

A HIERARCHICAL MULTIVARIATE SPATIO-TEMPORAL MODEL FOR CLUSTERED CLIMATE DATA WITH ANNUAL CYCLES

BY GIANLUCA MASTRANTONIO^{*,1}, GIOVANNA JONA LASINIO^{†,1},
ALESSIO POLLICE^{‡,1}, GIULIA CAPOTORTI[†], LORENZO TEODONIO[§],
GIULIO GENOVA[†] AND CARLO BLASI[†]

Politecnico di Torino^{*}, *Sapienza Università di Roma*[†], *Università degli Studi di Bari “Aldo Moro”*[‡] and *Ministry of Cultural Heritage and Activities and Tourism*[§]

We introduce a Bayesian multivariate hierarchical framework to estimate a space-time model for a joint series of monthly extreme temperatures and amounts of precipitation. Data are available for 360 monitoring stations over 60 years, with missing data affecting almost all series. Model components account for spatio-temporal correlation and annual cycles, dependence on covariates and between responses. Spatio-temporal dependence is modeled by the nearest neighbor Gaussian process (GP), response multivariate dependencies are represented by the linear model of coregionalization and effects of annual cycles are included by a circular representation of time. The proposed approach allows imputation of missing values and interpolation of climate surfaces at the national level. It also provides a characterization of the so called Italian ecoregions, namely broad and discrete ecologically homogeneous areas of similar potential as regards the climate, physiography, hydrography, vegetation and wildlife. To now, Italian ecoregions are hierarchically classified into 4 tiers that go from 2 Divisions to 35 Subsections and are defined by informed expert judgments. The current climatic characterization of Italian ecoregions is based on bioclimatic indices for the period 1955–2000.

1. Introduction. Climate elements and regimes, such as temperature, precipitation and their annual cycles, primarily affect the type and distribution of plants, animals and soils as well as their combination in complex ecosystems (Bailey (2004), Metzger et al. (2013)). The ecological classification of climate represents one of the basic steps for the definition and mapping of ecoregions, that is, of broad ecosystems occurring in discrete geographical areas (Bailey (1983), Loveland and Merchant (2004)). In keeping with these assumptions, a hierarchical classification of Italian ecoregions was recently obtained by combining climatic diagnostic features with distribution patterns of biological diversity and physical characteristics

Received March 2018; revised July 2018.

¹Supported by MIUR (Italian Ministry of Education, University and Scientific Research) (20154X8K23-SH3). This work was partially developed under the PRIN2015 supported project “Environmental processes and human activities: capturing their interactions via statistical methods (EPHStat).”

Key words and phrases. Multivariate process, cyclic effect, coregionalization, NNGP.

(a) $k = 5$ (b) $k = 2$ (c) $k = 3$ (d) $k = 4$

FIG. 1. Ecoregion hierarchical tiers and the climate monitoring network (a).

of the environment (Blasi et al. (2014)). The Italian ecoregions (see Figure 1) are arranged into four hierarchically nested tiers, which consist of two Divisions, seven Provinces, 11 Sections and 35 Subsections (Appendix A). The climatic features adopted for the diagnosis and description of the Italian ecoregions refer to thermo-pluviometric data and bioclimatic indices that date back to the period 1955–2000. The use of long instrumental time series, with a minimum record of at least 30 years, is usually required to place current weather and climate within a historical perspective (WMO (2016)).

The primary focus of this work is the characterization of Italian ecoregions in terms of current and past climatic conditions and involves summarizing climate variables at the ecoregion level, in order to evaluate climate impacts on ecosystems and formulate reliable biodiversity conservation strategies. To these aims several approaches have been proposed (see, e.g., Balint et al. (2011), Hannah et al. (2013), Pesaresi et al. (2014)), often based on interpolated climate surfaces derived from

the WorldClim database. The most recent release of WorldClim was obtained using the work of Fick and Hijmans (2017) that updates an older protocol by Hijmans et al. (2005). Data from between 9000 and 60,000 weather monitoring stations are interpolated using time-independent thin-plate splines with covariates including elevation, distance to the coast and satellite-derived variables (maximum and minimum land surface temperature as well as cloud cover obtained by the MODIS satellite platform). The database includes independent monthly spatial interpolations of temperature (minimum, maximum and average), precipitation, solar radiation, vapor pressure and wind speed at approximately 1 km² spatial resolution for the target temporal range 1970–2000. The authors adopt a multistep procedure to select the best performing model for each WorldClim region and climate variable. Although this solution allows for an improvement in terms of goodness of fit with respect to the previous WorldClim protocol, it does not prevent the known risk of local bias nor does it provide a rigorous assessment of estimates uncertainty (Faye et al. (2014)). As a matter of fact, the uncertainty of estimated climate variables considerably increases in areas characterized by large variation in elevation or with sparse weather monitoring stations.

1.1. *The available data.* Long-term climate data for a network of 360 weather monitoring stations were specifically collected for this work. This network returns a better representation of the geographical and orographic heterogeneity of Italy with respect to the one provided by WorldClim, which includes only 152 stations prevalently located in plain and hilly sectors. We consider monthly records of precipitation and min/max temperature at 360 monitoring stations over 60 years (1951–2010). The data were mostly obtained from National Institutions (ISPRA, CRA/CREA, Meteomont and ENEA) and local authorities (see Appendix B). Monthly records were obtained considering monthly cumulative precipitations and monthly averages of daily minimum and maximum air temperatures. Almost all time series are affected by variable amounts of missing data, as shown in Table 1, reporting summary statistics of the percentage of missing values at each monitoring station. The observed climate variables vary consistently with the 35 ecoregional Subsections approximately ordered from North to South, as is shown in Figure 2.

TABLE 1
Summary statistics of the percentage of missing values at each monitoring station

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Prec.	0.00	1.88	7.64	11.50	18.47	68.33
T. min	0.00	5.45	13.47	16.52	24.17	96.25
T. max	0.00	5.42	13.47	16.50	24.31	96.25

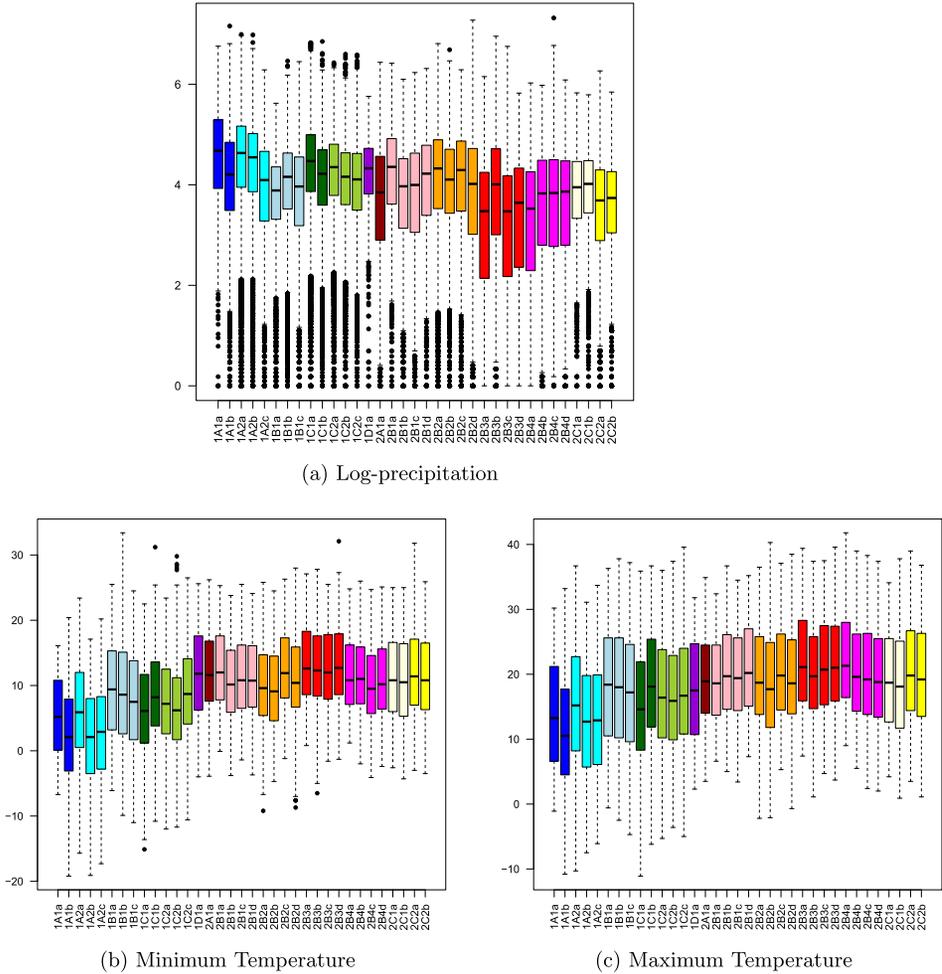


FIG. 2. Boxplots of precipitation, minimum and maximum temperature by ecoregions. Colors follow the Italian ecoregional Sections, see Figure 1 and Appendix A for boxplot labels.

1.2. *Spatio-temporal interpolation.* The secondary objective of our work is climate mapping with long monthly time series dating back to 1950s and characterized by variable amounts of missing data. We address this issue by a fully model-based approach, relying on a stochastic model that accounts for some fundamental features of the multivariate spatio-temporal field that generates the data, that is, correlation among climate variables and space-time variability. Estimation is embedded in the Bayesian hierarchical modeling framework that allows control over various sources of uncertainty. It is commonly assumed that spatio-temporal data are generated by a Gaussian process (GP) and that the covariance function captures space-time dependencies (Gelfand et al. (2010) and references therein). While the richness and flexibility of spatio-temporal stochastic process models

are indisputable, their computational feasibility and implementation pose some challenges for *large* datasets. The implementation of a fully likelihood-based estimation algorithm for a multivariate space-time GP model would imply storing and inverting the covariance matrix of the entire multivariate process. In our case this matrix has order $3 \times 12 \times 60 \times 360 = 7.776 \times 10^5$, then its storage is hardly feasible even for a computer cluster, further it is also numerically impossible to invert since rounding errors would likely cause the matrix to be ill-conditioned for practical purposes. In double precision each element occupies 8 bytes, then the covariance matrix for the data at hand would require about 269 Gb to be stored. As for matrix inversion, consider that a state of the art pc with 16 Gb RAM takes 0.002 seconds to invert a covariance matrix of order 100 using the `solve` function in *R* (R Core Team (2017)). It takes 1.2 seconds for an order 1000 covariance matrix and jumps to 8.7 seconds just doubling the matrix order to 2000. By simple cubic spline extrapolation we obtain that inverting an order 10^5 covariance matrix would require approximately 11 days (if that matrix could be stored)!

