Environmental and Ecological Statistics
https://doi.org/10.1007/s10651-024-00616-8

Climate model selection via conformal clustering of spatial functional data

Veronica Villani[1,3] · **Elvira Romano**[1] · **Jorge Mateu**[2]

Received: 20 October 2023 / Revised: 18 March 2024 / Accepted: 19 March 2024

Abstract
Climate model selection stands as a critical process in climate science and research. It involves choosing the most appropriate climate models to address specific research questions, simulating climate behaviour, or making projections about future climate conditions. This paper proposes a new approach, using spatial functional data analysis, to asses which of the 18 EURO CORDEX simulation models work better for predicting average temperatures in the Campania region (Italy). The method involves two key steps: first, using functional data analysis to process climate variables and select optimal models by a hierarchical clustering procedure; second, validating the chosen models by proposing a new conformal prediction approach to the anomalies associated to each cluster.

Keywords Climate model selection · Clustering · Conformal validation · Spatially dependent functional data

Handling Editor: Luiz Duczmal.

Elvira Romano and Jorge Mateu contributed equally to this work.

✉ Elvira Romano
elvira.romano@unicampania.it

Veronica Villani
v.villani@cira.it

Jorge Mateu
mateu@mat.uji.es

[1] Department of Mathematics and Physics, University of Campania Luigi Vanvitelli, 81100 Caserta, Italy

[2] Department of Mathematics, Universitat Jaume I, 12071 Castellón, Spain

[3] Department of Reliability and Safety of Systems and Infrastructure, CIRA (The Italian Aerospace Research Centre), 81043 Capua, Italy

Published online: 11 May 2024

1 Introduction

Selecting a representative climate model is a significant challenge in environmental science, as these models are crucial for understanding and forecasting interactions within the Earth's climate system. The diversity among available models, each with distinct assumptions, complexities, and spatio-temporal resolutions, complicates the identification of a universally good model.

Recent research has increasingly focused on addressing this challenge, especially concerning questions related to earth-system processes, climate change impacts, and adaptation. The options for simulating climate models typically involve the use of Global Climate Models (GCMs), Regional Climate Models (RCMs), or a combination of both. GCMs provide a global-scale perspective, while RCMs offer a more detailed view at the regional level. The combination of both models contributes to a more comprehensive understanding of climate dynamics, encompassing both global-scale patterns from GCMs and finer regional details from RCMs (Jacob et al. 2020). The simulated temporal evolution of future climate is subject to different uncertainties. In order to evaluate all the possible climate simulations according to different uncertainties, the use of the largest possible model ensemble analysing mean and standard deviation of climate models is suggested. A common approach in such studies is simply to average over all models (called the ensemble mean) with available data. This approach is justified by global scale results, generally examining only the mean climate, that show the "average model" is often the best. In the literature, different experimental approaches for assessing and selecting climate models are available. Some of them are called "past-performance" approaches, according to which climate models are selected based on their skill in representing, for the present climate, the trends of the variables of interest average and extreme values (Biemans et al. 2013; Pierce et al. 2009). Pierce et al. (2009) used an approach which consists in generating metrics of model skill to prequalify models based on their ability to simulate climate in the region or variable of interest. In particular, they evaluate whether the models selected in this way provide an estimate of climate change over the historical record that is closer to the observations than the models rejected on this basis. Some studies combined several performance measures such as root mean square errors (RMSE) (Chiew et al. 2009; Gleckler and Taylor 2008; Pitman and Perkins 2008; Winter and Nychka 2009), correlation coefficients (Murphy and Epstein 1989; Murphy 1996), and average of errors (Altinsoy and Yildirim 2015, 2016; Gleckler and Taylor 2008), to determine the climate model that is most representative of the observations. In most of these studies, researchers used a single value to represent the climate characteristics of the entire region. As a result, some significant deficiencies and errors of the models on smaller parts of the region of interest are not noticed. This is due to the averaging of the errors of the opposite signs when trying to get a simple figure. Furthermore, Bayesian/relative-likelihood approach was proposed to calibrate double counting in climate model studies in Steele and Werndl 2013, 2018. Other kinds of approaches, known as envelope approaches, aim at reducing the number

of models to be included in the ensemble allowing to represent a wide range of possible future scenarios (Houle et al. 2012; Sorg et al. 2014; Warszawski et al. 2014). Furthermore, to exclude highly dependent models, quantitative techniques are used that focus on distances and correlations between the output of different models (Masson and Knutti 2011). Among these, the most interesting ones propose a cluster analysis using different metrics and climate indices (Cannon 2015; Knutti and Sedl 2013; Sanderson et al. 2015).

The various methods outlined above exhibit certain limitations. More precisely, past-performance approaches focus solely on the models' capabilities in representing the current climate, overlooking future scenario characteristics. On the other hand, envelope approaches solely consider the convergence of climatic anomalies from individual models, neglecting their performance in the current context. Finally, quantitative techniques aimed at reducing ensemble interdependence do not incorporate measures of model performance. In this work we propose a new approach based on Spatial Functional Data Analysis (SFDA) (Delicado et al. 2010; Mateu and Romano 2017). SFDA is an extension of Functional Data Analysis (FDA) (Ramsay and Silverman 2005) that encompasses the statistical framework for the analysis of spatially correlated function-valued data. In SFDA, each sampled variable is treated as an individual realisation of an underlying spatial functional stochastic process. This approach is particularly relevant in the context of climate studies, where it is applied to analyse functions such as temperature, pressure, or humidity observed on a spatial grid, essentially representing one-dimensional functions with spatial locations. The strategy involves two main steps: model selection and validation of classified models. In the first step, time series of climate variables are represented as spatially located functions. A hierarchical clustering procedure with a spatio-functional distance, a convex combination of the spatial and functional dimensions, identifies similarities among these functions. Then spatial clusters of functional curves are compared in terms of a functional distance with all the simulated models, to build a ranking of models based on their skills in reproducing spatial functional properties of the selected climate variables.

In the second step, for model validation, future projections are used with a conformal prediction approach (Lei and Wasserman 2014), specifically introducing conformal prediction for clustering validation. This step uses anomalies associated with each cluster to construct conformal intervals, providing insights into how well selected models align with anticipated changes in the variable of interest. The chart in Fig. 1 shows the main steps of the proposed strategy.

The paper is organised as follows. Section 2 provides an overview of the data structure employed. Section 3 introduces the proposed methodology. In Sect. 4, the main results derived from the real case study are presented. The paper ends with some final remarks and conclusions.

