A Dirichlet process model for changepoint detection with multivariate bioclimatic data

Received: 5 February 2021 | Revised: 10 May 2021 | Accepted: 12 July 2021

DOI: 10.1002/env.2699

RESEARCH ARTICLE

WILEY

A Dirichlet process model for change-point detection with multivariate bioclimatic data

Gianluca Mastrantonio[1] | **Giovanna Jona Lasinio**[2] | **Alessio Pollice**[3] | **Lorenzo Teodonio**[4] | **Giulia Capotorti**[5]

[1]Department of Mathematical Sciences, Politecnico di Torino, Turin, Italy

[2]Department of Statistical Sciences, Sapienza Università di Roma, Rome, Italy

[3]Department of Economics and Finance, Università di Bari Aldo Moro, Bari, Italy

[4]Department of Environmental Biology, Sapienza Università di Roma, Rome, Italy

[5]ICRCPAL, Ministry of Cultural Heritage and Activities and Tourism, Rome, Italy

Correspondence
Gianluca Mastrantonio, Department of Mathematical Sciences, Politecnico di Torino, Turin, Italy.
Email: gianluca.mastrantonio@polito.it

Abstract

Motivated by real-world data of monthly values of precipitation, minimum, and maximum temperature recorded at 360 monitoring stations covering the Italian territory for 60 years (12×60 months), in this work we propose a change-point model for multiple multivariate time series, inspired by the hierarchical Dirichlet process. We assume that each station has its change-point structure and, as main novelties, we allow unknown subsets of the parameters in the data likelihood to stay unchanged before and after a change-point, that stations possibly share values of the same parameters and that the unknown number of weather regimes is estimated as a random quantity. Owing to the richness of the formalization, our proposal enables us to identify clusters of spatial units for each parameter, evaluate which parameters are more likely to change simultaneously, and distinguish between abrupt changes and smooth ones. The proposed model provides useful benchmarks to focus monitoring programs regarding ecosystem responses. Results are shown for the whole data, and a detailed description is given for three monitoring stations. Evidence of local behaviors includes highlighting differences in the potential vulnerability to climate change of the Mediterranean ecosystems from the Temperate ones and locating change trends distinguishing between continental plains and mountain ranges.

KEYWORDS

change-points, Dirichlet process, hierarchical model, multivariate process, thermopluviometric data

1 | INTRODUCTION

Climate elements and regimes, such as temperature, precipitation, their annual cycles and mutual relationships, primarily affect the type and distribution of plants, animals, and soils, as well as their combination in complex ecosystems and ecoregions (Bailey, 2004; Metzger et al., 2013). Consequently, exploring climate change and respective responses of biodiversity is of primary importance for natural capital conservation and sustainable development at multiple levels (Pecl et al., 2017). Notwithstanding the impacts of changes are being quite in depth investigated for species, many uncertainties still remain as regards ecosystems, that is, for groups of interacting species that live in the same environment, and as regards the ecosystem arrangement within ecoregions (Felton & Smith, 2017; Walther, 2010; Yu et al., 2019). Main

issues are posed by the enhanced resilience of complex ecological systems with respect to individual species, that probably determines long-term adaptive responses in terms of comprehensive species composition and/or spatial displacement (Frauendorf et al., 2019). Also for this reason, abrupt climate changes, potentially detectable by means of change-points (CPs), are supposed to have a stronger effect on ecosystem structure and functioning than gradual climate feature variations (Alley et al., 2003; Williams et al., 2011). Models able to provide more evidences on sharp temporal and spatial shifts in climate regimes are therefore needed, especially at intermediate scales between the site and the global or continental ones. The hypothesis we moved from is that the detection of CPs by means of a model that takes into account the similarities among spatial units and the interactions between different bioclimatic features may effectively support the inference of climate change impacts on ecosystems. Especially in territories with marked orographic and physiographic heterogeneity, climate change may actually occur unevenly across ecological regions and may differently affect the interdependence between joint climate parameters and biodiversity arrangement and distribution. Accordingly, the aim of the present work is to develop and disseminate an original methodology for i) defining CPs even on the basis of the correlation between temperatures and precipitation, besides their mean values and variances and for each of the meteorological stations occurring in the different ecoregions, ii) describing the spatial distribution of such CPs across ecoregions at a national level, based on common climatic trends in the respective stations, and iii) spotting similarities between ecoregions in bioclimatic terms, even independently from potential changes.

Our modeling proposal is applied to investigate the possible presence of CPs in thermopluviometric historical data over the Italian peninsula. We consider monthly records of precipitation and min/max temperature at 360 monitoring stations over 60 years (1951–2010). The full database has $3 \times 360 \times 60 \times 12$ entries, though almost all time series are affected by variable amounts of missing data (Mastrantonio et al., 2019). Series observed over a large time span are usually subject to changes of their structure and features, concerning both the first-order (means) and the second-order (variances and correlations) properties (Battaglia et al., 2019). Moreover, seasonality is always present in thermopluviometric data and cannot be disregarded.

CP problems can rely on a quite extensive literature and methods are often directly motivated by specific fields of study and areas of interest. The use of CP models for climate data has become widespread in the late years, and many different approaches have been proposed in both the classic and Bayesian framework, including the use of change detection statistics (Jandhyala et al., 2010), the estimation of piecewise linear trends (Tomè & Miranda, 2004), sequential and CUSUM tests for regime shifts (Robbins et al., 2011; Rodionov, 2004), information approaches (Beaulieu et al., 2012; Lu et al., 2010), two-phase time series regression models (Lund et al., 2007; Lund & Reeves, 2002), ML estimation (Bhattacharya, 1987; Hawkins, 2001; Killick et al., 2010), variable selection in high-dimensional regression (Li et al., 2021), with associated software implementations (James & Matteson, 2013; Lindeløv, 2020). As regards Bayesian methods, the seminal works of Carlin et al. (1992) and Chib (1998) have influenced much of the subsequent literature, recently reviewed in Peluso et al. (2019) to whom we refer the interested reader. To the authors knowledge, the works of Ko et al. (2015) and Peluso et al. (2019) contain previous attempts to exploit the flexibility of the Dirichlet process in the construction of CP models in the Bayesian framework. The above scholars testify the liveliness and variety of the research in the field of CP modeling.

The approach we propose is motivated by methods that consider time series as possibly broken down into time regimes composed of adjacent observations (Samé et al., 2011), assuming that observations belonging to the same regime follow a common distribution. Regimes are separated by *local change points*, that is, changes that may repeatedly occur in time due to seasonal or climate cycles, or by *global change points*, that is, changes that occur only once and determine a permanent shift in the distribution. Local and global CPs can be obtained by mixture-type models for model-based clustering, where mixture components act as generators of clusters or time regimes and classification is equivalent to the model fitting process. As we are interested in global CPs, we rely on the formulation of the Bayesian mixture-type CP model proposed by Chib (1998) and extend it to a multivariate Dirichlet process-based CP model (Ferguson, 1973). While the former requires to set a priori the maximum number of CPs, the Dirichlet process (DP) formulation allows to define this number as a random quantity to be estimated along with the model fitting. Our proposal is inspired by the single-hierarchical DP of Teh et al. (2006) that we generalize to a combination of multiple DPs, each corresponding to one of the scalar parameters of a multivariate time series of climate data. In classic mixture-type models (McLachlan & Peel, 2000) the density of each regime depends on multiple parameters and regime shifts correspond to changes in all of them. Peluso et al. (2019) proposed a nonhierarchical DP model that allows CPs due to only some of the parameters, but implies a priori knowledge of which parameters change jointly. Our formalization overcomes these limitations since, when a change point occurs, it allows us to determine which parameters are affected and how. To summarize, the main novel contributions of our proposal to the methodology of CP modeling have to do with: i) the definition of CP as due to an unspecified subset of the parameters of a multivariate time series, ii) the consequent identification of the parameters (i.e., of the distributional

features) corresponding to the CPs, iii) the estimate of the unknown number of CPs as a by-product of the model fitting process, iv) the fact that parameter values are possibly shared by different monitoring stations forming clusters of spatial units, and v) the ability to distinguish between abrupt changes and smooth ones.

In our proposal the data are a realization of a CP model based on the multivariate normal distribution with vector-valued parameters obtained as suitably modified DP draws. In particular, we assume that each monitoring station follows its own DP-based CP model with a trivariate normal density for the data, allowing stations to share subsets of the DP atoms. The model is estimated through the Markov chain Monte Carlo (MCMC) algorithm, after setting the minimum number of contiguous observations that form a regime. This allows to avoid too short regimes with little biological meaning and hard to interpret and justify in the Mediterranean context. Despite the complexity of the proposed model, the MCMC algorithm is easy to implement and mostly based on Gibbs steps that are defined by introducing suitable latent variables. The model capacity to recognize the hypothesized data structure is successfully tested with a simulated example. Compared with competitive approaches, the proposed model shows a better performance in terms of predictive ability as measured by continuous ranked probability score (CRPS).

The article is organized as follows. In Section 2 we give a brief account of available climate data and how they appear in the CP model. Section 3 provides the hierarchical formulation of the CP model with its components. The model performances are compared with similar competing approaches in Section 4.2. The results are discussed in detail for three selected stations in Section 4.3, while in Section 4.4 the results for all stations are shown. The article ends with some concluding remarks and directions for future researches in Section 5. The Appendix contains details of the MCMC algorithm, an example of how CPs are appropriately identified with artificial data, and the legend of the Italian ecoregion system.

2 | REPRESENTATION OF THE AVAILABLE DATA

The present work relies on monthly records of precipitation and min/max temperature over 60 years (1951–2010) from a network of 360 weather monitoring stations, spread over the Italian ecoregions (Figure 1). The ecoregions represent wide ecosystems occurring in discrete geographical areas (Bailey, 1983; Loveland & Merchant, 2004), and a hierarchical classification of Italian ecoregions was recently obtained by combining climatic diagnostic features with distribution patterns of biological diversity and other physical characteristics of the environment (Blasi et al., 2018). The Italian ecoregions are arranged into four hierarchically nested tiers, which consist of two divisions, seven provinces, 14 sections, and 33 subsections (see Tables C1 and C2). Raw data on precipitation and min/max temperature over the period 1951–2010 were mostly obtained from National Institutions (ISPRA, CRA/CREA, Meteomont, and ENEA) and local authorities (from each of the 20 Italian regions). Monthly records were obtained by monthly cumulative precipitations and monthly averages of daily minimum and maximum air temperatures from a network of selected meteorological stations belonging to the different ecoregions. Almost all time series are affected by variable amounts of missing data but, with respect to other databases (such as the WorldClim, providing fine resolution but mainly inferred data, Fick & Hijmans, 2017) these are based on truly observed data and better represent the specificity of the geographical and orographic heterogeneity of Italy. To guarantee an even representation of the different ecoregional sectors, also stations presenting a variable amount of missing data (up to a threshold of 50% of missing data for each of the parameters) have been included in the network (for further details on the database see Mastrantonio et al., 2019).

Let $Y^*_{1,t,\mathbf{s}}$, $Y^*_{2,t,\mathbf{s}}$, and $Y^*_{3,t,\mathbf{s}}$ be the random variables underlying the precipitation, minimum, and maximum temperature at time t, where $t \in \mathcal{T} \equiv \{1, 2, \dots, T\}$ and spatial location $\mathbf{s}$, where $\mathbf{s} \in \mathcal{S} \subset \mathbb{R}^2$. To simplify the model definition and the computations, as well as to avoid dimensionality issues, these variables are transformed. While standardization is sufficient for the minimum and maximum temperature $Y^*_{2,t,\mathbf{s}}$ and $Y^*_{3,t,\mathbf{s}}$, we have to consider that the precipitation $Y^*_{1,t,\mathbf{s}}$ is a positive real valued variable with zeros that account for observations with no precipitations. We indicate with $Y_{1,t,\mathbf{s}}$, $Y_{2,t,\mathbf{s}}$ and $Y_{3,t,\mathbf{s}}$ the transformed variables, where

$$\begin{cases} Y_{1,t,\mathbf{s}} = \frac{Y^*_{1,t,\mathbf{s}}}{\sqrt{S^2_{Y^*_1}}} & \text{if } Y^*_{1,t,\mathbf{s}} > 0, \\ Y_{1,t,\mathbf{s}} \leq 0 & \text{if } Y^*_{1,t,\mathbf{s}} = 0, \end{cases} \tag{1}$$

where $S^2_{Y^*_1}$ is the sample variance. We can then model $Y_{1,t,\mathbf{s}}$ instead of $Y^*_{1,t,\mathbf{s}}$, which has the advantage to be defined over $\mathbb{R}$, as the two standardized variables $Y_{2,t,\mathbf{s}}$ and $Y_{3,t,\mathbf{s}}$, and induces a bulk of probability at zero for $Y^*_{1,t,\mathbf{s}}$. Equation (1) is an easy

FIGURE 1 Italian ecoregions (sections) and the climate monitoring network

way to define a zero inflated distribution for $Y^*_{1,t,\mathbf{s}}$ since for any choice of the density of $Y_{1,t,\mathbf{s}}$ it ensures that the probability of $Y^*_{1,t,\mathbf{s}} = 0$ is equal to the cumulative distribution of $Y_{1,t,\mathbf{s}}$ evaluated at zero, while variables $Y^*_{1,t,\mathbf{s}}/\sqrt{S^2_{Y^*_1}}$ and $Y_{1,t,\mathbf{s}}$ have the same density on the positive real line. Notice that (1) is the same transformation used in the Tobit model (McDonald & Moffitt, 1980) which is a special case of a censored regression model. Indeed, values of $Y_{1,t,\mathbf{s}} \leq 0$ cannot be computed directly from the data and are considered missing, but it is easy to deal with this issue from the implementation point of view, as we show in the Appendix.

3 | THE CP MODEL

With the definition of the random variables $Y_{1,t,\mathbf{s}}$, $Y_{2,t,\mathbf{s}}$, and $Y_{3,t,\mathbf{s}}$ underlying the precipitation, minimum, and maximum temperature given in Section 2, the proposed model assumes that $\mathbf{Y}_{t,\mathbf{s}} = (Y_{1,t,\mathbf{s}}, Y_{2,t,\mathbf{s}}, Y_{3,t,\mathbf{s}})$ has a trivariate normal density at each monitoring station $\mathbf{s}$ and time t:

$$f(\mathbf{y}_{t,\mathbf{s}}|\psi_{t,\mathbf{s}}, \boldsymbol{\theta}_{t,\mathbf{s}}) = \phi_3\left(\mathbf{y}_{t,\mathbf{s}}|\psi_{t,\mathbf{s}} + \boldsymbol{\mu}_{t,\mathbf{s}}, \boldsymbol{\Sigma}_{t,\mathbf{s}}\right). \tag{2}$$

The vector-valued parameter $\boldsymbol{\theta}_{t,\mathbf{s}}$ is composed by the nine elements in $\boldsymbol{\mu}_{t,\mathbf{s}}$ and $\boldsymbol{\Sigma}_{t,\mathbf{s}}$. More precisely, $\theta_{t,\mathbf{s},i} = \mu_{t,\mathbf{s},i}$ with $i = 1, 2, 3$, $\theta_{t,\mathbf{s},i} = \sigma^2_{t,\mathbf{s},i-3}$ with $i = 4, 5, 6$ are the variances and the last three elements of $\boldsymbol{\theta}_{t,\mathbf{s}}$ are the correlation coefficients between the three climate variables: $\rho_{t,\mathbf{s},1,2}$, $\rho_{t,\mathbf{s},1,3}$ and $\rho_{t,\mathbf{s},2,3}$. The term $\psi_{t,\mathbf{s}}$ represents a seasonal component that we expect to be relevant in the available data, with $\psi_{t,\mathbf{s}} = \psi_{t+12,\mathbf{s}}$. As an identification constraint needed to jointly estimate the monthly effects $\psi_{t,\mathbf{s}}$ and mean effects $\boldsymbol{\mu}_{t,\mathbf{s}}$, we assume $\sum_{t=1}^{12}\psi_{t,\mathbf{s}} = 0$. This also ensures that the total contribution of the seasonal component over the 12 months is 0 and lets $\boldsymbol{\mu}_{t,\mathbf{s}}$ stand for the mean of $\mathbf{y}_{t,\mathbf{s}}$. This last point is critical for the interpretation, since it enables to compare and test if different monitoring stations have equal means.

