Understanding Deep Learning Decisions in Statistical Downscaling Models

Jorge Baño-Medina
bmedina@ifca.unican.es
Institute of Physics of Cantabria
Santander, Cantabria, Spain

ABSTRACT

Deep learning (DL) models are progressively being applied to climate applications due to their ability to learn complex nonlinear spatiotemporal patterns, typically present in the atmosphere. In particular, deep learning has landed on the downscaling field, providing high-resolution climate change projections crucial for sectorial applications. Despite their merits, they are still seen as "black boxes" generating distrust among the climate community and thus limiting their use in real applications. Therefore, there is a need to develop techniques that unravel the knowledge hidden in the neural models to 1) gain understanding about their decisions and 2) make these models more reliable to the community. In this study, we adopt a technique used in computer vision to visualize the decisions, to convolutional-based downscaling models. The results show comprehensive links learned by the network connecting the large-scale to the local-scale and prove the implicit feature selection that occurs within the hidden layers. To our knowledge, this is the first study that properly assesses a methodology to unravel the "black box", in particular information concerning the predictor-predictand link, in a downscaling application.

CCS CONCEPTS

• **Applied computing → Physics**.

KEYWORDS

neural networks, deep learning, statistical downscaling, understanding

ACM Reference Format:

Jorge Baño-Medina. 2020. Understanding Deep Learning Decisions in Statistical Downscaling Models. In *10th International Conference on Climate Informatics (CI2020), September 22–25, 2020, virtual, United Kingdom.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3429309.3429321

1 INTRODUCTION

In the last years, deep neural networks have been proposed for a variety of climate applications, from detecting extreme weather events in climate datasets [16], to hurricane tracking [1] or precipitation nowcasting [26], among others. The success of these first attempts together with the breakthroughs achieved by neural networks in other disciplines, has driven the climate community to outline future research benefiting from the spatio-temporal learning properties of deep models [22]. However, in a physical discipline as climate, neural networks are yet seen as "black boxes" generating distrust among scientists. To this end, a recent Nature publication [22] encouraged to research towards the understanding of deep nets in climate science, to make these models more reliable to the community. To date, very few studies in climate have dealt with interpretability [14],[17],[20]. To our knowledge, only one study especially focuses on this topic [20] that introduces a variety of interpretability techniques of different nature (permutation-based predictor importance, novelty detection, forward and backward selection, and "saliency" maps, among others) to a bundle of meteorological applications. Concerning the predictor-predictand link, the mentioned attempts basically rely on saliency maps that depict the influence of input variables on the output by simply computing the partial derivatives [27] or by class activation mapping [31]. Recently more sophisticated methods relying on prediction difference analysis [32] or causal inference [23] have proved successful in a variety of computer vision applications. Therefore proper algorithms and especially devoted research towards interpretability are still insufficient for climate applications.

In particular, deep learning has strongly landed on a well-known climate application called statistical downscaling [29],[24],[4], by learning nonlinear empirical relationships between large-scale atmospheric variables (predictors) and a record of historical observations at finer scales (predictands). Therefore, statistical downscaling (SD, see [18] for a review) aims to bridge the gap between the coarse resolution given by Global Climate Models (GCMs) and the local scale to provide stakeholders with regional climate information. The interpretability of the models is a relevant concern in statistical downscaling, especially when they are intended to downscale climate change projections, where several sources of uncertainty already exist (e.g., limitations of the spatial resolution and deficient parameterizations, among others).

To this end, several attempts of different nature have been proposed to gain comprehension regarding this application. In particular, different deep learning topologies of increasing levels of complexity were intercompared [4] in the largest-to-date downscaling intercomparison study [19], to gain understanding on the role of the elements typically conforming a neural net (e.g., convolutional and/or dense layers, nonlinearity in presence of ReLu activation functions). The latter study showed how convolutional model (see [15] for more details on this kind of models) can automatically deal with high-dimensional predictor spaces contrary to conventional SD techniques (e.g., generalized linear models or analogs [9]) that

CI2020, September 22–25, 2020, virtual, United Kingdom

ACM ISBN 978-1-4503-8848-1/20/09…$15.00
https://doi.org/10.1145/3429309.3429321

need to undergo restrictive feature selection/reduction steps (e.g., principal component analysis, Lasso regression). In addition, CNN models explained better the local variability, including the reproducibility at the extremes, due to their ability to infer nonlinear links present in the task. Further research was conducted to shed light on the internal behavior of convolutional topologies, especially regarding their extrapolation ability [3], by downscaling climate change scenarios, and the nature of their implicit regularization (i.e., responsible for their ability to operate in high-dimensional downscaling tasks) partially linked to the inductive bias of multitask models [2]. Other studies focus on the predictor-predictand link, such as [14]. Though only valid for very particular deep learning (DL hereafter) topologies, [14] uses class activation mapping [31] to visualize the importance of the predictor domain in the downscaling. Despite these few attempts, our knowledge of the internal behavior of DL models in downscaling is still very limited, and there are important questions that remain unanswered.

