UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE DE RIBEIRÃO PRETO
DEPARTAMENTO DE ECONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

FERNANDA VALENTE

Essays in spatial statistics

Ensaios em estatística espacial

Orientador: Prof. Dr. Márcio Poletti Laurini

RIBEIRÃO PRETO
2023

Prof. Dr. Carlos Gilberto Carlotti Junior
Reitor da Universidade de São Paulo

Prof. Dr. Fabio Augusto Reis Gomes
Diretor da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto

Prof. Dr. Milton Barossi Filho
Chefe do Departamento de Economia

Prof. Dr. Luciano Nakabashi
Coordenador do Programa de Pós Graduação em Economia - Área: Economia Aplicada

FERNANDA VALENTE

Essays in spatial statistics

Ensaios em estatística espacial

Tese submetida ao Programa de Pós-Graduação em Economia - Área: Economia aplicada da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto da Universidade de São Paulo, para a obtenção do título de Doutora em Ciências. Versão Corrigida. A original encontra-se disponível na FEA-RP/USP.

Orientador: Prof. Dr. Márcio Poletti Laurini

RIBEIRÃO PRETO
2023

Valente, Fernanda
Essays in spatial statistics / Fernanda Valente – Ribeirão Preto, 2023.
245f.: il.; 30 cm

Tese submetida ao Programa de Pós-Graduação em Economia - Área: Economia aplicada da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto da Universidade de São Paulo, para a obtenção do título de Doutora em Ciências. Versão Corrigida. A original encontra-se disponível na FEA-RP/USP. – Universidade de São Paulo
Orientador: Laurini, Márcio Poletti

1. Spatial Point Process. 2. Log Gaussian Cox Process. 3. INLA. 4. Environmentrics. 5. Climate Events. 6. COVID-19.

Aos meus amados pais.

ACKNOWLEDGMENTS

Agradeço, em primeiro lugar, aos meus amados pais, Eledir e Delson, que sempre me incentivaram a percorrer o caminho da academia e não mediram esforços para garantir uma educação de qualidade à mim e minha irmã. Eu jamais teria conseguido sem vocês.

À minha amada irmã, Amanda, pelo apoio, admiração e por ter nos dado, ao longo desta jornada, nosso maior presente, a nossa Isabelinha.

Ao meu amor e parceiro de vida, Kael, por me apoiar em todos os meus sonhos, vibrar pelas minhas conquistas e me dar suporte nos momentos mais difíceis desta jornada. Obrigada por me lembrar da minha força e me encorajar a seguir em frente, mesmo que isso significasse a distância de um oceano entre nós.

Aos amigos que ganhei no Programa, em especial à Patrícia Felini, que deixaram as necessárias horas de estudo mais divertidas e leves.

À minha querida amiga, Franciele Alba, por compartilhar comigo a incrível experiência de viver e estudar fora do Brasil. Guardo com carinho na memória todos os momentos que vivemos na 51, Watson Crescent.

Ao meu orientador, Márcio Laurini, pela excepcional orientação, disponibilidade e paciência em compartilhar seu conhecimento ao longo destes anos de trabalho.

Ao professor Finn Lindgren, por me receber na University of Edinburgh e gentilmente contribuir com nosso trabalho.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001 e do Instituto Escolhas através do Programa Cátedra Escolhas de Economia e Meio Ambiente.

"Torture the data and it will confess to anything"
(Ronald Coase)

ABSTRACT

VALENTE, F. (2023). **Essays in Spatial Statistics**. Doctoral Thesis. Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2023.

This thesis is composed by eight self-contained papers in spatial statistics. Five papers are applications of statistical decomposition for spatio-temporal analysis of changes in climate-related events. The sixth paper is an application of statistical decomposition for climatic time series to analyze the existence of changes in rainfall and temperature patterns. The two final papers deal with statistical decomposition for spatio-temporal analysis of coronavirus disease in the national and global scales.

Keywords: Spatial Point Process, Log Gaussian Cox Process, INLA, Environmentrics, Climate Events, COVID-19

RESUMO

VALENTE, F. (2023). **Ensaios em Estatística Espacial**. Tese (Doutorado). Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2023.

Esta tese é composta por oito artigos independentes em estatística espacial. Cinco artigos são aplicações de decomposição estatística para análises espaço-temporais de mudanças em eventos relacionados ao clima. O sexto artigo é uma aplicação de decomposição estatística para séries temporais climáticas para analisar a existência de mudanças nos padrões de chuva e temperatura. Os dois trabalhos finais tratam da decomposição estatística para análise espaço-temporal da doença do coronavírus em escalas nacional e global.

Palavras-chave: Spatial Point Process, Log Gaussian Cox Process, INLA, Environmentrics, Climate Events, COVID-19

LIST OF FIGURES

LIST OF TABLES

CONTENTS

GENERAL INTRODUCTION

This thesis is composed by eight self-contained papers in spatial statistics. Five papers are applications of statistical decomposition for spatio-temporal analysis of changes in climate-related events like fires and tornadoes. One paper is an application of statistical decomposition for climatic time series to analyze the existence of changes in rainfall and temperature patterns. The two final papers deal with statistical decompositions for spatio-temporal analysis of coronavirus disease.

In the first paper we propose to analyze the possible changes in the tornado occurrences in the Unites States. The results have provided evidences that trend in annual tornado occurrences in the United States have remained relatively constant, supporting previously reported findings. This paper was published in Econometrics, v. 8, n. 2, p. 25, 2020.

The second paper investigates the efficiency of a regulatory change in sugarcane production (Law Nº11.241/2002) that gradually banish the pre-harvest burning in the sugarcane production sector in the São Paulo state in Brazil, analyzing the occurrences of fires detected by satellite. The results have provided evidence that the trend component reflects the consistent reduction in fire occurrences, suggesting the efficiency of the mitigation actions implemented in the São Paulo state. This paper was published in Cleaner Engineering and Technology, v. 4, p. 100255, 2021.

In the third paper, we present a statistical decomposition for spatio-temporal analysis of changes in fire occurrence in Australia and its association with climate factors. We found evidence of variability in the trend results for fire occurrence, and also some evidence that this variation is related to climate patterns. This paper was published in Stochastic Environmental Research and Risk Assessment, v. 35, n. 9, p. 1759-1770, 2021.

The fourth paper studies the possible changes in the patterns of fire occurrence in the Legal Amazon, within the spatio-temporal point process framework. The results show that long-term movements of fire occurrence exhibits a marked decrease between the beginning of the sample and 2012, followed by an increase that extents to the end of the sample, which can be related to governance measures and market mechanisms. Additionally, our model was able to capture the spatial variability in the Brazilian Amazon, higher in regions where the climate has dry seasons.

Similarly, the fifth paper also analyses the changes in the patterns of fire occurrence based on remote sensing data, but now in the Brazilian Pantanal. The results obtained indicate that there are relevant variations in the trends of fires occurrence in the Brazilian Pantanal, the intensity of occurrence of fires is statistically higher in natural vegetation, and that a relevant part of the record of fires observed in the first three quarters of 2020 cannot be explained by climatic factors alone, possibly being caused by intentional human actions.

It is worth noting that, in general, the first five chapters of this thesis are grounded in a structural modeling framework that incorporates the potential for extracting trend, cycle, and seasonality components through a structural decomposition framework à la Harvey, while also integrating spatial elements. Specifically, given that the data under consideration comprises point pattern observations with associated geographic coordinates within a continuous space, we adopted the Log Gaussian Cox Process (LGCP), where the decomposition structure is incorporated within the intensity function of the LGCP, that is, the intensity function is defined

as a function of a spatial component plus latent factors.

The LGCP, as introduced by Møller, Syversveen and Waagepetersen (1998), represents a class of valuable models for addressing spatio-temporal point processes. Cox processes are non-homogeneous Poisson processes, where the intensity function is a stochastic process. LGCP, in particular, is a specific type of Cox process in which the log-intensity function follows a Gaussian random field. The advantage is this case is the fact that, according to Simpson et al. (2016), the LGCP formulation naturally integrates into the Bayesian hierarchical modeling framework which allows to perform analysis using integrated nested Laplace approximation method (RUE; MARTINO; CHOPIN, 2009).

The idea to use INLA to fit LGCP models started with Illian et al. (2010), who developed a Poisson approximation to the true LGCP likelihood based on regular grids and the counting number of points within each grid cell. However, Simpson et al. (2016) demonstrated that this approach could be highly inefficient because the quality of the likelihood approximation depends on the grid size, requiring the construction of an excessively fine grid, which is computationally intensive.

As an alternative, Simpson et al. (2016) proposed the use of a stochastic partial differential equations (SPDE) approach to approximate a Gaussian field as a Gaussian Markov random field, defined by sparse matrices due to its Markov properties. The key insight provided by Simpson et al. (2016) is that there's no need to construct an overly fine grid to approximate the LGCP likelihood effectively, making way for computationally efficient methods.

Regarding the structural decomposition components, the proposed structure allows us to identify long-term and transient effects of climate-related occurrences through non-stationary and stationary latent components, respectively. This, in turn, represents a method for analysing climate variability patterns through a spatio-temporal structure.

In the sixth paper we propose a novel statistical decomposition of climatic time series into long-term trend, seasonal and cycle components to analyze the existence of changes in rainfall and temperature patterns in the metropolitan area of São Paulo. Differently from the previous chapters, in this case, we do not consider the spatial dimension. However, in this scenario, we propose the same decomposition structure to extreme values, probability of rain and the duration of dry days. The results indicate a significant increase in the trend component of daily temperature and in the rain patterns. In addition, the analysis for annual temperature records and annual records of daily rainfall support the hypothesis of permanent changes in observed climatic patterns. On the other hand, there is no significant evidence of changes in the pattern of duration of dry days nor changes in the trend component for the annual maximum of dry days, indicating the stability of this component. This paper was published in Urban Climate, v. 41, p. 101077, 2022.

In the scenario of the final two chapters, the structural decomposition framework is applied within the time series and polygon-based spatial data context. The models applied are based on generalized decompositions of trend, seasonality and cycle components incorporating a time-varying spatial component, which is based on a conditional autoregressive structure, that is, the prior information for the number of occurrences in a certain region of interest is based on the information of occurrences in some definition od neighborhood.

The seventh paper proposes a method to estimate the trend in the cases of COVID-19, con-

trolling for the presence of measurement error. We applied time series decompositions analysis for the total number of deaths in Brazil and for the states of São Paulo and Amazonas, and a spatio-temporal analysis for all occurrences of deaths at the state level in Brazil, using two alternative specifications with global and regional components. This paper was published in Spatial and Spatio-temporal Epidemiology, v. 39, p. 100455, 2021.

The eighth and final paper also deals with the estimation of the trend of deaths by COVID-19, but now on a global scale. The proposed model was able to capture the patterns in the occurrence of deaths related to COVID-19, overcoming the problems observed in COVID-19 data. We found compelling evidence that spatio-temporal models are more accurate than univariate models to estimate the patterns of the occurrence of deaths. The findings suggested that the spatial dynamics have an important role in the COVID-19 epidemic process since the results provided evidence that spatio-temporal models are more accurate to estimate the general patterns of the occurrence of deaths related to COVID-19. This paper was published in BMJ open, v. 11, n. 8, p. e047002, 2021.

1 TORNADO OCCURRENCES IN THE UNITED STATES: A SPATIO-TEMPORAL POINT PROCESS APPROACH

Fernanda Valente Márcio Laurini

ABSTRACT

In this paper we analyze the tornado occurrences in the Unites States. To perform inference procedures for the spatio-temporal point process we adopt a dynamic representation of Log-Gaussian Cox Process. This representation is based on the decomposition of intensity function in components of trend, cycles, and spatial effects. In this model, spatial effects are also represented by a dynamic functional structure, which allows to analyze the possible changes in the spatio-temporal distribution of the occurrence of tornadoes due to possible changes in climate patterns. The model was estimated using Bayesian inference through the Integrated Nested Laplace Approximations. We use data from Storm Prediction Center's Severe Weather Database between 1954 and 2018, and the results have provided evidences, from new perspectives, that trend in annual tornado occurrences in the United States have remained relatively constant, supporting previously reported findings.

Keywords: Tornadoes in the United States; Spatio-temporal models; Point Process; Log-Gaussian Cox Process; INLA

VALENTE, Fernanda; LAURINI, Márcio. Tornado occurrences in the United States: A spatio-temporal point process approach. **Econometrics**, v. 8, n. 2, p. 25, 2020.

1.1 Introduction

According to the Intergovernmental Panel on Climate Change report (IPCC, 2014), the global average temperature has continued increasing in the last few decades. The climate changes, associated to natural or anthropological activities, can lead to changes in the likelihood of the occurrence and/or strength of severe weather and climate events (ALLEN et al., 2014). Severe events such as heat waves, droughts, tornadoes, and hurricanes cause properties damages and kill people. According to the Emergency Events Database (EM-DAT) of Centre for Research on the Epidemiology of Disasters (CRED), in 2018, there were 315 natural disaster events recorded, which caused 11.804 deaths and more than $130 billion dollars in economic losses across the world.

Tornadoes are extreme weather events defined by "a violently rotating column of air touching the ground, usually attached to the base of a thunderstorm"[1]. The United States experiences more tornadoes per year than any other country, with an average of over 1000 tornadoes each year, according to the National Oceanic and Atmospheric Administration (NOAA). In 2017, tornadoes caused almost $650 million in damages and 35 fatalities in the United States (NOAA).

[1] See https://www.weather.gov/phi/TornadoDefinition.

Given the impact of these events on society, understanding how tornado activity will respond to climate change is important in order to be prepared for possible impacts.

Previous studies have searched for relationships between tornado activity and global warming (LEE, 2012; DIFFENBAUGH; SCHERER; TRAPP, 2013; MOORE; DEBOER, 2019). Recent works analyze the variability of tornadoes occurrence in the United States. As a result, some studies have found that, when removing many non-metereological factors, there is no significant changes in the annual frequency of United States tornadoes (KUNKEL et al., 2013; TIPPETT et al., 2015). Some other studies have generally found that tornadoes are becoming more clustered over time. Specifically, Tippett, Lepore and Cohen (2016) found that the frequency of U.S. outbreaks with many tornadoes is increasing since the 1970s and that it is increasing faster for more extreme outbreaks. Moore (2017) found evidences that the number of (E)F1 and stronger tornado days has decreased whereas the number of high-frequency (30+) tornado days has increased, which tended to be focused further east of the Great Plains. Another empirical studies also propose to analyze if the spatial distribution of tornadoes has shifted over time (GENSINI; BROOKS, 2018; MOORE; MCGUIRE, 2019). Gensini and Brooks (2018) found evidences that a significant spatially-varying temporal trends in tornado frequency have occurred since 1979, with negative tendencies in portion of Great Plains and positive trends in portions of the Midwest and Southeast United States. Moore and McGuire (2019) also found that the spatial dispersion of tornadoes in the United States have changed between 1954 and 2017, specially in spring, summer and fall. In addition, their results suggest that the increased occurrence of tornado outbreaks is contributing to the decrease in dispersion.

Tornadoes can be represented by point processes, which are used to represent occurrences of events in space associated to their spatial coordinates and temporal instant, which corresponds to spatio-temporal point processes. The Log Gaussian Cox process (LGCP), proposed by Møller, Syversveen and Waagepetersen (1998), defines a class of useful models to deal with spatio-temporal point process. Cox processes are inhomogeneous Poisson processes, which the intensity function is a stochastic process and LGCP is a special case of Cox processes, where the log-intensity function is a Gaussian random field.

