Optimization of Convolutional Neural Network models for spatially coherent multi-site fire danger predictions

Óscar Mirones[1], Jorge Baño-Medina[2], Mario Santa Cruz[3], Swen Brands[1], and joaquin Bedia[4]

[1]Instituto de Física de Cantabria (IFCA), CSIC-Universidad de Cantabria
[2]Santander Meteorology Group, Institute of Physics of Cantabria, CSIC-University of Cantabria, Santander, Spain
[3]Predictia Intelligent Data Solutions S.L.
[4]University of Cantabria

November 14, 2023

Abstract

The accurate prediction of the Fire Weather Index (FWI), a multivariate climate index for wildfire risk characterization, is crucial for both wildfire management and climate-resilient planning. Moreover, consistent multisite fire danger predictions are key for targeted allocation of resources and early intervention in high-risk areas, as well as for "megafire" risk detection. In this regard, Convolutional Neural Networks (CNNs) are known to capture complex spatial patterns in climate data. This study compares different CNN architectures and traditional Statistical Downscaling (SD) methods (regression and analogs) for predicting daily FWI across diverse locations in Spain, considering marginal, distributional and spatial coherence measures for validation. Overall, the CNN-Multi-Site-Multi-Gaussian configuration, which explicitly accounts for the inter-site variability in the output layer structure, showed a superior performance. These insights provide a methodological guidance for the successful application of CNNs in the context wildfire risk assessment, enhancing wildfire response strategies and climate adaptation planning.

Optimization of Convolutional Neural Network models for spatially coherent multi-site fire danger predictions

Óscar Mirones[1], Jorge Baño-Medina[1], Mario Santa Cruz[2], Swen Brands[1], Joaquín Bedia[3,4]

[1]Instituto de Física de Cantabria (IFCA), CSIC-Universidad de Cantabria, Santander, Spain
[2]Predictia Intelligent Data Solutions S.L.
[3]Dept. Matemática Aplicada y Ciencias de la Computación (MACC), Universidad de Cantabria, Santander, Spain
[4]Grupo de Meteorología y Computación, Universidad de Cantabria, Unidad Asociada al CSIC, Santander, Spain

Key Points:

- Convolutional neural networks (CNNs) are compared with classical statistical downscaling methods for Fire Weather Index (FWI) prediction.
- The best CNN setup provides balanced results for all validation metrics, including accuracy, simulation of extremes and spatial consistency.
- Our findings provide a methodological basis for the development of more robust, spatially coherent regional future FWI scenarios.

Corresponding author: Óscar Mirones, `mironeso@unican.es`

Abstract

The accurate prediction of the Fire Weather Index (FWI), a multivariate climate index for wildfire risk characterization, is crucial for both wildfire management and climate-resilient planning. Moreover, consistent multisite fire danger predictions are key for targeted allocation of resources and early intervention in high-risk areas, as well as for “megafire” risk detection. In this regard, Convolutional Neural Networks (CNNs) are known to capture complex spatial patterns in climate data. This study compares different CNN architectures and traditional Statistical Downscaling (SD) methods (regression and analogs) for predicting daily FWI across diverse locations in Spain, considering marginal, distributional and spatial coherence measures for validation. Overall, the CNN-Multi-Site-Multi-Gaussian configuration, which explicitly accounts for the inter-site variability in the output layer structure, showed a superior performance. These insights provide a methodological guidance for the successful application of CNNs in the context wildfire risk assessment, enhancing wildfire response strategies and climate adaptation planning.

Keywords: deep learning, statistical downscaling, Generalized Linear Models, analogs, spatial structure, future wildfire risk assessment.

Plain Language Summary

This study focuses on the Fire Weather Index (FWI), a pivotal climate index for the assessment of wildfire risk. Accurate FWI predictions are vital for wildfire management. This study explores the viability of employing Convolutional Neural Networks (CNNs) as a Statistical Downscaling (SD) technique for precise FWI prediction across diverse locations in Spain in comparison with two conventional SD methodologies: Generalized Linear Models and analogs. Following a cross-validation scheme based on observed daily FWI data, we find that the CNN-Multi-Site-Multi-Gaussian (CNN-MSMG) configuration exhibits noteworthy proficiency in daily FWI prediction. This model explicitly incorporates the covariance structure of the predictands into the CNN architecture, yielding spatially consistent FWI predictions. Furthermore, CNN-MSMG has optimal properties for use in the context of climate change, providing a robust replication of extreme events and extrapolation capabilities if applied to novel climate scenarios. These findings have substantial implications for improving regional-to-local FWI scenarios used to inform vulnerability and impact assessment studies.

1 Introduction

Climate fire danger indices are key to assess and predict the risk of wildfire occurrence and severity. They are based on the integration of daily near-surface temperature, humidity, wind speed and precipitation records (de Groot et al., 2006), and thus provide more accurate wildfire risk forecasts than their input variables alone (see e.g. Dowdy et al., 2009; Fugioka et al., 2009). Beyond the near-term prediction horizon, fire danger indices are also useful to monitor changes in wildfire risk over time. As a result, downscaled fire danger scenarios are essential for vulnerability and adaptation strategies in regional to local applications, since General Circulation Model (GCM) outputs (Eyring et al., 2016) can’t provide actionable climate information at these spatial scales (Giorgi et al., 2009). Given their suitability for most impact studies, statistical downscaling (SD, Maraun & Widmann, 2018) of future fire weather scenarios is often required, including *perfect-prognosis* methods (Bedia et al., 2013, see Sec. 2.1) or bias-adjustment tools (Abatzoglou & Brown, 2012; Casanueva et al., 2018). In this case, there are three key aspects to focus on: (1) the reproducibility of extremes, as they can substantially increase wildfire impacts (Turco et al., 2018); (2) extrapolation capability is vital for predicting of out-of-sample values, since fire danger conditions are expected to change drastically in many regions (Bedia et al., 2015; Quilcaille et al., 2023), and (3) the ability to keep the predictand’s (FWI) spatial consistency is important to identify potentially hazardous fire risk scenarios affecting a wide geographical area, thereby increasing the odds of "fire clusters" with catastrophic potential (San-Miguel-Ayanz et al., 2013).

While most standard SD methods show good performance in at least one of these 3 aspects (Maraun et al., 2019), none of them is able to effectively accomplish all of them. In this context, the classical analog method (Lorenz, 1969; Zorita & von Storch, 1999; Brands et al., 2011) is still a competitive benchmark due to its ability to model both the extremes and the spatial structure (Widmann et al., 2019). However, if applied in its original form (Zorita & von Storch, 1999), this method fails to extrapolate beyond observed extremes, limiting its use for climate change applications (Bedia et al., 2013). In this sense, regression-based models are the better choice since they allow for better extrapolation (Baño-Medina et al., 2021; Balmaceda-Huarte et al., 2023) but, on the downside, they usually underestimate the extremes (Hertig et al., 2019). A further disadvantage of standard regression models (including Generalized Linear Models, GLMs) is their single-site structure unable to effectively model the spatial dependencies of the predictand variable(s). Other proposed alternatives combine the benefits of perfect-prog models and Weather Generators (PP-WG, see e.g.: Cannon, 2008; Carreau & Vrac, 2011), allowing to estimate the uncertainty of a local predictand variable and even to sample from the conditional distributions to recover the variability of the time series. To date, however, and with some exceptions (Legasa et al., 2023), most of these studies have focused on the estimation of uni-variate, single-site distributions, thereby not taking into account the spatial structure of the predictand nor its relationships with other predictand variables.

In this regard, deep learning methods, and in particular Convolutional Neural Networks (CNNs, LeCun et al., 1995) may offer a suitable alternative to meet these requirements with an adequate tuning. CNNs perform convolutions with learnable kernels over the spatial dimensions of atmospheric fields, inferring a non-linear mapping between low-resolution predictor fields and high-resolution predicand fields that has been shown to outperform conventional SD methods in many aspects (Baño-Medina et al., 2020). Regarding extrapolation ability, CNNs can produce plausible future climate change scenarios (Baño-Medina et al., 2021), comparable to those provided by dynamical downscaling (Baño-Medina et al., 2022). For a better reproducibility of extremes, parametric-CNNs (P-CNNs, Sec. 2.3) can estimate the parameters of conditional distributions given certain atmospheric conditions. As in the PP-WG approach, an adequate CNN architecture is able to estimate the parameters of the whole joint (multi-site) probability structure of the covariance matrix and can coherently reproduce the spatial structure of the predicted fire danger series across all predictand locations.

In this study we describe different CNN-based regression models for multi-site extreme fire danger assessment under climate change conditions, based on Canadian Fire Weather Index (van Wagner, 1987) records at 29 locations in Spain. We deploy three alternative CNN topologies based on the PP-WG approach that estimate either uni-variate or multi-variate Gaussian distributions on daily timescale. The validation is based on specific measures of extreme reproducibility and spatial coherence, using classical SD methods (analogs and GLMs) as benchmark.

2 Data and Methods

2.1 Predictor set

Perfect-prognosis SD establishes empirical relationships between 1) the variability of atmospheric variables operating on large scales, typically derived from a global reanalysis with a resolution similar to that offered by current global climate models (Eyring et al., 2016) and 2) the local-scale variability of the predictand of interest (here: FWI) as represented by in-situ observations or gridded observational datasets derived there-from. Once the SD model is calibrated, the learnt relationships can be applied to GCM (instead of reanalysis) predictors in order to derive local climate change projections if, ideally, the following requirements are fulfilled: The predictor variables should be realistically represented by the GCMs (Fernandez-Granja et al., 2021; Brands, 2022; Brands et al., 2023), should carry the climate change signal and be physically related with the local variable. In addition, the SD model should be capable to extrapolate the learnt relationships to altered/unobserved climate regimes (Gutiérrez et al., 2013). For the case of FWI downscaling, the predictor selection under such non-perfect circumstances has been explored in a previous study we built upon here (Bedia et al., 2013). Namely, we use daily-mean 2m air temperature, the zonal and meridional wind velocity components at 10m, as well as temperature and specific humidity a the 850 hPa pressure level, covering a spatial domain centered on the target region. These data have been retrieved from ERA-Interim (Dee et al., 2011) for the period 1985–2011 (see Table A2).

2.2 Predictand: Fire Weather Index observations

The FWI is a multivariable index, and therefore the downscaling approach must carefully consider the physical consistency of its input variables. When these are separately downscaled, inter-variable dependencies may be modified leading to spatio-temporal inconsistencies in the simulated output fields that would affect the coherence of the output FWI predictions (see e.g.: Vrac & Friederichs, 2015). This uncertainty source is here circumvented by using the FWI index, rather than its components, as sole predictand variable. To this end, in-situ observations from 29 weather stations of the Spanish Meteorological Agency (AEMET) were obtained, recording the required data for FWI calculation. The AEMET dataset provides instantaneous values of temperature, relative humidity and wind speed at 13:00 UTC, and last 24-h accumulated precipitation, recorded at 07:00 UTC. FWI calculation follows the methodology described by Bedia et al. (2013). For an optimal dataset completeness, we consider the calibration period 1985–2011.

2.3 Convolutional Neural Networks

To identify the key factors of the FWI spatial structure, we deploy three CNN architectures of increasing topological complexity (see Fig. 1). The backbone of these topologies builds on well tested CNNs known to outperform both analogs and GLMs in temperature and precipitation downscaling (Baño-Medina et al., 2020). The hidden structure consists of 3 convolutional layers followed by two fully-connected ones. The convolutional layers consist of a block of three layers with 50, 25, and 10 (3×3) kernels respectively, while the fully-connected (dense) layers each contain 50 neurons for the CNN-Multi-Site (CNN-MS) and CNN-Multi-Site-Gaussian (CNN-MSG) configuration, or 200 neurons for the CNN-Multi-Site-Multi-Gaussian (CNN-MSMG) version (see Fig. 1). A non-linear ReLU activation function is applied between the layers. The output, where we find the main differences across models, is a dense fully-connected network with a linear activation function in CNN-MSG and CNN-MSMG. In the case of CNN-MS, the output layer consists of 29 neurons, each neuron corresponding to a point location (Table A1), yielding deterministic FWI predictions at each site.

In order to improve the representation of FWI extremes, we introduce modifications to the P-CNN structures in CNN-MSG and CNN-MSMG. In CNN-MSG, the output is modeled stochastically using an *independent* Gaussian distribution to estimate the parameters of $\mathcal{N}(\mu, \sigma)$ (mean and standard deviation respectively). Thus, for each of the 29 stations, two pairs of neurons are added to the output layer, one for each parameter. CNN-MSMG, in turn, aims to estimate the parameters of a *multivariate* Gaussian distribution $\mathcal{N}(\mu, \Sigma)$. In this case, μ denotes the mean and Σ represents the covariance matrix. Therefore, the output layer consists of a pair of neuron vectors, with sizes 29 and 464 respectively. The 29 neurons represent the μ parameters, while the 464 neurons correspond to the number of unique parameters estimated in the covariance matrix Σ. The aim of this multivariate Gaussian setup is to describe the values at each location as a correlated set, unlike the outcome of an independent Gaussian distribution, where the predictions at each site are independent of each other. The general architecture scheme for each P-CNN configuration is outlined in Fig. 1. To avoid model overfitting, ensure robustness and optimize parameter tuning and CNN architectures, all SD models have been fit following a cross-validation procedure comprising 4 temporal blocks spanning the periods 1985-1991, 1992-1998, 1999-2004 and 2005-2011. The loss functions used are the Mean Square Error (MSE) for CNN-MS, the negative log-likelihood of the independent Gaussian distribution for CNN-MSG, and the negative log-likelihood of the multivariate Gaussian distribution for CNN-MSMG. The benchmark SD methods (analogs and GLMs) have been fitted following the same cross-validation scheme (see Appendix B for additional details on these methods).

2.4 Validation

Here, the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), as well as the quantile-quantile plot (QQ-plot), are used to validate the similarity of the temporal sequence and empirical distribution between the downscaled and observed daily FWI time series at a given station (Déqué, 2011). Apart from these classical marginal validation metrics, the focus is put on *spatial* coherence, as outlined in the following.

2.4.1 Location Correlogram

To qualitatively evaluate whether the distinct SD methods are able to reproduce the spatial correlation structure of the observed FWI, we use the *location correlogram* (Herdin et al., 2005). Firstly, the $n = 29$ observed daily *in-situ* FWI time series from the complete station network are correlated with each other for all possible combinations (i.e. $n \times \frac{n-1}{2}$ pairs) using Spearman's rank correlation coefficient and the resulting coefficients are plotted against the respective pairwise station distances. Then a local 2nd-order polynomial ("loess") is fitted to the scatter-plot, resulting in a curve that depicts the spatial correlation structure of the observed FWI. As a quantitative summary measure, we use the *correlation length* (CL), defined as the geographical distance corresponding to the point of intersection of a given correlation threshold with the fitted loess line. A threshold of $\rho = 0.4$ has proven to be most suitable for characterizing the spatial FWI structure in this study (Table Appendix C), and the overall results are robust to changes in this choice. After applying the same method to the downscaled time series from each of the thee SD methods, the CL bias between the simulated and observed spatial structure is calculated as an overall measure of the methods' capability to reproduce the spatial coherence of the observed FWI (Widmann et al., 2019).

2.4.2 Mutual information for FWI90

FWI extremes are particularly relevant for fire danger assessment. As a result, from the spatial consistency point of view, the users of downscaled FWI values will be primarily interested in a realistic representation of joint higher-percentile FWI exceedances among locations (see e.g. Bedia et al., 2014). To this aim, Mutual Information (MI) provides a suitable measure of the dependence between two random variables X, Y (here, predictions at two locations) that is unaffected by their marginal distributions and quantifies the amount of mutual information between them (see e.g. Hlinka et al., 2013). For two discrete random variables X and Y it is defined as:

$$MI(X,Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \cdot \log \left(\frac{p(x,y)}{p(x) \cdot p(y)} \right) \tag{1}$$

MI is zero if the two events are independent, i.e. if $p(X,Y) = p(X) \cdot p(Y)$, non-negative ($MI(X,Y) \geq 0$) and symmetric ($MI(X,Y) = MI(Y,X)$).