In this article, we focus on the predictor-predictand link and adapt a technique used in computer vision models to visualize their decisions (namely saliency maps), based on prediction difference analysis [32], to any SD neural-based model. In particular, we focus on the top-ranked deep models developed in [4] to downscale precipitation and temperature. Through this mechanism, we shed light on its internal behavior by measuring the influence of both predictor's domain and variables in the downscaling. Therefore the resulting saliency maps provide comprehensive information concerning the predictor-predictand link inferred by the deep model. To our knowledge, this is the first study that is specially designed to understand the decisions made by a neural network in a downscaling application. This will contribute to gain confidence in these models which in turn will help to maximize the potential benefits derived from the climate-DL synergy.

The paper is organized as follows: Sec.2 delves into the mathematics of neural networks (Sec.2.1) and prediction difference analysis (Sec.2.2). In Sec.3 and Sec.4 we describe the data and methodology used, respectively. For transparent science, we devote a section for reproducibility of the current study (Sec.5). Sec.6 discusses the resulting saliency maps and in Sec.7 we present the conclusions.

2 METHODS

In this section we first overview the basic concepts of neural networks emphasizing their particularities for the downscaling of temperature and precipitation (Sec. 2.1), followed by an explanation of the prediction difference analysis technique [32] and its mathematical adaptation to the variables of study (Sec. 2.2).

2.1 Deep Learning for Downscaling

Given a set of predictor variables x, neural networks aim to learn a nonlinear relationship $f^{\omega}(x)$, parameterized by a set of coefficients ω, that explains the predictand variable y (see [15] for a review). The set of coefficients ω is updated to fit the data according to a loss function (e.g., mean squared error) in an iterative learning process relying on the backpropagation algorithm and the gradient descent method. Therefore, it is a regression-based technique that infers nonlinear patterns among variables, being capable to approximate, in practice, any function.

$$y = f^{\omega}(x) \tag{1}$$

The output of the network is directly related to the loss function used to optimize the parameters. Due to the different nature of the meteorological variables, it is common to minimize the negative log-likelihood of the distribution that best fits the variable of interest (e.g., Gaussian's for temperature and Bernoulli-Gamma's for precipitation). Eq.1 is a valid formulation of the downscaling of temperature since minimizing the mean squared error is equivalent to minimize the negative log-likelihood of a Gaussian distribution (note that in this case, we would estimate the mean, μ, and variance, σ^2, of the conditional Gaussian distribution, hence $[\mu, \sigma^2] = f^{\omega}(x)$). However for precipitation, better fits are obtained if we minimize the negative log-likelihood of a Bernoulli-Gamma distribution [5],[4]. In this case, the net estimates the parameter p of a Bernoulli distribution along with the shape and the scale parameters of a Gamma distribution given by α and β, respectively. We reformulate Eq.1 such that:

$$[p, \alpha, \beta] = f^{\omega}(x) \tag{2}$$

2.2 Prediction Difference Analysis

In order to understand the decisions made by the deep models in terms of predictor importance in a downscaling framework, we rely on a visualization technique successfully tested in computer vision applications called prediction difference analysis [32]. The main idea behind [32] is to estimate the relevance of an input feature, x_j, by measuring how the predicted parameters, y (i.e., $y = \{p, \alpha, \beta\}$ for precipitation or $y = \{\mu, \sigma^2\}$ for temperature), change when it is unknown, y' (i.e., $y' = \{p', \alpha', \beta'\}$ for precipitation or $y' = \{\mu', \sigma^{2'}\}$ for temperature). This can be done by marginalizing the feature j:

$$y' = f^{\omega}(x_{\setminus j}) = \Sigma_{x_j} p(x_j|x_{\setminus j}) f^{\omega}(x_j, x_{\setminus j}) \tag{3}$$

Where $x_{\setminus j}$ refers to the set of all input features except x_j. According to [32], the probability function $p(x_j|x_{\setminus j})$ is adjusted with a conditional multivariate gaussian distribution. However, approximating the latter is normally not feasible and in [32] they simplify the term by conditioning only on a surrounding region of size LxL, given by $\hat{x}_{\setminus j}$, therefore retaining the local spatial correlations. Moreover, a multivariate analysis can be carried out by removing jointly a set of z features grouping a patch of size KxK. Hence,

$$y' = f^{\omega}(x_{\setminus z}) = \Sigma_{x_z} p(x_z|\hat{x}_{\setminus z}) f^{\omega}(x_z, x_{\setminus z})) \tag{4}$$

Fig.1 provides a simple and illustrative scheme explaining the above mentioned feature marginalization estimation for the domain of interest of this study.

In order to provide a quantitative value of the influence of x_z on a particular sample (i.e., day) i, we compute the difference in the expected value of the conditional daily distributions, or activation difference (hereafter AD), at the output locations of interest, hence:

$$AD_i = \mu_i' - \mu_i \tag{5}$$

$$AD_i = p_i'\alpha_i'\beta_i' - p_i\alpha_i\beta_i \tag{6}$$

where Eq.5 and Eq.6 are applicable to temperature and precipitation model outputs, respectively. Finally, we recover the influence on x_j by averaging the AD values over the x_z set of features containing x_j.

