RESEARCH ARTICLE

Interpolating climate variables by using INLA and the SPDE approach

Guido Fioravanti[1,2] 🄳 | **Sara Martino**[3] | **Michela Cameletti**[4] | **Andrea Toreti**[1]

[1]European Commission, Joint Research Centre, Ispra, Italy

[2]VAL-CLO, ISPRA, Rome, Italy

[3]Norwegian University of Science and Technology, Trondheim, Norway

[4]Department of Economics, University of Bergamo, Bergamo, Italy

Correspondence

Guido Fioravanti, European Commission, Joint Research Centre (JRC), Via E. Fermi 2749, 21027 Ispra, Italy.
Email: guido.fioravanti@ec.europa.eu

Abstract

Gridded observational products of the main climate parameters are essential in climate science. Current interpolation approaches, implemented to derive such products, often lack of a proper uncertainty propagation and representation. In this study, we introduce a Bayesian spatiotemporal approach based on the integrated nested Laplace approximation (INLA) and the stochastic partial differential equation (SPDE). The method is described and discussed by using a real case study based on high-resolution monthly 2-m maximum (Tmax) and minimum (Tmin) air temperature over Italy in 1961–2020. The INLA-SPDE based approach is able to properly take into account uncertainties in the final gridded products and offers interesting promising advantages to deal with non-stationary and non-Gaussian multisource data.

KEYWORDS

Bayesian approach, INLA, spatiotemporal correlation, SPDE, temperature climatology, uncertainty

1 | INTRODUCTION

Near-surface 2-m air temperature (hereafter Ta) provides essential information on the state of the climate system, its variability, extremes and its changes (Fang et al., 2022; Lian et al., 2017). It is easily understandable and it can be used for communicating effectively climate science and climate change to the general public. Thus, it should not come as a surprise that Ta plays an important role as a target of the Paris agreement and, more generally, for climate change related political processes (GCOS, 2017).

While the Sixth Assessment Report (AR6) of the IPCC (Masson-Delmotte et al., 2021) emphasizes that climate change will induce a general increase in the frequency and intensity of extreme events (such as heatwaves, heavy precipitation, droughts), it has to be noted that global warming is not homogeneous throughout the globe and some areas are seeing a faster rate of warming than others (Rantanen et al., 2022; Zittis et al., 2022). The Mediterranean area is a notable example of this fact (Tuel & Eltahir, 2020). Observational records and model-based studies clearly indicate the Mediterranean as a prominent climate change hot-spot, with a rate of warming which has accelerated particularly after the 1980s and is higher than the global average (Bruley et al., 2022; Cos et al., 2022; Lionello & Scarascia, 2018; Pastor et al., 2020; Zittis et al., 2019). Italy, located in the very middle of the Mediterranean basin, is particularly exposed to climate change impacts especially in the south where a considerable number of municipalities show low levels of resilience to disasters (Spano et al., 2020).

The collection and analysis of observational records from meteorological ground weather stations provides essential information for analysing past and current

trend, variability, as well as possible future climate scenarios. When interpolated into a regular grid, observational sparse data find use in a variety of scientific and operational applications such as statistical downscaling of Earth System Models and development of bias-adjusted climate change scenarios (Herrera et al., 2019; Tarek et al., 2021).

In recent decades, a number of gridded climate products has been produced. A selected collection of these products at the European (e.g., E-OBS) and global scale (e.g., CRU, GISTEMP, Berkeley Earth) is available through the Copernicus Climate Change Service (https://doi.org/10.24381/cds.11dedf0c). A review of the E-OBS dataset and its applications is available, among others, in Cornes et al. (2018), Ledesma and Futter (2017) and Raymond et al. (2017), while an example of evaluation of the accuracy of the E-OBS dataset against local observational measurements in Italy (Apulia region) can be found in My et al. (2022).

Several interpolation techniques are available to produce gridded maps from scattered in situ measurements of climate variables (Chen & Guo, 2017; Hofstra et al., 2008; Yang & Xing, 2021) and most of them are implemented in software packages for the R language (R Core Team, 2021). For example, gstat (Pebesma, 2004) is a well-known and user-friendly R package which implements both deterministic (inverse distance weighting [IDW]) and geostatistical (traditional kriging and its variants) methods. Geostatistical methods assume that the observed data are a realization of a continuously indexed spatial random field (Gelfand et al., 2010). In this respect, Gaussian random fields (GRFs) play a dominant role in spatial statistics and especially in the traditional field of geostatistics (Lindgren et al., 2011). Conversely, for deterministic methods no probabilistic model of data generation is taken into account. While geostatistical methods allow to attach an uncertainty measure to the individual spatial predictions, with deterministic methods is only possible to calculate an overall error rate (e.g., the root-mean-square error) by doing cross-validation (Sahu, 2021).

Uncertainty estimate is an important aspect to be considered when selecting and/or developing interpolation methods. Despite climate records not being immune to measurement and modelling uncertainties, scientists and practitioners typically have to deal with datasets for which the uncertainty information is generic, misleading, or absent (Merchant et al., 2017). Indeed, the available gridded datasets typically provide a single estimate for each specific location and time step (Tang et al., 2022), where a measure of uncertainty (the degree to which something is known; Nature, 2019) would allow to assess reliability and effects on derived products and information.

Although classical kriging is a very popular geostatistical method among climate practitioners, it is often neglected that its estimates of the prediction variance do not account for the uncertainty caused by inferring the covariance parameters from the data (Moyeed & Papritz, 2002; Pilz et al., 2005). In other words, by assuming that the spatial covariance function is known (Handcock & Stein, 1993; Helbert et al., 2009; Song et al., 2015) classical kriging suffers from uncertainty underestimation.

Concerning Italy, several exercises have been recently done and several products made available. For instance, Crespi et al. (2021) used a combination of two deterministic methods (IDW and PRISM; Daly et al., 2002) to develop high-resolution gridded datasets of daily mean temperature and precipitation records during 1980–2018 in northeastern Italy (Trentino-South Tyrol region). IDW was also applied in Caloiero et al. (2020) for southern Italy (Calabria region). Curci et al. (2021) employed a geostatistical approach (universal kriging) to generate interpolated maps for mean temperature and precipitation in central Italy (Abruzzo region) over the period 1930–2019. A comparison of IDW and geostatistical methods to produce monthly gridded precipitation over southern Italy (Calabria region) is discussed in Pellicone et al. (2018). Finally, Brunetti et al. (2014) applied regression analysis (local weighted linear regression [LWLR]) to obtain 1961–1990 monthly temperature climatologies for Italy. Although Brunetti et al. provide examples of uncertainty estimates (68% confidence interval) associated to each grid point, the proposed LWLR method does not explicitly account for the autocorrelation of data. Indeed, that method consists in an iterative application of local temperature-versus-elevation weighted regressions where the stations which contribute to the temperature estimates are searched within an a priori fixed spatial range of 200 km from a given grid cell.

In this study, we use a spatiotemporal regression model to interpolate Tmax and Tmin monthly mean data on a large space–time domain, namely the entire Italian territory for 60 years (1961–2020). We develop the statistical analysis in a Bayesian framework in order to incorporate all reasonable sources of uncertainty (including the parameters of the covariance function) in the final inferential summaries. Because of the large space–time domain, the inferential analysis has been implemented using the integrated nested Laplace approximation (INLA) approach (Rue et al., 2009) and the spatiotemporal covariance function has been derived from the solution of a stochastic partial differential equation (SPDE; Lindgren & Rue, 2015).

INLA is a deterministic algorithm which provides accurate approximation of the marginal posterior distributions of the model parameters based on Laplace and other numerical approximations and integration schemes. The SPDE approach provides a way to represent a continuous Gaussian Random Field through a

discretely indexed Gaussian Markov random field (GMRF; Lindgren et al., 2011). Computationally, GMRFs are much more efficient than GRFs as they are based on sparse matrices. The INLA-SPDE approach has been proved to be computationally faster than the simulation-based Markov chain Monte Carlo (MCMC) approach commonly used for Bayesian inference (Martino & Riebler, 2020). This is fundamental when models are complex or deal with massive datasets (Blangiardo & Cameletti, 2015).

The effective use of INLA and SPDE is documented in several studies based on environmental data, from ecological applications to disease mapping. A comprehensive list of updated references is available in Lindgren et al. (2022). Going through this list, it is illuminating to note the absence of applications in climate science. To bridge this gap, we here introduce INLA and SPDE by using and discussing a simple case study. In particular, we employ INLA and SPDE to develop a spatiotemporal model to derive gridded monthly temperature climatologies for Italy both for the most recent standard 30-year period (1991–2020) and three previous standard periods (1961–1990, 1971–2000, 1981–2010).

The input dataset is described in section 2. Section 3 introduces the statistical model, while section 4 discusses results, model validation and possible applications of the model estimates. Section 5 reviews the benefits of the INLA/SPDE approach.