2 Climate variables as spatial functional data

In climate analysis, the process of selecting the best models to represent a climate variable at a specific spatial point of interest often involves focusing on the average values across time in a spatial location of the ensemble. This means that the overall

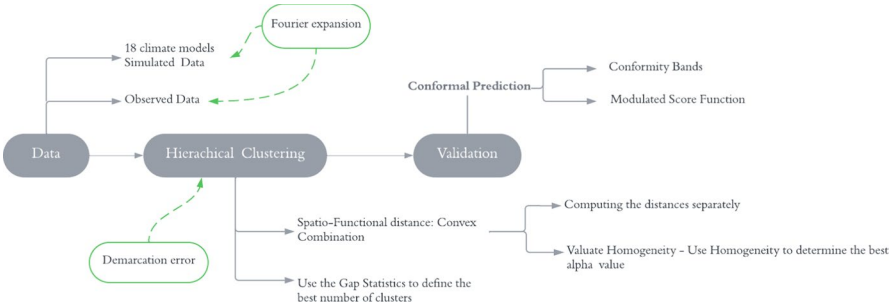

Fig. 1 The strategy step by step: model selection and validation

mean of the function is considered rather than the detailed behaviour over the time. The choice to adopt a spatial functional perspective for examining a climate variable and a simulated model overcomes this problem.

Climate variables and models can be represented as spatially located functions, to gain insights into the behaviour and performance of the model, such as identifying patterns and trends, characterising variability, and assessing model uncertainties.

Let $\left(\chi_{s_1}(t), \ldots, \ldots, \chi_{s_n}(t)\right)$ be an empirical sample at n spatial locations of a climate variable and $\left(\chi^m_{s_1}(t), \ldots, \ldots, \chi^m_{s_n}(t)\right)$ the corresponding simulated $m = 1, \ldots, M$ models. The n points $\{s_i\}_{i=1}^n$ in $D \subseteq \mathbf{R}^2$ identify the locations where the random functions χ_s and χ^m_s, spatial functional data, are located. These raw curves will be converted into functional data.

For a fixed site s_i, it is assumed that the observed data reflect a realisation of the following model:

$$\chi_{s_i}(t) = \mu_{s_i}(t) + \epsilon_{s_i}(t), \quad i = 1, \ldots, n \tag{1}$$

where $\epsilon_{s_i}(t)$ are zero-mean residuals with constant variance τ^2. For each $s_i \in D$, the function $\chi_{s_i}(t)$ is defined in $T = [a, b] \subset \mathbb{R}$ and it is assumed to belong to the Hilbert space $L_2(T)$.

With a set J of specified basis function $B_j(t)$ it is possible to estimate the true underlining function and representing it by the following linear combination

$$\chi_{s_i}(t) = \sum_{j=1}^{J} c_{ij} B_j(t) = \mathbf{c}_i^T \mathbf{B}(t), \quad i = 1, \ldots, n \tag{2}$$

where the c_{ij} are the coefficients (neither spatially correlated nor cross-correlated) estimated via a least squares approach, a weighted least squares approach, or a roughness penalty approach.

For each curve $\chi_{s_i}(t)$, the derivatives of these functions can be expressed as

$$\chi'_{s_i}(t) = \sum_{j=1}^{J} c_{ij} B'_j(t) = \mathbf{c}_i^T \mathbf{B}'(t), \quad i = 1, \ldots, n \tag{3}$$

and can be seen as spatial functional data that can reveal new insights (Ramsay and Silverman 2005).

In the context of climate variables, the use of derivatives becomes particularly valuable. Derivatives offer a detailed perspective on how climate-related functions change across geographical locations. This permits to identify key features and variations in climate patterns, which are crucial for comprehensive analysis and interpretation.

For instance, the derivatives of climate variables can provide insights into the rates of change, gradients, and functional trends across different locations. This information is essential for detecting spatial patterns such as temperature gradients, precipitation variations, or the evolution of climate-related phenomena over specific areas.

3 A two-step procedure for climate model selection

Our new approach includes the following two main steps:

- **Model selection** using a combination of hierarchical clustering based on a trimmed distance and a systematic approach for evaluating and ranking the performance of different climate models within specific clusters and their corresponding grid points.
- **Clustering validation** using a conformal prediction approach on anomalies associated to each classified model.

3.1 Model selection

3.1.1 Hierarchical clustering

Hierarchical clustering of spatially dependent functional data is an unsupervised clustering method that groups spatially located functional data with similar characteristics into clusters, based on their dissimilarities. The clustering process involves creating a tree-like hierarchy of nested clusters that can be visualised using a dendrogram. This approach is useful for exploring patterns in high-dimensional functional data and identifying subgroups with similar characteristics (Zhang and Parnell 2023).

Inspired by the approach proposed by Chavent et al. (2018), we propose a hierarchical clustering method based on a trimmed distance. Our main aim is to identify subsets of climate models that best capture the observed climate variables on a spatial domain. The clustering approach groups similar models together based on their functional similarities and allows to reduce the number of suitable models that represent the current information in both spatial and temporal components. The choice of the distance metric and linkage method can significantly impact the results of hierarchical clustering. Researchers often select these parameters based on the characteristics of the data and the specific goals of their analysis. We propose to use a

trimmed distance d as combination of a functional and spatial component between the functional derivatives. The use of derivatives in this framework aims to quantify the similarity between the rate of change of functional curves at different spatial locations.

Assuming a basis function representation for data, the convex spatio-functional distance is defined as:

$$d\Big(\chi_{s_i}(t), \chi_{s_j}(t)\Big) = \alpha d_t + (1-\alpha) d_s, \tag{4}$$

where d_t is a normalised functional distance taking into account the evolution of the trend in the temporal dimension, d_s is the normalised spatial distance between the spatial locations. The parameter $\alpha \in [0,1]$ is a mixing parameter allowing for the adjustment of the contribution of each component in the overall distance measure.

Formally we have:

$$d_t = \frac{1}{w_t}\sqrt{\int_T (\chi'_{s_i}(t) - \chi'_{s_j}(t))^2 dt}, \tag{5}$$

where $d_{t'} = \sqrt{\int_T (\chi'_{s_i}(t) - \chi'_{s_j}(t))^2 dt}$ is the distance between derivatives, $w_t = \max\{d_{t'}\}$. Using the expansion in (3) we have

$$d_t = \frac{1}{w_t}\sqrt{\int_T (\mathbf{c}_i - \mathbf{c}_j)^T \mathbf{B}'(t)\mathbf{B}'(t)^T(\mathbf{c}_i - \mathbf{c}_j) dt} \tag{6}$$

$$= \frac{1}{w_t}\sqrt{(\mathbf{c}_i - \mathbf{c}_j)^T \mathbf{G}(\mathbf{c}_i - \mathbf{c}_j)}, \tag{7}$$

where

$$G = \int_T \mathbf{B}'(t)\mathbf{B}'(t)^T dt$$

is the Gram matrix computed by an appropiately chosen numerical integration scheme tailored to the used basis system, and $\mathbf{c}_i, \mathbf{c}_j$ are the vectors of basis coefficients for the curves $\chi_{s_i}(t)$, $\chi_{s_j}(t)$. Then, the normalised spatial distance is defined as:

$$d_s = \frac{1}{w_s}\|s_i - s_j\|, \tag{8}$$

where w_s is the maximum value of the distance between all the ones obtained from the spatial coordinates, taken two-by-two. The parameter $\alpha \in [0,1]$ is chosen to look for a compromise between loss of functional and spatial homogeneity. These homogeneities can be quantified using the notion of inertia. Let W_γ be the within-cluster inertia in a cluster, derived from the distances d_t and d_s for $\gamma = t$ and $\gamma = s$,

respectively. Let P_k^α be the partition in k clusters and P_1^α the partition in one cluster. These measures are defined as follows:

$$Q_\gamma(P_k^\alpha) = 1 - \frac{W_\gamma(P_k^\alpha)}{W_\gamma(P_1^\alpha)} \in [0, 1], \tag{9}$$