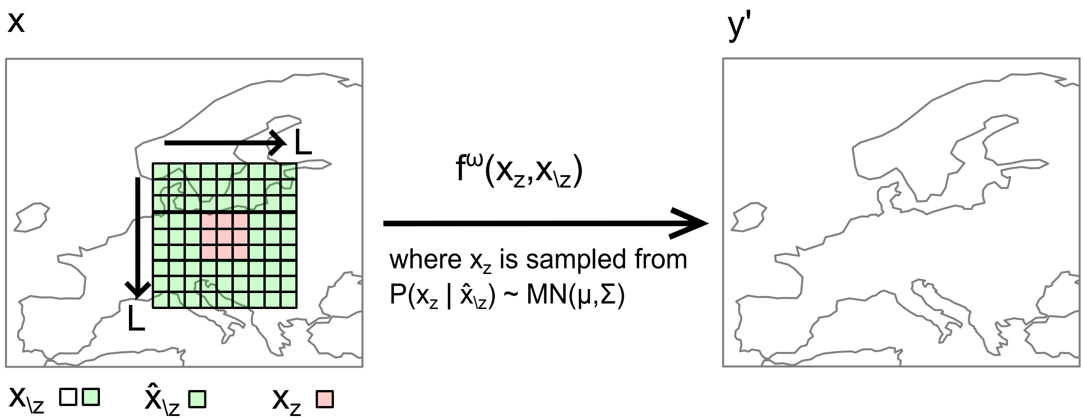

Figure 1: Scheme illustrating the marginalization of x_z of size KxK (red) conditioned on its surrounding region of size LxL, $\hat{x}_{\backslash z}$ (green), given an input feature channel x (for simplicity we only depict the case for 1 feature map). The input features, x_z and $x_{\backslash z}$, are fed to the neural network f^{ω}, where x_z is drawn from the multivariate normal (MN) distribution $P(x_z|\hat{x}_{\backslash z})$, described by μ and Σ, which are the estimated conditional parameters. This process is repeated by centering x_z in the rest of the predictor gridpoints.

3 DATA

In this study, we visualize the decisions for the top-ranked deep model developed in [4], which was framed in the cost action VALUE, a European initiative that provides an experimental framework for statistical downscaling [19]. According to VALUE, the predictor data used consists of 5 variables (temperature, specific humidity, geopotential and both zonal, and meridional wind) at 4 vertical levels (i.e., 250, 500, 750 and 1000 hPa), coming from the ERA-Interim reanalysis at a daily resolution. The predictor domain covers an area ranging from 36° to 72° in latitude and from −10° to 32° in longitude, resulting into a 19x22(x20, including the variables) regular grid. The models inferred an empirical relationship linking ERA-Interim's [7] low-resolution large-scale variables (i.e., 2°) to the E-OBS [6] high-resolution daily precipitation/temperature (i.e., 0.5°) for the period 1979-2008. We split the data on a train (1979-2002) and a test (2008) fold.

4 METHODOLOGY

The deep model that ranked best in [4] (hereafter, CNN) consisted of three convolutional hidden layers of 50, 25 and 1 (or 10 for temperature) feature maps, respectively, along with an input layer of 20 channels corresponding to the variables indicated in Sec. 3. The downscaling was only done over the E-OBS gridpoints located over land, leading to 3258 sites (note that for precipitation we estimate 3 parameters per predictand location, Eq.2, and thus the output layer consists on 9774 neurons). For more details on the architecture, we refer the reader to the original manuscript [4] and/or to the Santander Meteorology Group GitHub repository, that includes the full code to replicate the results.

Therefore the objective is to compute the parameters in Eq.1 (or Eq.2 for precipitation) and Eq.4 for the deep model described above. According to Sec.2, we estimate the parameters y' using a patch size of $K = 3$ centered on a pixel j, and average over $N = 30$ samples of a multivariate distribution conditioned on a region of size $L = 11$. Since our input layers consists on 20 large-scale atmospheric variables with a 19x22 latitude-longitude structure we marginalize over a total of 8360 features, obtaining the influence of every input neuron on the downscaling. Finally, we evaluate the predictor importance in the downscaling over a particular site, by plugging the outputs of Eq.1 (or Eq.2) and 4 in Eq.5 (or Eq.6).

In particular, we focus in 4 locations with different climatologies (or more precisely the predictands' gridpoints which are the closest to their coordinates): Paris, Rome, Copenhagen and a location in the Alps.