2 | DATA

2.1 | The study area

Italy, with latitude boundaries between 36°N and 48°N and longitude boundaries between 5°E and 20°E, is our study domain. The country is characterized by a narrow and long shape of about 7500 km of coast line which extends into the Mediterranean sea, two large mountain systems (the Alps to the north, and the Apennines which run northwest to south along the country), a large plain (the Po Valley with a surface of 46,000 km^2) and two major islands (Sicily and Sardinia). This complex morphology leads to variegated and contrasted climate regimes (from warm temperate with dry summer to polar) which strongly influences the variability of soil moisture and temperature (Lo Papa et al., 2020).

According to the technical report "Climate indicators in Italy" (ISPRA, 2021), which uses a dataset of 140 homogenized daily temperature time series from 1961 to 2020, the Italian-average mean-temperature anomalies (baseline period 1961–1990) are generally characterized by higher variability than global anomalies. The estimated warming trend for the annual Tmax and

Tmin Italian averages is +0.42 and +0.35°C·decade^{-1}, respectively. As regards Tmax, a positive anomaly of +1.82°C ranks 2020 as the warmest year in Italy since 1961 (2021 and 2022 are not included in this analysis).

2.2 | Temperature data

Our study uses daily maximum and minimum temperature records taken from SCIA (Desiato et al., 2011), a database which collects climate observations from a national network (the Italian Air Force synoptic network) and several meteorological, agro-meteorological and hydrographic networks across the 20 Italian regions. In SCIA, the raw temperature data are subjected to strict quality controls (ISPRA, 2016) for the detection of erroneous daily observations. Such quality controls are basically a subset of those described by Durre et al. (2010) for the Global Historical Climatology Network (GHCN) dataset (Menne et al., 2012).

Despite the large number of temperature records which populate the SCIA database, here, we only focus on those time series which supply data during 1961–2020 and with no missing data for the 2011–2020 decade. This choice is motivated by the fact that the 2011–2020 decade was particularly remarkable as regards global warming, being the warmest on record since 1961 both in Italy and at global scale (ISPRA, 2021; WMO, 2021).

The altitude of the selected weather stations varies between 0 (Villapiana Scalo, Marina di Ravenna, Fossalon stations) and 3488 m (Pian Rosa station) above mean sea level. Figure 1 displays their spatial distribution, while the inset on the left shows the number of available stations across the years. A visual inspection reveals that only a limited number of stations provides a complete coverage of the study period. Specifically, less than 250 stations are available over the first 30 years (1961–1990), while more than 500 stations are available since the early 2000s. This increase in the station number corresponds to the development of local hydrographic authorities and regional environmental agencies in Italy and the transition from mechanical to automatic station networks (Libertino et al., 2018).

The homogenization of temperature time series is a pre-requisite for applying statistical data analysis (Toreti et al., 2011). For the detection and adjustment of nonclimate perturbations (breakpoints) in temperature time series we used Climatol (Guijarro, 2019), a ready-to-use R package and an automatic well-established homogenization tool (Joelsson et al., 2022). Details and results of the homogenization study are not given here for lack of space. We refer interested readers to Fioravanti et al. (2019) for a more in-depth description both of the continuity and completeness criteria adopted for the selection
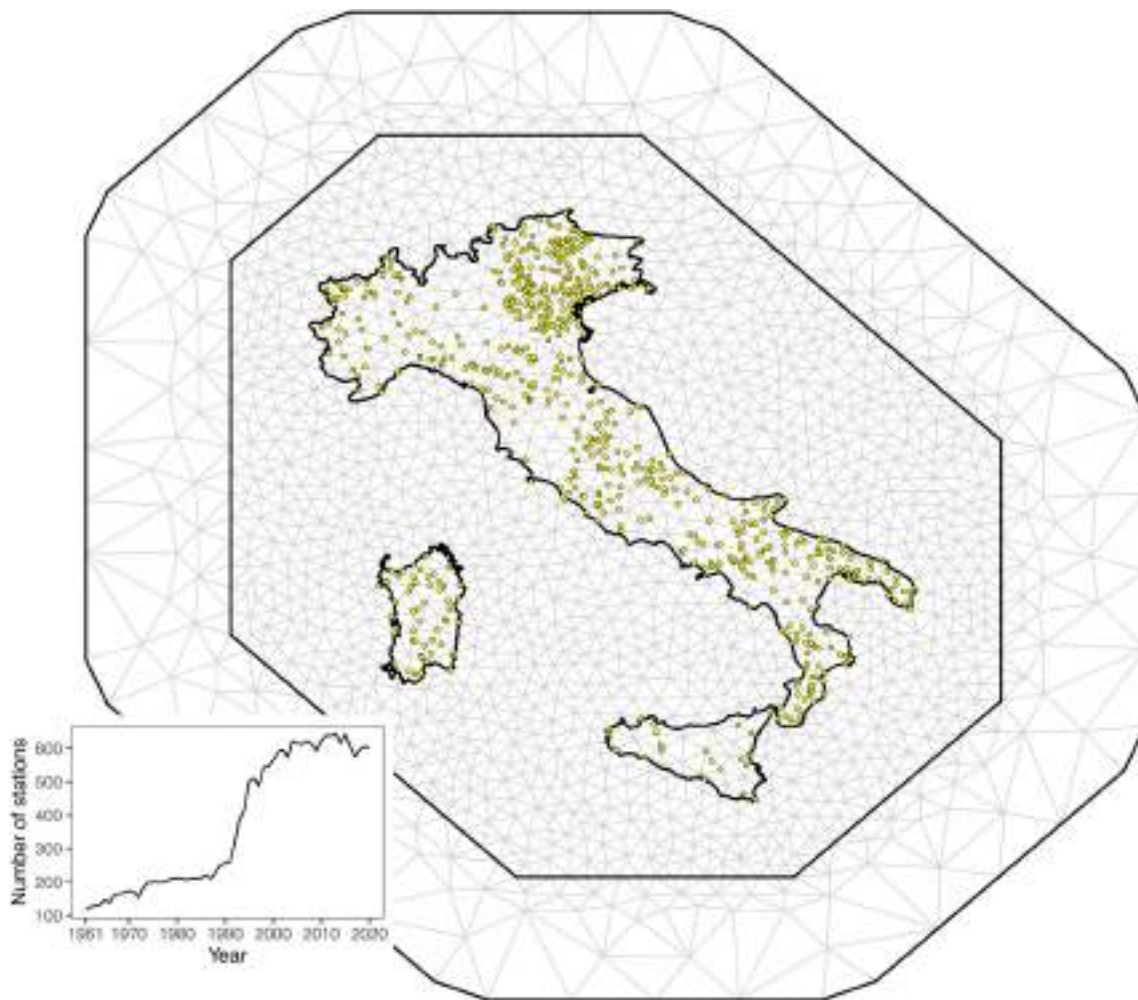
FIGURE 1 Spatial distribution and historical time series (inset, bottom-left) of the number of the selected input stations for the month of June (Tmax). The same results hold for Tmin and the other months of the year. In the background, the mesh used to build the SPDE approximation to the continuous Matérn field. Note, our study domain comprises only the Italian peninsula and the two major islands (Sicily and Sardinia) [Colour figure can be viewed at wileyonlinelibrary.com]

of the temperature time series and the homogenization process.

Tmax and Tmin monthly averages were calculated using only daily observations collected during the 1961–2020 period, no estimation or smoothing methods were applied to refill data gaps. The daily data were averaged by month for each year following the "10/4" completeness rule (WMO, 2017), which states that monthly values should not be calculated when (1) observations are missing for 11 or more days during the month; (2) observations are missing for a period of 5 or more consecutive days during the month.

2.3 | Terrain and location predictors

We used the following variables as predictors of the Tmax and Tmin monthly mean temperatures: latitude, elevation and distance to sea (Osborn et al., 2021). All these predictors have a clear and physically sound effect on air temperature.

For interpolation purposes, the predictors must be available as rectangular raster datasets (see Figure 2). The latitude raster simply contains the latitude value (Universal Transversal Mercator original coordinates scaled to kilometres) for each grid cell. The "distance to sea" raster contains the minimum Euclidean distance of each grid point to the Italian coastline. These two rasters were generated by using the geoprocessing capabilities of the R terra package (Maechler et al., 2021). Finally, the digital elevation model dataset EU-DEM v1.1 from the Land Monitoring Service of Copernicus (https://land.copernicus.eu/imagery-in-situ/eu-dem) was used as the elevation raster. The final spatial resolutions of the gridded predictors is 1 km × 1 km.