For $\gamma = s$, $Q_s(P_k^\alpha) = 1 - \frac{W_s(P_k^\alpha)}{W_s(P_1^\alpha)}$ defines the ratio between the within inertia in the k clusters based on the distance d_s and the total inertia. In other words Q_s represents the level of homogeneity of the P_k^α partition from a spatial point of view. We can easily understand that a higher value of this criteria implies greater homogeneity.

Similarly for $\gamma = t$, $Q_t(P_k^\alpha) = 1 - \frac{W_t(P_k^\alpha)}{W_t(P_1^\alpha)}$ defines the ratio between the within inertia in the k clusters based on the distance d_t and the total inertia; thus Q_t represents the level of homogeneity of the P_k^α partition from a functional point of view.

The parameter α is formally chosen by calculating separately $Q_t(\alpha)$ and $Q_s(\alpha)$ of the partitions obtained for a range of different values of α and several number of clusters k. A criteria based on the cross point between the curves $Q_t(\alpha)$ and $Q_s(\alpha)$ is then used to choose a value of α that is a compromise between the loss in functional homogeneity and the gain in spatial cohesion.

Once the value of α is chosen, we use the 'Gap statistic' method for estimating the number of clusters obtained by the trimmed distance. Gap statistic (Tibshiran et al. 2001) works by comparing the performance of a clustering algorithm with different values of k to a reference distribution, typically generated by a randomisation process. The idea is to measure how much better the clustering results are than what would be expected by random chance. The value that results in the largest gap statistic is considered the optimal number of clusters. This is the number of clusters that best captures the underlying structure in the data avoiding overfitting.

3.1.2 Cluster-based ranking of climate models

The clustered data are the reference against which the simulated climate models are systematically compared. This means that first the spatial locations of each cluster are identified, then for each of them the simulated variables are converted into spatial functional data for all the possible simulated models. This allows to draw parallels between the simulated data (models realisations) and the observed data. At this point a "demarcation score/error" function is defined, for each grid point, as the criteria to select the best set of models for each cluster. This function is obtained by "skill scores" quantifying how well the models represent the data.

Let C be a cluster with n_c grid points, and M the number of possible models. The "Integrate Root Mean Square Error" $\mathcal{I}$RMSE for each grid point and model m is given by:

$$\mathcal{I}\text{RMSE}_{s_i,m} = \sqrt{\int_T (\chi_{s_i}^{m'} - \chi_{s_i}')^2 \, dt}, \quad m = 1, \ldots, M \tag{10}$$

where $\chi_{s_i}^{m'}$ and $\chi_{s_i}^{'}$ are respectively the functional derivative of the simulated variable for model m and the observed ones at grid point s_i. The "demarcation line" between the best and the worst representative models in the clusters is obtained by the mean of $\mathcal{I}$RMSE for each spatial location s_i. It is then computed by:

$$\overline{\mathcal{I}\text{RMSE}_{s_i}} = \frac{\sum_{m=1}^{M} \mathcal{I}\text{RMSE}_{s_i,m}}{M}, \quad s_i = 1, \ldots, n_c \tag{11}$$

Thus, we obtain a function of the grid points in the clusters. The criteria to select the best model for each cluster is the following: if the error of a simulated climate variable at a given grid point is smaller than the demarcation line, at the same grid point, this model is considered successful and assigned the value of 1, otherwise 0 is assigned. Formally let $\delta_{s_i,m}$ be "a skill score function" defined for each grid point s_i and model m as:

$$\delta_{s_i,m} = \begin{cases} 1 \text{ if } IRMSE_{s_i,m} < \overline{IRMSE}_{s_i} \\ 0 \text{ otherwise} \end{cases}$$

For each model m, we compute the percentage of successful grid points within the cluster by:

$$\text{P}_m = \frac{\sum_{s_i=1}^{n_c} \delta_{s_i,m}}{n_c} \times 100.$$

The most representative models $\hat{m}$ for the spatio-functional cluster are the models with the maximum percentage P_m:

$$\hat{m} = \arg\max_m \left(\frac{\sum_{s_i=1}^{n_c} \delta_{s_i,m}}{n_c} \times 100 \right).$$

3.2 Clustering validation

Anomalies in climate change, often referred to as temperature anomalies, are a way to assess and communicate changes in temperature over time. These anomalies represent deviations from a long-term average temperature, typically based on a reference period. In this second step our main aim is to fit the anomalies on the clustering training samples, and then use the residuals on a held-out validation set to quantify the uncertainty in future predictions. Essentially, the selected climate models for each cluster are validated and further evaluated by assessing their performance on possible future scenarios using conformal prediction regions. We thus propose to combine the reliability and validity measures from conformal prediction with the quality of the clustering. Conformal clustering is based on the conformal prediction technique, thus, in this section, we will first

introduce conformal prediction, and then proceed to describe our idea starting from some previously proposed conformal clustering approach.

3.2.1 Conformal prediction

Conformal Prediction (CP) is a machine learning framework used for making predictions while providing a measure of the confidence or reliability of those predictions. It is used to estimate the probability that a prediction will be correct (Vovk and Glenn 2008). The key components of CP are training set, a nonconformity measure, significance level and finally the prediction interval. The algorithm starts with a labeled training dataset, which consists of input features and corresponding target values. This quantifies how unusual or different the test instance is compared to the training data. Then a rank of the nonconformity scores for all test instances in ascending order is provided. In this context, the nonconformity measure plays a fundamental role. It is a function that quantifies how much a specific prediction deviates from the other examples in the training set. Formally, it calculates the dissimilarity or nonconformity of a new input instance with respect to the training set. Thus for a given input instance and a significance level, that represents the probability that a prediction interval will contain the true target value, a prediction interval is defined with probability target values at least $(1 - \eta)$. CP provides the following main two formal guarantees (Fontana et al. 2023):

- Validity: The prediction intervals have a predefined coverage probability $(1 - \eta)$, meaning that they contain the true target value with at least this probability.
- Conservativeness: The prediction intervals are guaranteed to be valid even if the underlying model makes no assumptions about the data distribution.

The method's simplicity and versatility has enabled its extension to the analysis of functional data (Diquigiovanni et al. 2022) and spatial dependent functional data (Diana et al. 2023).