5 REPRODUCIBILITY

For transparency, we provide a Jupyter notebook with the code necessary to replicate the results presented herein, which can be found in the Santander Meteorology GitHub repository. We rely on climate4R [13], a set of R packages designed to handle climate data (e.g., loading, regridding, downscaling). In particular, we build on the climate4R package downscaleR.keras which integrates Keras (the state-of-the-art in deep learning libraries) into the climate4R framework. This package contains specific functions to infer the saliency maps for Keras models. We use a 64-bits machine with processor Intel® Core™ i7-6700 CPU @ 3.40GHz x 8 and a RAM memory of 15.6 GiB.

6 RESULTS

In this section, we analyze the saliency maps obtained using the prediction difference analysis technique on the CNN model. In Sec.6.1 we discuss the results for temperature whereas we present the analysis of precipitation in Sec.6.2.

6.1 Predictor Importance for Downscaling Temperature

In Fig.2 we depict the saliency maps per variable for a) Paris, b) Rome, c) Copenhagen and d) Alps. Every panel contains the averaged absolute (to avoid compensation) AD values over the test period for the set of 20 variables indicated in Sec.3. According to Fig.2, the temperature at 1000hPa is the most important variable in the downscaling with a notable difference in comparison with the other variables which have very little (e.g., the geopotential height at 700 and 1000 hPa in Paris) or null influence (e.g., the wind velocity, both meridional and zonal, at all locations). This indicates that the net has learned a very dependent link to the low-resolution temperature at 1000 hPa. This connection is in agreement with other studies that focus on predictor selection for statistical downscaling over Europe. Traditionally, downscaling techniques suffer from overfitting if fed with an undesirable large amount of predictors. To this end, several studies have studied the best set of predictors to downscale temperature over Europe, relying on sophisticated feature selection (e.g., stepwise, predictor-predictand correlation) or feature reduction techniques (e.g., principal component analysis [21]). These studies showed that the temperature was the main driver of the local temperature in the downscaling of both reanalysis [10][11] and global climate models [12]. Despite this accurate predictor sensitivity analysis, "human-guided" manipulation of the predictor space has not eliminated the uncertainties present in the choice of predictors having an impact in the climate change estimates [12]. Recently, the largest-to-date downscaling study intercompares different sets of predictors to address this source of uncertainty [9]. It was under this intercomparison study where CNN demonstrate exceptional ability to reproduce the local variability even though it was not subjected to any feature selection or reduction technique [4]. The analysis carried out on this study enables us to visualize the automatic feature selection (and even feature reduction since the hidden layers compress the input space into learnable non-linear subspaces) of the CNN. Therefore, the strong link between the large-scale and local-scale temperature is supported by the literature and has automatically been learned by the convolutional network.

Moreover, the gridpoints that influence the most the downscaling are located around the location of interest almost neglecting the influence of the rest of the predictor domain. The variation of this "relevant" spatial domain across sites is very little, remaining very similar in both shape and size. According to Fig.2, these areas typically encompass an area of 4x4 gridboxes.

6.2 Predictor Importance for Downscaling Precipitation

Fig.3 shows the saliency maps obtained in the CNN downscaling model of precipitation. As Fig.2, it consists of 4 panels (Paris, Rome, Copenhagen and Alps) containing the averaged absolute AD values over the test period for the set of 20 variables. Differently to temperature, in this case, we observe that both the relevant spatial domain and variables change across sites. In Paris the geopotential height at 850 hPa seems to be the most relevant variable with a minor (e.g., variables at 1000 hPa) or null influence (e.g., the temperature at 500,700 and 850 hPa) of other variables. In contrast, probably due to the presence of mountains and/or to the influence of the Mediterranean, the wind velocities at 700 and 850 hPa become important over the Alps and Rome (especially the vertical component of the wind in this last case) whilst being almost negligible in Paris and Copenhagen. The specific humidity at 1000 hPa is also relevant for all locations except in Paris where the geopotential height at 850hPa gains all the attention. The latter humidity combined with the temperature at 1000hPa is the main driver of precipitation in Copenhagen. In fact, this is the only location where the temperature becomes actually relevant. Commonly to the case of temperature, the choice of predictors is critical to reproduce daily precipitation and several studies have tried to address this topic and incorporate the uncertainties by comparing different predictor configurations [30][9][28]. However, this aspect in the downscaling setup is even more uncertain when it comes to precipitation [8], especially if we take into account the variability in the predictor importance across regions [28] or according to the selection criteria [30]. Despite this variability, a common set of predictors involves the use of geopotential height, humidity and temperature at different pressure levels [25][9].

The feature selection observed in Figs.2 and 3 in both variable and spatial domain, which is consistent with previous studies focused on the issue of feature selection for statistical downscaling over Europe, together with the posterior manipulation of these features throughout the neural layers into nonlinear hidden sub-spaces, would explain the capacity of CNN to deal with high-dimensional domains without overfitting, already hypothesized in the reference article [4] and confirmed in this study.