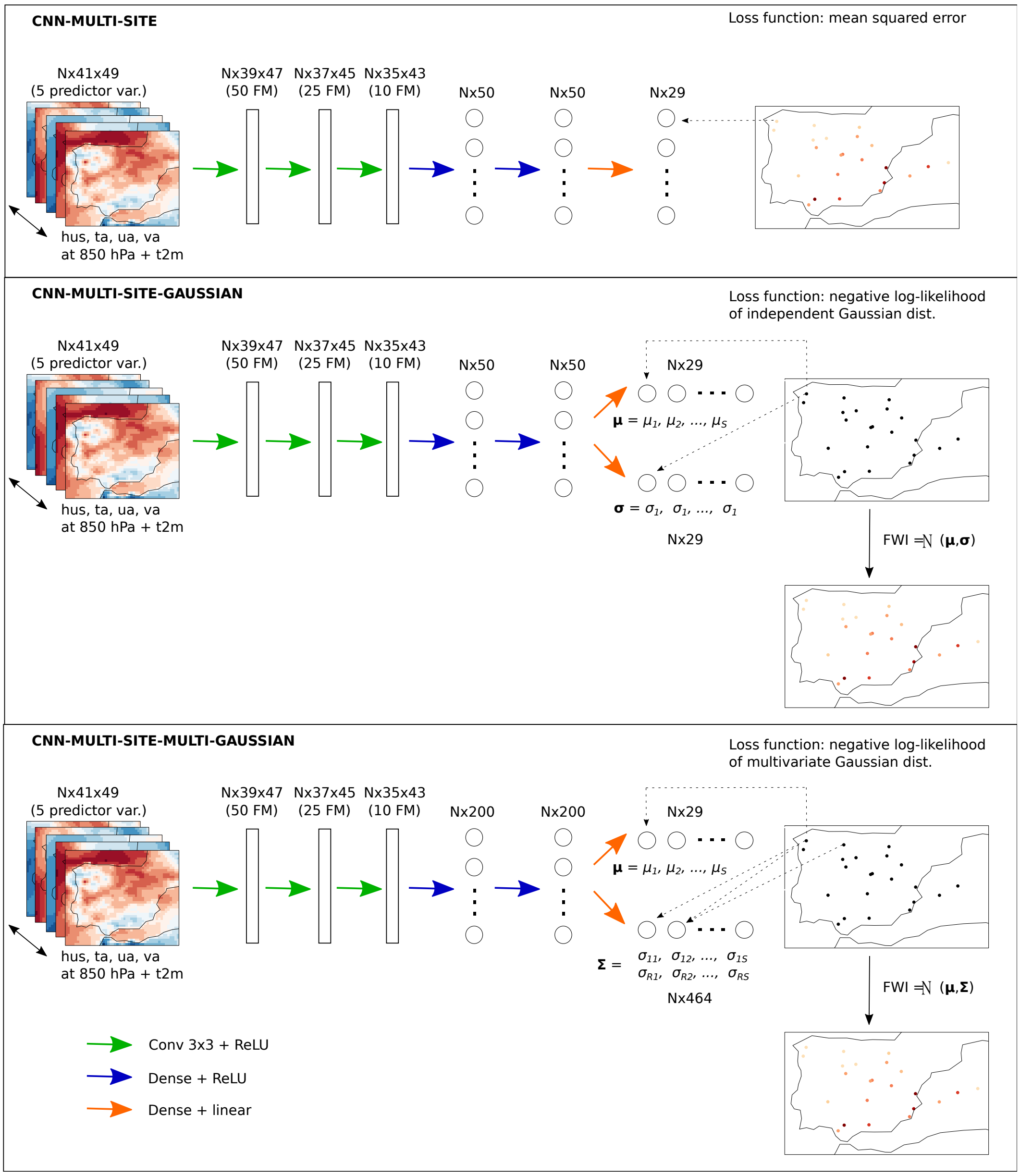

Figure 1. Scheme of the convolutional neural network architecture used in this study. The network includes a first block of three convolutional layers with 50, 25 and 10 (3x3) kernels, respectively, followed by two fully connected dense layers with 50 or 200 neurons each, depending on the model. For CNN-MSG and CNN-MSMG, the output is modeled through and independent Gaussian distribution and a multivariate Gaussian distribution respectively, and the corresponding parameters are estimated by the network, obtaining FWI as final product, either deterministically (CNN-MS) or stochastically (CNN-MSG and CNN-MSMG). The output layer is activated linearly while the previous layers of the network are activated non-linearly.

Here, we consider the binary variables X, Y at each location, stating whether the FWI values x_i, y_i lie above or below the 90th percentile for each pair of locations. We then calculate the MI for each pair of locations following the definition above (eq. 1). As for the correlograms (Sec. 2.4.1), we plot each MI_{ij} against the distance of the locations i, j and fit a degree-2 loess curve to the resulting scatter-plots. We then define MI thresholds for calculating the MI lengths (MIL) in observations and for the different downscaling methods. We use a MI threshold of 0.05, yielding results comparable to those obtained from CL analysis , and focusing on the identification of potential new information about each methods' performance (Fig. C2). As in CL analysis, the MIL biases are calculated as the difference between predicted and observed MILs.

3 Results

The results presented correspond to the generic June to September fire season, representative of the Iberian Peninsula (JJAS, see e.g.: Bedia et al., 2014). It's important to highlight that the models were calibrated using the entire annual dataset. However, a subset comprising the JJAS season was used to present the results relevant for fire danger assessment in this region. In the following subsections, we categorized the station network into three groups based on proximity to the sea and general climate conditions: Atlantic, Coastal Mediterranean, and Continental Mediterranean (see Table A1). The suitability of this classification for FWI aggregation is confirmed by the results obtained with the mutual information measure (Sec. 3.2.2).

3.1 Predictive accuracy and distributional similarity

In agreement with previous studies (Brands et al., 2011), the SD methods' accuracy is *generally* lower at continental sites than near the coast. Overall, all methods perform similarly with regard to predictive accuracy, summarized in terms of the RMSE of FWI90 predictions in Fig. C1 (Appendix C).

However, the distributional characteristics of the predictions differ largely among methods. The quantile-quantile (QQ) plots shown in Fig. 2 compare the observed and predicted empirical FWI distributions. While all methods perform well in predicting the mean FWI, disparities emerge at higher percentiles, crucial for fire danger analysis. The benchmarking analog approach produces best results for the right tail, closely followed by multivariate CNN-MSMG, showing similar results across regions. Conversely, GLM and CNN-MS consistently underestimate high percentile FWI events, failing to realistically represent most dangerous situations. CNN-MSG also achieves good results, comparable to CNN-MSMG, but is outperformed by the latter in the Coastal Mediterranean and Atlantic regions. Notably, in the Atlantic region, CNN-MSG unrealistically inflates the highest FWI percentiles and underestimates most of the FWI distribution. On the contrary, in the Continental Mediterranean region, CNN-MSG performs slightly better than CNN-MSMG for higher percentiles.

In order to obtain a quantitative measure of distributional deviance with respect to the observed distribution, we calculate the RMSE considering the differences between predicted and observed quantiles of i) the entire FWI times series and ii) the FWI time series values exceeding the station-specific 90th percentile (FWI90, Fig. 2). Excluding the results for the analog method, lowest RMSE values for both indicators are obtained either by CNN-MSG or CNN-MSMG. Regardless of the specific target region, the former approach demonstrates significantly better performance compared to the latter in terms of FWI, and only exhibits a slight decrease in performance for FWI90. Specifically, when emphasizing FWI90, the CNN-MS and GLM models exhibit noticeably poorer performance compared to the analog benchmark. In contrast, the results for CNN-MSG and CNN-MSMG models are considerably better in this regard.

Overall, the reference analog method performs best in representing the distribution of the daily FWI in most cases. CNN-MSG and CNN-MSMG perform slightly worse, with CNN-MSG slightly overemphasizing severe FWI frequencies in the Atlantic region. The analog method performs best overall, but it's applicability in climate changes studies is limited due to its inability to extrapolate predictions outside the observed range. Conversely, both CNN-MSG and CNN-MSMG are competitive alternative methods in terms of distributional similarity. In the Secs. 3.2.1 and 3.2.2, we assess whether these conclusions hold for the *spatial structure* of the simulated mean and extreme FWI fields.

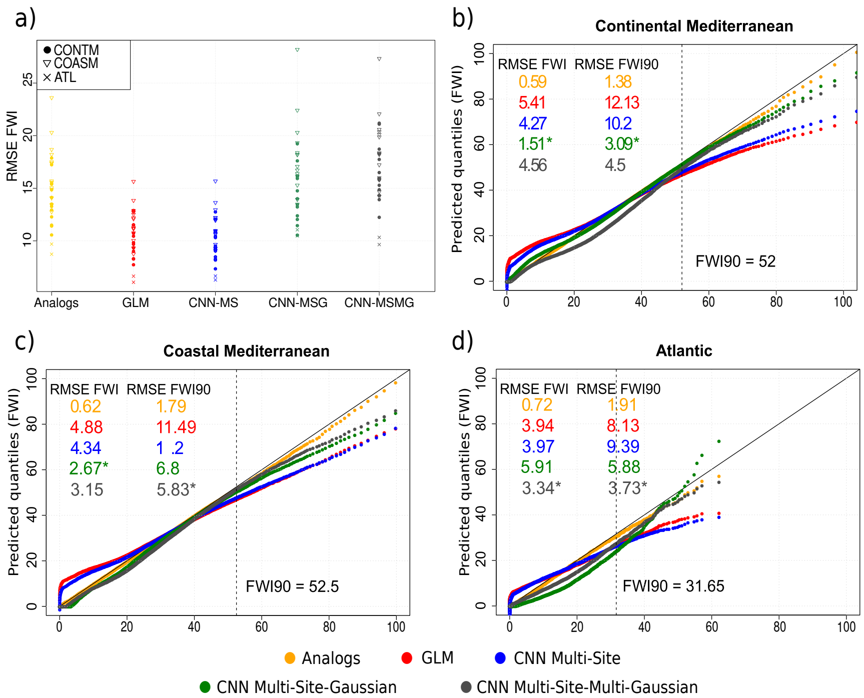

Figure 2. FWI RMSE per station and method (a) and Q-Q plots for the analog method, GLMs and the distinct CNN models (b - d). The figure is divided into 4 panels. The (a) panel refers to the RMSE for the simulated FWI per station and method distinguishing the regions by symbols. The remaining panels refer to the station subsets of (b) the Continental Mediterranean, (c) Coastal Mediterranean and (d) Atlantic regions. The method-specific distributional RMSE for the simulated FWI and FWI90 are indicated in the upper left corners of each panel and the best performing method is marked with an asterisk (excluding the benchmarking analog method). The dashed vertical line indicates the observed FWI90.

3.2 Spatial validation results

3.2.1 Dependence of inter-station relationships on distance

The temporal correlation coefficients' dependence on distance, analyzed as described in Sec. 3, is depicted in Fig. 3. As expected, the observed strength of the relations decreases exponentially with increasing distance between the stations and stabilizes around $rho = 0.1$, the CL for $\rho = 0.4$ being located at 208.30 km (panel *a)*, grey curve). The corresponding point clouds and polynomials for the SD methods are depicted in red in panels *b)* through *f)*, where the respective validation measures are also indicated (see upper right corners and also Appendix C). The exponential decay seen in the observations is reproduced more or less successfully by all SD methods except CNN-MSG, that produces far too weak short distance relationships, failing to reproduce any spatial structure in the data. The analog method is, as expected, most successful in reproducing the observed correlation structure, closely followed by CNN-MSMG, while GLM and CNN-MS consistently overestimate pairwise correlations. Among the suitable methods (i.e. excluding CNN-MSG), the medium-to-long-distance correlations are overestimated by all methods, particularly by GLM and CNN-MS. The stronger short-distance correlations are also generally overestimated, but to a lesser degree, and they are almost perfectly met by the analog method and closely approximated by CNN-MSMG.

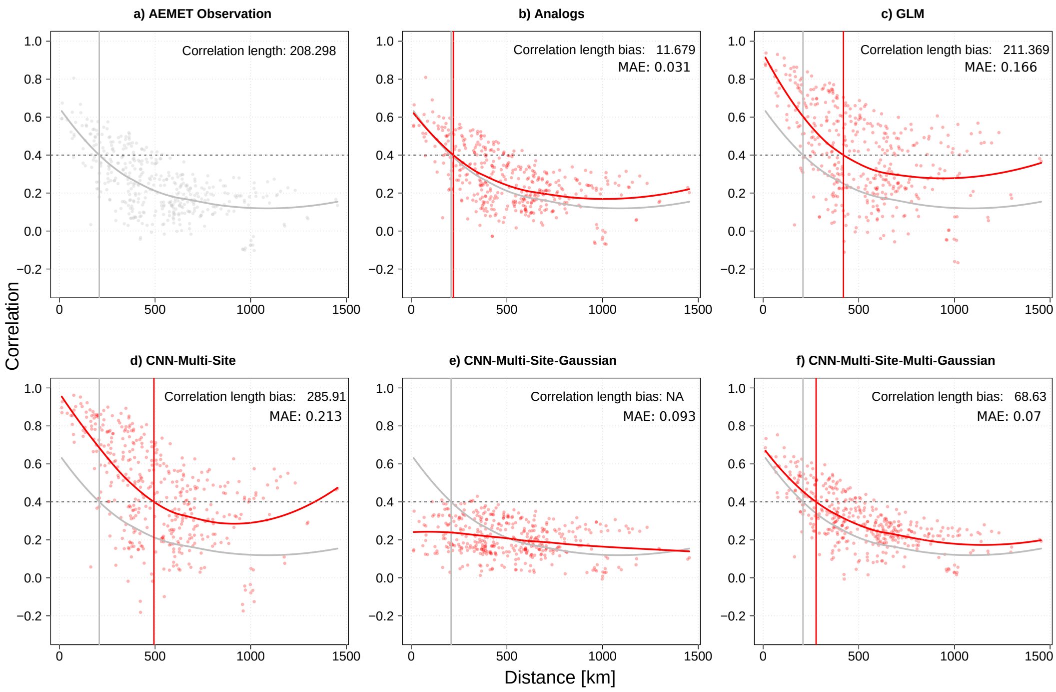

Figure 3. Correlograms illustrating the daily JJAS FWI dependence of the inter-station relationships, described by the Spearman correlation coefficients among all station pairs (y-axis), against their respective distances in kilometers (x-axis). The correlograms correspond to the observations (panel *a*) and to each SD method tested (panels *b* to *f*); the grey loess line of the observations correlogram is included in all panels for visual comparison. It is also displayed the observed Correlation Length (CL, panel *a)*) and the CL bias and MAE for each SD method (in panels *b* to *f*). Here, the MAE is calculated as the difference (in absolute value) between predicted and observed correlation coefficients for each station pair.

3.2.2 Mutual information for fire weather extremes

In Fig. 4 (upper panel), we present the mutual information (MI) values obtained from the observational network in the upper triangle (a1), compared with those produced by CNN-MSMG, the best performing SD method for this metric (the benchmarking analog method is excluded), in the lower triangle (a2). The stations are grouped into characteristic climate regimes as described in Sec. 3.1. Geographical proximity translates into higher MI values, as the case of Vigo and Santiago de Compostela (NW Iberia, $MI = 0.11$), Barajas and Retiro (Madrid, central Spain, $MI = 0.14$), or Valencia and Valencia-Airport (SE, Mediterranean coast, $MI = 0.13$). Furthermore, several climatologically homogeneous regions can be identified in the matrix, yielding visually discernible clusters of high MI values, e.g. the Soria-Valladolid-Salamanca-Zamora cluster pertaining to the central-north Iberian high plains. The MI pattern obtained from CNN-MSMG is similar to that seen in observations and is thus approximately symmetric (compare Fig. 4-a2 with a1). Nevertheless, CNN-MSMG somewhat overestimates the spatial dependencies, indicated by slightly higher MI values than those obtained from observations, and also reflected by regional clusters not seen in observations (e.g. Ciudad Real, Badajoz and Granada).