A growing literature on the interpolation of large spatial and spatio-temporal data is currently available, with some review papers (Heaton et al. (2017), Li et al. (2016), Jona Lasinio, Mastrantonio and Pollice (2013)) and books (Banerjee, Carlin and Gelfand (2015)) to which the interested reader is referred for details. Approaches for modeling purely spatial covariance matrices avoiding bottlenecks and computational issues due to the data size include low rank and covariance tapering models (Banerjee, Carlin and Gelfand (2015) and references therein), approximation of Gaussian Markov Random Fields (GMRF) by the Laplace transform and Stochastic Partial differential Equations (Blangiardo and Cameletti (2015)), products of lower dimensional conditional densities (see Datta et al. (2016a) and references therein) and composite likelihood estimation methods (Eidsvik et al. (2014)). Recently, multivariate extensions of covariance tapering and composite likelihood methods were proposed by Bevilacqua et al. (2016a) and Bevilacqua et al. (2016b). Examples of univariate spatio-temporal settings include the work of Finley, Banerjee and Gelfand (2012) who introduce dynamic low-rank spatio-temporal processes and Xu, Liang and Genton (2015) who adopt a GMRF-based approach. Spatio-temporal process models are usually defined with discrete time parameters, but the computational feasibility of models continuous in both space and time has also received some attention by Bai, Song and Raghunathan (2012) and Bevilacqua et al. (2012) who proposed to use composite likelihood methods for parameter estimation. A proposal to model large data continuous in both space and time was recently provided by Datta et al. (2016b) who extend the definition of the Nearest Neighbor Gaussian Process (NNGP) approximation for spatial data given in Datta et al. (2016a), the latter outperforming competitive state of the art approaches such as predictive processes (Banerjee et al. (2008)).

In this work we are going to introduce some relevant novelties in the application of the approach proposed by Datta et al. (2016b) to our multivariate spatio-temporal model setting and use a generalized NNGP approximation for the multiresponse interpolation of climate variables. To this aim, we combine the NNGP

with the linear model of coregionalization (Gelfand et al. (2004)) and with a circular representation of time (see for instance Shirota and Gelfand (2017)) that allows to incorporate the effects of annual cycles. Bayesian estimation offers a convenient framework for multivariate interpolation with continuous space-time data, leading to computationally feasible tools, a direct missing data imputation strategy and accurate evaluation of estimate and prediction uncertainties. The proposed approach is applied to the characterization of Italian ecoregions in terms of min/max temperature and precipitation and produces a criterion to choose the “best” ecoregion hierarchical classification tier as part of the model definition. Three model settings and four competing modeling approaches, covering some of the current literature on climate mapping with data from sparse monitoring networks, are then compared in terms of model prediction error and continuous ranked probability scores (CRPS). Comparison shows that the proposed multivariate spatio-temporal model outperforms all other approaches with the available data.

The paper is organized as follows. Section 2 is devoted to the definition of the multivariate coregionalization model for the Italian data, while Section 3 contains the NNGP definition and some details on the implementation of parameter estimation and random effects prediction. Results of estimation, prediction and model comparison are reported and commented on in Section 4, while Section 5 contains some final remarks and addresses possible future developments.

2. The model. Let $s \in \mathcal{S} \subset \mathbb{R}^d$, with $d = 2$, and $t \in \mathcal{T} \subset \mathbb{R}$ be spatial and temporal coordinates respectively, and let $Y_1^*(s, t)$, $Y_2^*(s, t)$ and $Y_3^*(s, t)$ represent the precipitation level, minimum and maximum temperature observed at (s, t) . Then these variables have the following constraints: $Y_1^*(s, t) \geq 0$ and $Y_3^*(s, t) \geq Y_2^*(s, t)$. To simplify modeling and computations, we prefer to work with latent variables defined over the entire real line \mathbb{R} , embedding the above constraints in the variable definitions. Latent variables $Y_1(s, t)$, $Y_2(s, t)$ and $Y_3(s, t)$ are defined as follows:

$$\begin{cases} Y_1(s, t) = Y_1^*(s, t) & \text{if } Y_1^*(s, t) > 0, \\ Y_1(s, t) \leq 0 & \text{if } Y_1^*(s, t) = 0, \end{cases}$$

$$Y_2(s, t) = Y_2^*(s, t),$$

$$\begin{cases} Y_3(s, t) = Y_3^*(s, t) - Y_2^*(s, t) & \text{if } Y_3^*(s, t) - Y_2^*(s, t) > 0, \\ Y_3(s, t) \leq 0 & \text{if } Y_3^*(s, t) - Y_2^*(s, t) = 0. \end{cases}$$

Each latent response Y_i , $i = 1, 2, 3$ is described by a combination of fixed and random terms:

$$(1) \quad Y_i(s, t) = \mathbf{X}(s)\boldsymbol{\beta}_{i,z_k(s)} + \omega_i(s, t) + \lambda_i(s, t) + \varepsilon_i(s, t)$$

with $\varepsilon_i(s, t) \stackrel{\text{iid}}{\sim} N(0, \sigma_{\varepsilon,i}^2)$. Here $\mathbf{X}(s) = (1, X(s))$ and $X(s)$ is the elevation of site s . The integer valued indicator $z_k(s) \subset \mathbb{Z}^+$ is the ecoregion label for the

k th ecoregion tier: with $k = 1$ we have one ecoregion covering the entire country, while $k = 5$ returns the finer classification with 35 ecoregions. In general, $z_1(\mathbf{s}) = 1$, $z_2(\mathbf{s}) \in \{1, 2\}$, $z_3(\mathbf{s}) \in \{1, 2, \dots, 7\}$, $z_4(\mathbf{s}) \in \{1, 2, \dots, 13\}$ and $z_5(\mathbf{s}) \in \{1, 2, \dots, 35\}$. Then $\beta_{i,z_k} = (\beta_{0,i,z_k}, \beta_{1,i,z_k})'$ are regression coefficients, varying with the ecoregion.

Let $h_{\text{space}} = \|s - s'\|$ and $h_{\text{time}} = |t - t'|$ indicate, respectively, spatial and temporal distances and let $h_{\text{circle}} = h_{\text{time}} \bmod L$ be a circular distance with period $L = 1$ year. The term $\lambda_i(\mathbf{s}, t)$ describes the monthly effect of the annual cycle, assumed to act independently on the three processes, that is, $\lambda_i(\mathbf{s}, t) \perp \lambda_{i'}(\mathbf{s}, t)$ for $i \neq i'$, and to follow a GP:

$$(2) \quad \lambda_i(\mathbf{s}, t) \sim \text{GP}(0, \sigma_{c,i}^2 \exp(-\phi_{c,i} h_{\text{circle}}) \mathbf{1}_0(h_{\text{space}})), \quad i = 1, 2, 3,$$

where $\mathbf{1}_0(h_{\text{space}}) = 1$ when $h_{\text{space}} = 0$ and 0 otherwise.

We define the random vector $\omega(\mathbf{s}, t) = (\omega_1(\mathbf{s}, t), \omega_2(\mathbf{s}, t), \omega_3(\mathbf{s}, t))'$ as

$$(3) \quad \omega(\mathbf{s}, t) = \mathbf{A}\mathbf{w}(\mathbf{s}, t),$$

where \mathbf{A} is a 3×3 matrix described below and $\mathbf{w}(\mathbf{s}, t) = (w_1(\mathbf{s}, t), w_2(\mathbf{s}, t), w_3(\mathbf{s}, t))'$ is a vector of zero mean GPs with independent components and spatio-temporal correlation function $C(h_{\text{space}}, h_{\text{time}}, \theta_i)$, with $w_i(\mathbf{s}, t) \perp \lambda_{i'}(\mathbf{s}, t)$ for all values of (\mathbf{s}, t) , i and i' . Since $\omega(\mathbf{s}, t)$ is a linear combination of independent GPs, it is a multivariate GP with dependent components. The functional form of $C(h_{\text{space}}, h_{\text{time}}, \theta_i)$ is given by the general nonseparable space-time correlation structure proposed by Gneiting (2002) in his equation (14), that is,

$$(4) \quad C(h_{\text{space}}, h_{\text{time}}; \theta_i) = \frac{1}{(\phi_{t,i} h_{\text{time}}^{2\alpha_i} + 1)^\tau} \exp\left(-\frac{\phi_{sp,i} h_{\text{space}}^{2\gamma_i}}{(\phi_{t,i} h_{\text{time}}^{2\alpha_i} + 1)^{\eta_i \gamma_i}}\right).$$