Let $\boldsymbol{\psi}$ and $\boldsymbol{\theta}$ be the vectors containing all the associated parameters, then the proposed model has the following likelihood:

$$f(\mathbf{y}|\boldsymbol{\psi},\boldsymbol{\theta}) = \prod_{\mathbf{s}\in S}\prod_{t\in\mathcal{T}} f(\mathbf{y}_{t,\mathbf{s}}|\psi_{t,\mathbf{s}},\boldsymbol{\theta}_{t,\mathbf{s}}).$$

The parameter $\boldsymbol{\theta}_{t,\mathbf{s}}$ is used to define the CP model by a modified version of the hierarchical DP (Teh et al., 2006), allowing the time series at different monitoring stations, and the same stations at different time-points, to share the same values for subsets of the nine elements of $\boldsymbol{\theta}_{t,\mathbf{s}}$. Intuitively, at each time-point and monitoring station, the nine-variate parameter vector $\boldsymbol{\theta}_{t,\mathbf{s}}$ is assumed to come from a DP mixture-type model. Mixture components correspond to different distributional features of the underlying three-variate Gaussian distribution, possibly connected with time regimes and clusters of spatial units of the climate variables. In the next section we introduce the definition of the discrete multivariate distribution that provides the main building block for the DP-CP model in Section 3.2.

3.1 | The discrete multivatiate distribution G_0

The DP is often used in Bayesian mixture modeling, since it allows to estimate the number of mixture components from the data, without exploiting information criteria. First, let us define independent DPs for the nine parameters $\theta_\ell \in \Theta_\ell$ of the multivariate normal distribution. A draw G_{θ_ℓ} from $\text{DP}(\gamma, H_{\theta_\ell})$ is a discrete distribution that depends on a *scaling* or *concentration parameter* $\gamma > 0$ and on the *base distribution* H_{θ_ℓ} over Θ_ℓ (Ferguson, 1973), that can be written as

$$G_{\theta_\ell} = \sum_{p\in\mathbb{N}} \nu_{\ell,p}\delta_{\eta_{\ell,p}},$$

where δ_x be the Dirac delta function. Notice that G_{θ_ℓ} is a sample from the $\text{DP}(\gamma, H_{\theta_\ell})$ and, as any DP sample, it can be represented using the sets of atoms $\{\eta_{\ell,p}\}_{p\in\mathbb{N}}$ and weights $\boldsymbol{\nu}_\ell = \{\nu_{\ell,p}\}_{p\in\mathbb{N}}$, where $\eta_{\ell,p}$ are i.i.d. from the base distribution H_{θ_ℓ} and $\boldsymbol{\nu}_\ell$ is a draw from an infinite-dimensional Dirichlet distribution whose parameters depend on γ and H_{θ_ℓ} (see Fox et al., 2011; Teh et al., 2006 for more details) and defined using the *stick-breaking representation* (Ferguson, 1973).

As a next step we combine the sets of atoms $\{\eta_{\ell,p}\}_{p\in\mathbb{N}}$ and weights $\boldsymbol{\nu}_\ell$ of each parameter θ_ℓ to define a multivariate discrete distribution that will be used as a base distribution of a further DP. Let $\boldsymbol{\eta}_k$ be a random vector-valued atom with $\boldsymbol{\eta}_k = \{\eta_{\ell,w_{\ell,k}}\}_{\ell=1}^9$, where $w_{\ell,k} = p$ means that the ℓth element of $\boldsymbol{\eta}_k$ is equal to $\eta_{\ell,p}$. Then, we can define a discrete multivariate distribution as follows:

$$G_0 = \sum_{k\in\mathbb{N}} \xi_k\delta_{\boldsymbol{\eta}_k}, \tag{3}$$

with random weights ξ_k obtained as products of the associated $\nu_{\ell,p}$, that is,

$$\xi_k = \prod_{\ell=1}^{9} \nu_{\ell,w_{\ell,k}}, \tag{4}$$

ensuring that two $\boldsymbol{\eta}_k$ cannot be exactly the same. The set $\{\boldsymbol{\eta}_k\}_{k\in\mathbb{N}}$ is then comprised of all possible combinations of the nine parameters without repetitions and all nine-variate atoms of G_0 contain three means, three variances, and three correlations, such that there are not two $\boldsymbol{\eta}_k$s that share the values of all the nine parameters. The discrete multivariate distribution G_0 is used as part of the DP-CP model, to generate random sets of vector-valued parameters for all time-points and monitoring stations. In the next sections we explain how these nonfully overlapping random sets potentially give rise to weather regimes, possibly shared by monitoring stations, and to CPs in the observed time series.

3.2 | The DP-CP model with no-return constraint and minimum regime length

A DP mixture-type model with components that can be shared among monitoring stations is designed introducing the station-specific discrete multivariate distribution

$$G_{\mathbf{s}} = \sum_{k \in \mathbb{N}} \pi_{\mathbf{s},k} \delta_{\boldsymbol{\eta}_k}, \tag{5}$$

as a draw from the Dirichlet process DP(α, G_0) with random set of atoms in $\{\eta_k\}_{k\in\mathbb{N}}$ i.i.d. from G_0 and associated random weights $\{\pi_{\mathbf{s},k}\}_{k\in\mathbb{N}}$ drawn from an infinite-dimensional Dirichlet distribution whose parameters depend on α and G_0 (Teh et al., 2006). The atoms of the distribution $G_{\mathbf{s}}$ (i.e., the combinations of mean, variance, and correlation parameters) contain all the possible values that $\boldsymbol{\theta}_{t,\mathbf{s}}$ can assume and they are thus random objects generated by construction of the discrete multivariate distribution G_0. Notice that, as in the hierarchical DP of Teh et al. (2006), since the multivariate base distribution G_0 is discrete and the same for each station, all $G_{\mathbf{s}}$ share the same atoms $\boldsymbol{\eta}_k \in \Theta$ with different associated weights.

Time regimes are constrained by the definition of global CP given in Section 1 as follows: a mixture-type CP model has the characteristic that if at any time-point t the process moves from the kth mixture component to a new one, at any time greater than t the process cannot go back to component k. This *no-return constraint* that characterizes the time evolution of the system at each monitoring station $\mathbf{s}$ is introduced defining $\boldsymbol{\pi}_{\mathbf{s}}(\mathcal{I})$ as the *restriction* of a probability vector $\boldsymbol{\pi}_{\mathbf{s}}$ for a set of unique indices $\mathcal{I} \subset \mathbb{N}$, namely,

$$\pi_{\mathbf{s},k}(\mathcal{I}) = \begin{cases} 0 & \text{if } k \notin \mathcal{I}, \\ \frac{\pi_{\mathbf{s},k}}{\sum_{j\in\mathcal{I}} \pi_{\mathbf{s},j}} & \text{if } k \in \mathcal{I}, \end{cases}$$

that is, the elements with indices not in $\mathcal{I}$ are set to zero and the remaining elements are scaled so to obtain a unit total probability mass. The corresponding restriction of the discrete multivariate distribution $G_{\mathbf{s}}$ in (5) is then

$$G_{\mathbf{s}}(\mathcal{I}) = \sum_{k \in \mathbb{N}} \pi_{\mathbf{s},k}(\mathcal{I}) \delta_{\boldsymbol{\eta}_k}.$$

The distribution $G_{\mathbf{s}}(\mathcal{I})$ assigns nonzero probabilities to values of $\boldsymbol{\theta}_{t,\mathbf{s}}$ in the set $\{\boldsymbol{\eta}_k\}_{k\in\mathcal{I}}$. Once we introduce the index variable $z_{t,\mathbf{s}}$ so that $z_{t,\mathbf{s}} = k$ when $\boldsymbol{\theta}_{t,\mathbf{s}} = \boldsymbol{\eta}_k$, with $z_{0,\mathbf{s}} = \emptyset$, the no-return constraint then consists in letting $\mathcal{I}_{t,\mathbf{s}} = \{k \in \mathbb{N} | z_{l,\mathbf{s}} \neq k, l = 1, 2, \ldots, t\} \cup z_{t,\mathbf{s}}$, with $\mathcal{I}_{0,\mathbf{s}} \equiv \mathbb{N}$. Finally, setting to m the *minimum regime length*, we obtain the nonparametric CP model as follows:

$$\boldsymbol{\theta}_{t,\mathbf{s}} | \boldsymbol{\theta}_{t-1,\mathbf{s}}, \ldots, \boldsymbol{\theta}_{1,\mathbf{s}}, G_{\mathbf{s}} \sim \begin{cases} \delta_{\boldsymbol{\eta}_{z_{t-1,\mathbf{s}}}} & \text{if } n^{t-1}_{\mathbf{s},z_{t-1,\mathbf{s}}} < m \\ G_{\mathbf{s}}(\mathcal{I}_{t-1,\mathbf{s}}) & \text{otherwise} \end{cases}, \tag{6}$$

where $n^t_{\mathbf{s},k} = \sum_{i=1}^{t} \mathbb{I}(z_{i,\mathbf{s}} = k)$, assuming $n^0_{\mathbf{s},k} = 0$. Equation (6) implies that all regimes are composed of at least m time-points since, if the regime occupied at time $t-1$ is shorter, it gives $\boldsymbol{\theta}_{t,\mathbf{s}} = \boldsymbol{\eta}_{z_{t-1,\mathbf{s}}}$ with probability one. If m or more time points are in the current regime, then $\boldsymbol{\theta}_{t,\mathbf{s}}$ is drawn from $G_{\mathbf{s}}(\mathcal{I}_{t-1,\mathbf{s}})$ which is composed by the atom $\boldsymbol{\eta}_{z_{t-1,\mathbf{s}}}$ and all the other atoms not previously observed at station $\mathbf{s}$.

Equation (6) defines the Dirichlet process change point (DP-CP) model for the parameter vector $\boldsymbol{\theta}_{\mathbf{s}}$ at station $\mathbf{s}$:

$$\boldsymbol{\theta}_{\mathbf{s}} | \alpha, G_0 \sim \text{ DP-CP}(\alpha, G_0, m),$$

with scale parameter α, base distribution G_0 and minimum length m. This is a DP mixture-type model that allows to discretize the time series of the parameters in $\boldsymbol{\theta}_{\mathbf{s}}$ in a finite number of regimes of at least m time-points, complying with the no-return constraint. At each monitoring station, weather regimes can share some of the values of the parameters allowing CPs in subsets of the whole set of parameters. Analogously, as $G_{\mathbf{s}}$ and $G_{\mathbf{s}'}$ have the same set of atoms, the time series at the two monitoring stations $\mathbf{s}$ and $\mathbf{s}'$ can share some or all the values of the parameters, allowing for the detection of similarities between the two corresponding stations. The possibility of identifying clusters of spatial units derives from the generalization of this consideration to more than two monitoring stations. Such clusters are not assumed to be informed by geographic distances or boundaries.

3.3 | The hierarchical formulation and some operative remarks

Let $\boldsymbol{\psi}$ and $\boldsymbol{\theta}$ be the vectors containing all the associated parameters and $\boldsymbol{\theta}_{\mathbf{s}} = \{\boldsymbol{\theta}_{t,\mathbf{s}}\}_{t\in\mathcal{T}}$. The proposed model is then defined by the following hierarchy:

$$f(\mathbf{y}|\boldsymbol{\psi},\boldsymbol{\theta}) = \prod_{\mathbf{s}\in S}\prod_{t\in\mathcal{T}} f(\mathbf{y}_{t,\mathbf{s}}|\psi_{t,\mathbf{s}},\boldsymbol{\theta}_{t,\mathbf{s}}),$$
$$\boldsymbol{\theta}_{\mathbf{s}}|\alpha, G_0 \sim \text{DP-CP}(\alpha, G_0, m), \quad \mathbf{s}\in S,$$
$$G_{\theta_\ell}|\gamma, H_{\theta_\ell} \sim \text{DP}(\gamma, H_{\theta_\ell}), \quad \ell = 1, \ldots, 9,$$

where G_0 is defined as in Equation (3) and priors for the elements of $\boldsymbol{\psi}$, $\boldsymbol{\theta}$, α, and γ are specified in Section 4.

As we said before, since two different $\boldsymbol{\eta}_k$ can share some of their components, the occurrence of a change point does not imply that all nine parameters change. The minimal requirement for a CP detection is that at least one of the elements of the parameter set changes. Every time a CP occurs, the mixture-type model generates a new regime with a set of parameters that is different from the previous ones. Although this means that something has changed in the distribution, from a practical point of view there may be cases where this change is too small to be considered a real CP. To single out the events that correspond to *abrupt* changes in the time evolution of at least one parameter, and fully address the definition of global CP given in Section 1, we say that we have a *disconnecting* CP (d-CP) in a certain time interval when the 95% highest posterior density (HPD) bounds of the corresponding parameter before and after the interval do not overlap. Conversely, *not-disconnecting* CPs (nd-CP) correspond to smooth changes, as complementary to d-CPs. As a matter of fact, the latter definition requires postprocessing the model output and gives rise to a final step that completes the quest for global CPs, see Section 4.1. The difference between d-CPs and nd-CPs is illustrated in detail in Section 4.3 for two monitoring stations.

Finally, the constraint induced by the minimum regime length m in (6) would result in the impossibility for a CP to be observed in the first m time-points of the time series. To avoid that, the set of time indices $\mathcal{T}$ was augmented with m time-points before the first observation and the associated $\mathbf{y}_{t,\mathbf{s}}$ were considered as missing data.

4 | THE DP CHANGE-POINT MODEL FOR THERMOPLUVIOMETRIC DATA

The DP change point model presented in the previous sections is here applied to the thermopluviometric Italian historical data described in Sections 1 and 2. In the following we provide the results of several estimation runs for alternative specifications of the model, all obtained using 80,000 MCMC iterations, with 60,000 burnin and thinning by 4. Implementation involved parallel computing (OpenMP Architecture Review Board, 2008) and computations were performed on the TeraStat cluster (Petrillo & Guerriro, 2014). Priors were set to $N(0, 100)$ for the means $\mu_{t,\mathbf{s},i}$ and seasonal parameters $\psi_{t,\mathbf{s}}$, InverseGamma(1, 1) for the variances $\sigma_{t,\mathbf{s},i}$, uniform over the space of n.n.d. correlation matrices for $\rho_{t,\mathbf{s},i,i'}$ and Gamma(1, 1) for the DP hyperparameters α and γ, with $i, i' = 1, 2, 3$, $t \in \mathcal{T}$ and $\mathbf{s} \in S$.

In the model we assume that observations are conditionally independent, given the parameters and monthly effects; this assumption may be questionable as usually an explicit modelization of time dependence is preferred. However, from Figure 2, that shows several lags of the sample autocorrelation function for the data and the model residuals, it is clear that our proposal captures the temporal correlation present in the original data. With these premises, and using the explicit definition of the seasonal component provided by the term $\psi_{t,\mathbf{s}}$, the lack of a specific parametric formulation of the temporal dependence does not have a considerable impact on the quality of the resulting model.