7 CONCLUSIONS

To our knowledge, this is the first study that presents a proper methodology to unravel the "black box" of deep learning models in the context of downscaling, which is crucial to gain confidence in these techniques. Overall, the prediction difference analysis technique has permitted us to better understand CNN deep model, in particular information regarding the predictor-predictand link. We conclude that the neural network of study learns adjustable windows optimum for every site, performing an implicit site-dependent feature selection of the input space. The latter is responsible for the ability of CNNs to handle high-dimensional predictor domains without leading to overfitting. The input variables exert a different influence depending on the location and on the predictand variable of interest. In particular, we observe a very strong dependence of the local temperature with the large-scale temperature at 1000 hPa at all sites. Differently, the input features relevant for the downscaling of precipitation are site-dependent, reflecting the variety of drivers that trigger this variable. Besides, the predictor-predictand links visualized through saliency maps containing the relevant input features may be used as prior knowledge for future predictor configurations in other statistical models. This is very relevant since the manipulation of the input space with feature selection or reduction approaches is a very time-consuming process and a high source of uncertainty in well-established statistical downscaling techniques (e.g., analogs or generalized linear models) leading in most cases to a relevant loss of predictor information damaging the downscaling. Regarding other applications, these maps could even

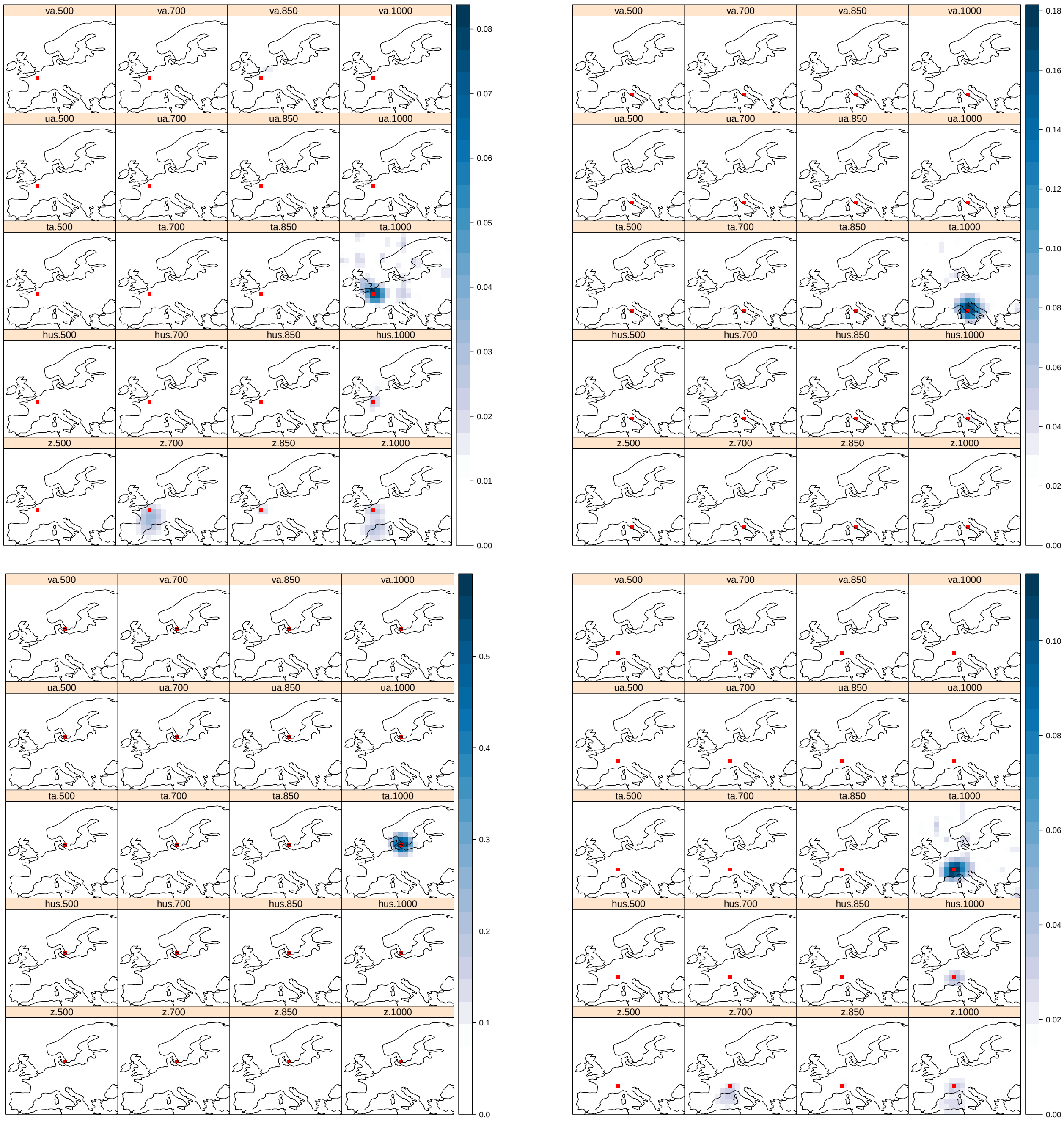

Figure 2: The averaged saliency maps per variable (in absolute values, to avoid compensation) over the test period (366 maps in total), for a) Paris, b) Rome c) Copenhagen and d) Alps. Every pannel contains the saliency maps that depicts the importance of the meridional (va) and zonal (ua) wind velocity, the air temperature (ta), the specific humidity (hus) and the geopotential height (z, in rows) at 500, 700, 850 and 1000 hPa (in columns) in the downscaling of temperature.

serve to unravel unknown links in the atmosphere, contributing not only to gain confidence but also to a better comprehension of the climate system.