Recently Illian et al. (2010) proposed to use integrated nested Laplace approximations (INLA) to fit LGCP models. They construct a Poisson approximation to the true LGCP likelihood based on regular lattices, and counting the number of points in each cell. However, Simpson et al. (2016) show that this approach could be very inefficient since the quality of the likelihood approximation depends on the size of the grid, i.e., it is necessary to construct a much fine grid, which is computationally intense. Therefore, Simpson et al. (2016) propose to use stochastic partial differential equations (SPDE) approach to approximate a Gaussian Field (GF) to a Gaussian Markov Random Fields (GMRF), which is defined by sparse matrices, due to Markov properties. The main result provided by Simpson et al. (2016) is that is not necessary to construct too fine grids in order to approximate LGCP likelihood, allowing computationally effective methods. Moreover, according to Simpson et al. (2016), the LGCP formulation fits naturally within the Bayesian hierarchical modelling framework in combination with Integrated Nested Laplace Approximations (INLA) approach, proposed by Rue, Martino and Chopin (2009), avoiding convergence problems related to Markov Chain Monte Carlo (MCMC) methods.

The paper herein discuss how to perform inference procedures for spatio-temporal processes

using a dynamic representation of a LGCP. This representation is based on the modeling of the intensity function from decomposition of components in trend, cycles, covariates and spatial effects. We also assume that spatial effects are based on a dynamic formulation, which allows it to change over time, based on an autoregressive functional structure. Our model follows the approach introduced by Laurini (2019) which proposes a method for decomposition of trend, cycle and seasonal components in spatio-temporal models, where the spatial component is based on a continuous projections of spatial covariance functions. Indeed, our proposed model can be seen as an extension of Laurini (2019) approach to spatial point process, where the dynamics in point process are captured by persistent term and mean-reverting components, plus the spatial term, which is time-varying by the autoregressive group structure. The persistent term is modeled as a first order random walk for a latent component whereas the cyclic component is based on a second-order latent autoregressive structure. This formulation is useful since it allows to identify permanent changes in the intensity of occurrences over time and also to capture cyclical effects in time series. We apply this methodology in order to model the spatio-temporal distribution of the tornado occurrences in the United States, taken into account the different intensities of the occurrences, based on Fujita (prior to 2007) and Enhanced Fujita (2007 and posterior) damaging rate scales, which ranges from (E)F0 (lowest damage indicator) to (E)F5 (highest damage indicator). The results indicate that the trend in annual tornado occurrences in the United States have remained relatively constant, supporting previously reported findings, e.g., Kunkel et al. (2013) and Moore (2017).

This article is organized as follow. Section 2 contains a description of the statistical approach. Section 3 presents the data. Section 4 shows the results with discussion. Section 5 concludes.

1.2 Material and Methods

To model the spatio-temporal distribution of the occurrence of tornadoes, we use the structure of point processes, which is a mathematical way of describing random events distributed in some abstract space. In our application the space used is the product of Cartesian space defined by the longitude and latitude of the occurrence of the tornado by a space defined by the time of occurrence of the event. In this work we assume that time t describes the number of tornadoes that occurred in the year. It is possible to use more complex structures to represent the process, such as the distribution of tornadoes in the terrestrial sphere or other sample intervals, at the cost of greater computational complexity.

An essential structure for modeling point processes is the so-called spatial Poisson process. In this process the number of occurrences in a certain limited region of space is characterized by a Poisson distribution with an intensity parameter λ. Assuming constant λ we have a homogeneous Poisson process, which would generate a random but regular distribution of points in space. As this process is very restrictive, a way of making possible a non-homogeneous distribution in space is through a function of deterministic intensity, where we usually assume that the intensity of the process is a deterministic function of a set of fixed and observed covariates in the same space.

This structure is interesting, but it may have several important empirical limitations. It is necessary to have a structure of covariates observed at each point in the space to assess the likelihood of the process. This is an important limitation since for practical purposes we

need covariates observed continuously for their use, and in many cases these covariates are not observed or are constructed from methods of interpolation of discrete observations in space. The second limitation is that this structure does not incorporate the possible sources of uncertainty associated with this process, such as measurement errors and omission of relevant variables in the model.

Another extremely important limitation is that the Poisson process is conditionally independent, i. e., conditional on the covariates of the process, the achievements are independent. This does not allow incorporating processes of spatial (or temporal) dependence generated by latent processes or processes of dependence generated by covariates omitted from the system.

An alternative is to allow the dependency function in a point process to be stochastic, allowing to incorporate spatio-temporal dependency processes in the occurrences and other forms of randomness in the process including structures of random effects. These effects can be structured in such a way as to capture possible spatial and temporal dependency structures in the punctual process, allowing control for the effects of dependence generated by the omission of relevant covariates, measurement errors and also as an approximation to point processes that are truly dependent outside of the class of Poisson point process.

We use in this work the structure of Log Gaussian Cox Process (LGCP), which are point processes defined by a function of stochastic intensity defined by the class of processes known as Gaussian Markov Random Process (RUE; HELD, 2005a). This class is defined by a structure of Gaussian latent effects, assuming a Markovian dependence structure which allows the use of efficient methods of computational representation and inference. Below we define the LGCP structure used in our work.

1.2.1 Spatio-Temporal Log-Gaussian Cox Process

The model used in this work is a spatio-temporal formulation of point processes with stochastic intensity, using a decomposition of the intensity function into components that vary in time and space. The central idea is to represent the possible temporal evolution through a component of stochastic tendency, which serves to identify permanent changes in the process intensity function, and a cycle component, which represents persistent patterns but with reversion to the mean, in the process intensity function. This interpretation is analogous to the so-called decomposition of unobserved components in time series, as discussed in Harvey (1989). The trend component is modeled as a local level process (first order random walk for a latent component), and thus changes in this level indicate permanent impacts. Note that this formulation is extremely useful for identifying possible changes in the intensity of occurrences over time, which would be associated with permanent changes in the occurrence of tornadoes associated with climate change. The mean reversion component is based on a second-order latent autoregressive structure, a parsimonious way of capturing cyclical effects in time series. This structure is important as it allows estimating the impact of non-permanent shocks on temporal patterns, being especially useful in modeling climate processes as it allows estimating the aggregate effects of phenomena such as droughts, cyclical changes in ocean temperature and other periodic patterns. A detailed discussion of this decomposition applied to climate change patterns can be found at Laurini (2019). The spatial distribution of the intensity function is captured through a random

effect, associated with a spatial covariance function projected continuously in space. This effect represents the spatial variation of the occurrence of tornadoes, and is a simple way to estimate regions with high and low intensities of occurrence of events. However, it is possible to make this spatial effect also time-varying, allowing to estimate possible changes in the spatial distribution of the occurrence of tornadoes, which again could be associated with possible effects of climate change on the spatial distribution of tornadoes. The decomposition proposed in this article is therefore especially useful for analyzing possible changes in the temporal and spatial patterns of the occurrence of events. It captures changes in the average number of occurrences through the trend component, and changes in the spatial distribution pattern through the time varying spatial random effect.

The model can be represented as a spatial LGCP where the intensity function is modelled as follows:

$$\begin{aligned} Y(s,t) &= Poisson(|e(s,t)|exp(\lambda(s,t)), \\ \lambda(s,t) &= \alpha + \mu_t + c_t + \xi(s,t) \\ \mu_t &= \mu_{t-1} + \eta_\mu \\ c_t &= \theta_1 c_{t-1} + \theta_2 c_{t-2} + \eta_c \\ \xi(s,t) &= \Phi\xi(s,t-1) + \omega(s,t) \\ Cov(\omega(s,t)) &= C(h) \end{aligned} \tag{1.1}$$

where $Y(s,t)$ is the number of occurrences in a region s in time t, $e(s)$ is the exposure offset for the region s, α is the intercept, μ_t is the long term trend, c_t is a cycle component represented by an second-order autoregressive process with complex roots, and $\xi(s,t)$ is the spatial random effects represented by the Gaussian process $\omega(s,t)$ continuously projected in space and given by

$$Cov(\omega(s,t)\omega(s',t')) = \begin{cases} 0 & \text{if } t \neq t' \\ \sigma^2 C(h) & \text{if } t = t' \end{cases} \qquad \text{for } s \neq s' \tag{1.2}$$

where $C(h)$ is a covariance function of the Matérn class, which can be written as

$$C(h) = \frac{2^{1-\nu}}{\Gamma(\nu)}(\kappa||h||)^\nu K_\nu(\kappa||h||) \tag{1.3}$$

where $h = ||s - s'||$ is the Euclidean distance between locations s and s', $\kappa > 0$ is a spatial scale parameter, $\nu > 0$ is the smoothness parameter and K_ν is a modified Bessel function.

The Matérn covariance function is a very useful covariance structure in spatial modeling. It is very flexible; for example, the Gaussian and exponential covariance functions can be obtained as particular cases of that function. Additionally, it has good approximation properties for other spatial dependency structures, as discussed in **??**), and computational advantages that will be discussed later.

The marginal variance σ^2 is defined by:

$$\sigma^2 = \frac{\Gamma(\nu)}{4\pi\kappa^{2\nu}\tau^2\Gamma(\nu + \frac{d}{2})} \tag{1.4}$$

where τ is a scaling parameter and d is the space dimension. Following Lindgren, Rue and Lindström (2011), we adopt a parameterization in terms of $\log\tau$ and $\log\kappa$ for the covariance

function:

$$\log \tau = \frac{1}{2} \log \left(\frac{\Gamma(\nu)}{\Gamma(\alpha)(4\pi)^{d/2}} \right) - \log \sigma - \nu \log \rho$$
$$\log \kappa = \frac{\log(8\nu)}{2} - \log \rho \tag{1.5}$$

where $\rho = \frac{(8\nu)^{1/2}}{\kappa}$. This representation is advantageous since, conditional on the value of ν, it is necessary to estimate only two parameters.

Considering a bounded region $\Omega \in R^2$, it follows that the likelihood for an LGCP associated with data $Y = \{s_i \in \Omega : i = 1, \ldots, n; t = 1, \ldots, T\}$ is of the form

$$\pi(Y|\lambda) = \exp\left(|\Omega| - \int_{\Omega} \lambda(s,t) ds \right) \prod_{t=1}^{T} \prod_{i=1}^{n_t} \lambda(s_i, t). \tag{1.6}$$

Due to the doubly-stochastic property of the intensity function, the likelihood in (1.6) is analytically intractable. Since the term $\omega(s,t)$ corresponds to a GF with Matérn covariance, it is possible to use the SPDE approach, proposed by Lindgren, Rue and Lindström (2011), to approximate the initial GF to a GMRF, which has very good computational properties due to Markov structure, providing a sparse representation of the spatial effect through a sparse precision matrix (KRAINSKI et al., 2018).

The first main important result, provided by Whittle (1954) and extensively used by Lindgren, Rue and Lindström (2011), is that a GF x(s) with the Matérn covariance function is a stationary solution to the linear fractional SPDE

$$(\kappa - \Delta)^{\alpha/2} x(s) = W(s), \quad s \in^d, \quad \alpha = \nu + d/2, \quad \kappa > 0, \quad \nu > 0 \tag{1.7}$$

where $\Delta = \sum_{i=1}^{d} \frac{\partial^2}{\partial s_i^2}$ is the Laplacian operator and $W(s)$ is a spatial white noise. Therefore, in order to find a GMRF approximation of a GF, we first need to find the stochastic weak solution of SPDE (1.7).

Lindgren, Rue and Lindström (2011) proposed to use Finite Elements Method (FEM) to construct an approximated solution of SPDE. By proposing FEM, Lindgren, Rue and Lindström (2011) provided a solution for the case of irregular grids, since one of the big advantages of the FEM method is the irregularity, i.e., the domain can be divided into a irregular non-intersecting set of elements. The approximation of SPDE solution is given by

$$x(s,t) \approx \tilde{x}(s,t) = \sum_{j=1}^{n} w_j \varphi_j(s,t) \tag{1.8}$$

where n is the number of vertices of the triangulation, $\{w_j\}_{j=1}^{n}$ are the weights with Gaussian distribution and $\{\varphi_j\}_{j=1}^{n}$ are the basis functions defined for each node on the mesh.

Replacing the GF $\omega(s,t)$ by the GMRF approximation $\tilde{\omega}(s,t)$ in equation (1.1), we obtain

$$\tilde{\xi}(s,t) = \Theta \tilde{\xi}(s, t-1) + \tilde{\omega}(s,t). \tag{1.9}$$

Simpson et al. (2016) show that by replacing $\xi(s,t)$ by $\tilde{\xi}(s,t)$ in the intensity function (1.1) and approximating the integral (1.6) by a quadrature rule, it results that the approximate

likelihood consists of $(n + n_t)T$ independent Poisson random variables, where n is the number of vertices and n_t is the number of observed point processes.

According to Simpson et al. (2016), the LGCP formulation fits naturally within the Bayesian hierarchical modelling framework and are latent Gaussian models, therefore, it may be fitted using the let approach of Rue, Martino and Chopin (2009).

1.2.2 Data

We used in this work a subset of the Storm Prediction Center's Severe Weather Database file. Although the database covers the 1950-2018 period, we propose to subset the 1954-2018 period in order to remove the undercount of tornadoes in the first years, as discussed by Agee and Childs (2014). In order to facilitate the visualization of the results, we use a yearly aggregation of the daily data. The database also includes damage rating scales, Fujita (F) - prior to 2007 - and Enhanced Fujita (EF) - beginning in 2007. Both scales range from 0 (lowest damage indicator) to 5 (highest damage indicator). Table 1.1 shows some descriptive statistics for the annual counts of tornadoes in each category, and Figure 1.1 show the number of tornado occurrences reported in each year and each intensity.

In this figure it is possible to observe an increasing in the number of (E)F0 tornadoes after 1990, which can be related to non-meteorological factor (MOORE, 2017). On the other hand, (E)F1 and stronger tornadoes appear to be are stationary. In addition, it is possible to observe a slightly decreasing in the number of (E)F1 and (E)F2 tornadoes after 1974, which can be related changes in to tornado intensity classfication adopted by National Weather Service, as discussed by Moore (2017). The occurrences of (E)F5 tornadoes are extremely rare, as can be seen in Table 1.1, therefore, hereafter we just consider (E)F4 and weaker tornadoes.

	Total	Mean	SD	Min	Max
(E)F0	28844	443.754	267.438	86	1186
(E)F1	20760	319.385	86.037	170	614
(E)F2	8726	134.246	53.257	61	299
(E)F3	2320	35.692	17.064	10	94
(E)F4	520	8.000	5.804	0	30
(E)F5	54	0.831	1.387	0	7

Table 1.1 – Summary Statistics

1.3 Results

In order to apply the inference procedures discussed in Section 1.2, the first step is to define a triangulation mesh of the interest region. Here, we define a triangulation mesh with 439 triangles covering all the United States area. A discussion of criteria for choosing mesh in estimating LGCP models can be found in Simpson et al. (2016). As discussed in this work, the use of the continuous approximation for LGCP avoids the use of very fine meshes in the process approximation, and thus allows to represent and estimate the parameters and random effects in a computationally efficient way. This characteristic is especially important in space-time models, where the computational representation of the process is very intensive in memory.