All these predictors have been rescaled to have zero mean and standard deviation equal to one to avoid
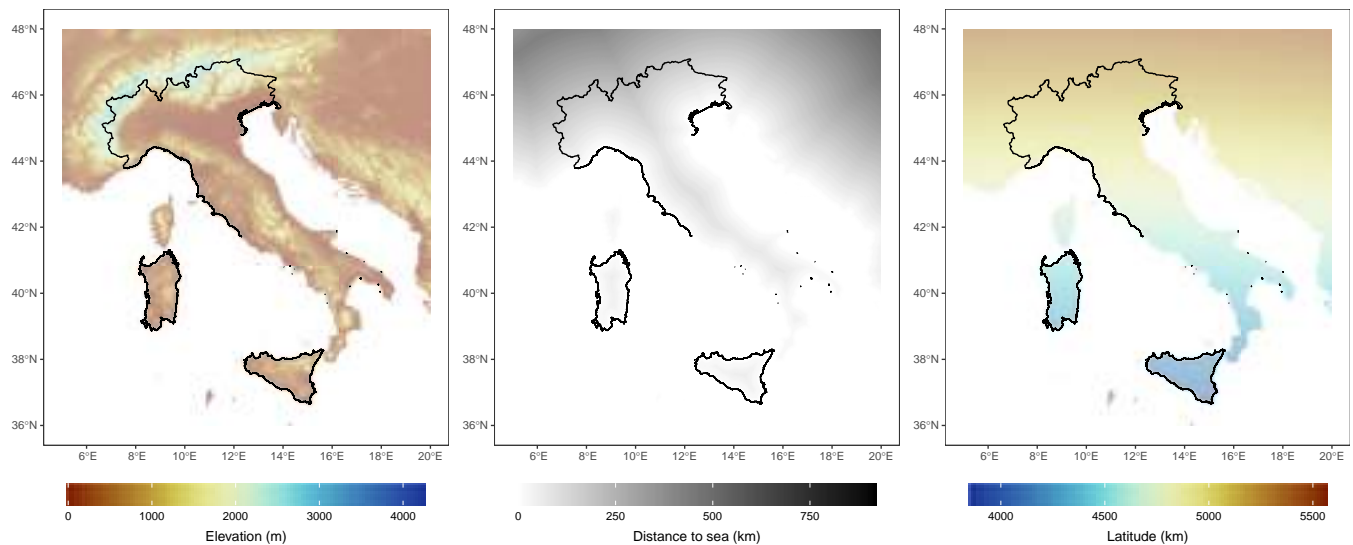
FIGURE 2 Raster maps of the spatial predictors used for interpolating Tmax and Tmin monthly mean records over the period 1961–2020 [Colour figure can be viewed at wileyonlinelibrary.com]

numerical problems and ease interpretability of their effects on temperature data.

3 | METHODOLOGY

3.1 | Modelling monthly temperatures with INLA and the SPDE

Here, we introduce the mathematics which defines our statistical model. Before that, we clearly motivate why an explicit spatiotemporal approach is needed: to this end, we start from a simple regression model.

Let $y^m(t,s)$ denote the realization of a space–time process $Y^m(t,s)$ which represents the monthly Tmax (Tmin) mean temperatures at year $t=1961,...,2020$ of month $m=1,...,12$ at location $s \in \mathcal{S}$, where $\mathcal{S}$ is the domain of interest. A linear model where the covariates account for the temporal and spatial trend can be written as

$$y^m(t,s) = \mu^m + \alpha^m\, t + \mathbf{x}(s)\beta^{mT} + \varepsilon^m(t,s). \quad (1)$$

Here μ^m is a month specific intercept, $\mathbf{x}(s)$ is the vector of observed covariates (introduced in section 2.3), β^m is the vector of coefficients to be estimated. We also include a linear time trend with coefficient α^m and a Gaussian distributed independent measurement error $\varepsilon^m(t,s)$ with mean 0 and standard deviation σ_ε^m. We fit the same model independently for each month and for both Tmax and Tmin (24 models in total), as from our exploratory analysis (not shown) we observed that the impact of each predictor varies across months. As we

are interested in the long term trend for each month, modelling each months' time series independently considerably simplifies the model allowing to ignore the seasonal cycle of temperatures and the strong temporal dependence between consecutive months. To ease the notation in the rest of the paper we omit the month index m.

We fit a simple multiple linear regression model in Equation (1) to our data and, only for exploratory purposes, use spatiotemporal variograms (Wikle et al., 2019) to investigate the independence assumption in model residuals $\varepsilon(t,s)$.

Figure 3 shows the variograms for the residuals of model 1 (solid lines) for all the 12 months fitted with Tmax data (similar results hold for Tmin, not shown). Spatial correlation is clearly visible in all panels, moreover, the variograms show that the residual spatial correlation varies across months. For example, September and October are characterized by large range spatial dependence with the variance increasing slowly, while December and January are characterized by a clear spatial range of 150 km. When it comes to time, all months exhibit week temporal correlation. In summary, the results of Figure 3 indicate that the standard assumption of independent and identically distributed errors is not tenable for our dataset. For a more general review of potential pitfalls that can seriously affect the results of ordinary least square regression we suggest Zuur et al. (2010).

We therefore modify the model including random effects terms that explicitly account for spatial and temporal correlation among temperature records,
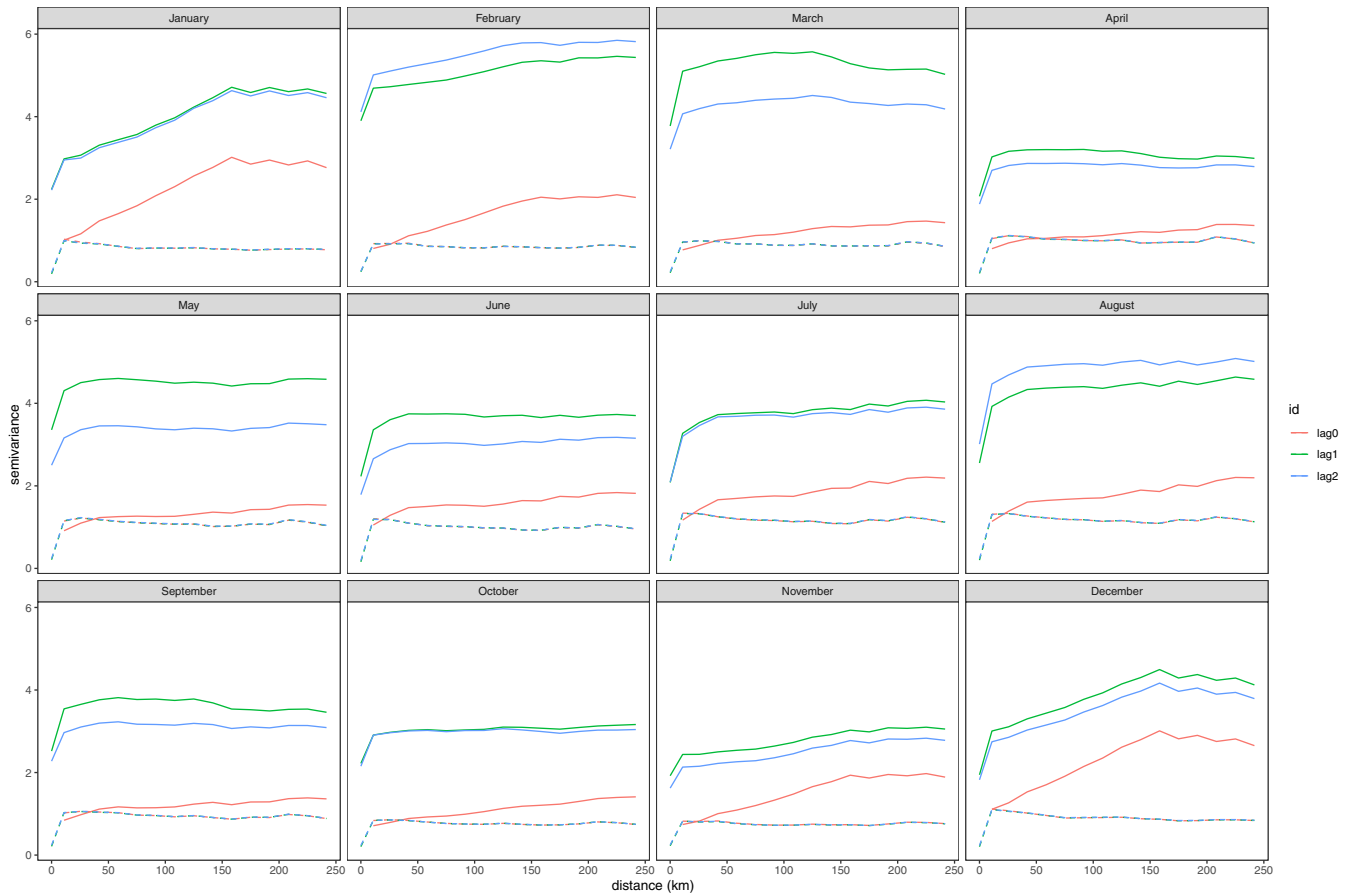
FIGURE 3 Monthly spatiotemporal variograms for the residuals of the linear model in Equation (1) (solid lines) and the spatiotemporal model in Equation (2) (dashed lines). Results refer to the Tmax parameter. Similar results hold for Tmin [Colour figure can be viewed at wileyonlinelibrary.com]

$$y(t,s) = \mu + \alpha\, t + \mathbf{x}(s)\beta' + \omega(t,s) + z(t) + \varepsilon(t,s). \qquad (2)$$

Here, $z(t) \sim \mathcal{N}\left(0, \sigma_z^2\right)$, $t = 1961,\ldots,2020$, is a temporal random effect meant to account for the extra temporal variability (colder or warmer than average years) that is not captured by the linear time trend. A similar effect is used in Serrano-Notivoli et al. (2019) for the estimation of daily temperatures in Spain.