In Fig. 4b and c, we illustrate the MI biases relative to the observations for CNN-MS and CNN-MSG (b1, b2), as well as for CNN-MSMG and GLM (c1, c2). We focus on station pairs with MI values $\geq$ 0.05 in observations, thus discarding already independent station pairs (blank matrix cells). Since the MI bias of the analog method is negligible for all station pairs, the corresponding results are shown in Appendix Appendix C. The MI bias of CNN-MSMG is below the 0.05 threshold for most station pairs, with a few exceptions with both positive or negative values (Fig. 4c1). CNN-MS exhibits a consistent positive bias, consistently overestimating the spatial dependence of extreme FWI events (Fig. 4b1). CNN-MSG, in turn, systematically underestimates these dependencies (Fig. 4b2), yielding a lower bias magnitude than for CNN-MS. The GLM approach tends to overestimate these dependencies, albeit to a lesser extent than CNN-MS (compare Fig. 4c2 with b1).

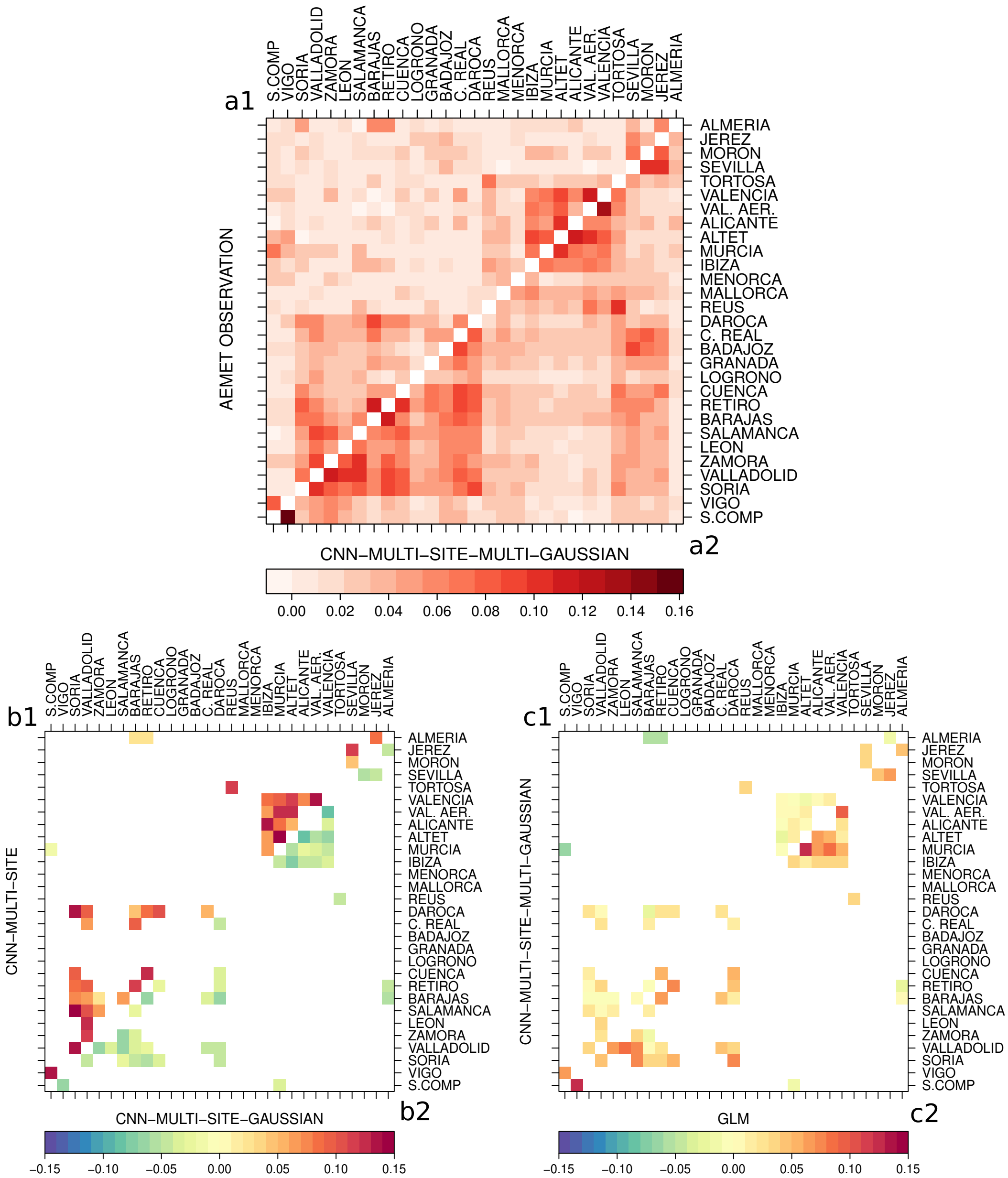

Figure 4. MI matrices for FWI90 events during the fire season (JJAS) obtained from observations (upper triangle in upper panel, a1) and from the best performing SD model (CNN-MSMG, lower triangle in upper panel, a2). Panels b an c show the MI biases for 4 remaining SD methods with respect to the observations (b1: CNN-MS, b2: CNN-MSG, c1: CNN-MSMG, c2: GLM). In b and c panels, only station pairs with MI ≥ 0.05 in the observations are shown.

4 Conclusions

We conducted a comprehensive comparison of various Convolutional Neural Network (CNN) architectures in contrast to two established statistical downscaling (SD) methods, specifically Generalized Linear Models (GLMs) and analogs. Our assessment focused on evaluating their performance in terms of predictive accuracy, distributional congruence, and spatial coherence for Fire Weather Index (FWI) predictions across 29 locations in Spain.

Among the diverse CNN architectures scrutinized, CNN-MSMG demonstrated the most favorable outcomes across these validation criteria. This setup considers the multivariate nature of the predictions in the output layer yielding a predicted covariance matrix that explicitly accounts for the inter-site variability. It exhibited a notable capacity to accurately represent observed FWI distributions at both single-point and multi-site scales, closely aligning with the outcomes of the benchmarking analogs method. Notably, the analogs method, by design, upholds multisite spatial consistency without alteration, at the cost of limitations for extrapolation in climate change conditions that can be overcome by the rest of methods tested. In contrast CNN-MS (multisite CNN) and GLMs yielded poorer predictive accuracy and consistently overestimated the spatial dependence among sites. In turn, CNN-MSG (multisite Gaussian) attained good results in terms of single-site validation, but proved inefficient in modelling the spatial structure, essentially behaving like a single-site weather generator.

The results presented emphasize the importance of parameter tuning for CNN development in the context of statistical downscaling in order to produce credible predictions. In the particular case of FWI, an adequate tuning is needed in order to ensure actionable climate information for the prevention of wildfire impacts, and this study provides a methodological guidance for the successful application of CNNs to this aim.

5 Open Research

We follow the FAIR principles (Findability, Accesibility, Interoperability and Reuse, Wilkinson et al. (2016)) and publish the code (DOI: `10.5281/zenodo.8387558`) and the data (DOI: `10.5281/zenodo.8381437`) required to replicate the results presented in this manuscript. We build on the R based (R Core Team, 2020) framework *climate4R* (Iturbide et al., 2019) to digest, manipulate, downscale (see also Bedia et al., 2020) and visualize (Frías et al., 2018) the climate data. For the deep learning models, we lean on *downscaleR.keras*, a library that integrates *tensorflow* (Abadi et al., 2015) and *keras* (Gulli & Pal, 2017) into the *climate4R* framework (Baño-Medina et al., 2020).

Appendix A Input Data

Table A1 is a summary of the AEMET weather station database. We also indicate their corresponding climatic zone, according to the spatial aggregation summarizing the results in Sec. 3.1 of the main text. The *Short name* column indicates the abbreviated labels used throughout the article figures.

Station name	**Short name**	**Lon**	**Lat**	**Altitude**	**Climatic region**
REUS–AEROPUERTO	REUS	1.18	41.15	71	COASM
SANTIAGO DE COMPOSTELA–LABACOLLA	S.COMP	-8.41	42.89	370	ATL
VIGO–PEINADOR	VIGO	-8.62	42.24	261	ATL
SORIA	SORIA	-2.48	41.77	1082	CONTM
VALLADOLID	VALLADOLID	-4.75	41.64	735	CONTM
ZAMORA	ZAMORA	-5.73	41.52	656	CONTM
LEÓN–VIRGEN DEL CAMINO	LEÓN	-5.65	42.59	916	CONTM
SALAMANCA–MATACÁN	SALAMANCA	-5.50	40.96	790	CONTM
MADRID–BARAJAS	BARAJAS	-3.56	40.47	609	CONTM
MADRID–RETIRO	RETIRO	-3.68	40.41	667	CONTM
CIUDAD REAL	C. REAL	-3.92	38.99	628	CONTM
BADAJOZ–TALAVERA LA REAL	BADAJOZ	-6.81	38.88	185	CONTM
GRANADA–AEROPUERTO	GRANADA	-3.79	37.19	567	CONTM
SEVILLA–SAN PABLO	SEVILLA	-5.88	37.42	34	COASM
MORÓN DE LA FRONTERA	MORÓN	-5.61	37.16	87	COASM
JEREZ DE LA FRONTERA–AEROPUERTO	JEREZ	-6.06	36.75	27	COASM
ALMERÍA–AEROPUERTO	ALMERÍA	-2.36	36.85	21	COASM
MURCIA–SAN JAVIER	MURCIA	-0.80	37.79	4	COASM
ALICANTE–EL ALTET	ALTET	-0.57	38.28	43	COASM
ALICANTE	ALICANTE	-0.49	38.37	81	COASM
CUENCA	CUENCA	-2.14	40.07	945	CONTM
VALENCIA–AEROPUERTO	VAL. AER.	-0.47	39.49	69	COASM
VALENCIA	VALENCIA	-0.37	39.48	11	COASM
LOGROÑO–AGONCILLO	LOGROÑO	-2.33	42.45	353	CONTM
DAROCA	DAROCA	-1.41	41.11	779	CONTM
TORTOSA	TORTOSA	0.49	40.82	44	COASM
PALMA DE MALLORCA–SON SAN JUAN	MALLORCA	2.74	39.56	8	COASM
MENORCA–MAÓ	MENORCA	4.22	39.85	91	COASM
IBIZA/ES CODOLA	IBIZA	1.38	38.88	6	COASM

Table A1. Selected stations of the Spanish AEMET network, indicating their position in decimal degrees and meters above sea level (Datum WGS-84). The abbreviations corresponding to the climatic regions in the column are as follows: ATL for Atlantic, COASM for Coastal Mediterranean, and CONTM for Continental Mediterranean.

Table A2 provides a summary of the reanalysis fields used as predictors in this study. The predictor set has been chosen following the methodology for FWI downscaling presented by Bedia et al. (2013), but replacing relative humidity by specific humidity, the former being not directly available in some model simulation databases. The spatial extent of these fields covers a bounding box centered over the Iberian Peninsula, limited by the geographical coordinates $-10°/15°$E, $35°/45°$N.

Appendix B Benchmarking SD methods

We next provide further methodological details on the standard SD methods used as benchmarks in this study. Both are implemented in the R package `downscaleR` (Bedia et al., 2020), part of the climate4R framework for climate data analysis and visualization (Iturbide et al., 2019, `https://github.com/SantanderMetGroup/climate4R`).

Code	**Name**	**units**
T2M	Air Temperature at surface	K
T850	Air Temperature at 850 hPa	K
HUS850	Specific humidity at 850 hPa	$g\,kg^{-1}$
UA850	U-wind at 850 hPa	$m\,s^{-1}$
VA850	V-wind at 850 hPa	$m\,s^{-1}$

Table A2. Predictor variables used in this study, selected from the predictor combination proposed for statistical downscaling of FWI in Bedia et al. (2013). Note that for convenience, relative humidity at 850 hPa has been replaced by specific humidity, more commonly available in GCM datasets. All fields are daily mean values.

B1 Generalized lineal models

GLMs (Nelder & Wedderburn, 1972) are an extension of the classical linear regression that models the expected value of a random predictand variable for different types of probability distributions and link functions. This makes them a versatile tool for modeling a wide range of data types and situations, and therefore extensively used in SD applications (see e.g.: Chandler & Wheater, 2002; Gutiérrez et al., 2019). Here, the response variable is assumed to follow a Gaussian distribution. The relationship between the linear predictor $g(\mu)$ and the expected value of FWI is defined by the identity link function, so the linear predictor directly models the mean FWI, where $g(\mu)$ is defined as $g(\mu) = \mathbf{X}\beta$, where $\mathbf{X}$ is the design matrix containing the predictor variables (Sec. 2.1), and β is the vector of coefficients, estimated by maximum likelihood based on the probability density function of the Gaussian distribution using a least-squares iterative algorithm implemented in the R package `stats`(R Core Team, 2020). Furthermore, predictor configuration is such that only local information is used for training at each site. Here, an optimal number of 16 closest grid-points to each predictand point-location are retained to construct the local predictor set (Bedia et al., 2020), after testing different neighbourhood sizes using cross-validation (Sec. 2.3).

B2 Analogs

The analog method is a simple yet powerful downscaling technique which assumes that similar (or analog) atmospheric patterns (predictor set $\mathbf{X}$) over a region originates similar local meteorological outcomes (daily FWI) for a particular location or set of locations (Sec. 2.2). In this study, we use the standard deterministic nearest neighbor method analog technique based on the Euclidean distance, considering the complete fields to compute distances and only the first closest nearest closest analog for prediction (San-Martín et al., 2016), similar to the standard 'ANALOG' method of the VALUE intercomparison experiment (described in Gutiérrez et al., 2019, A.2), and considering the implementation described in Bedia et al. (2020). Note that using the complete fields as predictors ensures the maximum spatial coherence of the predictions among stations, since the same analog dates are chosen in each case for every point-location (see e.g. Widmann et al., 2019).

Appendix C Results

This section contains additional results as indicated in the figure captions.

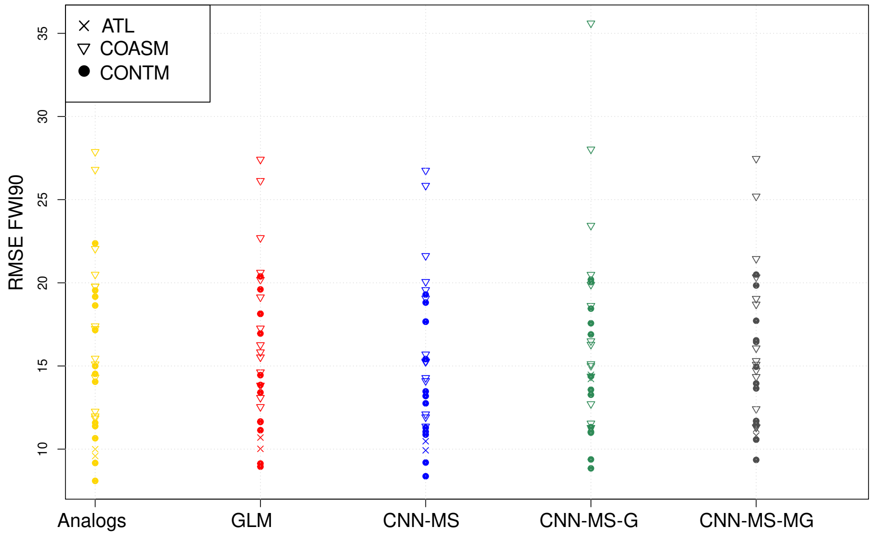

Figure C1. RMSE for the simulated FWI90 per station and method distinguishing the regions by symbols.

	CL	MIL	CL Bias	MIL Bias
AEMET_13UTC_FWI	208.30	168.22		
Analogs			11.68	-2.45
CNN-MS			285.91	259.81
CNN-MSG			*NA*	*NA*
CNN-MSMG			**68.63**	**119.67**
GLM			211.37	146.92

Table C1. The columns display the CL and MIL values for the reference observations, as well as the CL and MIL biases for the models, measured in kilometers (km). The lowest CL and MIL biases (excluding the benchmarking analogs method) are highlighted in boldface.

