Ecological Informatics 82 (2024) 102674

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Mapping local climates in highly heterogeneous mountain regions: Interpolation of meteorological station data vs. downscaling of macroclimate grids

Daniele Delle Monache [a,*], Giuseppe Martino [a], Andrea Chiocchio [a], Antonino Siclari [b,c], Roberta Bisconti [a], Luigi Maiorano [d], Daniele Canestrelli [a]

[a] Department of Ecological and Biological Sciences, Tuscia University, Viterbo, Italy
[b] Aspromonte National Park, Santo Stefano in Aspromonte, Italy
[c] Città Metropolitana di Reggio Calabria, Piazza Italia, 89100 Reggio Calabria, Italy
[d] Department of Biology and Biotechnologies «Charles Darwin», Sapienza University, Rome, Italy

ARTICLE INFO

Keywords:
High-resolution local climates
Topographic heterogeneity
Interpolation
Downscaling
Climate products

ABSTRACT

Topography modulates air and soil temperature and gives rise to local climates that may strongly deviate from regional macroclimates. These local climatic conditions determine the distribution of many organisms but are hardly accounted for by coarse-grained macroclimate grids. Downscaling macroclimate grids or interpolating high-resolution climate surfaces from weather station data has thus become a staple in ecological research. Against this background, topographically complex territories pose a challenge, as downscaling and interpolation techniques are susceptible to high topographic heterogeneity. Simultaneously, these regions will act as refugia and preserve biodiversity amidst the challenges of climate change. This contrast renders choosing the technique to derive local climatic conditions under high topographic complexity critically relevant.

In this study, we compared how interpolation and downscaling techniques predict the local climates of topographically heterogeneous territories. We interpolated monthly average temperatures from weather stations in a Southern Italy mountain region and downscaled WorldClim and CHELSA macroclimatic grids from their native 30 arcsecs to 10 m resolution. We carried out the interpolation and downscaling procedures by employing four commonly used statistical algorithms and by including physiographic descriptors (altitude, northness, eastness, distance to the coast, and monthly average of daily clear-sky insolation time) known to determine local climatic conditions. We compared the techniques' predictive performance via leave-one-out cross-validation. We extended the comparison over the study area to identify where the techniques most strongly diverged and evaluated how interpolation and downscaling techniques fare against environmental extrapolation.

Although the interpolations scored the best values across all error metrics, interpolating weather station data and downscaling WorldClim yielded essentially equivalent predictions of local climates in the cross-validations. Differences between the two techniques emerged during the spatial comparison, where downscaling WorldClim produced biased and less detailed monthly climate surfaces. On the contrary, weather station interpolation was affected by environmental extrapolation in the study area's most internal and topographically heterogeneous landscapes. Downscaling CHELSA severely over-smoothed local climatic conditions during both cross-validations and spatial comparisons, possibly due to lacking climate variability in the native product itself.

We demonstrated that technique and data source choice affect the prediction of local climates in topographically heterogeneous territories. Despite its limitations, we argue that interpolating weather station data better accounted for local-scale topographic effects. These effects form an integral part of the microrefugia that will shelter mountain communities from warming climates. We thus recommend carefully considering the techniques chosen to unravel the intricate climatic mosaics of topographically heterogeneous territories.

* Corresponding author.
E-mail address: daniele.dellemonache@unitus.it (D. Delle Monache).

https://doi.org/10.1016/j.ecoinf.2024.102674
Received 30 November 2023; Received in revised form 6 June 2024; Accepted 7 June 2024
Available online 8 June 2024

1. Introduction

Amidst the socio-economic consequences of globalization, climate change is arguably the most impactful on the world's ecosystems (Rosenzweig et al., 2008). Climate change prompts many terrestrial organisms to redistribute upslope and poleward to follow their climatic niche (Chen et al., 2011; Parmesan, 2006). It also forces many organisms to modify their physiology and seasonal activity to remain synchronous with environmental conditions favorable for growth and reproduction (Canestrelli et al., 2017; Parmesan and Yohe, 2003). Although climate change is occurring globally, it impacts some ecosystems more than others; in this respect, mountain ecosystems seem to be among the most vulnerable ones. The rate of warming increases with elevation, meaning that high-elevation environments are exposed to faster thermal changes than low-elevation ones (McCullough et al., 2016; Pepin et al., 2015). Such thermal changes strongly affect mountain biological communities. For instance, in a recent and extensive review concerning the European Alps (Vitasse et al., 2021), several taxa were shown to have shifted spring phenology and distribution range over recent decades. In addition, most of them struggled to keep track of the broad-scale isotherm shifts induced by climate warming.

At the same time, mountain species may experience lower extinction rates under warming climates than lowland species (Scherrer and Körner, 2011). The reason is the outstanding variety of temperature contrasts that mountain ranges house over short linear distances. These contrasts are comparable to those found along much broader elevational or latitudinal gradients (Scherrer and Körner, 2010). They facilitate vulnerable taxa coping with climate changes, allowing them to track the isotherm shifts easily by short-range migration (Sandel et al., 2011). Mountain ranges offered refuge to cold-adapted species during interglacials (Camacho-Sanchez et al., 2018; Stewart et al., 2010). Under warming climates, we should thus deem the preservation of mountain ranges of paramount importance. In particular, we should focus on preserving those small-scale climate niches expected to shelter biodiversity more effectively from climate change (Lenoir et al., 2013; Luoto and Heikkinen, 2008).

These areas of unusual microclimates relative to their surroundings are known as microrefugia or cryptic refugia (Ashcroft, 2010; Rull, 2009). Microrefugia are now finally recognized as game-changers for biodiversity conservation under climate change, and emerging technologies are increasingly facilitating their identification. Early microrefugia identification was qualitative, with microrefugia being equated with elements of the territory like boulder fields (Shoo et al., 2010) and old, climatically buffered, infertile landscapes (OCBILs sensu Hopper, 2009). Also, microrefugia were equated with areas that house stable climatic conditions becoming increasingly rare (Dobrowski, 2011) or that harbor climate-relict populations (Hampe and Jump, 2011). Instead, only one decade after these first attempts, microrefugia are identified by quantitatively modeling their climatic components. We can locate them in many ways, for example, via spatial interpolation of field measurements (Frey et al., 2016; Milling et al., 2018) or mechanistic models of fine-scale microclimates (Kearney and Porter, 2017; Maclean et al., 2019).

The growing availability of macroclimate grids is also coming in handy. These products provide information about free-air macroclimates concerning large shares of the Earth's surface and are commonly used to model species distributions over large spatial extents (Elith and Leathwick, 2009). Unfortunately, macroclimate grids are usually too coarse to capture the range of microclimates available at smaller scales (Potter et al., 2013). Thus, macroclimate grids might need to be downscaled to the appropriate spatial scales before they can be effectively used (Lenoir et al., 2017). High-resolution free-air temperature products are routinely obtained by linking macroclimates to the physiographic descriptors that drive local climatic conditions (McCullough et al., 2016). Then, after modeling the offset between local climates and microclimates (Lenoir et al., 2017; Maclean et al., 2019) or relying on available maps of temperature offsets (Haesen et al., 2021), these free-air climatic conditions can help characterizing the in-situ microclimates that organisms experience. However, downscaling free-air temperatures propagate the underlying uncertainty of the macroclimate grids; these products are built upon the interpolation of weather station data. In this respect, using weather station data to produce high-resolution climate surfaces directly might prevent the uncertainty from affecting the estimation of local climates.

Weather station data have been successfully employed in many studies so far. For example, Aalto et al. (2017) integrated weather station data with topographic information. They witnessed substantial improvements in estimating local climates compared to a more conventional approach involving only geographical location, elevation, and water cover. Meineri and Hylander (2017) similarly linked weather station data with topography via three different linear mixed models to obtain fine-grain climatic grids. They then supplemented these grids with coarser climate data to produce more refined species distribution models (SDMs) for 78 vascular alpine plants over Sweden. Lastly, Stark and Fridley (2022) compared SDMs generated using interpolated meteorological station data at two spatial scales, with one based on a previously validated microclimate model (Fridley, 2009). Their results reaffirm the importance of explicitly considering microclimates when projecting organismal response to climate change.

Ultimately, interpolating weather station data and downscaling macroclimate grids are valuable allies for mapping local climatic conditions. However, both techniques are likely to be challenged by high topographic complexity. On the one hand, weather stations tend to be located in low-elevation and relatively flat terrains, which puts predicting local climates in otherwise impervious mountain ranges under threat of environmental extrapolation (Bolstad et al., 1998; Daly et al., 1994; Phillips et al., 1992). Weather stations must also be densely distributed across the territory to be successfully interpolated (Guisan and Zimmermann, 2000; Lookingbill and Urban, 2003). On the other hand, downscaling techniques are likewise susceptible to extrapolation since they are calibrated on coarse-resolution products that lack the topographic information of the territory. These techniques might thus fail to encompass the fine-scale relationships between local climates and their physiographic drivers when the latter interact in extremely complex ways, as in highly heterogeneous mountain ranges (Barry and Blanken, 2016).

That is where we come to a paradox; interpolation and downscaling techniques could fail precisely where their contribution would be most decisive. Assessing how interpolating weather station data and downscaling macroclimate grids compare in highly heterogeneous mountain ranges is therefore urgently needed. Here, we leverage a highly climatically complex and topographically heterogeneous mountain range in Southern Italy to perform this comparison by interpolating monthly climate data from 33 weather stations and by downscaling two of the most widely used macroclimate grids, namely WorldClim (version 2.1; Fick and Hijmans, 2017) and CHELSA (version 1.1; Karger et al., 2017). We started by following a pointwise leave-one-out cross-validation approach to compare the techniques' predictive performance at the weather stations' locations. We then extended the comparison over the entire study area by constructing monthly climate surfaces for each technique, which we confronted to highlight the regions where the techniques most strongly diverged. Finally, we compared the robustness of interpolation and downscaling techniques to environmental extrapolation by investigating how similar their local climate estimates were and whether they remained coherent as the extrapolation increased.

2. Materials and methods

2.1. Study area

Our study area encompasses the Aspromonte region, located at the southern tip of the Apennines in Calabria, south Italy (Fig. 1). This

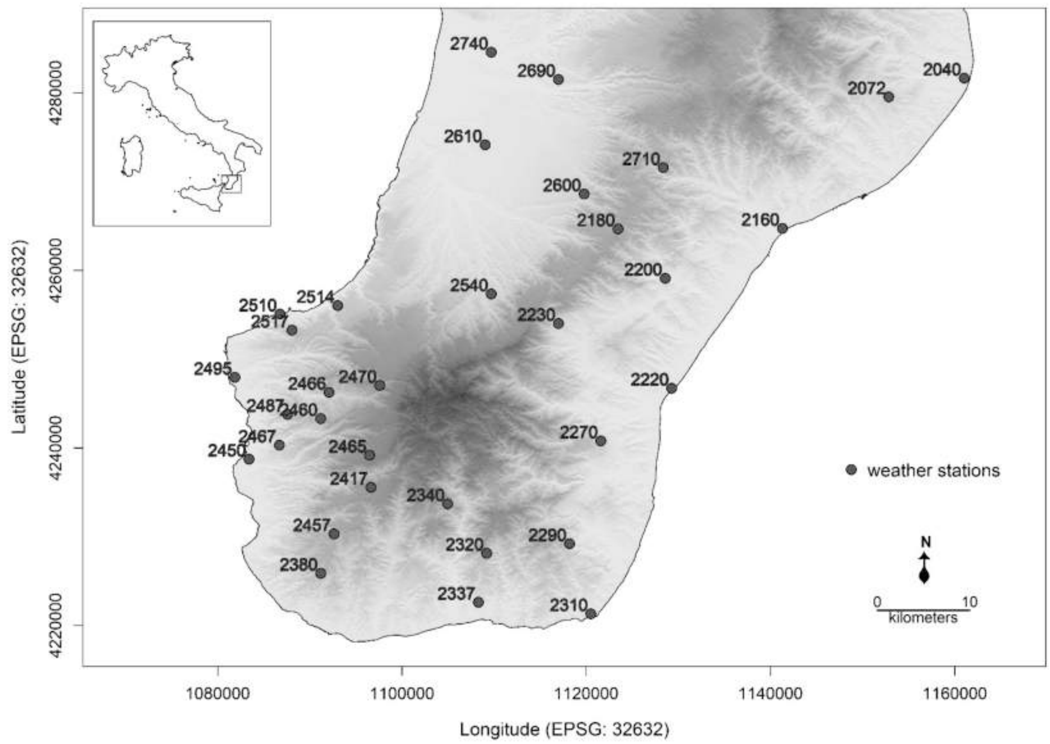

Fig. 1. Locations of the weather stations employed in the present study. We indicated the weather station's ID above each location. The inset displays the borders of the Italian peninsula, with those of the study area highlighted in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

region is home to the homonymous Aspromonte mountain range, which, by its conformation and topography, features a solid inner-outer macroclimatic gradient. Its rainfall distribution is typically Mediterranean, with abundant precipitation during autumn, winter, and spring that becomes scarce in summer; annual average rainfall over the region amounts to about 1000 mm, 83% of which is recorded from October to April (Colacino et al., 1997). Due to moderate-altitude reliefs reaching 1957 m asl, interior areas are colder and wetter than the peripheral; they indeed belong to a temperate oceanic bioclimate, while those along the coast to a Mediterranean rain-seasonal oceanic bioclimate (Brullo et al., 2001). Annual average temperatures reach about 18 °C along the coast and 13 °C inland; annual average rainfall above 800 m usually reaches 1100 mm while drops to 900 mm in valleys and coastal areas. On top of this macroclimatic gradient, the prevalently north-south disposition of the Aspromonte adds another east-west gradient in local climatic conditions. Westerlies cause humid air masses to be intercepted by the western side of the Aspromonte, which leaves a rain shadow on its eastern side. Consequently, local climates on the west are markedly wetter and cooler (annual average rainfall: 900 mm), and those on the east are warmer and drier (annual average rainfall: 700 mm). These inner-outer and east-west gradients give rise to a remarkable range of climate niches over short horizontal distances. In this light, it is no surprise that the Aspromonte mountain range represents a biodiversity hotspot for plant and animal communities at all levels of biological organization (Spampinato et al., 2008; Zampiglia et al., 2019). The region is also identified as a glacial refuge and constitutes a hotspot of genetic diversity for many temperate species (Chiocchio et al., 2024; Martino et al., 2022).

2.2. Weather station data

From the regional environmental agency for Calabria (http://www.arpacal.it/), we retrieved daily average temperatures for 33 weather stations from 1923 to 2020. As these climate time series suffered from severe data deficiencies (Table A.1 in Appendices), we imputed the missing temperatures using the bootstrap algorithm implemented in the Amelia II R package (Honaker et al., 2011). The algorithm leverages the correlation among incomplete time series, which is expected in weather station data, to produce multiple completed time series via multiple imputation. We ran the algorithm to impute 500 complete replicas of each weather station's time series. Then, for each weather station, we filled in the missing temperatures by taking the average of these replicas. We assessed the quality of the imputation procedures by checking the overimputation diagnostic plots and calculating the R^2 and root-mean-square errors (RMSE) of the overimputations (Fig. A.1, Fig. A.2, and Fig. A.3 in Appendices). After the imputation procedures, we converted the daily average temperatures for the 33 weather stations into monthly average temperatures. Then, we averaged a subset of the monthly average temperatures (1970–2000) to obtain 12 datasets valid for 1970–2000, one for each month.

2.3. Weather station interpolation

We interpolated weather station data through four statistical algorithms: a generalized additive model (GAM), an artificial single-hidden-layer neural network (ANN), a random forest (RF), and a generalized boosting regression model (GBM). We employed physiographic descriptors known to drive local climatic conditions as predictors in the algorithms; these descriptors were altitude, northness, eastness, distance to the coast, and monthly average of daily clear-sky insolation time. We derived most of these descriptors (altitude, northness, eastness, distance to the coast) from a digital terrain model at 10 m resolution in R, version 4.1.3 (R Core Team, 2022). Monthly averages of daily clear-sky insolation times were calculated by averaging the daily daylight lengths provided by the r.sun routine available in GRASS GIS, version 7.8.5. We fitted GAMs by defining a cubic regression spline for each descriptor and using the mgcv R package (Wood, 2011). We fitted ANNs via the nnet R package (Venables and Ripley, 2002) after performing cross-validation over five decays (0, 0.2, 0.4, 0.6, 0.8, 1) and sizes (1, 2, 3, 4, 5). We fitted RFs using the randomForest R package (Liaw and Wiener, 2002) with 500 trees, five as the minimum size of terminal nodes, and by sampling all descriptors for splitting at each node; these RFs were thus equivalent to bagged decision trees. Finally, we fitted GBMs via the gbm R package (Greenwell et al., 2020) by ensembling a maximum of 10,000 trees, with shrinkage and interaction parameters equal to 0.01 and 1, respectively; each tree was thus equivalent to a decision stump. After model calibration, we identified the most relevant physiographic descriptors for each algorithm and each month. For GAMs, we evaluated the predictors' relative importance by considering their F statistics. We assessed variable importance in ANNs by employing the Olden method (Olden et al., 2004) implemented in the NeuralNetTools R package (Beck, 2018). GBMs' variable importance was quantified by retrieving the computed relative influences, normalized to sum to 100. We finally assessed RFs' variable importance by computing the increase in node purity associated with each predictor.

To ensure that the weather stations sampled sufficient portions of landscape complexity, we first synthesized all five physiographic descriptors via principal component analysis (PCA) and quantified the contribution of each principal component (PC) to the explained variance. Then, we extracted its first two PCs and inspected the spatial sparsity of the weather stations' locations in the resulting monthly bivariate plots. For each month, we also evaluated the proportion of overlap between the study area's distribution of PCA scores along all five PCs and the sampling distribution created by the weather stations. These overlapping proportions were calculated using the overlapping R package (Pastore et al., 2022). Finally, we inspected where the interpolations extrapolated by deriving each month's surface of the Shape extrapolation metric. The Shape method measures the degree of extrapolation in environmental space by calculating the distance between training and projection data; higher Shape values thus indicate higher extrapolation degrees (Velazco et al., 2023). We calculated Shape values by considering the Mahalanobis distance and using the flexsdm R package (Velazco et al., 2022).