4.1 | The optimal minimum regime length *m* and model output postprocessing

As a preliminary step, our model implementation requires to set the minimum regime length m. This kind of requirement is not new in the field of CP detection, see, for example, Rodionov (2004) where a cut-off regime length of 10 years is considered for the time series of the Pacific Decadal Oscillation index. Regime lengths may be influenced by global atmospheric patterns and cycles. In southern Europe, and specifically in the Mediterranean region, natural cycles related to solar activity or global atmospheric dynamics can influence the climate pattern (Luque-Espinar et al., 2017). Interannual cycles are related to seasonal changes and are accounted for by the model term $\psi_{t,\mathbf{s}}$, while biennial cycles may be linked to quasi-Biennial oscillation (see Luque-Espinar et al., 2017, and references therein). More uncertainty affects the interpretation of longer cycles, for example, 5–6 years cycles may be a harmonic component of the 11-year sunspot cycle or they may be due to *Èl Niño* effects. We tested $m \in \{1, 24, 60\}$ since these values were considered plausible, respectively, corresponding to no constraints, and minimum regime lengths of 2 and 5 years, compatible with the underlying physical processes. The $m = 1$ option was tested to confirm that constraining the regime length affects model inferences. Notice

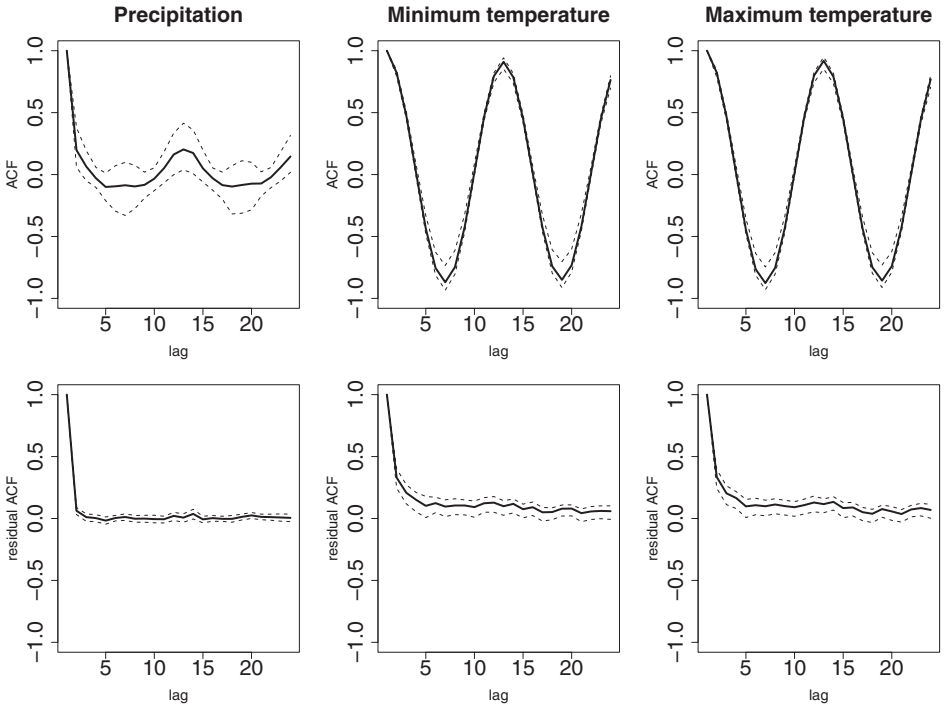


FIGURE 2 Autocorrelation functions, averaged over the 360 stations, for the observed data (first row) and the model residuals (second row). The dashed lines are the credible intervals and the solid lines the mean values

that a large value of m is bound to strongly affect inferences, for example, $m = 120$ implies that only a maximum of five CPs can be found in the 60 years observed time series. On the other hand, if $m \leq 12$ a seasonal variation can be confused with a CP. The 5-year time window is often chosen in the description of climatic phenomena, see, for example, the Global Climate Change NASA web site (https://climate.nasa.gov/vital-signs/global-temperature/) where the 5-year average variation of global surface temperatures are reported, or Collins et al. (2014) where again a 5-year time-window is adopted to compare changes in precipitation and temperatures and Mudelsee (2019) where "The bandwidth of 5 years was predefined to inspect mid- and shorter-term (decadal-scale) variations, such as the warming in the years around 1940, and to smooth away faster variations."

For each time-point and variable, we randomly selected 10% of the available stations and used them as a validation set, which was then comprised of 22,628 stations and time-points for $\mathbf{Y}_1$, 21,321 for $\mathbf{Y}_2$ and 21,272 for $\mathbf{Y}_3$. The values of the three response variables at these stations were then considered missing and models with $m \in \{1, 24, 60\}$ were estimated. The posterior samples of the variables in the validation set were used to compute the *continuous ranked probability scores* (Gneiting & Raftery, 2007) for each of the three thermopluviometric variables. The CRPS is frequently used in order to assess the accuracy of probabilistic forecasting models and can be thought of as the mean square error between the predicted and the true cumulative density functions. For a generic element of the validation set y the CRPS is computed as

$$\text{CRPS}(y) = \int_{-\infty}^{\infty} [F(x) - \mathbb{1}(y - x)]^2 dx,$$

where $F()$ is the posterior predictive distribution and $\mathbb{1}$ is the Heaviside step function. CRPS values, averaged over the validation set, are presented in Table 1. For the precipitation and minimum temperature $m = 60$ is clearly preferable, while the maximum temperature suggests $m = 24$. To choose a unique value of m for the three variables, we checked how many times the maximum temperature had smaller CRPS with $m = 24$ than with $m = 60$ in the validation set. This occurred with a relative frequency of 0.51, showing there is not strong evidence that $m = 24$ is preferable to $m = 60$. Then, in the following (Sections 4.2–4.4), we show results obtained fitting the model with $m = 60$ on the entire dataset.

Given the large number of stations (360) and multiple parameters (9), we define an automatic procedure to postprocess the model output and identify d-CPs. With the proposed model, the time when a CP occurs is itself a random variable

TABLE 1 Mean CRPS of the proposed model under different values of the minimum length of a regime (m)

	Prec.	Tmin.	Tmax
$m = 1$	1.0671	0.7059	0.8965
$m = 24$	1.0454	0.7033	0.8947
$m = 60$	1.0220	0.7018	0.8951

with some associated uncertainty. Then, even in the presence of d-CPs, it is very likely that adjacent time-points have overlapping HPD intervals. Therefore the comparison of HPD intervals for adjacent time-points would almost never detect a d-CP. Thus, we consider 12 time-points to be a time distance useful for HPD interval comparison to spot a d-CP, as it allows to compare the same months at 1-year distance, accounting for the seasonal variability. The automatic procedure begins by defining the indicator variable $\lambda_{t,\mathbf{s},\ell}$ that assumes value 1 if the 95% HPD intervals of parameter ℓ at times t and $t-12$ do not overlap, and 0 otherwise. Then, given a sequence of time-adjacent $\lambda_{t,\mathbf{s},\ell} = 1$, the time t corresponding to the first $\lambda_{t,\mathbf{s},\ell} = 1$ is defined to be a d-CP. The available data contain a large number of time-points with missing values for some or all the thermopluviometric variables. As a matter of fact, these time-points are not eligible to be d-CPs for the parameters of the corresponding missing variables. To take this into account, we adopt the conservative strategy of discarding CPs corresponding to time-points where the relative climate variables are missing. Notice that if, for example, only the minimum temperature is missing at time t, this will not affect the detection of a d-CP for the precipitation mean at the same time.

4.2 | Model comparison

As was mentioned in Section 1, the model we propose introduces some new features in standard CP modeling, such as sharing values for subsets of the parameters among stations and regimes. The same ideas can also be implemented in mixture models, where the no-return constraint is not accounted for. In this section we show that these particular features, in conjunction with the time constraint, produce a better description of the data with respect to competing models. Here, again, we use the same validation set described in the previous section and compare model performances in terms of CRPS. We compared the proposed DP-CP model with the three models described below.

4.2.1 | Ind-CP

In this case we assume that the data come from a CP model with $m = 60$, based on the trivariate normal density. Here stations and regimes cannot share the same parameter values, but the no-return constraint still holds. This can be easily achieved assuming that

$$G_{\mathbf{s}} \sim \mathrm{DP}(\gamma, H), \tag{7}$$

where $H = \prod_{\ell=1}^{9} H_{\theta_\ell} \in \Theta$, in Equation (6). Notice that for each $\mathbf{s}$ the base distribution $G_{\mathbf{s}}$ is constructed independently, then, since H is a continuous distribution, there are no atoms shared among stations.

4.2.2 | Ind-Mixture

This model is similar to the *Ind-CP* and does not allow that stations and regimes share the same parameter values. In this case the underlying classification is not based on a CP model but rather on a DP mixture model (without no-return constraint) with at least $m = 60$ observations in each regime. In more details $G_{\mathbf{s}}$ is here defined as in Equation (7) and $G_{\mathbf{s}}$ replaces $G_{\mathbf{s}}(\mathcal{I}_{t-1,\mathbf{s}})$ in Equation (6).

TABLE 2 Average CRPS for the competing models

	Prec.	Tmin.	Tmax
Proposed model	**1.022**	**0.7018**	**0.8951**
Ind-CP	1.052	0.7102	0.9023
Ind-Mixture	1.072	0.7102	0.8991
Shared-Mixture	1.034	0.7024	0.8983

4.2.3 | Shared-Mixture

This model is similar to the previous one since it is still a DP mixture model, but here $G_{\mathbf{s}}$ is defined as in Equation (5). This means that atoms are shared between stations and regimes.

As shown in Table 2, the proposed model outperforms all the others for the three response variables. Moreover, the comparison of *Ind-CP* with our proposal and of *Ind-Mixture* with *Shared-Mixture* proves that sharing atoms always improves the model performance. From a computational point of view, all four models require equivalent intense computations to be estimated.

4.3 | The Muravera, Ozieri, and Monte Cimone monitoring stations

The model output allows a very wide set of inferential usages, here we start by illustrating how to read results at the monitoring station level. As a proof of concept, we describe some of the results concerning two monitoring stations with a complex dynamic and one where the presence of a CP is very clear. The first two are Muravera and Ozieri, located in the Sardinia island at 19 and 390-m above sea level, respectively. The two stations are at the southern (Muravera) and northern (Ozieri) tips of the relative Ecoregion (2B4: Sardegna Section). Muravera is located close to the coast and shows a Mediterranean hot and dry bioclimatic character, while Ozieri is deeply inland and shows a Mediterranean wet bioclimatic character. It is known that at the beginning of the 1980s the intensity of the humid circulation in the Tyrrhenian sea increased, affecting the precipitation in coastal areas. This phenomenon did not touch the inland stations such as Ozieri. We then expect a larger number of possible changes in the Muravera than in the Ozieri station. However, these changes may not all be d-CPs, but they might rather correspond to smooth changes and/or missing values.

At the Muravera station the posterior mode of the number of occupied regimes $K_{\mathbf{s}}$ is 9, with probability 0.42, and 95% HPD interval [8, 11]. In the case of the Ozieri station the posterior mode of $K_{\mathbf{s}}$ is equal to 4, with probability 0.64, and the 95% HPD interval is [2, 6]. As we stated in the previous section, changes in regimes only correspond to potential d-CPs. Figure 3 shows the posterior 95% HPD intervals of the three mean parameters at each time-point, and allows to spot the difference between d-CPs and nd-CPs, as defined in Section 4.1. Notice that, due to the multimodality of the posteriors, the 95% HPD intervals are often composed by disjoint intervals, introducing a further complication in the identification of d-CPs. In Figure 3c, for example, we observe changes in the posterior distribution of the mean maximum temperature at Muravera at times ≈150 (June 1962) and ≈250 (October 1970). While in the first case the 95% HDP intervals before and after the change do not overlap, the converse is true at time 250. We can interpret the first as an abrupt change in the time series and call it a d-CP, while the second implies a small change in the corresponding parameter and is a nd-CP. There are quite a few missing data at the Ozieri monitoring station (Figure 3, second row) and, unlike at Muravera, we cannot spot CPs at times ≈150 or ≈250. For the mean of the minimum temperature in Figure 3e we are not sure when a d-CP occurs around time 400, due to the missing data, even if the posterior is really different before and after. Then, it is not possible to identify a d-CP in this case. At the two mentioned stations, our procedure detects only one d-CP at time 150 (June 1962) for the maximum temperature at the Muravera station.

The third monitoring station we consider is Monte Cimone, Aeronautica, where a clear increase in temperatures and precipitation occurs around the last two decades of the 20th century. The station is located at 2165-m above the sea level, at the uppermost sector of the Italian Apennine Ecoregion (Ecoregion 1C1: Northern and northwestern Apennine Section), and shows an Alpine cold bioclimatic character. In Figure 4 we report the observed time series and the HPD bounds of the three mean parameters for the Monte Cimone station. The model returns four CP's with probability 0.48. After checking for the presence of missing data at the CP location, we find that December 1999 is confirmed as a d-CP for the mean of the three variables, for the variance of maximum temperature, for the correlations between temperatures, and

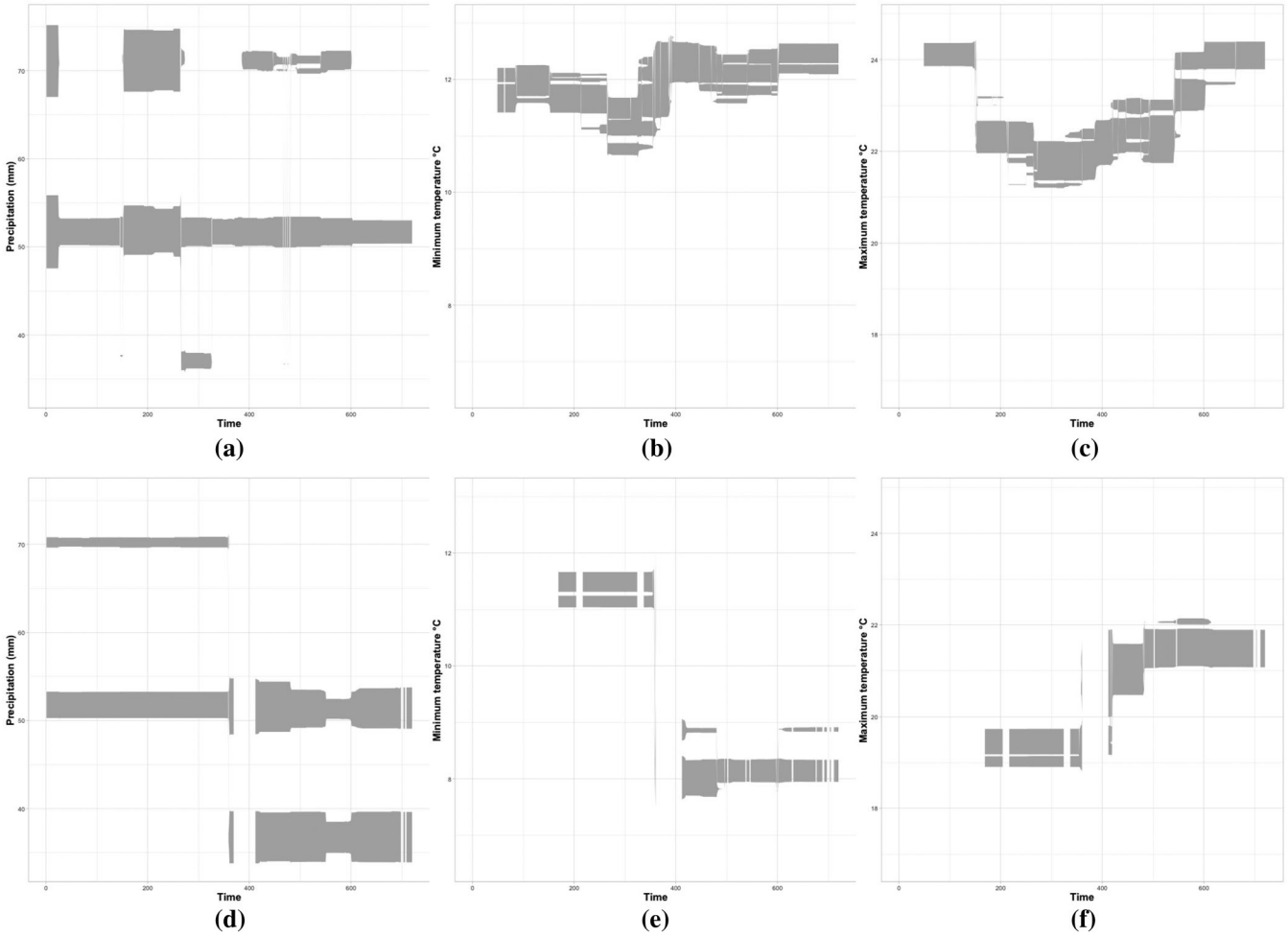


FIGURE 3 95% HPD bounds of the mean parameter for precipitation (first column), minimum temperature (second column), and maximum temperature (third column) at each time-point, for the Muravera (first row) and Ozieri (second row) monitoring stations. 95% HPD bounds are shown only for time-points with nonmissing values

for the correlation between precipitation and minimum temperature. The precipitation's variance instead, shows a d-CP on January 1952. HPD plots of correlations and variances are not reported for the sake of brevity, but they are available from the authors upon request

As stations can share the atoms of the underlying DP processes, the model output allows to check if one of the nine parameters has the same value at two monitoring stations. The relative frequency of the times that the value of the parameter ℓ is the same in the posterior samples for any pair of stations $\mathbf{s}_1$ and $\mathbf{s}_2$ at time t is hereby denoted by $\pi_{\mathbf{s}_1,\mathbf{s}_2,t,\ell}$ and referred to as *similarity*. As expected, the similarity of the mean of the maximum temperature for the Muravera and Ozieri stations (Figure 5a) is quite low at all time-points. According to the intensity of the humid circulation in the area, after December 1979 (time-point 360) the mean of the precipitation shows a decreasing similarity between the two stations (Figure 5b).