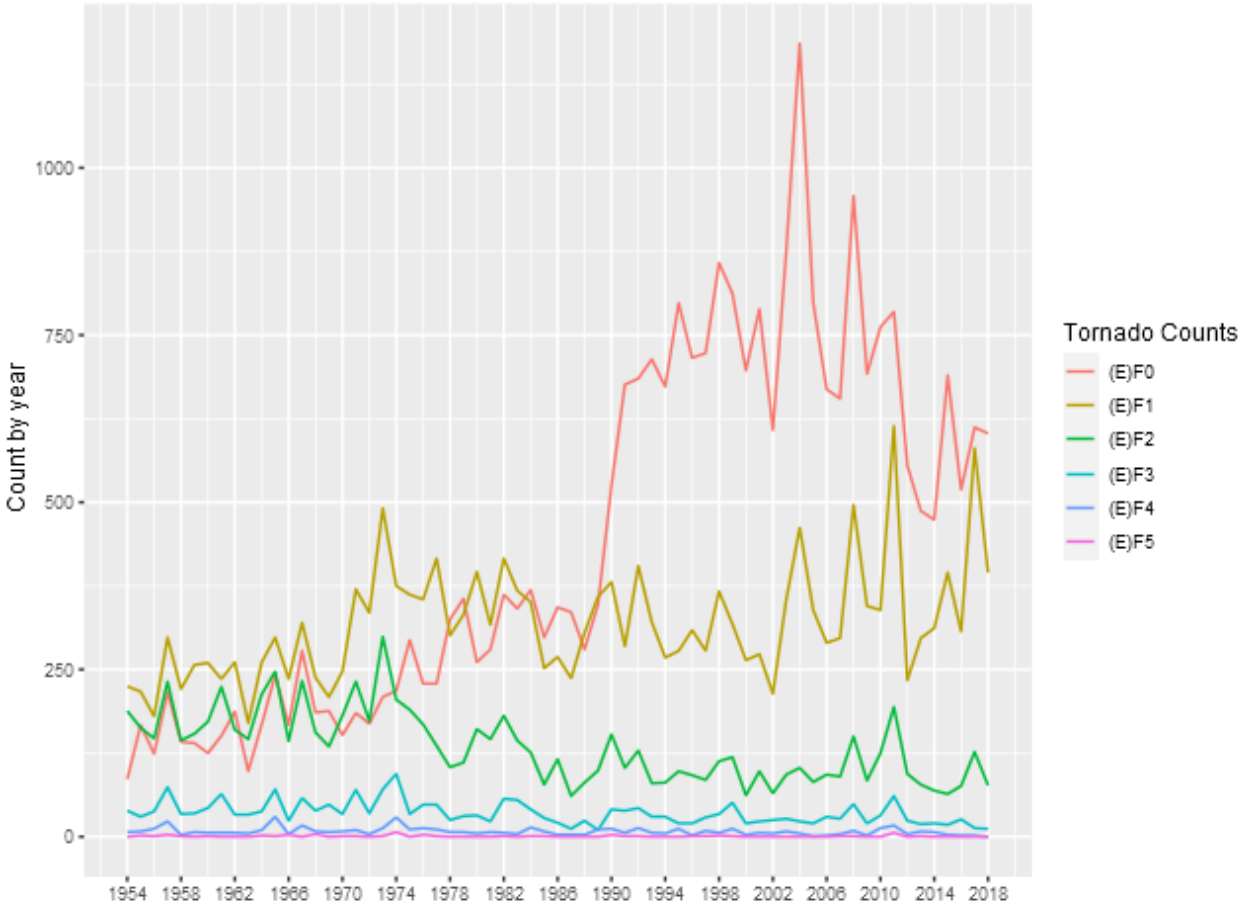


Figure 1.1 – Tornado Counts by Year

The second step is to define a set of knots over time to built a temporal mesh. For the time domain, we define a temporal mesh based on the number of observed years, 65. In order to obtain the space-time aggregation, we find to which polygon belongs each data point in the spatial mesh and to which part of the time belongs each data point. Hereafter, we use these both identification index sets to aggregate the data.

As described in equation 1.1, the parameters to be estimated are the parameters associated to the intercept (α), the precision of the trend component ($1/\eta_\mu$) and cycle component ($1/\eta_c$). The parameters of the second-order autoregressive process of the cycle as parameterized as partial autocorrelatios (PACF1 and PACF2), whereas the parameters of spatial covariance are represented by $\log \tau$ and $\log \kappa$ and the parameter of spatial time dependence (Φ).

In Table 1.2 we present the estimated posterior distribution for the parameters associated with the model given by equation 1.1, for the tornado classifications. The intercept parameter estimates the unconditional mean of the estimated log intensity function, and we can see that the values are consistent with the observed magnitude of the number of tornadoes in each classification. The precision parameters report the variability associated with the trend and cycle processes. Greater precision is associated with less variability in the component, and again the results are consistent with the general patterns of the number of tornadoes occurring in each category analyzed. The partial correlation parameters are related to the autoregressive

Table 1.2 – Posterior distribution of estimated parameters

	Mean	SD	0.025q	0.5q	0.975q	Mode
(E)F0 tornadoes						
Intercept	-0.91	0.07	-1.05	-0.91	-0.77	-0.91
Precision for trend	82.81	13.45	59.62	81.70	112.33	79.48
Precision for cycle	35.56	6.40	23.60	35.47	48.45	35.58
PACF1 for cycle	-0.00	0.08	-0.17	-0.00	0.16	0.01
PACF2 for cycle	-0.05	0.11	-0.25	-0.05	0.19	-0.07
Log τ	-0.77	0.04	-0.85	-0.77	-0.70	-0.77
Log κ	-0.67	0.04	-0.74	-0.68	-0.60	-0.68
Group Φ	0.90	0.01	0.88	0.90	0.91	0.90
(E)F1 tornadoes						
Intercept	-1.07	0.12	-1.30	-1.07	-0.85	-1.07
Precision for trend	597.09	101.69	390.67	603.27	775.28	630.87
Precision for cycle	19.15	2.02	15.16	19.19	23.04	19.44
PACF1 for cycle	0.11	0.06	0.01	0.11	0.23	0.09
PACF2 for cycle	-0.02	0.05	-0.11	-0.02	0.08	-0.02
Log τ	-0.51	0.04	-0.59	-0.51	-0.42	-0.51
Log κ	-0.98	0.04	-1.07	-0.98	-0.90	-0.97
Group Φ	0.93	0.01	0.92	0.93	0.95	0.93
(E)F2 tornadoes						
Intercept	-1.60	0.15	-1.89	-1.60	-1.30	-1.60
Precision for trend	514.76	136.17	262.99	513.72	781.10	512.46
Precision for cycle	76.14	15.58	50.31	74.52	111.19	71.35
PACF1 for cycle	0.22	0.12	0.01	0.21	0.48	0.16
PACF2 for cycle	-0.06	0.18	-0.44	-0.05	0.27	0.01
Log τ	-0.29	0.06	-0.40	-0.29	-0.18	-0.29
Log κ	-1.24	0.05	-1.35	-1.24	-1.13	-1.24
Group Φ	0.88	0.04	0.77	0.89	0.93	0.92
(E)F3 tornadoes						
Intercept	-2.47	0.22	-2.92	-2.47	-2.03	-2.46
Precision for trend	1133.60	396.70	460.09	1108.60	1969.11	1027.53
Precision for cycle	32.85	9.22	18.78	31.49	54.61	28.94
PACF1 for cycle	-0.14	0.14	-0.39	-0.15	0.17	-0.20
PACF2 for cycle	-0.06	0.21	-0.48	-0.05	0.33	0.00
Log τ	-0.06	0.09	-0.25	-0.05	0.11	-0.04
Log κ	-1.51	0.09	-1.68	-1.51	-1.32	-1.53
Group Φ	0.93	0.01	0.90	0.93	0.95	0.93
(E)F4 tornadoes						
Intercept	-3.62	0.37	-4.38	-3.60	-2.93	-3.57
Precision for trend	24904.71	25883.49	2088.27	17256.64	93210.40	5798.00
Precision for cycle	13.30	8.86	4.14	10.85	36.69	7.75
PACF1 for cycle	-0.35	0.27	-0.80	-0.38	0.24	-0.45
PACF2 for cycle	-0.22	0.31	-0.76	-0.23	0.40	-0.24
Log τ	-0.32	0.13	-0.57	-0.32	-0.06	-0.33
Log κ	-1.49	0.13	-1.76	-1.49	-1.23	-1.48
Group Φ	0.95	0.01	0.92	0.95	0.97	0.96

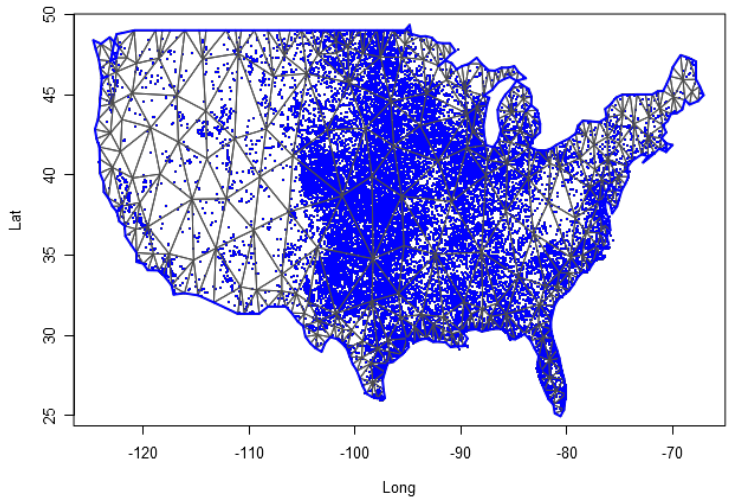


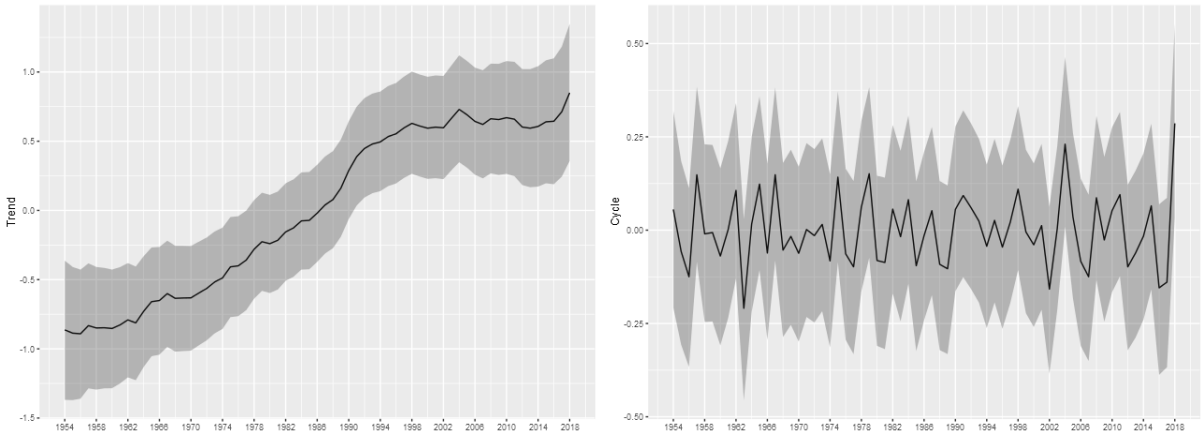
Figure 1.2 – Spatial Mesh and (E)F0 Tornadoes

parameters in the AR (2) representation of the cycle. In all estimates, the results indicate the presence of a cyclic component, since the roots of the lag polynomial associated with the estimated partial correlation coefficients fall in the complex region. The estimated period for the cycle components for (E)F0, (E)F1, (E)F2, (E)F3 and (E)F4 tornadoes are, respectively, 3.98, 5.63, 5.59, 2.78 and 3.09 years. These periods are consistent with those observed for climatic phenomena associated with interactions between sea temperature and the atmosphere, such as El Niño and La Niña, as discussed in Cook et al. (2017). Note that the interpretation of this cycle component is an approximation for the effects of all shocks with persistent but non-permanent effects that affect the occurrence of tornadoes.

The estimated trend and cycle components for each tornado intensity are shown in Figures 1.3-1.7, which present the posterior mean of the estimated components with 95% Bayesian credibility interval. The most important result is related to trend component. Considering (E)F0 tornadoes, it is possible to observe a growth pattern in the trend component between 1962 and 2000, whereas between 2000 and 2016 it remains stable. From 2016 can be seen a growth pattern again. This result is consistent to results commonly found in the literature and, as discussed by Moore (2017), part of this effect can also be attributed to non-meteorological factors such as changes in reporting practices and technology.

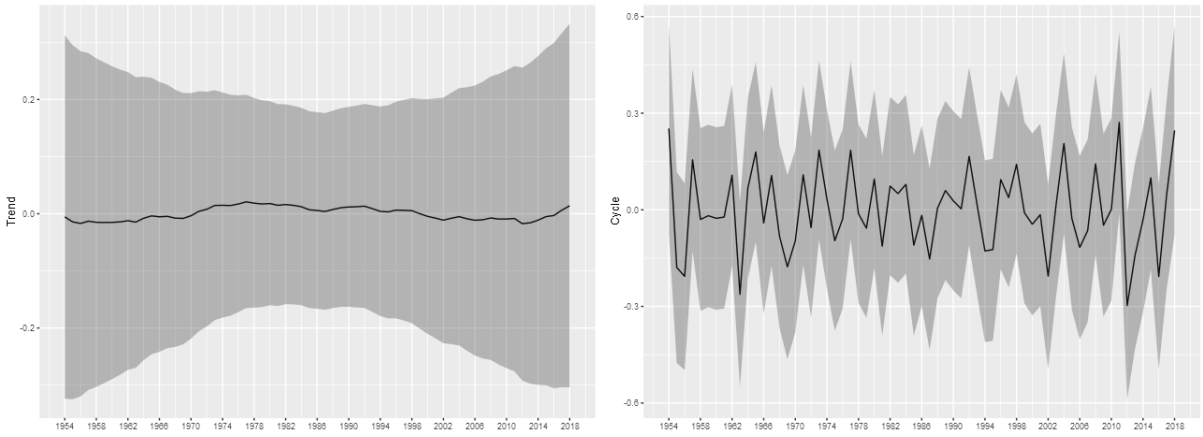
For the tornadoes of (E)F1-(E)F4 classification, it is not possible to infer permanent changes in the occurrence intensity. The estimated trend components do not show relevant variations, being all estimated close to zero, with the credibility intervals always covering this value. The results obtained through the modeling carried out in this article do not indicate the presence of relevant changes in the trend of occurrences for these categories of tornadoes. Considering (E)F1-(E)F4 tornadoes, the results are consistent with stationary patterns in the temporal counts of tornado occurrences, consistent with results found in Kunkel et al. (2013) and Moore (2017).