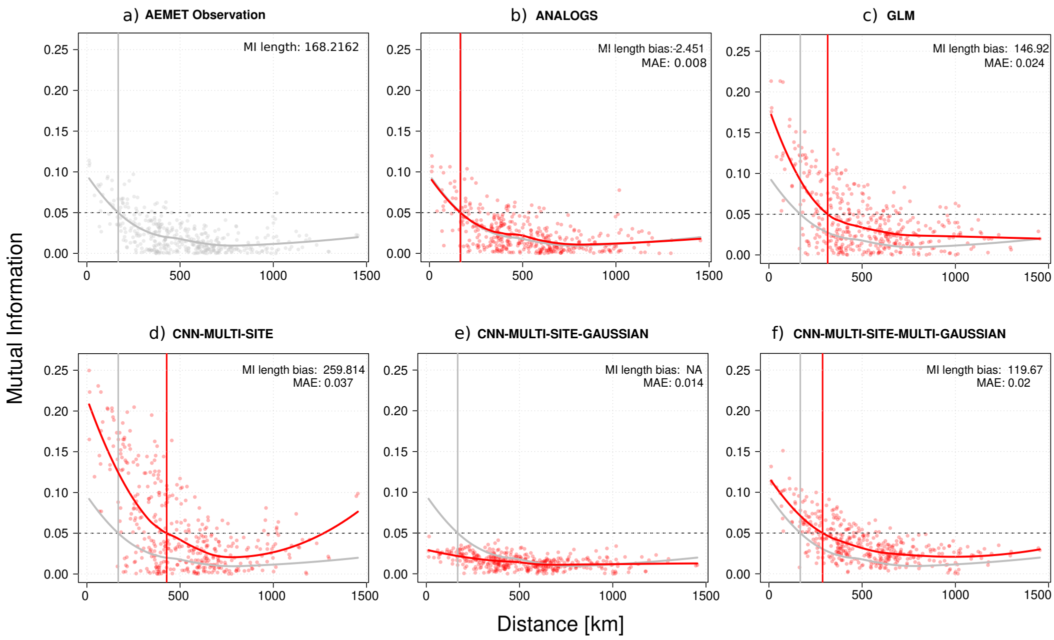

Figure C2. Mutual Information diagrams for FWI90 for fire season (JJAS) showing the mutual information of the FWI90 time series for each pair of stations against their geographical distances. The MI and MI length for the reference observations are shown in the upper left panel. In the rest of the panels, the MI length bias and the MAE are indicated at the top right of the panel.

Acknowledgments

This paper is part of the R+D+i project CORDyS (PID2020-116595RB-I00) with funding from the Spanish Ministry of Science MCIN/AEI/10.13039/501100011033. O.M. has received research support from grant PID2020-116595RB-I00 funded by MCIN/AEI /10.13039/501100011033. SB and JB acknowledge funding by the Ministry for the Ecological Transition and the Demographic Challenge (MITECO) and the European Commission NextGenerationEU (Regulation EU 2020/2094), through CSIC's Interdisciplinary Thematic Platform Clima (PTI-Clima).

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., … Zheng, X. (2015). *TensorFlow: Large-scale machine learning on heterogeneous systems.* Retrieved from `https://www.tensorflow.org/` (Software available from tensorflow.org)

Abatzoglou, J., & Brown, T. (2012). A comparison of statistical downscaling methods suited for wildfire applications. *International Journal of Climatology*, *32*, 772–780. doi: 10.1002/joc.2312

Balmaceda-Huarte, R., Baño-Medina, J., Olmo, M. E., & Bettolli, M. L. (2023). On the use of convolutional neural networks for downscaling daily temperatures over southern south america in a climate change scenario. *Climate Dynamics*, 1–15.

Baño-Medina, J., Manzanas, R., Cimadevilla, E., Fernández, J., González-Abad, J., Cofiño, A. S., & Gutiérrez, J. M. (2022, September). Downscaling multi-model climate projection ensembles with deep learning (DeepESD): contribution to CORDEX EUR-44. *Geoscientific Model Development*, *15*(17), 6747–6758. Retrieved 2023-03-24, from `https://gmd.copernicus.org/articles/15/6747/2022/` (Publisher: Copernicus GmbH) doi: 10.5194/gmd-15-6747-2022

Baño-Medina, J., Manzanas, R., & Gutiérrez, J. M. (2020, April). Configuration and intercomparison of deep learning neural models for statistical downscaling. *Geoscientific Model Development*, *13*(4), 2109–2124. Retrieved 2023-03-24, from `https://gmd.copernicus.org/articles/13/2109/2020/` (Publisher: Copernicus GmbH) doi: 10.5194/gmd-13-2109-2020

Baño-Medina, J., Manzanas, R., & Gutiérrez, J. M. (2021, December). On the suitability of deep convolutional neural networks for continental-wide downscaling of climate change projections. *Climate Dynamics*, *57*(11), 2941–2951. Retrieved 2023-03-24, from `https://doi.org/10.1007/s00382-021-05847-0` doi: 10.1007/s00382-021-05847-0

Bedia, J., Baño-Medina, J., Legasa, M. N., Iturbide, M., Manzanas, R., Herrera, S., … Gutiérrez, J. M. (2020, April). Statistical downscaling with the downscaleR package (v3.1.0): contribution to the VALUE intercomparison experiment. *Geoscientific Model Development*, *13*(3), 1711–1735. Retrieved 2020-04-03, from `https://www.geosci-model-dev.net/13/1711/2020/` doi: 10.5194/gmd-13-1711-2020

Bedia, J., Herrera, S., Camia, A., Moreno, J. M., & Gutierrez, J. M. (2014). Forest Fire Danger Projections in the Mediterranean using ENSEMBLES Regional Climate Change Scenarios. *Climatic Change*, *122*, 185–199. doi: 10.1007/s10584-013-1005-z

Bedia, J., Herrera, S., Gutierrez, J., Benali, A., Brands, S., Mota, B., & Moreno, J. (2015). Global patterns in the sensitivity of burned area to fire-weather: implications for climate change. *Agricultural and Forest Meteorology*, *214–215*, 369–379. doi: 10.1016/j.agrformet.2015.09.002

Bedia, J., Herrera, S., San-Martin, D., Koutsias, N., & Gutiérrez, J. M. (2013, September). Robust projections of Fire Weather Index in the Mediterranean using statistical downscaling. *Climatic Change*, *120*, 229–247. Retrieved from http://link.springer.com/article/10.1007/s10584-013-0787-3 doi: 10.1007/s10584-013-0787-3

Brands, S. (2022). A circulation-based performance atlas of the cmip5 and 6 models for regional climate studies in the northern hemisphere mid-to-high latitudes. *Geoscientific Model Development*, *15*(4), 1375–1411. Retrieved from https://gmd.copernicus.org/articles/15/1375/2022/ doi: 10.5194/gmd-15-1375-2022

Brands, S., Fernández-Granja, J. A., Bedia, J., Casanueva, A., & Fernández, J. (2023). A global climate model performance atlas for the southern hemisphere extratropics based on regional atmospheric circulation patterns. *Geophysical Research Letters*, *50*(10), e2023GL103531. doi: https://doi.org/10.1029/2023GL103531

Brands, S., Taboada, J., Cofiño, A., Sauter, T., & Schneider, C. (2011, 08). Statistical downscaling of daily temperatures in the nw iberian peninsula from global climate models: Validation and future scenarios. *Climate Research*, *48*, 163-176. doi: 10.3354/cr00906

Cannon, A. J. (2008). Probabilistic multisite precipitation downscaling by an expanded bernoulli–gamma density network. *Journal of Hydrometeorology*, *9*(6), 1284–1300.

Carreau, J., & Vrac, M. (2011). Stochastic downscaling of precipitation with neural network conditional mixture models. *Water Resources Research*, *47*(10).

Casanueva, A., Bedia, J., Herrera, S., Fernández, J., & Gutiérrez, J. M. (2018). Direct and component-wise bias correction of multi-variate climate indices: the percentile adjustment function diagnostic tool. *Climatic Change*, *147*, 411–425. doi: 10.1007/s10584-018-2167-5

Chandler, R. E., & Wheater, H. S. (2002, October). Analysis of rainfall variability using generalized linear models: A case study from the west of Ireland: GENERALIZED LINEAR MODELING OF DAILY RAINFALL. *Water Resources Research*, *38*(10), 1–11. Retrieved 2019-06-24, from http://doi.wiley.com/10.1029/2001WR000906 doi: 10.1029/2001WR000906

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., … Vitart, F. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q J R Meteorol Soc*, *137*, 553–597. doi: 10.1002/qj.828

de Groot, W. J., Goldammer, J. G., Keenan, T., Brady, M. A., Lynham, T. J., Justice, C. O., … O'Loughlin, K. (2006). Developing a global early warning system for wildland fire. *Forest Ecology and Management*, *234*, S10.

Dowdy, A. J., Centre for Australian Weather and Climate Research, Australia, Bureau of Meteorology, & CSIRO. (2009). *Australian fire weather as represented by the McArthur Forest Fire Danger Index and the Canadian Forest Fire Weather Index.* Melbourne: Centre for Australian Weather and Climate Research,.

Déqué, M. (2011). Deterministic forecasts of continuous variables. In *Forecast verification* (p. 77-94). John Wiley and Sons, Ltd. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119960003.ch5 doi: https://doi.org/10.1002/9781119960003.ch5

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. doi: 10.5194/gmd-9-1937-2016

Fernandez-Granja, J. A., Casanueva, A., Bedia, J., & Fernández, J. (2021). Improved atmospheric circulation over europe by the new generation of cmip6 earth system models. *Climate Dynamics*, *56*, 3527–3540. doi: 10.1007/s00382-021-05652-9

Frías, M. D., Iturbide, M., Manzanas, R., Bedia, J., Fernández, J., Herrera, S., …

Gutiérrez, J. M. (2018, January). An R package to visualize and communicate uncertainty in seasonal climate prediction. *Environmental Modelling & Software*, *99*, 101–110. Retrieved 2017-10-28, from https://www.sciencedirect.com/science/article/pii/S1364815217305157 doi: 10.1016/j.envsoft.2017.09.008

Fugioka, F. M., Gill, A., Viegas, D. X., & Wotton, B. (2009). Fire Danger and Fire Behavior Modeling Systems in Australia, Europe, and North America. In A. Bytnerowicz, M. Arbaugh, A. Riebau, & C. Andersen (Eds.), *Developments in Environmental Science.* The Netherlands: Elsevier B.V.

Giorgi, F., Jones, C., & Asrar, G. R. (2009). Addressing climate information needs at the regional level: the CORDEX framework. *WMO Bulletin*, *58*(3), 175–183. Retrieved 2023-09-06, from https://public.wmo.int/en/bulletin/addressing-climate-information-needs-regional-level-cordex-framework

Gulli, A., & Pal, S. (2017). *Deep learning with keras.* Packt Publishing Ltd.

Gutiérrez, J. M., Maraun, D., Widmann, M., Huth, R., Hertig, E., Benestad, R., ... others (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), 3750–3785.

Gutiérrez, J. M., San-Martín, D., Brands, S., Manzanas, R., & Herrera, S. (2013). Reassessing statistical downscaling techniques for their robust application under climate change conditions. *Journal of Climate*, *26*(1), 171 - 188. doi: https://doi.org/10.1175/JCLI-D-11-00687.1

Herdin, M., Czink, N., Ozcelik, H., & Bonek, E. (2005). Correlation matrix distance, a meaningful measure for evaluation of non-stationary MIMO channels. In *Vehicular technology conference, 2005. vtc 2005-spring. 2005 ieee 61st* (Vol. 1, pp. 136–140).

Hertig, E., Maraun, D., Bartholy, J., Pongracz, R., Vrac, M., Mares, I., ... Soares, P. M. M. (2019). Comparison of statistical downscaling methods with respect to extreme events over Europe: Validation results from the perfect predictor experiment of the COST Action VALUE. *International Journal of Climatology*, *39*, 3846–3867. Retrieved 2018-07-30, from https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.5469 doi: 10.1002/joc.5469

Hlinka, J., Hartman, D., Vejmelka, M., Runge, J., Marwan, N., Kurths, J., & Palus, M. (2013). Reliability of Inference of Directed Climate Networks Using Conditional Mutual Information. *Entropy*, *15*(6), 2023–2045. doi: 10.3390/e15062023

Iturbide, M., Bedia, J., Herrera, S., Baño-Medina, J., Fernández, J., Frías, M. D., ... Gutiérrez, J. M. (2019, January). The R-based climate4R open framework for reproducible climate data access and post-processing. *Environmental Modelling & Software*, *111*, 42–54. Retrieved 2023-03-24, from https://www.sciencedirect.com/science/article/pii/S1364815218303049 doi: 10.1016/j.envsoft.2018.09.009

LeCun, Y., Bengio, Y., & Laboratories, T. B. (1995). Convolutional Networks for Images, Speech, and Time-Series. *The Handbook of Brain Theory and Neural Networks*.

Legasa, M. N., Chandler, R. E., & Manzanas, R. (2023, April). *Bayesian Network-Informed Conditional Random Forests for Probabilistic Multisite Downscaling of Precipitation Occurrence* (preprint). Preprints. Retrieved from https://essopenarchive.org/users/603734/articles/633943-bayesian-network-informed-conditional-random-forests-for-probabilistic-multisite-downscaling-of-precipitation-occurrence?commit=66f034dbe288a09503eccfcdd6a4431b4aaba042 doi: 10.22541/essoar.168167381.15060857/v1

Lorenz, E. (1969). Atmospheric predictability as revealed by naturally occurring analogues. *Journal of the Atmospheric Sciences*, *26*, 636–&.

Maraun, D., & Widmann, M. (2018). *Statistical Downscaling and Bias Correction* *for Climate Research* (1st ed.). Cambridge University Press. Retrieved 2022-05-23, from https://www.cambridge.org/core/product/identifier/9781107588783/type/book doi: 10.1017/9781107588783

Maraun, D., Widmann, M., & Gutiérrez, J. M. (2019, July). Statistical downscaling skill under present climate conditions: A synthesis of the VALUE perfect predictor experiment. *International Journal of Climatology*, *39*(9), 3692–3703. doi: 10.1002/joc.5877

Nelder, J., & Wedderburn, R. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society Series A-General*, *135*(3), 370–&. doi: 10.2307/2344614

Quilcaille, Y., Batibeniz, F., Ribeiro, A. F. S., Padrón, R. S., & Seneviratne, S. I. (2023, May). Fire weather index data under historical and shared socioeconomic pathway projections in the 6th phase of the Coupled Model Intercomparison Project from 1850 to 2100. *Earth System Science Data*, *15*(5), 2153–2177. Retrieved 2023-11-02, from https://essd.copernicus.org/articles/15/2153/2023/ (Publisher: Copernicus GmbH) doi: 10.5194/essd-15-2153-2023

R Core Team. (2020). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/

San-Martín, D., Manzanas, R., Brands, S., Herrera, S., & Gutiérrez, J. M. (2016, October). Reassessing Model Uncertainty for Regional Projections of Precipitation with an Ensemble of Statistical Downscaling Methods. *Journal of Climate*, *30*(1), 203–223. Retrieved 2019-06-27, from https://journals.ametsoc.org/doi/full/10.1175/JCLI-D-16-0366.1 doi: 10.1175/JCLI-D-16-0366.1

San-Miguel-Ayanz, J., Moreno, J. M., & Camia, A. (2013, April). Analysis of large fires in European Mediterranean landscapes: Lessons learned and perspectives. *Forest Ecology and Management*, *294*, 11–22. Retrieved 2013-09-10, from http://linkinghub.elsevier.com/retrieve/pii/S0378112712006561 doi: 10.1016/j.foreco.2012.10.050

Turco, M., Rosa-Cánovas, J. J., Bedia, J., Jerez, S., Montávez, J. P., Llasat, M. C., & Provenzale, A. (2018, December). Exacerbated fires in Mediterranean Europe due to anthropogenic warming projected with non-stationary climate-fire models. *Nature Communications*, *9*, 3821. Retrieved 2018-09-28, from http://www.nature.com/articles/s41467-018-06358-z doi: 10.1038/s41467-018-06358-z

van Wagner, C. E. (1987). *Development and structure of the Canadian Forest Fire Weather Index* (Forestry Tech. Rep. No. 35). Ottawa, Canada: Canadian Forestry Service.