The term $\omega(t,s)$ is a spatially correlated process meant to capture the residual dependence beside that described by the covariates. We assume $\omega(t,s)$ to be independent in time but correlated in space with covariance function,

$$\mathrm{Cov}(\omega(t,s), \omega(t',s')) = \begin{cases} 0, & \text{for } t \neq t' \\ \mathcal{C}(h), & \text{for } t = t' \end{cases}, \qquad (3)$$

where $h = \left\| s - s' \right\|$ is the Euclidean distance between sites s and s'.

A common specification for the purely spatial covariance function $\mathcal{C}(h)$ is the Matérn function,

$$\mathcal{C}(h) = \sigma_\omega^2 \frac{1}{\Gamma(\nu)2^{\nu-1}} (k\,h)^\nu K_\nu(k\,h), \qquad (4)$$

where σ_ω^2 is the marginal variance and $K_\nu(\cdot)$ the Bessel function of second kind and order $\nu > 0$. The parameter ν measures the degree of spatial smoothness of the process. As this parameter is hard to estimate, we fixed it to one as suggested by Blangiardo and Cameletti (2015). The term $k > 0$ is a scaling parameter related to the range ρ, that is, the distance at which the spatial correlation becomes negligible (close to 0.1). The larger ρ is, the more dependent the spatial process is. In Equation (4) ρ is defined as $\frac{\sqrt{8\nu}}{k}$, an empirically derived definition proposed by Lindgren et al. (2011). To represent the continuous field $\omega(t,s)$ as a GMRF, we used the SPDE approach, which is based on the finite element method (fem). In our case, the triangulation used for fem is the one shown in Figure 1. In order to obtain accurate approximations of the underlying GRF, the triangular mesh must be dense enough to capture the spatial variability of the monthly mean temperatures. It is noteworthy to observe that we

constructed a mesh which is rather dense over the study domain and sparser in the outer region. The purpose of the outer region is twofold: avoid boundary effects and, at the same time, reduce the computational costs.

3.2 | Priors definition

We finalize our Bayesian model by defining priors on all model parameters: the coefficients α and β, the standard deviations $\sigma_\varepsilon, \sigma_z, \sigma_\omega$ and the range ρ of the Matérn function. We used vague independent Gaussian priors for α and β and log-Gamma priors on the log precision parameters (equal to the log of the inverse of the variances $\sigma_\varepsilon, \sigma_z$).

For the parameters of the Matérn field, we used penalized complexity (PC) priors (Simpson et al., 2017). PC priors are designed to penalize model complexity and avoid overfitting. For ρ and σ_ω we used the joint PC prior suggested in Fuglstad et al. (2019). The prior parameters are expressed as

$$\mathrm{Prob}\left(\rho < u_\rho\right) = \alpha_\rho; \mathrm{Prob}\left(\sigma_\omega > u_{\sigma_\omega}\right) = \alpha_{\sigma_\omega},$$

where we set $u_\rho = 700$, $\alpha_\rho = 0.7$, $u_{\sigma_\omega} = 1$ and $\alpha_{\sigma_\omega} = 0.6$. Since the large scale spatial dependence is explained by the

elevation and location covariates, it is reasonable to assume the range of the residual spatial dependence to be smaller than 700 km.

4 | RESULTS

In this section we first look at the residuals of the model given by Equation (2) to check if the assumption of error independence is now more reasonably satisfied. We then discuss parameter estimates for the 24 monthly models and assess model performance through cross-validation. Finally, we use the interpolated monthly surfaces of Tmax and Tmin to calculate climate normals for the following standard 30-year periods: 1961–1990, 1971–2000, 1981–2010 and 1991–2020. A comparison between 1991–2020 and the previous standard periods is then presented.

4.1 | Variograms for the residuals of the spatiotemporal model

The dashed lines in Figure 3 refer to the spatiotemporal variograms for the Tmax residuals of model of Equation (2). The plot shows that the Matérn field allowed to capture the spatiotemporal signal: the
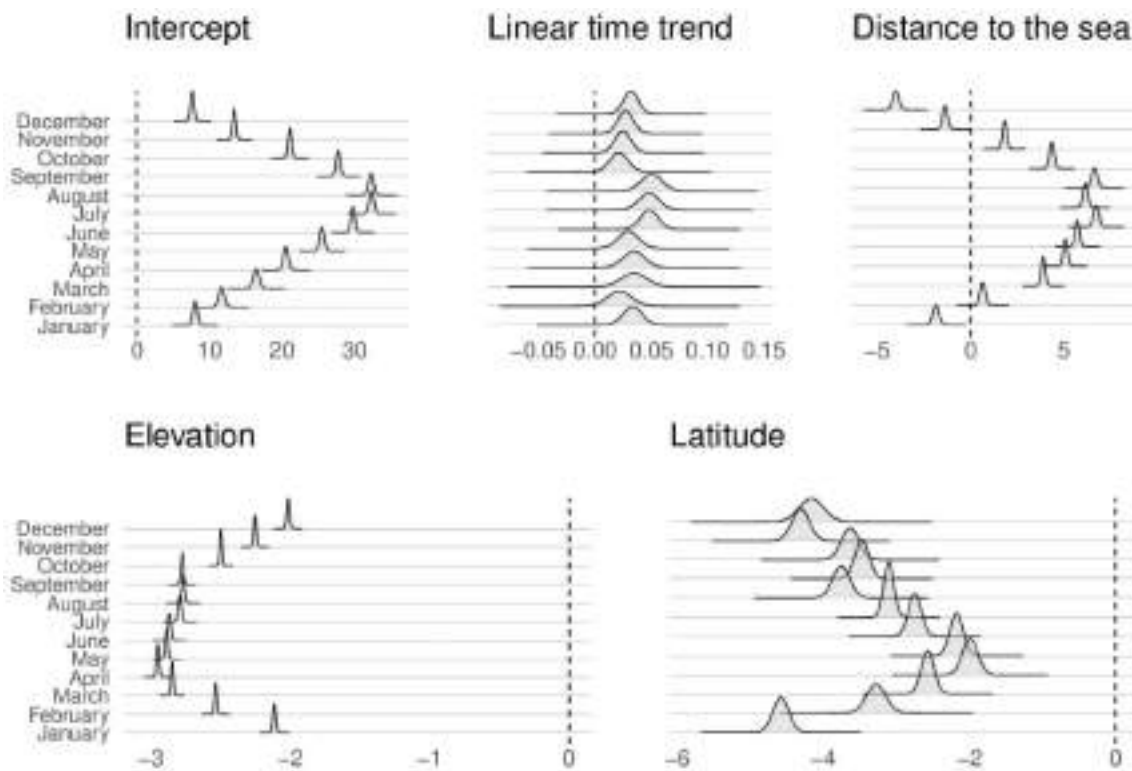

FIGURE 4 Tmax: marginal posterior distributions for the intercept μ, the linear time trend coefficient α and fixed effects β. The shaded colour refers to fixed effects for which the whole 95% credible interval is above or below zero (fixed effects which are significantly different from zero)
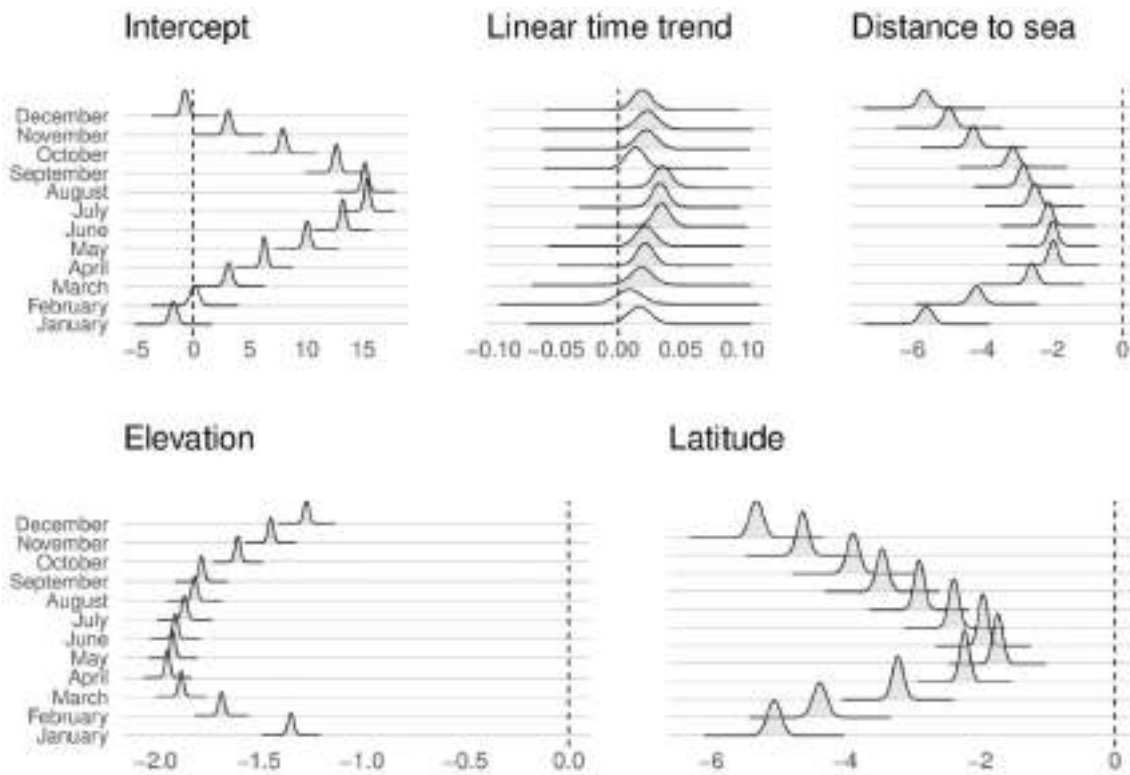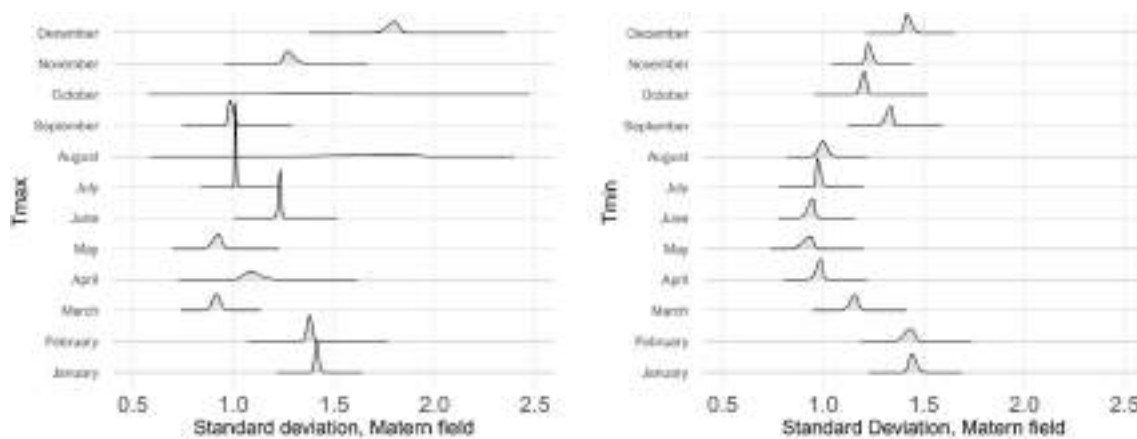
FIGURE 5 Tmin: marginal posterior distributions for the intercept μ, the linear time trend α and fixed effects β. The shaded colour refers to fixed effects for which the whole 95% credible interval is above or below zero (fixed effects which are significantly different from zero)