Nonnegative scaling parameters $\phi_{t,i}$ and $\phi_{sp,i}$ are associated to time and space respectively, smoothness parameters α_i and γ_i take values in $(0, 1]$, the space-time interaction parameter η_i ranges in $[0, 1]$ and $\tau \geq d/2$. Following Gneiting (2002), we set $\tau = 1$, $\alpha = 1$ and $\gamma = 0.5$. Attractively, as η_i decreases towards zero, we achieve separability in space and time. Now, letting $\mathbf{T}_i = \mathbf{a}_i \mathbf{a}_i'$, where \mathbf{a}_i is the i th column of \mathbf{A} , the covariance matrix for the process ω at different times and locations is given by:

$$\text{Cov}(\omega(s_l, t_l), \omega(s_q, t_q)) = \sum_{i=1}^3 \mathbf{T}_i C(\|s_l - s_q\|, |t_l - t_q|; \theta_i)$$

with $\mathbf{A}\mathbf{A}' = \text{Cov}(\omega(\mathbf{s}, t)) = \Sigma, \forall(\mathbf{s}, t)$. Remark that different specifications of matrix \mathbf{A} in equation (3) can define the same covariance matrix Σ , with specific consequences on the process structure (Gelfand et al. (2004)). Hence a careful definition of matrix \mathbf{A} is required. A popular choice is the Cholesky decomposition of the symmetric matrix Σ that produces a lower diagonal matrix. In this setting this

decomposition induces an artificial ordering of the response variables, given that the correlation structure of ω_1 depends only on $C(\cdot, \cdot; \theta_1)$, the one of ω_2 depends on $C(\cdot, \cdot; \theta_1)$ and $C(\cdot, \cdot; \theta_2)$, while the correlation of ω_3 depends on $C(\cdot, \cdot; \theta_1)$, $C(\cdot, \cdot; \theta_2)$ and $C(\cdot, \cdot; \theta_3)$. To avoid this artifact, we propose to decompose Σ with a different approach: let $\Xi = \text{diag}(\xi_1, \xi_2, \xi_3)$ be the diagonal matrix of the square rooted eigenvalues of Σ and Ψ be the orthogonal matrix of its eigenvectors, such that $\Psi' \Psi = \mathbf{I}$, we then let $\mathbf{A} = \Psi \Xi \Psi'$. Such matrix \mathbf{A} is symmetric by construction and its elements do not depend on the ordering of the eigenvalues. Assume that \mathbf{D} is a 3×3 matrix that changes the ordering of the elements of $\omega(\mathbf{s}, t)$. Then the covariance matrix of $\mathbf{D}\omega$ is $\mathbf{D}\Sigma\mathbf{D}'$ with the columns of matrix $\mathbf{D}\Psi$ as eigenvectors. Now let $\mathbf{D}\Sigma\mathbf{D}' = \mathbf{A}_* \mathbf{A}'_*$, then $\mathbf{A}_* = \mathbf{D}\Psi \Xi \Psi' \mathbf{D}' = \mathbf{D}\mathbf{A}\mathbf{D}'$, proving that \mathbf{A}_* has the same values of \mathbf{A} but arranged according to the reordering matrix \mathbf{D} .

For ease of interpretation, remark that differences between ecoregions are specified by the regressive part, spatial and temporal dependence is ruled by $\omega_i(\mathbf{s}, t)$, the seasonal component is described in $\lambda_i(\mathbf{s}, t)$, the correlation between responses is defined by \mathbf{A} and nonstructured effects are summarized in $\varepsilon_i(\mathbf{s}, t)$.

3. NNGP. For a parsimonious implementation of the model estimation algorithm, we define $\omega_i^*(\mathbf{s}, t) = \omega_i(\mathbf{s}, t) + \lambda_i(\mathbf{s}, t)$ and use this new variable in what follows. Letting N be the number of observations in space and time, we denote their locations by (s_n, t_n) , $n = 1, \dots, N$ and we let $\omega_n^* = (\omega_1^*(s_n, t_n), \omega_2^*(s_n, t_n), \omega_3^*(s_n, t_n))'$ with $\omega^* = (\omega_1^*, \dots, \omega_N^*)'$. If $f(\cdot)$ is a generic Gaussian density function, then the joint finite dimensional distribution of the whole observed GP is given by

$$(5) \quad f(\omega^*) = \prod_{n=1}^N f(\omega_n^* | \omega_{n-1}^*, \dots, \omega_1^*)$$

with $\omega_0^* = \emptyset$. Notice that, though there is no univocal definition of a space-time ordering of observed locations, (5) is a valid representation of the joint density for any given ordering.

Let $\Omega_n = (\omega_{n-1}^*, \dots, \omega_1^*)'$ be the conditional set of ω_n^* in (5) and let $\Omega_n(m) \subseteq \Omega_n$ be a subset that contains at most m elements of Ω_n . With the NNGP, the joint finite dimensional distribution in (5) is approximated by

$$\prod_{n=1}^N f(\omega_n^* | \Omega_n(m)).$$

As shown by Datta et al. (2016a), the quality of the approximation increases with m and the best results are achieved if we choose the m elements of Ω_n that have the higher correlation with ω_n^* . To implement the NNGP, three decisions are required:

- how to order the observations;
- how to choose the value of m ;
- how to choose the elements of $\Omega_n(m) \subseteq \Omega_n$.

The ordering. A natural ordering is immediately available for the time dimension, but there is not a unique way to order observations in space at a given time. The way we order spatial locations has a strong influence on the definition of how candidate locations enter $\Omega_n(m)$ (see also [Katzfuss and Guinness \(2017\)](#) and [Guinness \(2018\)](#) for some theoretical implications of the neighbor choice and alternative proposals, as the MMD ordering). Here we follow [Datta et al. \(2016a\)](#) and order locations first according to the longitude coordinate, then according to latitude. This ensures that Ω_n includes observations spatially and temporally close to ω_n^* .

The value of m . Compared to the size of the problem, the number of neighbors m should be small in order to obtain a computational gain. [Datta et al. \(2016b\)](#) showed that, assuming that the elements of $\Omega_n(m)$ are “close enough” (correlated or geographically close) to ω_n^* , choosing $m \in \{10, \dots, 20\}$ produces an approximation almost indistinguishable from the original process.

The elements of $\Omega_n(m)$. Again [Datta et al. \(2016b\)](#) suggest that the best choice for $\Omega_n(m)$ is to take the m elements that have higher correlation with ω_n^* . In a univariate purely spatial or temporal setting, where correlation decreases with the distance, the optimal choice for the elements of $\Omega_n(m)$ would consider observations spatially/temporally closer to ω_n^* . In a spatio-temporal setting with nonseparable correlation function there is not a one to one correspondence between distance and correlation, since a spatio-temporal distance is not uniquely defined. In a univariate spatio-temporal setting, [Datta et al. \(2016b\)](#) propose an adaptive approach in which $\Omega_n(m)$ is defined at each MCMC iteration as the set that has the higher correlation with ω_n^* . The choice of $\Omega_n(m)$ based on correlation would imply considering all possible sets of m neighbors at each point for each MCMC iteration. In this work we prefer not to follow this approach since it is not efficient in terms of parameter estimation and computational time. We apply the NNGP approximation to the GP ω^* whose covariance function is given by the sum of ω 's and λ 's covariances. The numerator of the correlation function needed for the implementation of Datta's adaptive neighbor is $\sigma_{c,i}^2 C_{\lambda,i} + \sigma_{\omega,i}^2 C_{\omega,i}$, where $C_{\lambda,i}$ is the correlation function of the monthly effect and $C_{\omega,i}$ is the Gneiting correlation function. Notice that if $\sigma_{c,i}^2 \gg \sigma_{\omega,i}^2$, and given that $C_{\lambda,i}$ is equal to zero if $h_{\text{space}} \neq 0$, observations at different spatial locations will have low correlation and will not be chosen as neighbors, while they are required to learn about the spatial dependence structure. Paradoxically, in this case, we would end up to choose only time neighbors. If we choose the number of neighbors separately for the two components ω and λ , we would have to simulate both components at all times and sites, with a considerable increase in the computational effort.

Hence we propose the following. We define a spatio-temporal distance as

$$(6) \quad \sqrt{\left(h_{\text{space}} \frac{1}{150}\right)^2 + h_{\text{time}}^2},$$

and we include in $\Omega_n(m)$ the m locations with smaller distances from ω_n . Notice that, as the spatial and temporal dimensions have different scales, we adjust the spatio-temporal distance heuristically, assuming that one year has the same weight as 150 kms. Our choice is justified by the following considerations: in each neighborhood we want to include information on the spatial dependence, the time dependence and the cross-correlation structure, furthermore we need information on the annual cyclical component. Equation (6) ensures that the generic point (s_n, t_n) has approximately m neighbors, \sqrt{m} of which are observed at the same time and at different locations, \sqrt{m} share the same spatial location and are observed at different times and the remaining are observed at different times and locations. Furthermore, in order to learn about the annual cyclical component, we may have to modify the points at the boundaries of $\Omega_n(m)$. This is done in such a way that, for example, the neighborhood of location $(s, \text{January } 2000)$ includes location $(s, \text{January } 1999)$ and $(s, \text{February } 1999)$.

3.1. *Implementation details.* In Section 2 we assumed $E(\omega^*) = \mathbf{0}_N$, then using standard results from the multivariate normal theory, we can write

$$f(\omega_n^* | \Omega_n(m)) = \phi_3(\omega_n^* | \mathbf{B}_n \Omega_n(m), \mathbf{F}_n),$$

where $\phi_3(\omega_n^* | \mathbf{B}_n \Omega_n(m), \mathbf{F}_n)$ is the 3-variate normal distribution with mean $\mathbf{B}_n \Omega_n(m)$ and covariance matrix \mathbf{F}_n . Parameters \mathbf{B}_n and \mathbf{F}_n depend on the Gneiting and circular correlation function parameters, on Σ and on the distances between the spatio-temporal locations in $(\omega_n, \Omega_n(m))$.

Our data are observed over 360 spatial locations, and each point is observed for 720 times, that is, 12 months \times 60 years. Given the set of m values we explored ($m = 10, 15, 20$), the maximum temporal distance between ω_n^* and the elements of $\Omega_n(m)$ is equal to one year, with the exception of the first 12 months in the database that have a maximum distance of less than one year. Starting from the 13th time-point onwards, the parameters \mathbf{B}_n and \mathbf{F}_n at the n th spatial location are the same, since they depend on the Gneiting and circular correlation function parameters, on Σ and on the temporal distances between locations in $(\omega_n^*, \Omega_n(m))$, that do not change. Then we only need to compute \mathbf{B}_n and \mathbf{F}_n for the first 13 times and 360 spatial locations, thus obtaining a huge computational gain. Notice that in this setting also the computation of the full conditionals of ω_n is simplified as it implies only a time window of two years at each spatial location. Given that after the first 12 months the distances between time points start to repeat, we need to compute only the full conditionals for the first 13 months and the last 12 months, as from the 13th to the $(720-12)$ th month we have a full constant 24 months time window.