4.4 | Summary of model results for all monitoring stations

In Figure 6 stations with at least one d-CP over the entire time-period are reported for the nine parameters. Many stations have no d-CPs (all of them for the correlation between precipitation and maximum temperature), while d-CPs are found more often for the mean of the two temperatures and the for correlation between them. Owing to the intrinsic variability of the precipitation in Mediterranean contexts (Dükeloh & Jacobeit, 2003), d-CPs for this variable sporadically occur in the Mediterranean ecoregions, involving exclusively areas in mountain sectors, while being relatively more widespread in the

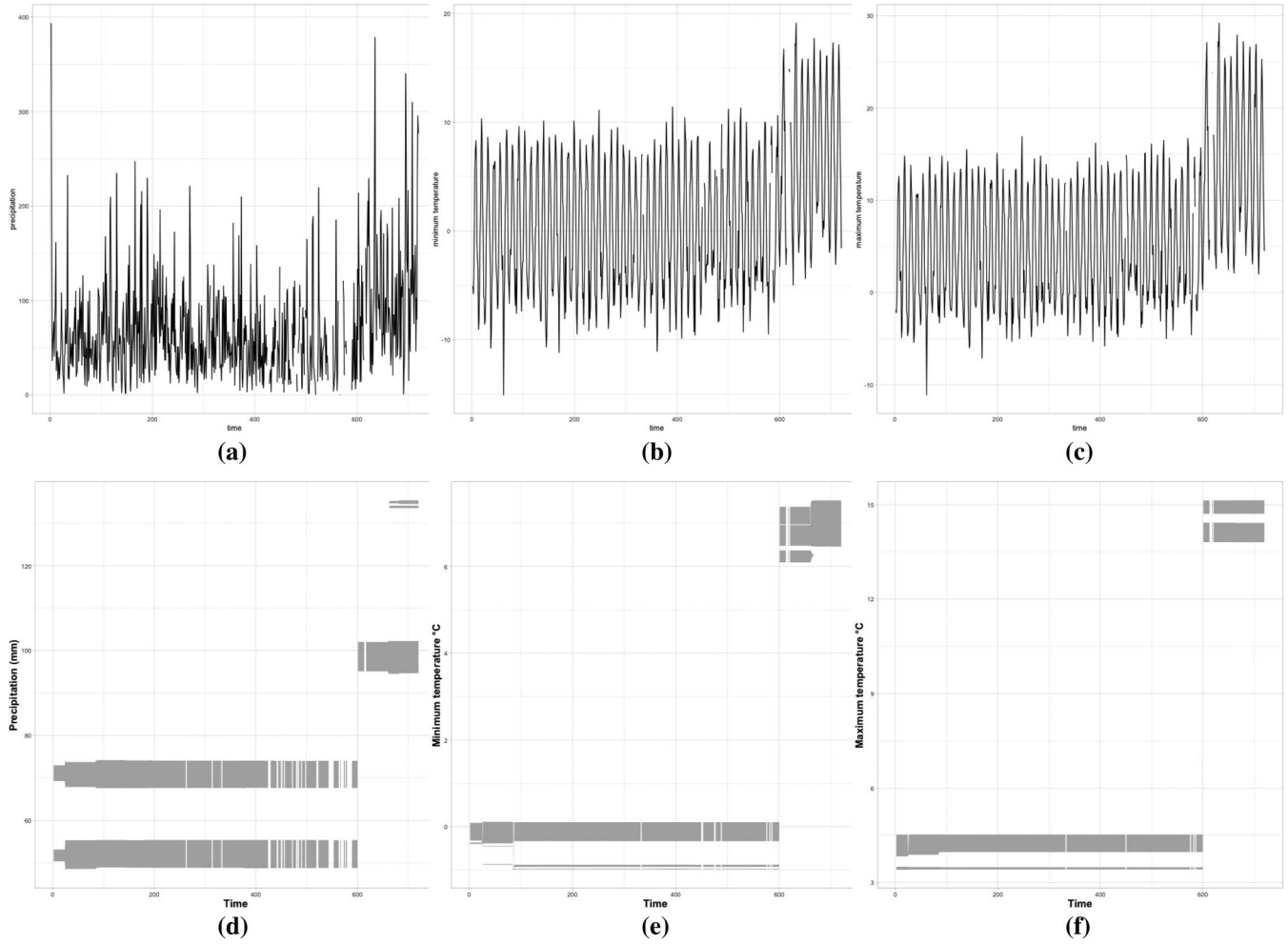


FIGURE 4 Time series (first row) and 95% HPD bounds (second row) of the mean parameters for precipitation (first column), minimum temperature (second column), and maximum temperature (third column) at the Monte Cimone monitoring station

Temperate ecoregions (Alpine and Apennine sections). Similarly, as regards temperatures, fewer changes are observed in the more continental Sectors (especially 1B1, Po Plain Section, and 2C1, Central Adriatic Section), which are intrinsically subject to daily and seasonal temperature variations.

We can also evaluate how many d-CPs are observed simultaneously for each pair of parameters ℓ_1 and ℓ_2. For this purpose we can use the indicator variable $\lambda_{t,\mathbf{s},\ell}$ defined in Section 4.1. Let $\mathcal{P}_{\mathbf{s},\ell_i}$ be the set of temporal indices that have $\lambda_{t,\mathbf{s},\ell_i} = 1$ and let $n_{\mathcal{P},\mathbf{s},\ell_i}$ be its cardinality at location $\mathbf{s}$. The proportion of times in which $\lambda_{t,\mathbf{s},\ell_2} = 1$ conditionally on $\lambda_{t,\mathbf{s},\ell_1} = 1$ is given by:

$$\frac{\sum_{\mathbf{s}\in S}\sum_{t\in\mathcal{P}_{\mathbf{s},\ell_1}} I(\lambda_{t,\mathbf{s},\ell_2}=1)}{\sum_{\mathbf{s}\in S} n_{\mathcal{P},\mathbf{s},\ell_1}}. \tag{8}$$

Equation (8) gives the proportion of times and stations where ℓ_2 has a d-CP conditional on ℓ_1 having a d-CP. For each pair of variables and parameters, these indices are depicted in Figure 7, where ℓ_1 represents the row of the matrix and ℓ_2 the column. As a proof of concept, notice that whenever we have a d-CP in the correlation between the minimum temperature and the precipitation there is a large probability of observing a d-CP in the correlation between the two temperatures (row 7, column 9) and the mean of the minimum temperature (row 7, column 2). Figure 7 also shows that d-CPs in the mean of the two temperatures are often caused by those in the mean precipitation (row 1, columns 2, 3), but the converse is not true (rows 1, 2, column 1). As expected, Figure 7 shows that the means of the two temperatures often have concurrent d-CPs (row 2 column 3, row 3 column 2).

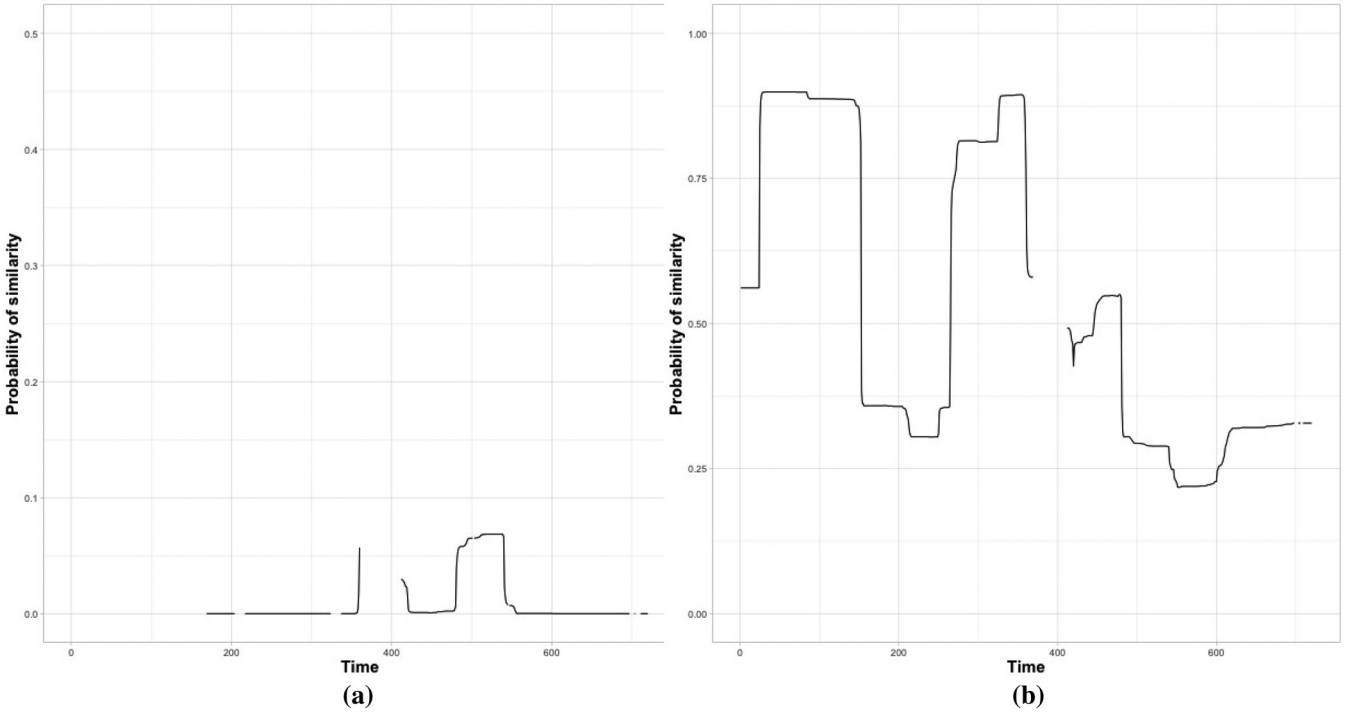


FIGURE 5 Time series of similarities between the Muravera and Ozieri stations for the mean of maximum temperature (a) and of the precipitation (b)

Given the large number of stations and time-points, we summarize the similarities $\pi_{\mathbf{s}_1,\mathbf{s}_2,t,\ell}$ between stations $\mathbf{s}_1$ and $\mathbf{s}_2$ for parameter ℓ at time t by regressing their logistic transformation $\log(\pi_{\mathbf{s}_1,\mathbf{s}_2,t,\ell}/(1-\pi_{\mathbf{s}_1,\mathbf{s}_2,t,\ell}))$ on the spatial distances Δ_{sp} between stations $\mathbf{s}_1$ and $\mathbf{s}_2$, the differences in their elevation Δ_{el} and a grouping factor that identifies the combination of the ecoregions (14 sections) of the two stations. Notice that, for each parameter ℓ, we consider the couple of stations $(\mathbf{s}_1, \mathbf{s}_2)$ at time t only if both do not have missing values in the corresponding variable. All the regression coefficients of Δ_{sp} and Δ_{el} suggest that the similarity decreases with the spatial distance and with the elevation difference. In Table 3 we report the coefficients referred to the mean and variance parameters. Notice that, while the distance between stations seems to have approximately the same effect for the three mean parameters, similarities are more sensitive to distance effects for the precipitation variance than for the temperature ones. Indeed rainfall and snow events are less linked to geographical variables than temperature (Ninyerola et al., 2000). As expected, differences in elevation affect the similarity of the mean temperatures more than they do with the precipitation.

For each of the nine parameters, Figure 8 describes the relative variation of the logit of the similarities between pairs of ecoregions predicted by the regression model using the mean spatial distance and mean elevation difference between the ecoregions. Notice that, even if we do not use spatial information in the model, we can see in Figure 8 that regions that are spatially close tend to be more similar, which confirms that the results we obtain are sound.

According to the diagnostic criteria adopted for drawing ecoregion boundaries (Blasi et al., 2014), the similarity in biophysical features between ecoregional sections is generally expected to increase with their belonging to the same higher tier of the classification (province or division). Results in Figure 8 allow this hypothesis to be verified and detailed as regards climatic characteristics. For example, dissimilarities between the Italian Temperate division and the Mediterranean division (set of ecoregions with codes alternatively beginning with "1" or "2") emerged to be more marked in terms of mean temperatures (Figure 8b,c) rather than of mean precipitation (Figure 8a), for the same reason above-mentioned. At this level, precipitation means become differential just for a subset of the Temperate ecoregions (namely, the Alpine sections 1A1 and 1A2, characterized by higher precipitation values and clearly dissimilar from all the Mediterranean sections) or, alternatively, for a subset of the Mediterranean ecoregions (namely, the southernmost sections 2B3, 2B4, 2C2, characterized by a very marked summer decrease and clearly dissimilar from most of the Temperate sections). As regards provinces (codes denoted by the first number plus the first letter) within the same division, the Alpine Province (codes beginning with "1A") is quite well fitting within the Temperate division in terms of maximum temperatures variance (Figure 8f) and correlation between temperatures (Figure 8i), while emerged to be dissimilar from the other Temperate

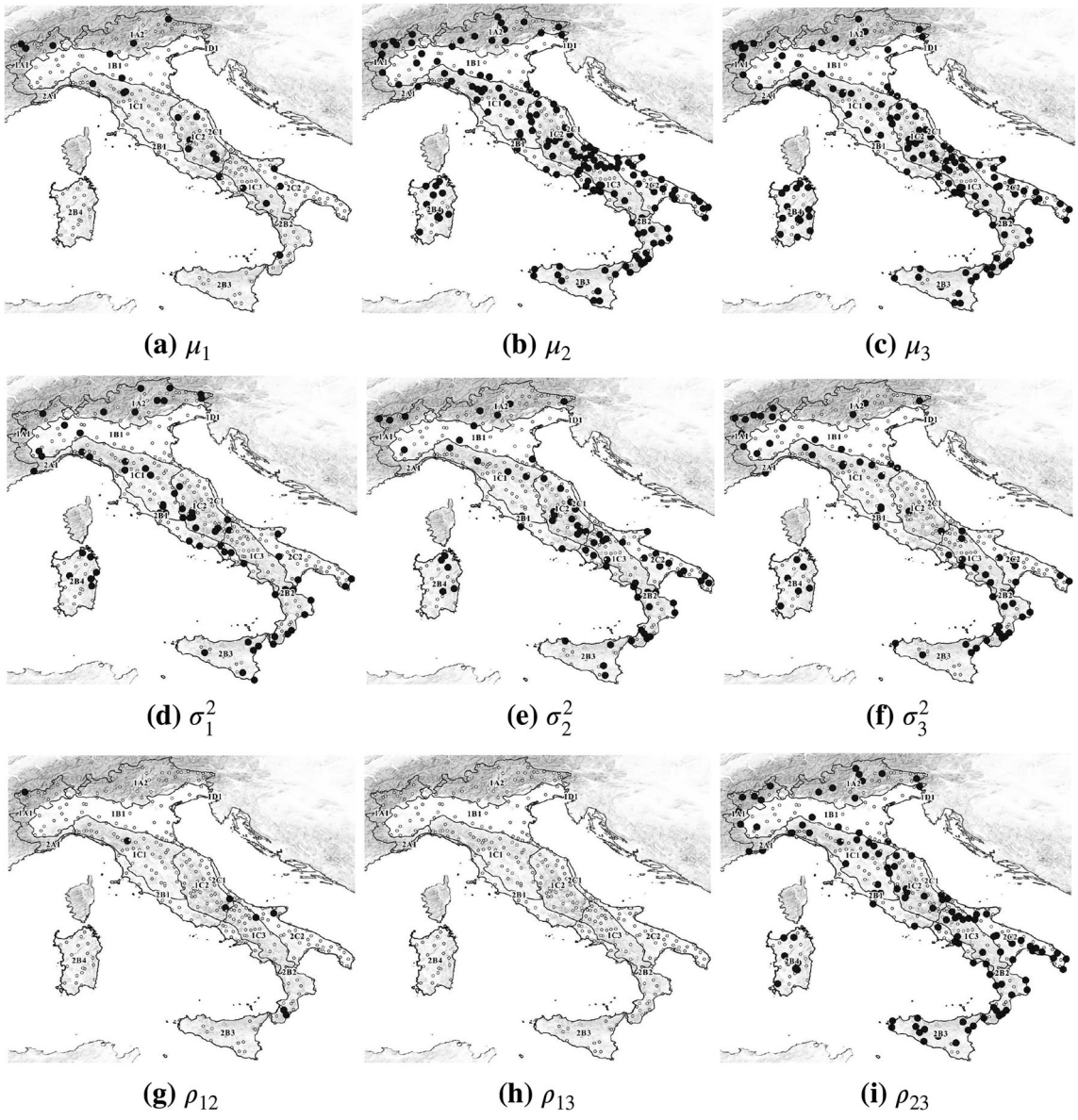


FIGURE 6 Stations with at least one d-CP, for the nine parameters

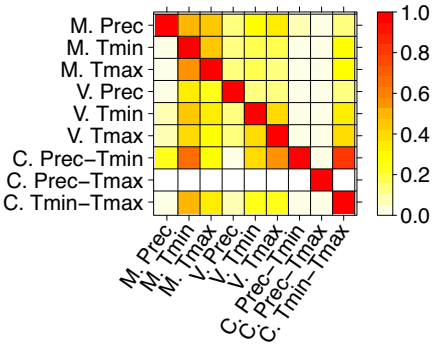


FIGURE 7 Proportion of times and stations where the parameter ℓ_2 (columns) has a d-CP conditional on ℓ_1 (rows) having a d-CP. Each square represents the proportion of times and stations where $\lambda_{t,\mathbf{s},\ell}$ of the column element is equal to one, in the subset of times and stations where $\lambda_{t,\mathbf{s},\ell}$ of the row element is equal to 1