3.2.2 Clustering validation via conformal prediction

Conformal prediction, traditionally employed in several supervised learning tasks such as classification and regression, has recently seen attempts to extend its application to the unsupervised learning task of clustering (Cherubin et al. 2015; Nouretdinov et al. 2019). In Cherubin et al. (2015) a clustering method solely relying on conformal prediction, similar to hierarchical clustering, is proposed. This method allows for the control of the number of instances that remain unclustered by specifying a desired confidence level. In Nouretdinov et al. (2019) a conformal prediction with traditional clustering approaches, such as k-means or density-based clustering, is introduced with the aim of overcoming various clustering challenges, including model parameter tuning, cluster merging, and accommodating clusters of diverse shapes and sizes. We propose to generalise this last approach to the spatial functional framework.

The monthly climatic anomalies, obtained as the difference of the monthly temperature values relating to a future period, in accordance with the defined scenario, compared to the reference period, are converted into a spatial functional form. To account for future changes in the variable of interest, we compute the conformal interval of these functional anomalies for each climate model representative of the k clusters.

So, given a nominal miscoverage level $\eta \in (0,\ 1)$, we define a conformal band $C^{\eta} \subset \mathcal{L}_2(T)$ for the set of anomalies curves $(\chi^{\alpha}_{s_1}(t), , \ldots , \chi^{a}_{s_{nc}}(t))$ in the clusters.

Originally, the idea of CP is trying all possible curves for the test object to see how well these curves conform to a set of training samples. We construct a sequential prediction of possible anomalies on a spatial grid which conform to the mean anomalies in the clusters and thus conforms the model. Our approach includes the following main steps:

- Let $\{\chi^{a}_{s_j}(t), j = 1, \ldots , n\}$ be a set of anomalies of each site s_j. We split randomly into a training and a detection set.
- Consider the same partition of the spatial grid point obtained by the clustering procedure for the training sample and fix the center of the cluster as reference point.
- Define the augumented data set with a new anomaly $\chi^{a}_{s_{j+1}}$ and compute the nonconformity function D:

$$\mathcal{D}(\chi^{a}_{s_j}(t)) = \left\| \bar{\chi}_{s_i}(t) - \chi^{a}_{s_j}(t) \right\| \; j = 1, \ldots , n_c \tag{12}$$

where $\bar{\chi}_{s_i}(t)$ is the mean of the anomalies in cluster C. This function quantifies the deviation or unusualness of the anomaly $\chi^{a}_{s_j}(t)$ relative to the previous instances $\chi_{s_j}(t) j = 1, \ldots , n_c$ within the clusters.

- Repeat the above steps for each anomaly, define the distribution $\pi\left(\chi^{a}_{s_j}\right)$ of D, called distribution scores, and set $\hat{C}^{(\eta)} = \{\chi^{a}_{s_j} | \pi(\chi^{a}_{s_j}) \geq \eta\}$
- group the elements of $\hat{C}^{(\eta)}$ such that $\chi^{a}_{s_j}$ and $\chi^{a}_{s'_j}$ are neighbours to the same average model.

Finally, prediction intervals are computed using a significance level along with a set of nonconformity scores. The intervals are calculated roughly by taking the $\eta - th$ percentile of the distribution scores

$$C^{(\eta)} = \left\{ \chi^{a}_{s_j} \in \mathcal{L}_2(T) : \; \chi^{a}_{s_j} \in \left[\hat{\bar{\chi}}_{s_i}(t) - r^{\pi} S(t),\ \hat{\bar{\chi}}_{s_i}(t) + r^{\pi} S(t) \right] \right\} \tag{13}$$

where $\bar{\chi}_{s_i}(t)$is the mean of the anomalies in the clusters and the centre of the prediction band, r^{π} is the ray of the prediction, and $S(t)$ that is the functional standard deviation of all the anomalies in the clusters, is a modulation function. In particular r^{π} is the value of $(1 - \eta)$ quantile of the distribution values $\{P_i : \; i = n_c + 1, \ldots , n\}$, where $P_i = \mathcal{D}(\chi^{a}_{s_j}(t)) = \left\| \bar{\chi}_{s_i}(t) - \chi^{a}_{s_j}(t) \right\|$. The conformal prediction region provides a

measure of the thrust worthiness of each model in predicting future outcomes, which can help guiding a decision-making in the face of climate change.

4 Data and results

4.1 Climatic data

The observed data belong to the E-OBS dataset (Haylock et al. 2008). This dataset is employed for the analysis of the climatic period 1971–2005, providing monthly precipitation and monthly mean temperature data on a regular grid with a horizontal resolution of approximately 12 km. Specifically, we used the E-OBS 25.0e version (Cornes et al. 2018), released in April 2022. We focused on monthly mean temperature data for the period 1971–2005. The dataset encompasses a grid of time points covering the entire 35-year period, consisting of 420 grid points. Climate analysis is performed by CORDEX regional climate model (RCM) simulations available over the European domain (EURO-CORDEX) at a resolution 0.11*degree* (EUR-11, about 12.5 km) forced by different global climate models (GCM) (Jacob et al. 2020; Kotlarski et al. 2014). The climate simulations comprise 18 GCM-RCM combinations conducted within the EURO-CORDEX framework, considering both the historical experiment and the IPCC RCP8.5 scenarios (Moss et al. 2010). The eighteen EURO-CORDEX models listed in Table 1 were used (where r1i1p1, r3i1p1, and r12i1p1 represent ensemble members in the driving global model calculation). For

Table 1 List of the EURO-CORDEX simulations considered

Model ID	Driving GCM	GCM member	RCM name
cm5_cclm4	CNRM-CM5	r1i1p1	CCLM4
cm5_aladin53	CNRM-CM5	r1i1p1	ALADIN53
cm5_rca4	CNRM-CM5	r1i1p1	RCA4
earthr12_cclm4	EC-EARTH	r12i1p1	CCLM4
earthr12_racmo22e	EC-EARTH	r12i1p1	RACMO22E
earthr12_rca4	EC-EARTH	r12i1p1	RCA4
earthr1_racmo22e	EC-EARTH	r1i1p1	RACMO22E
earthr3_hirham5	EC-EARTH	r3i1p1	HIRHAM5
esr1_cclm4	HadGEM2-ES	r1i1p1	CCLM4
esr1_rca4	HadGEM2-ES	r1i1p1	RCA4
esr1_racmo22e	HadGEM2-ES	r1i1p1	RACMO22E
lrr1_cclm4	MPI-ESM-LR	r1i1p1	CCLM4
lrr1_remo2009	MPI-ESM-LR	r1i1p1	REMO2009
lrr1_rca4	MPI-ESM-LR	r1i1p1	RCA4
lrr2_remo2009	MPI-ESM-LR	r2i1p1	REMO2009
m_hirham5	NorESM1-M	r1i1p1	HIRHAM5
mr_rca4	IPSL-CM5A-MR	r1i1p1	RCA4
mr_wrf331f	IPSL-CM5A-MR	r1i1p1	WRF331F

clarity, it is noted that 'r' stands for realisation (the starting point of the calculation), 'i' for initialisation method, and 'p' for physics version. We specifically used the monthly mean temperature data simulated by these eighteen EURO-CORDEX models.