FIGURE 6 Marginal posterior distributions for the standard deviation of the Matérn field $\omega(t,s)$

uncorrelated residuals return variograms which are more or less flat both along the spatial and temporal dimension. This result holds for Tmin as well (not shown).

4.2 | Parameter estimates

The plots in Figures 4 and 5 reproduce the marginal posterior distributions for the intercept μ, the linear time

effect α and the covariate coefficients β for Tmax and Tmin monthly means, respectively.

The plots show both the intercept and the spatial predictors (elevation, latitude and distance to sea) have a clear seasonal behaviour, whatever the parameter (Tmax/Tmin). These seasonal-varying effects support our initial hypothesis that a monthly regression analysis could improve the accuracy of the final estimates. Interestingly, both latitude and elevation have a negative

FIGURE 7 Marginal posterior distributions for the spatial range ρ of the Matérn field ω
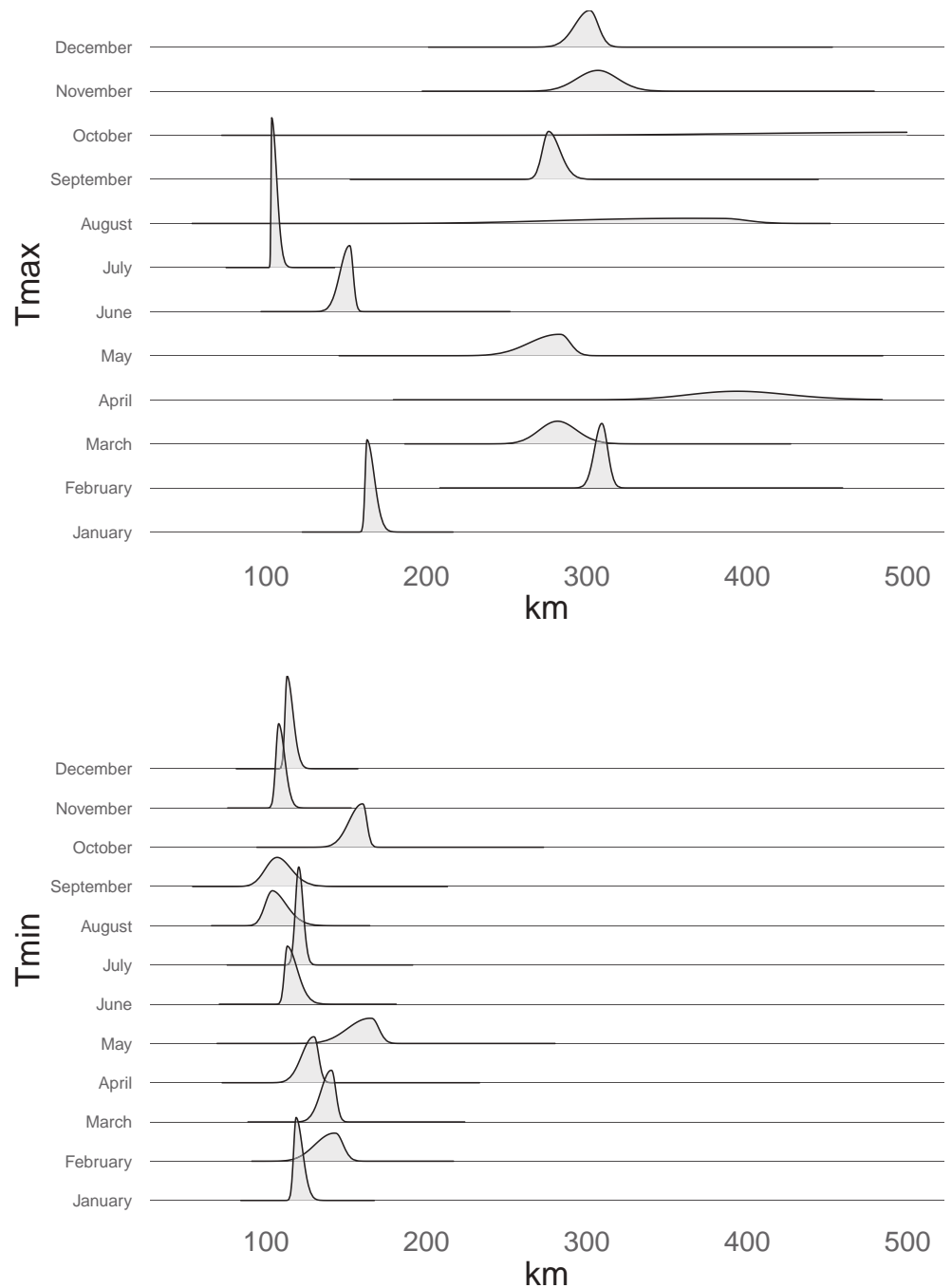
effect on temperatures across the months, while the sign of the distance to sea effect depends on the variable: when Tmax is considered, distance to sea has a negative effect during the winter season and positive otherwise. Conversely, a negative effect on Tmin monthly means characterizes distance to sea across the year.

The linear time effect α indicates a clear increase of the Tmax and Tmin monthly mean temperatures during 1961–2020, more marked during the summer season (June–August). Further, a significant positive effect characterizes the linear time trend of Tmax across the year,

while a weaker linear time trend effect characterizes the winter season (January and February) and the month of September for Tmin.

The marginal posterior densities of the standard deviation σ_ω and spatial range ρ of the Matérn field ω are reproduced in Figures 6 and 7. As observed above (section 3), the larger ρ is, the more dependent the spatial process is. First, we observe that, unlike the spatial predictors in Figures 4 and 5, here the estimated densities are characterized by very skewed shapes and do not exhibit any pronounced seasonal effect. Second, the