Vrac, M., & Friederichs, P. (2015). Multivariate—Intervariable, Spatial, and Temporal—Bias Correction*. *Journal of Climate*, *28*(1), 218–237. Retrieved 2015-04-28, from http://journals.ametsoc.org/doi/abs/10.1175/JCLI-D-14-00059.1

Widmann, M., Bedia, J., Gutiérrez, J., Bosshard, T., Hertig, E., Maraun, D., … Huth, R. (2019, January). Validation of spatial variability in downscaling results from the VALUE perfect predictor experiment. *International Journal of Climatology*, *39*, 3819–3845. Retrieved 2019-02-07, from http://doi.wiley.com/10.1002/joc.6024 doi: 10.1002/joc.6024

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., … others (2016). The fair guiding principles for scientific data management and stewardship. *Scientific data*, *3*(1), 1–9.

Zorita, E., & von Storch, H. (1999). The analog method as a simple statistical downscaling technique: Comparison with more complicated methods. *Journal of Climate*, *12*(8), 2474 - 2489. doi: https://doi.org/10.1175/1520-0442(1999)012⟨2474:TAMAAS⟩2.0.CO;2

Optimization of Convolutional Neural Network models for spatially coherent multi-site fire danger predictions

Óscar Mirones[1], Jorge Baño-Medina[1], Mario Santa Cruz[2], Swen Brands[1], Joaquín Bedia[3,4]

[1]Instituto de Física de Cantabria (IFCA), CSIC-Universidad de Cantabria, Santander, Spain
[2]Predictia Intelligent Data Solutions S.L.
[3]Dept. Matemática Aplicada y Ciencias de la Computación (MACC), Universidad de Cantabria, Santander, Spain
[4]Grupo de Meteorología y Computación, Universidad de Cantabria, Unidad Asociada al CSIC, Santander, Spain

Key Points:

- Convolutional neural networks (CNNs) are compared with classical statistical downscaling methods for Fire Weather Index (FWI) prediction.
- The best CNN setup provides balanced results for all validation metrics, including accuracy, simulation of extremes and spatial consistency.
- Our findings provide a methodological basis for the development of more robust, spatially coherent regional future FWI scenarios.

Corresponding author: Óscar Mirones, `mironeso@unican.es`

Abstract

The accurate prediction of the Fire Weather Index (FWI), a multivariate climate index for wildfire risk characterization, is crucial for both wildfire management and climate-resilient planning. Moreover, consistent multisite fire danger predictions are key for targeted allocation of resources and early intervention in high-risk areas, as well as for "megafire" risk detection. In this regard, Convolutional Neural Networks (CNNs) are known to capture complex spatial patterns in climate data. This study compares different CNN architectures and traditional Statistical Downscaling (SD) methods (regression and analogs) for predicting daily FWI across diverse locations in Spain, considering marginal, distributional and spatial coherence measures for validation. Overall, the CNN-Multi-Site-Multi-Gaussian configuration, which explicitly accounts for the inter-site variability in the output layer structure, showed a superior performance. These insights provide a methodological guidance for the successful application of CNNs in the context wildfire risk assessment, enhancing wildfire response strategies and climate adaptation planning.

Keywords: deep learning, statistical downscaling, Generalized Linear Models, analogs, spatial structure, future wildfire risk assessment.

Plain Language Summary

This study focuses on the Fire Weather Index (FWI), a pivotal climate index for the assessment of wildfire risk. Accurate FWI predictions are vital for wildfire management. This study explores the viability of employing Convolutional Neural Networks (CNNs) as a Statistical Downscaling (SD) technique for precise FWI prediction across diverse locations in Spain in comparison with two conventional SD methodologies: Generalized Linear Models and analogs. Following a cross-validation scheme based on observed daily FWI data, we find that the CNN-Multi-Site-Multi-Gaussian (CNN-MSMG) configuration exhibits noteworthy proficiency in daily FWI prediction. This model explicitly incorporates the covariance structure of the predictands into the CNN architecture, yielding spatially consistent FWI predictions. Furthermore, CNN-MSMG has optimal properties for use in the context of climate change, providing a robust replication of extreme events and extrapolation capabilities if applied to novel climate scenarios. These findings have substantial implications for improving regional-to-local FWI scenarios used to inform vulnerability and impact assessment studies.

1 Introduction

Climate fire danger indices are key to assess and predict the risk of wildfire occurrence and severity. They are based on the integration of daily near-surface temperature, humidity, wind speed and precipitation records (de Groot et al., 2006), and thus provide more accurate wildfire risk forecasts than their input variables alone (see e.g. Dowdy et al., 2009; Fugioka et al., 2009). Beyond the near-term prediction horizon, fire danger indices are also useful to monitor changes in wildfire risk over time. As a result, downscaled fire danger scenarios are essential for vulnerability and adaptation strategies in regional to local applications, since General Circulation Model (GCM) outputs (Eyring et al., 2016) can't provide actionable climate information at these spatial scales (Giorgi et al., 2009). Given their suitability for most impact studies, statistical downscaling (SD, Maraun & Widmann, 2018) of future fire weather scenarios is often required, including *perfect-prognosis* methods (Bedia et al., 2013, see Sec. 2.1) or bias-adjustment tools (Abatzoglou & Brown, 2012; Casanueva et al., 2018). In this case, there are three key aspects to focus on: (1) the reproducibility of extremes, as they can substantially increase wildfire impacts (Turco et al., 2018); (2) extrapolation capability is vital for predicting of out-of-sample values, since fire danger conditions are expected to change drastically in many regions (Bedia et al., 2015; Quilcaille et al., 2023), and (3) the ability to keep the predictand's (FWI) spatial consistency is important to identify potentially hazardous fire risk scenarios affecting a wide geographical area, thereby increasing the odds of “fire clusters” with catastrophic potential (San-Miguel-Ayanz et al., 2013).

While most standard SD methods show good performance in at least one of these 3 aspects (Maraun et al., 2019), none of them is able to effectively accomplish all of them. In this context, the classical analog method (Lorenz, 1969; Zorita & von Storch, 1999; Brands et al., 2011) is still a competitive benchmark due to its ability to model both the extremes and the spatial structure (Widmann et al., 2019). However, if applied in its original form (Zorita & von Storch, 1999), this method fails to extrapolate beyond observed extremes, limiting its use for climate change applications (Bedia et al., 2013). In this sense, regression-based models are the better choice since they allow for better extrapolation (Baño-Medina et al., 2021; Balmaceda-Huarte et al., 2023) but, on the downside, they usually underestimate the extremes (Hertig et al., 2019). A further disadvantage of standard regression models (including Generalized Linear Models, GLMs) is their single-site structure unable to effectively model the spatial dependencies of the predictand variable(s). Other proposed alternatives combine the benefits of perfect-prog models and Weather Generators (PP-WG, see e.g.: Cannon, 2008; Carreau & Vrac, 2011), allowing to estimate the uncertainty of a local predictand variable and even to sample from the conditional distributions to recover the variability of the time series. To date, however, and with some exceptions (Legasa et al., 2023), most of these studies have focused on the estimation of uni-variate, single-site distributions, thereby not taking into account the spatial structure of the predictand nor its relationships with other predictand variables.

In this regard, deep learning methods, and in particular Convolutional Neural Networks (CNNs, LeCun et al., 1995) may offer a suitable alternative to meet these requirements with an adequate tuning. CNNs perform convolutions with learnable kernels over the spatial dimensions of atmospheric fields, inferring a non-linear mapping between low-resolution predictor fields and high-resolution predicand fields that has been shown to outperform conventional SD methods in many aspects (Baño-Medina et al., 2020). Regarding extrapolation ability, CNNs can produce plausible future climate change scenarios (Baño-Medina et al., 2021), comparable to those provided by dynamical downscaling (Baño-Medina et al., 2022). For a better reproducibility of extremes, parametric-CNNs (P-CNNs, Sec. 2.3) can estimate the parameters of conditional distributions given certain atmospheric conditions. As in the PP-WG approach, an adequate CNN architecture is able to estimate the parameters of the whole joint (multi-site) probability structure of the covariance matrix and can coherently reproduce the spatial structure of the predicted fire danger series across all predictand locations.

In this study we describe different CNN-based regression models for multi-site extreme fire danger assessment under climate change conditions, based on Canadian Fire Weather Index (van Wagner, 1987) records at 29 locations in Spain. We deploy three alternative CNN topologies based on the PP-WG approach that estimate either uni-variate or multi-variate Gaussian distributions on daily timescale. The validation is based on specific measures of extreme reproducibility and spatial coherence, using classical SD methods (analogs and GLMs) as benchmark.

2 Data and Methods

2.1 Predictor set

Perfect-prognosis SD establishes empirical relationships between 1) the variability of atmospheric variables operating on large scales, typically derived from a global reanalysis with a resolution similar to that offered by current global climate models (Eyring et al., 2016) and 2) the local-scale variability of the predictand of interest (here: FWI) as represented by in-situ observations or gridded observational datasets derived there-from. Once the SD model is calibrated, the learnt relationships can be applied to GCM (instead of reanalysis) predictors in order to derive local climate change projections if, ideally, the following requirements are fulfilled: The predictor variables should be realistically represented by the GCMs (Fernandez-Granja et al., 2021; Brands, 2022; Brands et al., 2023), should carry the climate change signal and be physically related with the local variable. In addition, the SD model should be capable to extrapolate the learnt relationships to altered/unobserved climate regimes (Gutiérrez et al., 2013). For the case of FWI downscaling, the predictor selection under such non-perfect circumstances has been explored in a previous study we built upon here (Bedia et al., 2013). Namely, we use daily-mean 2m air temperature, the zonal and meridional wind velocity components at 10m, as well as temperature and specific humidity a the 850 hPa pressure level, covering a spatial domain centered on the target region. These data have been retrieved from ERA-Interim (Dee et al., 2011) for the period 1985–2011 (see Table A2).

2.2 Predictand: Fire Weather Index observations

The FWI is a multivariable index, and therefore the downscaling approach must carefully consider the physical consistency of its input variables. When these are separately downscaled, inter-variable dependencies may be modified leading to spatio-temporal inconsistencies in the simulated output fields that would affect the coherence of the output FWI predictions (see e.g.: Vrac & Friederichs, 2015). This uncertainty source is here circumvented by using the FWI index, rather than its components, as sole predictand variable. To this end, in-situ observations from 29 weather stations of the Spanish Meteorological Agency (AEMET) were obtained, recording the required data for FWI calculation. The AEMET dataset provides instantaneous values of temperature, relative humidity and wind speed at 13:00 UTC, and last 24-h accumulated precipitation, recorded at 07:00 UTC. FWI calculation follows the methodology described by Bedia et al. (2013). For an optimal dataset completeness, we consider the calibration period 1985–2011.

2.3 Convolutional Neural Networks

To identify the key factors of the FWI spatial structure, we deploy three CNN architectures of increasing topological complexity (see Fig. 1). The backbone of these topologies builds on well tested CNNs known to outperform both analogs and GLMs in temperature and precipitation downscaling (Baño-Medina et al., 2020). The hidden structure consists of 3 convolutional layers followed by two fully-connected ones. The convolutional layers consist of a block of three layers with 50, 25, and 10 (3×3) kernels respectively, while the fully-connected (dense) layers each contain 50 neurons for the CNN-Multi-Site (CNN-MS) and CNN-Multi-Site-Gaussian (CNN-MSG) configuration, or 200 neurons for the CNN-Multi-Site-Multi-Gaussian (CNN-MSMG) version (see Fig. 1). A nonlinear ReLU activation function is applied between the layers. The output, where we find the main differences across models, is a dense fully-connected network with a linear activation function in CNN-MSG and CNN-MSMG. In the case of CNN-MS, the output layer consists of 29 neurons, each neuron corresponding to a point location (Table A1), yielding deterministic FWI predictions at each site.

In order to improve the representation of FWI extremes, we introduce modifications to the P-CNN structures in CNN-MSG and CNN-MSMG. In CNN-MSG, the output is modeled stochastically using an *independent* Gaussian distribution to estimate the parameters of $\mathcal{N}(\mu, \sigma)$ (mean and standard deviation respectively). Thus, for each of the 29 stations, two pairs of neurons are added to the output layer, one for each parameter. CNN-MSMG, in turn, aims to estimate the parameters of a *multivariate* Gaussian distribution $\mathcal{N}(\mu, \Sigma)$. In this case, μ denotes the mean and Σ represents the covariance matrix. Therefore, the output layer consists of a pair of neuron vectors, with sizes 29 and 464 respectively. The 29 neurons represent the μ parameters, while the 464 neurons correspond to the number of unique parameters estimated in the covariance matrix Σ. The aim of this multivariate Gaussian setup is to describe the values at each location as a correlated set, unlike the outcome of an independent Gaussian distribution, where the predictions at each site are independent of each other. The general architecture scheme for each P-CNN configuration is outlined in Fig. 1. To avoid model overfitting, ensure robustness and optimize parameter tuning and CNN architectures, all SD models have been fit following a cross-validation procedure comprising 4 temporal blocks spanning the periods 1985-1991, 1992-1998, 1999-2004 and 2005-2011. The loss functions used are the Mean Square Error (MSE) for CNN-MS, the negative log-likelihood of the independent Gaussian distribution for CNN-MSG, and the negative log-likelihood of the multivariate Gaussian distribution for CNN-MSMG. The benchmark SD methods (analogs and GLMs) have been fitted following the same cross-validation scheme (see Appendix B for additional details on these methods).

2.4 Validation

Here, the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), as well as the quantile-quantile plot (QQ-plot), are used to validate the similarity of the temporal sequence and empirical distribution between the downscaled and observed daily FWI time series at a given station (Déqué, 2011). Apart from these classical marginal validation metrics, the focus is put on *spatial* coherence, as outlined in the following.

2.4.1 Location Correlogram

To qualitatively evaluate whether the distinct SD methods are able to reproduce the spatial correlation structure of the observed FWI, we use the *location correlogram* (Herdin et al., 2005). Firstly, the $n = 29$ observed daily *in-situ* FWI time series from the complete station network are correlated with each other for all possible combinations (i.e. $n \times \frac{n-1}{2}$ pairs) using Spearman's rank correlation coefficient and the resulting coefficients are plotted against the respective pairwise station distances. Then a local 2nd-order polynomial ("loess") is fitted to the scatter-plot, resulting in a curve that depicts the spatial correlation structure of the observed FWI. As a quantitative summary measure, we use the *correlation length* (CL), defined as the geographical distance corresponding to the point of intersection of a given correlation threshold with the fitted loess line. A threshold of $\rho = 0.4$ has proven to be most suitable for characterizing the spatial FWI structure in this study (Table Appendix C), and the overall results are robust to changes in this choice. After applying the same method to the downscaled time series from each of the thee SD methods, the CL bias between the simulated and observed spatial structure is calculated as an overall measure of the methods' capability to reproduce the spatial coherence of the observed FWI (Widmann et al., 2019).

2.4.2 Mutual information for FWI90

FWI extremes are particularly relevant for fire danger assessment. As a result, from the spatial consistency point of view, the users of downscaled FWI values will be primarily interested in a realistic representation of joint higher-percentile FWI exceedances among locations (see e.g. Bedia et al., 2014). To this aim, Mutual Information (MI) provides a suitable measure of the dependence between two random variables X, Y (here, predictions at two locations) that is unaffected by their marginal distributions and quantifies the amount of mutual information between them (see e.g. Hlinka et al., 2013). For two discrete random variables X and Y it is defined as:

$$MI(X,Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \cdot \log\left(\frac{p(x,y)}{p(x) \cdot p(y)}\right) \tag{1}$$

MI is zero if the two events are independent, i.e. if $p(X,Y) = p(X) \cdot p(Y)$, non-negative ($MI(X,Y) \geq 0$) and symmetric ($MI(X,Y) = MI(Y,X)$).