2.4. WorldClim and CHELSA data

We downloaded version 2.1 of WorldClim's historical monthly climate rasters ("tavg 30s") from the WorldClim website (https://www.worldclim.org/). These 12 rasters cover 1970–2000 and come at 30 arcsec resolution (~ 1 km^2). For CHELSA, we downloaded the CHELSA-EUR11 (https://chelsa-climate.org/high-resolution-climate-data-for-europe/) high-resolution (30 arcsecs; ~ 1 km^2) monthly average temperature rasters for each year from 1981 to 2005 ("CHELSA_EUR11_tas_mon_1981–2005_V1.1"). We then averaged these monthly average temperatures, similar to what we did for the weather station data, to produce 12 rasters of monthly climate covering 1981–2005.

2.5. WorldClim and CHELSA downscaling

Before downscaling WorldClim and CHELSA, we cropped the monthly climate rasters to the extent of the study area (Fig. 1). We resampled the altitude, distance to the coast, and monthly average of daily clear-sky insolation time rasters we employed in the interpolation to match the extent and resolution (30 arcsecs) of the WorldClim and CHELSA monthly climate data. Northness and eastness rasters at 30 arcsec resolution were derived from the resampled altitude raster. Then, we fitted the same four statistical algorithms (GAM, ANN, RF, GBM) we employed when interpolating weather station data to the set of centroids of each monthly climate raster, constituting our calibration points. We calibrated each algorithm by replicating the same modeling procedures we followed during the interpolations and using the resampled altitude, northness, eastness, distance to the coast, and monthly average of daily clear-sky insolation time as physiographic descriptors. As with the interpolations, we finally inspected where WorldClim and CHELSA downscalings extrapolated by comparing the training data at the calibration points with the projection data over the study area. By doing so, we obtained 12 surfaces of the Shape extrapolation metric for both WorldClim and CHELSA, one for each month.

2.6. Methodological comparison

For each monthly climate dataset, we assessed the quality of the interpolations via leave-one-out cross-validation. Therefore, each weather station was excluded once from the dataset and used as a validation point. The quality of the downscalings was evaluated by calibrating the algorithms on the set of centroids of each monthly climate raster and then cross-validating them on the weather stations' locations. The predictive performance of the techniques was compared by using four validation metrics, namely the Kling-Gupta efficiency (KGE) scores, percent biases (*pbias*), mean absolute errors (MAE), and root-mean-square errors (RMSE). KGE is a composite measure that simultaneously accounts for correlation, variability errors, and bias errors among observed and predicted data (Gupta et al., 2009) and was calculated as follows:

$$KGE = 1 - \sqrt{(\rho - 1)^2 + \left(\frac{\sigma_{pred}}{\sigma_{obs}} - 1\right)^2 + \left(\frac{\mu_{pred}}{\mu_{obs}} - 1\right)^2}$$

where ρ is the Pearson correlation coefficient between observed and predicted temperatures, σ_{pred} is the standard deviation of the predicted temperatures, σ_{obs} is the standard deviation of the observed temperatures, μ_{pred} is the mean of the predicted temperatures, and μ_{obs} is the mean of the observed temperatures. A KGE score equal to 1 indicates perfect concordance between observed and predicted temperatures, while KGE scores between 0 and 0.5 indicate poor model performance (Rogelis et al., 2016). We chose percent biases to highlight the propensity for the predicted temperatures to be larger or smaller than those measured by the weather stations. Low *pbias* values reflect accurate model prediction; the optimal *pbias* value is 0. We calculated percent biases as follows:

$$pbias = 100 \times \frac{\sum (t_{obs} - t_{pred})}{\sum t_{obs}}$$

where t_{obs} and t_{pred} are observed and predicted temperatures, respectively. Consequently, underestimation corresponds with positive *pbias* values, whereas overestimation with negative *pbias* values. MAE and RMSE are two commonly used validation metrics, and we calculated them as follows:

$$MAE = \frac{1}{n}\left(\sum |t_{obs} - t_{pred}|\right) \quad RMSE = \sqrt{\frac{1}{n}\left(\sum (t_{obs} - t_{pred})^2\right)}$$

where n (= 33) is the number of weather stations.

The predictions of the interpolations and the downscalings were also compared over the entire study area to highlight in which environmental conditions the techniques diverged the most. Hence, we constructed four monthly climate surfaces per month and technique, one for each algorithm. Then, we averaged the surfaces over the algorithms to obtain a monthly climate surface for each month and each technique. These surfaces were then compared by calculating the pairwise differences and inspecting the spatial distribution of these differences across the territory.

Finally, we assessed the impact of environmental extrapolation on the interpolation and downscaling procedures by comparing the local climate estimates of each pair of techniques at increasing degrees of extrapolation. For each month, we extracted the Shape values for each pair of techniques at 100,000 random locations. Then, we correlated these Shape values with the absolute differences between local climate predictions to check whether a joint increase in the degree of environmental extrapolation would lead to a more significant divergence between pairs of techniques.

3. Results

3.1. Methodological comparison

The imputation procedures reliably filled the missing temperatures (average $R^2 = 98.017\%$; average RMSE = 0.72832 °C). Among the three techniques, interpolating weather station data produced estimates that more closely matched the temperatures measured by the weather stations (Fig. A.4 in Appendices). Downscaling WorldClim produced reliable estimates, although they less closely matched the measured temperatures than those predicted by the interpolations (Fig. A.5 in Appendices). On the contrary, downscaling CHELSA produced over-smoothed underestimates of local monthly climates (Fig. A.6 in Appendices). Interpolating weather station data and downscaling WorldClim yielded similarly high average KGE scores. Nevertheless, WorldClim's average KGE scores were noticeably lower during fall, winter, and early spring (Fig. 2a). Interpolating weather station data and downscaling WorldClim also compared in terms of average *pbias*, although mainly from May to October (hereafter referred to as "hot season") when WorldClim's average *pbias* were only slightly positive. During the cold season, WorldClim's average *pbias* decreased towards negative values, while weather stations' average *pbias* remained close to zero (Fig. 2b). Interpolating weather station data yielded slightly lower values of average MAE and RMSE (Fig. 2c and Fig. 2d). Finally, downscaling CHELSA resulted in significantly lower average KGE scores (Fig. 2a), consistently positive average *pbias* (Fig. 2b), and average MAE and RMSE that almost doubled those we obtained from the interpolations and the WorldClim downscaling (Fig. 2c and Fig. 2d). These shortcomings could be partly attributed to the fact that CHELSA data cover a different period (1985–2000) than WorldClim and weather station data (1970–2000).

Downscaling WorldClim produced smoother and less detailed monthly climate surfaces than interpolating weather station data; this was particularly evident during the cold season (Fig. 3c and Fig. 3e) but remained true during the hot season (Fig. 4c and Fig. 4e). During the cold season, downscaling WorldClim also resulted in lower temperatures in valleys and at low altitudes along the coast (Fig. 3f and Fig. 3g). Temperature differences reversed moving along the slopes, where downscaling WorldClim yielded higher temperatures compared to weather station interpolation (Fig. 3f and Fig. 3g), thus echoing the lower average *pbias* values that emerged in the pointwise cross-validations. During the hot season, downscaling WorldClim led to higher temperatures near the coast and lower temperatures across most of the study area (Fig. 4f and Fig. 4g) compared to interpolating weather station data; these results were in line with WorldClim downscaling scoring higher average *pbias* values during these months. Throughout the year, weather station interpolation and WorldClim downscaling consistently diverged at higher altitudes: interpolating weather station data yielded higher temperatures that exceeded WorldClim's by >3 °C during the hot season (Fig. 3f and Fig. 4f). Finally, similarly to how it behaved in the pointwise cross-validations, CHELSA downscaling produced heavily over-smoothed monthly climate surfaces (Fig. 3h and Fig. 4h). Consequently, temperature differences between CHELSA downscaling and the other two techniques closely matched the thermal gradients of the study area both during the cold season (Fig. 3b and Fig. 3d) and the hot season (Fig. 4b and Fig. 4d).

At the 100,000 random locations we sampled each month, downscaling WorldClim and interpolating weather station data yielded similar monthly climate estimates (Fig. A.7 in Appendices). Monthly estimates diverged as the Shape values (i.e., degrees of extrapolation) jointly increased but seldom exceeded 2.5 °C. Divergence was highest during the hot season. Monthly climate estimates were far less similar between weather station interpolation and CHELSA downscaling (Fig. A.8 in Appendices) and between WorldClim and CHELSA downscalings (Fig. A.9 in Appendices). No clear divergence at increasing Shape values could be observed for these two pairs of techniques. Finally, the Shape values of weather station interpolation and WorldClim and CHELSA downscalings only moderately correlated ($R^2 \approx 0.30$; Fig. A.7 and A.8 in Appendices); on the contrary, WorldClim and CHELSA downscalings extrapolated nearly identically ($R^2 \approx 0.80$; Fig. A.9 in Appendices).

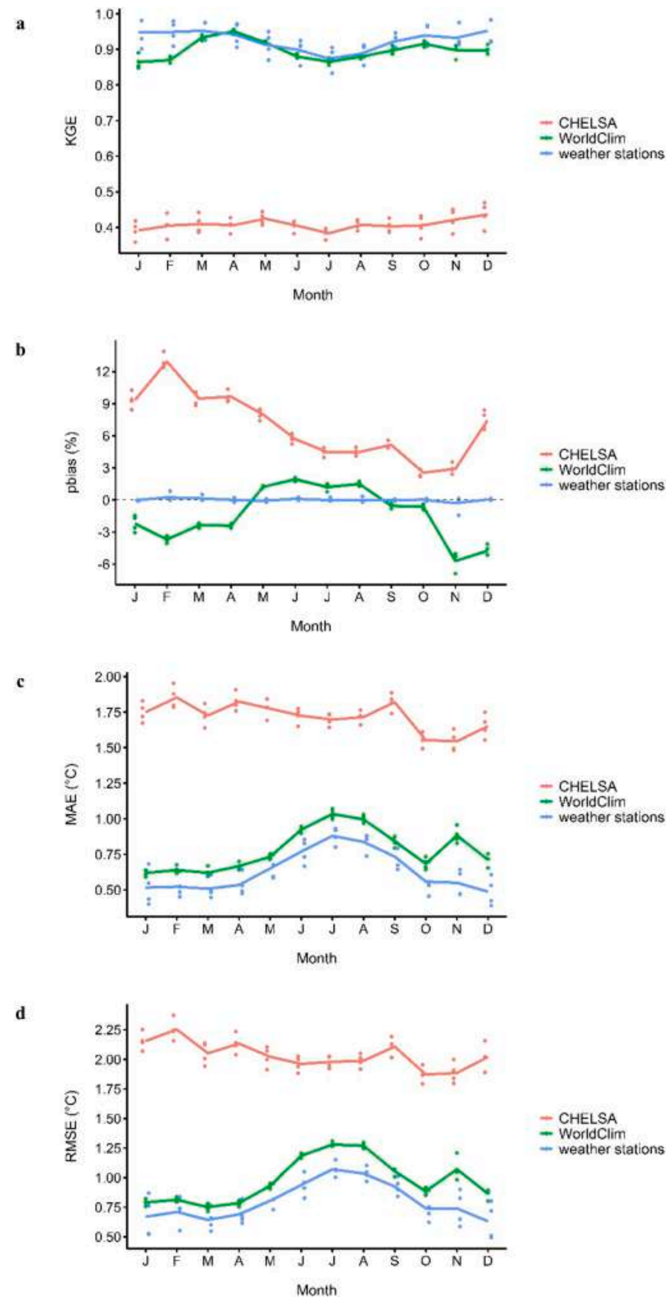

Fig. 2. Seasonal patterns in average KGE scores (**a**), *pbias* (**b**), MAE (**c**), and RMSE (**d**) that we obtained when predicting monthly average temperatures using each technique. Lines and points represent average scores and those obtained by each algorithm, respectively.

3.2. Weather station interpolation

During the hot season, KGE scores decreased for all the algorithms except ANN. For most of the year, GBM ranked as the least efficient algorithm. Nevertheless, KGE scores remained high (> 0.80; Fig. A.10a

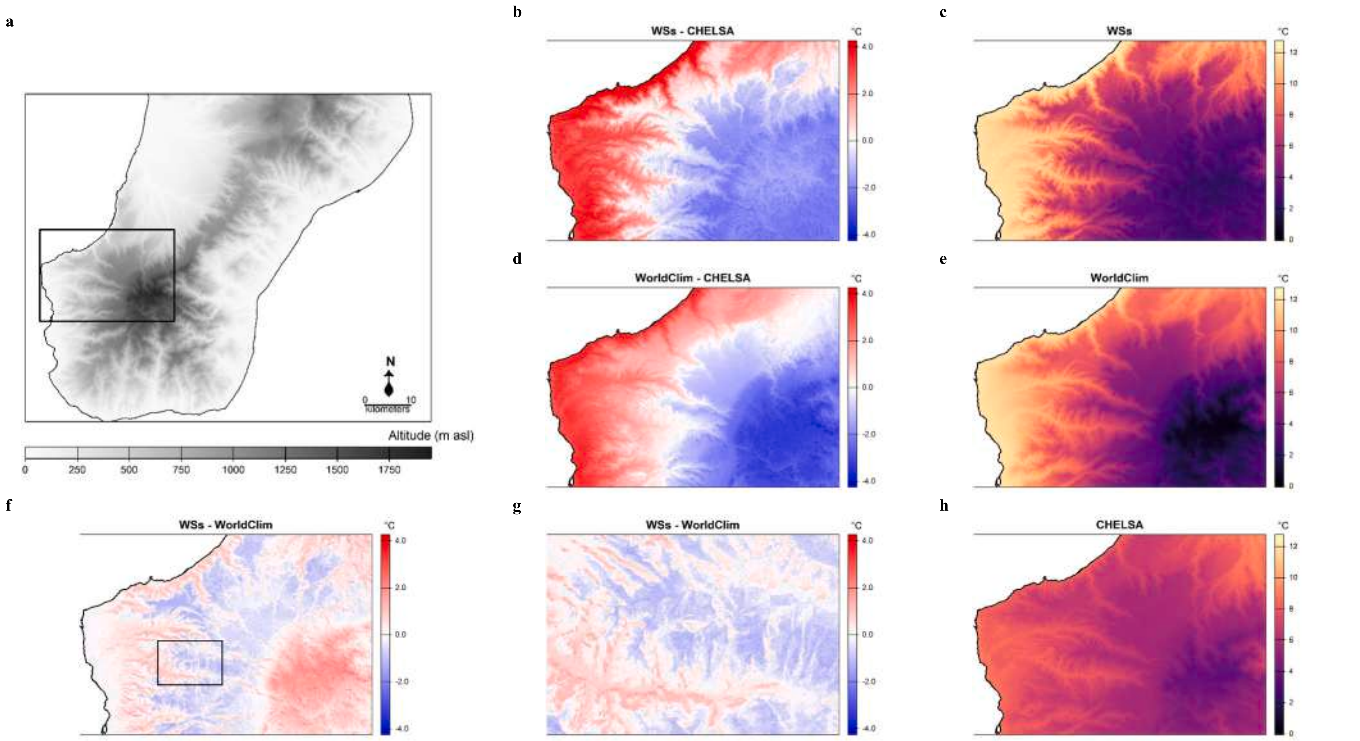

Fig. 3. Average January temperatures predicted by interpolating weather station (WS) data and downscaling WorldClim and CHELSA. The contours of the geographical area used for comparison are shown in subplot (**a**). Subplots (**b**), (**d**), and (**f**) display the differences between the temperature surfaces of WSs interpolation and WorldClim downscaling, WSs interpolation and CHELSA downscaling, and WorldClim and CHELSA downscaling, respectively. Subplots (**c**), (**e**), and (**h**) display the temperature surfaces predicted by WSs interpolation and WorldClim and CHELSA downscaling, respectively. Subplot (**g**) shows the details of the area contoured in subplot (**f**).