To show the importance of the trend and cycle components in the analysis of the tornado occurrence patterns, we show in Figure 1.8 the predicted value for the tornado count (E)F0 in



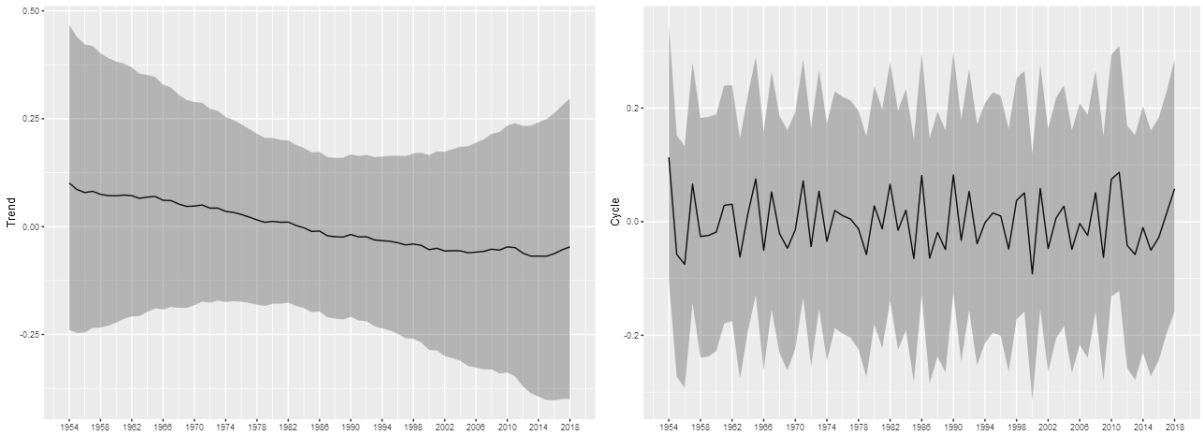
(a) Trend (b) Cycle

Figure 1.3 – Trend and Cycle decomposition - (E)F0



(a) Trend (b) Cycle

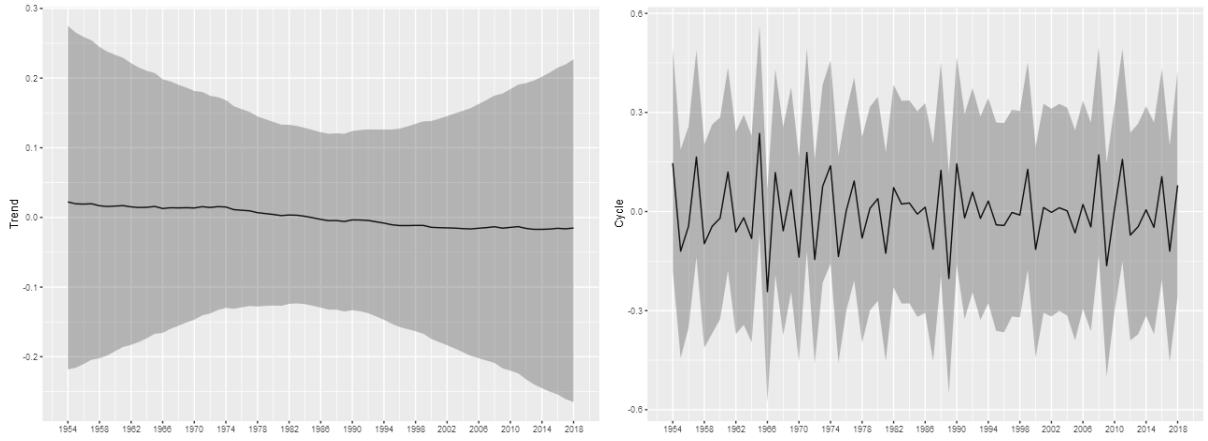
Figure 1.4 – Trend and Cycle decomposition - (E)F1



(a) Trend (b) Cycle

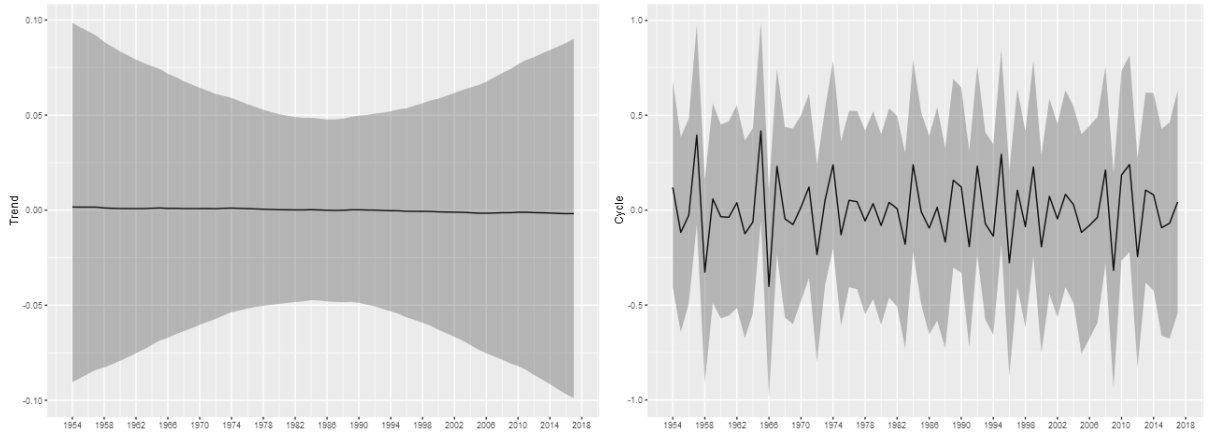
Figure 1.5 – Trend and Cycle decomposition - (E)F2

each year given by the sum of the trend, cycle and intercept components of model, and also the credibility interval of this sum. This represents the prediction of the total count ignoring the



(a) Trend (b) Cycle

Figure 1.6 – Trend and Cycle decomposition - (E)F3



(a) Trend (b) Cycle

Figure 1.7 – Trend and Cycle decomposition - (E)F4

spatial random effect, obtained by aggregating the exposure of the entire area multiplied by the exponential of the sum of these components. The analogous figures for the other classifications (Figures 1.12-1.15) are shown in the Appendix.

We can see that for the (E)F0-(E)F1 classifications, the random effects of trend and cycle explain a large part of the variability observed in the total tornado count. For the other classifications, the spatial component is more relevant in explaining the count observed each year, indicating a greater spatial heterogeneity in the intensity of occurrences.

The spatial heterogeneity of the tornado occurrences can be seen through the estimated spatial random effect $\tilde{\xi}(s,t)$. In this section we show in Figure 1.9 the estimated posterior distribution of spatial random effect for (E)F0 tornadoes, and in the Appendix the results for the other classifications in Figures 1.16-1.19.

In these figures it is possible to observe the variability captured by the spatial random effects, especially in the region known as "Tornado Alley", which is a nickname given to an area in the southern plains of the central United States that experiences a high frequency of tornadoes each year (NOAA). In addition, after 1980s, it is also possible to observe an increasing in the

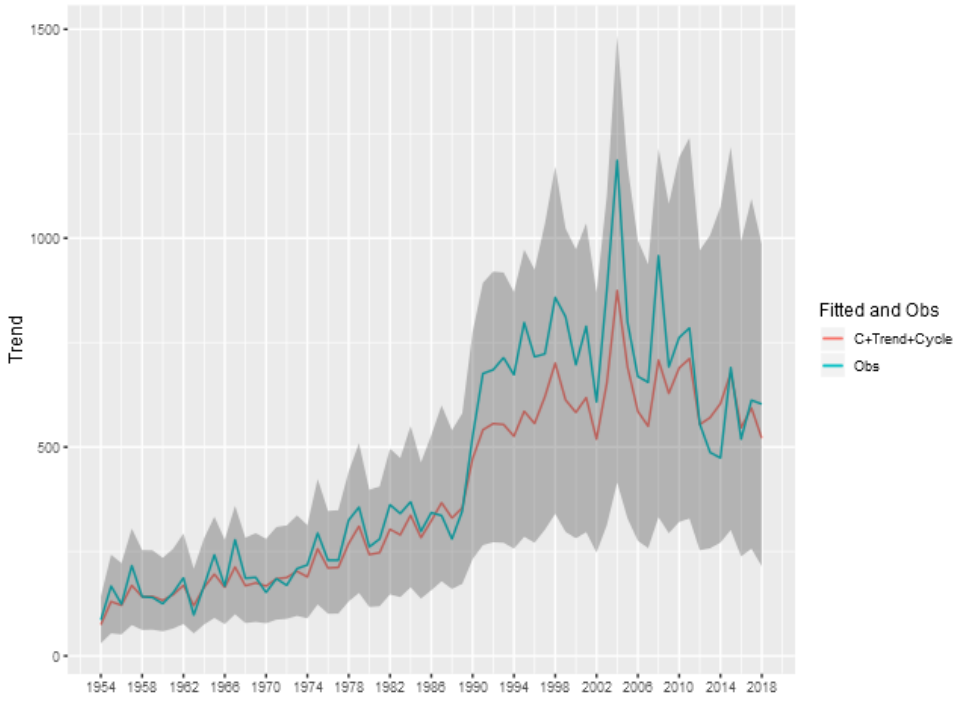


Figure 1.8 – Fitted and Observed Tornadoes without consider the spatial effects - (E)F0

variability in the Southeast region, for all intensities considered, which is consistent with finds reported in the literature, e.g., Moore and DeBoer (2019) and Gensini and Brooks (2018).

In Figure 1.10 we show the estimated intensity function for the (E)F0 classification tornadoes, given by the sum of all random components (trend, cycle and spatial) in each year. The analogous results for the other classifications are shown in the Figures 1.20-1.23 in the Appendix.

In each year shown in these figures we also show the observed occurrences of tornadoes as black dots. We can observe that the estimated intensity log function adequately explains the spatio-temporal variation observed in the tornado count for all classifications, indicating that the space-time LGCP model proposed in this work has an adequate fit to the analyzed pattern of occurrences.

1.3.1 Zero-inflated Poisson

Data were taken for different intensities of tornadoes, based on Fujita (F) and Enhanced Fujita (EF) damage rating scales. Historically, stronger tornadoes are less likely to occur, which led to the database having many zero counts. For classifications with very few occurrences, a possible modification is to change the likelihood function to account for the possibility of an excess of zeros. Mixed-distribution models, such as Zero-Inflated-Poisson (ZIP) can be used in such cases. According to Lambert (1992), ZIP is a model for count with excess zeros, which assumes that with probability p the observation is 0, and with probability $1 - p$, a Poisson is observed.

We consider two different types of ZIP models: type 0 and type 1. The type 0 likelihood is

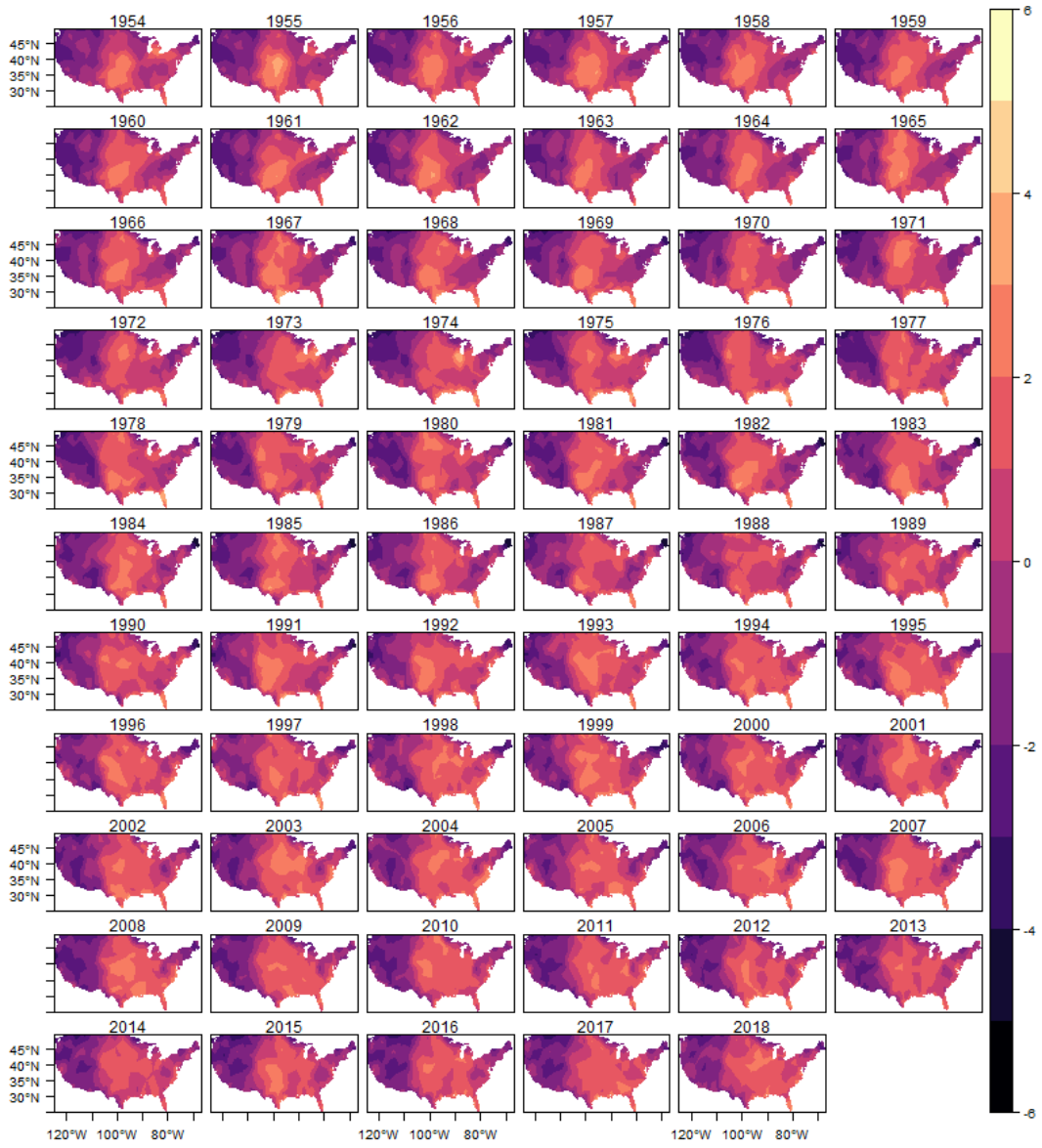


Figure 1.9 – Spatial Random Effects - (E)F0

defined as

$$Prob(y|\ldots) = p \times 1_{[y=0]} + (1-p) \times \text{Poisson}(y|y>0) \tag{1.10}$$

where p is a hyperparameter where

$$p = \frac{\exp(\theta)}{1+\exp(\theta)} \tag{1.11}$$

and θ is the internal representation of p, meaning that the initial value and prior is given by θ. On the other hand, the type 1 is defined as

$$Prob(y|\ldots) = p \times 1_{[y=0]} + (1-p) \times \text{Poisson}(y) \tag{1.12}$$

where p is a hyperparameter where

$$p = \frac{\exp(\theta)}{1+\exp(\theta)} \tag{1.13}$$

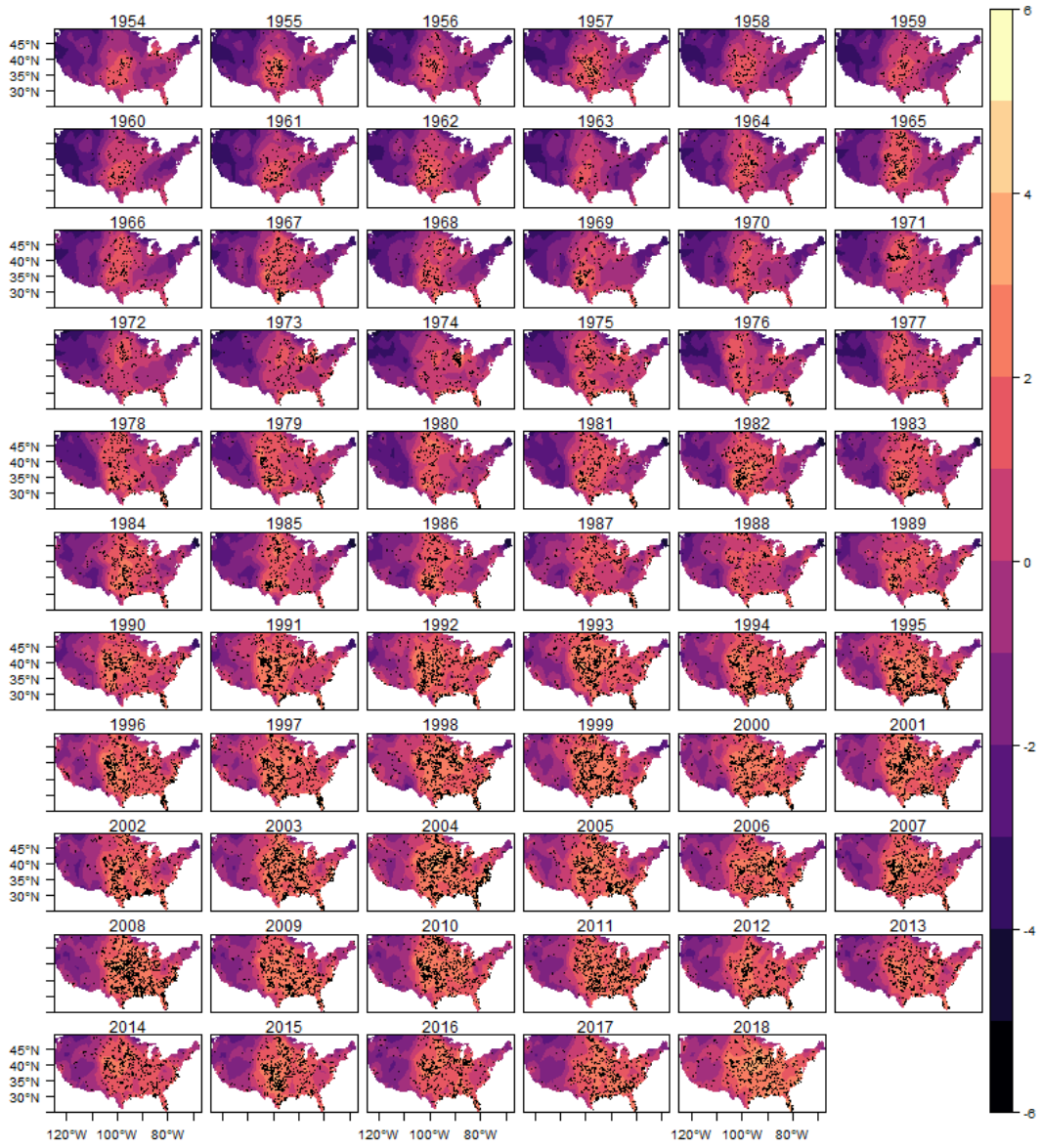


Figure 1.10 – Fitted Log-Intensity and Observed Tornadoes - (E)F0
Note: Black points indicate observed tornado occurrences.

where θ has the same meaning above-mentioned. As discussed by Serra et al. (2014), the only difference between type 0 and 1 is the conditioning on $y > 0$ for type 0, which means that for type 0 the probability that y is equal 0 is p, whereas for type 1, the same probability is $p + (1 + p)Poisson(y)$.