TABLE 1 Cross-validation monthly results: average performance measures over the three validation datasets

|	RMSE °C		Correlation		Bias °C	
	Tmax	Tmin	Tmax	Tmin	Tmax	Tmin
January	1.13	1.54	0.96	0.95	0.13	0.02
February	1.05	1.46	0.97	0.95	0.18	0.03
March	1.04	1.36	0.96	0.94	0.23	0.06
April	1.07	1.3	0.96	0.94	0.27	0.11
May	1.15	1.33	0.96	0.94	0.29	0.07
June	1.21	1.48	0.96	0.92	0.29	0.03
July	1.33	1.63	0.95	0.92	0.30	−0.01
August	1.3	1.63	0.95	0.92	0.29	−0.03
September	1.13	1.42	0.96	0.94	0.24	−0.02
October	1.01	1.4	0.97	0.94	0.21	0
November	1	1.37	0.97	0.95	0.15	0.03
December	1.12	1.49	0.97	0.95	0.12	0.04
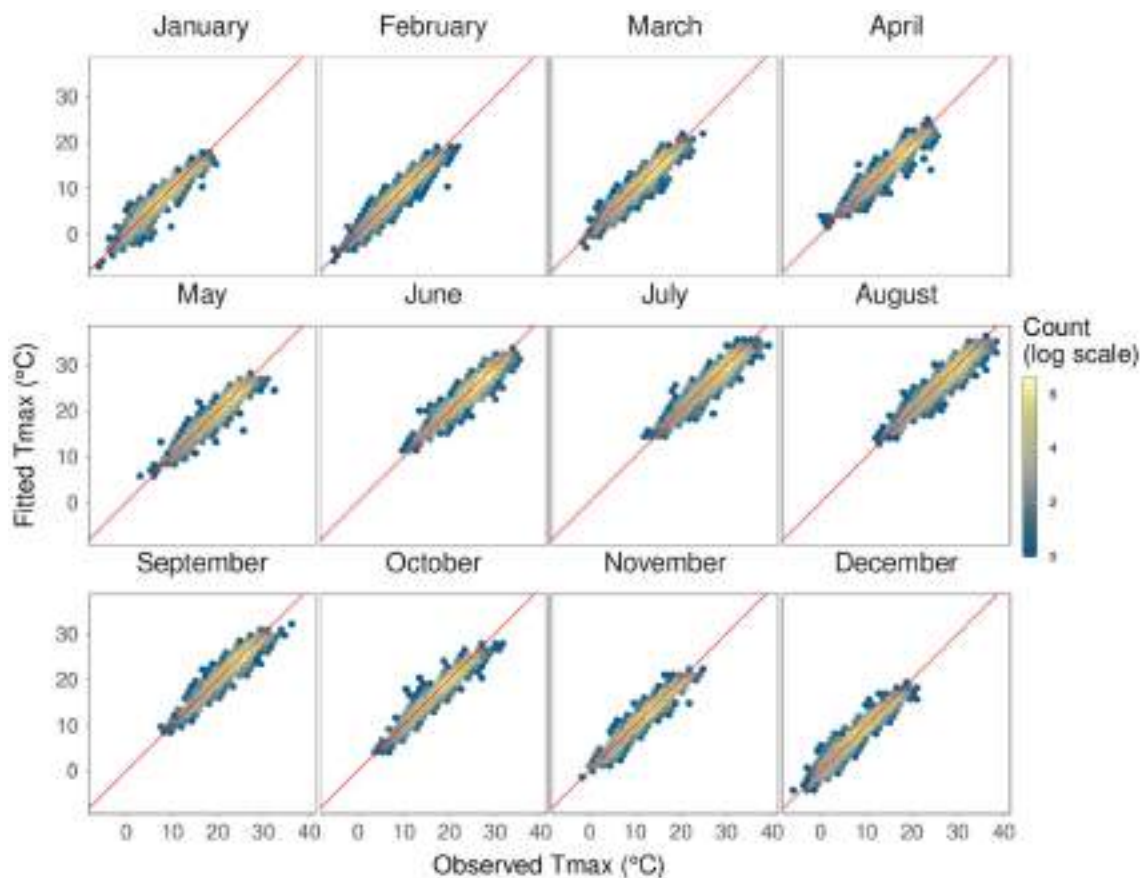
FIGURE 8 Agreement between modelled (means of the posterior predictive distributions) and observed Tmax monthly means. Lighter colours indicate areas with higher points concentrations. The solid line is the 1:1 line as a reference [Colour figure can be viewed at wileyonlinelibrary.com]

visual inspection of the posterior densities suggests that Tmax data show much more spatial structure than Tmin: the posterior mean of ρ varies between 100 and 160 km for Tmin and between 100 and 500 km for Tmax. At the same time, both the standard deviation and the spatial range of Tmax shows greater variability than Tmin.
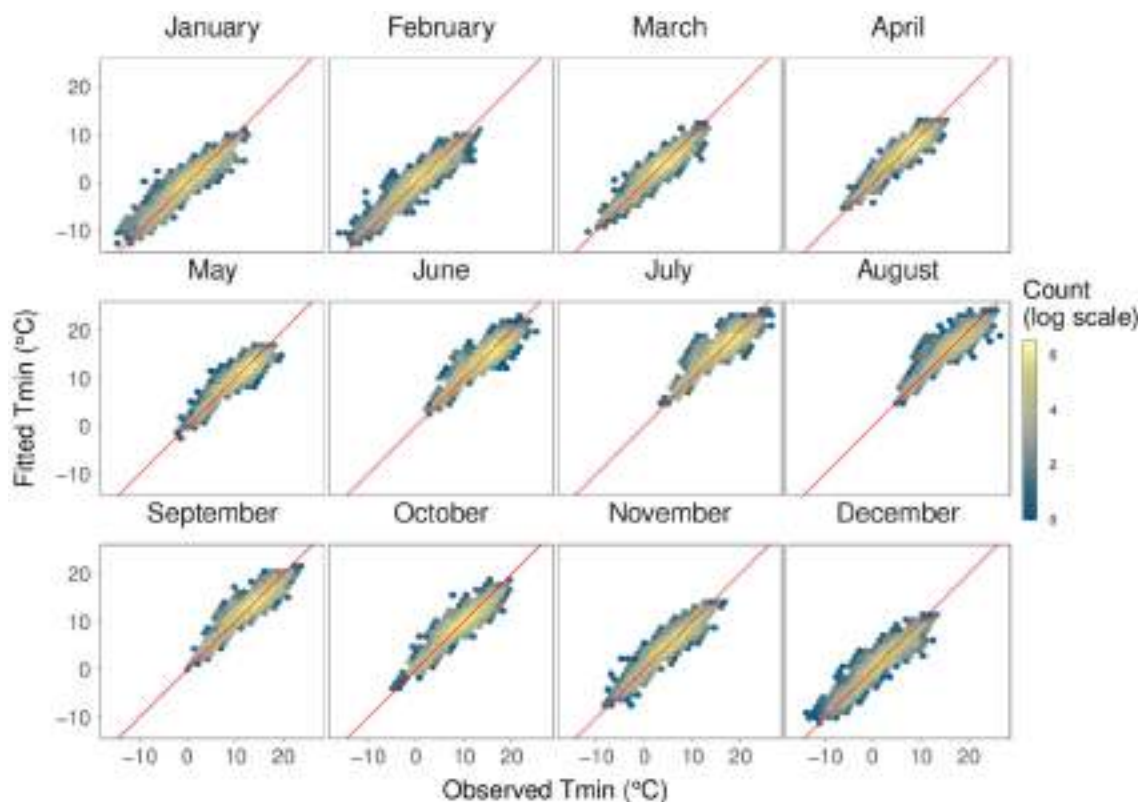
FIGURE 9 Agreement between modelled (means of the posterior predictive distributions) and observed Tmin monthly means. Lighter colours indicate areas with higher points concentrations. The solid line is the 1:1 line as a reference [Colour figure can be viewed at wileyonlinelibrary.com]

We conclude this section observing that the estimated densities of the spatial range are consistent with the variograms reproduced in Figure 3 for the residuals of model 1. For example, the diffuse posterior density for the month of October indicates that for Tmax the spatial correlation decays slowly at long distance. Its posterior mean is 488 km and its 95% credible interval is quite wide with bounds of 323 and 616 km. Indeed, the corresponding variograms increase slowly and level off after 250 km (the maximum spatial distance reported in Figure 3). Conversely, the posterior distribution for the month of January is characterized by a posterior mean of 165 km and a narrow 95% credible interval with bounds of 160 and 173 km. A spatial range of approximately 160 km can be inferred by a visual inspection of the variogram of Figure 3.

4.3 | Model validation

To investigate if the model generalizes well or suffers from overfitting/underfitting, we implemented the following validation strategy. First, we identified the weather stations which provide data across all months both for Tmax and Tmin (it can happen that a station does not meet completeness criteria for a specific month and/or variable). These

stations were stratified according to their network of origin (see section 2.2). Within each network, we sampled randomly 10% of the stations in order to define a validation dataset (67 stations). The remainder of the stations constitutes the training dataset. Finally, we fitted the 24 monthly models to the training stations and the fitted models were used to predict the monthly mean temperatures on the validation stations. The above procedure was repeated using three different validation and training datasets.

As performance measures we choose the following statistics comparing predicted and observed values: (1) the Pearson correlation coefficient; (2) the root-mean-square error (RMSE); (3) the bias (calculated as difference between fitted and observed values). For each station, these statistics are computed by comparing the observed Tmax and Tmin monthly means and the means of the corresponding posterior predictive distributions (PPDs; McElreath, 2020) computed using the spatiotemporal model. For each of the three validation datasets, the average of each performance measure over all stations was computed. Table 1 reports the global model performance, that is, the average of each score over the three different validations datasets.