For both the E-OBS observation dataset and for each EURO-CORDEX model, the monthly mean temperature data relating to the grid points covering the Campania region, which is the area of interest for this work, are considered. This means that for each spatial location, identified by the latitude-longitude coordinates, we have not only the known data that belong to the E-OBS dataset, but also different simulated data, consequence of the fact that each of the eighteen models can be evaluated in that specific location.

4.2 Results

The two-step strategy based on SFDA presented earlier has been used to select the most representative climate models for predicting temperature in the area of interest. In the first step of the analysis, we cluster the monthly temperature data from E-OBS v25 using the proposed trimmed distance. Subsequently, we assess the performance of different climate models within specific clusters and their associated grid points using a ranking procedure.

This step includes the construction of spatio-functional data by means of smoothing where we select 100 Fourier basis functions based on a cross-validation criteria. Subsequently, a preprocessing step is performed to obtain the optimal value of the trimmed parameter α.

For a given number of $k = 9$ clusters, obtained by a clustering algorithm based on the distance d_t, we begin with a predefined grid G for α values within the range [0, 1].

For each $\alpha_j \in G$, we apply a hierarchical clustering algorithm based on a Ward criteria to create a partition consisting of $k = 9$ clusters. We then evaluate the quality of each partition $P_{k_j}^{\alpha}$ using the criteria on $Q_s(P_j^{\alpha})$ defined in (9) and observing visually how much the spatial partition deviates from the spatial partition P_1. In the same way we work on the temporal dimension. The visual representation of the relationship between α_j and $Q_s(P_j^{\alpha})$ and between $\alpha_j$and $Q_t(P_j^{\alpha})$ (Fig. 2) allows to select the appropriate value for α from the grid G. The choice of alpha is like adjusting a balance between functional features and spatial closeness. It is computed separately calculating functional homogeneity and geographic homogeneity for partitions obtained across different α-values and a fixed number of clusters k obtained by the functional classification. Figure 2 shows the proportion of explained inertia calculated with d_t (the functional distances) is equal to 1 when $\alpha = 0$ and decreases when alpha increases (black line). On the contrary, the proportion of explained inertia calculated with d_s (the spatial distances) is equal to 1 when $\alpha = 1$ and decreases when alpha decreases (red line). By comparing the curves of $Q_t(P_j^{\alpha})$ and $Q_s(P_j^{\alpha})$, we determine that $\alpha = 0.6$ strikes a balance between the loss in functional homogeneity and the gain in spatial cohesion. In particular we see that the proportion of explained inertia calculated with d_t is equal to 0.80 when $\alpha = 0$. On the contrary the

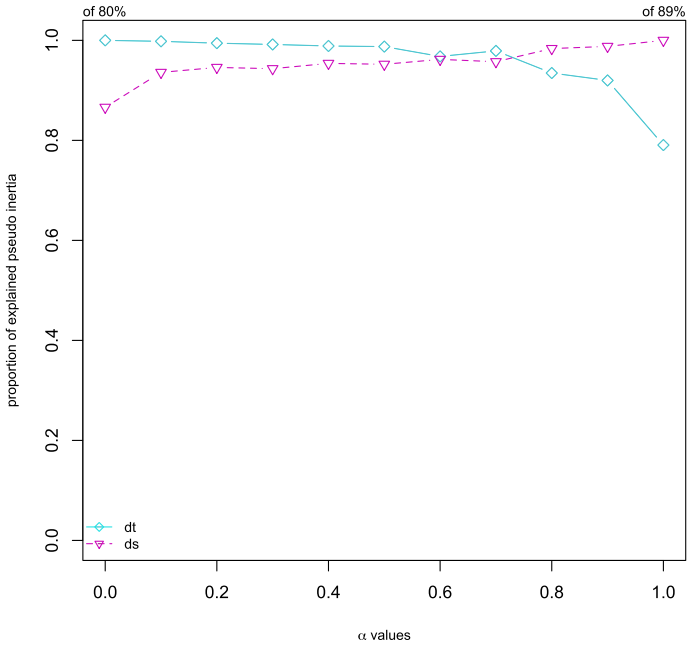

Fig. 2 Comparison plot: the variations in explained inertia curves, $Q_t(P_j^\alpha)$ and $Q_s(P_j^\alpha)$, across different values of $\alpha \in [0, 1]$

proportion of explained inertia calculated with d_s (the geographical distances) is equal to 0.87 when $\alpha = 1$.

The final step involves determining the optimal number of clusters. This task is accomplished by applying the gap statistic criteria applied on a hierarchical clustering based on the trimmed distance. In this specific case, the optimal number of clusters is found to be $k = 9$, as can be seen in Fig. 3. The choice of $k = 9$ is based on a balance between the gap statistic and the stability observed in the clustering results. After 10 it is possible to observe a constant trend.

The visual representation of this partition is illustrated in Fig. 4, where each cluster is distinguished by a unique color. The cohesion of the clusters is balanced by the choice of the $\alpha = 0.6$ value. This suggests that the selected parameter has been effective in creating meaningful and balanced clusters, as evidenced by the visual representation in the figure.

After classifying the observed functional dataset into 9 spatio-functional clusters, the next step involves investigating the monthly temperature data associated with each of the 18 EURO-CORDEX regional climate models for the period spanning 1971–2005.

For each model, the grid points of the observation dataset are identified and classified on one of the 9 detected clusters. To establish a demarcation line for determining the most representative models within each cluster, we compute the average Euclidean distance (error) between the simulated and observed values for

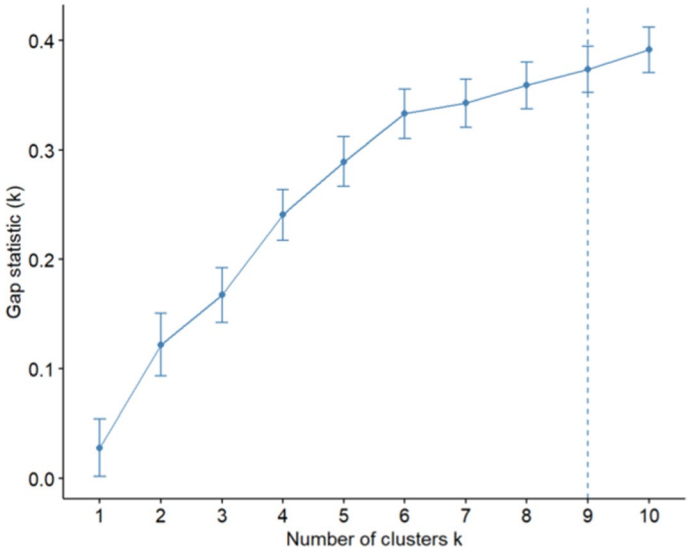

Fig. 3 Gap statistics computed for different values of k using the trimmed distance based on the optimal value of $\alpha = 0.6$