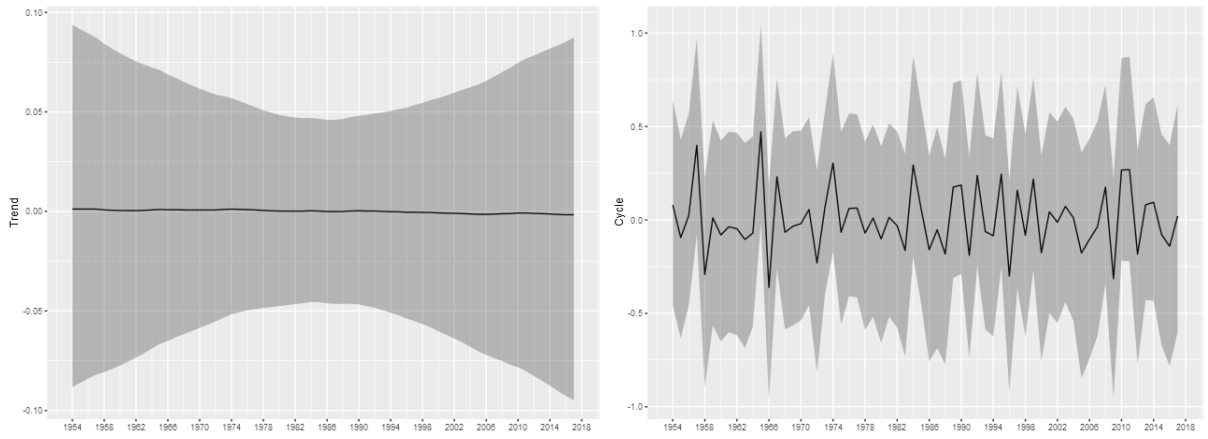
Table 1.3 shows the estimated posterior distribution for the parameters associated to ZIP model for (E)F4 tornadoes, considering type 1. For space issue we show only the results obtained with type 1, which had best results, in terms of fit, in relation to type 0. The results obtained with type 0 are available with the authors. As in Table 1.2, the precision parameters represent the variability associated with trend and cycle components. The results indicate great precision associated to trend component whereas the precision of cycle component are relatively minor.

Figure 1.11 shows the estimated trend and cycle components for the (E)F4 tornadoes, considering the ZIP model with type 1, where it is possible to observe the absence of permanent

Table 1.3 – Posterior distribution of estimated parameters - ZIP model Type 1

	Mean	SD	0.025q	0.5q	0.975q	Mode
Intercept	-3.202	0.636	-4.542	-3.179	-1.985	-3.137
ZIP parameter	0.480	5.00e-02	0.381	0.480	0.576	0.482
Precision for trend	22484.136	2.01e+04	2716.213	16904.253	76040.634	7676.415
Precision for cycle	10.103	5.95e+00	3.092	8.670	25.554	6.482
PACF1 for cycle	-0.143	3.16e-01	-0.711	-0.152	0.478	-0.167
PACF2 for cycle	-0.047	1.85e-01	-0.417	-0.042	0.301	-0.017
Log τ	0.021	1.69e-01	-0.311	0.021	0.353	0.022
Log κ	-1.787	2.05e-01	-2.198	-1.785	-1.390	-1.775
Group Φ	0.983	8.00e-03	0.964	0.985	0.994	0.987

changes in the occurrence intensity. In addition, it is important to note that the ZIP model with type 1 results do not present significant changes in the trend and cycle components, relatively to those results obtained with Poisson distribution and shown in Figures 1.4-1.7.



(a) Trend (b) Cycle

Figure 1.11 – Trend and Cycle decomposition - (E)F4 - ZIP Type 1

1.4 Conclusions

Climate changes, associated to natural or human activities, can lead to changes in the likelihood of the occurrences of severe weather events, such as heat waves, droughts, tornadoes and hurricanes. The Unites States experiences more tornadoes per year than any other country, which are responsible for deaths and damages. Given the impact of tornadoes on society, understanding how these events are responding to climate changes is important in order to be prepared. Previous studies have searched for relationships between tornado activity and climate changes, for example Lee (2012), Diffenbaugh, Scherer and Trapp (2013) and Moore and McGuire (2019). This present paper contributes to this literature by analyzing tornado occurrences in the United States following a method proposed by Laurini (2019). Therefore, to perform inference procedures for the spatio-temporal point process we adopt a dynamic representation of Log-Gaussian Cox Process. This representation is based on the decomposition of intensity function in components of trend, cycles, and spatial effects. In this model, spatial effects are also

represented by a dynamic functional structure, which allows to analyze the possible change in the spatial distribution. The decomposition proposed in this article is therefore especially useful by analyzing possible changes in the temporal and spatial patterns of the occurrence of events, since it captures changes in the average number of occurrences through the trend component, and changes in the spatial distribution pattern through the time varying spatial random effect. We use daily data from Storm Prediction Center's Severe Weather Database between 1954 and 2018. The results have provided evidences, from new perspectives, that trend in annual tornado occurrences in the United States have remained relatively constant, supporting previously reported findings.

1.5 Appendix

1.5.1 Fitted and Observed Tornadoes Without Spatial Component

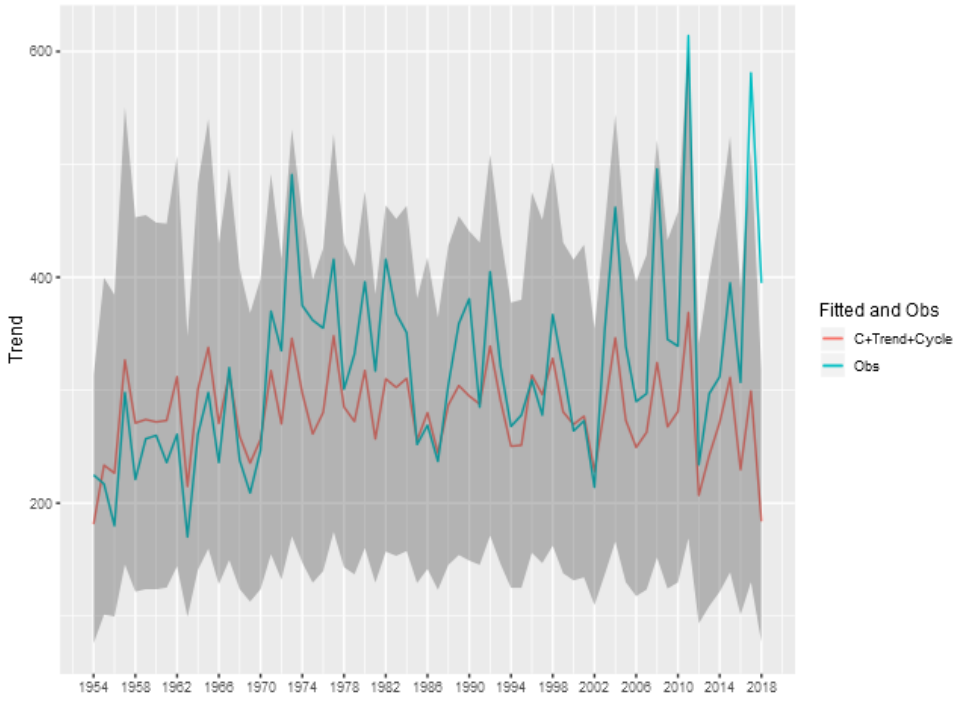


Figure 1.12 – Fitted and Observed Tornadoes without consider the spatial effects - (E)F1

1.5.2 Spatial Random Effects

1.5.3 Fitted Log-Intensity and Observed Tornadoes

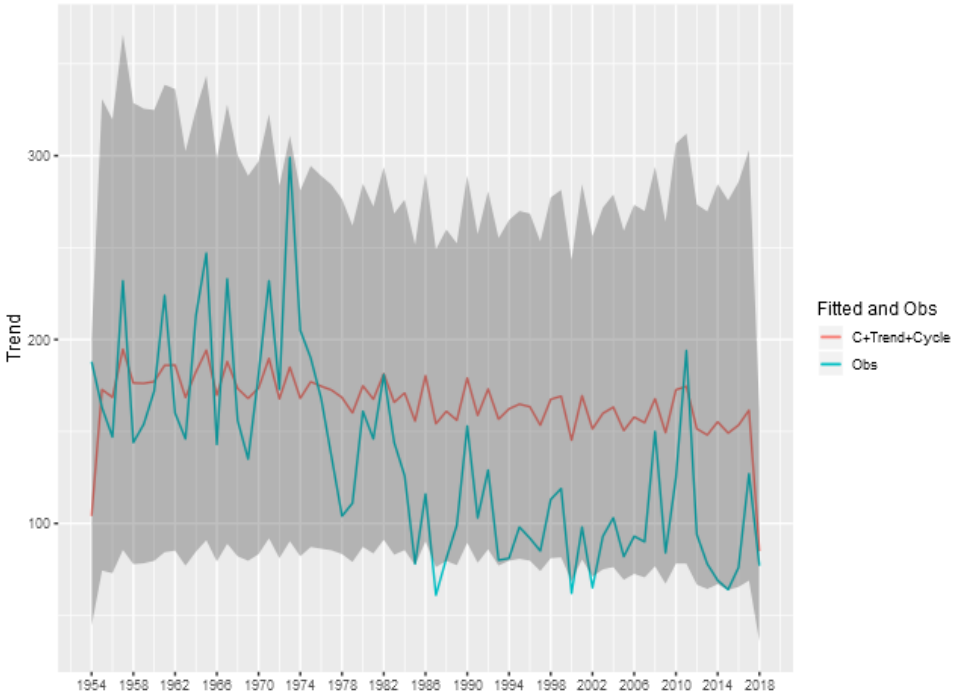


Figure 1.13 – Fitted and Observed Tornadoes without consider the spatial effects - (E)F2

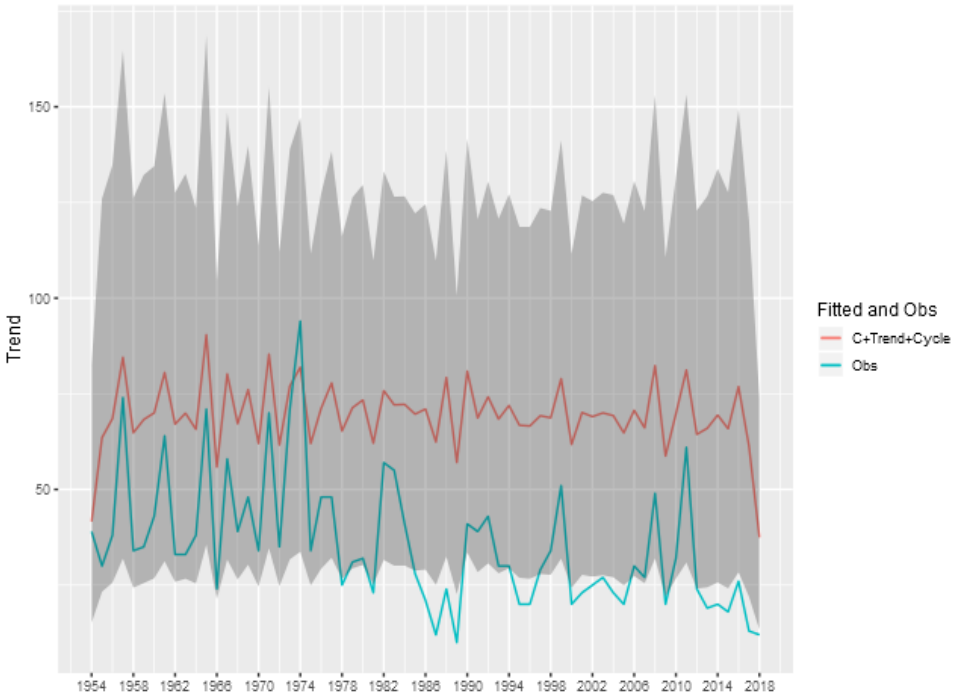


Figure 1.14 – Fitted and Observed Tornadoes without consider the spatial effects - (E)F3

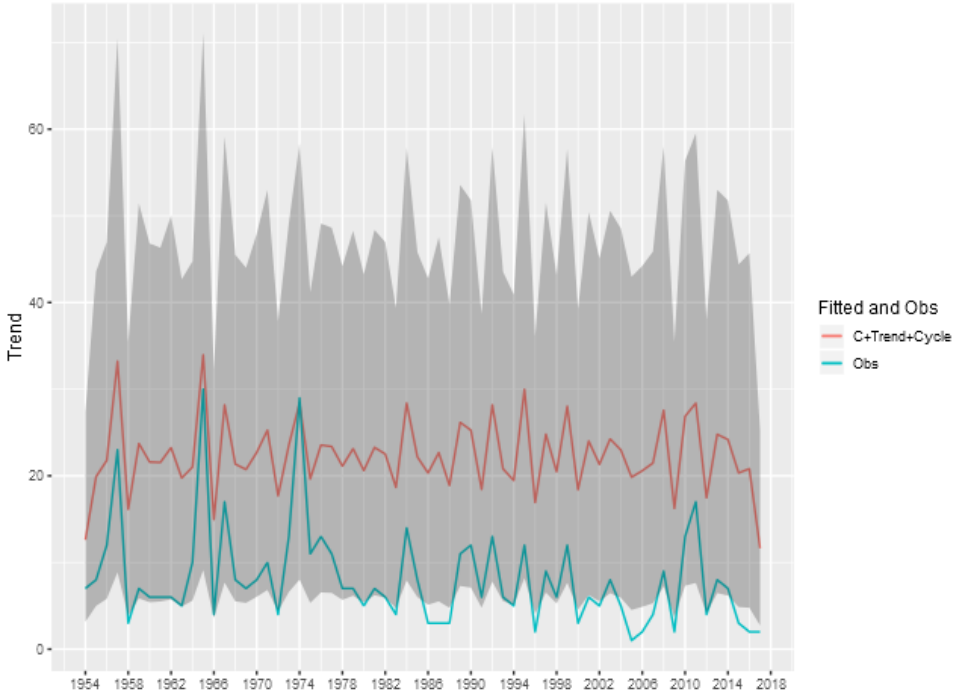


Figure 1.15 – Fitted and Observed Tornadoes without consider the spatial effects - (E)F4