The results of Table 1 suggest a good model performance across the months both for Tmax and Tmin. The Pearson correlation coefficient is always greater than 0.9, while the
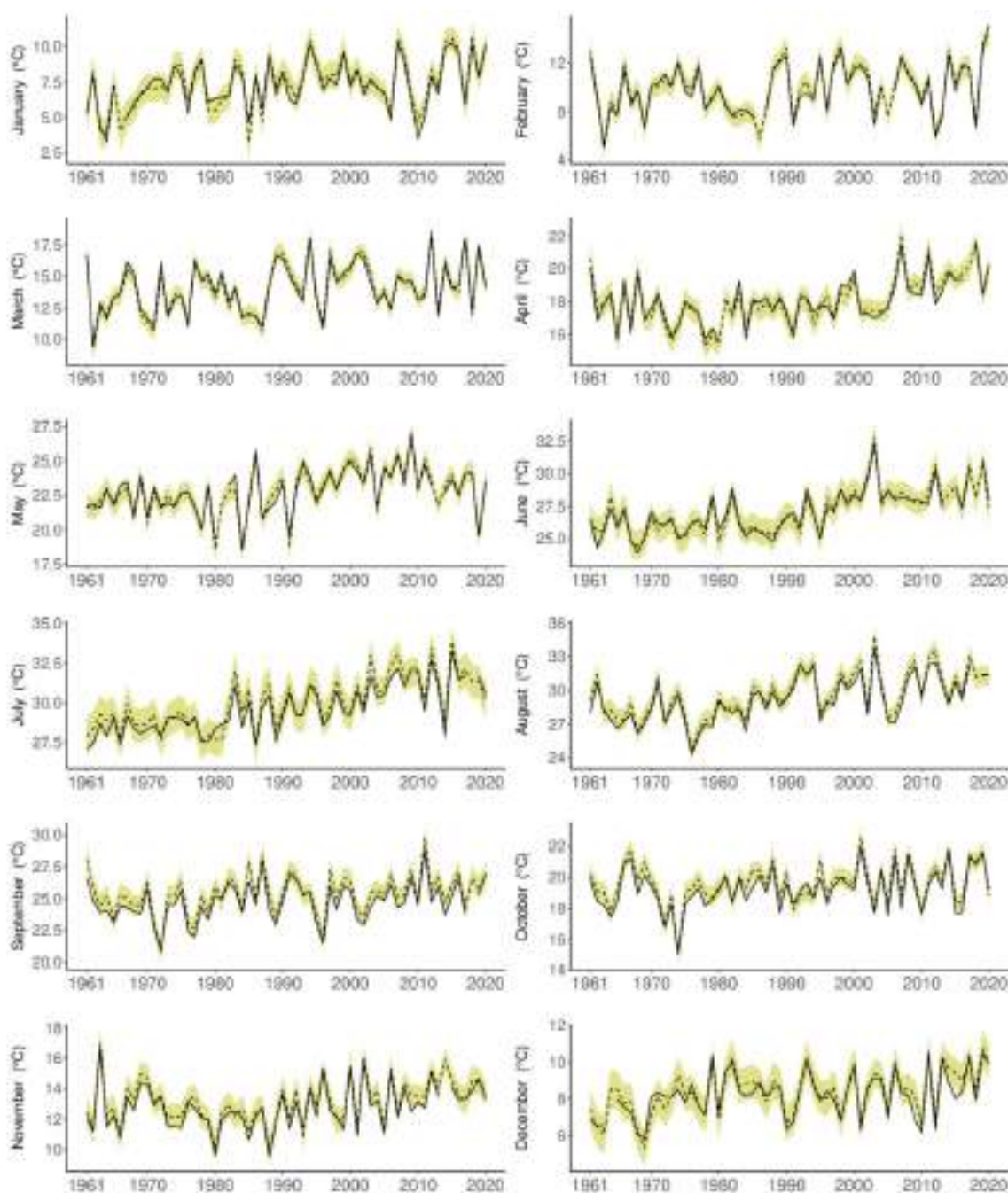
FIGURE 10 Maximum temperature monthly time series for the station of "Forlì urbana" (Emilia-Romagna, north of Italy). Observed (solid lines) versus fitted values (dashed lines). The 95% credible interval of each series is represented as a yellow band [Colour figure can be viewed at wileyonlinelibrary.com]

bias ranges between 0 and 0.3°C in absolute value. Finally, the RMSE is always lower than 1.5°C with the exception of the months of January, July and August for Tmin.

The scatterplots in Figures 8 and 9 show, for the validation datasets, the distribution of the fitted values (means of the PPDs) versus the observed values for Tmax and Tmin, respectively. The points spread uniformly along the diagonal line, showing good agreement between observed and modelled data.

Figures 10 and 11 compare observed and predicted annual time series for two illustrative stations chosen from the validation dataset. The plots indicate that the Tmax and Tmin monthly models are able to reproduce the temporal variability of the weather stations in the validation dataset. Each estimate (mean of the PPD) comes with an associated measure of uncertainty: the yellow band marks the bounds of the 95% prediction interval (Figures 10 and 11). In general, almost all of the observed
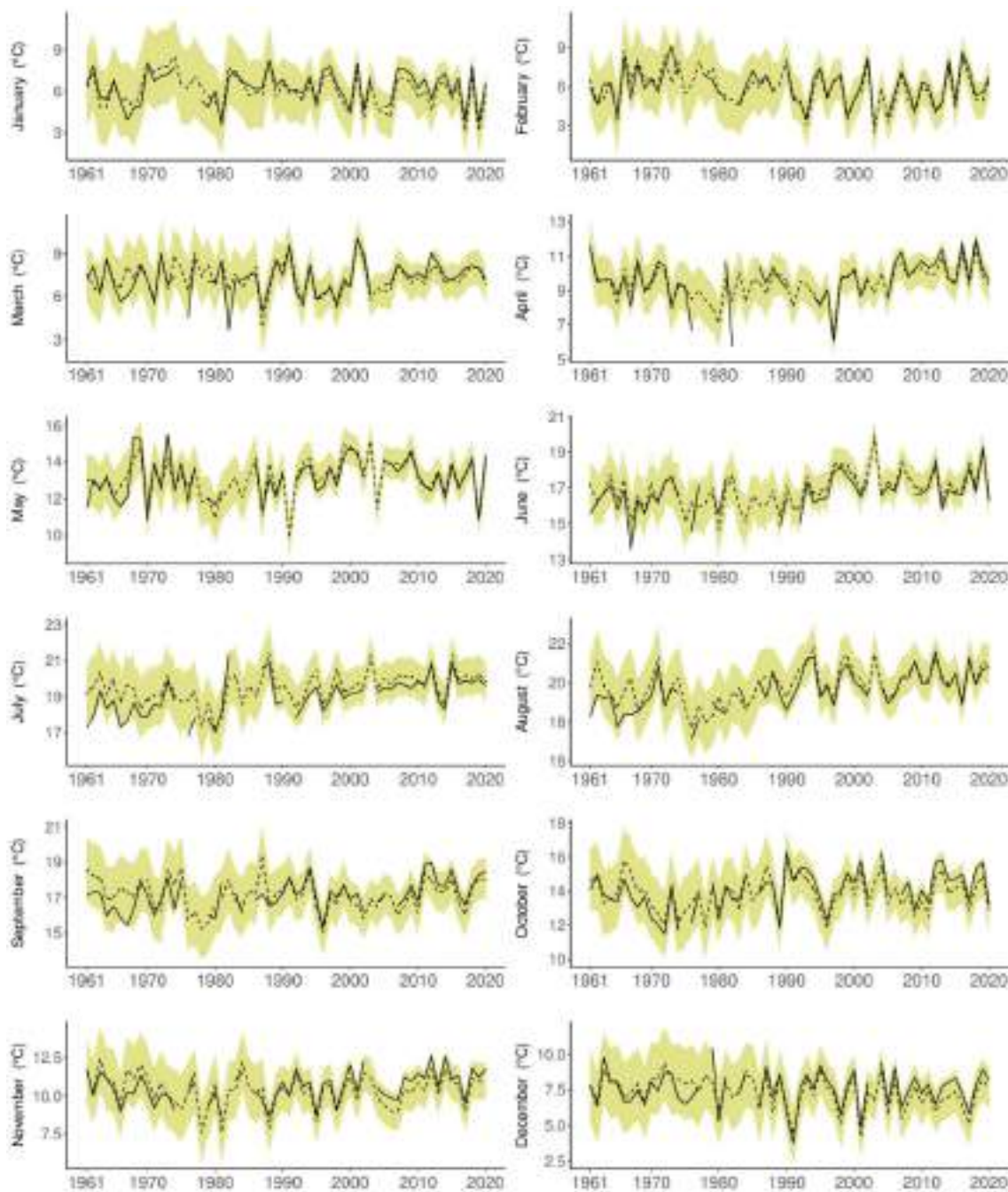
FIGURE 11 Minimum temperature monthly time series for the station of Mileto (Calabria, south of Italy). Observed (solid lines) versus fitted values (dashed lines) [Colour figure can be viewed at wileyonlinelibrary.com]

temperatures fall within their 95% prediction interval. Note, in this example the 95% prediction intervals are quite wide for the station of Mileto in Figure 11.

4.4 | Tmax and Tmin monthly climatologies

Once the model is fitted, it is possible to sample from the posterior distribution of all parameters. Each sample would correspond to parameters values that are plausible under our model assumptions and observed data. We use such simulated values in order to generate gridded maps of temperatures over the space–time domain considered in the paper: the Italian territory in the period 1961–2020. For each month and year, we generate 1000 such maps. These represent an ensamble that embeds in a natural way the various uncertainties in the model.