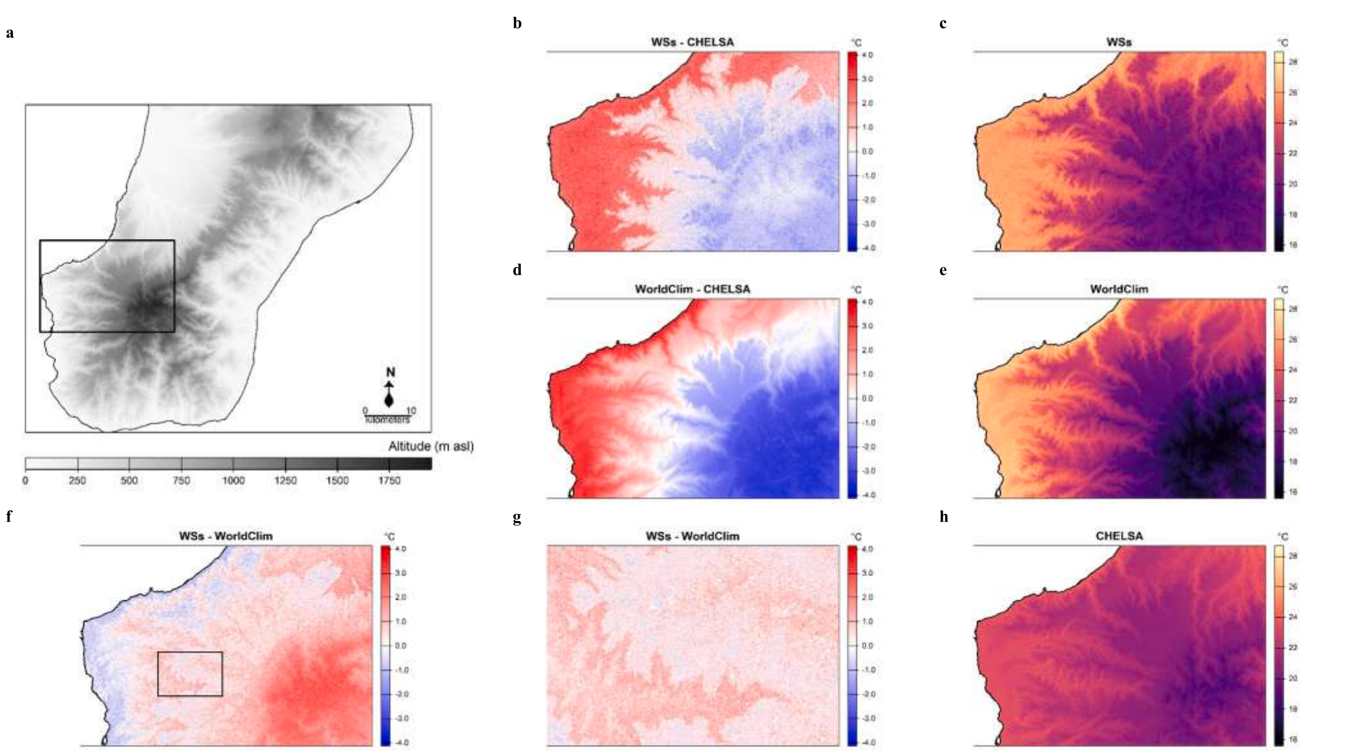

Fig. 4. Average July temperatures predicted by interpolating weather station (WS) data and downscaling WorldClim and CHELSA. The contours of the geographical area used for comparison are shown in subplot (**a**). Subplots (**b**), (**d**), and (**f**) display the differences between the temperature surfaces of WSs interpolation and WorldClim downscaling, WSs interpolation and CHELSA downscaling, and WorldClim and CHELSA downscaling, respectively. Subplots (**c**), (**e**), and (**h**) display the temperature surfaces predicted by WSs interpolation and WorldClim and CHELSA downscaling, respectively. Subplot (**g**) shows the details of the area contoured in subplot (**f**).

in Appendices). Like KGE scores, MAE and RMSE increased for all the algorithms during the hot season, although both metrics showed limited variability (0.4–1.3 °C; Fig. A.10c and Fig. A.10d in Appendices). Percent biases were always remarkably close to zero, with little differences between algorithms (Fig. A.10b in Appendices). Altitude was the most relevant physiographic descriptor; however, in all algorithms but ANN, its relative importance decreased during the hot season. In these same months, the relative importance of the remaining descriptors generally increased (Fig. A.11 in Appendices).

Principal component analysis indicated that weather stations sampled moderate amounts of landscape complexity: the study area's distributions of PCA scores and the sampling distributions of the weather stations mostly overlapped, especially along the first two PCs, which accounted for about 55% of the total variance (Table A.2 and Fig. A.14 in Appendices). Most of the unexplained variance, and thus the environmental extrapolation, involved the innermost regions of the study area, which are those at higher altitudes and more heterogeneous topographically. This pattern was especially evident during the cold season but remained consistent throughout the year (Fig. A.15 in Appendices).

3.3. WorldClim and CHELSA downscaling

KGE scores did not follow any seasonal trend for WorldClim (Fig. A.10e in Appendices) or CHELSA (Fig. A.10i in Appendices). However, while WorldClim's KGE scores were consistently high (> 0.80), CHELSA's KGE scores indicated poor model performance (< 0.50) all year round and across all algorithms. WorldClim's percent biases were slightly positive during the hot season and negative for the remaining year. In this respect, all the algorithms behaved homogeneously (Fig. A.10f in Appendices). The algorithms behaved homogeneously for CHELSA, too, but always showed significantly positive percent biases (Fig. A.10j in Appendices). Like weather stations, WorldClim's MAE and RMSE increased during the hot season (Fig. A.10g and Fig. A.10h in Appendices). We did not observe this same increase when downscaling CHELSA (Fig. A.10k and Fig. A.10l in Appendices). Altitude was also the most relevant physiographic descriptor for WorldClim and CHELSA downscalings. Contrary to weather station interpolation, its relative importance decreased moderately during the hot season, especially for WorldClim's GAMs and CHELSA's GAMs and RFs. When downscaling macroclimate grids, the relative importance of some other descriptors showed seasonal variation, too, albeit to a far lesser extent than for weather station interpolation. In particular, distance to the coast was the only descriptor whose importance increased when downscaling WorldClim during the hot season; for CHELSA downscaling, the importance of distance to the coast and insolation time displayed reversed seasonal patterns compared to the other techniques, decreasing during the hot season. No seasonal patterns could be observed for WorldClim's and CHELSA's ANNs, as the behavior of these algorithms was remarkably unstable (Fig. A.12 and Fig. A.13 in Appendices).

Compared to the interpolations, WorldClim and CHELSA downscalings extrapolated more diffusely: except for the northwestern lowland region, extrapolation covered the study area almost homogeneously. It also followed no clear seasonal patterns (Fig. A.16 and Fig. A.17 in Appendices).

4. Discussion

In this study, we aimed to assess how interpolation and downscaling techniques compare in predicting local monthly climates in highly heterogeneous mountain ranges. We interpolated weather station data and downscaled two of the most widely used macroclimate grids – WorldClim and CHELSA – to check if the data source could affect the downscaling performance. Although it was impossible to directly validate the techniques against observed climates over the entire study area, approaching the comparison from different angles allowed us to draw several valuable insights.

4.1. Do interpolating weather station data and downscaling macroclimate grids compare in heterogeneous mountain ranges?

Firstly, we compared the interpolation and downscaling techniques by validating them on the only available observed climates, namely the weather stations' measurements. In this context, the interpolation of weather station data ranked as the best-performing technique across all validation metrics (Fig. 2). WorldClim downscaling closely followed, scoring RMSE/MAE values only slightly higher (~ 0.15 °C; Fig. 2c and Fig. 2d) than the interpolations. Downscaling WorldClim also yielded KGE values that rivaled those obtained when interpolating weather station data (Fig. 2a). However, the marginally superior performance of the interpolations should be taken with due caution. Despite capturing good shares of landscape complexity (Table A.2 and Fig. A.14 in Appendices), the weather stations we employed cannot be considered fully independent as they left the most internal and heterogeneous regions above 1250 m asl uncovered. Performing leave-one-out cross-validation in such areas would likely have worsened the interpolations' performance. So, it might be fair to say that interpolating weather station data and downscaling WorldClim emerged, concerning this initial pointwise comparison, as essentially equivalent.

Were then interpolating weather station data and downscaling WorldClim also equivalent over the territory? Here, the answer gets more nuanced. On the one hand, the two techniques produced monthly climate surfaces that were remarkably close, rarely differing by >2 °C (Fig. 3 and Fig. 4). They demonstrated they were jointly robust against environmental extrapolation since their temperature estimates remained coherent as the algorithms increasingly extrapolated (Fig. A.7 in Appendices). Interpolations and WorldClim downscaling correctly identified altitude as the most relevant physiographic descriptor (Fig. A.11 and Fig. A.12 in Appendices); at such a high spatial resolution (10 m), we indeed expected the temperature lapse rate to be the predominant driver of local climatic conditions (Joly et al., 2018). More interestingly, both techniques detected decreases in altitude's relative importance during the hot season. They also exhibited more divergent monthly climates during this season (Fig. A.7 in Appendices), lower KGE scores (Fig. 2a), and higher RMSE/MAE values (Fig. 2c and Fig. 2d). These results indicate increased spatial variability in local climates, a phenomenon witnessed in many other geographical contexts (Caesar et al., 2006; Hopkinson et al., 2012; Jones and Briffa, 1992; Srivastava et al., 2009). The reasons for this could be topographical. The proximity of the Aspromonte relief to the sea might affect cloud formation, thereby promoting uneven cloud cover patterns that might induce maximum temperatures to vary more during summer. This phenomenon has, for example, been observed in the Iberian Peninsula (Pena-Angulo et al., 2015). Also, increased evaporation during summer might bolster the formation of precipitation along windward slopes, further enhancing spatial variability in local climates by modifying latent heat fluxes (Martin-Vide, 2004). So, regardless of the reasons explaining them, these shared patterns provided evidence that interpolating weather station data and downscaling WorldClim behaved similarly also over the study area.

On the other hand, the two techniques differed in several crucial aspects. One of the study's premises was that downscaling techniques could have difficulties capturing climate variability in topographically complex landscapes. Our results confirm that these techniques can struggle to account for topographic effects. Indeed, downscaling WorldClim yielded monthly climate surfaces that were smoother than those obtained from the interpolations (Fig. 3 and Fig. 4). Compared to weather station interpolation, this technique also produced consistent underestimates across most of the study area and overestimates along the coast during the hot season (Fig. 4), as well as underestimates in valleys and overestimates along slopes during the cold season (Fig. 3).

Importantly, these biases were reflected by the cross-validations, too, namely by consistently negative average *pbias* values (i.e., underestimation) during the hot season and positive average *pbias* values (i.e., overestimation) during the cold season (Fig. 2a). That downscaling techniques could have encountered some difficulties was anticipated. As mentioned, downscaling macroclimate grids cannot catch climate-forcing effects not initially present in the coarse-scale products (Lembrechts et al., 2019). WorldClim's grid values, for instance, are initially derived by interpolating weather station measurements with the help of altitude and distance to the coast as topographical predictors (Fick and Hijmans, 2017); interestingly, these two descriptors were the only relevant ones when we downscaled WorldClim (Fig. A.12 in Appendices). In this respect, our results underline that these difficulties can have practical impacts in highly heterogeneous mountain ranges, introducing biases in local climate estimates and leading to an overall loss of topographical detail.

Despite producing the most detailed and accurate temperature predictions (Fig. A.4 in Appendices), interpolating weather station data also had its share of difficulties. This technique, as expected, extrapolated in the most challenging regions of the study area from a topographical standpoint, namely the innermost high-altitude landscapes (Fig. A.15 in Appendices). While it was impossible to quantify its impact against observed climates directly, such extrapolation affected the technique's performance. Indeed, weather station interpolation produced monthly climates that were several degrees (up to >3 °C during the hot season; Fig. 4) warmer than those from WorldClim downscaling. Weather stations are not new to this problem; temperature overestimation in high-altitude environments has been witnessed several times (Bolstad et al., 1998; Daly et al., 1994). One possible solution could be limiting the predictors' interactions and the flexibility of their response curves when calibrating the algorithms to prevent them from predicting unrealistic climates in the extrapolations. However, a price for doing this might be paid when trying to keep up with topographic heterogeneity. Too rigid model parameterizations could fail to encompass the locality and complexity of climate-forcing in heterogeneous mountainous landscapes. Temperature lapse rates are notoriously variable in these environments due to local phenomena such as cold-air pools and other topographically induced thermal inversions (Gudiksen et al., 1992; Whiteman et al., 2001). Correctly modeling these lapse rates is critical, as altitude was found to be the most important variable for all the techniques, although less predominantly for the interpolation of weather stations (Fig. A.11, A.12, and A.13 in Appendices). Also, wind fields are affected by the position and topographic conformation of the mountain reliefs, leading to local phenomena like foehn effects that further complicate the thermal gradients of these regions (Drechsel and Mayr, 2008; Dujardin and Lehning, 2022; Ha et al., 2009).

Finally, cross-validations and pairwise spatial comparisons showed that downscaling CHELSA led to over-smoothed and biased temperature estimates. Positive average *pbias* values (Fig. 2b) indicated that the technique systematically underestimated local climates at the weather stations' locations, KGE scores (Fig. 2a) and RMSE/MAE values (Fig. 2c and Fig. 2d) never compared with those of the other two techniques, and monthly climate surfaces displayed severely limited spatial climate variability (Fig. 3 and Fig. 4). We expected that CHELSA, relying on more refined routines integrating small-scale orographic effects, would perform better in highly heterogeneous mountain ranges than WorldClim; instead, it produced monthly climate surfaces that came to differ by >5 °C from the interpolations and WorldClim downscaling. Most of CHELSA's poor performance results from a lack of spatial climate variability in the native CHELSA product itself. Indeed, when we compared the empirical temperature lapse rates for the native CHELSA and WorldClim products in January and July, we witnessed much lower temperature range between maxima in lowlands and minima at higher altitudes in CHELSA compared to WorldClim (Fig. A.18 in Appendices). These results were echoed by CHELSA's over-smoothed monthly climate surfaces for both January (Fig. 3h) and July (Fig. 4h). As mentioned earlier, the quality of the downscaling predictions is clearly limited by the quality of the native coarse-grained product. Lembrechts et al. (2019) incurred a similar limitation when they employed CHELSA bioclimatic variables at native (30 arcsecs) and downscaled (30 m) resolutions to model the distribution of 50 plant species, albeit to a far lesser degree. Although they obtained comparable performance between SDMs calibrated with CHELSA and with the topoclimatic interpolation of weather station data from Aalto et al. (2017), they nonetheless gained no substantial benefit from downscaling CHELSA compared to using its coarse-grained version. It is finally worth recalling that CHELSA data covering a different period (1985–2000) than WorldClim and the weather stations (1970–2000) could contribute to CHELSA's lacking performance.

4.2. *What could be the implications for species conservation under climate change?*

Topographically complex territories constitute decisive cornerstones for species conservation under climate change. They house significant variation in microclimatic conditions, which provides microrefugia for vulnerable species to endure warming. Topographically complex territories exhibit greater stability in natural populations, as observed in the case of various British butterfly species (Oliver et al., 2010). Suggitt et al. (2018) showed that, after accounting for anthropogenic pressure, several climate-threatened species across England, including vascular plants, bryophytes, lepidopterans, and coleopterans, have suffered lower population losses in areas exhibiting higher topographic heterogeneity in recent decades. Therefore, failing to account for local topographic effects hinders our ability to locate microrefugia crucial to population survival. For instance, incorporating topographic heterogeneity into SDMs also substantially increased the projected persistence of European tree species (Randin et al., 2009). Similarly, disregarding topographic heterogeneity inflated the projection of future butterfly species losses in mountainous regions (Luoto and Heikkinen, 2008). Finally, including topographically-induced temperature spatial variability improved niche estimates in SDMs calibrated over topographically heterogeneous regions (Karger et al., 2023).

This study compared interpolation and downscaling techniques by predicting monthly average temperatures. However, minima and maxima are equally, if not more, relevant in shaping the niche of many organisms; weather and climate extremes also contribute to setting the bounds of organisms' range limits (Germain and Lutz, 2020; Ummenhofer and Meehl, 2017). Understanding which techniques to downscale these additional climates with would further boost our modeling capabilities. Due to their complex fine-scale topographic effects, this is especially true for topographically heterogeneous mountain landscapes (Barry and Blanken, 2016).

In vegetated areas with complex topographies, topography interacts with soil characteristics and vegetation structure to generate mosaics of even smaller-scale microclimates. These interactions can take particularly complex forms and delimit the thermal niches most organisms inhabiting forests or other vegetated areas experience (Geiger et al., 2009; Rita et al., 2021). Microclimates can substantially differ from free-air local climates (Barry and Blanken, 2016); as such, interpolations and downscalings like ours failed to capture microclimatic variation at ground level on several occasions (Lembrechts et al., 2019; Slavich et al., 2014). Mapping microclimates is, in fact, the prerogative of specific statistical routines that range from the interpolation of in-situ microclimate measurements (Ashcroft and Gollan, 2012; Vanwalleghem and Meentemeyer, 2009) to the mechanistic modeling of near-the-ground microclimates (Kearney and Porter, 2017; Maclean and Klinges, 2021). As mentioned, some routines require local climates to be estimated before mapping microclimates, which are later derived by modeling the offset between local climates and microclimates directly (Lenoir et al., 2017; Maclean et al., 2019) or by leveraging available maps of temperature offsets like ForestTemp (Haesen et al., 2021). As forest covers will enhance local buffering from future macroclimate warming (De Lombaerde et al., 2022), how to best map forest microclimates is increasingly studied (Maclean and Klinges, 2021; Zellweger et al., 2019). Although deriving local climatic conditions through interpolations and downscalings is often not enough, our results suggest that this step should not be overlooked; indeed, choosing the most suitable technique might significantly improve the temperature estimates and thus prevent statistical uncertainty from propagating down the prediction process.

5. Conclusions

High topographic heterogeneity severely challenges our attempts to derive high-resolution local climate surfaces. In this respect, are interpolation and downscaling techniques equally performant? Interpolating weather station data managed to capture the fine-scale variability in local climatic conditions; downscaling macroclimatic grids, on the contrary, provided biased and less detailed monthly climate surfaces. Downscaling performance was also found to heavily depend on the choice of the macroclimate grid since downscaling WorldClim yielded far better temperature estimates than CHELSA. Instead, which algorithm to operate with had a lower impact on the accuracy of the predictions. Given how critical accounting for fine-scale climatic variability is for reliable niche modeling, our study suggests that interpolating weather station data might have the edge over downscaling macroclimate grids in topographically heterogeneous landscapes. However, its advantage was not clear-cut; interpolations suffered from environmental extrapolation in the highest and most heterogeneous landscapes. It is riveting to note how even our simple comparative study elicited far-from-trivial considerations. After all, the stakes are very high: identifying and protecting microrefugia directly depends on us being able to portray the mosaic of fine-grain climatic conditions that topography, vegetation, and soil collectively contribute to creating. In topographically heterogeneous landscapes, such as the Aspromonte mountain range, biodiversity at all levels of organization manages to flourish precisely due to the complexity of these climatic mosaics. Thus, understanding the relationships that natural populations establish with the landscape requires us to carefully select the statistical techniques at our disposal. Only by doing so can we protect those microrefugia that may allow vulnerable mountain communities to endure future climate change.