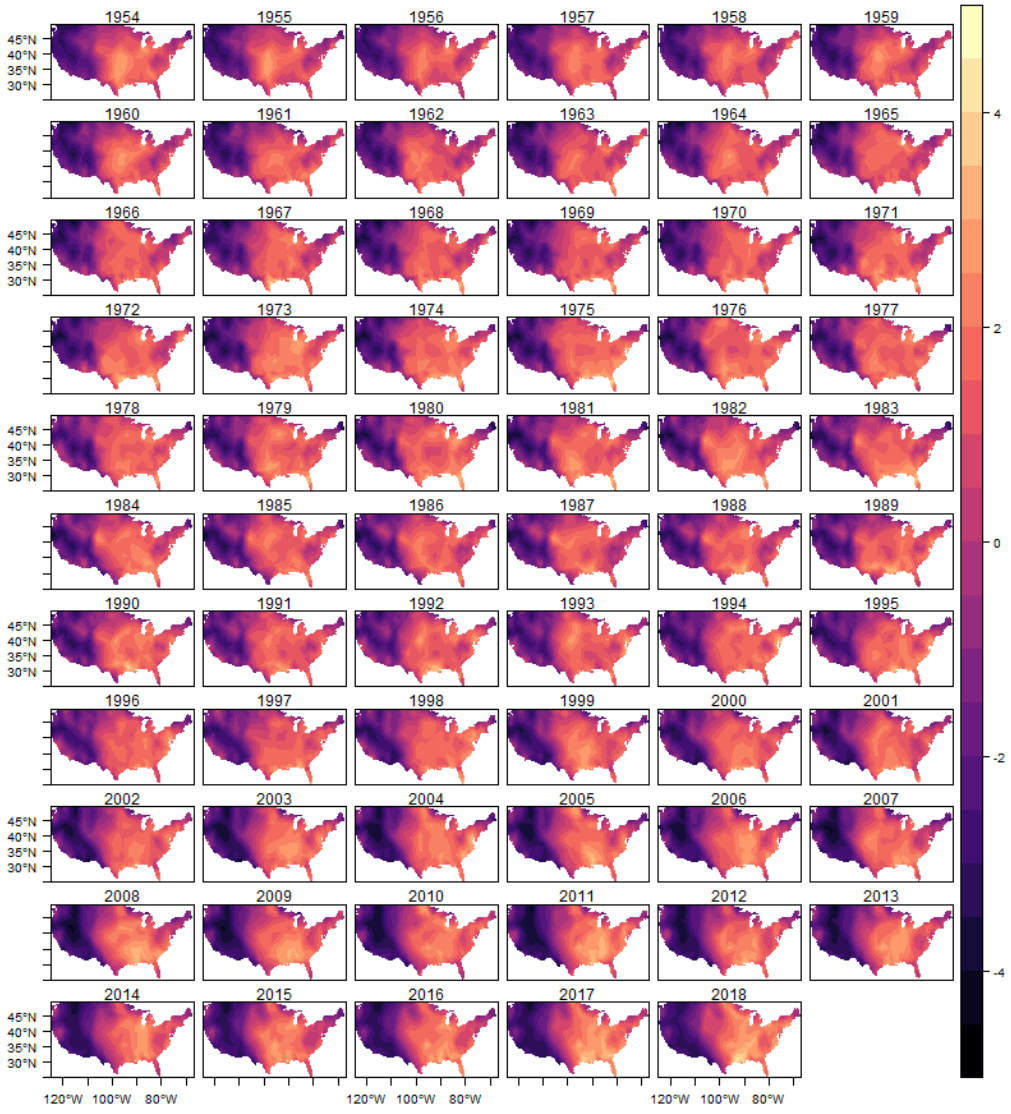


Figure 1.16 – Spatial Random Effects - (E)F1

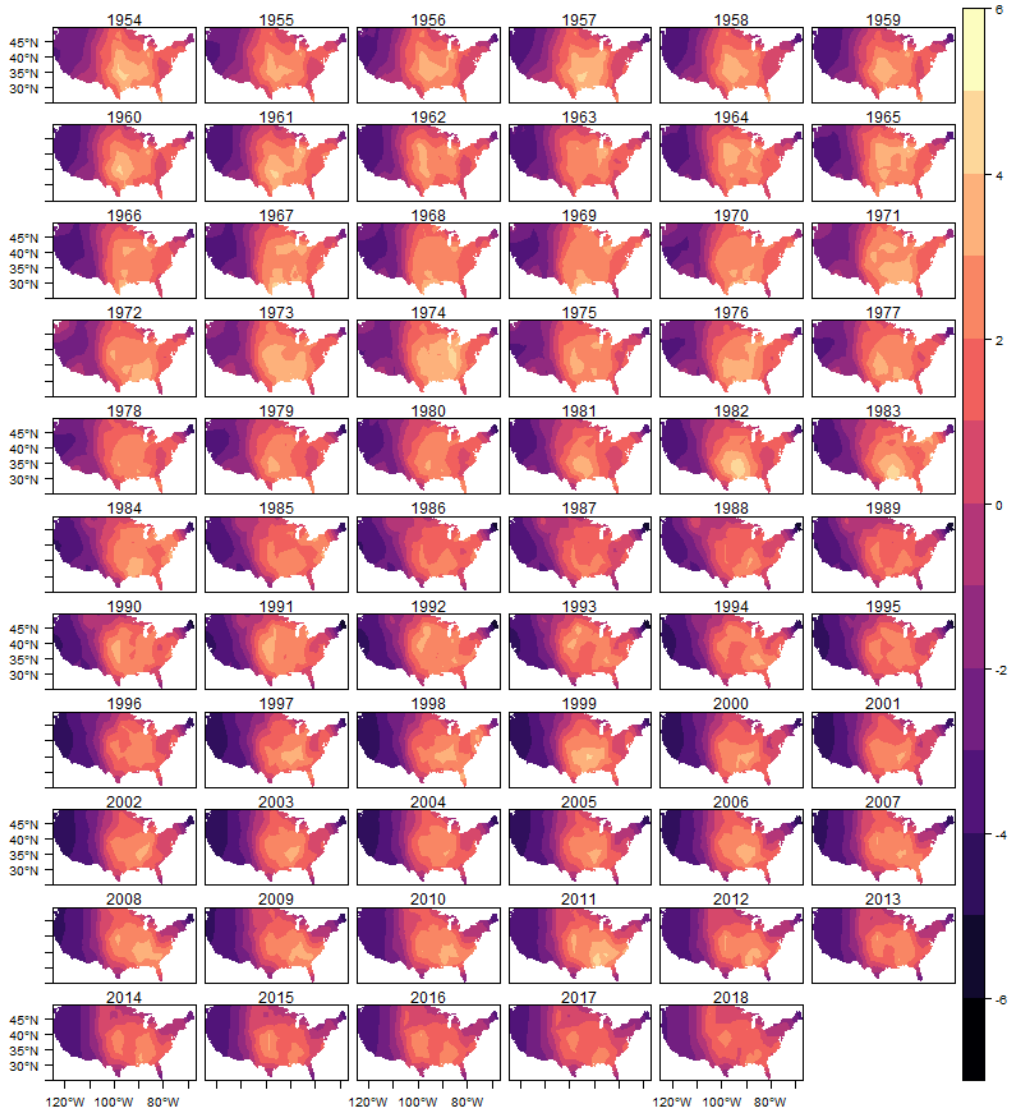


Figure 1.17 – Spatial Random Effects - (E)F2

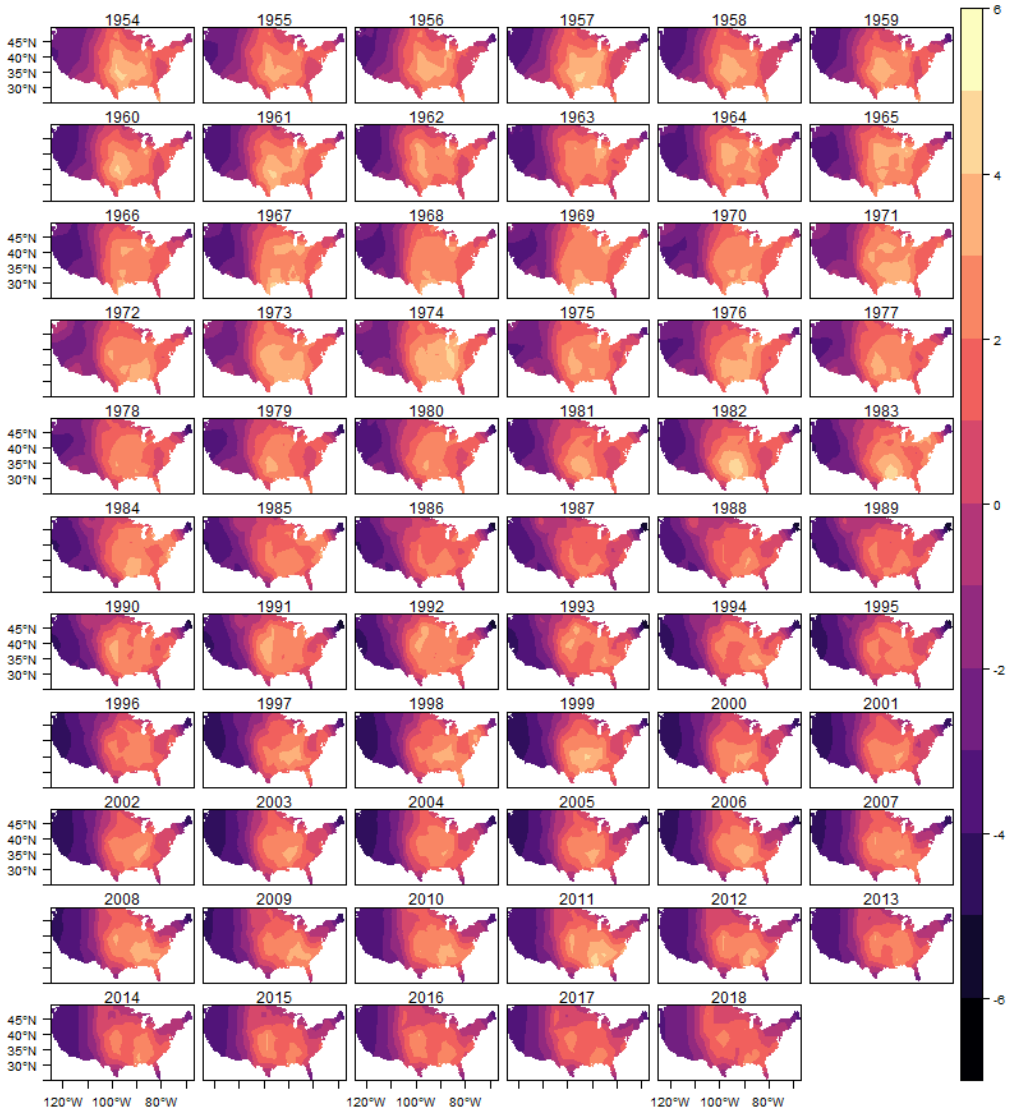


Figure 1.18 – Spatial Random Effects - (E)F3

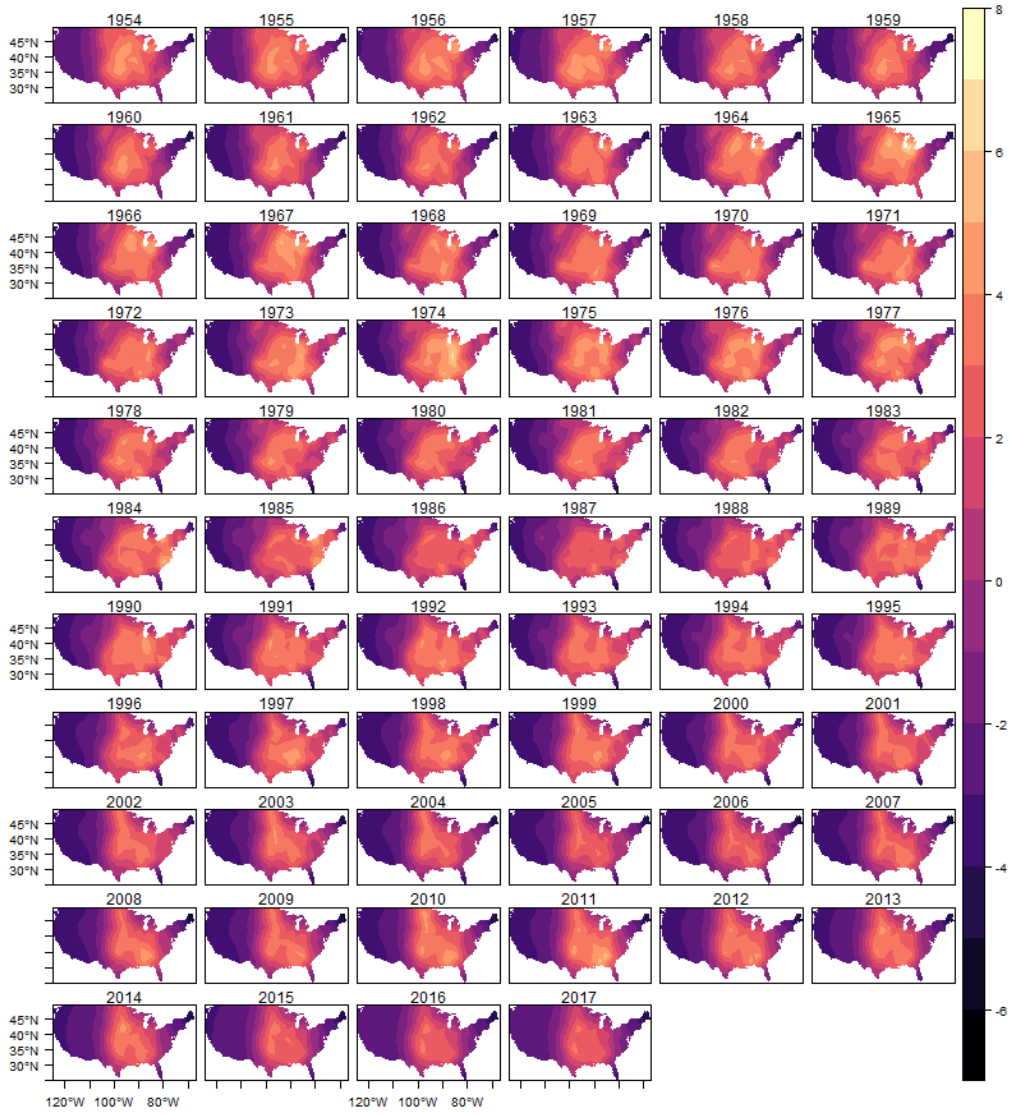


Figure 1.19 – Spatial Random Effects - (E)F4

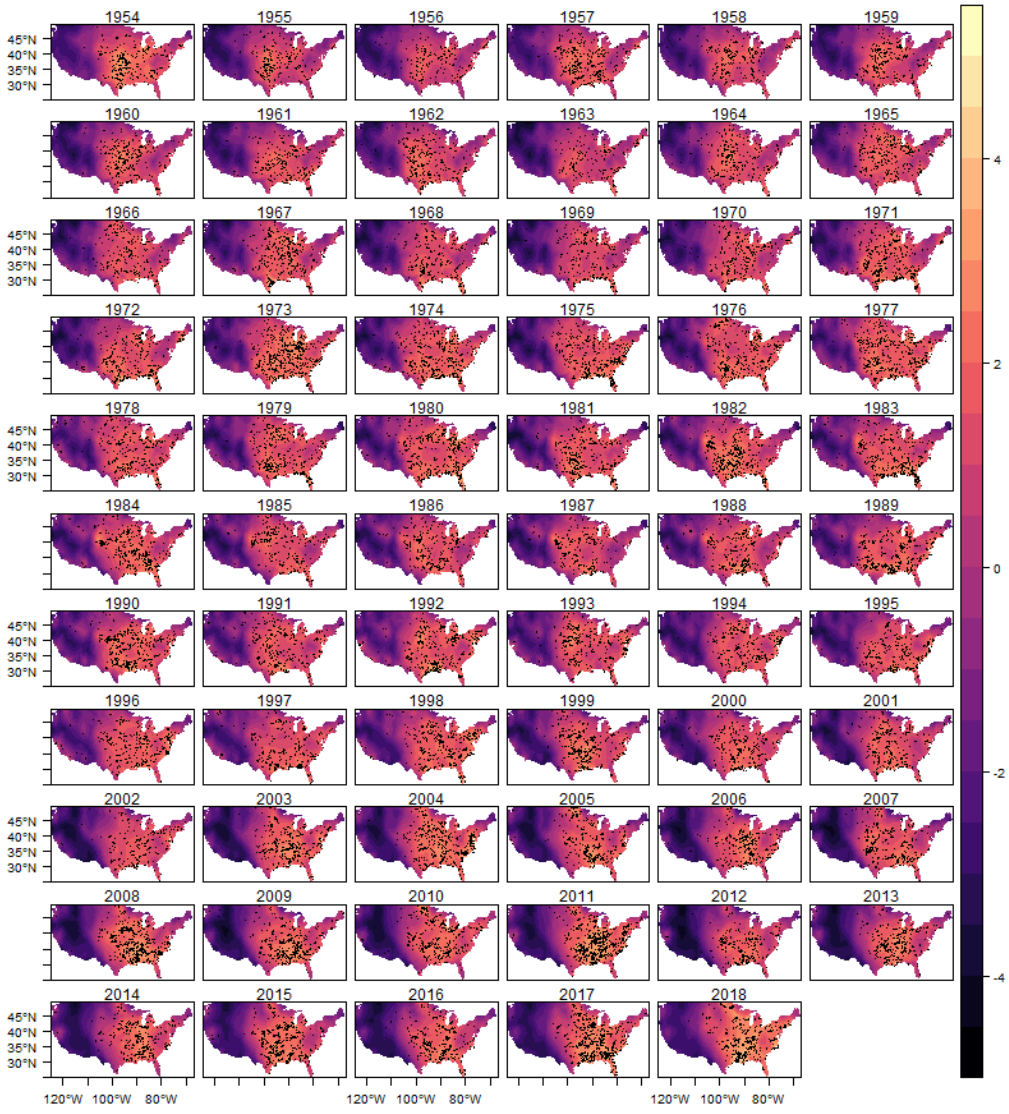


Figure 1.20 – Fitted Log-Intensity and Observed Tornadoes - (E)F1

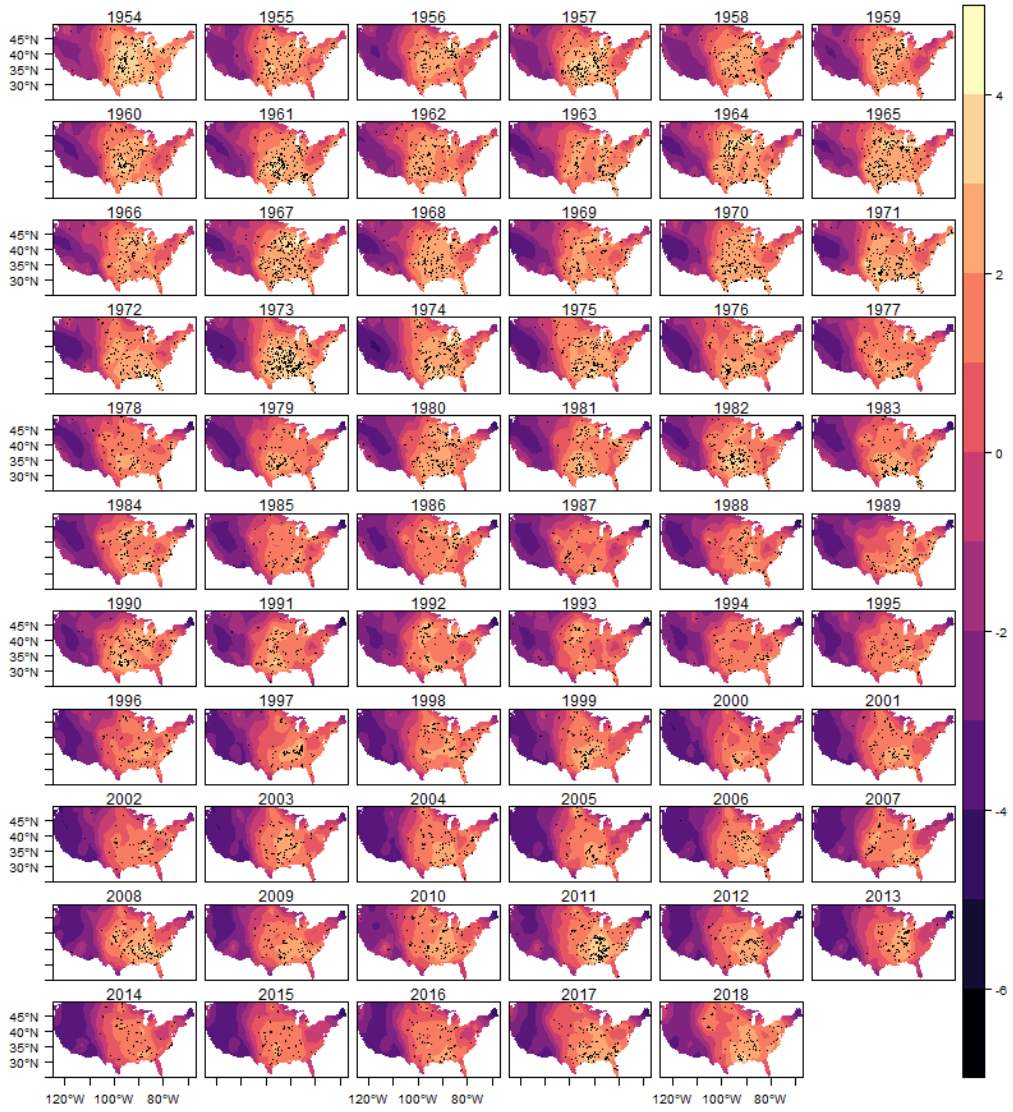


Figure 1.21 – Fitted Log-Intensity and Observed Tornadoes - (E)F2

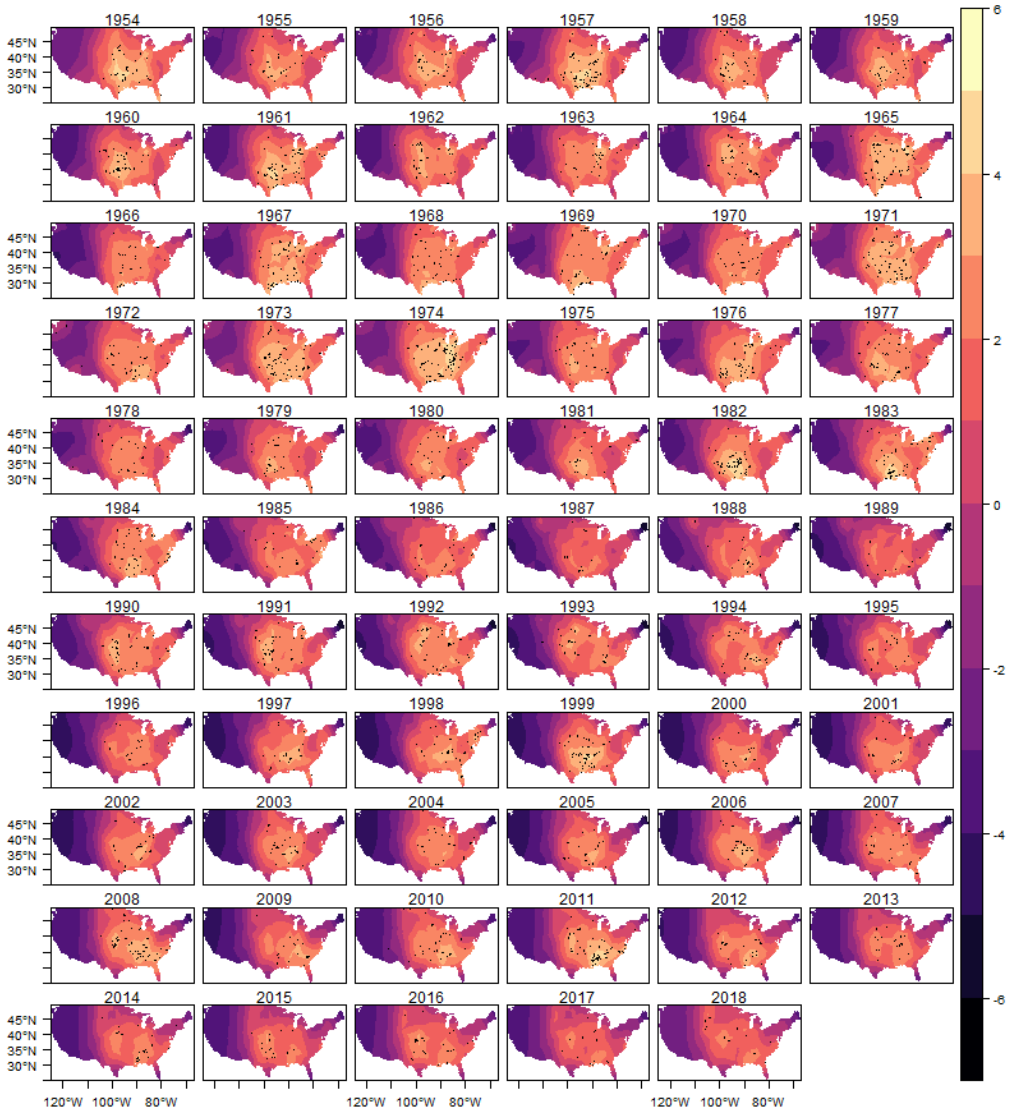


Figure 1.22 – Fitted Log-Intensity and Observed Tornadoes - (E)F3

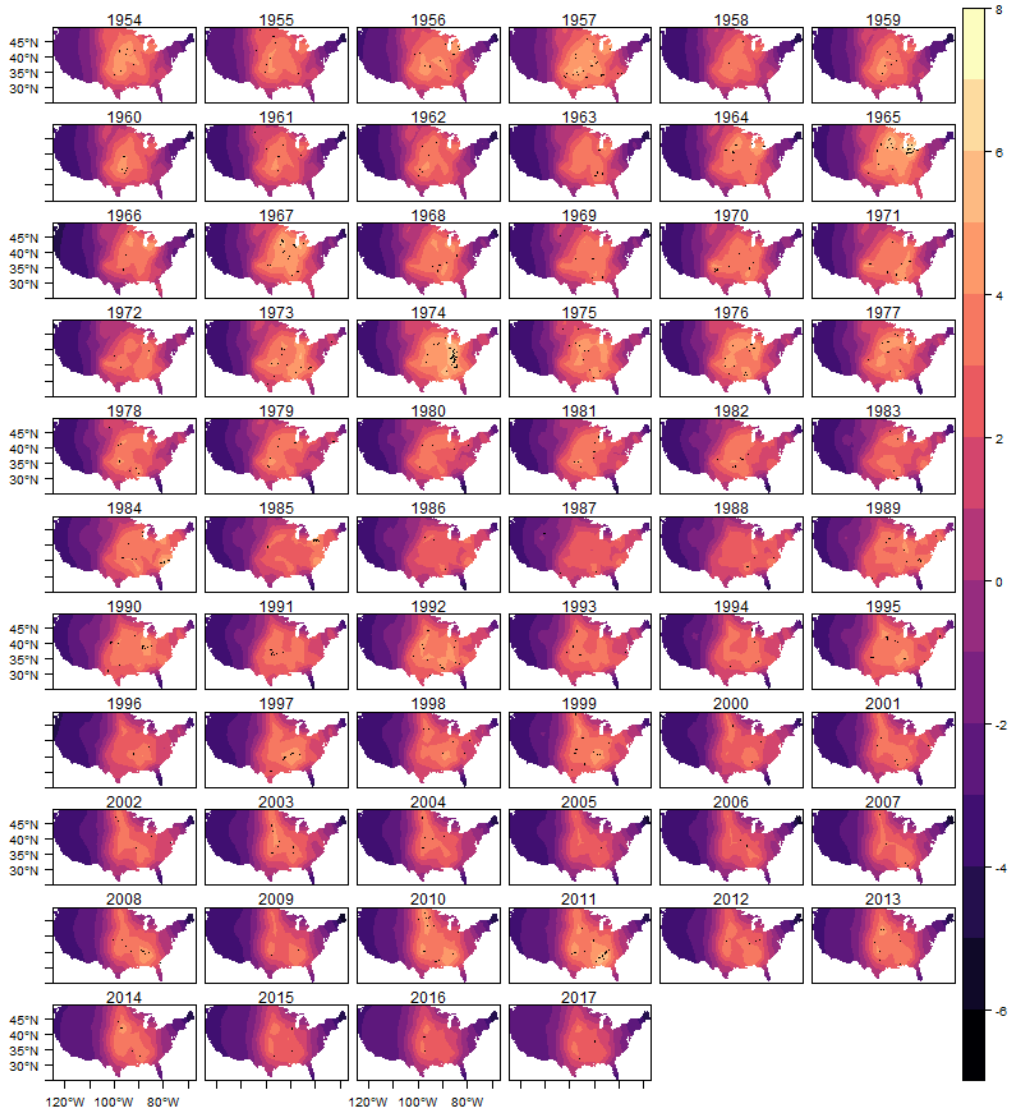


Figure 1.23 – Fitted Log-Intensity and Observed Tornadoes - (E)F4

2 PRE-HARVEST SUGARCANE BURNING: A STATISTICAL ANALYSIS OF THE ENVIRONMENTAL IMPACTS OF A REGULATORY CHANGE IN THE ENERGY SECTOR

Fernanda Valente Márcio Laurini

ABSTRACT

We evaluate the efficiency of a regulatory change in sugarcane production (Law Nº11.241/2002) that gradually banish the pre-harvest burning in the sugarcane production sector in the São Paulo state in Brazil, analyzing the occurrences of fires detected by satellite. To estimate the spatio-temporal dynamics of fire occurrence we use a structural representation in a log-Gaussian Cox process, decomposing the intensity function in components of trend, seasonality, cycles, covariates and spatial effects. The results have provided evidence that the trend component reflects the consistent reduction in fire occurrences, suggesting the efficiency of the mitigation actions implemented in the São Paulo state, and indicating a permanent advance in the environmental sustainability of this form of renewable energy production.

Keywords: Sugarcane; Regulatory Changes; Spatio-temporal models; Log-Gaussian Cox Process

VALENTE, F.; LAURINI, M. Pre-harvest sugarcane burning: A statistical analysis of the environmental impacts of a regulatory change in the energy sector. **Cleaner Engineering and Technology**, v. 4, p. 100255, 2021.