ACKNOWLEDGMENTS

We acknowledge the E-OBS dataset from the EU-FP6 project UERRA (http://www.uerra.eu) and the Copernicus Climate Change Service,

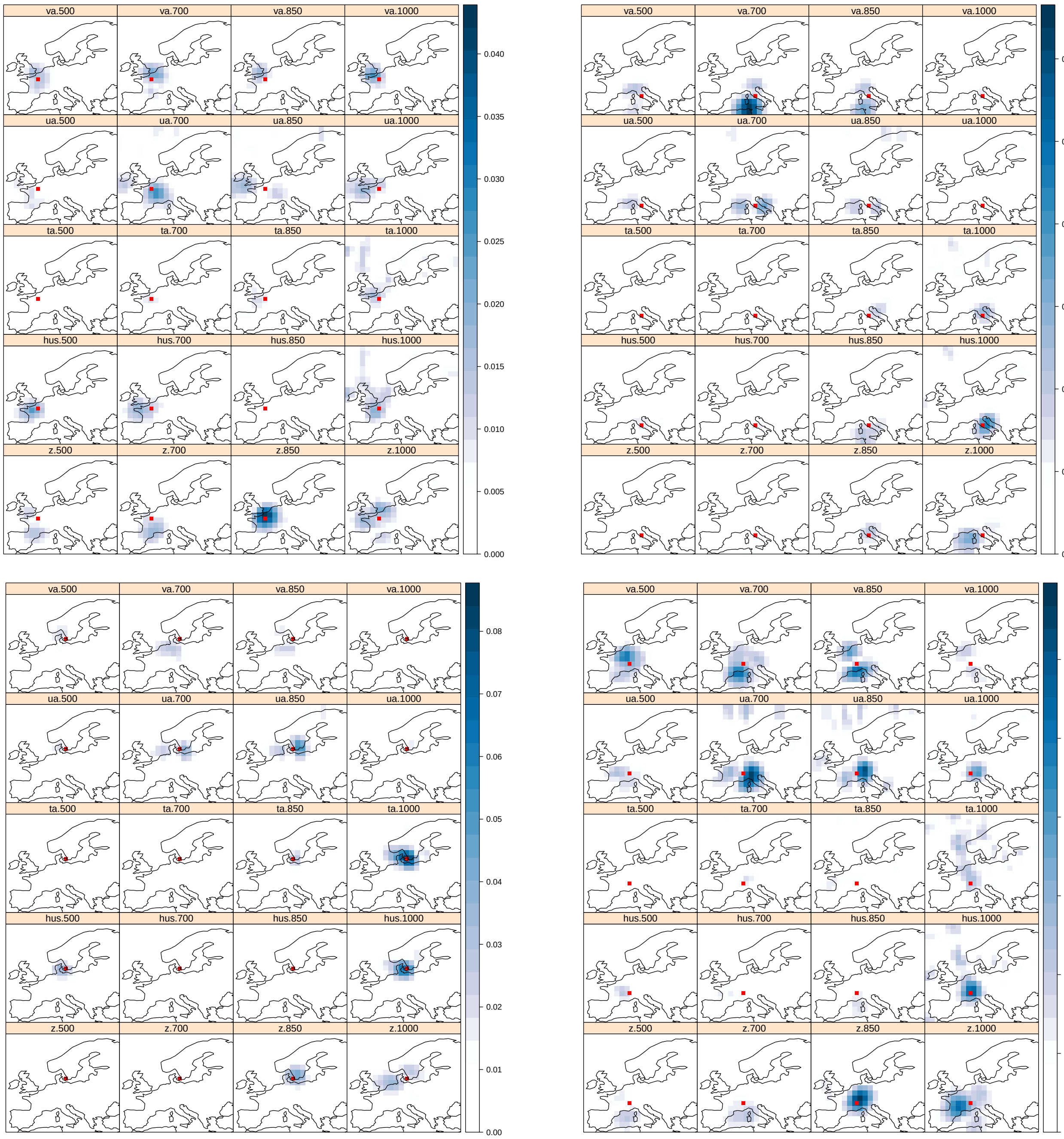

Figure 3: Same as Fig.2 but for the precipitation downscaling models.

and the data providers in the ECA&D project (https://www.ecad.eu). I also acknowledge Dr. Rodrigo García Manzanas and Dr. José Manuel Gutiérrez for their advice during the elaboration of the manuscript.