Grid prediction. Let $\mathbf{Y}_0 = (\mathbf{Y}(s_0, t_1), \dots, \mathbf{Y}(s_0, t_{720}))'$ be the 3-variate time series of Y 's at a spatial grid point and let define $\Omega_{y,j} = (\mathbf{y}_{0,j-1}, \dots, \mathbf{y}_{0,1}, \mathbf{y})$ as the conditioning set of $\mathbf{y}_{0,j}$, where $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_N)'$ is the set of all data

from the monitoring network and $\mathbf{y}_{0,j} = (y_1(\mathbf{s}_0, t_j), y_2(\mathbf{s}_0, t_j), y_3(\mathbf{s}_0, t_j))'$, with $j \in \{1, \dots, 720\}$. In a similar vein, we define $\Omega_{\mathbf{y},j}(m)$ as the set of m nearest neighbors of $\mathbf{y}_{0,j}$, based on the distance (6). Notice that, since $\Omega_{\mathbf{y},j}$ contains all sampled spatio-temporal locations, the set of m nearest neighbors $\Omega_{\mathbf{y},j}(m)$ can contain temporal indexes that are even higher than t_j , for example, in the set of neighbors of the first time of a grid point, there can be points in the second or third time. We want to obtain samples of \mathbf{Y}_0 from the predictive density

$$(7) \quad f(\mathbf{y}_0|\mathbf{y}^O) = \int \prod_{j=1}^{720} f(\mathbf{y}_{0,j}|\Omega_{\mathbf{y},j}, \boldsymbol{\theta}) f(\mathbf{y}^M, \boldsymbol{\theta}|\mathbf{y}^O) d\boldsymbol{\theta} d\mathbf{y}^M,$$

where \mathbf{y}^M and \mathbf{y}^O are subsets of \mathbf{y} composed of, respectively, missing and observed data, $\boldsymbol{\theta}$ contains all model parameters and $f(\mathbf{y}^M, \boldsymbol{\theta}|\mathbf{y}^O)$ is the posterior distribution. Our interest is also in the prediction of the annual cyclical component $\boldsymbol{\lambda}_0 = (\boldsymbol{\lambda}_{0,1}, \dots, \boldsymbol{\lambda}_{0,12})$, where $\boldsymbol{\lambda}_{0,\ell} = (\lambda_1(s_0, \ell), \lambda_2(s_0, \ell), \lambda_3(s_0, \ell))'$, with $\ell = 1, \dots, 12$. We then sample from the following predictive density:

$$(8) \quad f(\boldsymbol{\lambda}_0|\mathbf{y}^O) = \int f(\boldsymbol{\lambda}_0|\mathbf{y}_0, \mathbf{y}, \boldsymbol{\theta}) \prod_{j=1}^{720} f(\mathbf{y}_{0,j}|\Omega_{\mathbf{y},j}, \boldsymbol{\theta}) f(\mathbf{y}^M, \boldsymbol{\theta}|\mathbf{y}^O) d\boldsymbol{\theta} d\mathbf{y}^M d\mathbf{y}_0.$$

The density $f(\mathbf{y}_{0,j}|\Omega_{\mathbf{y},j}, \boldsymbol{\theta})$ in (7) and (8) is a trivariate normal, but as with equation (5), estimation of its parameters requires the computation/inversion of a covariance matrix of dimension $N + j$ (Banerjee, Carlin and Gelfand (2015)). We then use the NNGP to approximate the predictive densities and substitute $\Omega_{\mathbf{y},j}(m)$ to $\Omega_{\mathbf{y},N}$ in both expressions. After model fitting, posterior samples from (7) and (8) can be obtained using standard Monte Carlo integration.

4. Results and discussion. Notice that the response variables were rescaled and standardized to improve the efficiency of the MCMC estimation algorithm. In order to preserve the information about the zeroes, the monthly precipitation amount was simply rescaled by its standard deviation while Y_2 and Y_3 were both standardized. Results are presented according to the model (transformed) scale, except for the model-choice indices and Figure 3.

4.1. *Model choice and posterior estimates.* We estimated nine different models, varying the number of neighbors in the NNGP and the ecoregional hierarchical tier: $m = \{10, 15, 20\}$ and $k \in \{3, 4, 5\}$, respectively. Weakly informative priors were used throughout, namely $N(0, 100)$ for regression parameters, $IG(1, 1)$ for all variances, $IW(3, \mathbf{I})$ for $\boldsymbol{\Sigma}$, $U(0, 1)$ for the separability parameter, $\phi_{sp} \sim U(0.05, 6)$ and $\phi_t, \phi_c \sim U(3, 36)$ that correspond to practical ranges in the intervals [6, 60] km and [1, 12] months, respectively. The MCMC was implemented with 100,000 iterations, a burn-in phase of 70,000 and thinning by 12, keeping 2500 samples for posterior inferences. Posteriors estimates were obtained in about three days and

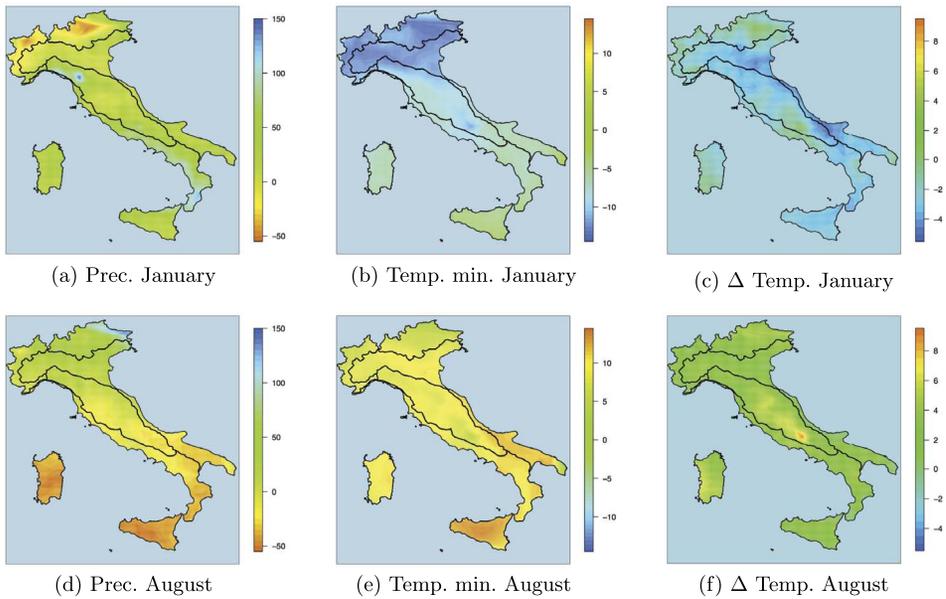


FIG. 3. Maps of the monthly effects of January (a), (b), (c) and August (d), (e), (f) on the annual cycles of the three components of the process.

were implemented on the TeraStat cluster (Ferraro Petrillo and Raimato (2014)) that allows for fast computing with a limitation on the number of processes that can be launched simultaneously.

Due to conjugacy, the MCMC algorithm samples regression coefficients, nuggets, missing data, spatio-temporal and circular processes using Gibbs steps, that is, from their full conditionals, while the other parameters are sampled all together with a Metropolis step using the adaptive proposal of Andrieu and Thoms (2008), algorithm 4.

The choice among alternative specifications of the same model was performed using the Deviance Information Criterion (DIC) (Spiegelhalter et al. (2002)) and the CRPS (Grimm et al. (2006)) reported in Tables 2 and 3. Notice that while DICs are referred to the multivariate response, CRPSs are obtained for each of the three process components. As expected, the largest number of neighbors always returns the smallest DIC value for a given k , while CRPS has a different behavior for the precipitation and temperature. Precipitation prefers more local features ($m = 10$ versus $m = 20$) than temperatures. We choose the model structure suggested by the majority of criteria that corresponds to $k = 3$ and $m = 20$. The “best” model suggests to aggregate ecoregions into seven distinct Provinces (see Figure 1), that is, the most aggregated and general ecoregional tier among those considered with the model implementation. This result is consistent with a principle widely adopted by hierarchical approaches to the ecological classification of land. This basic principle states that climate acts as a primary environmental factor in determining the

TABLE 2
Model choice, DIC values for different choices of the hierarchical ecoregional tier (k) and neighborhood size (m) in the NNGP approximation. DIC values are to be read as $\times 10^4$

		<i>m</i>		
		10	15	20
<i>k</i>	3	459.35	343.49	336.21
	4	479.78	451.81	385.38
	5	679.39	553.75	543.26

broad-scale ecosystem variation, while factors such as geomorphology and soil features assume an equal or greater importance than climate at local spatial aggregation levels (Bailey (2004), Múcher et al. (2010)). In the following, we are going to report on parameter estimates and predictions obtained with the chosen model.