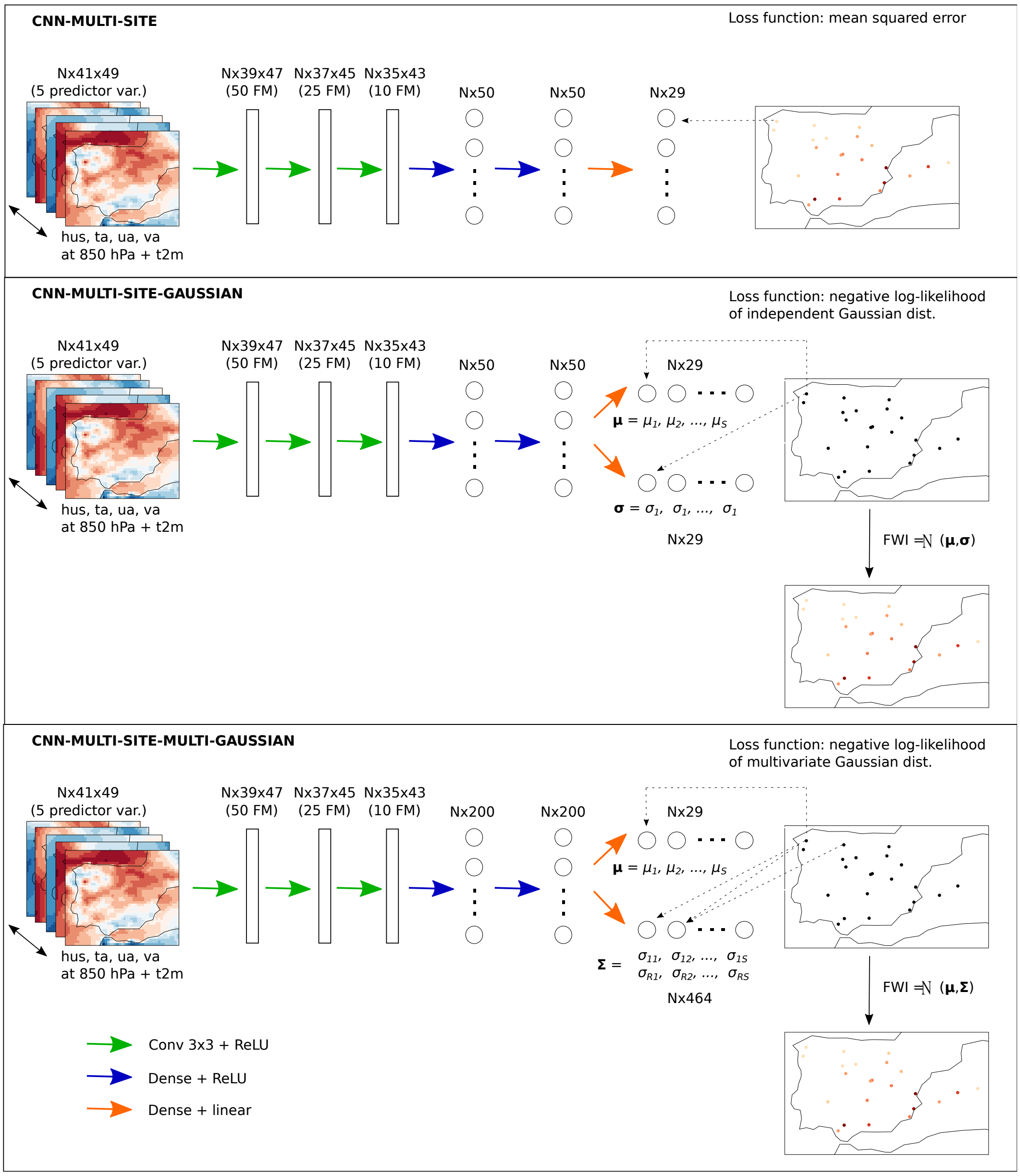

Figure 1. Scheme of the convolutional neural network architecture used in this study. The network includes a first block of three convolutional layers with 50, 25 and 10 (3x3) kernels, respectively, followed by two fully connected dense layers with 50 or 200 neurons each, depending on the model. For CNN-MSG and CNN-MSMG, the output is modeled through and independent Gaussian distribution and a multivariate Gaussian distribution respectively, and the corresponding parameters are estimated by the network, obtaining FWI as final product, either deterministically (CNN-MS) or stochastically (CNN-MSG and CNN-MSMG). The output layer is activated linearly while the previous layers of the network are activated non-linearly.

Here, we consider the binary variables X, Y at each location, stating whether the FWI values x_i, y_i lie above or below the 90th percentile for each pair of locations. We then calculate the MI for each pair of locations following the definition above (eq. 1). As for the correlograms (Sec. 2.4.1), we plot each MI_{ij} against the distance of the locations i, j and fit a degree-2 loess curve to the resulting scatter-plots. We then define MI thresholds for calculating the MI lengths (MIL) in observations and for the different downscaling methods. We use a MI threshold of 0.05, yielding results comparable to those obtained from CL analysis , and focusing on the identification of potential new information about each methods' performance (Fig. C2). As in CL analysis, the MIL biases are calculated as the difference between predicted and observed MILs.

3 Results

The results presented correspond to the generic June to September fire season, representative of the Iberian Peninsula (JJAS, see e.g.: Bedia et al., 2014). It's important to highlight that the models were calibrated using the entire annual dataset. However, a subset comprising the JJAS season was used to present the results relevant for fire danger assessment in this region. In the following subsections, we categorized the station network into three groups based on proximity to the sea and general climate conditions: Atlantic, Coastal Mediterranean, and Continental Mediterranean (see Table A1). The suitability of this classification for FWI aggregation is confirmed by the results obtained with the mutual information measure (Sec. 3.2.2).

3.1 Predictive accuracy and distributional similarity

In agreement with previous studies (Brands et al., 2011), the SD methods' accuracy is *generally* lower at continental sites than near the coast. Overall, all methods perform similarly with regard to predictive accuracy, summarized in terms of the RMSE of FWI90 predictions in Fig. C1 (Appendix C).

However, the distributional characteristics of the predictions differ largely among methods. The quantile-quantile (QQ) plots shown in Fig. 2 compare the observed and predicted empirical FWI distributions. While all methods perform well in predicting the mean FWI, disparities emerge at higher percentiles, crucial for fire danger analysis. The benchmarking analog approach produces best results for the right tail, closely followed by multivariate CNN-MSMG, showing similar results across regions. Conversely, GLM and CNN-MS consistently underestimate high percentile FWI events, failing to realistically represent most dangerous situations. CNN-MSG also achieves good results, comparable to CNN-MSMG, but is outperformed by the latter in the Coastal Mediterranean and Atlantic regions. Notably, in the Atlantic region, CNN-MSG unrealistically inflates the highest FWI percentiles and underestimates most of the FWI distribution. On the contrary, in the Continental Mediterranean region, CNN-MSG performs slightly better than CNN-MSMG for higher percentiles.

In order to obtain a quantitative measure of distributional deviance with respect to the observed distribution, we calculate the RMSE considering the differences between predicted and observed quantiles of i) the entire FWI times series and ii) the FWI time series values exceeding the station-specific 90th percentile (FWI90, Fig. 2). Excluding the results for the analog method, lowest RMSE values for both indicators are obtained either by CNN-MSG or CNN-MSMG. Regardless of the specific target region, the former approach demonstrates significantly better performance compared to the latter in terms of FWI, and only exhibits a slight decrease in performance for FWI90. Specifically, when emphasizing FWI90, the CNN-MS and GLM models exhibit noticeably poorer performance compared to the analog benchmark. In contrast, the results for CNN-MSG and CNN-MSMG models are considerably better in this regard.

Overall, the reference analog method performs best in representing the distribution of the daily FWI in most cases. CNN-MSG and CNN-MSMG perform slightly worse, with CNN-MSG slightly overemphasizing severe FWI frequencies in the Atlantic region. The analog method performs best overall, but it's applicability in climate changes studies is limited due to its inability to extrapolate predictions outside the observed range. Conversely, both CNN-MSG and CNN-MSMG are competitive alternative methods in terms of distributional similarity. In the Secs. 3.2.1 and 3.2.2, we assess whether these conclusions hold for the *spatial structure* of the simulated mean and extreme FWI fields.

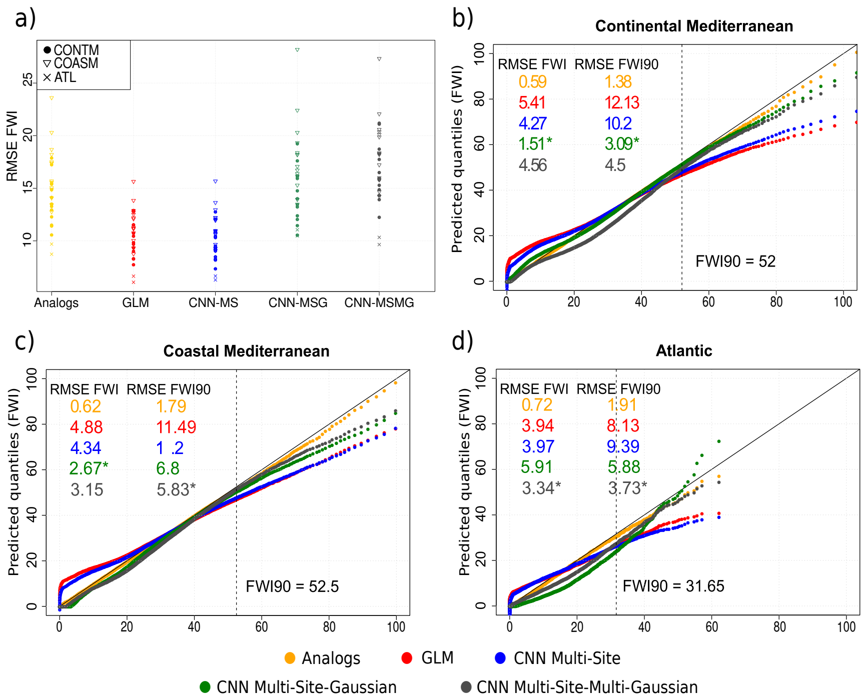

Figure 2. FWI RMSE per station and method (a) and Q-Q plots for the analog method, GLMs and the distinct CNN models (b - d). The figure is divided into 4 panels. The (a) panel refers to the RMSE for the simulated FWI per station and method distinguishing the regions by symbols. The remaining panels refer to the station subsets of (b) the Continental Mediterranean, (c) Coastal Mediterranean and (d) Atlantic regions. The method-specific distributional RMSE for the simulated FWI and FWI90 are indicated in the upper left corners of each panel and the best performing method is marked with an asterisk (excluding the benchmarking analog method). The dashed vertical line indicates the observed FWI90.

3.2 Spatial validation results

3.2.1 Dependence of inter-station relationships on distance

The temporal correlation coefficients' dependence on distance, analyzed as described in Sec. 3, is depicted in Fig. 3. As expected, the observed strength of the relations decreases exponentially with increasing distance between the stations and stabilizes around $rho = 0.1$, the CL for $\rho = 0.4$ being located at 208.30 km (panel *a)*, grey curve). The corresponding point clouds and polynomials for the SD methods are depicted in red in panels *b)* through *f)*, where the respective validation measures are also indicated (see upper right corners and also Appendix C). The exponential decay seen in the observations is reproduced more or less successfully by all SD methods except CNN-MSG, that produces far too weak short distance relationships, failing to reproduce any spatial structure in the data. The analog method is, as expected, most successful in reproducing the observed correlation structure, closely followed by CNN-MSMG, while GLM and CNN-MS consistently overestimate pairwise correlations. Among the suitable methods (i.e. excluding CNN-MSG), the medium-to-long-distance correlations are overestimated by all methods, particularly by GLM and CNN-MS. The stronger short-distance correlations are also generally overestimated, but to a lesser degree, and they are almost perfectly met by the analog method and closely approximated by CNN-MSMG.

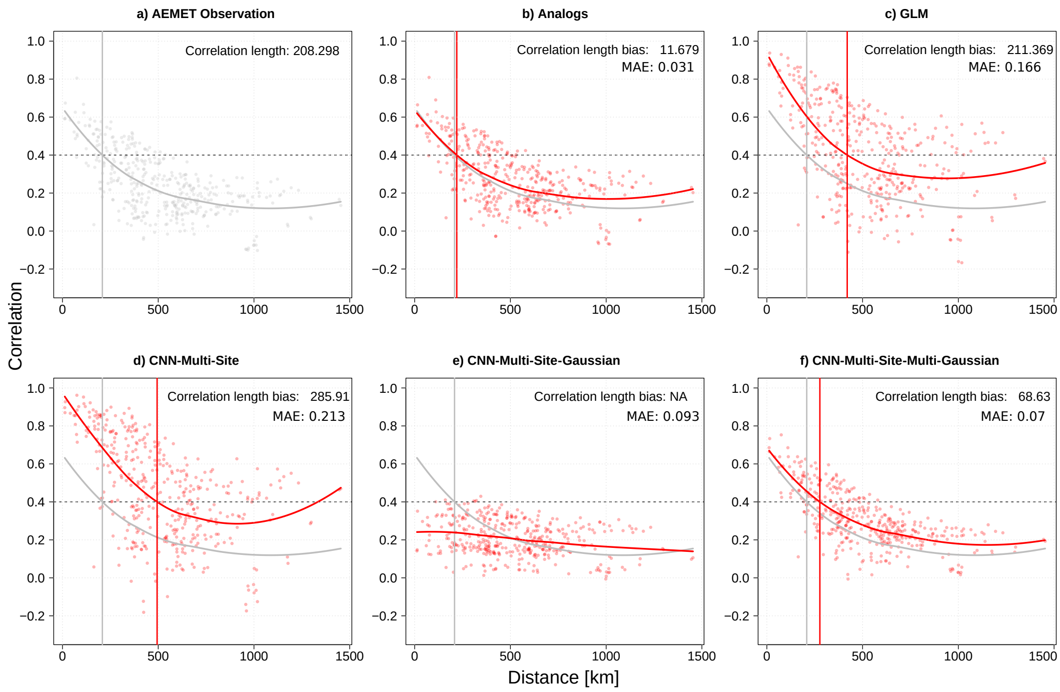

Figure 3. Correlograms illustrating the daily JJAS FWI dependence of the inter-station relationships, described by the Spearman correlation coefficients among all station pairs (y-axis), against their respective distances in kilometers (x-axis). The correlograms correspond to the observations (panel *a*) and to each SD method tested (panels *b* to *f*); the grey loess line of the observations correlogram is included in all panels for visual comparison. It is also displayed the observed Correlation Length (CL, panel *a*)) and the CL bias and MAE for each SD method (in panels *b* to *f*). Here, the MAE is calculated as the difference (in absolute value) between predicted and observed correlation coefficients for each station pair.

3.2.2 Mutual information for fire weather extremes

In Fig. 4 (upper panel), we present the mutual information (MI) values obtained from the observational network in the upper triangle (a1), compared with those produced by CNN-MSMG, the best performing SD method for this metric (the benchmarking analog method is excluded), in the lower triangle (a2). The stations are grouped into characteristic climate regimes as described in Sec. 3.1. Geographical proximity translates into higher MI values, as the case of Vigo and Santiago de Compostela (NW Iberia, $MI = 0.11$), Barajas and Retiro (Madrid, central Spain, $MI = 0.14$), or Valencia and Valencia-Airport (SE, Mediterranean coast, $MI = 0.13$). Furthermore, several climatologically homogeneous regions can be identified in the matrix, yielding visually discernible clusters of high MI values, e.g. the Soria-Valladolid-Salamanca-Zamora cluster pertaining to the central-north Iberian high plains. The MI pattern obtained from CNN-MSMG is similar to that seen in observations and is thus approximately symmetric (compare Fig. 4-a2 with a1). Nevertheless, CNN-MSMG somewhat overestimates the spatial dependencies, indicated by slightly higher MI values than those obtained from observations, and also reflected by regional clusters not seen in observations (e.g. Ciudad Real, Badajoz and Granada).

In Fig. 4b and c, we illustrate the MI biases relative to the observations for CNN-MS and CNN-MSG (b1, b2), as well as for CNN-MSMG and GLM (c1, c2). We focus on station pairs with MI values $\geq$ 0.05 in observations, thus discarding already independent station pairs (blank matrix cells). Since the MI bias of the analog method is negligible for all station pairs, the corresponding results are shown in Appendix Appendix C. The MI bias of CNN-MSMG is below the 0.05 threshold for most station pairs, with a few exceptions with both positive or negative values (Fig. 4c1). CNN-MS exhibits a consistent positive bias, consistently overestimating the spatial dependence of extreme FWI events (Fig. 4b1). CNN-MSG, in turn, systematically underestimates these dependencies (Fig. 4b2), yielding a lower bias magnitude than for CNN-MS. The GLM approach tends to overestimate these dependencies, albeit to a lesser extent than CNN-MS (compare Fig. 4c2 with b1).