REFERENCES

[1] Sheila Alemany, Jonathan Beltran, Adrian Perez, and Sam Ganzfried. 2018. Predicting Hurricane Trajectories using a Recurrent Neural Network. *arXiv:1802.02548 [physics, stat]* (Feb. 2018). http://arxiv.org/abs/1802.02548 arXiv: 1802.02548.

[2] Jorge Baño-Medina and José Manuel Gutiérrez. 2019. The Importance of Inductive Bias in Convolutional Models for Statistical Downscaling. In *Proceedings of the 9th International Workshop on Climate Informatics: CI 2019.* https://doi.org/doi:10.5065/y82j-f154

[3] Jorge Baño-Medina, Rodrigo Manzanas, and José Manuel Gutiérrez. [n.d.]. On the suitability of deep convolutional neural networks for downscaling climate

change projections. *Submitted to Climate Dynamics* 4 ([n. d.]).

[4] Jorge Baño-Medina, Rodrigo Manzanas, and José Manuel Gutiérrez. 2020. Configuration and intercomparison of deep learning neural models for statistical downscaling. *Geoscientific Model Development* 13, 4 (April 2020), 2109–2124. https://doi.org/10.5194/gmd-13-2109-2020 Publisher: Copernicus GmbH.

[5] Alex J. Cannon. 2008. Probabilistic Multisite Precipitation Downscaling by an Expanded Bernoulli–Gamma Density Network. *Journal of Hydrometeorology* 9, 6 (Dec. 2008), 1284–1300. https://doi.org/10.1175/2008JHM960.1

[6] Richard C. Cornes, Gerard van der Schrier, Else J. M. van den Besselaar, and Philip D. Jones. 2018. An Ensemble Version of the E-OBS Temperature and Precipitation Data Sets. *Journal of Geophysical Research: Atmospheres* 123, 17 (Sept. 2018), 9391–9409. https://doi.org/10.1029/2017JD028200

[7] D. P. Dee, S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A. J. Geer, L. Haimberger, S. B. Healy, H. Hersbach, E. V. Hólm, L. Isaksen, P. Kållberg, M. Köhler, M. Matricardi, A. P. McNally, B. M. Monge-Sanz, J.-J. Morcrette, B.-K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J.-N. Thépaut, and F. Vitart. 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society* 137, 656 (April 2011), 553–597. https://doi.org/10.1002/qj.828

[8] Guobin Fu, Stephen P. Charles, Francis H. S. Chiew, Marie Ekström, and Nick J. Potter. 2018. Uncertainties of statistical downscaling from predictor selection: Equifinality and transferability. *Atmospheric Research* 203 (May 2018), 130–140. https://doi.org/10.1016/j.atmosres.2017.12.008

[9] J. M. Gutiérrez, D. Maraun, M. Widmann, R. Huth, E. Hertig, R. Benestad, O. Roessler, J. Wibig, R. Wilcke, S. Kotlarski, D. San Martín, S. Herrera, J. Bedia, A. Casanueva, R. Manzanas, M. Iturbide, M. Vrac, M. Dubrovsky, J. Ribalaygua, J. Pórtoles, O. Räty, J. Räisänen, B. Hingray, D. Raynaud, M. J. Casado, P. Ramos, T. Zerenner, M. Turco, T. Bosshard, P. Štěpánek, J. Bartholy, R. Pongracz, D. E. Keller, A. M. Fischer, R. M. Cardoso, P. M. M. Soares, B. Czernecki, and C. Pagé. 2019. An intercomparison of a large ensemble of statistical downscaling methods over Europe: Results from the VALUE perfect predictor cross-validation experiment. *International Journal of Climatology* 39, 9 (2019), 3750–3785.

[10] R Huth. 1999. Statistical downscaling in central Europe: evaluation of methods and potential predictors. *Climate Research* 13 (1999), 91–101. https://doi.org/10.3354/cr013091

[11] Radan Huth. 2002. Statistical Downscaling of Daily Temperature in Central Europe. *JOURNAL OF CLIMATE* 15 (2002), 12.

[12] Radan Huth. 2004. Sensitivity of Local Daily Temperature Change Estimates to the Selection of Downscaling Models and Predictors. *JOURNAL OF CLIMATE* 17 (2004), 13.

[13] M. Iturbide, J. Bedia, S. Herrera, J. Baño-Medina, J. Fernández, M. D. Frías, R. Manzanas, D. San-Martín, E. Cimadevilla, A. S. Cofiño, and J. M. Gutiérrez. 2019. The R-based climate4R open framework for reproducible climate data access and post-processing. *Environmental Modelling & Software* 111 (Jan. 2019), 42–54. https://doi.org/10.1016/j.envsoft.2018.09.009

[14] Pablo Rozas Larraondo, Iñaki Inza, and Jose A Lozano. 2017. Automating weather forecasts based on convolutional networks. *Proceedings of the ICML Workshop on Deep Structured Prediction, PMLR. Vol. 70* (2017).