TABLE 3
CRPS values for different choices of the hierarchical ecoregional tier (k) and neighborhood size (m) in the NNGP approximation

<i>Y</i> ₁		<i>m</i>		
		10	15	20
<i>k</i>	3	19.366	19.315	19.872
	4	18.718	19.541	20.510
	5	19.570	19.189	19.820

<i>Y</i> ₂		<i>m</i>		
		10	15	20
<i>k</i>	3	0.423	0.406	0.400
	4	0.429	0.402	0.411
	5	0.426	0.402	0.411

<i>Y</i> ₃		<i>m</i>		
		10	15	20
<i>k</i>	3	0.475	0.460	0.456
	4	0.484	0.458	0.464
	5	0.487	0.461	0.477

TABLE 4
Posterior estimates of the GP and annual cyclical component parameters, as in equations (4) and (2), respectively

		ϕ_{sp}	ϕ_t	ϕ_c	η
Y_1	Est.	0.188	28.979	15.210	0.774
	(CI)	(0.184 0.192)	(28.871 29.072)	(14.972 15.394)	(0.774 0.775)
Y_2	Est.	0.138	9.628	10.176	0.943
	(CI)	(0.137 0.140)	(9.476 9.750)	(10.102 10.239)	(0.942 0.943)
Y_3	Est.	0.431	23.814	9.760	0.166
	(CI)	(0.429 0.432)	(23.576 23.995)	(9.690 9.835)	(0.165 0.168)

		σ_c^2	σ_ε^2	σ_ω^2
Y_1	Est.	0.617	0.176	0.413
	(CI)	(0.612 0.623)	(0.175 0.178)	(0.409 0.416)
Y_2	Est.	6.968	0.008	0.050
	(CI)	(6.830 7.092)	(0.008 0.008)	(0.050 0.051)
Y_3	Est.	2.799	0.062	0.525
	(CI)	(2.647 2.919)	(0.061 0.062)	(0.519 0.532)

Posterior estimates of the GP and annual cyclical component parameters (see equations (2) and (4)) are reported in Table 4 with their 95% credible intervals (CI). In Table 5 we show the proportion of variance due to the space-time ($\sigma_{\omega,i}^2$ is the i th element of the diagonal of Σ), the cyclical and the residual component for each response variable, in order to appreciate the relevance of each of the three components in explaining the total variation. It is worth noticing that for the minimum temperature Y_2 almost the entire variation can be ascribed to the cyclical component, while for the thermal excursion Y_3 only 15.5% is due to the spatio-temporal term and a negligible contribution comes from the residual part. The precipitation Y_1 has a different behavior: a large portion of the total variation (51%) is due to the cyclical component, the space-time dynamic accounts for 34.2%, while

TABLE 5
Proportions of the space-time, cyclical and residual components of the variance for each climate variable

	σ_ω^2	σ_c^2	σ_ε^2
Y_1	0.342	0.512	0.146
Y_2	0.007	0.992	0.001
Y_3	0.155	0.827	0.018

a quite relevant 14.6% of the total variation is left unexplained. Compatibly with the physics of the phenomena, precipitation dynamics seem to be more affected by small scale events with respect to Y_2 and Y_3 , causing higher residual variability.

Table 4 shows that the three climate variables have nonseparable space-time dynamics, as CIs for the η parameter are never close to 0. Practical ranges and covariances of the three components in (1) provide useful information on the extent of the spatial, temporal and annual cyclical dependence. The spatial practical ranges of Y_1 , Y_2 and Y_3 are respectively 15.95 km, 21.676 km and 6.967 km, suggesting a less homogeneous spatial behavior of the temperature range with respect to the other two variables. In terms of time dependence, we have the following practical ranges: 37.78 days, 113.73 days and 45.98 days for Y_1 , Y_2 and Y_3 , respectively. These values highlight a similar extension of the temporal correlation for the precipitation and the thermal excursion. Finally, as expected, the annual cyclical effect ϕ_c has similar behavior for the second and third variable, with practical ranges 71.99 days (Y_1), 107.60 days (Y_2) and 112.19 days (Y_3): annual cycles are longer and almost seasonal (four months long, as expected) for the minimum temperature and the temperature range, while a shorter cycle is estimated for the precipitation. Measures of the correlation between climate variables are also obtained and are all far from zero. Precipitation and minimum temperature are positively correlated (0.210 with 95% CI (0.208, 0.213)), while precipitation and temperature range are negatively correlated (-0.214 with 95% CI ($-0.218, -0.211$)). As expected, minimum temperature and temperature range are negatively correlated with a stronger relation (-0.493 with 95% CI ($-0.499, -0.497$)).

In Tables 6 and 7 the posterior estimates of the intercepts and regression coefficients for each of seven ecoregion Provinces are reported with their 95% CIs. All estimates for the fourth Province (1D) are not relevant and very different from the other values, due to the presence of only one monitoring station in the given area. This suggests to aggregate the fourth Province to one of its neighbors for future investigations, as recently tested in the first report on the Italian natural capital.² Estimates of the model intercepts β_0 's allow to analyze the average behavior of each component in the specific Province. The only estimate that shows a value close to zero is for the temperature range in ecoregion 1B, the Po Plain Province, suggesting a very small temperature range in that area. All Provinces are well characterized with some overlapping of CIs for each variable, suggesting similarities between areas. A similar behavior in terms of precipitation (Y_1) can be found in Provinces 1B, 2A, 2B and 2C ($j = 2, 5, 6, 7$), while 1A, 1B and 2C ($j = 1, 2, 7$) show similarities in terms of minimum temperature (Y_2) and only 1A and 2C ($j = 1, 7$) show interval estimates overlapping for the temperature range

²Italian Natural Capital Committee (INCC), 2017. 1st Report on the State of Natural Capital in Italy (synthesis). Available at: http://www.minambiente.it/sites/default/files/archivio/allegati/sviluppo_sostenibile/sintesi_raccomandazioni_primo_rapporto_capitale_naturale_english_version.pdf.

TABLE 6

Point estimates of the intercepts at each Province for each GP component and relative 95% CIs. Provinces are coded as follows 1 = 1A, 2 = 1B, 3 = 1C, 4 = 1D, 5 = 2A, 6 = 2B and 7 = 2C

		$\beta_{0,1}$	$\beta_{0,2}$	$\beta_{0,3}$	$\beta_{0,4}$
Y ₁	Est.	1.348	0.741	0.965	-0.050
	(CI)	(1.280 1.400)	(0.668 0.794)	(0.924 1.003)	(-8.155 7.777)
Y ₂	Est.	0.207	0.225	0.094	-0.005
	(CI)	(0.189 0.234)	(0.207 0.240)	(0.084 0.106)	(-1.746 1.758)
Y ₃	Est.	0.155	0.008	0.779	0.083
	(CI)	(0.055 0.201)	(-0.069 0.097)	(0.737 0.844)	(-4.919 4.748)

		$\beta_{0,5}$	$\beta_{0,6}$	$\beta_{0,7}$
Y ₁	Est.	0.633	0.740	0.783
	(CI)	(0.520 0.787)	(0.710 0.774)	(0.745 0.838)
Y ₂	Est.	0.878	0.457	0.277
	(CI)	(0.838 0.910)	(0.445 0.464)	(0.263 0.295)
Y ₃	Est.	-1.965	-0.070	0.196
	(CI)	(-2.038 -1.856)	(-0.125 -0.038)	(0.127 0.264)

TABLE 7

Point estimates of the regression coefficients of the elevation (multiplied by 10³) at each Province for each GP component and relative 95% CIs. Provinces are coded as follows 1 = 1A, 2 = 1B, 3 = 1C, 4 = 1D, 5 = 2A, 6 = 2B and 7 = 2C

		$\beta_{1,1}$	$\beta_{1,2}$	$\beta_{1,3}$	$\beta_{1,4}$
Y ₁	Est.	0.029	0.498	0.248	63.022
	(CI)	(-0.022 0.068)	(0.189 0.742)	(0.208 0.294)	(-650.237 788.956)
Y ₂	Est.	-0.747	-0.322	-0.452	73.423
	(CI)	(-0.761 -0.729)	(-0.414 -0.242)	(-0.470 -0.437)	(-86.057 231.301)
Y ₃	Est.	-0.366	-0.094	-1.073	-117.311
	(CI)	(-0.428 -0.305)	(-0.648 0.423)	(-1.177 -1.023)	(-544.667 338.024)

		$\beta_{1,5}$	$\beta_{1,6}$	$\beta_{1,7}$
Y ₁	Est.	-0.532	0.548	0.416
	(CI)	(-1.250 0.241)	(0.502 0.592)	(0.272 0.546)
Y ₂	Est.	-2.802	-0.842	-0.452
	(CI)	(-3.038 -2.541)	(-0.856 -0.826)	(-0.498 -0.405)
Y ₃	Est.	8.256	0.040	-0.736
	(CI)	(6.891 9.607)	(-0.042 0.112)	(-0.985 -0.570)

(Y_3). Notice that Province 1B covers the Po Plain and is a very heterogeneous area where a transition from the continental to the Mediterranean behavior occurs. The relation with the elevation described by the estimates of the regression coefficients β_1 's often admits the zero value in the 95% CI. This is likely linked to the presence of a latitudinal gradient in the area. For example, in the Po Plain Province (1B) a large area is divided by the Po river in a Northern sector with continental regime and a Southern sector with Apennines regime, as already highlighted for the effects of the cyclical components. The Alpine province (1A) is associated to regression coefficients that are all quite far from zero and this can be linked to the absence of a latitudinal gradient, being the region only affected by a longitudinal variation. Moreover the area is characterized by a considerable relief energy (large elevation gradient).

4.2. *Predictions and out of sample validation.* After model fitting, we used posterior samples to predict the values of the responses, and hence the cyclical component of the model $\lambda_i(s, t)$, on a square lattice of 3305 spatial points with 15 km side over 720 time points. Given the limitation on the number of parallel processes that can be launched on the Terastat cluster grid predictions were all implemented on the Bari ReCaS Data Center. ReCaS provides a computing power of 128 servers each with 64 cores and 256 Gb of RAM.