TABLE 3 Estimated regressive coefficients referred to similarities of mean and variance parameters (with confidence intervals)

	Δ_{sp}	Δ_{el}
Mean prec.	−1.313	−0.061
	(−1.323, −1.302)	(−0.610, −0.604)
Mean min. temp.	−1.447	−1.125
	(−1.454, −1.439)	(-1.127 -1.123)
Mean max. temp.	−1.464	−1.937
	(−1.472, −1.456)	(−1.939, −1.935)
Var. prec.	−2.244	−0.477
	(−2.255, −2.232)	(−0.481, −0.473)
Var. min. temp.	−1.395	−1.058
	(−1.405, −1.386)	(-1.062, −1.056)
Var. max. temp.	−1.757	−0.325
	(−1.767, −1.747)	(−0.322, −0.329)

provinces for mean temperatures (Figure 8b,c) and for precipitation means and variance (Figure 8a,d). This distinctiveness may be ascribed, on the one hand, to longer and deeper winter frost with respect to that occurring in the Po Plain Province at lower altitudes (code beginning with "1B") and in the Apennine Province at lower latitudes (codes beginning with "1C"), and, on the other hand, to higher values and continental regime for precipitation (with winter minimum) with respect to lower values and summer minimums occurring in the peninsular sectors at lower latitudes. As regards sections, some transitional characters could instead be detected. For example, the Central Adriatic Section (2C1) is quite well fitting within the Adriatic Province (codes beginning with "2C") of the Mediterranean division for all the parameters, but many similarities with the Po Plain and Apennine Provinces of the Temperate division also emerged (especially in terms of mean and variance of precipitation, mean and variance of maximum temperatures, and all the correlations).

5 | CONCLUSIONS

In this work we propose a model-based CP detection procedure for large multivariate time series. The motivating example spreads from a climate dataset with monthly values of precipitation, minimum, and maximum temperature recorded in Italy from a network of 360 monitoring stations over 60 years. Our proposal introduces some remarkable novelties in CP modeling: we do not force all parameters to change at a CP; changing parameters are not set a priori but their identification is part of the model inference; monitoring stations can share values of the parameters; at each monitoring station the number of CPs is a random quantity that is estimated. As a by-product, the model output allows to identify clusters of spatiotemporal observations that share values of some or all the parameters, that is, to identify time regimes or groups of stations that are similar all over the observed time-window. Model output postprocessing also informs on which parameters are more likely to change simultaneously, and provides a procedure to distinguish abrupt from smooth changes. Actually, the model enabled to support the identification of abrupt changes in individual climate features and in their mutual relationships, which potentially impair the resilience of Italian natural ecosystems.

It is known that the presence of positive autocorrelation in the time series is a key issue which can eventually degrade climate CP detection. Nevertheless, explicitly accounting for time dependence in the model formulation would obviously add considerably to the computational complexity. However, we showed that our proposal is capable of handling time dependence effectively (see Figure 2). The high-computational complexity of our proposal is counterbalanced by the richness of the model output. Indeed, it allows to comment on the behavior of single time series and on their joint behavior in both time and space. Local and general features are easily highlighted and we obtain a very relevant insight into the analyzed phenomena. A unique feature of our proposal is that it allows to postprocess the model output at different spatial scales, thus giving the possibility of evaluating the similarities between stations and/or areas in terms of CPs. A further

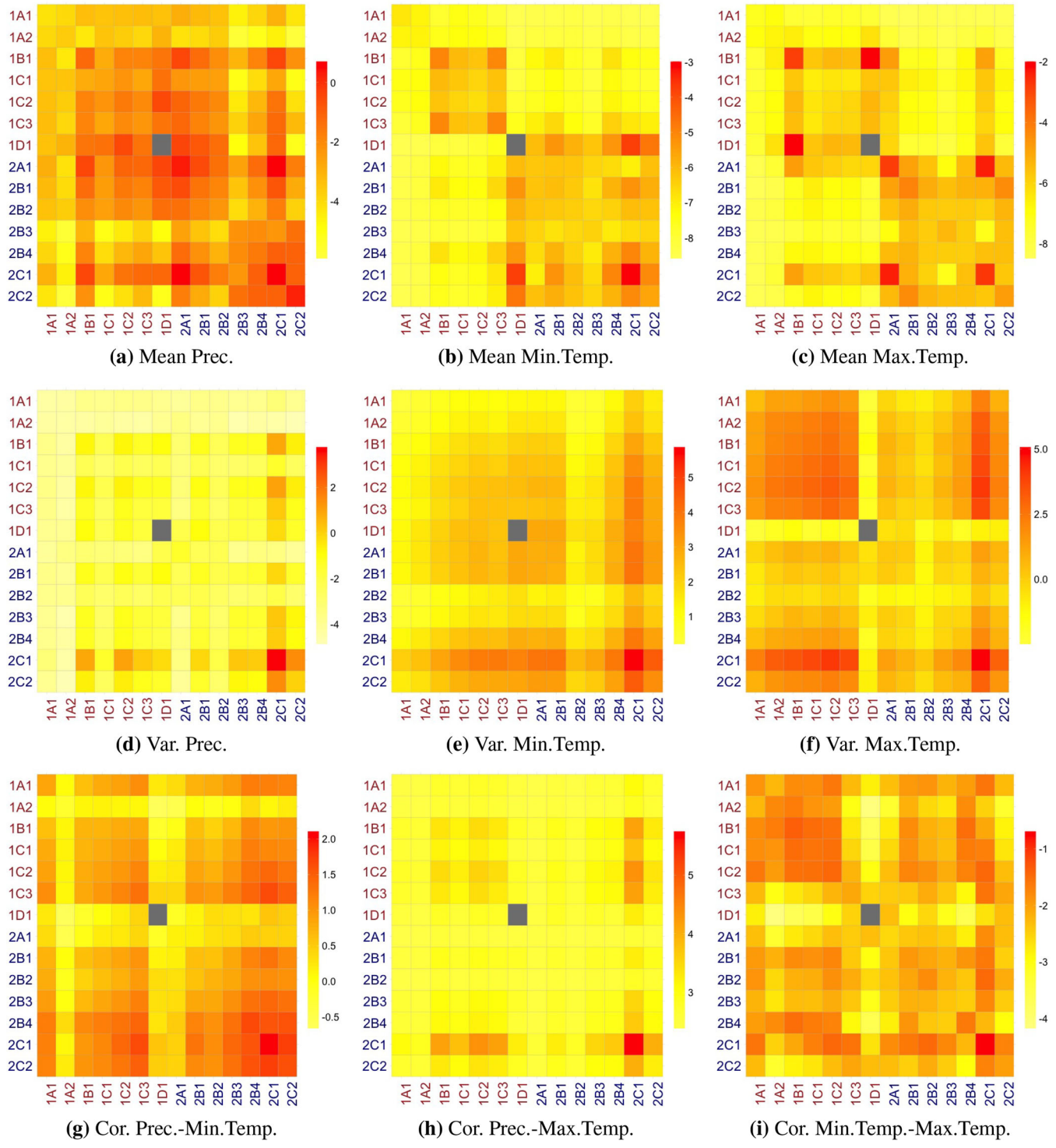


FIGURE 8 Predictive values of the logit of the similarities between pairs of ecoregions (sections), obtained using the mean value of Δ_{sp} and Δ_{el}. As ecoregion 1D1 is composed of one station we do not compute the self-similarity, leaving the relative square grey. To ease the comparison between ecoregions, the color scale is made specific for each panel

improvement on standard CP search protocols is that the choice of the minimum regime length can be done in a rigorous and verifiable way using the available data, as shown in Section 4.1, reducing the level of discretion in the model setting.

The inferences allowed by the proposed model significantly improve the bioclimatic characterization of the Italian ecoregions and, therefore, the understanding of complex climate patterns in a country of great physiographic heterogeneity. In particular, with the support of observed rather than interpolated data, it makes possible to better define i) the similarity/dissimilarity within and between ecoregions in terms of climate mean and variability, ii) the effects of the different correlation between climatic variables on vegetation patterns, with a focus at the national/subnational scale, and iii) the vulnerability of sensitive ecosystems to local climate change trends, in accordance with the IUCN Red List criteria (Keith et al., 2015). Actually, the presented outcomes provide useful benchmarks to better focus monitoring programs as regards ecosystem responses (e.g., Chelli et al., 2017). It may be possible, for example, to determine according to which parameters the potential vulnerability to climate change of the Mediterranean ecosystems differs from that of the Temperate ones. Otherwise, the detection of d-CP occurrence across ecoregional sections provides a better spatial locationing of change trends with respect to available models at broader scales, for example, by distinguishing d-CP frequency between continental plains and mountain ranges within the Temperate context.

Finally, although the model is presented in the context of a CP problem, the same ideas can be used in other mixture-type models, where it is of interest to evaluate which elements of a vector-valued parameter differ between regimes or sets of observations, allowing to enhance common and different features.

ACKNOWLEDGMENT

The authors acknowledge the support of the Italian Ministry of Education, University, and Research (MIUR), grant *Dipartimenti di Eccellenza*, CUP: E11G18000350001, conferred to Dipartimento di Scienze Matematiche - DISMA, Politecnico di Torino. Open Access Funding provided by Politecnico di Torino within the CRUI-CARE Agreement. [Correction added on 19 May 2022, after first online publication: CRUI funding statement has been added.]

The data that support the findings of this study are available from http://www.scia.isprambiente.it/wwwrootscia/Home_new.html. Restrictions apply to the availability of these data, which were used under license for this study.

ORCID

Gianluca Mastrantonio https://orcid.org/0000-0002-2963-6729

REFERENCES

Alley, R. B., Marotzke, J., Nordhaus, W. D., Overpeck, J. T., Peteet, D. M., Pielke, R. A., Pierrehumbert, R. T., Rhines, P. B., Stocker, T. F., Talley, L. D., & Wallace, J. M. (2003). Abrupt climate change. *Science*, *299*(5615), 2005–2010.

Bailey, R. G. (1983). Delineation of ecosystem regions. *Environmental Management*, *7*(4), 365–373.

Bailey, R. G. (2004). Identifying ecoregion boundaries. *Environmental Management*, *34*(Suppl 1), S14–S26.

Battaglia, F., Cucina, D., & Rizzo, M. (2019). Parsimonious periodic autoregressive models for time series with evolving trend and seasonality. *Statistics and Computing*, *30*(1), 77–91.

Beaulieu, C., Chen, J., & Sarmiento, J. L. (2012). Change-point analysis as a tool to detect abrupt climate variations. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *370*(1962), 1228–1249.

Bhattacharya, P. (1987). Maximum likelihood estimation of a change-point in the distribution of independent random variables: General multiparameter case. *Journal of Multivariate Analysis*, *23*(2), 183–208.

Blasi, C., Capotorti, G., Copiz, R., Guida, D., Mollo, B., Smiraglia, D., & Zavattero, L. (2014). Classification and mapping of the ecoregions of Italy. *Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology*, *148*(6), 1255–1345.

Blasi, C., Capotorti, G., Copiz, R., & Mollo, B. (2018). A first revision of the Italian ecoregion map. *Plant Biosystems*, *152*(6), 1201–1204.

Carlin, B. P., Gelfand, A. E., & Smith, A. F. M. (1992). Hierarchical Bayesian analysis of changepoint problems. *Journal of the Royal Statistical Society Series C (Applied Statistics)*, *41*(2), 389–405.

Chelli, S., Wellstein, C., Campetella, G., Canullo, R., Tonin, R., Zerbe, S., & Gerdol, R. (2017). Climate change response of vegetation across climatic zones in Italy. *Climate Research*, *71*(3), 249–262.

Chib, S. (1998). Estimation and comparison of multiple change-point models. *Journal of Econometrics*, *86*(2), 221–241.

Collins, M., Knutti, R., Arblaster, J., Dufresne, J-L., Fichefet, T., Friedlingstein, P., Gao, X., Gutowski, W., Johns, T., Krinner, G., Shongwe, M., Tebaldi, C., Weaver, A., & Wehner, M. (2014). *Long-term climate change: Projections, commitments and irreversibility*. In T.F. Stocker, D. Qin, G-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex & P.M. Midgley, (eds.), *Climate change 2013 – The physical science basis: Working group i contribution to the fifth assessment report of the intergovernmental panel on climate change* (pp. 1029–1136). Cambridge University Press.

Dükeloh, A., & Jacobeit, J. (2003). Circulation dynamics of mediterranean precipitation variability 1948-98. *International Journal of Climatology*, *23*(15), 1843–1866.

Felton, A. J., & Smith, M. D. (2017). Integrating plant ecological responses to climate extremes from individual to ecosystem levels. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1723), 20160142.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, *1*(2), 209–230.

Fick, S. E., & Hijmans, R. J. (2017). Worldclim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315.

Fox, E. B., Sudderth, E. B., Jordan, M. I., & Willsky, A. S. (2011). A sticky HDP-HMM with application to speaker Diarization. *The Annals of Applied Statistics*, *5*(2A), 1020–1056.

Frauendorf, T. C., MacKenzie, R. A., Tingley, R. W., III, Frazier, A. G., Riney, M. H., & El-Sabaawi, R. W. (2019). Evaluating ecosystem effects of climate change on tropical Island streams using high spatial and temporal resolution sampling regimes. *Global Change Biology*, *25*(4), 1344–1357.

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*(477), 359–378.

Hawkins, D. M. (2001). Fitting multiple change-point models to data. *Computational Statistics and Data Analysis*, *37*(3), 323–341.

James, N. A., & Matteson, D. S. (2013). ecp: An r package for nonparametric multiple change point analysis of multivariate data. *Journal of Statistical Software*, *62*(7), 1–25.

Jandhyala, V. K., Fotopoulos, S. B., & You, J. (2010). Change-point analysis of mean annual rainfall data from Tucumán, Argentina. *Environmetrics*, *21*(7-8), 687–697.

Keith, D. A., Rodríguez, J. P., Brooks, T. M., Burgman, M. A., Barrow, E. G., Bland, L., Comer, P. J., Franklin, J., Link, J., McCarthy, M. A., Miller, R. M., Murray, N. J., Nel, J., Nicholson, E., Oliveira-Miranda, M. A., Regan, T. J., Rodríguez-Clark, K. M., Rouget, M., & Spalding, M. D. (2015). The IUCN red list of ecosystems: Motivations, challenges, and applications. *Conservation Letters*, *8*(3), 214–226.

Killick, R., Eckley, I. A., Ewans, K., & Jonathan, P. (2010). Detection of changes in variance of oceanographic time-series using changepoint analysis. *Ocean Engineering*, *37*(13), 1120–1126.

Ko, S. I. M., Chong, T. T. L., & Ghosh, P. (2015). Dirichlet process hidden Markov multiple change-point model. *Bayesian Analysis*, *10*(2), 275–296.

Li, Y., Chan, N. H., Yau, C. Y., & Zhang, R. (2021). Group orthogonal greedy algorithm for change-point estimation of multivariate time series. *Journal of Statistical Planning and Inference*, *212*, 14–33.

Lindeløv, J. K. (2020). mcp: An r package for regression with multiple change points. *OSF Preprints*.

Loveland, T. R., & Merchant, J. M. (2004). Ecoregions and ecoregionalization: Geographical and ecological perspectives. *Environmental Management*, *34*(Suppl 1), S1.

Lu, Q., Lund, R., & Lee, T. C. M. (2010). An MDL approach to the climate segmentation problem. *The Annals of Applied Statistics*, *4*(1), 299–319.

Lund, R., & Reeves, J. (2002). Detection of undocumented changepoints: A revision of the two-phase regression model. *Journal of Climate*, *15*(17), 2547–2554.