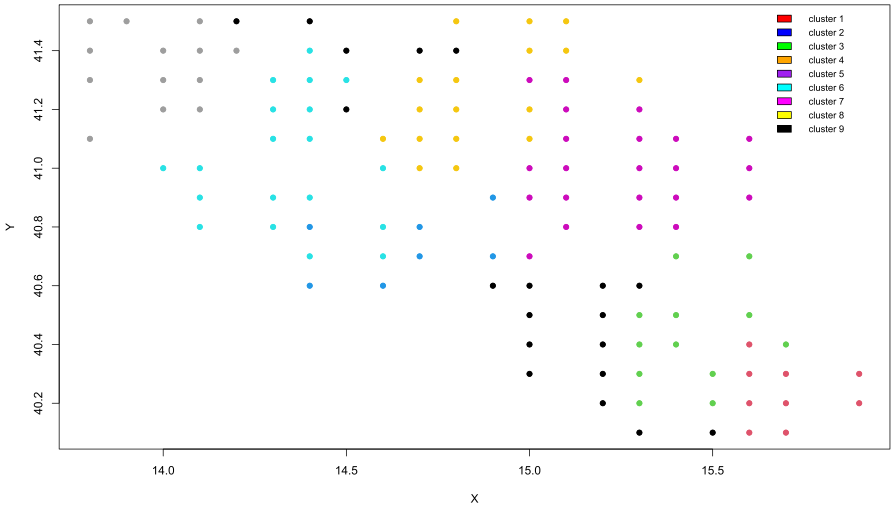

Fig. 4 Final clusters organized by latitude-longitude coordinates and color-coded based on their respective cluster assignments for visual grouping

all 18 climate models at each grid point within the cluster. Within each cluster, the climate models with the highest count of grid points having an error less than the defined demarcation error are chosen as the most representative models for that particular cluster.

Table 2 Summary of clusters: number of grid points and selected climate models

Cluster ID	N. grid points	N. selected climate models
1	6	6
2	18	4
3	13	3
4	10	9
5	8	1
6	15	6
7	15	2
8	10	8
9	23	10

Table 2 shows the composition of each of the 9 clusters obtained, that is, the number of grid points (on 118 N. grid points) that fall into each cluster and the number of climate models that are representative for each cluster. The tables containing the complete names of the climate models found to be representative for each cluster are shown in Appendix A.

Moreover, the temperature profiles of the clusters are illustrated in Fig. 5. These graphs show the temperature behaviour at each grid point within the cluster, presenting data for both the observed dataset and the chosen climate models for clusters 5 and 7. It is evident from these visualisations that the spatio-temporal patterns in the climate models selected for clusters 5 and 7 closely mirror the trends observed in the dataset. Hence, it can be deduced that the chosen climate models faithfully capture the temperature patterns observed in the dataset.

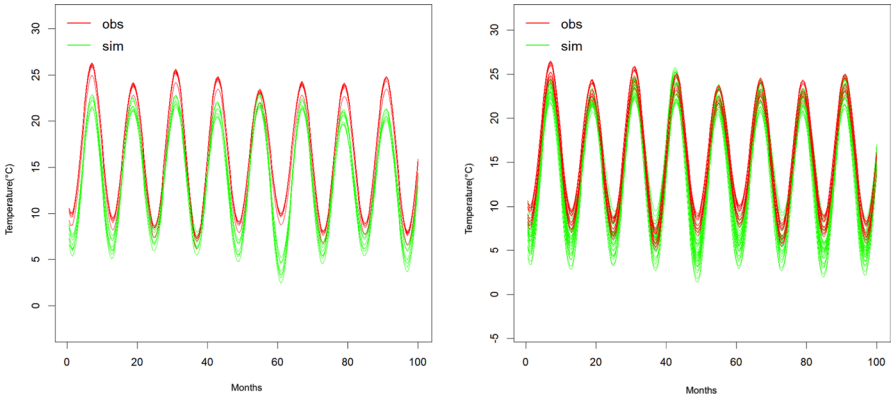

Fig. 5 Temperature curves for clusters 5 (on the left) and 7 (on the right): observed dataset (in red) and selected climate models (in green)

4.3 Model evaluation

In terms of model evaluation, and for each cluster within the Campania region, the selection process focuses on identifying and choosing regional climate models that accurately represent the climatic conditions and variations of this specific area. This selection is related both to historical climate data to capture past trends and future climate projections to anticipate potential changes. For each climate model representing the clusters, monthly climatic anomalies are calculated. These anomalies represent the difference between monthly temperature values for the future period (2036–2065) under the RCP8.5 scenario and a reference period (1981–2010). These calculated anomalies are then converted into a functional form by using (3). Following this, conformal intervals for the derived functional anomalies are determined for each cluster. These intervals help identifying models that do not conform to the average behaviour observed in all models. Models that fall outside the conformal region are excluded from further consideration. This process helps refine the selection of climate models for more in-depth analysis or decision-making.

To illustrate this process, Fig. 6 displays the evolution of the derived functional anomalies of the mean temperature of the EURO-CORDEX models selected in the first step (black lines) for 1 year, in the period 2036–2065 (RCP8.5) compared to the period 1981–2010 for cluster 5 and 7 respectively. The blue line represents the mean value of the derived functional anomalies of all models, and the red lines represent the extremes of conformal interval for $\eta = 95\%$. Since the derived functional anomalies fall within the conformal interval, these models are representative of the average behaviour across all models. This indicates that their characteristics and performance closely align with the collective behaviour observed in the entire set of models in the clusters.

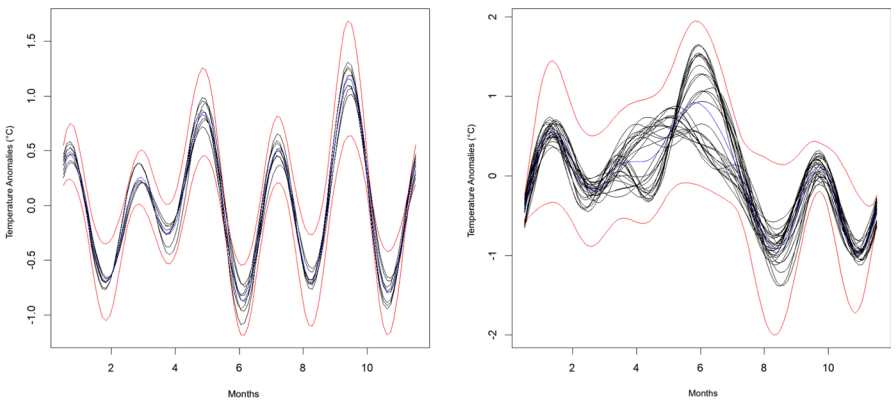

Fig. 6 Conformal intervals obtained considering the functional derived anomalies, in one year, from the climate models that are representative for clusters 5 and 7

5 Conclusions

Climate models are complex structures that can predict, with a certain level of accuracy, the variations of climate variables across the Earth's atmosphere. Indeed, models can try to predict how climate change can affect parts of the natural world through the study of some key variables.