We can then use such simulated maps to create 1000 long-term (30-year) monthly averages (climatologies), for
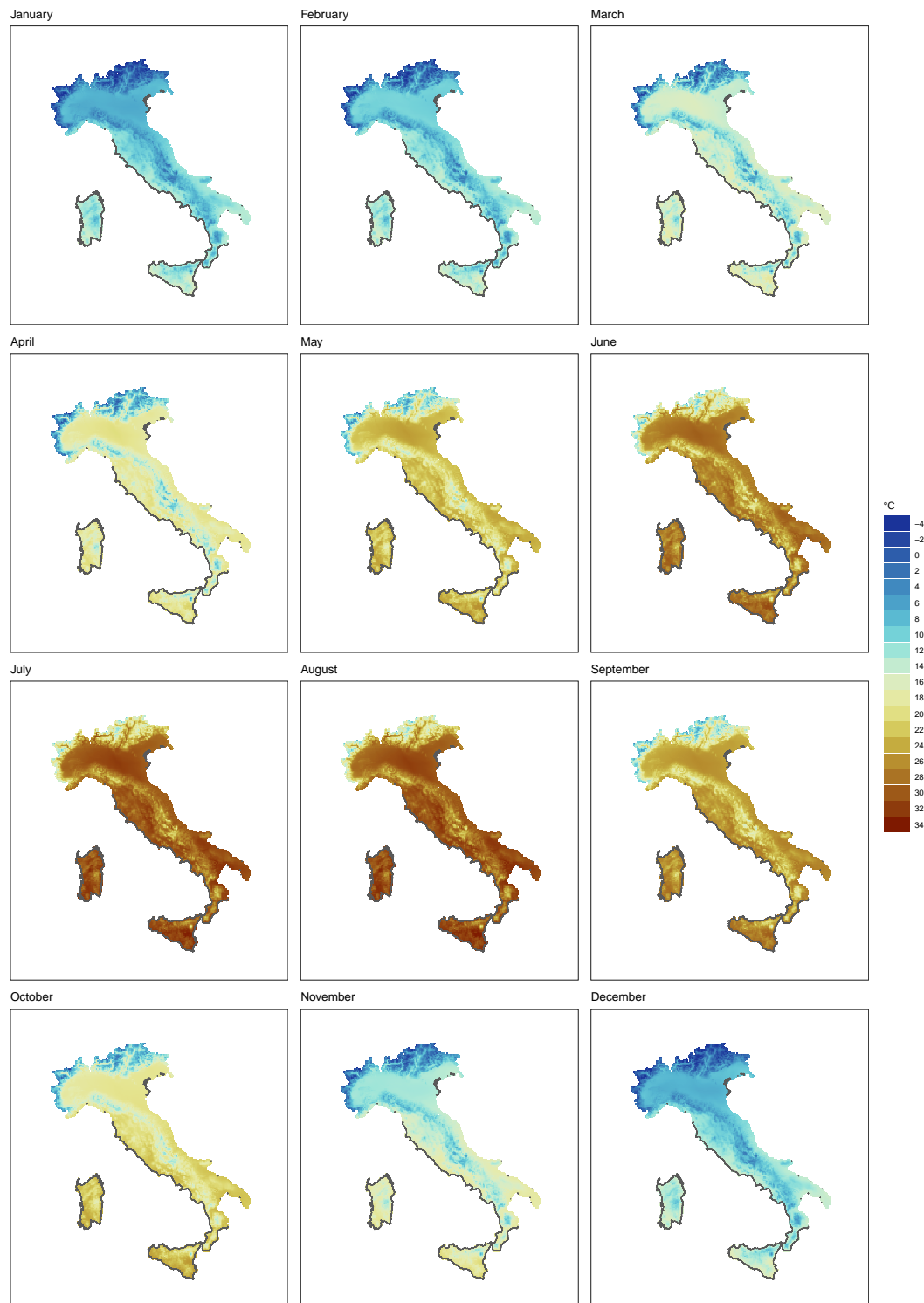
FIGURE 12   Mean of the sample distribution of the monthly climatologies in Italy for the period 1991–2020 (Tmax) [Colour figure can be viewed at wileyonlinelibrary.com]

example, for the period 1991–2020. The 1000 climatologies represent a distribution of gridded maps, which we summarize using their mean and standard deviation. In Figures 12 and 13 the resulting means are plotted for the period 1991–2020 for Tmax and Tmin, respectively.

The corresponding standard deviations maps are in Figure 14 (Tmax only).

These maps point to a global low degree of uncertainty in our predictions, with values of the standard deviation lower that 0.3°C. As expected, greater
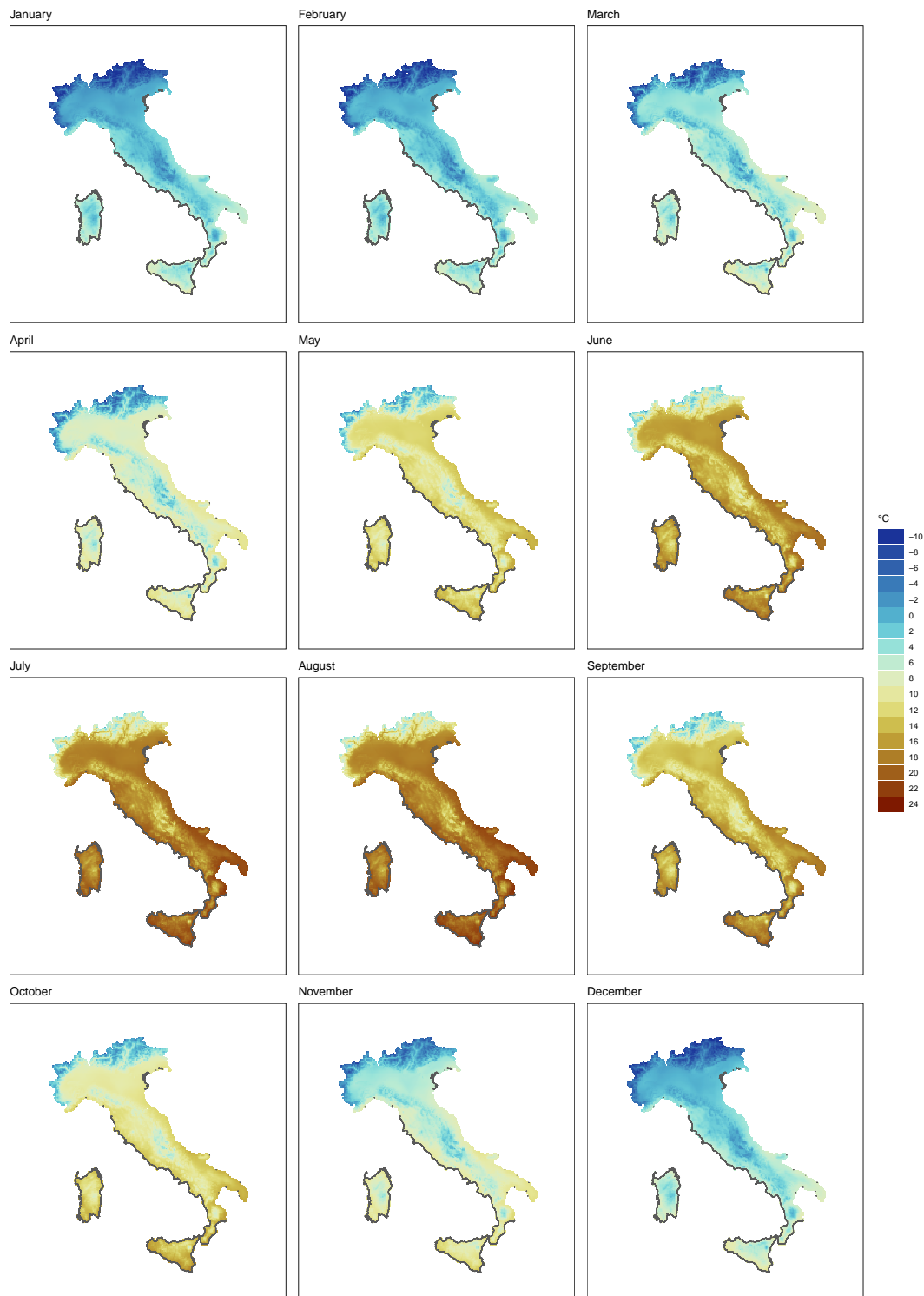
FIGURE 13   Mean of the sample distribution of the monthly climatologies in Italy for the period 1991–2020 (Tmin) [Colour figure can be viewed at wileyonlinelibrary.com]

uncertainty affects those areas where the density of the weather stations is poorer (see the spatial distribution of the input stations in Figure 1).

Another way to use our ensamble of gridded maps is to compute "difference" maps (1991–2020 vs. 1961–1990, 1991–2020 vs. 1971–2000, 1991–2020 vs. 1981–2010). We follow the same procedure as before, that is, we compute 1000 differences for each month and year and then summarize them using mean and standard deviation.