In Figure 3 examples of two maps of predicted monthly effects of the annual cyclical component are reported. The maps for the months of January and August have been chosen as representative of the factors affecting the composition of ecosystems and their distribution over the Italian territory. Above all, these factors include moisture availability in the different seasons, winter cold and summer drought. First, the model was able to show some interesting seasonal patterns of precipitation (Figure 3(a) and 3(d)). These include: (i) the continental regime of the Alpine Province, the only region with larger precipitation values in summer than in winter months; (ii) the transitional character of the Po Plain Province towards a more Mediterranean regime, with lower summer precipitation; (iii) the very clear latitudinal gradient in both the Apennines and peninsular Tyrrhenian Provinces, which mainly reflects the varying distance from the coast of the mountain reliefs. More local patterns are suggested as well, that however need a deeper investigation at the Section and/or Subsection ecoregional levels. These include the longitudinal summer gradient in precipitation between Eastern and Western Alps and the marked summer precipitation decrease in some Southern peninsular and main island sectors. Second, winter cold (Figure 3(b)) clearly characterizes both the Alpine and Po Plain Provinces within the Temperate Division. On the contrary, the variable behavior within the Apennines Province should be further investigated in order to elaborate on the differences with the Tyrrhenian Province of the Mediterranean Region. Patterns that need to be characterized at lower ecoregional levels emerged in this case as well. These include the latitudinal gradient along the Adriatic Province and the differences between the two main Tyrrhenian

islands. Among several features, the third component of the process shows the relevance of reduced winter temperatures and their variation in characterizing the thermic continentality of the Po Plain and Adriatic Provinces. It also confirms that higher temperatures occur in both the Tyrrhenian and the Adriatic Mediterranean Provinces.

To assess the out of sample predictive capability of the proposed modeling approach, we built a validation set by randomly choosing 10% of available spatial observations at each time point (between 24 and 35 validation points for each time) for each variable (overall, 22,628 points for Y_1 , 21,321 points for Y_2 and 21,272 for Y_3). Three settings of the proposed model and five alternative competing modeling approaches for climate mapping with data from sparse monitoring networks were employed to obtain predictions for the validation sets. The predictive capability was measured by the rooted mean squared errors (RMSE) of the hold-out samples. Here we briefly describe the chosen models, while details on the models and implementation settings can be found in Appendix C:

M1 Independent interpolation of each climate variable at each time point by three-variate thin plate splines (TPS) for the spatial coordinates and elevation.

M2 Independent interpolation of each climate variable by a spatio-temporal generalized additive model (GAM) with a decennial three-variate thin plate regression spline (see Wood (2017)) component for the spatial coordinates and elevation, two independent univariate thin plate regression splines terms for smooth effects of years and months and fixed effects of the seven Provinces. Alternative model specifications had higher AIC and RMSE and smaller deviance explained.

M3 Independent interpolation of each climate variable at each time point by a spatial GAM with a bivariate thin plate regression spline on the coordinates and a thin plate regression spline on the elevation term. Provinces were included as fixed effects only in the precipitation model, as they did not seem to affect the temperature spatial behavior at this temporal resolution. Again, model choice was guided by the values of AIC, RMSE and deviance explained.

M4 Independent interpolation of each climate variable at each time point by local linear regression over nearest neighbors (NN) with elevation as explanatory variable. We adopted 10 neighbors for the precipitation and 15 for the temperature. The neighborhood dimension was chosen as the one minimizing the RMSE.

M5 Independent interpolation of each climate variable at each time point by Bayesian Gaussian Kriging with a linear trend on the spatial coordinates and elevation, parameterized as in Diggle and Ribeiro (2002). Specific modeling features are reported in Appendix C and were selected as those producing the smallest RMSE.

M6 Univariate version (uncorrelated responses) of the model in equation (1) without annual cyclical effects.

M7 Univariate version (uncorrelated responses) of the model in equation (1) with annual cyclical effects.

TABLE 8

Rooted mean squared error and CRPS computed over the validation set, in bold the smallest values for each criterion. In italics the Mean Absolute Error computed for the models where CRPS could not be evaluated

	RMSE			CRPS		
	prec (mm)	tmin (C°)	tmax (C°)	prec (mm)	tmin (C°)	tmax (C°)
M1	44.17	1.84	1.69	<i>26.37</i>	<i>1.37</i>	<i>1.22</i>
M2	48.75	1.91	1.85	23.60	1.18	1.04
M3	44.50	1.86	1.79	36.69	3.92	4.67
M4	46.32	1.87	1.83	<i>27.61</i>	<i>1.40</i>	<i>1.28</i>
M5	41.39	1.81	1.78	36.72	3.91	4.67
M6	61.04	0.89	0.93	34.07	0.42	0.72
M7	39.50	0.86	0.91	19.88	0.41	0.47
M8	38.29	0.84	0.89	19.87	0.40	0.46

M8 Multivariate spatio-temporal model in equation (1).

In Table 8 we list RMSEs and CRPS³ for the eight alternatives and the three climatic response variables. As a further benchmark, consider that Fick and Hijmans (2017) report the following cross-validation RMSEs for WorldClim predictions: 49.46 mm for precipitation, 1.40 C° for minimum temperature and 1.30 C° for maximum temperature.

The proposed multivariate spatio-temporal model outperforms all other approaches with the data available in this case. The introduction of model features such as seasonality, nonseparability and correlation between climate response variables considerably helps in improving all considered criteria. Among the selected alternatives, only GAMs include ecoregions as model terms and allow to analyze their characteristics in terms of climate variables, however models M2 and M3 are not satisfactory in terms of prediction error, CRPS and missing data imputation.

5. Concluding remarks and future developments. In this paper we present a multivariate generalization of the NNGP model proposed in Datta et al. (2016b). Our proposal originates from the currently most efficient approach suggested by a sequence of comparative studies. Datta et al. (2016a) compare NNGP based spatial models with predictive process models in terms of computational complexity and estimation performance, concluding in favour of NNGPs. In Datta et al. (2016b) Dynamic NNGPs (DNNGP) are compared to full rank GPs and low rank Gaussian Predictive Processes (Banerjee et al. (2008)) detecting notable improvements of

³Remark that following Gneiting and Raftery (2007), Mean Absolute Errors are computed for the models where CRPS could not be evaluated and are highlighted in italics.

NNGPs performance over low-rank methods. Additional simulation experiments proving a better predictive performance of DNNNGPs over Local Approximation GPs (Gramacy and Apley (2015)) are provided in the supplemental material of the same article. Low-rank models were also considered in Banerjee et al. (2008) and favorably compared to predictive process models in terms of model accuracy, complexity and efficiency. Our proposal combines the computational efficiency of NNGPs with several new ideas for handling complex structures typical of climate data. We use the linear model of coregionalization to account for multivariate spatio-temporal dependencies, a circular representation of the time index to define the annual cyclical term and propose an efficient implementation that allows estimation of model parameters with a large amount of data. Compared to alternative approaches, the explicit consideration of the underlying process features (such as correlation among climate responses, space-time variability and cyclical components) improves the prediction error. In particular these features make up for auxiliary information, such as satellite data as used in WorldClim. Indeed, many auxiliary information sources could be considered, such as MODIS, or TRMM based climatologies and/or observations. Unfortunately none of these goes back far enough to comply with the chosen time window (1950–2010) and allows interpolation of climate data at any desired resolution. The richness of the model output allows to characterize the Italian ecoregions with respect to precipitation, minimum and maximum temperature, returning information on the cyclical trend, spatial and temporal correlation. While the exponential function is almost mandatory when modeling correlation on a circular time scale (Gneiting (2013)), a possible generalization of the current model structure to less smooth phenomena is obtained by changing the values of α and γ in equation (4). The current model settings are well suited for relatively smooth phenomena such as monthly summaries of climatic variables and should be revised if different time aggregations are chosen. Notice that the model architecture is very flexible and can be adapted to a wide variety of meteo-climatic studies.

The future will find us working on a more detailed bioclimatic characterization of the Italian ecoregions, obtaining parameter estimates for all available ecoregional tiers, including Divisions, Sections and Subsections. As new ecoregional boundaries have recently been proposed mainly based on biogeographic and physiographic considerations (Blasi et al., unpublished data), the model could be applied to develop a climatic characterization of the new strata, comparing results to those reported in this paper. Further developments will include considering different specifications of matrix \mathbf{A} and hence using different type of priors for Σ following the discussion in Daniels and Kass (1999) and Daniels and Pourahmadi (2009). Rewriting the estimation algorithm using a parallel architecture would considerably speed up computations and will be addressed as a next step. Notable contributions along these lines related to similar settings include the very recent works of Finley et al. (2017) and Chung et al. (2018).

APPENDIX A: ITALIAN ECOREGIONS

- 1 Temperate Division
 - 1A Alpine Province
 - 1A1 Western Alps Section
 - 1A1a Alpi Marittime Subsection
 - 1A1b Northwestern Alps Subsection
 - 1A2 Central and Eastern Alps Section
 - 1A2a Pre-Alps Subsection
 - 1A2b Dolomiti and Carnia Subsection
 - 1A2c Northeastern Alps Subsection
 - 1B Po Plain Province
 - 1B1 Po Plain Section
 - 1B1a Lagoon Subsection
 - 1B1b Central Plain Subsection
 - 1B1c Western Po Basin Subsection
 - 1C Apennine Province
 - 1C1 Northern and Western Apennine Section
 - 1C1a Toscana and Emilia-Romagna Subsection
 - 1C1b Tuscan Basin Subsection
 - 1C2 Central and Southern Apennine Section
 - 1C2a Umbria and Marche Apennine Subsection
 - 1C2b Lazio and Abruzzo Apennine Subsection
 - 1C2c Campania Apennine Subsection
 - 1D Italian part of Illyrian Province
- 2 Mediterranean Division
 - 2A Italian part of Ligurian-Provencal Province
 - 2B Tyrrhenian Province
 - 2B1 Northern and Central Tyrrhenian Section
 - 2B1a Eastern Liguria Subsection
 - 2B1b Maremma Subsection
 - 2B1c Roman Area Subsection
 - 2B1d Southern Lazio Subsection
 - 2B2 Southern Tyrrhenian Section
 - 2B2a Western Campania Subsection
 - 2B2b Lucania Subsection
 - 2B2c Cilento Subsection
 - 2B2d Calabria Subsection
 - 2B3 Sicilia Section
 - 2B3a Iblei Subsection
 - 2B3b Sicilia Mountains Subsection
 - 2B3c Central Sicilia Subsection
 - 2B3d Western Sicilia Subsection