Lund, R., Wang, X. L., Lu, Q. Q., Reeves, J., Gallagher, C., & Feng, Y. (2007). Changepoint detection in periodic and autocorrelated time series. *Journal of Climate*, *20*(20), 5178–5190.

Luque-Espinar, J. A., Mateos, R. M., García-Moreno, I., Pardo-Igúzquiza, E., & Herrera, G. (2017). Spectral analysis of climate cycles to predict rainfall induced landslides in the Western Mediterranean (Majorca, Spain). *Natural Hazards*, *89*(3), 985–1007.

Mastrantonio, G., Jona Lasinio, G., Pollice, A., Capotorti, G., Teodonio, L., Genova, G., & Blasi, C. (2019). A hierarchical multivariate spatio-temporal model for clustered climate data with annual cycles. *The Annals of Applied Statistics*, *13*(2), 797–823.

McDonald, J. F., & Moffitt, R. A. (1980). The uses of tobit analysis. *The Review of Economics and Statistics*, *62*(2), 318–321.

McLachlan, C., & Peel, D. (2000). *Finite mixture models*. John Wiley & Sons.

Metzger, M. J., Bunce, R. G. H., Jongman, R. H. G., Sayre, R., Trabucco, A., & Zomer, R. (2013). A high-resolution bioclimate map of the world: A unifying framework for global biodiversity research and monitoring. *Global Ecology and Biogeography*, *22*(5), 630–638.

Mudelsee, M. (2019). Trend analysis of climate time series: A review of methods. *Earth-Science Reviews*, *190*, 310–322.

Ninyerola, M., Pons, X., & Roure, J. M. (2000). A methodological approach of climatological modelling of air temperature and precipitation through gis techniques. *International Journal of Climatology*, *20*(14), 1823–1841.

OpenMP Architecture Review Board (2008). OpenMP application program interface version 3.0.

Pecl, G. T., Araújo, M. B., Bell, J. D., Blanchard, J., Bonebrake, T. C., Chen, I.-C., Clark, T. D., Colwell, R. K., Danielsen, F., Evengård, B., Falconi, L., Ferrier, S., Frusher, S., Garcia, R. A., Griffis, R. B., Hobday, A. J., Janion-Scheepers, C., Jarzyna, M. A., Jennings, S., … Williams, S. E. (2017). Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. *Science*, *355*(6332), 1–9.

Peluso, S., Chib, S., & Mira, A. (2019). Semiparametric multivariate and multiple change-point modeling. *Bayesian Anal*, *14*(3), 727–751.

Petrillo, F. U., & Guerriro, R. (2014). *Terastat computer cluster for high performance computing*. Department of Statistical Science Sapienza University of Rome. http://www.dss.uniroma1.it/en/node/6554

Pitman, J. (2002). *Combinatorial stochastic processes. Technical report 621, Lecture notes for St. Flour course*. Department of Statistics.

Robbins, M., Gallagher, C., Lund, R., & Aue, A. (2011). Mean shift testing in correlated data. *Journal of Time Series Analysis*, *32*(5), 498–511.

Rodionov, S. N. (2004). A sequential algorithm for testing climate regime shifts. *Geophysical Research Letters*, *31*(9), 1–4.

Samé, A., Chamroukhi, F., Govaert, G., & Aknin, P. (2011). Model-based clustering and segmentation of time series with changes in regime. *Advances in Data Analysis and Classification*, *5*(4), 301–321.

Teh, Y. W., Jordan, M. I., Beal, M. J., & Blei, D. M. (2006). Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, *101*(476), 1566–1581.

Tomè, A. R., & Miranda, P. M. A. (2004). Piecewise linear fitting and trend changing points of climate parameters. *Geophysical Research Letters*, *31*(2), 1–4.

van Dyk, D. A., & Park, T. (2008). Partially collapsed gibbs samplers. *Journal of the American Statistical Association*, *103*(482), 790–796.

Van Gael, J., Saatci, Y., Teh, Y. W., & Ghahramani, Z. (2008). Beam sampling for the infinite hidden Markov model. *Proceedings of the 25th International Conference on Machine Learning, ICML '08* (pp. 1088–1095), New York, NY: ACM.

Walther, G.-R. (2010). Community and ecosystem responses to recent climate change. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1549), 2019–2024.

Williams, J. W., Blois, J. L., & Shuman, B. N. (2011). Extrinsic and intrinsic forcing of abrupt ecological change: Case studies from the late quaternary. *Journal of Ecology*, *99*(3), 664–677.

Yu, D., Liu, Y., Shi, P., & Wu, J. (2019). Projecting impacts of climate change on global terrestrial ecoregions. *Ecological Indicators*, *103*, 114–123.

How to cite this article: Mastrantonio, G., Jona Lasinio, G., Pollice, A., Teodonio, L., & Capotorti, G. (2022). A Dirichlet process model for change-point detection with multivariate bioclimatic data. *Environmetrics*, *33*(1), e2699. https://doi.org/10.1002/env.2699

APPENDIX A. THE ALGORITHM

When using MCMC to estimate mixture-type models, extra care has to be given to simulate missing data without decreasing the convergence speed. This is mainly due to the tendency to obtain regimes fully composed of missing values, with the corresponding sets of likelihood parameters estimated by nonobserved data. An obvious consequence is an increase in the number of occupied regimes A strategy would imply integrating out the missing data, but this would make the sampling of the likelihood parameters more time consuming and complex under the proposed model, with no real advantage in terms of the quality of the estimates. We then adopt a mixed-strategy where, at each iteration, i) we first sample all parameters except for $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ using a likelihood marginalized with respect to the missing data, ii) we then simulate the missing values and, finally, iii) we draw samples of $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ from their full conditionals based on the full likelihood. Working with the marginalized-likelihood mitigates the problem of clusters composed by missing values, while likelihood parameters are easily sampled from the full likelihood. The proposed strategy produces a valid MCMC algorithm as shown in van Dyk and Park (2008) (see for example their example 8, or section 3).

In the following, we first show the form of the likelihood based only on nonmissing data, then the MCMC steps are discussed in detail.

A.0.1 Likelihood over nonmissing data

For the model response variable $\mathbf{y}$, obtained standardizing and transforming the observed data $\mathbf{y}^*$ by (1), we have two kinds of missing values:

1. some or all the components of $\mathbf{y}_{t,\mathbf{s}}$ are missing because we have missing values in $\mathbf{y}^*_{t,\mathbf{s}}$;
2. the value of $y_{1,t,\mathbf{s}}$ is missing because $y^*_{1,t,\mathbf{s}}$ is equal to zero.

Let $\mathbf{y}^{\text{obs}}_{t,\mathbf{s}}$ contain only nonmissing variables observed at time t and location $\mathbf{s}$. Notice that $\mathbf{y}^{\text{obs}}_{t,\mathbf{s}}$ can also be an empty set. Conditional independence of vectors $\mathbf{y}_{t,\mathbf{s}}$ given the model components and normality imply that the likelihood of $\mathbf{y}^{\text{obs}}|\boldsymbol{\psi},\boldsymbol{\theta}$ is simply specified by the densities $f(\mathbf{y}^{\text{obs}}_{t,\mathbf{s}}|\psi_{t,\mathbf{s}},\boldsymbol{\theta}_{t,\mathbf{s}})$, with self-explicative notation. With the first kind of missing data $f(\mathbf{y}^{\text{obs}}_{t,\mathbf{s}}|\psi_{t,\mathbf{s}},\boldsymbol{\theta}_{t,\mathbf{s}})$ is given by

$$\phi_{\text{obs}}\left(\mathbf{y}^{\text{obs}}_{t,\mathbf{s}}|\psi^{\text{obs}}_{\ell,\mathbf{s}}+\boldsymbol{\mu}^{\text{obs}}_{t,\mathbf{s}},\boldsymbol{\Sigma}^{\text{obs}}_{t,\mathbf{s}}\right),$$

assuming $\phi_{\text{obs}}(\cdot,\cdot,\cdot)=1$ when obs $\equiv \emptyset$.

With regard to the second kind of missing data, Equation (1) tells that the probability of $y^*_{t,\mathbf{s}}=0$ corresponds to the probability that $y_{1,t,\mathbf{s}}\leq 0$, then in this case $f(\mathbf{y}^{\text{obs}}_{t,\mathbf{s}}|\psi_{t,\mathbf{s}},\boldsymbol{\theta}_{t,\mathbf{s}})$ is given by:

$$\int_{-\infty}^{0}\phi_{\text{obs}}\left(\mathbf{y}^{\text{obs}}_{t,\mathbf{s}}|\psi^{\text{obs}}_{\ell,\mathbf{s}}+\boldsymbol{\mu}^{\text{obs}}_{t,\mathbf{s}},\boldsymbol{\Sigma}^{\text{obs}}_{t,\mathbf{s}}\right)dy_{1,t,\mathbf{s}}.$$

A.0.2 MCMC steps

Sampling $\theta_{t,\mathbf{s}}$

We define $\mathcal{K}_{\mathbf{s}}^{o}$ as the set containing the $K_{\mathbf{s}}$ indices of $\boldsymbol{\eta}_k$ that have been observed at station $\mathbf{s}$, ordered by their temporal appearance, and $\mathcal{K}_{\mathbf{s}}$ as the ordered set of all natural numbers with the elements of $\mathcal{K}_{\mathbf{s}}^{o}$ in the first $K_{\mathbf{s}}$ positions. We also define $k_{\mathbf{s}}$, $l_{\mathbf{s}}$, and $1_{\mathbf{s}}$ as, respectively the kth, lth and first element of $\mathcal{K}_{\mathbf{s}}$.

As in the standard mixture-type model, it is generally easier to work with $\mathbf{z}_{\mathbf{s}}$ to sample $\boldsymbol{\theta}_{t,\mathbf{s}}$ since the relation $\boldsymbol{\theta}_{t,\mathbf{s}} = \boldsymbol{\eta}_{z_{t,\mathbf{s}}}$ holds. We use a two-step strategy, composed of a *merging* step and a *splitting* step. Let $k_{\mathbf{s}}^{+}$ and $k_{\mathbf{s}}^{-}$ be the predecessor and the successor of $k_{\mathbf{s}}$ in $\mathcal{K}_{\mathbf{s}}$; notice that they respectively correspond to an empty regime if $k_{\mathbf{s}}$ is the first or last element of $\mathcal{K}_{\mathbf{s}}^{o}$. In the merging step three possible moves are available for the $k_{\mathbf{s}}$th regime: it is merged with the $k_{\mathbf{s}}^{-}$th regime with probability proportional to p^{-}; it is merged with the $k_{\mathbf{s}}^{+}$th regime with probability proportional to p^{+}; it is not merged with probability proportional to p.

To compute the probabilities, we use (6).

Then, we have

$$p^{-} = \frac{\pi_{\mathbf{s},k_{\mathbf{s}}}^{n_{\mathbf{s},k_{\mathbf{s}}}^{T}} \prod_{t:z_{t,\mathbf{s}}=k_{\mathbf{s}}} f(\mathbf{y}_{t,\mathbf{s}}^{\text{obs}}|\boldsymbol{\psi}_{\mathbf{s}}, \boldsymbol{\eta}_{k_{\mathbf{s}}^{-}})}{\left(\sum_{l=k-1}^{\infty}\pi_{\mathbf{s},l_{\mathbf{s}}}\right)^{n_{\mathbf{s},k_{\mathbf{s}}}^{T}} \prod_{l=k+1}^{K_{\mathbf{s}}}\left(\pi_{\mathbf{s},k_{\mathbf{s}}} + \sum_{l'=l}^{\infty}\pi_{\mathbf{s},l'_{\mathbf{s}}}\right)^{n_{\mathbf{s},l_{\mathbf{s}}}^{T}-m+1-\mathbb{I}(l=K_{\mathbf{s}})}}, \tag{A1}$$

$$p = \frac{\pi_{\mathbf{s},k_{\mathbf{s}}}^{n_{\mathbf{s},k_{\mathbf{s}}}^{T}-m+1-\mathbb{I}(k_{\mathbf{s}}=K_{\mathbf{s}})} \prod_{t:z_{t,\mathbf{s}}=k_{\mathbf{s}}} f(\mathbf{y}_{t,\mathbf{s}}^{\text{obs}}|\boldsymbol{\psi}_{\mathbf{s}}, \boldsymbol{\eta}_{k_{\mathbf{s}}})}{\prod_{l=k}^{K_{\mathbf{s}}}\left(\pi_{\mathbf{s},k_{\mathbf{s}}} + \sum_{l'=l}^{\infty}\pi_{\mathbf{s},l'_{\mathbf{s}}}\right)^{n_{\mathbf{s},l_{\mathbf{s}}}^{T}-m+1-\mathbb{I}(l=K_{\mathbf{s}})}},$$

$$p^{+} = \frac{\pi_{\mathbf{s},k_{\mathbf{s}}}^{n_{\mathbf{s},k_{\mathbf{s}}}^{T}} \prod_{t:z_{t,\mathbf{s}}=k_{\mathbf{s}}} f(\mathbf{y}_{t,\mathbf{s}}^{\text{obs}}|\boldsymbol{\psi}_{\mathbf{s}}, \boldsymbol{\eta}_{k_{\mathbf{s}}^{+}})}{\left(\sum_{l=k}^{\infty}\pi_{\mathbf{s},l_{\mathbf{s}}}\right)^{n_{\mathbf{s},k_{\mathbf{s}}}^{T}} \prod_{l=k+1}^{K_{\mathbf{s}}}\left(\pi_{\mathbf{s},k_{\mathbf{s}}} + \sum_{l'=l}^{\infty}\pi_{\mathbf{s},l'_{\mathbf{s}}}\right)^{n_{\mathbf{s},l_{\mathbf{s}}}^{T}-m+1-\mathbb{I}(l=K_{\mathbf{s}})}}, \tag{A2}$$

if $k = K_{\mathbf{s}}$, that is, $k_{\mathbf{s}}$ is the last regime, we assume $p^{+} = 0$ while, if $k = 1$, that is, $k_{\mathbf{s}}$ is the first regime, then $p^{-} = 0$. The algorithm is applied beginning to merge the first regime. If it is merged it then tries to merge the newly created regime, otherwise it proceeds with the next.

The splitting step is based on the same probabilities used for the merging step. Allowing for a slight abuse of notation, assume we take a generic regime ad split it into two parts. The first one has index $k_{\mathbf{s}}$, selected with probability proportional to $\pi_{\mathbf{s},k_{\mathbf{s}}}$ among the indices not yet observed at station $\mathbf{s}$, and has $n_{\mathbf{s},k_{\mathbf{s}}}^{T}$ elements. If all terms in Equations (A1)–(A2) are obtained based on the segmentation of the time series including the $k_{\mathbf{s}}$th regime, Equations (A1)–(A2) can be used to compute the splitting probabilities. Under this setting, p^{+} is the probability that no changes occur, p is the probability of the splitting, while p^{-} merges the newly created regime with its predecessor. The algorithm is applied starting with the first observation ($t = 1$) and each t is assumed to be the last element of regime $k_{\mathbf{s}}$. If $k_{\mathbf{s}}$ is merged with $k_{\mathbf{s}}^{-}$, then the we start again with t, otherwise with $t + 1$. If $n_{\mathbf{s},k_{\mathbf{s}}}^{T} < m$, then we assume $p = 0$.

Sampling $\boldsymbol{\pi}_{\mathbf{s}}$

Following the work of Teh et al. (2006), since G_0 is discrete and $G_{\mathbf{s}} \sim \text{DP}(\alpha, G_0)$, we have $\boldsymbol{\pi}_{\mathbf{s}} \sim \text{DP}(\alpha, \boldsymbol{\beta})$. Let suppose that

$$\tilde{\pi}_{\mathbf{s},k_{\mathbf{s}}}|\alpha, \boldsymbol{\xi} \sim B\left(\alpha\xi_{k_{\mathbf{s}}}, \alpha\sum_{l=k+1}^{\infty}\xi_{l_{\mathbf{s}}}\right), \quad k \in \mathcal{K}_{\mathbf{s}}$$

then, following Teh et al. (2006), the following relations, called stick-breaking construction, exist between $\boldsymbol{\pi}_{\mathbf{s}}$ and $\tilde{\boldsymbol{\pi}}_{\mathbf{s}}$:

$$\pi_{\mathbf{s},1_{\mathbf{s}}} = \tilde{\pi}_{\mathbf{s},1_{\mathbf{s}}},$$
$$\pi_{\mathbf{s},k_{\mathbf{s}}} = \tilde{\pi}_{\mathbf{s},k_{\mathbf{s}}}\prod_{l=1}^{k-1}\left(1 - \tilde{\pi}_{\mathbf{s},l_{\mathbf{s}}}\right).$$

Since there is a one-to-one relation between $\tilde{\boldsymbol{\pi}}_{\mathbf{s}}$ and $\boldsymbol{\pi}_{\mathbf{s}}$ we can sample the former and then transform it into a sample of the latter.