Funding sources

This work was funded by the Tuscia University and the Aspromonte National Park. Research project implemented under the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.4 - Call for tender No. 3138 of 16 December 2021, rectified by Decree n.3175 of 18 December 2021 of Italian Ministry of University and Research funded by the European Union - Next Generation EU. Project code CN_00000033, Concession Decree No. 1034 of 17 June 2022 adopted by the Italian Ministry of University and Research, CUP J83C22000860007, Project title “National Biodiversity Future Center - NBFC”, and under the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.5, funded by the European Union - Next Generation EU, Project code ECS 0000024 Rome Technopole, CUP B83C22002820006.

CRediT authorship contribution statement

Daniele Delle Monache: Data curation, Investigation, Methodology, Software, Visualization, Writing – original draft. **Giuseppe Martino:** Data curation, Writing – review & editing. **Andrea Chiocchio:** Investigation, Supervision, Writing – review & editing. **Antonino Siclari:** Funding acquisition, Writing – review & editing. **Roberta Bisconti:** Visualization, Writing – review & editing. **Luigi Maiorano:** Methodology, Supervision, Validation, Visualization, Writing – review & editing. **Daniele Canestrelli:** Conceptualization, Funding acquisition, Supervision, Validation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Climate data can be downloaded from the website of the regional environmental agency for Calabria (http://www.arpacal.it/) and from the WorldClim (https://www.worldclim.org/) and CHELSA (https://chelsa-climate.org/) websites.

Acknowledgements

This paper is dedicated to the memory of Sergio Tralongo, who supported this research as past director of the Aspromonte National Park.

Appendix A. Appendices

Table A.1
Coordinates, altitude, daily average time series length, and proportion of missing data for 1970–2000 for each weather station employed in the present study.

ID	Coordinates (EPSG: 32632)		Altitude (m a.s.l.)	Daily average time series		Missing data for 1970–2000
	longitude	latitude		start	end	
2040	1,161,029	4,281,649	74.97	1989-01-19	2021-01-31	62.06%
2072	1,152,849	4,279,535	393.49	2004-10-14	2021-01-31	100%
2160	1,141,307	4,264,722	13.03	1999-05-06	2021-01-31	94.65%
2180	1,123,456	4,264,655	944.89	1963-01-01	2021-01-31	22.38%
2200	1,128,582	4,259,092	163.40	1992-03-25	2021-01-31	73.05%
2220	1,129,293	4,246,703	8.97	2005-01-01	2021-01-31	100%
2230	1,116,979	4,254,000	309.00	1992-03-25	2021-01-31	72.52%
2270	1,121,600	4,240,769	406.67	1992-03-24	2021-01-31	72.53%
2290	1,118,187	4,229,187	456.32	1992-03-24	2021-01-31	71.79%
2310	1,120,507	4,221,293	102.96	1988-06-10	2021-01-31	61.90%
2320	1,109,195	4,228,125	960.13	2005-01-01	2021-01-31	100%
2337	1,108,304	4,222,600	202.55	2014-05-01	2021-01-31	100%

(*continued on next page*)

Table A.1 (*continued*)

ID	Coordinates (EPSG: 32632)		Altitude (m a.s.l.)	Daily average time series		Missing data for 1970–2000
	longitude	latitude		start	end	
2340	1,104,962	4,233,682	972.76	1992-04-10	2021-01-31	71.84%
2380	1,091,178	4,225,861	708.85	2001-12-04	2021-01-31	100%
2417	1,096,637	4,235,553	1124.54	2014-05-01	2021-01-31	100%
2450	1,083,389	4,238,723	67.11	1924-01-01	2021-01-31	14.09%
2457	1,092,630	4,230,294	902.05	2014-05-01	2021-01-31	100%
2460	1,091,155	4,243,293	758.06	2014-05-01	2021-01-31	100%
2465	1,096,479	4,239,176	1238.84	1999-05-08	2021-01-31	94.94%
2466	1,092,099	4,246,248	547.11	2004-10-14	2021-01-31	100%
2467	1,086,702	4,240,290	387.55	2014-05-01	2021-01-31	100%
2470	1,097,587	4,247,042	1157.10	1939-04-26	2021-01-31	42.71%
2487	1,087,578	4,243,767	551.72	2014-05-01	2021-01-31	100%
2495	1,081,864	4,247,930	21.45	2005-01-01	2021-01-31	100%
2510	1,086,756	4,255,067	68.64	1988-09-08	2021-01-31	60.66%
2514	1,093,014	4,256,028	555.04	2005-11-01	2021-01-31	100%
2517	1,088,060	4,253,266	660.75	2014-05-01	2021-01-31	100%
2540	1,109,687	4,257,321	454.58	1988-07-15	2021-01-31	61.55%
2600	1,119,769	4,268,599	484.03	1924-01-01	2021-01-31	10.10%
2610	1,109,031	4,274,130	85.70	2005-01-01	2021-01-31	100%
2690	1,116,977	4,281,494	175.47	1988-10-01	2021-01-31	61.01%
2710	1,128,374	4,271,555	701.19	1992-03-26	2021-01-31	71.97%
2740	1,109,681	4,284,554	25.31	1940-05-16	2021-01-31	46.06%

Table A.2
Proportions of overlap between the study area's distribution of PCA scores and the sampling distribution of the weather stations, and proportions of variance (i.e., landscape complexity) explained by the PC axes.

		PC 1	PC 2	PC 3	PC 4	PC 5
January	overlap	84.74%	79.05%	73.02%	59.50%	64.51%
	variance	35.39%	19.94%	19.25%	19.19%	6.22%
February	overlap	85.51%	93.25%	78.50%	87.29%	63.33%
	variance	35.54%	19.48%	19.37%	19.33%	6.26%
March	overlap	83.97%	83.18%	71.61%	71.69%	63.02%
	variance	35.90%	19.72%	19.65%	18.39%	6.34%
April	overlap	83.94%	83.37%	71.36%	73.47%	63.02%
	variance	36.05%	19.97%	19.90%	17.68%	6.40%
May	overlap	84.49%	83.49%	71.35%	74.88%	62.50%
	variance	36.06%	19.99%	19.93%	17.63%	6.39%
June	overlap	84.28%	83.49%	71.33%	74.13%	62.08%
	variance	36.17%	20.28%	20.21%	16.86%	6.48%
July	overlap	84.36%	83.49%	71.23%	74.66%	62.53%
	variance	36.04%	20.05%	19.98%	17.53%	6.40%
August	overlap	84.25%	83.43%	71.42%	74.12%	62.45%
	variance	36.10%	20.02%	19.95%	17.52%	6.41%
September	overlap	83.83%	83.27%	71.26%	72.32%	63.21%
	variance	36.00%	19.89%	19.82%	17.90%	6.38%
October	overlap	85.69%	83.57%	71.74%	69.60%	62.97%
	variance	35.69%	19.61%	19.54%	18.85%	6.32%
November	overlap	85.03%	80.18%	73.05%	58.61%	63.98%
	variance	35.41%	19.89%	19.27%	19.21%	6.23%
December	overlap	83.83%	81.97%	70.30%	56.78%	64.68%
	variance	35.43%	20.48%	19.00%	18.94%	6.14%

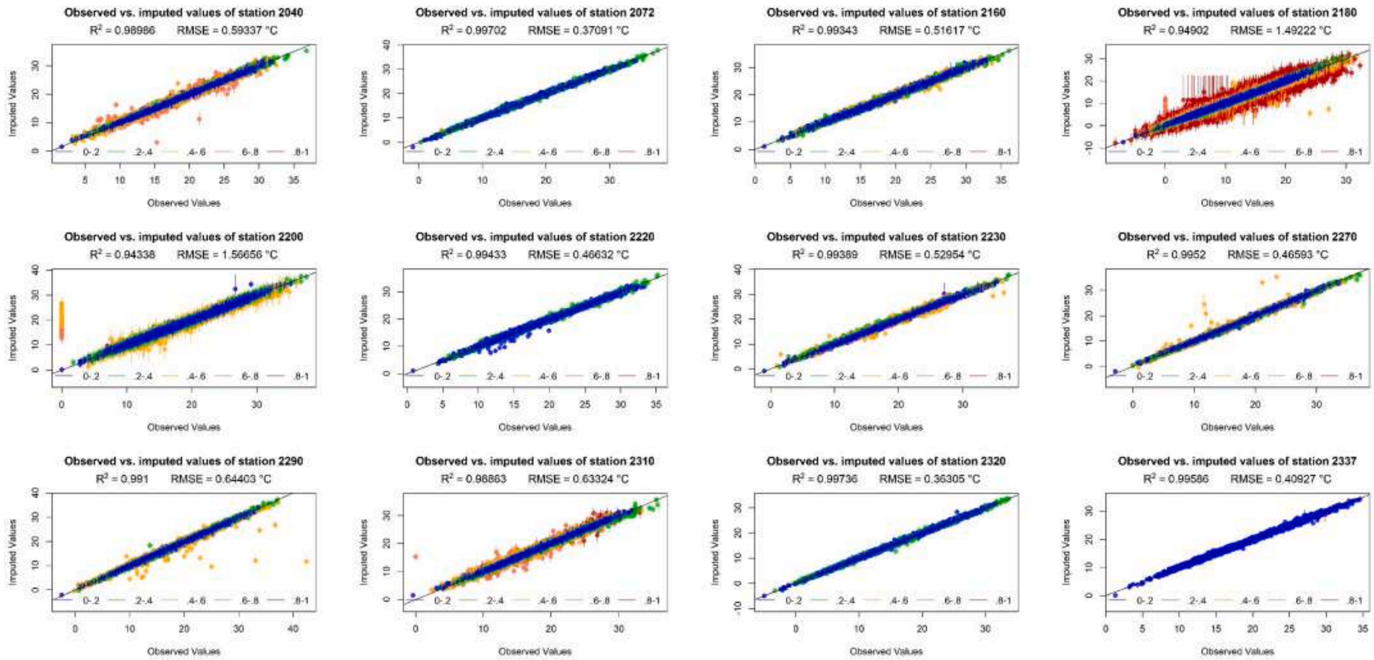

Fig. A.1. Overimputation diagnostic plots for stations 2040 to 2337. For each weather station, we treated each observed temperature as missing and drew 20 imputed replicas for each: dots represent the mean of these replicas, while the vertical lines are their 90% confidence intervals. The diagonal line indicates perfect concordance between observed and imputed temperatures; the dots would all fall on the line if the imputations were perfect. For each observed temperature of a weather station, the color of dots and vertical lines displays the proportion of measurements that are missing from other weather stations; the legend reports this information in decimals. As expected, imputations get more precise as the proportion of missing measurements from other weather stations decreases. R^2 and RMSE values measure the degree of association between observed and imputed temperatures.

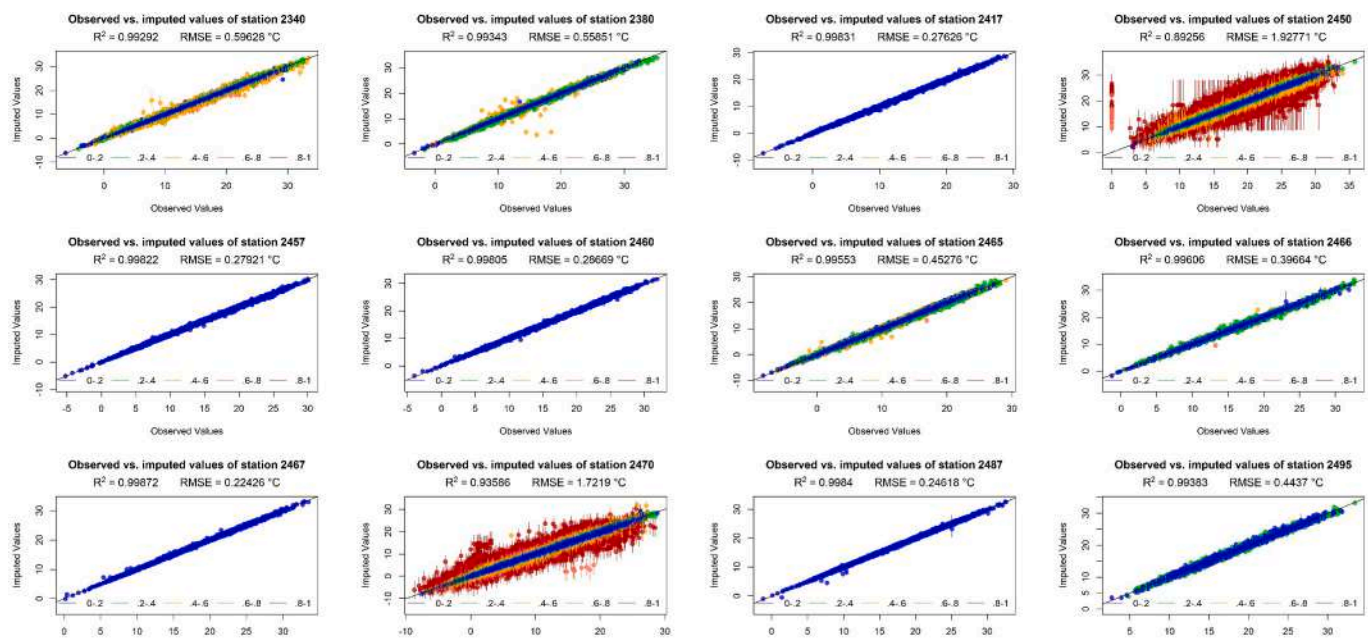

Fig. A.2. Overimputation diagnostic plots for stations 2340 to 2495. For each weather station, we treated each observed temperature as missing and drew 20 imputed replicas for each: dots represent the mean of these replicas, while the vertical lines are their 90% confidence intervals. The diagonal line indicates perfect concordance between observed and imputed temperatures; the dots would all fall on the line if the imputations were perfect. For each observed temperature of a weather station, the color of dots and vertical lines displays the proportion of measurements that are missing from other weather stations; the legend reports this information in decimals. As expected, imputations get more precise as the proportion of missing measurements from other weather stations decreases. R^2 and RMSE values measure the degree of association between observed and imputed temperatures.

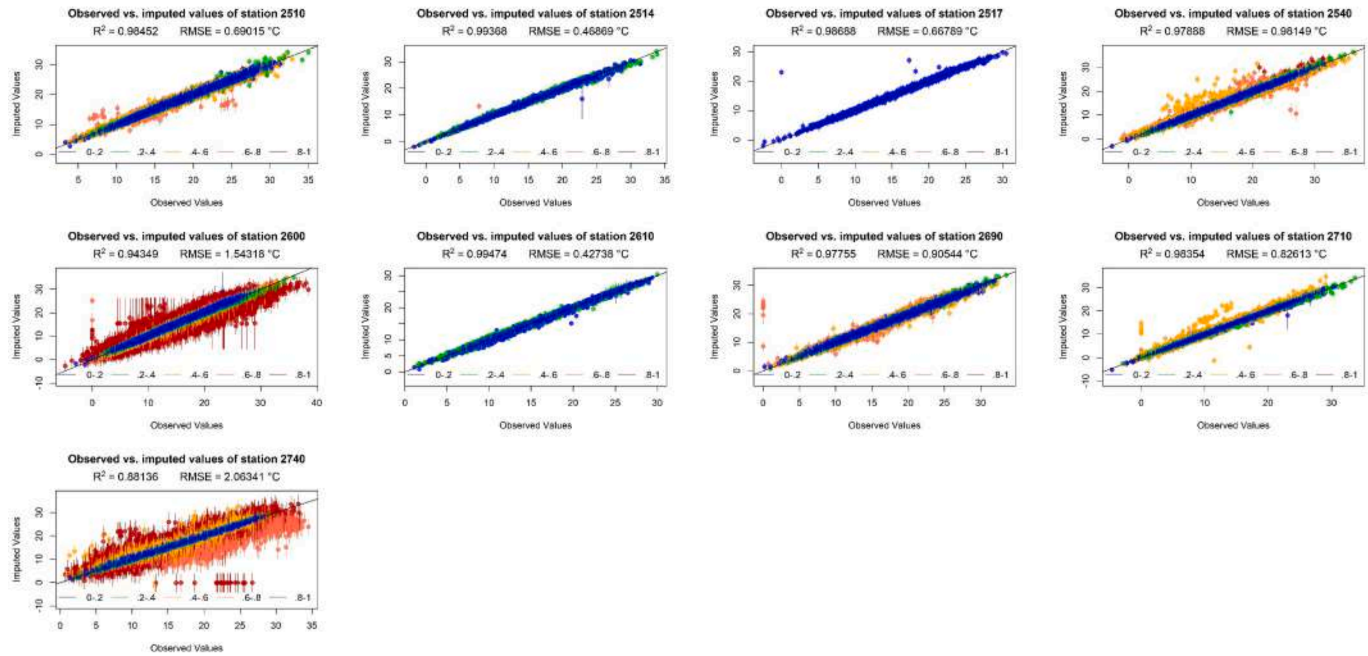

Fig. A.3. Overimputation diagnostic plots for stations 2510 to 2740. For each weather station, we treated each observed temperature as missing and drew 20 imputed replicas for each: dots represent the mean of these replicas, while the vertical lines are their 90% confidence intervals. The diagonal line indicates perfect concordance between observed and imputed temperatures; the dots would all fall on the line if the imputations were perfect. For each observed temperature of a weather station, the color of dots and vertical lines displays the proportion of measurements that are missing from other weather stations; the legend reports this information in decimals. As expected, imputations get more precise as the proportion of missing measurements from other weather stations decreases. R^2 and RMSE values measure the degree of association between observed and imputed temperatures.