- 2B4 Sardegna Section
 - 2B4a Southwestern Sardegna Subsection
 - 2B4b Northwestern Sardegna Subsection
 - 2B4c Southeastern Sardegna Subsection
 - 2B4d Northeastern Sardegna Subsection
- 2C Adriatic Province
 - 2C1 Central Adriatic Section
 - 2C1a Abruzzo and Molise Adriatic Subsection
 - 2C1b Marche Adriatic Subsection
 - 2C2 Southern Adriatic Section
 - 2C2a Murge and Salento Subsection
 - 2C2b Gargano Subsection

APPENDIX B: DATA SOURCES

Data sources, organized by region: Abruzzo: Regione Abruzzo, direzione Lavori Pubblici e Protezione Civile; Basilicata: Regione Basilicata, Ufficio Protezione Civile; Calabria: Regione Calabria, ARPACAL, Centro funzionale multi-rischi; Campania: Regione Campania, Direzione generale Protezione Civile; Emilia Romagna: ARPA Emilia Romagna; Friuli Venezia Giulia: ARPA Friuli Venezia Giulia, Protezione Civile Regionale; Lazio: Regione Lazio, Servizio Integrato Agrometeorologico; Liguria: Arpa Liguria; Lombardia: Arpa Lombardia, Protezione Civile Regionale; Marche: Regione Marche, Servizio Agrometeorologico Regionale; Molise: Protezione Civile Regionale; Piemonte: Arpa Piemonte; Puglia: Regione Puglia, Arpa, Protezione Civile Regionale; Sardegna: Regione Sardegna, Arpa Sardegna; Sicilia: Regione Sicilia, Osservatorio Acque, Assessorato dell'Energia e dei servizi di pubblica utilità, dipartimento dell'Acqua e dei rifiuti; Trentino: Provincia di Trento, Centro funzionale Protezione Civile; Toscana: Regione Toscana, Settore idrologico regionale; Umbria: Regione Umbria, Centro funzionale decentrato di monitoraggio meteo-idrologico; Valle d'Aosta: Arpa Valle D'Aosta; Veneto: Arpa Veneto.

APPENDIX C: MODEL COMPARISONS

In this section we add some details on the models compared in the out of sample validation assessment given in Section 4.2. As a general remark, notice that with models M6–M8 we consider the response variables Y_1 , Y_2 and Y_3 defined as in Section 2, rescaled and standardized as we addressed in Section 4. With univariate models M1–M5, Y_1 and Y_2 are again as above, but the Y_3 response is simply the standardized maximum monthly temperature.

M1 Independent interpolation of each climate variable at each time point by three-variate thin plate splines for the spatial coordinates and elevation. At time t ,

the response variable Y_i is expressed as a three-dimensional thin plate spline function (tps) with arguments the UTM coordinates in s and the site elevation $X(s)$, all expressed in km: $Y_i(s, t) = \text{tps}(s, X)$. Smoothness of the function is chosen via generalized cross-validation as implemented in the function `Tps` of the `fields` package (Nychka et al. (2017)) in the *R* software.

M2 Independent interpolation of each climate variable by a generalized additive model (GAM) with a decennial three-variate thin plate regression spline component for the spatial coordinates and elevation, two independent univariate thin plate regression spline terms for smooth effects of years and months and fixed effects of the seven Provinces:

$$Y_i(s, t) = s_{d_t}(s, X(s)) + s(\text{year}_t) + s(\text{month}_{\text{time}}) + \mathbf{X}\boldsymbol{\beta}, \quad i = 1, \dots, 3,$$

where $s(\cdot)$ denotes thin plate regression splines as in Wood (2017), d_t is the index of the decennial time window where t belongs, \mathbf{X} is a matrix of seven dummy variables associated to ecoregions at tier level 3 and $\boldsymbol{\beta}$ is a vector of regression parameters. The above model specification was chosen among several competitive ones on the basis of smallest AIC and RMSE, and largest deviance explained. Estimates were obtained by the quadratically penalised likelihood type approach implemented in the function `gam` of the `mgcv` package (Wood (2017)) in the *R* software. The degree of smoothness of model terms is estimated as part of fitting.

M3 Independent interpolation of each climate variable at each time point by a GAM with a bivariate thin plate regression spline component for the spatial coordinates and an independent thin plate regression spline term for the elevation. Provinces were included as fixed effects only in the precipitation model, as they did not seem to affect the temperature spatial behavior at this temporal resolution:

$$Y_1(s, t) = s(s) + s(X(s)) + \mathbf{X}\boldsymbol{\beta},$$

$$Y_i(s, t) = s(s) + s(X(s)), \quad i = 2, 3.$$

Again, model choice was guided by the values of AIC, RMSE and deviance explained. Estimates were obtained by the quadratically penalised likelihood type approach implemented in the function `gam` of the `mgcv` package (Wood (2017)) in the *R* software. The degree of smoothness of model terms is estimated as part of fitting.

M4 Independent interpolation of each climate variable at each time point by local linear regression over nearest neighbors (NN) with elevation as explanatory variable. The method was implemented with function `gstat` of the `gstat` *R* package (Pebesma (2004), Gräler, Pebesma and Heuvelink (2016)), setting the inverse distance power to zero. We adopted 10 neighbors for the precipitation and 15 for the temperature measurements. The neighborhood dimension was chosen as the one minimizing the RMSE.

M5 Independent interpolation of each climate variable at each time point by Bayesian Gaussian Kriging parameterized and implemented as in the *R* package

geOR (Diggle and Ribeiro (2002)). The same model was used for all variables and time points, with the following features selected as producing the smallest RMSE:

- linear trend for the spatial coordinates and elevation expressed as UTM-km and km, respectively;
- exponential covariance structure;
- flat prior for the trend coefficients: $N(0, 100)$;
- uninformative reciprocal prior for the process variance: $p(\sigma^2) \propto \frac{1}{\sigma^2}$;
- discrete uniform priors for the range parameter ψ with $\psi \in \{0, 50, 100, 150, 200\}$ and for the relative nugget τ_{rel}^2 with $\tau_{\text{rel}}^2 \in \{0, 0.5, 1, 1.5, 2\}$ in km.

M6 Univariate version of the model in equation (1) with uncorrelated responses and no annual cyclical effects.

M7 Univariate version of the model in equation (1) with uncorrelated responses and annual cyclical effects.

M8 Multivariate spatio-temporal model in equation (1).

All scripts are available from the authors upon request.

Acknowledgments. The authors would like to thank Tilman Gneiting for his very valuable suggestions on an earlier version of this manuscript.

REFERENCES

- ANDRIEU, C. and THOMS, J. (2008). A tutorial on adaptive MCMC. *Stat. Comput.* **18** 343–373. [MR2461882](#)
- BAI, Y., SONG, P. X.-K. and RAGHUNATHAN, T. E. (2012). Joint composite estimating functions in spatiotemporal models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 799–824. [MR2988907](#)
- BAILEY, R. G. (1983). Delineation of ecosystem regions. *Environ. Manag.* **7** 365–373.
- BAILEY, R. G. (2004). Identifying ecoregion boundaries. *Environ. Manag.* **34** S14–S26.
- BALINT, P. J., STEWART, R. E., DESAI, A. and WALTERS, L. C. (2011). *Wicked Environmental Problems*. Island Press, Washington, DC.
- BANERJEE, S., CARLIN, B. P. and GELFAND, A. E. (2015). *Hierarchical Modeling and Analysis for Spatial Data*, 2nd ed. *Monographs on Statistics and Applied Probability* **135**. CRC Press, Boca Raton, FL. [MR3362184](#)
- BANERJEE, S., GELFAND, A. E., FINLEY, A. O. and SANG, H. (2008). Gaussian predictive process models for large spatial data sets. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 825–848. [MR2523906](#)
- BEVILACQUA, M., GAETAN, C., MATEU, J. and PORCU, E. (2012). Estimating space and space-time covariance functions for large data sets: A weighted composite likelihood approach. *J. Amer. Statist. Assoc.* **107** 268–280. [MR2949358](#)
- BEVILACQUA, M., FASSÒ, A., GAETAN, C., PORCU, E. and VELANDIA, D. (2016a). Covariance tapering for multivariate Gaussian random fields estimation. *Stat. Methods Appl.* **25** 21–37. [MR3460484](#)
- BEVILACQUA, M., ALEGRIA, A., VELANDIA, D. and PORCU, E. (2016b). Composite likelihood inference for multivariate Gaussian random fields. *J. Agric. Biol. Environ. Stat.* **21** 448–469. [MR3542081](#)
- BLANGIARDO, M. and CAMELETTI, M. (2015). *Spatial and Spatio-Temporal Bayesian Models with R-INLA*. Wiley, Chichester. [MR3364017](#)