Working with $\tilde{\boldsymbol{\pi}}_{\mathbf{s}}$ is easier since $\tilde{\pi}_{\mathbf{s},k_{\mathbf{s}}}$ is the probability to select/stay in the $k_{\mathbf{s}}$th regime. The full conditional of $\tilde{\boldsymbol{\pi}}_{\mathbf{s}}$ is then easily obtained as:

$$\begin{cases} B\left(\alpha\xi_{k_{\mathbf{s}}} + n^T_{\mathbf{s},k_{\mathbf{s}}} - m + 1, \alpha\sum_{l=k+1}^{\infty}\xi_{l_{\mathbf{s}}} + 1 - \mathbb{I}(k = K_{\mathbf{s}})\right) & \text{if } k_{\mathbf{s}} \in \mathcal{K}_{\mathbf{s}} \\ B\left(\alpha\xi_{k_{\mathbf{s}}}, \alpha\sum_{l=k+1}^{\infty}\xi_{l_{\mathbf{s}}}\right) & \text{otherwise} \end{cases}.$$

Sampling ν_ℓ

To define an efficient sampling scheme we introduce a stick-breaking construction for the distribution $G_{\mathbf{s}}$, assuming

$$\tilde{\lambda}_{\mathbf{s},i}|\alpha \sim B(1,\alpha), \quad i \in \mathbb{N}$$

and

$$\tilde{\boldsymbol{\eta}}_{\mathbf{s},i}|G_0 \sim G_0. \tag{A3}$$

Then also the distribution $G^{\lambda}_{\mathbf{s}} = \sum_{i\in\mathbb{N}} \lambda_{\mathbf{s},i}\delta_{\tilde{\boldsymbol{\eta}}_{\mathbf{s},i}}$, with

$$\lambda_{\mathbf{s},1} = \tilde{\lambda}_{\mathbf{s},1},$$
$$\lambda_{\mathbf{s},i} = \tilde{\lambda}_{\mathbf{s},i}\prod_{j=1}^{i-1}\left(1-\tilde{\pi}_{\mathbf{s},j}\right)$$

is DP(α, G_0), as well as $G_{\mathbf{s}}$, and $G^{\lambda}_{\mathbf{s}} \overset{d}{=} G_{\mathbf{s}}$, where $\overset{d}{=}$ indicates equality in distribution. Indeed, since G_0 is discrete some of the atoms $\tilde{\boldsymbol{\eta}}_{\mathbf{s},i}$ are the same The unique elements in $\tilde{\boldsymbol{\eta}}_{\mathbf{s},i}$ are the same as those in $\boldsymbol{\eta}_{\mathbf{s},k}$. Then, the following holds:

$$\pi_{\mathbf{s},k} = \sum_{i\in\mathbb{N}} \lambda_{\mathbf{s},i}\mathbb{I}(\tilde{\boldsymbol{\eta}}_{\mathbf{s},i} = \boldsymbol{\eta}_{\mathbf{s},k}). \tag{A4}$$

It is then possible to uniquely derive $G_{\mathbf{s}}$ from $G^{\lambda}_{\mathbf{s}}$, meaning that working with the latter is equivalent of working with the former.

To understand why this new parametrization is useful, note that every time we select or decide to stay in a regime k, this is done with probability proportional to $\pi_{\mathbf{s},k}$ while, given (A4), with the new parametrization this is done with probability proportional to one of the $\lambda_{\mathbf{s},i}$ associated to $\pi_{\mathbf{s},k}$, that is, we select one of the $\lambda_{\mathbf{s},i}$s at each time-point. Then, $\lambda_{\mathbf{s},i}$s can be seen as the probabilities of a DP-mixture model with number of observations equal to $n^T_{\mathbf{s},k} - m + 1$, that is, the number of "free" observations in regime k and station $\mathbf{s}$, with scaling parameter α (the one of $\lambda_{\mathbf{s},i}$).

We indicate the number of occupied regimes in this new mixture as $d_{\mathbf{s},k}$ which can be computed as

$$d_{\mathbf{s},k} = \sum_{i\in\mathbb{N}} \mathbb{I}(\tilde{\boldsymbol{\eta}}_{\mathbf{s},i} = \boldsymbol{\eta}_{\mathbf{s},k}).$$

Using the Chinese restaurant process representation of a DP-mixture (Pitman, 2002), we can easily sample $d_{\mathbf{s},k}$ with the following steps:

$$x_{\mathbf{s},k,i} \sim B\left(\frac{\alpha}{i-1+\alpha}\right), \quad i = 1, \ldots, n^T_{\mathbf{s},k} - m + 1$$

and then

$$d_{\mathbf{s},k} = \sum_{i=1}^{n^T_{\mathbf{s},k}-m+1} x_{\mathbf{s},k,i}.$$

Note that $d_{\mathbf{s},k}$ has to be computed only for the nonempty regimes of the CP-model. Given $d_{\mathbf{s},k}$, the parameter $\boldsymbol{\nu}_\ell$ can be sampled quite easily. From (A3) we see that parameter $\boldsymbol{\eta}_{\mathbf{s},k}$ is sampled $d_{\mathbf{s},k}$ times in regime k and station $\mathbf{s}$, with probability given by $\xi_k^{d_{\mathbf{s},k}} = \prod_{\ell=1}^{9} \nu_{\ell,w_{\ell,k}}^{d_{\mathbf{s},k}}$, see Equation (4). Then, let define

$$d_p^\ell = \sum_{\mathbf{s}\in\mathcal{S}} \sum_{k\in\mathcal{K}_\mathbf{s}} d_{\mathbf{s},k} \mathbb{I}(w_{\ell,k} = p),$$

as the total number that the pth component of the ℓth parameter is selected and let n^ℓ be the number of elements of $\boldsymbol{\nu}_\ell$ that have $d_p^\ell > 0$. Without loss of generality, we also assume that the first n^ℓ elements of $\boldsymbol{\nu}_\ell$ are those with $d_p^\ell > 0$. Since $\boldsymbol{\nu}_\ell$ is the vector of weights of a draw from a DP, we have that

$$\boldsymbol{\nu}_\ell^{n^\ell} \sim \text{Dir}(d_1^\ell, d_2^\ell, \ldots, d_{n^\ell}^\ell, \gamma),$$

where $\boldsymbol{\nu}_\ell^{n^\ell}$ contains the first n^ℓ elements of $\boldsymbol{\nu}_\ell$. As in the standard DP-mixture model, the posterior is a Dirichlet distribution with parameters of the first n elements given by the number of "observations," while the last one is the scalar parameter of the DP and it is used to generate a new, not yet observed, component.

Sampling $w_{\ell,k}$

Let suppose that station $\mathbf{s}$ is in the kth regime, with parameter $\boldsymbol{\eta}_k$. If we indicate with $\boldsymbol{\eta}_k^{\ell,i}$ the vector $\boldsymbol{\eta}_k$ with $\eta_{\ell,i}^*$ corresponding to the ℓth element, then, using the DP-mixture representation based on λ, it is easy to see that the probability that the ℓth element is equal to $\eta_{\ell,i}^*$ is proportional to

$$\nu_{\ell,i}^{d_{\mathbf{s},k}} \prod_{t:z_{t,\mathbf{s}}=k} f(\mathbf{y}_{t,\mathbf{s}}^{\text{obs}} | \boldsymbol{\psi}_\mathbf{s}, \boldsymbol{\eta}_k^{\ell,i}) \tag{A5}$$

if this is coherent with the CP time dynamic, meaning that, once the ℓth element is changed, two regimes cannot have the same vector of parameters $\boldsymbol{\eta}_k$, otherwise the probability is zero.

Expression (A5) has to be evaluated for all infinite possible values of $w_{\ell,k}$. To solve the problem we use the beam sampling scheme (Van Gael et al., 2008), which can be easily implemented in this context. The idea is to introduce the additional variable

$$u_{\mathbf{s},k,\ell} \sim U(0, \nu_{\ell,p^*})$$

and then, conditioning on $u_{\mathbf{s},k,\ell}$, we have to evaluate (A5) multiplied by the density of $u_{\mathbf{s},k,\ell}$, that is ν_{ℓ,p^*}^{-1}, only if $\nu_{\ell,i} > u_{\mathbf{s},k,\ell}$, drastically reducing the number of times (A5) has to be computed.

Sampling α

Here again we can use the representation of the CP model in terms of DP-mixtures. Conditioning on all $\boldsymbol{\pi}_{\mathbf{s},k}$s and $d_{\mathbf{s},k}$s, the full conditional of α is proportional to

$$f(\alpha) \prod_{\mathbf{s}\in\mathcal{S}} \prod_{k\in\mathcal{K}^o} f(d_{\mathbf{s},k} | \alpha, n_{\mathbf{s},k}^T), \tag{A6}$$

where $f(\alpha)$ is the prior distribution. The full conditional in (A6) has the same structure of the full conditional for the DP parameter obtained by Fox et al. (2011) (section E.1 of the supplementary online material). Using their approach, we assume a $G(a_\alpha, b_\alpha)$ prior and define the following latent variables

$$r_{\mathbf{s},k}^{1,\alpha} \sim B(\alpha + 1, n_{\mathbf{s},k}^t - m + 1),$$

$$r_{\mathbf{s},k}^{2,\alpha} \sim \text{Bern}\left(\frac{n_{\mathbf{s},k}^t - m + 1}{n_{\mathbf{s},k}^t - m + 1 + \alpha} \right),$$

for $k \in \mathcal{K}_{\mathbf{s}}^{o}$ and $\mathbf{s} \in \mathcal{S}$. Then, a posterior sample of α is obtained sampling from

$$G\left(a_\alpha + \sum_{\mathbf{s}\in\mathcal{S}}\sum_{k\in\mathcal{K}_{\mathbf{s}}^{o}}(d_{\mathbf{s},k} - r_{\mathbf{s},k}^{2,\alpha}), b_\alpha - \sum_{\mathbf{s}\in\mathcal{S}}\sum_{k\in\mathcal{K}_{\mathbf{s}}^{o}}\log r_{\mathbf{s},k}^{1,\alpha}\right).$$

Sampling γ

As a matter of fact, parameter γ is responsible for the number of unique values of $\boldsymbol{\nu}_\ell$ used in the model. If a $G(a_\gamma, b_\gamma)$ prior is assumed for γ, a reasoning similar to the one used to sample α can be used to obtain its full conditional. Letting $d = \sum_{\mathbf{s}\in\mathcal{S}}\sum_{k\in\mathcal{K}_{\mathbf{s}}} d_{\mathbf{s},k}$, the full conditional of γ is proportional to

$$f(\gamma)\prod_{\ell=1}^{9} f(n^\ell|\alpha, d_\ell),$$

which has the same structure of (A6). Then, if we define the following latent variables:

$$r_\ell^{1,\gamma} \sim B(\gamma+1, d),$$
$$r_\ell^{2,\gamma} \sim \text{Bern}\left(\frac{d}{d+\gamma}\right),$$

for $\ell = 1, \ldots 9$, a posterior sample for γ is obtained from

$$G\left(a_\gamma + \sum_{\ell=1}^{9}(n^\ell - r_\ell^{2,\gamma}), b_\gamma - \sum_{\ell=1}^{9}\log r_\ell^{1,\gamma}\right).$$

Sampling $\mathbf{y}^{\text{miss}}$

The conditional independence of $\mathbf{y}_{t,\mathbf{s}}$s makes sampling the missing data quite easy. If the missing value is not due to zero precipitations sampling is straightforward, since it can be simulated from the Normal conditional distribution. Conversely, in the case of zero precipitations, we have to ensure that the simulated value is below zero and this is attained by a truncated normal conditional distribution.

Sampling $\eta_{\ell,i}^*$

Samples of the $\eta_{\ell,i}^*$ parameters are obtained by Metropolis steps. Even though the means could be updated using Gibbs sampling, Metropolis proved to speed up convergence.

Let the set $B_{\mathbf{s},i}^{\ell}$ contain the temporal indices corresponding to station $\mathbf{s}$ having parameter $\eta_{\ell,i}^*$ and let $\boldsymbol{\theta}_{t,\mathbf{s}}^{\text{prop}}$ be the vector-valued parameter $\boldsymbol{\theta}_{t,\mathbf{s}}$ with the ℓth value replaced by the proposed $\eta_{\ell,i}^*$. Then, the model contribution to the Metropolis ratio is equal to

$$\frac{\prod_{\mathbf{s}\in\mathcal{S}}\prod_{t\in B_{\mathbf{s},i}^{\ell}} f(\mathbf{y}_{t,\mathbf{s}}|\psi_{t,\mathbf{s}}, \boldsymbol{\theta}_{t,\mathbf{s}}^{\text{prop}})}{\prod_{\mathbf{s}\in\mathcal{S}}\prod_{t\in B_{\mathbf{s},i}^{\ell}} f(\mathbf{y}_{t,\mathbf{s}}|\psi_{t,\mathbf{s}}, \boldsymbol{\theta}_{t,\mathbf{s}})}. \tag{A7}$$

To complete the Metropolis ratio, we must multiply (A7) for the prior and proposal density ratios.

Since we are working with a trivariate normal density, not all possible combinations of the correlation parameters produce a valid nonnegative matrix and if a nonvalid value is proposed, the trivariate normal density is equal to zero and it is then never accepted.

Sampling $\boldsymbol{\psi}_{\mathbf{s}}$

Sampling $\boldsymbol{\psi}_{\mathbf{s}}$ is straightforward if we introduce the following new variable

$$\psi_{j,\mathbf{s}} = \psi_{j,\mathbf{s}}^* - \frac{\sum_{j=1}^{12}\psi_{j,\mathbf{s}}^*}{12}. \tag{A8}$$

Notice that the equality $\psi^*_{j,\mathbf{s}} = \psi^*_{j+12,\mathbf{s}}$ holds as $\psi_{j,\mathbf{s}} = \psi_{j+12,\mathbf{s}}$. These new variable definition is needed since the sum-to-zero constraint on $\psi_{j,\mathbf{s}}$ makes the sample challenging, while $\psi^*_{j,\mathbf{s}}$ has no such constraint. The likelihood can then be written as

$$\phi_3\left(\mathbf{y}_{t,\mathbf{s}}|\psi^*_{j,\mathbf{s}} - \frac{\sum_{j=1}^{12}\psi^*_{j,\mathbf{s}}}{12} + \boldsymbol{\mu}_{t,\mathbf{s}}, \boldsymbol{\Sigma}_{t,\mathbf{s}}\right)$$

and $\psi^*_{j,\mathbf{s}}$s can be envisioned as regressive coefficients, that is,

$$\psi^*_{j,\mathbf{s}} - \frac{\sum_{j=1}^{12}\psi^*_{j,\mathbf{s}}}{12} = X_{j \mod 12}(\psi^*_{1,\mathbf{s}}, \dots, \psi^*_{12,\mathbf{s}})',$$

where $X_{j \mod 12}$ is a row vector of dimension 12 with all elements equal to $-1/12$ except the $(j \mod 12)$th, which is $1 - 1/12$. A posterior sample of $(\psi^*_{1,\mathbf{s}}, \dots, \psi^*_{12,\mathbf{s}})'$ can be then obtained using the standard sample of regressive coefficients for a normal density if we assume a normal prior for all $\psi_{j,\mathbf{s}}$s, and consequently for $\psi^*_{j,\mathbf{s}}$s. Having sampled $(\psi^*_{1,\mathbf{s}}, \dots, \psi^*_{12,\mathbf{s}})'$ we can compute $(\psi_{1,\mathbf{s}}, \dots, \psi_{12,\mathbf{s}})'$ using (A8).