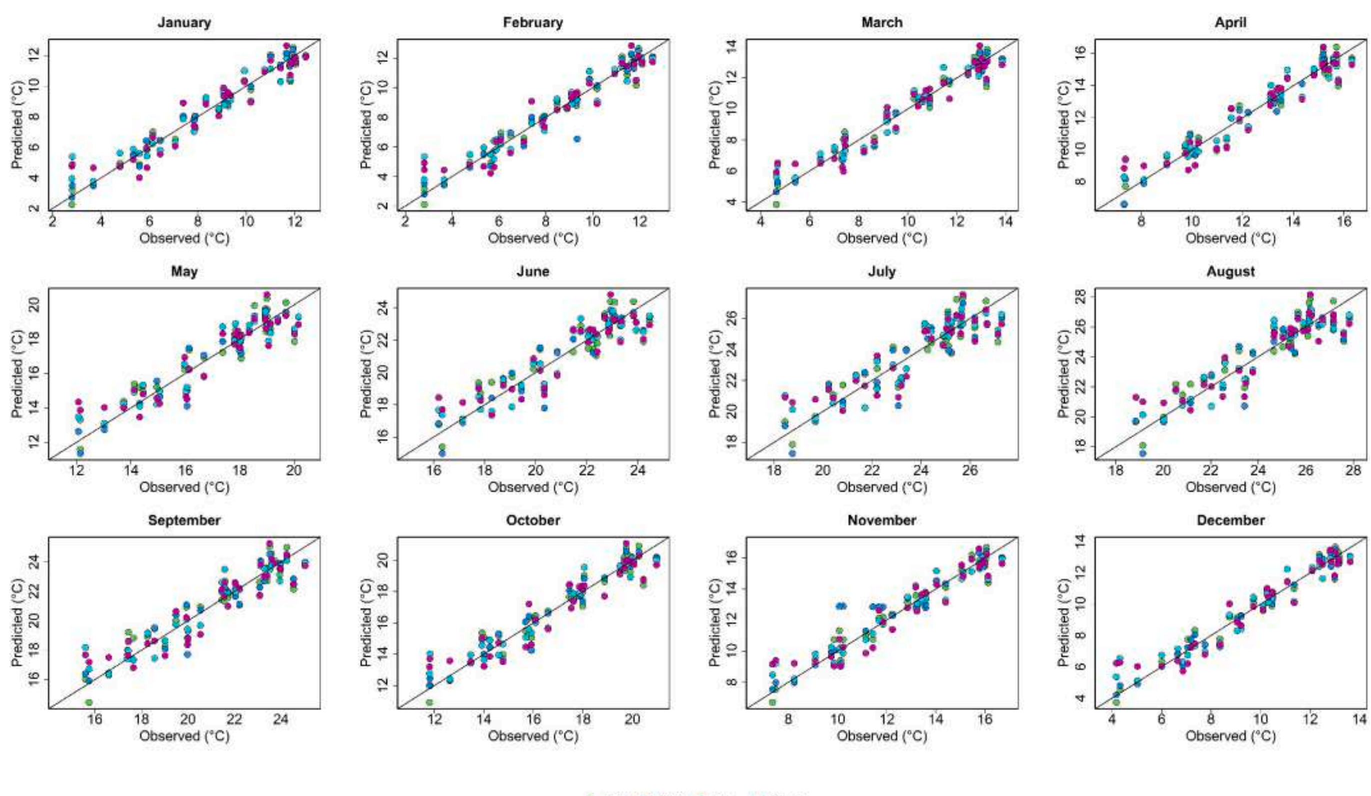

Fig. A.4. Observed monthly climates at the weather stations' locations and those predicted when interpolating weather station data.

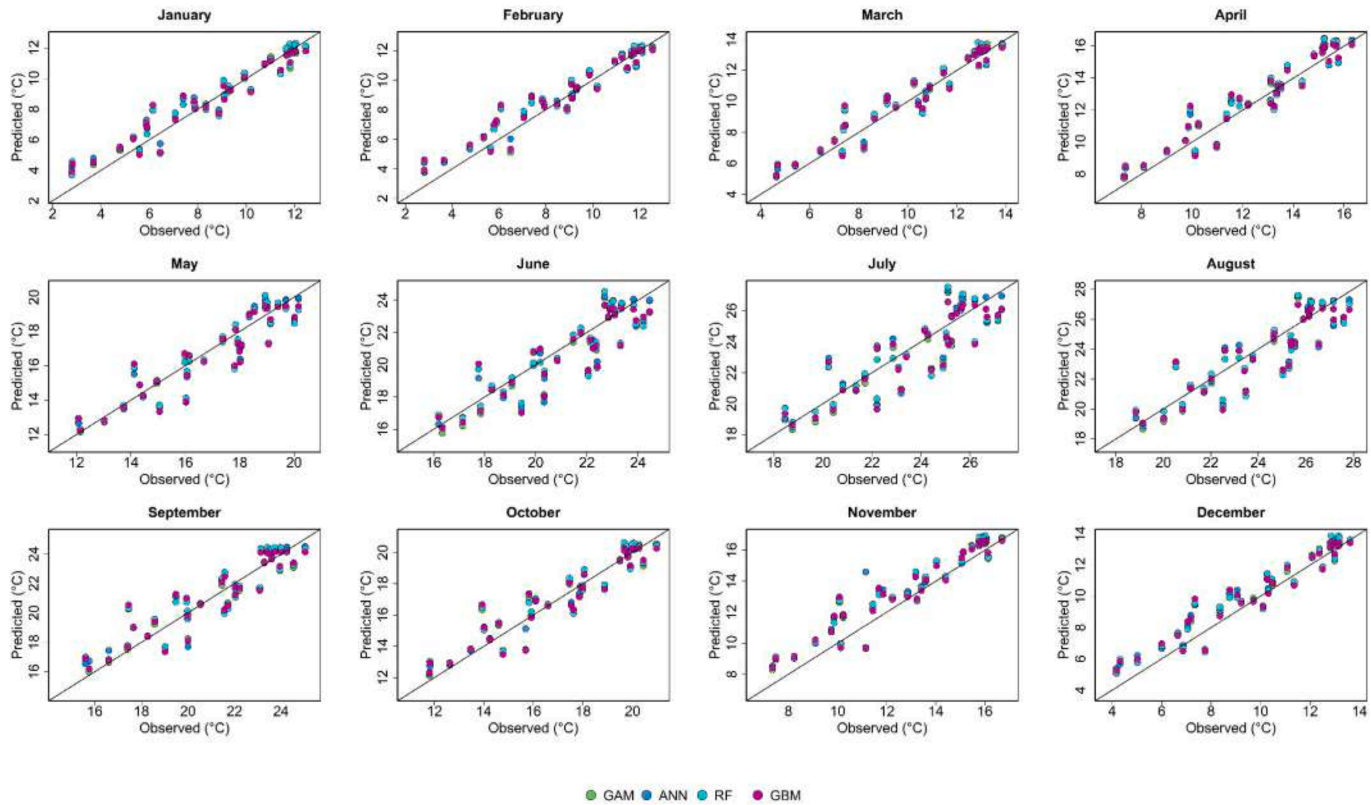

Fig. A.5. Observed monthly climates at the weather stations' locations and those predicted when downscaling WorldClim.

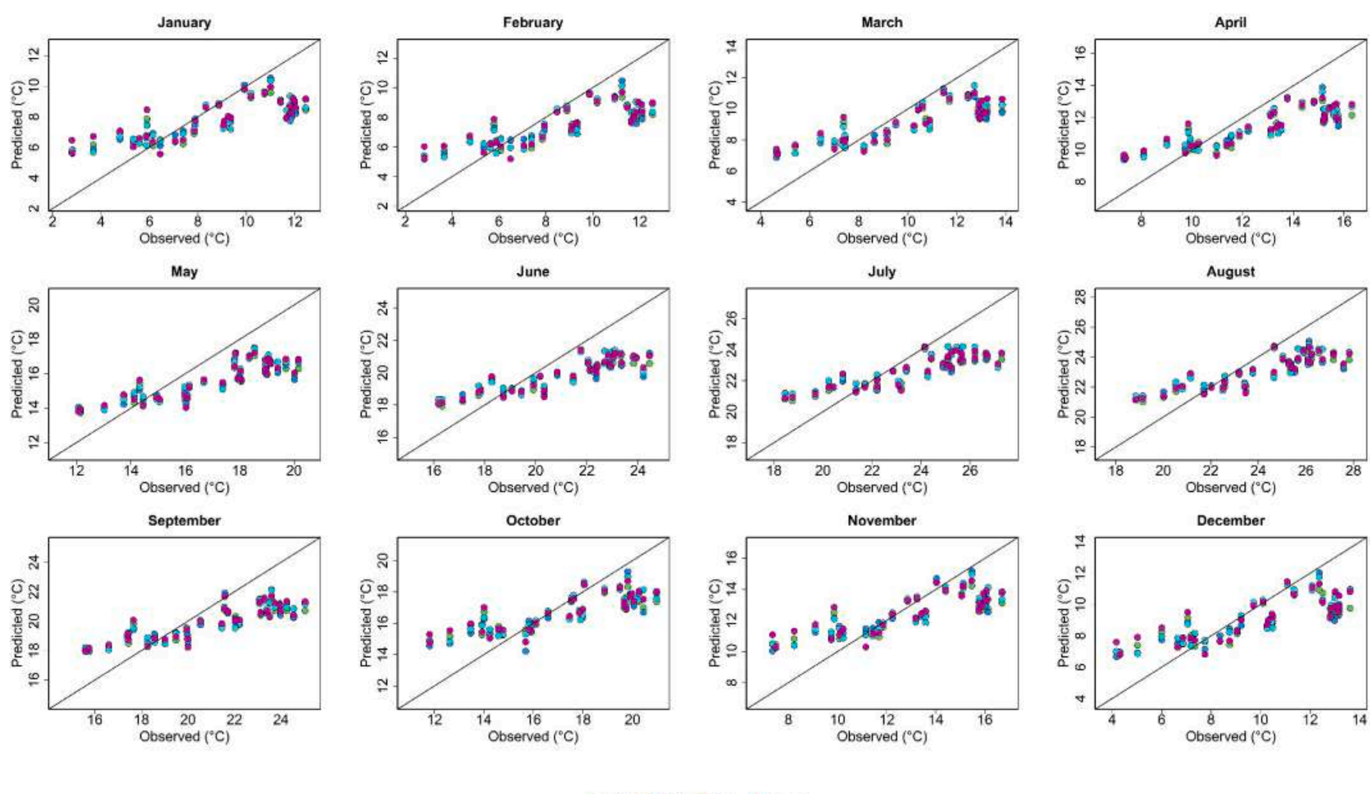

Fig. A.6. Observed monthly climates at the weather stations' locations and those predicted when downscaling CHELSA.

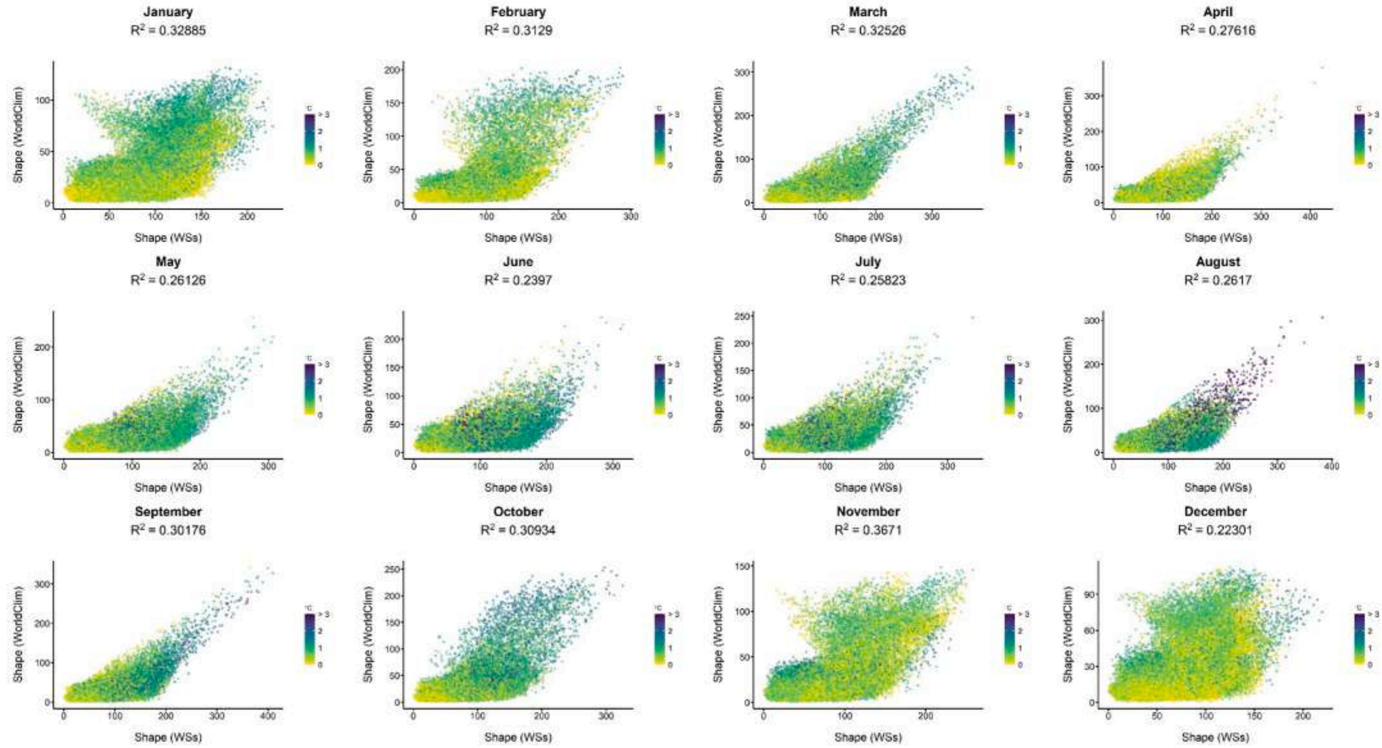

Fig. A.7. Coherence of monthly climate estimates when interpolating weather station (WS) data and downscaling WorldClim, with respect to the degree of environmental extrapolation. We quantified environmental extrapolation by calculating the Shape metric at 100,000 random points; higher Shape values indicate higher extrapolation degrees. Color gradients indicate the absolute differences between predictions under WSs interpolation and WorldClim downscaling at those 100,000 points; lighter colors indicate stronger coherence of monthly climate estimates. We capped temperature absolute differences at 3 °C for graphical purposes. R^2 values measure the degree of association between the Shape values of WSs and WorldClim.

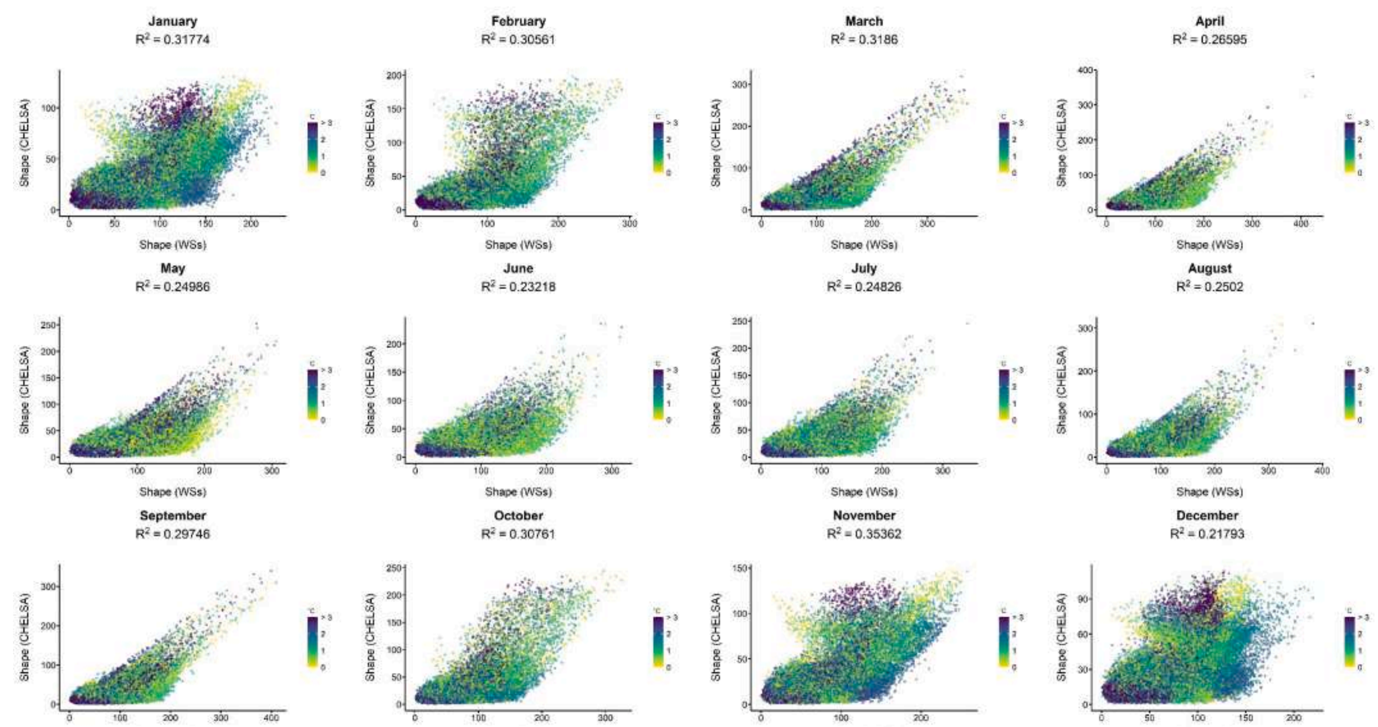

Fig. A.8. Coherence of monthly climate estimates when interpolating weather station (WS) data and downscaling CHELSA, with respect to the degree of environmental extrapolation. We quantified environmental extrapolation by calculating the Shape metric at 100,000 random points; higher Shape values indicate higher extrapolation degrees. Color gradients indicate the absolute differences between predictions under WSs interpolation and CHELSA downscaling at those 100,000 points; lighter colors indicate stronger coherence of monthly climate estimates. We capped temperature absolute differences at 3 °C for graphical purposes. R^2 values measure the degree of association between the Shape values of WSs and CHELSA.