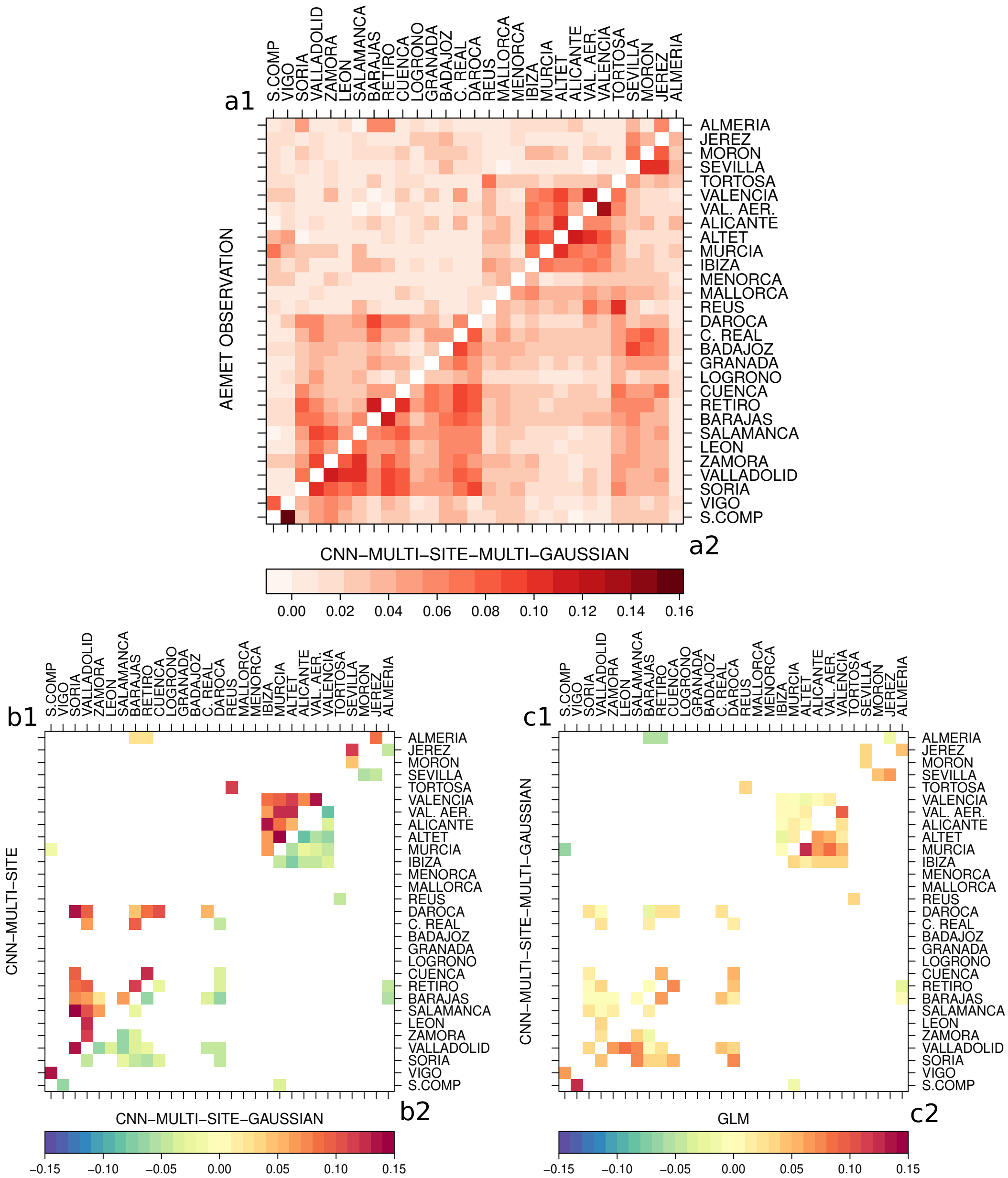

Figure 4. MI matrices for FWI90 events during the fire season (JJAS) obtained from observations (upper triangle in upper panel, a1) and from the best performing SD model (CNN-MSMG, lower triangle in upper panel, a2). Panels b an c show the MI biases for 4 remaining SD methods with respect to the observations (b1: CNN-MS, b2: CNN-MSG, c1: CNN-MSMG, c2: GLM). In b and c panels, only station pairs with MI ≥ 0.05 in the observations are shown.

4 Conclusions

We conducted a comprehensive comparison of various Convolutional Neural Network (CNN) architectures in contrast to two established statistical downscaling (SD) methods, specifically Generalized Linear Models (GLMs) and analogs. Our assessment focused on evaluating their performance in terms of predictive accuracy, distributional congruence, and spatial coherence for Fire Weather Index (FWI) predictions across 29 locations in Spain.

Among the diverse CNN architectures scrutinized, CNN-MSMG demonstrated the most favorable outcomes across these validation criteria. This setup considers the multivariate nature of the predictions in the output layer yielding a predicted covariance matrix that explicitly accounts for the inter-site variability. It exhibited a notable capacity to accurately represent observed FWI distributions at both single-point and multisite scales, closely aligning with the outcomes of the benchmarking analogs method. Notably, the analogs method, by design, upholds multisite spatial consistency without alteration, at the cost of limitations for extrapolation in climate change conditions that can be overcome by the rest of methods tested. In contrast CNN-MS (multisite CNN) and GLMs yielded poorer predictive accuracy and consistently overestimated the spatial dependence among sites. In turn, CNN-MSG (multisite Gaussian) attained good results in terms of single-site validation, but proved inefficient in modelling the spatial structure, essentially behaving like a single-site weather generator.

The results presented emphasize the importance of parameter tuning for CNN development in the context of statistical downscaling in order to produce credible predictions. In the particular case of FWI, an adequate tuning is needed in order to ensure actionable climate information for the prevention of wildfire impacts, and this study provides a methodological guidance for the successful application of CNNs to this aim.

5 Open Research

We follow the FAIR principles (Findability, Accesibility, Interoperability and Reuse, Wilkinson et al. (2016)) and publish the code (DOI: `10.5281/zenodo.8387558`) and the data (DOI: `10.5281/zenodo.8381437`) required to replicate the results presented in this manuscript. We build on the R based (R Core Team, 2020) framework *climate4R* (Iturbide et al., 2019) to digest, manipulate, downscale (see also Bedia et al., 2020) and visualize (Frías et al., 2018) the climate data. For the deep learning models, we lean on *downscaleR.keras*, a library that integrates *tensorflow* (Abadi et al., 2015) and *keras* (Gulli & Pal, 2017) into the *climate4R* framework (Baño-Medina et al., 2020).

Appendix A Input Data

Table A1 is a summary of the AEMET weather station database. We also indicate their corresponding climatic zone, according to the spatial aggregation summarizing the results in Sec. 3.1 of the main text. The *Short name* column indicates the abbreviated labels used throughout the article figures.

Station name	Short name	Lon	Lat	Altitude	Climatic region
REUS–AEROPUERTO	REUS	1.18	41.15	71	COASM
SANTIAGO DE COMPOSTELA–LABACOLLA	S.COMP	-8.41	42.89	370	ATL
VIGO–PEINADOR	VIGO	-8.62	42.24	261	ATL
SORIA	SORIA	-2.48	41.77	1082	CONTM
VALLADOLID	VALLADOLID	-4.75	41.64	735	CONTM
ZAMORA	ZAMORA	-5.73	41.52	656	CONTM
LEÓN–VIRGEN DEL CAMINO	LEÓN	-5.65	42.59	916	CONTM
SALAMANCA–MATACÁN	SALAMANCA	-5.50	40.96	790	CONTM
MADRID–BARAJAS	BARAJAS	-3.56	40.47	609	CONTM
MADRID–RETIRO	RETIRO	-3.68	40.41	667	CONTM
CIUDAD REAL	C. REAL	-3.92	38.99	628	CONTM
BADAJOZ–TALAVERA LA REAL	BADAJOZ	-6.81	38.88	185	CONTM
GRANADA–AEROPUERTO	GRANADA	-3.79	37.19	567	CONTM
SEVILLA–SAN PABLO	SEVILLA	-5.88	37.42	34	COASM
MORÓN DE LA FRONTERA	MORÓN	-5.61	37.16	87	COASM
JEREZ DE LA FRONTERA–AEROPUERTO	JEREZ	-6.06	36.75	27	COASM
ALMERÍA–AEROPUERTO	ALMERÍA	-2.36	36.85	21	COASM
MURCIA–SAN JAVIER	MURCIA	-0.80	37.79	4	COASM
ALICANTE–EL ALTET	ALTET	-0.57	38.28	43	COASM
ALICANTE	ALICANTE	-0.49	38.37	81	COASM
CUENCA	CUENCA	-2.14	40.07	945	CONTM
VALENCIA–AEROPUERTO	VAL. AER.	-0.47	39.49	69	COASM
VALENCIA	VALENCIA	-0.37	39.48	11	COASM
LOGROÑO–AGONCILLO	LOGROÑO	-2.33	42.45	353	CONTM
DAROCA	DAROCA	-1.41	41.11	779	CONTM
TORTOSA	TORTOSA	0.49	40.82	44	COASM
PALMA DE MALLORCA–SON SAN JUAN	MALLORCA	2.74	39.56	8	COASM
MENORCA–MAÓ	MENORCA	4.22	39.85	91	COASM
IBIZA/ES CODOLA	IBIZA	1.38	38.88	6	COASM

Table A1. Selected stations of the Spanish AEMET network, indicating their position in decimal degrees and meters above sea level (Datum WGS-84). The abbreviations corresponding to the climatic regions in the column are as follows: ATL for Atlantic, COASM for Coastal Mediterranean, and CONTM for Continental Mediterranean.

Table A2 provides a summary of the reanalysis fields used as predictors in this study. The predictor set has been chosen following the methodology for FWI downscaling presented by Bedia et al. (2013), but replacing relative humidity by specific humidity, the former being not directly available in some model simulation databases. The spatial extent of these fields covers a bounding box centered over the Iberian Peninsula, limited by the geographical coordinates $-10°/15°$E, $35°/45°$N.

Appendix B Benchmarking SD methods

We next provide further methodological details on the standard SD methods used as benchmarks in this study. Both are implemented in the R package `downscaleR` (Bedia et al., 2020), part of the climate4R framework for climate data analysis and visualization (Iturbide et al., 2019, `https://github.com/SantanderMetGroup/climate4R`).

Code	**Name**	**units**
T2M	Air Temperature at surface	K
T850	Air Temperature at 850 hPa	K
HUS850	Specific humidity at 850 hPa	$g\,kg^{-1}$
UA850	U-wind at 850 hPa	$m\,s^{-1}$
VA850	V-wind at 850 hPa	$m\,s^{-1}$

Table A2. Predictor variables used in this study, selected from the predictor combination proposed for statistical downscaling of FWI in Bedia et al. (2013). Note that for convenience, relative humidity at 850 hPa has been replaced by specific humidity, more commonly available in GCM datasets. All fields are daily mean values.

B1 Generalized lineal models

GLMs (Nelder & Wedderburn, 1972) are an extension of the classical linear regression that models the expected value of a random predictand variable for different types of probability distributions and link functions. This makes them a versatile tool for modeling a wide range of data types and situations, and therefore extensively used in SD applications (see e.g.: Chandler & Wheater, 2002; Gutiérrez et al., 2019). Here, the response variable is assumed to follow a Gaussian distribution. The relationship between the linear predictor $g(\mu)$ and the expected value of FWI is defined by the identity link function, so the linear predictor directly models the mean FWI, where $g(\mu)$ is defined as $g(\mu) = \mathbf{X}\beta$, where $\mathbf{X}$ is the design matrix containing the predictor variables (Sec. 2.1), and β is the vector of coefficients, estimated by maximum likelihood based on the probability density function of the Gaussian distribution using a least-squares iterative algorithm implemented in the R package `stats`(R Core Team, 2020). Furthermore, predictor configuration is such that only local information is used for training at each site. Here, an optimal number of 16 closest grid-points to each predictand point-location are retained to construct the local predictor set (Bedia et al., 2020), after testing different neighbourhood sizes using cross-validation (Sec. 2.3).

B2 Analogs

The analog method is a simple yet powerful downscaling technique which assumes that similar (or analog) atmospheric patterns (predictor set $\mathbf{X}$) over a region originates similar local meteorological outcomes (daily FWI) for a particular location or set of locations (Sec. 2.2). In this study, we use the standard deterministic nearest neighbor method analog technique based on the Euclidean distance, considering the complete fields to compute distances and only the first closest nearest closest analog for prediction (San-Martín et al., 2016), similar to the standard 'ANALOG' method of the VALUE intercomparison experiment (described in Gutiérrez et al., 2019, A.2), and considering the implementation described in Bedia et al. (2020). Note that using the complete fields as predictors ensures the maximum spatial coherence of the predictions among stations, since the same analog dates are chosen in each case for every point-location (see e.g. Widmann et al., 2019).

Appendix C Results

This section contains additional results as indicated in the figure captions.

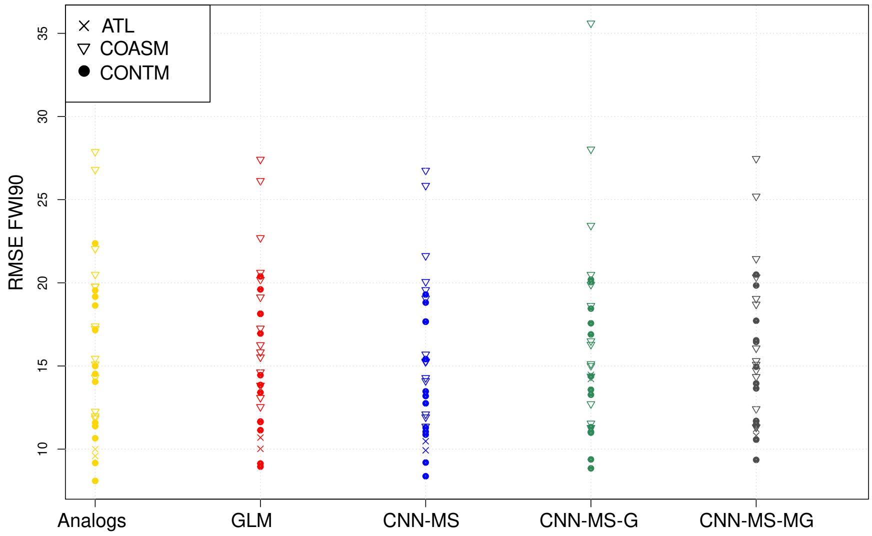

Figure C1. RMSE for the simulated FWI90 per station and method distinguishing the regions by symbols.

	CL	MIL	CL Bias	MIL Bias
AEMET_13UTC_FWI	208.30	168.22		
Analogs			11.68	-2.45
CNN-MS			285.91	259.81
CNN-MSG			*NA*	*NA*
CNN-MSMG			**68.63**	**119.67**
GLM			211.37	146.92

Table C1. The columns display the CL and MIL values for the reference observations, as well as the CL and MIL biases for the models, measured in kilometers (km). The lowest CL and MIL biases (excluding the benchmarking analogs method) are highlighted in boldface.

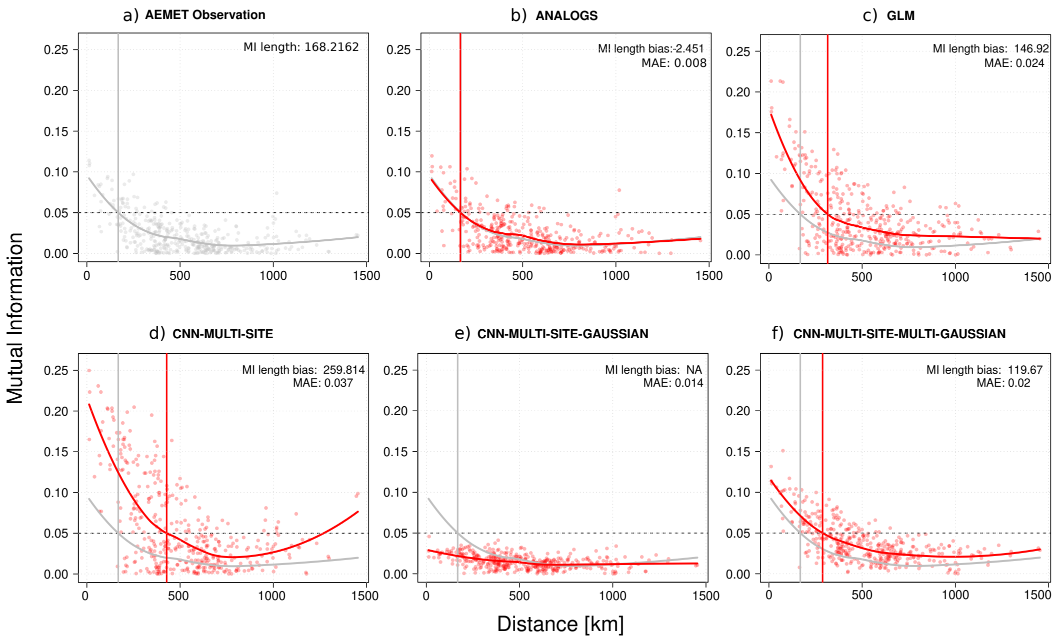

Figure C2. Mutual Information diagrams for FWI90 for fire season (JJAS) showing the mutual information of the FWI90 time series for each pair of stations against their geographical distances. The MI and MI length for the reference observations are shown in the upper left panel. In the rest of the panels, the MI length bias and the MAE are indicated at the top right of the panel.