In this work, the monthly mean temperature was calculated with both the monthly series of the observed dataset available over the Campania region and with the data of 18 EURO-CORDEX models. The mutual dissimilarities between the models and the time series related to the observations were evaluated in space and time by implementing conformal clustering method for spatially dependent functional data. One of the important advantages of the proposed approach is that it gives a chance to look deeply and in detail to specific parts of the area of interest where their models produce less than perfect results compared to other climate models in the literature.

The proposed strategy offers several advantages over other model selection approaches. As a matter of fact, by using a hierarchical clustering method based on a trimmed distance, we can better account for the spatial dependencies and variability of climate variables, which are often highly correlated and exhibit complex spatial patterns. The use of skill scores and ranking criteria to select the most representative models, for each cluster, further improves the accuracy and reliability of the selected models. Moreover, the use of conformal prediction regions allows us to quantify the uncertainty associated with future model predictions, which is critical to make informed decisions in the face of climate change. By providing a measure of the reliability and accuracy of each selected model, the conformal prediction regions can help decision-makers to better understand the range of possible outcomes.

Appendix A: List of selected EURO-CORDEX climate models for each cluster

See Tables 3, 4, 5, 6, 7, 8, 9, 10 and 11.

Table 3 List of selected EURO-CORDEX climate models for cluster 1

Model ID	Driving GCM	GCM member	RCM name
cm5_rca4	CNRM-CM5	r1i1p1	RCA4
earthr12_cclm4	EC-EARTH	r12i1p1	CCLM4
earthr12_racmo22e	EC-EARTH	r12i1p1	RACMO22E
lrr1_cclm4	MPI-ESM-LR	r1i1p1	CCLM4
lrr1_remo2009	MPI-ESM-LR	r1i1p1	REMO2009
lrr1_rca4	MPI-ESM-LR	r1i1p1	RCA4

Table 4 List of selected EURO-CORDEX climate models for cluster 2

Model ID	Driving GCM	GCM member	RCM name
cm5_cclm4	CNRM-CM5	r1i1p1	CCLM4
cm5_rca4	CNRM-CM5	r1i1p1	RCA4
lrr1_cclm4	MPI-ESM-LR	r1i1p1	CCLM4
lrr1_rca4	MPI-ESM-LR	r1i1p1	RCA4

Table 5 List of selected EURO-CORDEX climate models for cluster 3

Model ID	Driving GCM	GCM member	RCM name
cm5_rca4	CNRM-CM5	r1i1p1	RCA4
lrr1_remo2009	MPI-ESM-LR	r1i1p1	REMO2009
lrr1_rca4	MPI-ESM-LR	r1i1p1	RCA4

Table 6 List of selected EURO-CORDEX climate models for cluster 4

Model ID	Driving GCM	GCM member	RCM name
cm5_cclm4	CNRM-CM5	r1i1p1	CCLM4
cm5_rca4	CNRM-CM5	r1i1p1	RCA4
earthr12_cclm4	EC-EARTH	r12i1p1	CCLM4
earthr12_racmo22e	EC-EARTH	r12i1p1	RACMO22E
earthr12_rca4	EC-EARTH	r12i1p1	RCA4
earthr1_racmo22e	EC-EARTH	r1i1p1	RACMO22E
lrr1_cclm4	MPI-ESM-LR	r1i1p1	CCLM4
lrr1_remo2009	MPI-ESM-LR	r1i1p1	REMO2009
mr_wrf331f	IPSL-CM5A-MR	r1i1p1	WRF331F

Table 7 List of selected EURO-CORDEX climate models for cluster 5

Model ID	Driving GCM	GCM member	RCM name
cm5_rca4	CNRM-CM5	r1i1p1	RCA4

Table 8 List of selected EURO-CORDEX climate models for cluster 6

Model ID	Driving GCM	GCM member	RCM name
cm5_rca4	CNRM-CM5	r1i1p1	RCA4
earthr12_cclm4	EC-EARTH	r12i1p1	CCLM4
lrr1_cclm4	MPI-ESM-LR	r1i1p1	CCLM4
lrr1_remo2009	MPI-ESM-LR	r1i1p1	REMO2009
lrr1_rca4	MPI-ESM-LR	r1i1p1	RCA4
mr_wrf331f	IPSL-CM5A-MR	r1i1p1	WRF331F

Table 9 List of selected EURO-CORDEX climate models for cluster 7

Model ID	Driving GCM	GCM member	RCM name
lrr1_remo2009	MPI-ESM-LR	r1i1p1	REMO2009
lrr1_rca4	MPI-ESM-LR	r1i1p1	RCA4

Table 10 List of selected EURO-CORDEX climate models for cluster 8

Model ID	Driving GCM	GCM member	RCM name
cm5_rca4	CNRM-CM5	r1i1p1	RCA4
earthr12_cclm4	EC-EARTH	r12i1p1	CCLM4
earthr12_rca4	EC-EARTH	r12i1p1	RCA4
earthr1_racmo22e	EC-EARTH	r1i1p1	RACMO22E
earthr3_hirham5	EC-EARTH	r3i1p1	HIRHAM5
lrr1_cclm4	MPI-ESM-LR	r1i1p1	CCLM4
lrr1_remo2009	MPI-ESM-LR	r1i1p1	REMO2009
lrr1_rca4	MPI-ESM-LR	r1i1p1	RCA4

Table 11 List of selected EURO-CORDEX climate models for cluster 9

Model ID	Driving GCM	GCM member	RCM name
cm5_cclm4	CNRM-CM5	r1i1p1	CCLM4
cm5_rca4	CNRM-CM5	r1i1p1	RCA4
earthr12_cclm4	EC-EARTH	r12i1p1	CCLM4
earthr12_racmo22e	EC-EARTH	r12i1p1	RACMO22E
earthr12_rca4	EC-EARTH	r12i1p1	RCA4
earthr1_racmo22e	EC-EARTH	r1i1p1	RACMO22E
lrr1_cclm4	MPI-ESM-LR	r1i1p1	CCLM4
lrr1_remo2009	MPI-ESM-LR	r1i1p1	REMO2009
lrr1_rca4	MPI-ESM-LR	r1i1p1	RCA4
mr_wrf331f	IPSL-CM5A-MR	r1i1p1	WRF331F

Author Contributions All authors reviewed the results and approved the final version of the manuscript

Funding Open access funding provided by Università degli Studi della Campania Luigi Vanvitelli within the CRUI-CARE Agreement.