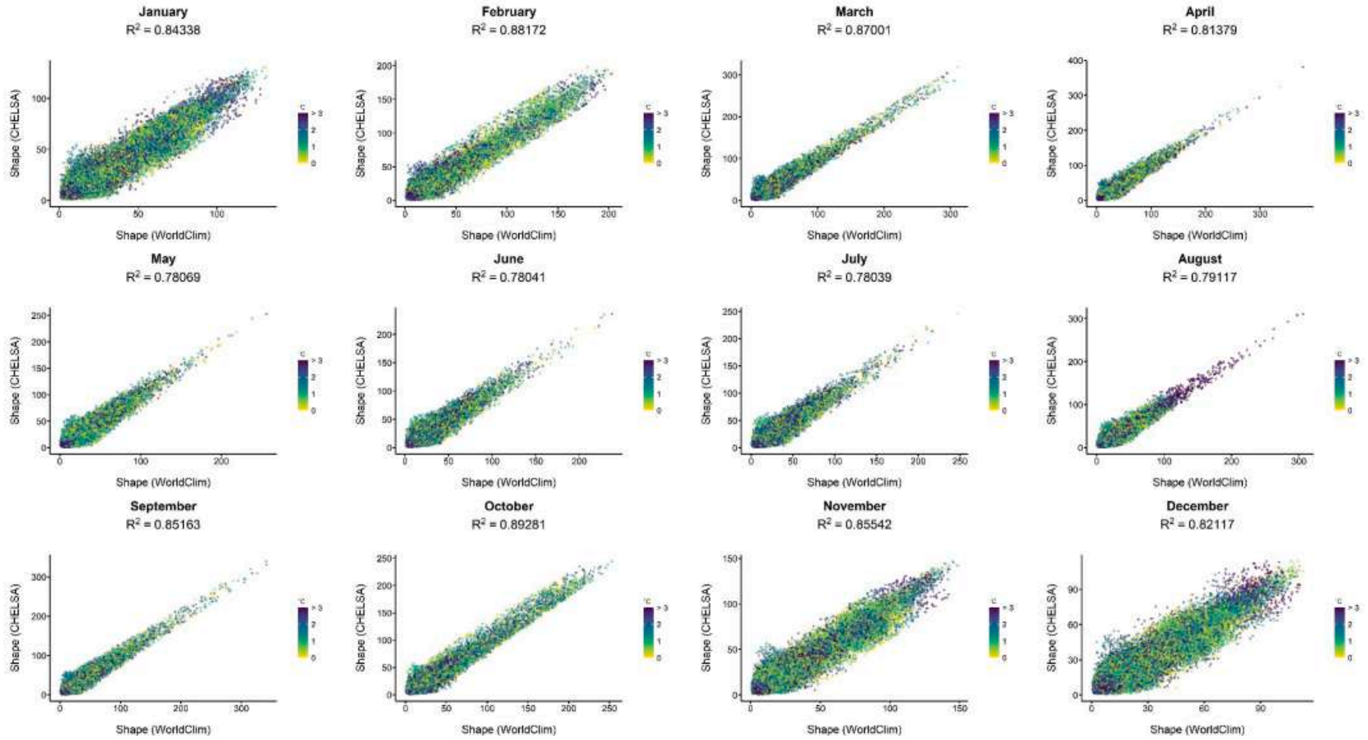

Fig. A.9. Coherence of monthly climate estimates when downscaling WorldClim and CHELSA, with respect to the degree of environmental extrapolation. We quantified environmental extrapolation by calculating the Shape metric at 100,000 random points; higher Shape values indicate higher extrapolation degrees. Color gradients indicate the absolute differences between predictions under WorldClim and CHELSA downscaling at those 100,000 points; lighter colors indicate stronger coherence of monthly climate estimates. We capped temperature absolute differences at 3 °C for graphical purposes. R^2 values measure the degree of association between the Shape values of WorldClim and CHELSA.

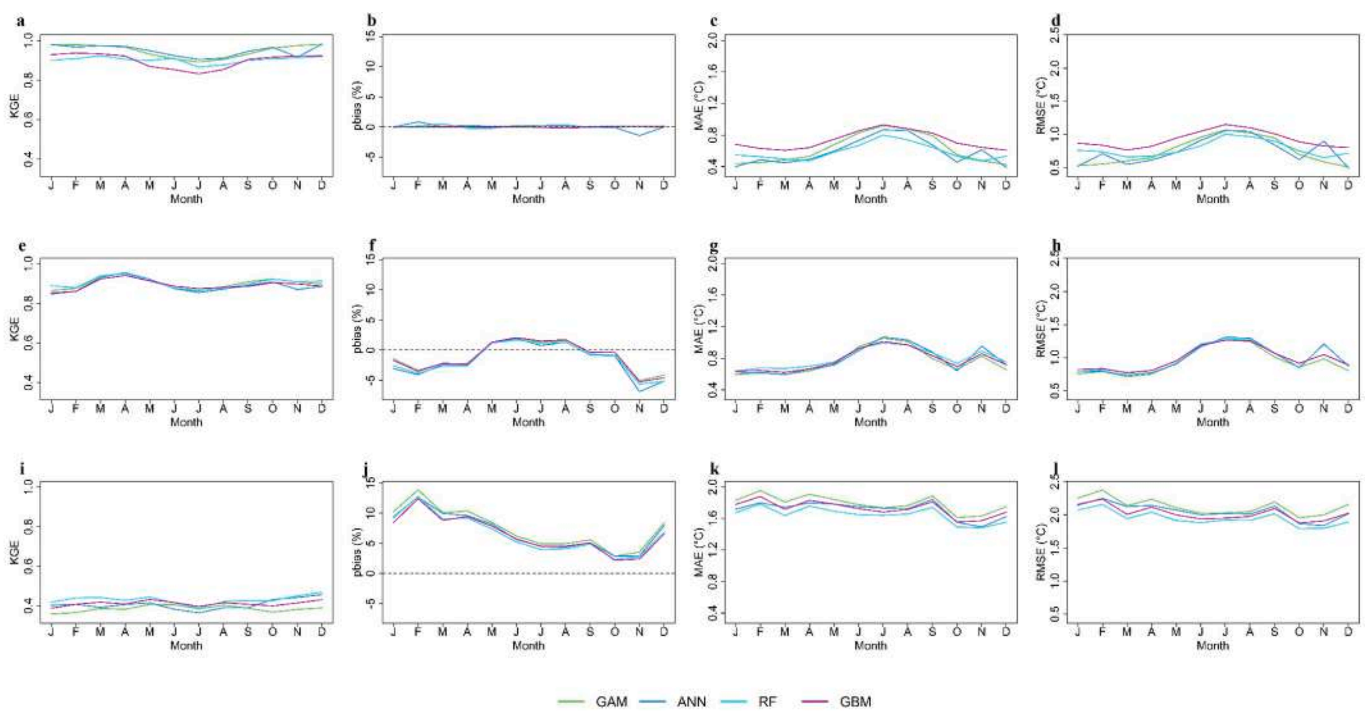

Fig. A.10. Seasonal patterns in KGE scores (first column), *pbias* (second column), MAE (third column), and RMSE (fourth column) that we obtained in the cross-validations when employing each algorithm to interpolate weather station data (top row) and downscale WorldClim (middle row) and CHELSA (bottom row). Model performance improves as KGE values approach 1; KGE values between 0 and 0.5 indicate poor model performance. Low *pbias* values reflect accurate model prediction; the optimal *pbias* value is 0. Model performance increases as MAE and RMSE values decrease.

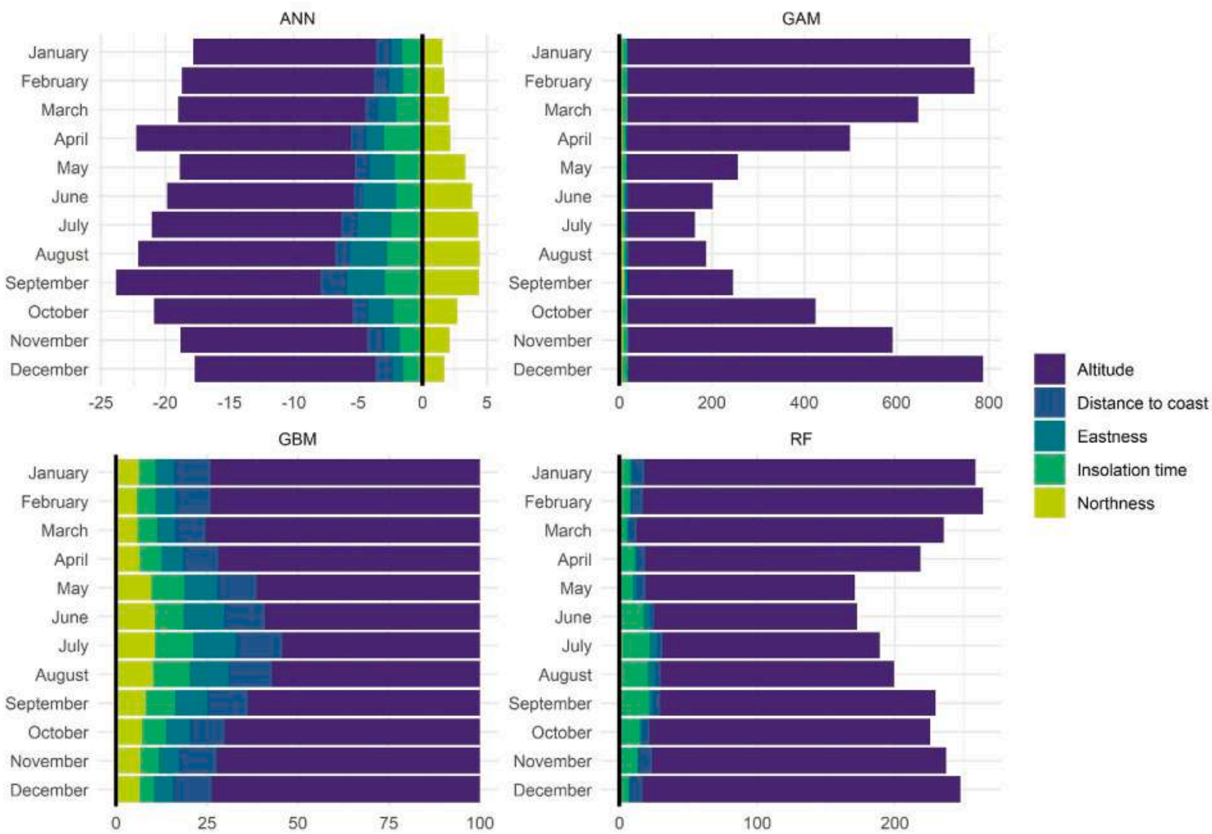

Fig. A.11. Variable importance plots for the interpolation of weather station data. ANN's top-left subplot displays the Olden relative variable importance, GAM's top-right subplot the F statistic for each predictor, GBM's bottom-left subplot the computed relative influences, normalized to sum to 100, and RF's bottom-right subplot the increase in node purity associated with each predictor.

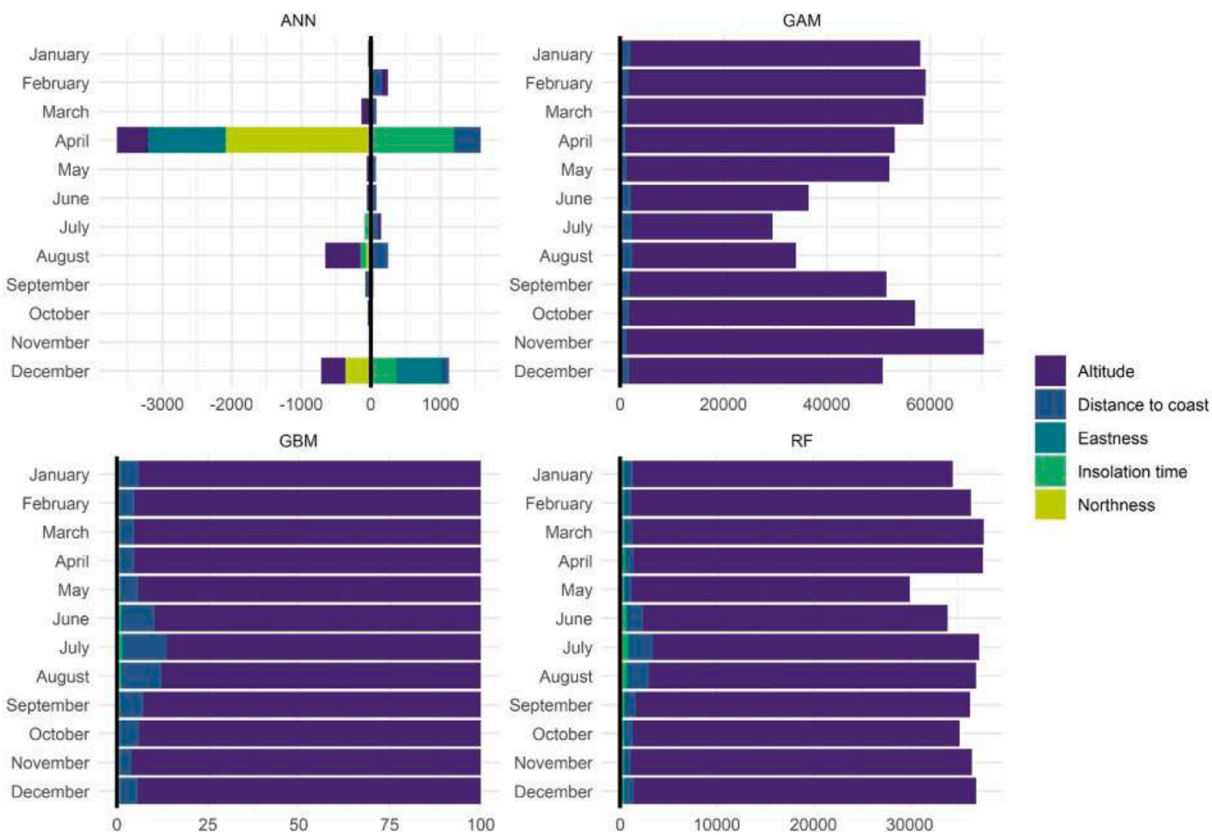

Fig. A.12. Variable importance plots for the WorldClim downscaling. ANN's top-left subplot displays the Olden relative variable importance, GAM's top-right subplot the F statistic for each predictor, GBM's bottom-left subplot the computed relative influences, normalized to sum to 100, and RF's bottom-right subplot the increase in node purity associated with each predictor.

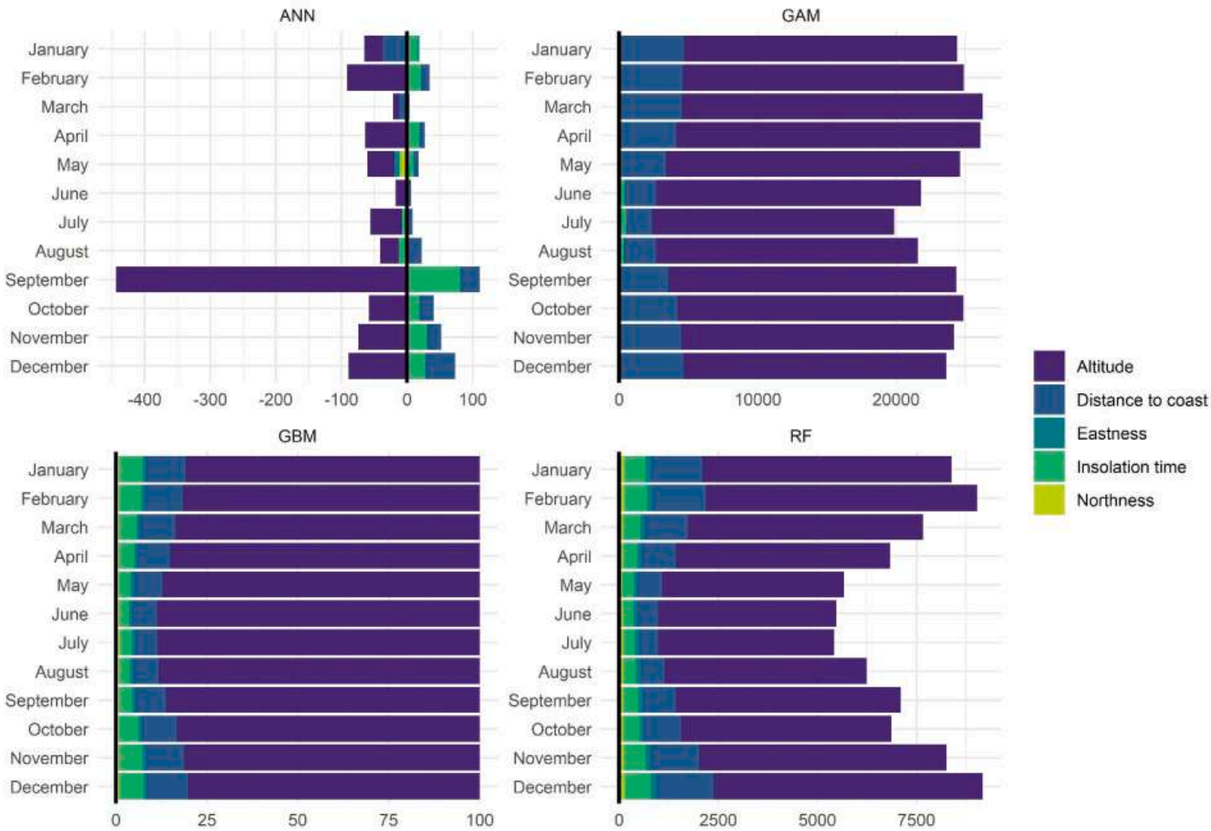

Fig. A.13. Variable importance plots for the CHELSA downscaling. ANN's top-left subplot displays the Olden relative variable importance, GAM's top-right subplot the F statistic for each predictor, GBM's bottom-left subplot the computed relative influences, normalized to sum to 100, and RF's bottom-right subplot the increase in node purity associated with each predictor.