Acknowledgments

This paper is part of the R+D+i project CORDyS (PID2020-116595RB-I00) with funding from the Spanish Ministry of Science MCIN/AEI/10.13039/501100011033. O.M. has received research support from grant PID2020-116595RB-I00 funded by MCIN/AEI /10.13039/501100011033. SB and JB acknowledge funding by the Ministry for the Ecological Transition and the Demographic Challenge (MITECO) and the European Commission NextGenerationEU (Regulation EU 2020/2094), through CSIC's Interdisciplinary Thematic Platform Clima (PTI-Clima).

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., … Zheng, X. (2015). *TensorFlow: Large-scale machine learning on heterogeneous systems.* Retrieved from https://www.tensorflow.org/ (Software available from tensorflow.org)

Abatzoglou, J., & Brown, T. (2012). A comparison of statistical downscaling methods suited for wildfire applications. *International Journal of Climatology*, *32*, 772–780. doi: 10.1002/joc.2312

Balmaceda-Huarte, R., Baño-Medina, J., Olmo, M. E., & Bettolli, M. L. (2023). On the use of convolutional neural networks for downscaling daily temperatures over southern south america in a climate change scenario. *Climate Dynamics*, 1–15.

Baño-Medina, J., Manzanas, R., Cimadevilla, E., Fernández, J., González-Abad, J., Cofiño, A. S., & Gutiérrez, J. M. (2022, September). Downscaling multi-model climate projection ensembles with deep learning (DeepESD): contribution to CORDEX EUR-44. *Geoscientific Model Development*, *15*(17), 6747–6758. Retrieved 2023-03-24, from https://gmd.copernicus.org/articles/15/6747/2022/ (Publisher: Copernicus GmbH) doi: 10.5194/gmd-15-6747-2022

Baño-Medina, J., Manzanas, R., & Gutiérrez, J. M. (2020, April). Configuration and intercomparison of deep learning neural models for statistical downscaling. *Geoscientific Model Development*, *13*(4), 2109–2124. Retrieved 2023-03-24, from https://gmd.copernicus.org/articles/13/2109/2020/ (Publisher: Copernicus GmbH) doi: 10.5194/gmd-13-2109-2020

Baño-Medina, J., Manzanas, R., & Gutiérrez, J. M. (2021, December). On the suitability of deep convolutional neural networks for continental-wide downscaling of climate change projections. *Climate Dynamics*, *57*(11), 2941–2951. Retrieved 2023-03-24, from https://doi.org/10.1007/s00382-021-05847-0 doi: 10.1007/s00382-021-05847-0

Bedia, J., Baño-Medina, J., Legasa, M. N., Iturbide, M., Manzanas, R., Herrera, S., … Gutiérrez, J. M. (2020, April). Statistical downscaling with the downscaleR package (v3.1.0): contribution to the VALUE intercomparison experiment. *Geoscientific Model Development*, *13*(3), 1711–1735. Retrieved 2020-04-03, from https://www.geosci-model-dev.net/13/1711/2020/ doi: 10.5194/gmd-13-1711-2020

Bedia, J., Herrera, S., Camia, A., Moreno, J. M., & Gutierrez, J. M. (2014). Forest Fire Danger Projections in the Mediterranean using ENSEMBLES Regional Climate Change Scenarios. *Climatic Change*, *122*, 185–199. doi: 10.1007/s10584-013-1005-z

Bedia, J., Herrera, S., Gutierrez, J., Benali, A., Brands, S., Mota, B., & Moreno, J. (2015). Global patterns in the sensitivity of burned area to fire-weather: implications for climate change. *Agricultural and Forest Meteorology*, *214–215*, 369–379. doi: 10.1016/j.agrformet.2015.09.002

Bedia, J., Herrera, S., San-Martin, D., Koutsias, N., & Gutiérrez, J. M. (2013, September). Robust projections of Fire Weather Index in the Mediterranean using statistical downscaling. *Climatic Change*, *120*, 229–247. Retrieved from http://link.springer.com/article/10.1007/s10584-013-0787-3 doi: 10.1007/s10584-013-0787-3

Brands, S. (2022). A circulation-based performance atlas of the cmip5 and 6 models for regional climate studies in the northern hemisphere mid-to-high latitudes. *Geoscientific Model Development*, *15*(4), 1375–1411. Retrieved from https://gmd.copernicus.org/articles/15/1375/2022/ doi: 10.5194/gmd-15-1375-2022

Brands, S., Fernández-Granja, J. A., Bedia, J., Casanueva, A., & Fernández, J. (2023). A global climate model performance atlas for the southern hemisphere extratropics based on regional atmospheric circulation patterns. *Geophysical Research Letters*, *50*(10), e2023GL103531. doi: https://doi.org/10.1029/2023GL103531

Brands, S., Taboada, J., Cofiño, A., Sauter, T., & Schneider, C. (2011, 08). Statistical downscaling of daily temperatures in the nw iberian peninsula from global climate models: Validation and future scenarios. *Climate Research*, *48*, 163-176. doi: 10.3354/cr00906

Cannon, A. J. (2008). Probabilistic multisite precipitation downscaling by an expanded bernoulli–gamma density network. *Journal of Hydrometeorology*, *9*(6), 1284–1300.

Carreau, J., & Vrac, M. (2011). Stochastic downscaling of precipitation with neural network conditional mixture models. *Water Resources Research*, *47*(10).

Casanueva, A., Bedia, J., Herrera, S., Fernández, J., & Gutiérrez, J. M. (2018). Direct and component-wise bias correction of multi-variate climate indices: the percentile adjustment function diagnostic tool. *Climatic Change*, *147*, 411–425. doi: 10.1007/s10584-018-2167-5

Chandler, R. E., & Wheater, H. S. (2002, October). Analysis of rainfall variability using generalized linear models: A case study from the west of Ireland: GENERALIZED LINEAR MODELING OF DAILY RAINFALL. *Water Resources Research*, *38*(10), 1–11. Retrieved 2019-06-24, from http://doi.wiley.com/10.1029/2001WR000906 doi: 10.1029/2001WR000906

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., ... Vitart, F. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q J R Meteorol Soc*, *137*, 553–597. doi: 10.1002/qj.828

de Groot, W. J., Goldammer, J. G., Keenan, T., Brady, M. A., Lynham, T. J., Justice, C. O., ... O'Loughlin, K. (2006). Developing a global early warning system for wildland fire. *Forest Ecology and Management*, *234*, S10.

Dowdy, A. J., Centre for Australian Weather and Climate Research, Australia, Bureau of Meteorology, & CSIRO. (2009). *Australian fire weather as represented by the McArthur Forest Fire Danger Index and the Canadian Forest Fire Weather Index.* Melbourne: Centre for Australian Weather and Climate Research,.

Déqué, M. (2011). Deterministic forecasts of continuous variables. In *Forecast verification* (p. 77-94). John Wiley and Sons, Ltd. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119960003.ch5 doi: https://doi.org/10.1002/9781119960003.ch5

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. doi: 10.5194/gmd-9-1937-2016

Fernandez-Granja, J. A., Casanueva, A., Bedia, J., & Fernández, J. (2021). Improved atmospheric circulation over europe by the new generation of cmip6 earth system models. *Climate Dynamics*, *56*, 3527–3540. doi: 10.1007/s00382-021-05652-9

Frías, M. D., Iturbide, M., Manzanas, R., Bedia, J., Fernández, J., Herrera, S., ...

Gutiérrez, J. M. (2018, January). An R package to visualize and communicate uncertainty in seasonal climate prediction. *Environmental Modelling & Software*, *99*, 101–110. Retrieved 2017-10-28, from `https://www.sciencedirect.com/science/article/pii/S1364815217305157` doi: 10.1016/j.envsoft.2017.09.008

Fugioka, F. M., Gill, A., Viegas, D. X., & Wotton, B. (2009). Fire Danger and Fire Behavior Modeling Systems in Australia, Europe, and North America. In A. Bytnerowicz, M. Arbaugh, A. Riebau, & C. Andersen (Eds.), *Developments in Environmental Science.* The Netherlands: Elsevier B.V.

Giorgi, F., Jones, C., & Asrar, G. R. (2009). Addressing climate information needs at the regional level: the CORDEX framework. *WMO Bulletin*, *58*(3), 175–183. Retrieved 2023-09-06, from `https://public.wmo.int/en/bulletin/addressing-climate-information-needs-regional-level-cordex-framework`

Gulli, A., & Pal, S. (2017). *Deep learning with keras.* Packt Publishing Ltd.

Gutiérrez, J. M., Maraun, D., Widmann, M., Huth, R., Hertig, E., Benestad, R., ... others (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), 3750–3785.

Gutiérrez, J. M., San-Martín, D., Brands, S., Manzanas, R., & Herrera, S. (2013). Reassessing statistical downscaling techniques for their robust application under climate change conditions. *Journal of Climate*, *26*(1), 171 - 188. doi: https://doi.org/10.1175/JCLI-D-11-00687.1

Herdin, M., Czink, N., Ozcelik, H., & Bonek, E. (2005). Correlation matrix distance, a meaningful measure for evaluation of non-stationary MIMO channels. In *Vehicular technology conference, 2005. vtc 2005-spring. 2005 ieee 61st* (Vol. 1, pp. 136–140).

Hertig, E., Maraun, D., Bartholy, J., Pongracz, R., Vrac, M., Mares, I., ... Soares, P. M. M. (2019). Comparison of statistical downscaling methods with respect to extreme events over Europe: Validation results from the perfect predictor experiment of the COST Action VALUE. *International Journal of Climatology*, *39*, 3846–3867. Retrieved 2018-07-30, from `https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.5469` doi: 10.1002/joc.5469

Hlinka, J., Hartman, D., Vejmelka, M., Runge, J., Marwan, N., Kurths, J., & Palus, M. (2013). Reliability of Inference of Directed Climate Networks Using Conditional Mutual Information. *Entropy*, *15*(6), 2023–2045. doi: 10.3390/e15062023

Iturbide, M., Bedia, J., Herrera, S., Baño-Medina, J., Fernández, J., Frías, M. D., ... Gutiérrez, J. M. (2019, January). The R-based climate4R open framework for reproducible climate data access and post-processing. *Environmental Modelling & Software*, *111*, 42–54. Retrieved 2023-03-24, from `https://www.sciencedirect.com/science/article/pii/S1364815218303049` doi: 10.1016/j.envsoft.2018.09.009

LeCun, Y., Bengio, Y., & Laboratories, T. B. (1995). Convolutional Networks for Images, Speech, and Time-Series. *The Handbook of Brain Theory and Neural Networks*.

Legasa, M. N., Chandler, R. E., & Manzanas, R. (2023, April). *Bayesian Network-Informed Conditional Random Forests for Probabilistic Multisite Downscaling of Precipitation Occurrence* (preprint). Preprints. Retrieved from `https://essopenarchive.org/users/603734/articles/633943-bayesian-network-informed-conditional-random-forests-for-probabilistic-multisite-downscaling-of-precipitation-occurrence?commit=66f034dbe288a09503eccfcdd6a4431b4aaba042` doi: 10.22541/essoar.168167381.15060857/v1

Lorenz, E. (1969). Atmospheric predictability as revealed by naturally occurring analogues. *Journal of the Atmospheric Sciences*, *26*, 636–&.

Maraun, D., & Widmann, M. (2018). *Statistical Downscaling and Bias Correction* *for Climate Research* (1st ed.). Cambridge University Press. Retrieved 2022-05-23, from https://www.cambridge.org/core/product/identifier/9781107588783/type/book doi: 10.1017/9781107588783

Maraun, D., Widmann, M., & Gutiérrez, J. M. (2019, July). Statistical downscaling skill under present climate conditions: A synthesis of the VALUE perfect predictor experiment. *International Journal of Climatology*, *39*(9), 3692–3703. doi: 10.1002/joc.5877

Nelder, J., & Wedderburn, R. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society Series A-General*, *135*(3), 370–&. doi: 10.2307/2344614

Quilcaille, Y., Batibeniz, F., Ribeiro, A. F. S., Padrón, R. S., & Seneviratne, S. I. (2023, May). Fire weather index data under historical and shared socioeconomic pathway projections in the 6th phase of the Coupled Model Intercomparison Project from 1850 to 2100. *Earth System Science Data*, *15*(5), 2153–2177. Retrieved 2023-11-02, from https://essd.copernicus.org/articles/15/2153/2023/ (Publisher: Copernicus GmbH) doi: 10.5194/essd-15-2153-2023

R Core Team. (2020). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/

San-Martín, D., Manzanas, R., Brands, S., Herrera, S., & Gutiérrez, J. M. (2016, October). Reassessing Model Uncertainty for Regional Projections of Precipitation with an Ensemble of Statistical Downscaling Methods. *Journal of Climate*, *30*(1), 203–223. Retrieved 2019-06-27, from https://journals.ametsoc.org/doi/full/10.1175/JCLI-D-16-0366.1 doi: 10.1175/JCLI-D-16-0366.1

San-Miguel-Ayanz, J., Moreno, J. M., & Camia, A. (2013, April). Analysis of large fires in European Mediterranean landscapes: Lessons learned and perspectives. *Forest Ecology and Management*, *294*, 11–22. Retrieved 2013-09-10, from http://linkinghub.elsevier.com/retrieve/pii/S0378112712006561 doi: 10.1016/j.foreco.2012.10.050

Turco, M., Rosa-Cánovas, J. J., Bedia, J., Jerez, S., Montávez, J. P., Llasat, M. C., & Provenzale, A. (2018, December). Exacerbated fires in Mediterranean Europe due to anthropogenic warming projected with non-stationary climate-fire models. *Nature Communications*, *9*, 3821. Retrieved 2018-09-28, from http://www.nature.com/articles/s41467-018-06358-z doi: 10.1038/s41467-018-06358-z

van Wagner, C. E. (1987). *Development and structure of the Canadian Forest Fire Weather Index* (Forestry Tech. Rep. No. 35). Ottawa, Canada: Canadian Forestry Service.

Vrac, M., & Friederichs, P. (2015). Multivariate—Intervariable, Spatial, and Temporal—Bias Correction*. *Journal of Climate*, *28*(1), 218–237. Retrieved 2015-04-28, from http://journals.ametsoc.org/doi/abs/10.1175/JCLI-D-14-00059.1

Widmann, M., Bedia, J., Gutiérrez, J., Bosshard, T., Hertig, E., Maraun, D., ... Huth, R. (2019, January). Validation of spatial variability in downscaling results from the VALUE perfect predictor experiment. *International Journal of Climatology*, *39*, 3819–3845. Retrieved 2019-02-07, from http://doi.wiley.com/10.1002/joc.6024 doi: 10.1002/joc.6024

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... others (2016). The fair guiding principles for scientific data management and stewardship. *Scientific data*, *3*(1), 1–9.

Zorita, E., & von Storch, H. (1999). The analog method as a simple statistical downscaling technique: Comparison with more complicated methods. *Journal of Climate*, *12*(8), 2474 - 2489. doi: https://doi.org/10.1175/1520-0442(1999)012⟨2474:TAMAAS⟩2.0.CO;2