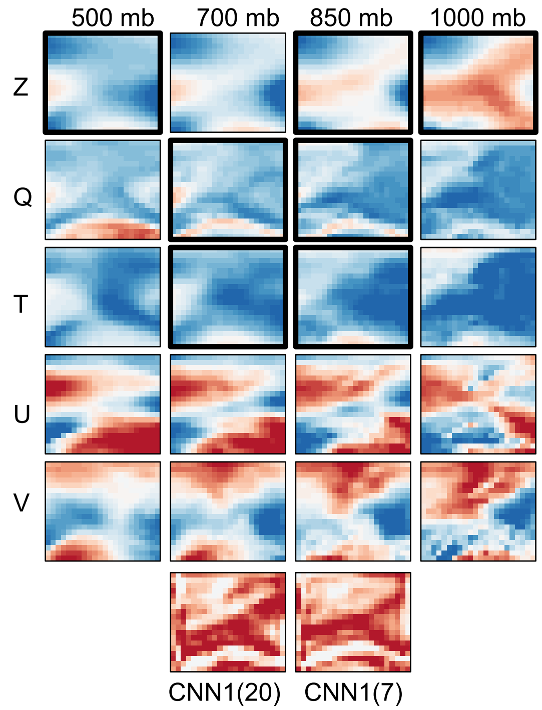

Fig. 5. Twenty predictors considered in VALUE (above) and the resulting features in the last convolution layer (below) for the CNN1(7) model trained with seven variables (those with thick boxes) and with the full twenty predictor set, CNN1(20).

IV. CONCLUSIONS

We have shown that (deep) convolutional neural networks provide a suitable statistical downscaling methodology for large (continental) domains, handling redundancy and performing geographical and variable selection/transformation. The analysis have focused on a simple illustrative example (precipitation occurrence) and promising results are obtained from the comparison of state-of-the art statistical downscaling methods (from the VALUE initiative). Further analysis is required in order to check the general validity of the results.

ACKNOWLEDGMENTS

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

REFERENCES

[1] D. Maraun, F. Wetterhall, A. M. Ireson, R. E. Chandler, E. J. Kendon, M. Widmann, S. Brienen, H. W. Rust, T. Sauter, M. Themel, V. K. C. Venema, K. P. Chun, C. M. Goodess, R. G. Jones, C. Onof, M. Vrac, and I. Thiele-Eich, "Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user," vol. 48, no. 3, pp. n/a–n/a.

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

[3] J. M. Gutiérrez, D. Maraun, M. Widmann, and et al., "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*, 2018.

[4] S. Tripathi, 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.

[5] 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.

[6] 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.

[7] D. San-Martín, R. Manzanas, S. Brands, S. Herrera, and J. M. Gutiérrez, "Reassessing Model Uncertainty for Regional Projections of Precipitation with an Ensemble of Statistical Downscaling Methods," *Journal of Climate*, vol. 30, pp. 203–223, Oct. 2016.

[8] 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.

[9] 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.

[10] 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.

[11] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural networks*, vol. 61, pp. 85–117, 2015.

[12] 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.

[13] 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.

[14] S. Rasp, M. S. Pritchard, and P. Gentine, "Deep learning to represent subgrid processes in climate models," *Proceedings of the National Academy of Sciences*, p. 201810286, Sept. 2018.

[15] 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.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE international conference on computer vision*, pp. 1026–1034, 2015.

[17] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pp. 249–256, 2010.

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

[19] I. Mason, "A model for assessment of weather forecasts," *Aust. Meteor. Mag*, vol. 30, no. 4, pp. 291–303, 1982.