[15] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *Nature* 521, 7553 (May 2015), 436–444. https://doi.org/10.1038/nature14539

[16] Yunjie Liu, Evan Racah, Prabhat, Joaquin Correa, Amir Khosrowshahi, David Lavers, Kenneth Kunkel, Michael Wehner, and William Collins. 2016. Application of Deep Convolutional Neural Networks for Detecting Extreme Weather in Climate Datasets. (May 2016). https://arxiv.org/abs/1605.01156

[17] Ankur Mahesh, Maximilian Evans, Garima Jain, Mattias Castillo, Aranildo Lima, Brent Lunghino, Himanshu Gupta, Carlos Gaitan, Jarrett K Hunt, Omeed Tavasoli, Patrick T Brown, and V Balaji. [n.d.]. Forecasting El Niño with Convolutional and Recurrent Neural Networks. ([n. d.]).

[18] Douglas Maraun and Martin Widmann. 2018. *Statistical Downscaling and Bias Correction for Climate Research.* Cambridge University Press.

[19] Douglas Maraun, Martin Widmann, José M. Gutiérrez, Sven Kotlarski, Richard E. Chandler, Elke Hertig, Joanna Wibig, Radan Huth, and Renate A.I. Wilcke. 2015. VALUE: A framework to validate downscaling approaches for climate change studies. *Earth's Future* 3, 1 (Jan. 2015), 1–14. https://doi.org/10.1002/2014EF000259

[20] Amy McGovern, Ryan Lagerquist, David John Gagne, G. Eli Jergensen, Kimberly L. Elmore, Cameron R. Homeyer, and Travis Smith. 2019. Making the Black Box More Transparent: Understanding the Physical Implications of Machine Learning. *Bulletin of the American Meteorological Society* 100, 11 (Nov. 2019), 2175–2199. https://doi.org/10.1175/BAMS-D-18-0195.1

[21] Rudolph W. Preisendorfer. 1988. *Principal Component Analysis in Meteorology and Oceanography.* Elsevier.

[22] Markus Reichstein, Gustau Camps-Valls, Bjorn Stevens, Martin Jung, Joachim Denzler, Nuno Carvalhais, and Prabhat. 2019. Deep learning and process understanding for data-driven Earth system science. *Nature* 566, 7743 (Feb. 2019), 195–204. https://doi.org/10.1038/s41586-019-0912-1

[23] Christian Reimers, Jakob Runge, and Joachim Denzler. 2019. Using Causal Inference to Globally Understand Black Box Predictors Beyond Saliency Maps. *Proceedings of the 9th International Workshop on Climate Informatics* (2019), 6.

[24] E. Rocha Rodrigues, I. Oliveira, R. Cunha, and M. Netto. 2018. DeepDownscale: A Deep Learning Strategy for High-Resolution Weather Forecast. In *2018 IEEE 14th International Conference on e-Science (e-Science).* 415–422. https://doi.org/10.1109/eScience.2018.00130

[25] J. Schmidli, C. M. Goodess, C. Frei, M. R. Haylock, Y. Hundecha, J. Ribalaygua, and T. Schmith. 2007. Statistical and dynamical downscaling of precipitation: An evaluation and comparison of scenarios for the European Alps. *Journal of Geophysical Research: Atmospheres* 112, D4 (2007). https://doi.org/10.1029/2005JD007026

[26] Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai Kin Wong, and Wang-chun WOO. 2015. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. *Advances in neural information processing systems* (June 2015).

[27] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. 2014. Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. (April 2014).

[28] P. M. M. Soares, D. Maraun, S. Brands, M. W. Jury, J. M. Gutiérrez, D. San-Martín, E. Hertig, R. Huth, A. Belušić Vozila, Rita M. Cardoso, S. Kotlarski, P. Drobinski, and A. Obermann-Hellhund. 2019. Process-based evaluation of the VALUE perfect predictor experiment of statistical downscaling methods. *International Journal of Climatology* 39, 9 (2019), 3868–3893. https://doi.org/10.1002/joc.5911

[29] Thomas Vandal, Evan Kodra, Sangram Ganguly, Andrew Michaelis, Ramakrishna Nemani, and Auroop R. Ganguly. 2017. 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.* ACM Press, Halifax, NS, Canada, 1663–1672. https://doi.org/10.1145/3097983.3098004

[30] Chunli Yang, Ninglian Wang, Shijin Wang, and Liang Zhou. 2018. Performance comparison of three predictor selection methods for statistical downscaling of daily precipitation. *Theoretical and Applied Climatology* 131, 1 (Jan. 2018), 43–54. https://doi.org/10.1007/s00704-016-1956-x

[31] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. 2016. Learning Deep Features for Discriminative Localization. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* IEEE, Las Vegas, NV, USA, 2921–2929. https://doi.org/10.1109/CVPR.2016.319

[32] Luisa M. Zintgraf, Taco S. Cohen, Tameem Adel, and Max Welling. 2017. Visualizing Deep Neural Network Decisions: Prediction Difference Analysis. *ICLR* (Feb. 2017).