- BLASI, C., CAPOTORTI, G., COPIZ, R., GUIDA, D., MOLLO, B., SMIRAGLIA, D. and ZAVAT-TERO, L. (2014). Classification and mapping of the ecoregions of Italy. *Plant Biosyst.* **148** 1255–1345.
- CHUNG, M., BINOIS, M., GRAMACY, R. B., MOQUIN, D. J., SMITH, A. P. and SMITH, A. M. (2018). Parameter and uncertainty estimation for dynamical systems using surrogate stochastic processes. Preprint. Available at [arXiv:1802.00852](https://arxiv.org/abs/1802.00852).
- DANIELS, M. J. and KASS, R. E. (1999). Nonconjugate Bayesian estimation of covariance matrices and its use in hierarchical models. *J. Amer. Statist. Assoc.* **94** 1254–1263. [MR1731487](https://doi.org/10.1080/01621459.1999.10476877)
- DANIELS, M. J. and POURAHMADI, M. (2009). Modeling covariance matrices via partial autocorrelations. *J. Multivariate Anal.* **100** 2352–2363. [MR2560376](https://doi.org/10.1016/j.jmva.2009.05.005)
- DATTA, A., BANERJEE, S., FINLEY, A. O. and GELFAND, A. E. (2016a). Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *J. Amer. Statist. Assoc.* **111** 800–812. [MR3538706](https://doi.org/10.1080/01621459.2016.1191475)
- DATTA, A., BANERJEE, S., FINLEY, A. O., HAMM, N. A. S. and SCHAAP, M. (2016b). Nonseparable dynamic nearest neighbor Gaussian process models for large spatio-temporal data with an application to particulate matter analysis. *Ann. Appl. Stat.* **10** 1286–1316. [MR3553225](https://doi.org/10.1214/15-AOS1325)
- DIGGLE, P. J. and RIBEIRO, P. J. JR. (2002). Bayesian inference in Gaussian model-based geostatistics. *Geogr. Environ. Model.* **6** 129–146.
- EIDSVIK, J., SHABY, B. A., REICH, B. J., WHEELER, M. and NIEMI, J. (2014). Estimation and prediction in spatial models with block composite likelihoods. *J. Comput. Graph. Statist.* **23** 295–315. [MR3215812](https://doi.org/10.1080/10618600.2014.910541)
- FAYE, E., HERRERA, M., BELLOMO, L., SILVAIN, J.-F. and DANGLES, O. (2014). Strong discrepancies between local temperature mapping and interpolated climatic grids in tropical mountainous agricultural landscapes. *PLoS ONE* **9** e105541.
- FERRARO PETRILLO, U. and RAIMATO, G. (2014). TeraStat Computer cluster for high performance computing, Dept. Statistical Science, Sapienza Univ. Rome. Available at <http://www.dss.uniroma1.it/en/node/6554>.
- FICK, S. E. and HIJMANS, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **37** 4302–4315.
- FINLEY, A. O., BANERJEE, S. and GELFAND, A. E. (2012). Bayesian dynamic modeling for large space-time datasets using Gaussian predictive processes. *J. Geogr. Syst.* **14** 29–47.
- FINLEY, A. O., DATTA, A., COOK, B. C., MORTON, D. C., ANDERSEN, H. E. and BANERJEE, S. (2017). Efficient algorithms for Bayesian nearest neighbor Gaussian processes. Preprint. Available at [arXiv:1702.00434](https://arxiv.org/abs/1702.00434).
- GELFAND, A. E., SCHMIDT, A. M., BANERJEE, S. and SIRMANS, C. F. (2004). Nonstationary multivariate process modeling through spatially varying coregionalization. *TEST* **13** 263–312. [MR2154003](https://doi.org/10.1080/10618600.2004.10554003)
- GELFAND, A. E., DIGGLE, P. J., FUENTES, M. and GUTTORP, P., eds. (2010). *Handbook of Spatial Statistics. Chapman & Hall/CRC Handbooks of Modern Statistical Methods*. CRC Press, Boca Raton, FL. [MR2761512](https://doi.org/10.1002/9781118009007)
- GNEITING, T. (2002). Nonseparable, stationary covariance functions for space-time data. *J. Amer. Statist. Assoc.* **97** 590–600. [MR1941475](https://doi.org/10.1198/0162145020000001475)
- GNEITING, T. (2013). Strictly and non-strictly positive definite functions on spheres. *Bernoulli* **19** 1327–1349. [MR3102554](https://doi.org/10.1080/10236192.2013.8102554)
- GNEITING, T. and RAFTERY, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *J. Amer. Statist. Assoc.* **102** 359–378. [MR2345548](https://doi.org/10.1198/016214506000002345548)
- GRÄLER, B., PEBESMA, E. and HEUVELINK, G. (2016). Spatio-temporal interpolation using gstat. *RFID J.* **8** 204–218.
- GRAMACY, R. B. and APLEY, D. W. (2015). Local Gaussian process approximation for large computer experiments. *J. Comput. Graph. Statist.* **24** 561–578. [MR3357395](https://doi.org/10.1080/10618600.2015.10557395)

- GRIMIT, E. P., GNEITING, T., BERROCAL, V. J. and JOHNSON, N. A. (2006). The continuous ranked probability score for circular variables and its application to mesoscale forecast ensemble verification. *Q. J. R. Meteorol. Soc.* **132** 2925–2942.
- GUINNESS, J. (2018). Permutation and grouping methods for sharpening Gaussian process approximations. *Technometrics*. To appear.
- HANNAH, L., ROEHRDANZ, P. R., IKEGAMI, M., SHEPARD, A. V., SHAW, M. R., TABOR, G., ZHI, L., MARQUET, P. A. and HIJMANS, R. J. (2013). Climate change, wine, and conservation. *Proc. Natl. Acad. Sci. USA* **110** 6907–6912.
- HEATON, M. J., DATTA, A., FINLEY, A., FURRER, R., GUHANIYOGI, R., GERBER, F., GRAMACY, R. B., HAMMERLING, D., KATZFUSS, M., LINDGREN, F., NYCHKA, D. W., SUN, F. and ZAMMIT-MANGION, A. (2017). A case study competition among methods for analyzing large spatial data. Preprint. Available at [arXiv:1710.05013](https://arxiv.org/abs/1710.05013).
- HIJMANS, R. J., CAMERON, S. E., PARRA, J. L., JONES, P. G. and JARVIS, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25** 1965–1978.
- JONA LASINIO, G., MASTRANTONIO, G. and POLLICE, A. (2013). Discussing the “big n problem.” *Stat. Methods Appl.* **22** 97–112. [MR3038184](https://doi.org/10.1007/s10260-013-0184-4)
- KATZFUSS, M. and GUINNESS, J. (2017). A general framework for Vecchia approximations of Gaussian processes. Preprint. Available at [arXiv:1708.06302](https://arxiv.org/abs/1708.06302).
- LI, S., DRAGICEVIC, S., CASTRO, F. A., SESTER, M., WINTER, S., COLTEKIN, A., PETTIT, C., JIANG, B., HAWORTH, J., STEIN, A. and CHENG, T. (2016). Geospatial big data handling theory and methods: A review and research challenges. *ISPRS J. Photogramm. Remote Sens.* **115** 119–133.
- LOVELAND, T. R. and MERCHANT, J. M. (2004). Ecoregions and ecoregionalization: Geographical and ecological perspectives. *Environ. Manag.* **34** S1.
- METZGER, M. J., BUNCE, R. G. H., JONGMAN, R. H. G., SAYRE, R., TRABUCCO, A. and ZOMER, R. (2013). A high-resolution bioclimate map of the world: A unifying framework for global biodiversity research and monitoring. *Glob. Ecol. Biogeogr.* **22** 630–638.
- MÜCHER, C. A., KLIJN, J. A., WASCHER, D. M. and SCHAMINÉE, J. H. J. (2010). A new European landscape classification (LANMAP): A transparent, flexible and user-oriented methodology to distinguish landscapes. *Ecol. Indicators* **10** 87–103.
- NYCHKA, D., FURRER, R., PAIGE, J. and SAIN, S. (2017). fields: Tools for spatial data. R package version 9.6, Univ. Corporation for Atmospheric Research, Boulder, CO.
- PEBESMA, E. J. (2004). Multivariable geostatistics in S: The gstat package. *Comput. Geosci.* **30** 683–691.
- PESARESI, S., GALDENZI, D., BIONDI, E. and CASAVECCHIA, S. (2014). Bioclimate of Italy: Application of the worldwide bioclimatic classification system. *J. Maps* **10** 538–553.
- R CORE TEAM (2017). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- SHIROTA, S. and GELFAND, A. E. (2017). Space and circular time log Gaussian Cox processes with application to crime event data. *Ann. Appl. Stat.* **11** 481–503. [MR3693535](https://doi.org/10.1214/16-AAS015)
- SPIEGELHALTER, D. J., BEST, N. G., CARLIN, B. P. and VAN DER LINDE, A. (2002). Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 583–639. [MR1979380](https://doi.org/10.1111/j.1467-9868.2002.01979.x)
- WMO (2016). *Guidelines on Best Practices for Climate Data Rescue*. WMO—Guidelines **1182**. World Meteorological Organization, Geneva.
- WOOD, S. N. (2017). *Generalized Additive Models: An Introduction with R*, 2nd ed. *Texts in Statistical Science Series*. CRC Press/CRC, Boca Raton, FL. [MR2206355](https://doi.org/10.1201/9781498710584)
- XU, G., LIANG, F. and GENTON, M. G. (2015). A Bayesian spatio-temporal geostatistical model with an auxiliary lattice for large datasets. *Statist. Sinica* **25** 61–79. [MR3328803](https://doi.org/10.1007/s11464-015-0461-4)

G. MASTRANTONIO
DEPARTMENT OF MATHEMATICAL SCIENCE
POLITECNICO DI TORINO
CORSO DUCA DEGLI ABRUZZI, 24
10129 TORINO
ITALY
E-MAIL: gianluca.mastrantonio@polito.it

A. POLLICE
DEPARTMENT OF ECONOMICS AND FINANCE
UNIVERSITÀ DEGLI STUDI DI BARI "ALDO MORO"
PIAZZA UMBERTO I
70121 BARI
ITALY

G. JONA LASINIO
DEPARTMENT OF STATISTICAL SCIENCES
SAPIENZA UNIVERSITÀ DI ROMA
PIAZZALE ALDO MORO 5
00185 ROMA
ITALY

G. CAPOTORTI
G. GENOVA
C. BLASI
DEPARTMENT OF ENVIRONMENTAL BIOLOGY
SAPIENZA UNIVERSITÀ DI ROMA
PIAZZALE ALDO MORO 5
00185 ROMA
ITALY

L. TEODONIO
ICRCPAL
MINISTRY OF CULTURAL HERITAGE
AND ACTIVITIES AND TOURISM
VIA DEL COLLEGIO ROMANO 27
00186 ROMA
ITALY