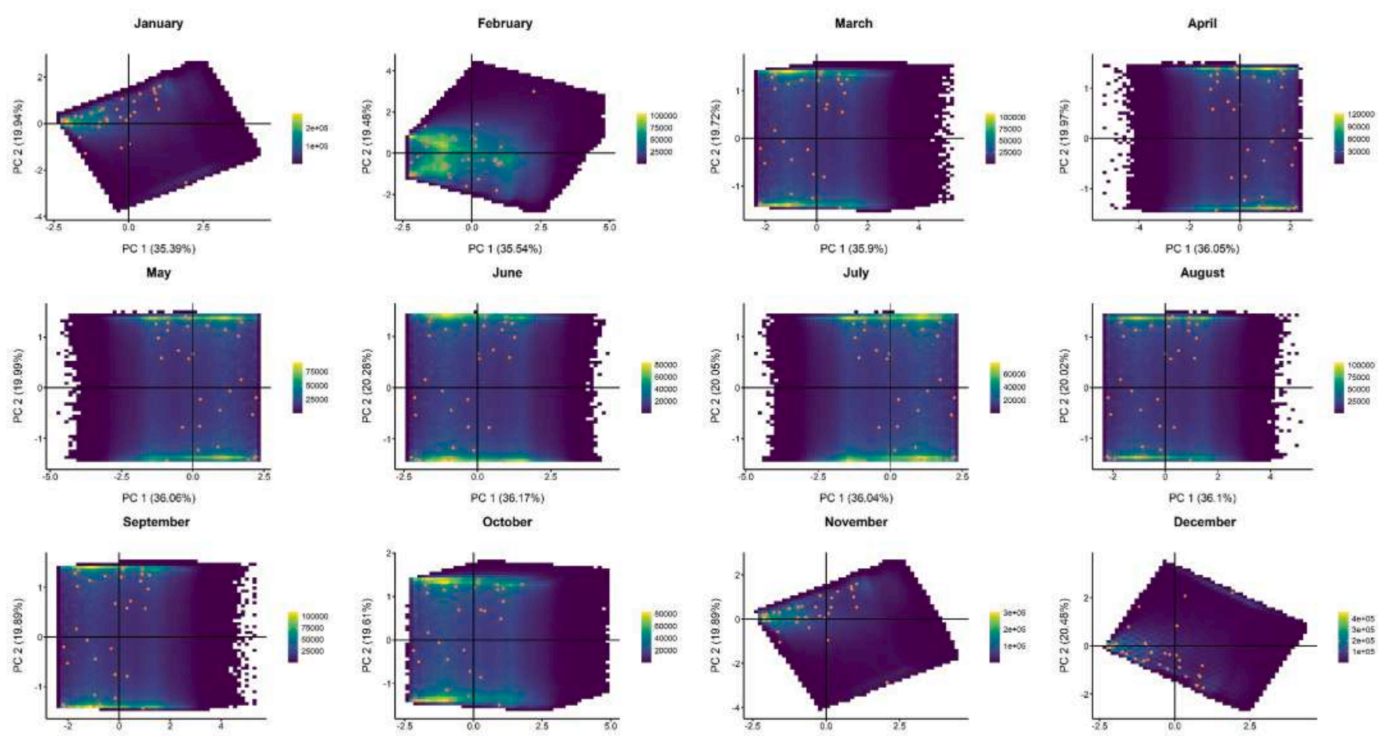

Fig. A.14. Monthly bivariate plots of the PCAs that synthesize landscape complexity. Orange points indicate weather stations' locations.

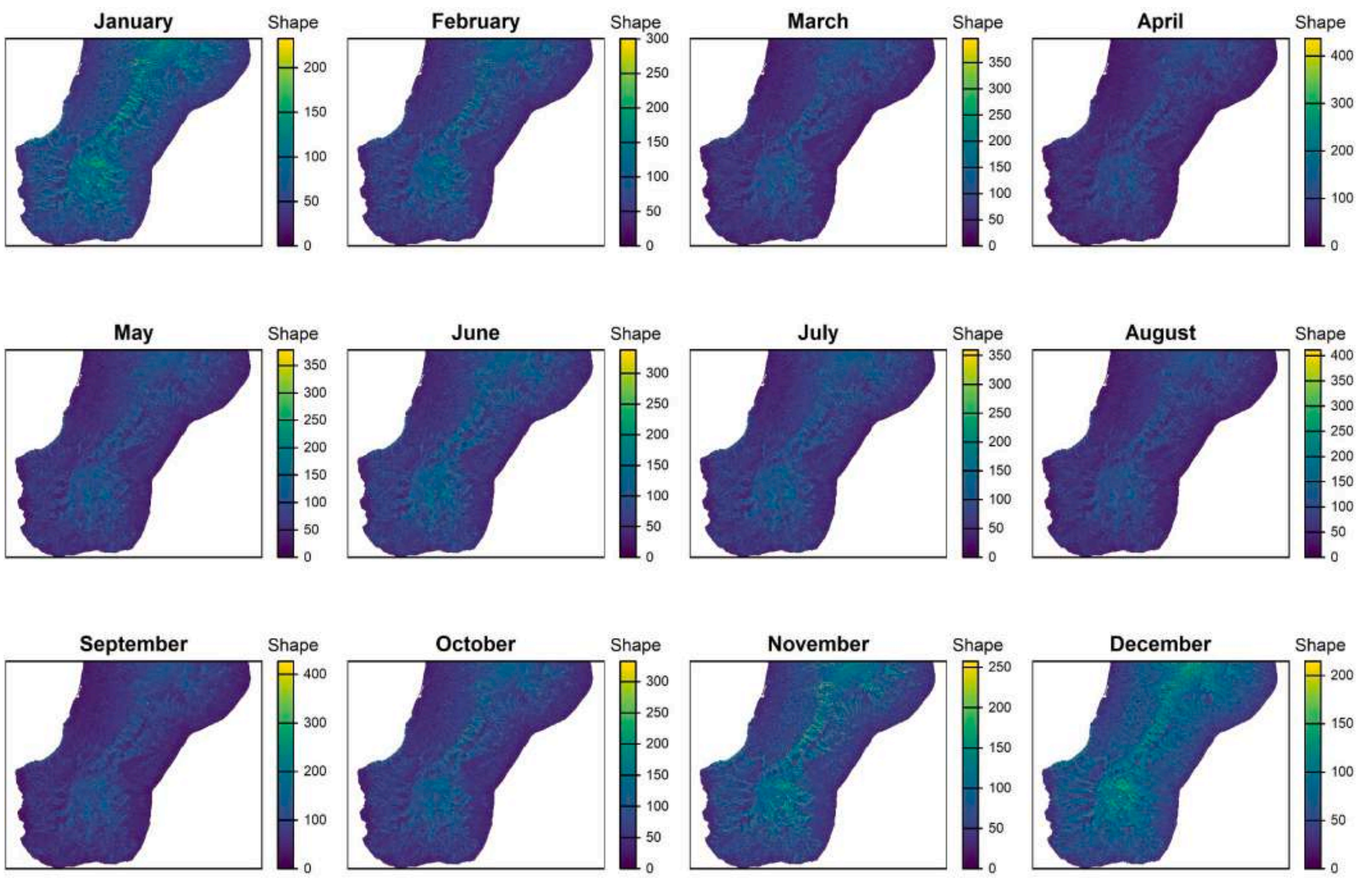

Fig. A.15. Seasonal patterns of environmental extrapolation when interpolating weather station data. Extrapolation was quantified by the Shape method; higher Shape values indicate higher extrapolation degrees.

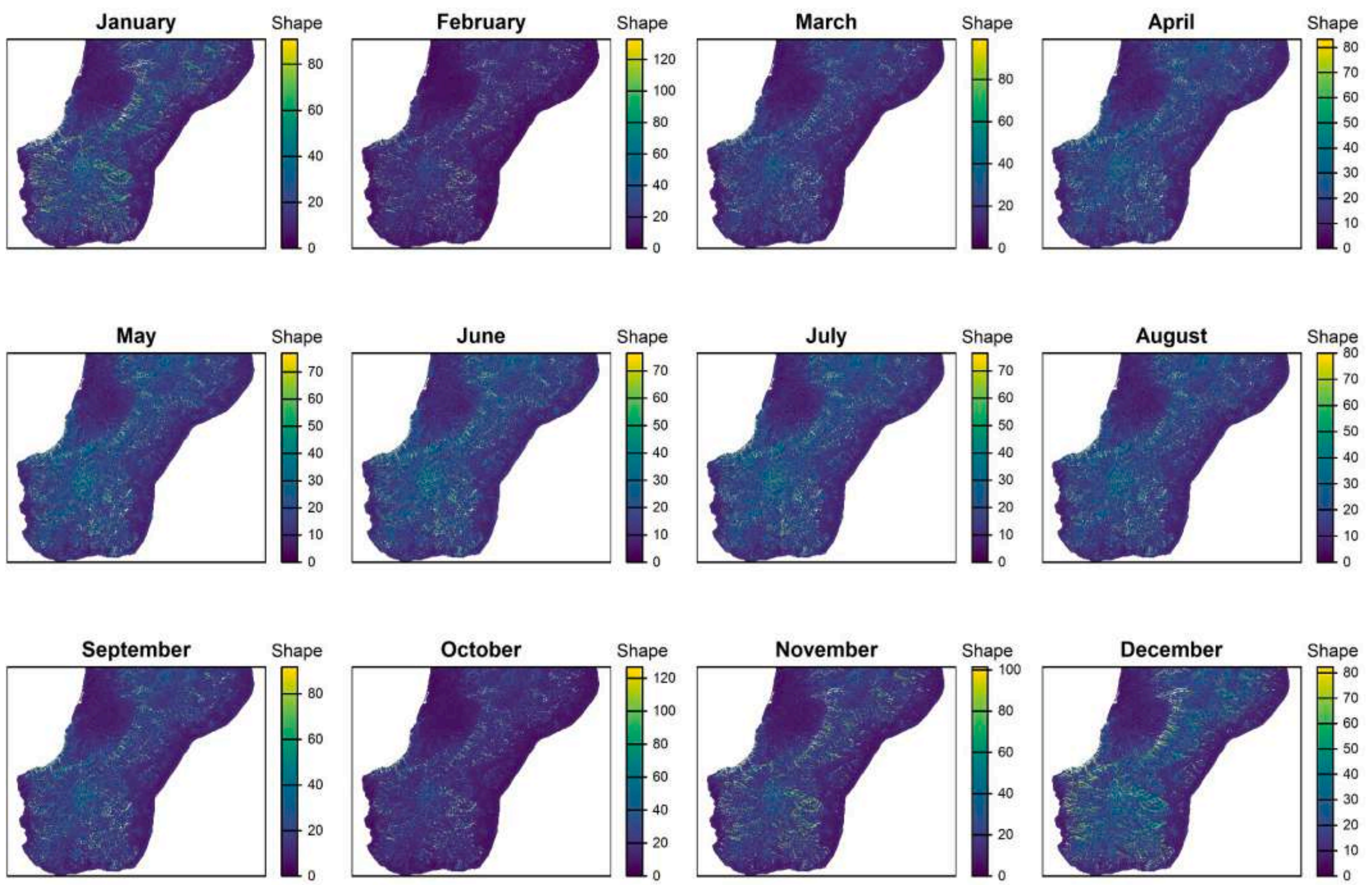

Fig. A.16. Seasonal patterns of environmental extrapolation when downscaling WorldClim. Extrapolation was quantified by the Shape method; higher Shape values indicate higher extrapolation degrees. For graphical purposes, Shape values were capped at their 99th percentile.

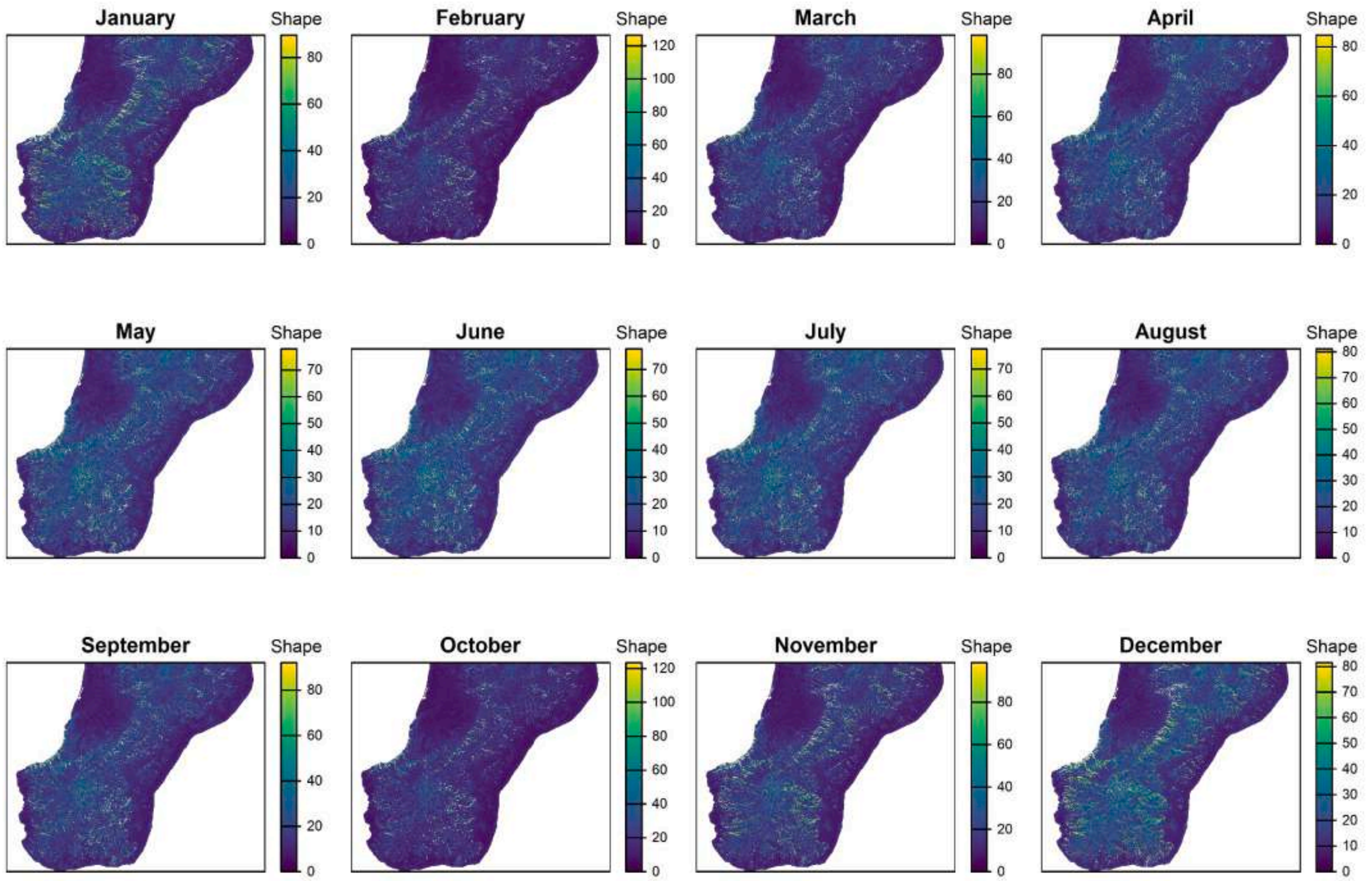

Fig. A.17. Seasonal patterns of environmental extrapolation when downscaling CHELSA. Extrapolation was quantified by the Shape method; higher Shape values indicate higher extrapolation degrees. For graphical purposes, Shape values were capped at their 99th percentile.

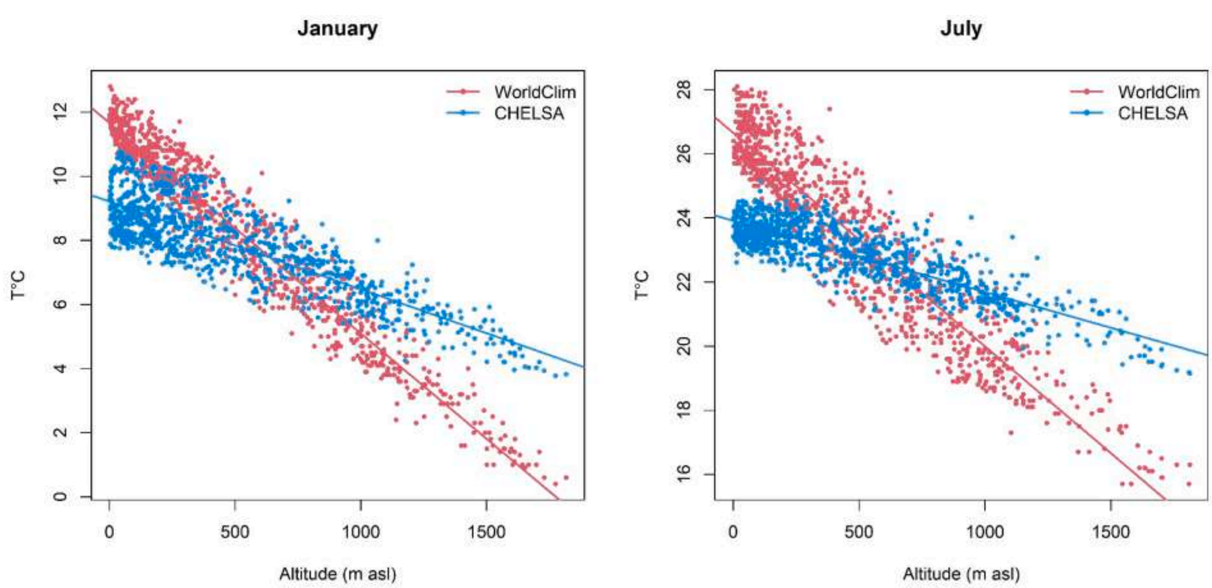

Fig. A.18. Empirical temperature lapse rates of the native WorldClim and CHELSA macroclimate grids for January and July. Points indicate the altitude and coarse-grained temperature at 1000 random locations, while lines show the corresponding trend curves outputted by linear regression models.