Declarations

Conflict of interest The authors declare no conflict of interest

References

Altinsoy H, Yildirim HA (2015) Labor productivity losses over western Turkey in the twenty-first century as a result of alteration in WBGT. Int J Biometeorol 59(4):463–471

Altinsoy H, Yildirim HA (2016) Wet bulb globe temperature across Western Turkey according to the ENSEMBLES project. Int J Glob Warm 9(1):66–80

Biemans H, Speelman LH, Ludwig F, Moors EJ, Wiltshire AJ, Kumar P, Gerten D, Kabat P (2013) Selecting global climate models for regional climate change studies. Proc Natl Acad Sci USA 106(21):8441–8446. https://doi.org/10.1073/pnas.0900094106

Cannon AJ (2015) Selecting GCM scenarios that span the range of changes in a multimodel ensemble: application to CMIP5 climate extremes indices. J Clim 28:1260–1267

Chavent M, Kuentz-Simonet V, Labenne A, Saracco J (2018) ClustGeo: an R package for hierarchical clustering with spatial constraints. Comput Stat 33:1799–1822. https://doi.org/10.1007/s00180-018-0791-1

Cherubin G, Nouretdinov J, Gammerman A, Jordaney R, Wang Z, Papini D, Cavallaro L (2015) Conformal clustering and its application to botnet traffic. In: International symposium on statistical learning and data sciences. Springer, pp 313–322

Chiew FHS, Teng J, Vaze J, Kirono DGC (2009) Influence of global climate model selection on runoff impact assessment. J Hydrol 379:172–180

Cornes R, Van der Schrier G, Van den Besselaar EJM, Jones PD (2018) An ensemble version of the E-OBS temperature and precipitation datasets. J Geophys Res Atmos 123:9391–9409

Delicado P, Giraldo R, Comas C, Mateu J (2010) Statistics for spatial functional data: some recent contributions. Environmetric 21:224–239

Diana A, Romano E, Irpino A (2023) Distribution free prediction for geographically weighted functional regression models. Spatial Stat 57:100765. https://doi.org/10.1016/j.spasta.2023.100765. (**ISSN 2211-6753**)

Diquigiovanni J, Fontana M, Vantini S (2022) Conformal prediction bands for multivariate functional data. J Multivar Data Anal 189(C):104879

Fontana M, Zeni G, Vantini S (2023) Conformal prediction: a unified review of theory and new challenges. Bernoulli. https://doi.org/10.3150/21-BEJ1447

Gleckler PJ, Taylor KE, Doutriaux C (2008) Performance metrics for climate models. J Geophys Res 113(D6)

Haylock M, Hofstra N, Klein Tank A, Klok E, Jones P, New M (2008) A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. J Geophys Res. https://doi.org/10.1029/2008JD010201

Hennemuth TI, Jacob D, Keup-Thiel E, Kotlarski S, Nikulin GO et al. (2017) Guidance for EUROCORDEX climate projections data use. Version1.0-2017.08. https://www.euro-cordex.net/imperia/md/content/csc/cordex/euro-cordex-guide lines-version1.0-2017.08.pdf

Houle D, Bouffard A, Duchesne L, Logan T, Harvey R (2012) Projections of future soil temperature and water content for three Southern Quebec forested sites. J Clim 25(21):7690–7701. https://doi.org/10.1175/JCLI-D-11-00440.1

Jacob D, Teichmann C, Sobolowski S et al (2020) Regional climate downscaling over Europe: perspectives from the EURO-CORDEX community. Reg Environ Change 20(2):20–51

Knutti R, Sedl J (2013) Robustness and uncertainties in the new CMIP5 climate model projections. Nat Clim Change 3:369–373. https://doi.org/10.1038/nclimate1716

Kotlarski S, Keuler K, Christensen OB et al (2014) Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble. Geosci Model Dev 7:1297–1333. https://doi.org/10.5194/gmd-7-1297-2014

Lei J, Wasserman L (2014) Distribution-free prediction bands for non-parametric regression. J R Stat Soc Ser B (Stat Methodol) 76(1):71–96

Masson D, Knutti R (2011) Climate model genealogy. Geophys Res Lett 38:L08703. https://doi.org/10.1029/2011GL046864

Mateu J, Giraldo R (2021) Geostatistical functional data analysis. Wiley, New York

Mateu J, Romano E (2017) Advances in spatial functional statistics. Stoch Environ Res Risk Assess 31:1–6

Moss RH et al (2010) The next generation of scenarios for climate change research and assessment. Nature 463:747–756

Murphy AH (1996) General decompositions of MSE-based skill scores: measures of some basic aspects of forecast quality. Mon Weather Rev 124:2353–2369

Murphy AM, Epstein ES (1989) Skill scores and correlation coefficients in model verification. Mon Weather Rev 117(3):572–581

Nouretdinov I, Gammerman J, Fontana M, Rehal D (2019) Multilevel conformal clustering: a distribution-free technique for clustering and anomaly detection. Neurocomputing 397:279–291

Pierce DW, Barnett TP, Santer BD, Gleckler PJ (2009) Selecting global climate models for regional climate change studies. Proc Natl Acad Sci USA 106(21):8441–8446. https://doi.org/10.1073/pnas.0900094106

Pitman AJ, Perkins SE (2008) Regional projections of future seasonal and annual changes in rainfall and temperature over Australia based on skill-selected AR4 models. Earth Interact 12:1–50

Ramsay J, Silverman B (2005) Functional data analysis. Springer, New York

Sanderson B, Knutti R, Caldwell P (2015) A representative democracy to reduce interdependency in a multimodel ensemble. J Clim 28:5171–5194. https://doi.org/10.1175/JCLI-D-14-00362.1

Sorg A, Huss M, Rohrer M, Stoffel M (2014) The days of plenty might soon be over in glacierized Central Asian catchments. Environ Res Lett. https://doi.org/10.1088/1748-9326/9/10/104018

Steele K, Werndl C (2013) Climate models, calibration, and confirmation. Br J Philos Sci 64(3):609–635

Steele K, Werndl C (2018) Model-selection theory: the need for a more nuanced picture of use-novelty and double-counting. Br J Philos Sci 69:351–375

Tibshirani R et al (2001) Estimating the number of clusters in a data set via the gap statistic. J R Stat Soc Ser B (Stat Methodol) 63(2):411–423

Vovk V, Glenn S et al (2008) A tutorial on conformal prediction. J Mach Learn Res 9:371–421

Warszawski L, Frieler K, Huber V, Piontek F, Serdeczny O, Schewe J (2014) The inter-sectoral impact model intercomparison project (ISI-MIP): project framework. Proc Natl Acad Sci USA 111(9):3228–3232. https://doi.org/10.1073/pnas.1312330110

Winter CL, Nychka D (2009) Forecasting skill of model averages. Stoch Environ Res Risk A 24(5):633–638. https://doi.org/10.1007/s00477-009-0350-y

Zhang M, Parnell A (2023) Review of clustering methods for functional data. ACM Trans Knowl Discov Data 17(7):91. https://doi.org/10.1145/3581789