APPENDIX B. SIMULATED EXAMPLE

In this section we use the distribution model in (2) to simulate a multivariate dataset with components Y_1, Y_2, Y_3 at 30 monitoring stations, with 360 time-points, providing a latent grouping structure that generates clusters of spatial units and time regimes with the no-return constraint (according to the change-point model). Then, for each station we assume that the minimum regime length is $m = 60$ and simulate it equal to 360 with probability 0.3 and to $m + x$ with probability

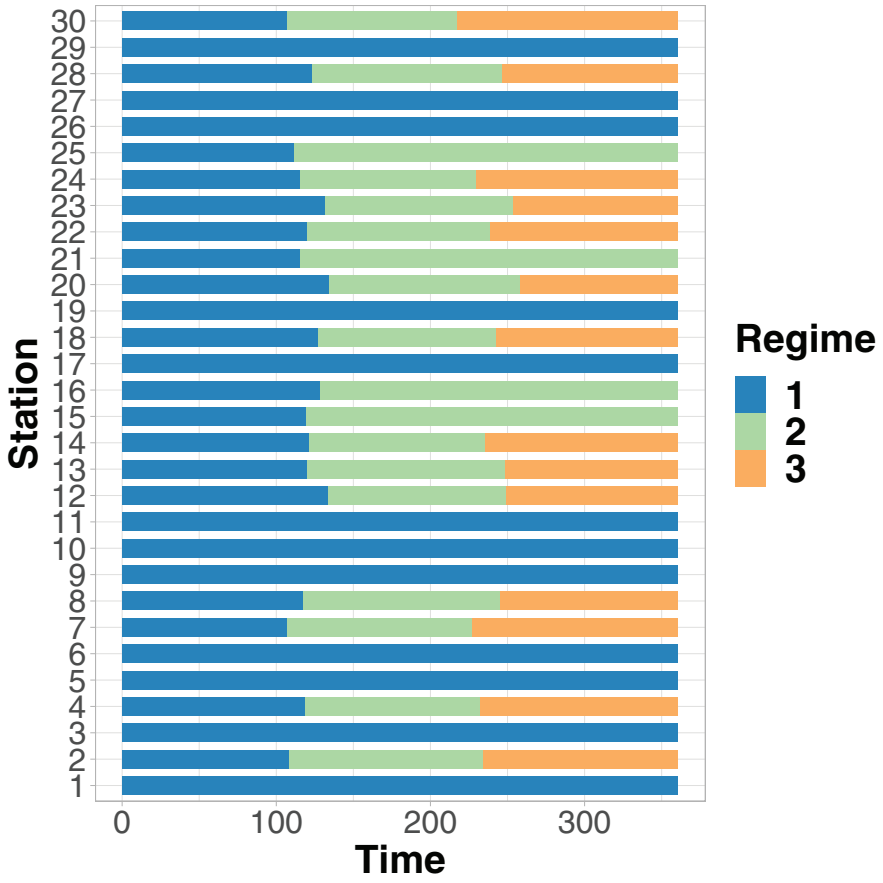


FIGURE B1 Time regimes with simulated lengths at each of 30 monitoring stations

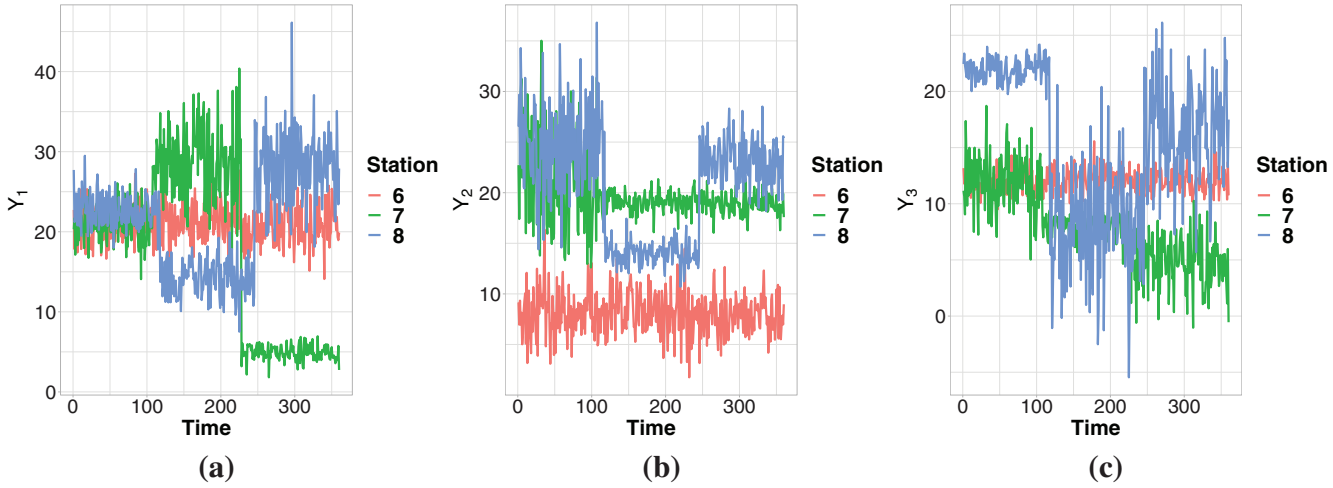


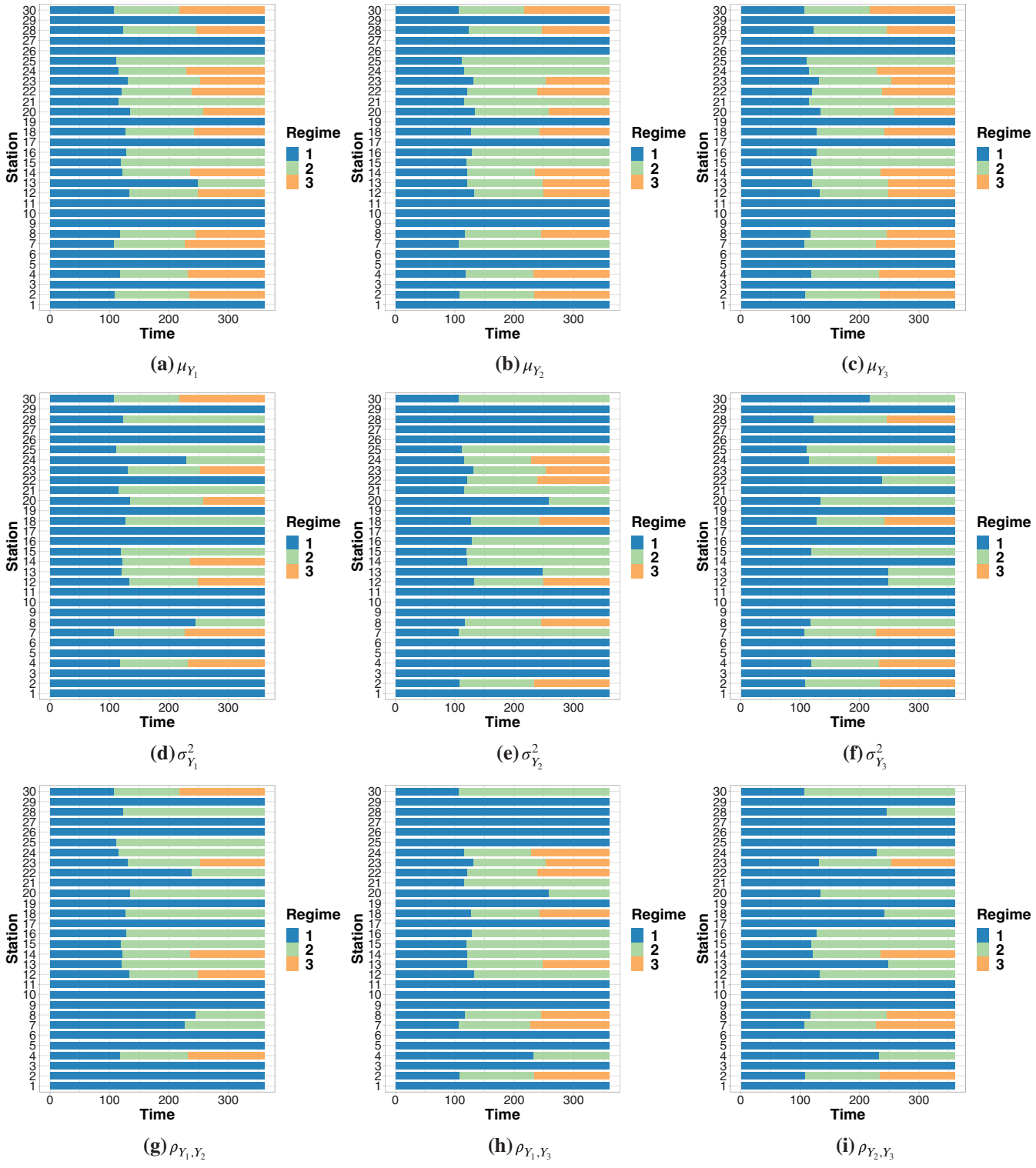
FIGURE B2 Examples of simulated data at three monitoring stations

0.7, where x is a Poisson random variate with mean 60. In Figure B1 we show the time regimes with simulated lengths at each of 30 monitoring stations.

We restrict the possible values of the nine parameters so that each value can be sampled multiple times and generate regimes and groups of stations with exactly the same values of some or all the parameters. In details, the means $(\mu_{Y_1}, \mu_{Y_2}, \mu_{Y_3})$ assume values in the set $D_\mu = \{5, \dots, 30\}$, the variances $(\sigma^2_{Y_1}, \sigma^2_{Y_2}, \sigma^2_{Y_3})$ have values in $D_\sigma = \{1, 5, 20\}$, and the correlations $(\rho_{Y_i,Y_j}, i, j = 1, 2, 3, i \neq j)$ can be equal to $D_\rho = \{0, 0.5, 0.85\}$. At each time point and monitoring station values of the parameters are sampled from these sets with probabilities equal to $\frac{1}{|D_h|}$, where $h = \mu, \sigma, \rho$ and $|D_h|$ is the cardinality of the set D_h. We ensure that the no-return constraint is satisfied and that the correlation matrix is nonnegative definite. The elements of the cyclical components $\boldsymbol{\psi}_{t,\mathbf{s}}$ are sampled from a normal distribution with mean 0 and variance 0.5. Examples of the simulated variables at three stations are shown in Figure B2, while the parameters-specific regimes are in Figure B3. We estimate the model using the same priors, iterations, burnin, and thinning used for the real data application (Section 4).

To evaluate if the DP-CP model is able to detect the CPs, we use $P(\boldsymbol{\theta}_{t,s} \neq \boldsymbol{\theta}_{t+1,s}|\mathbf{y})$ and $P(\boldsymbol{\theta}_{t,s,\ell} \neq \boldsymbol{\theta}_{t+1,s,\ell}|\mathbf{y})$ with $\ell = 1, \dots, 9$, that is, the posterior probabilities that the entire set of parameters, or a specific one, change value between two consecutive time-points. As a matter of fact, we obtain $P(\boldsymbol{\theta}_{t,s} \neq \boldsymbol{\theta}_{t+1,s}|\mathbf{y}) \in [0, 0.09)$ and $P(\boldsymbol{\theta}_{t,s,\ell} \neq \boldsymbol{\theta}_{t+1,s,\ell}|\mathbf{y}) \in [0, 0.16)$ for the time-points and stations where $\boldsymbol{\theta}_{t,s} = \boldsymbol{\theta}_{t+1,s}$ and $\boldsymbol{\theta}_{t,s,\ell} = \boldsymbol{\theta}_{t+1,s,\ell}$, respectively, and $P(\boldsymbol{\theta}_{t,s} \neq \boldsymbol{\theta}_{t+1,s}|\mathbf{y}) \in (0.96, 1]$ and $P(\boldsymbol{\theta}_{t,s,\ell} \neq \boldsymbol{\theta}_{t+1,s,\ell}|\mathbf{y}) \in (0.81, 1]$ otherwise. These results tell us that when there is a change in the data, our algorithm is very much able to detect it. Moreover, the DP-CP model was able to detect the exact time when the CP occurred for 30 among the 32 CPs of Figure B1, while the detected CP was off by only 1 time-point for the remaining two.

The data are simulated with a (small) finite number of possible values for each parameter, then there are several time points sharing the same values across stations and we want to evaluate if our proposal is able to detect such similarities. To this end we compute the similarities $\pi_{s_i,s_j,t,\ell}$ introduced in Section 4.4, which measure the similarities between stations s_i and s_j at time t for parameter ℓ. For each $\ell \in 1, \dots, 9$ we divide the similarities $\pi_{s_i,s_j,t,\ell}$ into two sets: one composed by the time-points and stations where $\boldsymbol{\theta}_{t,s_i,\ell} = \boldsymbol{\theta}_{t,s_j,\ell}$ and another where $\boldsymbol{\theta}_{t,s_i,\ell} \neq \boldsymbol{\theta}_{t,s_j,\ell}$, that is, the two sets are respectively composed of the time-points and stations where the true ℓth parameter is the same or different. For all parameters, 95% of the values of $\pi_{s_i,s_j,t,\ell}$ are in $[0, 0.08)$ if the true parameters are different, while when $\boldsymbol{\theta}_{t,s_i,\ell} = \boldsymbol{\theta}_{t,s_j,\ell}$ we have that 95% of the values of $\pi_{s_i,s_j,t,\ell}$ are in $(0.91, 1]$, $(0.89, 1]$ and $(0.72, 1]$ for the means, variances, and correlations, respectively. These results highlight that we are able to estimate with great accuracy when parameter values are not shared by any two stations, while when they have the same value we are more confident on the results for the means and variances, while equal correlations are a little harder to spot. In terms of parameter estimates we obtain that 97.3% of the 95% HPD intervals at each time-point and station contains the true value used to simulate the data.



(a) μ_{Y_1} (b) μ_{Y_2} (c) μ_{Y_3}

(d) $\sigma^2_{Y_1}$ (e) $\sigma^2_{Y_2}$ (f) $\sigma^2_{Y_3}$

(g) ρ_{Y_1,Y_2} (h) ρ_{Y_1,Y_3} (i) ρ_{Y_2,Y_3}

FIGURE B3 Parameter-specific regimes at each of 30 monitoring stations

APPENDIX C. LEGEND OF THE ITALIAN ECOREGION SYSTEM

TABLE C1 Legend of the Italian ecoregion system 1

Ecoregion code and name	Area (km^2)
1 Temperate Division	189,266
1A Alpine Province	54,502
1A1 Western Alps Section	17,940
1A1a Alpi Marittime Subsection	4023
1A1b Northwestern Alps Subsection	13,917
1A2 Central and Eastern Alps Section	36,561
1A2a Pre-Alps Subsection	15,769
1A2b Dolomiti and Carnia Subsection	8249
1A2c Northeastern Alps Subsection	12,543
1B Po Plain Province	49,851
1B1 Po Plain Section	49,851
1B1a Lagoon Subsection	7461
1B1b Central Plain Subsection	33,108
1B1c Western Po Basin Subsection	9282
1C Apennine Province	84,633
1C1Northern and Northwestern Apennine Section	38,800
1C1a Toscana and Emilia-Romagna Apennine Subsection	17,206
1C1b Tuscan Basin Subsection	21,594
1C2 Central Apennine Section	26,398
1C2a Umbria and Marche Apennine Subsection	10,483
1C2b Lazio and Abruzzo Apennine Subsection	11,453
1C2c Marche and Abruzzo Sub-Apennine Subsection	4462
1C3 Southern Apennine Section	19,435
1C3a Campania Apennine Subsection	10,126
1C3b Lucania Apennine Subsection	9309
1D Italian part of the Illyrian Province	281

TABLE C2 Legend of the Italian ecoregion system 2

Ecoregion code and name	Area (km^2)
2 Mediterranean Division	112,849
2A Italian part of Ligurian-Provencal Province	1053
2B Tyrrhenian Province	85,203
2B1 Northern and Central Tyrrhenian Section	15,231
2B1a Eastern Liguria Subsection	699
2B1b Maremma Subsection	6165
2B1c Roman Area Subsection	4577
2B1d Southern Lazio Subsection	3790
2B2 Southern Tyrrhenian Section	20,054
2B2a Western Campania Subsection	3336
2B2b Cilento Subsection	3132
2B2c Calabria Subsection	13,586
2B3 Sicilia Section	25,832
2B3a Iblei Subsection	3709
2B3b Sicilia Mountains Subsection	7823
2B3c Central Sicilia Subsection	7794
2B3d Western Sicilia Subsection	6506
2B4 Sardegna Section	24,086
2B4a Southwestern Sardegna Subsection	5007
2B4b Northwestern Sardegna Subsection	4957
2B4c Southeastern Sardegna Subsection	11,564
2B4d Northeastern Sardegna Subsection	2557
2C Adriatic Province	26,592
2C1 Central Adriatic Section	2170
2C1a Marche and Abruzzo Coastal Subsection	2170
2C2 Southern Adriatic Section	24,422
2C2a Gargano Subsection	7007
2C2b Murge and Salento Subsection	17,415