References

Aalto, J., Riihimäki, H., Meineri, E., Hylander, K., Luoto, M., 2017. Revealing topoclimatic heterogeneity using meteorological station data. Int. J. Climatol. 37, 544–556.

Ashcroft, M.B., 2010. Identifying refugia from climate change. J. Biogeogr. 37 (8), 1407–1413.

Ashcroft, M.B., Gollan, J.R., 2012. Fine-resolution (25 m) topoclimatic grids of near-surface (5 cm) extreme temperatures and humidities across various habitats in a large (200× 300 km) and diverse region. Int. J. Climatol. 32 (14), 2134–2148.

Barry, R.G., Blanken, P.D., 2016. Microclimate and local climate. Cambridge University Press.

Beck, M.W., 2018. NeuralNetTools: visualization and analysis tools for neural networks. J. Stat. Softw. 85 (11), 1.
Bolstad, P.V., Swift, L., Collins, F., Régnière, J., 1998. Measured and predicted air temperatures at basin to regional scales in the southern Appalachian mountains. Agric. For. Meteorol. 91 (3–4), 161–176.
Brullo, S., Scelsi, F., Spampinato, G., 2001. La Vegetazione dell'Aspromonte. Studio fitosociologico. Laruffa editore. Reggio Calabria.
Caesar, J., Alexander, L., Vose, R., 2006. Large-scale changes in observed daily maximum and minimum temperatures: creation and analysis of a new gridded data set. J. Geophys. Res. Atmos. 111 (D5).
Camacho-Sanchez, M., Quintanilla, I., Hawkins, M.T., Tuh, F.Y., Wells, K., Maldonado, J. E., Leonard, J.A., 2018. Interglacial refugia on tropical mountains: novel insights from the summit rat (*Rattus baluensis*), a Borneo mountain endemic. Divers. Distrib. 24 (9), 1252–1266.
Canestrelli, D., Bisconti, R., Chiocchio, A., Maiorano, L., Zampiglia, M., Nascetti, G., 2017. Climate change promotes hybridisation between deeply divergent species. PeerJ 5, e3072.
Chen, I.C., Hill, J.K., Ohlemüller, R., Roy, D.B., Thomas, C.D., 2011. Rapid range shifts of species associated with high levels of climate warming. Science 333 (6045), 1024–1026.
Chiocchio, A., Maiorano, L., Pezzarossa, A., Bisconti, R., Canestrelli, D., 2024. From the mountains to the sea: Rethinking Mediterranean glacial refugia as dynamic entities. Journal of Biogeography 51, 956–967.
Colacino, M., Conte, M., Piervitali, E., 1997. Elementi di climatologia della Calabria. IFA-CNR, Rome.
Daly, C., Neilson, R.P., Phillips, D.L., 1994. A statistical-topographic model for mapping climatological precipitation over mountainous terrain. J. Appl. Meteorol. Climatol. 33 (2), 140–158.
De Lombaerde, E., Vangansbeke, P., Lenoir, J., Van Meerbeek, K., Lembrechts, J., Rodríguez-Sánchez, F., De Frenne, P., 2022. Maintaining forest cover to enhance temperature buffering under future climate change. Sci. Total Environ. 810, 151338.
Dobrowski, S.Z., 2011. A climatic basis for microrefugia: the influence of terrain on climate. Glob. Chang. Biol. 17 (2), 1022–1035.
Drechsel, S., Mayr, G.J., 2008. Objective forecasting of foehn winds for a subgrid-scale alpine valley. Weather Forecast. 23 (2), 205–218.
Dujardin, J., Lehning, M., 2022. Wind-topo: downscaling near-surface wind fields to high-resolution topography in highly complex terrain with deep learning. Q. J. R. Meteorol. Soc. 148 (744), 1368–1388.
Elith, J., Leathwick, J.R., 2009. Species distribution models: ecological explanation and prediction across space and time. Annu. Rev. Ecol. Evol. Syst. 40 (1), 677–697.
Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol. 37 (12), 4302–4315.
Frey, S.J., Hadley, A.S., Johnson, S.L., Schulze, M., Jones, J.A., Betts, M.G., 2016. Spatial models reveal the microclimatic buffering capacity of old-growth forests. Sci. Adv. 2 (4), e1501392.
Fridley, J.D., 2009. Downscaling climate over complex terrain: high finescale (< 1000 m) spatial variation of near-ground temperatures in a montane forested landscape (Great Smoky Mountains). J. Appl. Meteorol. Climatol. 48 (5), 1033–1049.
Geiger, R., Aron, R.H., Todhunter, P., 2009. The Climate near the Ground. Rowman & Littlefield.
Germain, S.J., Lutz, J.A., 2020. Climate extremes may be more important than climate means when predicting species range shifts. Clim. Chang. 163 (1), 579–598.
Greenwell, B., Boehmke, B., Cunningham, J., GBM Developers, 2020. gbm: Generalized Boosted Regression Models. R Package Version 2.1.8.
Gudiksen, P.H., Leone, J.M., King, C.W., Ruffieux, D., Neff, W.D., 1992. Measurements and modeling of the effects of ambient meteorology on nocturnal drainage flows. J. Appl. Meteorol. Climatol. 31 (9), 1023–1032.
Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecol. Model. 135 (2–3), 147–186.
Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. J. Hydrol. 377 (1–2), 80–91.
Ha, K.J., Shin, S.H., Mahrt, L., 2009. Spatial variation of the regional wind field with land–sea contrasts and complex topography. J. Appl. Meteorol. Climatol. 48 (9), 1929–1939.
Haesen, S., Lembrechts, J.J., De Frenne, P., Lenoir, J., Aalto, J., Ashcroft, M.B., Van Meerbeek, K., 2021. ForestTemp–sub-canopy microclimate temperatures of European forests. Glob. Chang. Biol. 27 (23), 6307–6319.
Hampe, A., Jump, A.S., 2011. Climate relicts: past, present, future. Annu. Rev. Ecol. Evol. Syst. 42, 313–333.
Honaker, J., King, G., Blackwell, M., 2011. Amelia II: a program for missing data. J. Stat. Softw. 45, 1–47.
Hopkinson, R.F., Hutchinson, M.F., McKenney, D.W., Milewska, E.J., Papadopol, P., 2012. Optimizing input data for gridding climate normals for Canada. J. Appl. Meteorol. Climatol. 51 (8), 1508–1518.
Hopper, S.D., 2009. OCBIL theory: towards an integrated understanding of the evolution, ecology and conservation of biodiversity on old, climatically buffered, infertile landscapes. Plant Soil 322 (1), 49–86.
Joly, D., Castel, T., Pohl, B., Richard, Y., 2018. Influence of spatial information resolution on the relation between elevation and temperature. Int. J. Climatol. 38 (15), 5677–5688.
Jones, P.D., Briffa, K.R., 1992. Global surface air temperature variations during the twentieth century: part 1, spatial, temporal and seasonal details. The Holocene 2 (2), 165–179.
Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Kessler, M., 2017. Climatologies at high resolution for the Earth's land surface areas. Sci. Data 4 (1), 1–20.
Karger, D.N., Saladin, B., Wüest, R.O., Graham, C.H., Zurell, D., Mo, L., Zimmermann, N. E., 2023. Interannual climate variability improves niche estimates for ectothermic but not endothermic species. Sci. Rep. 13 (1), 12538.
Kearney, M.R., Porter, W.P., 2017. NicheMapR – an R package for biophysical modelling: the microclimate model. Ecography 40 (5), 664–674.
Lembrechts, J.J., Lenoir, J., Roth, N., Hattab, T., Milbau, A., Haider, S., Nijs, I., 2019. Comparing temperature data sources for use in species distribution models: from in-situ logging to remote sensing. Glob. Ecol. Biogeogr. 28 (11), 1578–1596.
Lenoir, J., Graae, B.J., Aarrestad, P.A., Alsos, I.G., Armbruster, W.S., Austrheim, G., Svenning, J.C., 2013. Local temperatures inferred from plant communities suggest strong spatial buffering of climate warming across Northern Europe. Glob. Chang. Biol. 19 (5), 1470–1481.
Lenoir, J., Hattab, T., Pierre, G., 2017. Climatic microrefugia under anthropogenic climate change: implications for species redistribution. Ecography 40 (2), 253–266.
Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R News 2 (3), 18–22.
Lookingbill, T.R., Urban, D.L., 2003. Spatial estimation of air temperature differences for landscape-scale studies in montane environments. Agric. For. Meteorol. 114 (3–4), 141–151.
Luoto, M., Heikkinen, R.K., 2008. Disregarding topographical heterogeneity biases species turnover assessments based on bioclimatic models. Glob. Chang. Biol. 14 (3), 483–494.
Maclean, I.M., Klinges, D.H., 2021. Microclimc: a mechanistic model of above, below and within-canopy microclimate. Ecol. Model. 451, 109567.
Maclean, I.M., Mosedale, J.R., Bennie, J.J., 2019. Microclima: an r package for modelling meso-and microclimate. Methods Ecol. Evol. 10 (2), 280–290.
Martino, G., Chiocchio, A., Siclari, A., Canestrelli, D., 2022. Distribution and conservation status of threatened endemic amphibians within the Aspromonte mountain region, a hotspot of Mediterranean biodiversity. Nat. Conserv. 50, 1–22.
Martin-Vide, J., 2004. Spatial distribution of a daily precipitation concentration index in peninsular Spain. Int. J. Climatol. 24 (8), 959–971.
McCullough, I.M., Davis, F.W., Dingman, J.R., Flint, L.E., Flint, A.L., Serra-Diaz, J.M., Franklin, J., 2016. High and dry: high elevations disproportionately exposed to regional climate change in Mediterranean-climate landscapes. Landsc. Ecol. 31 (5), 1063–1075.
Meineri, E., Hylander, K., 2017. Fine-grain, large-domain climate models based on climate station and comprehensive topographic information improve microrefugia detection. Ecography 40 (8), 1003–1013.
Milling, C.R., Rachlow, J.L., Olsoy, P.J., Chappell, M.A., Johnson, T.R., Forbey, J.S., Thornton, D.H., 2018. Habitat structure modifies microclimate: an approach for mapping fine-scale thermal refuge. Methods Ecol. Evol. 9 (6), 1648–1657.
Olden, J.D., Joy, M.K., Death, R.G., 2004. An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. Ecol. Model. 178 (3–4), 389–397.
Oliver, T., Roy, D.B., Hill, J.K., Brereton, T., Thomas, C.D., 2010. Heterogeneous landscapes promote population stability. Ecol. Lett. 13 (4), 473–484.
Parmesan, C., 2006. Ecological and evolutionary responses to recent climate change. Annu. Rev. Ecol. Evol. Syst. 637–669.
Parmesan, C., Yohe, G., 2003. A globally coherent fingerprint of climate change impacts across natural systems. Nature 421 (6918), 37–42.
Pastore, M., Di Loro, P.A., Mingione, M., Calcagni, A., 2022. Overlapping: Estimation of Overlapping in Empirical Distributions. R Package Version 2.1.
Pena-Angulo, D., Cortesi, N., Brunetti, M., González-Hidalgo, J.C., 2015. Spatial variability of maximum and minimum monthly temperature in Spain during 1981–2010 evaluated by correlation decay distance (CDD). Theor. Appl. Climatol. 122 (1), 35–45.
Pepin, N., Bradley, R.S., Diaz, H.F., Baraër, M., Caceres, E.B., Forsythe, N., Mountain Research Initiative EDW Working Group, 2015. Elevation-dependent warming in mountain regions of the world. Nat. Clim. Chang. 5 (5), 424–430.
Phillips, D.L., Dolph, J., Marks, D., 1992. A comparison of geostatistical procedures for spatial analysis of precipitation in mountainous terrain. Agric. For. Meteorol. 58 (1–2), 119–141.
Potter, K.A., Arthur Woods, H., Pincebourde, S., 2013. Microclimatic challenges in global change biology. Glob. Chang. Biol. 19 (10), 2932–2939.
R Core Team, 2022. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Randin, C.F., Engler, R., Normand, S., Zappa, M., Zimmermann, N.E., Pearman, P.B., Guisan, A., 2009. Climate change and plant distribution: local models predict high-elevation persistence. Glob. Chang. Biol. 15 (6), 1557–1569.
Rita, A., Bonanomi, G., Allevato, E., Borghetti, M., Cesarano, G., Mogavero, V., Saracino, A., 2021. Topography modulates near-ground microclimate in the Mediterranean *Fagus sylvatica* treeline. Sci. Rep. 11 (1), 8122.
Rogelis, M.C., Werner, M., Obregón, N., Wright, N., 2016. Hydrological model assessment for flood early warning in a tropical high mountain basin. Hydrol. Earth Syst. Sci. Discuss. 1–36.
Rosenzweig, C., Karoly, D., Vicarelli, M., Neofotis, P., Wu, Q., Casassa, G., Imeson, A., 2008. Attributing physical and biological impacts to anthropogenic climate change. Nature 453 (7193), 353–357.
Rull, V., 2009. Microrefugia. J. Biogeogr. 36 (3), 481–484.
Sandel, B., Arge, L., Dalsgaard, B., Davies, R.G., Gaston, K.J., Sutherland, W.J., Svenning, J.C., 2011. The influence of late quaternary climate-change velocity on species endemism. Science 334 (6056), 660–664.

Scherrer, D., Körner, C., 2010. Infra-red thermometry of alpine landscapes challenges climatic warming projections. Glob. Chang. Biol. 16 (9), 2602–2613.
Scherrer, D., Körner, C., 2011. Topographically controlled thermal-habitat differentiation buffers alpine plant diversity against climate warming. J. Biogeogr. 38 (2), 406–416.
Shoo, L.P., Storlie, C., Williams, Y.M., Williams, S.E., 2010. Potential for mountaintop boulder fields to buffer species against extreme heat stress under climate change. Int. J. Biometeorol. 54 (4), 475–478.
Slavich, E., Warton, D.I., Ashcroft, M.B., Gollan, J.R., Ramp, D., 2014. Topoclimate versus macroclimate: how does climate mapping methodology affect species distribution models and climate change projections? Divers. Distrib. 20 (8), 952–963.
Spampinato, G., Cameriere, P., Caridi, D., Crisafulli, A., 2008. Carta della biodiversità vegetale del Parco Nazionale dell'Aspromonte (Italia meridionale). Quaderni di Botanica Ambientale e Applicata 19, 3–36.
Srivastava, A.K., Rajeevan, M., Kshirsagar, S.R., 2009. Development of a high resolution daily gridded temperature data set (1969–2005) for the Indian region. Atmos. Sci. Lett. 10 (4), 249–254.
Stark, J.R., Fridley, J.D., 2022. Microclimate-based species distribution models in complex forested terrain indicate widespread cryptic refugia under climate change. Glob. Ecol. Biogeogr. 31 (3), 562–575.
Stewart, J.R., Lister, A.M., Barnes, I., Dalén, L., 2010. Refugia revisited: individualistic responses of species in space and time. Proc. R. Soc. B Biol. Sci. 277 (1682), 661–671.
Suggitt, A.J., Wilson, R.J., Isaac, N.J., Beale, C.M., Auffret, A.G., August, T., Maclean, I., 2018. Extinction risk from climate change is reduced by microclimatic buffering. Nat. Clim. Chang. 8 (8), 713–717.
Ummenhofer, C.C., Meehl, G.A., 2017. Extreme weather and climate events with ecological relevance: a review. Philos. Trans. R. Soc. B 372 (1723), 20160135.
Vanwalleghem, T., Meentemeyer, R.K., 2009. Predicting forest microclimate in heterogeneous landscapes. Ecosystems 12, 1158–1172.
Velazco, S.J.E., Rose, M.B., de Andrade, A.F.A., Minoli, I., Franklin, J., 2022. Flexsdm: an r package for supporting a comprehensive and flexible species distribution modelling workflow. Methods Ecol. Evol. 13 (8), 1661–1669.
Velazco, S.J.E., Rose, M.B., De Marco Jr, P., Regan, H.M., Franklin, J., 2023. How far can I extrapolate my species distribution model? Exploring shape, a novel method. Ecography e06992.
Venables, W.N., Ripley, B.D., 2002. Statistics and Computing. Modern Applied Statistics with S. Springers, New York, USA.
Vitasse, Y., Ursenbacher, S., Klein, G., Bohnenstengel, T., Chittaro, Y., Delestrade, A., Lenoir, J., 2021. Phenological and elevational shifts of plants, animals and fungi under climate change in the European Alps. Biol. Rev. 96 (5), 1816–1835.
Whiteman, C.D., Zhong, S., Shaw, W.J., Hubbe, J.M., Bian, X., Mittelstadt, J., 2001. Cold pools in the Columbia Basin. Weather Forecast. 16 (4), 432–447.
Wood, S.N., 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. J. R. Stat. Soc. Ser. B Stat Methodol. 73 (1), 3–36.
Zampiglia, M., Bisconti, R., Maiorano, L., Aloise, G., Siclari, A., Pellegrino, F., Canestrelli, D., 2019. Drilling down hotspots of intraspecific diversity to bring them into on-ground conservation of threatened species. Front. Ecol. Evol. 7, 205.
Zellweger, F., De Frenne, P., Lenoir, J., Rocchini, D., Coomes, D., 2019. Advances in microclimate ecology arising from remote sensing. Trends Ecol. Evol. 34 (4), 327–341.