2.1 Introduction

The sugarcane industry has an important role in the world, with the world's largest crop by product, and the three most important countries in sugarcane production are Brazil, India, and China (Food and Agriculture Organization of the United Nations, 2017). In Brazil, the state of São Paulo is the dominant sugarcane producer, representing 55% of the national production in 2017/2018 (Brazilian Sugarcane Industry Association, 2018). After 2003, with the adoption of flex-fuel vehicles in Brazil, the sugarcane production in the country has expanded sharply, increasing from 320 million tons in the harvest of 2002/2003 to 641 million tons in 2017/2018 (Brazilian Sugarcane Industry Association, 2018). Also, the production growth was driven by the increased international interest in alternative energy sources to mitigate global warming, promoting the use of ethanol to replace fossil fuels to reduce emissions of greenhouse gases (GHG). Although ethanol is considered as clean fuel, because it originates from sustainable sources, concerns on its sustainability has raised, mostly due to the way of the sugarcane is harvested, which can be done by manual labor with a necessary pre-harvest sugarcane straw burning, or mechanically, with or without pre-harvest burning. Previous studies have shown that burning of sugarcane residues is responsible for a large emission of pollutant gases in atmosphere with significant environmental effects, such as impacts on air quality (PARAISO; GOUVEIA,

2015), GHG emissions (RIBEIRO, 2008), and human health (CANÇADO et al., 2006), and also the degradation of soils and destruction or damage of high-biodiversity areas (e.g., Goldemberg, Coelho and Guardabassi (2008)).

Due to environmental and economic reasons, mitigation actions regarding the sugarcane pre-harvest burning are necessary. Based on this, the São Paulo state implemented a legislation (Law Nº11.241/2002) to gradually banish the burning practice by 2031. The law establishes two different calendars for the prohibition of burning, according to soil slope. The deadline for terrains with land slope less or equal 12% is 2021, whereas for areas with slope greater than 12% the deadline is 2031. However, to anticipate the end of this burning practice to 2017, a "Green Protocol" was established in 2007, setting 2014 as a target year for properties with land slope less or equal 12% and 2017 for non-mechanizable areas.

While these policies promote a new technology pattern in the Brazilian sugarcane production, it is still necessary to assess the effectiveness of these legal restrictions. In fact, there are evidence, based on remote sensing data monitoring, that the Green Protocol's expected goals has been fulfilled, with an increase of the green harvest in the sugarcane sector, especially on recently expanded sugarcane fields (AGUIAR et al., 2011). However, based on the fact that the regulatory changes in the São Paulo state was implemented gradually, the trend analysis may be an important tool to evaluate the efficiency of the Law Nº11.241/2002, which takes into consideration the prior knowledge of the producers about the deadline imposed by the law. Nonetheless, it is worth noting that it is necessary a more complex model to properly identify trends and changes in patterns over time of a space-time event, such as the possible fire occurrences caused by the sugarcane pre-harvest burning. To fulfill this task, an estimation of the permanent and periodic components in spatio-temporal models can be used, following the definition of Laurini (2019). This structure is used in the modelling of tornado occurrences in Valente and Laurini (2020), and wildfire events in Valente and Laurini (2021b). The main idea of the model is obtaining a more robust estimate of long-term trends in the intensity function of the spatial counting process of fire incidence, controlling for transient effects and possible covariates. The trend estimation allows verifying whether the implementation of Law Nº11.241/2002 led to permanent reductions in the intensity of fires in the São Paulo state in Brazil.

In this sense, we aim to contribute to the literature by assessing the efficiency of regulatory changes in the sugarcane production sector, by analyzing the process of change in the pattern of fire occurrences in the São Paulo state. Since the data which we use in this work are daily reports of fires occurrences, which contains information such as location and time of observed occurrences, one way to model fires occurrences is through spatio-temporal point process models. In particular, we used a dynamic representation of a log Gaussian Cox process (LGCP), where the intensity function is modeled through decomposition of components in trend, seasonality, cycles, covariates and spatial effects (LAURINI, 2019). We use two representations for the spatial random effects in the model. In the first we assume that spatial effects are time varying, based on an autoregressive functional structure, and the second representation is based on a non-separable spatio-temporal covariance function. The results indicate that the trend component reflects the consistent reduction in fire occurrences related to Sugarcane Burning Law (Law Nº11.241/2002).

2.2 Material and Methods

Our objective in this work is to analyze the existence of changes in long-term trends in the incidence of fires in the São Paulo state that may be related to the progressive application of Law Nº11.241/2002. For this, we need an appropriate statistical method to model space-time processes for counting events, in this case, the occurrence of fires that are normally associated with burning cane straw. In this case, traditional trend analysis methods are not appropriate, since the process is located in the spatial continuum, and it is a counting process, which does not allow the use of distributions associated with continuous variables, such as the Gaussian distribution normally used in trend analysis.

A common structure used to deal with point pattern data is the Poisson process. However, this structure is limited even if one assumes a non-homogeneous spatial intensity through some deterministic function. The limitations are related to the lack of possible sources of uncertainty and the fact that the Poisson process is conditionally independent. As an alternative, it is possible to allow the Poisson intensity to be a stochastic function, known as Cox process. In particular, the Log Gaussian Cox process (LGCP) (MØLLER; SYVERSVEEN; WAAGEPETERSEN, 1998) is a special case of the Cox process, where the log of the intensity function is given by a Gaussian random field. The LGCP framework is very useful for both theoretical and empirical motivations. In special, it allows to introduce more complex stochastic structures in the stochastic intensity function, controlling for general processes of spatial dependence, and also controlling for omitted variables with spatial dependence. In addition, it is possible to distinguish long-term changes and periodic components through a decomposition of the intensity function into trend, seasonality, cycle, covariates and spatial effects (LAURINI, 2019), which is very interesting in the context of climate-related issues. Below we give some details about the structure of this model.

From the point of view of inference procedures, the main difficulty associated with LGCP is the likelihood evaluation, which is analytically intractable due to the integral of the intensity function. One way to circumvent this problem is approximate the LGCP likelihood through the stochastic partial differential equation (SPDE) approach (SIMPSON et al., 2016), based on approximating the Gaussian random field (GRF) generated by a Matérn covariance function with a latent Gaussian Markov random field (GMRF) (LINDGREN; RUE; LINDSTRÖM, 2011), which can be estimated within the integrated nested Laplace approximations (let) framework (RUE; MARTINO; CHOPIN, 2009). One of the main advantages of SPDE approach is the fact that GMRF are defined by sparse matrices, which makes the process computationally effective. The SPDE approach has been widely used as a useful tool to deal with spatio-temporal models in different contexts, e.g., to model the wildfire occurrences in Catalonia, to assess the dynamics of fuel prices in Brazil (LAURINI, 2017), and to model the evolution of tornado occurrences in the United States (VALENTE; LAURINI, 2020) and fire occurrences in Australia (VALENTE; LAURINI, 2021b).

In order to provide a clearer idea of the method adopted herein, we fist present a brief description of the SPDE approach, considering a general LGCP. The results below are quite technical, and can be skipped by readers interested only in the empirical results. Let consider a

LGCP where the intensity function is defined by:

$$\begin{aligned} Y(s,t) &= Poisson(|e(s,t)|exp(\lambda(s,t)) \\ \log\lambda(s,t) &= z(s,t)\beta + \xi(s,t) \\ \xi(s,t) &= f(\omega(s,t)) \end{aligned} \tag{2.1}$$

where $Y(s,t) \equiv \{y(s,t)|(s,t) \in D \times T \in \mathbb{R}^2 \times \mathbb{R}\}$ are the observations of occurrences in a region s in time t, $e(s,t)$ is the exposure offset for the s region, $z(s,t)$ is a set of covariates continuously observed in the location s and time t, β is a set of parameters related to the covariates, $\xi(s,t)$ are the spatial random effects, and $\omega(s,t)$ is a spatially continuous Gaussian process, with zero-mean Gaussian distribution and the spatio-temporal covariance function defined as

$$Cov[\omega(s,t),\omega(s',t')] = \begin{cases} 0 & \text{if } t \neq t' \\ \sigma^2 C(h) & \text{if } t = t' \end{cases} \qquad \text{for } s \neq s'. \tag{2.2}$$

It follows that $Var(\omega(s,t)) = \sigma^2$, and we assume that the purely spatial correlation function $C(h)$ is a Matérn covariance function, which depends on the location s and s' only through the Euclidean distance, and is defined by the function

$$C(h) = \frac{2^{1-\nu}}{\Gamma(\nu)}(\kappa||h||)^{\nu}K_{\nu}(\kappa||h||) \tag{2.3}$$

where $h = ||s - s'||$ is the Euclidean distance between points s and s', $\kappa > 0$ is parameter for the spatial scale, $\nu > 0$ is a parameter controlling the smoothness of the process and K_{ν} is the modified Bessel function. With this structure, the marginal variance σ^2 of the Matérn process is given by:

$$\sigma^2 = \frac{\Gamma(\nu)}{4\pi\kappa^{2\nu}\tau^2\Gamma(\nu + \frac{d}{2})} \tag{2.4}$$

where τ is also a scale parameter and d is the spatial dimension. We follow the parameterization proposed by Lindgren, Rue and Lindström (2011), using $\log\tau$ and $\log\kappa$ in the covariance function. Conditional on a fixed ν, this representation depends on only two parameters:

$$\begin{aligned} \log\tau &= \frac{1}{2}\log\left(\frac{\Gamma(\nu)}{\Gamma(\alpha)(4\pi)^{d/2}}\right) - \log\sigma - \nu\log\rho \\ \log\kappa &= \frac{\log(8\nu)}{2} - \log\rho \end{aligned} \tag{2.5}$$

with $\rho = \frac{(8\nu)^{1/2}}{\kappa}$.

To approximate the LGCP likelihood we use the so-called SPDE approach (SIMPSON et al., 2016). This method is based on the equivalence of the $\omega(s,t)$ Matérn covariance function with the solution of a stochastic partial differential equation, and approximating the solution of this SPDE with a GMRF. The GF $\omega(s)$, assuming a Matérn covariance, is equivalent to the stationary solution to the following linear fractional SPDE (WHITTLE, 1954; LINDGREN; RUE; LINDSTRÖM, 2011).

$$(\kappa - \Delta)^{\alpha/2}\omega(s) = W(s), \quad s \in R^d, \quad \alpha = \nu + d/2, \quad \kappa > 0, \quad \nu > 0 \tag{2.6}$$

with $\Delta = \sum_{i=1}^{d} \frac{\partial^2}{\partial s_i^2}$ being a Laplacian operator and $W(s)$ is a spatial white noise. The method is based on a GMRF approximation of stochastic weak solution of SPDE (2.6), which is obtained by the use of Finite Elements Method (FEM) to solve the SPDE (LINDGREN; RUE; LINDSTRÖM, 2011) given by

$$\omega(s,t) \approx \tilde{\omega}(s,t) = \sum_{j=1}^{n} w_j \varphi_j(s,t) \tag{2.7}$$

where n is the number of vertices of the mesh, $\{w_j\}_{j=1}^{n}$ are solution weights following a Gaussian distribution and $\{\varphi_j\}_{j=1}^{n}$ are the basis functions localized in the nodes of the mesh, which was chosen to be piecewise linear on each triangle, that is equal to 1 at the jth vertex of the mesh and is 0 at all other vertices. See Simpson, Lindgren and Rue (2012) for a discussion of the choice of basis functions in spatial models. Therefore, it is possible to observe that the equation (2.7) is a link between the GF and GMRF, where $\{w_j\}$ has a Markovian structure (LINDGREN; RUE; LINDSTRÖM, 2011).

By replacing the GF $\omega(s,t)$ with the GMRF approximation $\tilde{\omega}(s,t)$ in equation (2.1), we have an approximate likelihood for the LGCP, which consists of $(n+n_t)T$ independent Poisson random variables, where n is the number of vertices in the mesh and n_t is the number of observed fires (SIMPSON et al., 2016). Under the GMRF structure, it is possible to estimate the model within the Bayesian framework using the Integrated Nested Laplace Approximation (INLA) framework. This methodology allows the use of deterministic approximations to perform the estimation of latent parameters and components in models with an additive structure. These methods are computationally efficient and are comparable to Markov Chain Monte Carlo methods with a very large number of samples. See Rue, Martino and Chopin (2009) for a detailed discussion on this method, and Krainski et al. (2018) for application in spatial and spatio-temporal point processes.

In order to analyze the process of change in the pattern of fires in the São Paulo state, we first consider a particular version of the Equation (2.1), with a dynamic version of a spatial LGCP where the intensity function is modelled as follows:

$$\begin{aligned}
Y(s,t) &= Poisson(|e(s,t)|exp(\lambda(s,t)) \\
\log \lambda(s,t) &= \mu_t + s_t + c_t + z(s,t)\beta + \xi(s,t) \\
\mu_t &= \mu_{t-1} + \eta_\mu \\
s_t &= s_{t-1} + s_{t-2} + \ldots + s_{t-m} + \eta_s \\
c_t &= \theta_1 c_{t-1} + \theta_2 c_{t-2} + \eta_c \\
\xi(s,t) &= \Theta\xi(s,t-1) + \omega(s,t), \quad 0 < \Theta < 1.
\end{aligned} \tag{2.8}$$

where $Y(s,t)$ is the number of occurrences in a region s in time t, $e(s,t)$ is the exposure offset for the region s, μ_t is the long term trend, s_t represents the seasonal components, c_t is a cycle component represented by a second-order autoregressive process with complex roots, $z(s,t)$ is a set of covariates observed in the location s and period t, η_μ, η_c and η_s are nonspatial independent innovations with $\eta_\mu \sim N(0, \sigma^2_{\eta_\mu})$, $\eta_c \sim N(0, \sigma^2_{\eta_c})$ and $\eta_s \sim N(0, \sigma^2_{\eta_s})$, and $\xi(s,t)$ are defined as spatio-temporal Gaussian fields that change in time with first order autoregressive processes, with Θ being a temporal correlation parameter, and $\xi(s,1) \sim N\left(0, \frac{\sigma^2}{1-\Theta^2}\right)$. The Gaussian

process $\omega(s,t)$ is continuously projected in space with a Matérn covariance, as previously defined. This structure was also used to model fire occurrences in Australia (VALENTE; LAURINI, 2021b).

It is worth noting that the proposed model is a separable space-time model, which is defined as a Matérn/SPDE model for the spatial domain and an first-order autoregressive model for the time domain. In this case, the precision matrix in the Gaussian random field is given by the Kronecker product between the spatial and temporal covariance matrices. In particular, the spatially correlated innovations $\omega(s,t)$ introduced in the equation (2.8) can be replaced by the GMRF $\hat{\omega}(s,t)$, where the temporally independent precision matrix $\mathbb{Q}_S$ can be obtained through the SPDE representation, and its dimension n is given the number of the vertices in the mesh. On the other hand, the T-dimensional precision matrix $\mathbb{Q}_T$ of the temporal first-order autoregressive process can be written as

$$\mathbb{Q}_T = \begin{pmatrix} 1/\sigma^2 & -\Theta/\sigma^2 & & & \\ -\Theta/\sigma^2 & (1+\Theta^2)/\sigma^2 & & & \\ & & \ddots & & \\ & & & (1+\Theta^2)/\sigma^2 & -\Theta/\sigma^2 \\ & & & -\Theta/\sigma^2 & (1+\Theta^2)/\sigma^2 \end{pmatrix} \tag{2.9}$$

and it follows that $\xi \sim N(0, \mathbb{Q}^{-1})$, where $\mathbb{Q} = \mathbb{Q}_T \otimes \mathbb{Q}_S$.

In the Section 2.4 we introduce an alternative formulation based on a non-separable spatio-temporal generalization of the Matérn covariance function.

In summary, our goal is to capture the intensity function of a counting process, using a Poisson distribution with stochastic intensity. In this representation, the intensity function is decomposed into components of trend, seasonality, cycle, plus the effect of covariates and the so-called spatial random effect, which captures the local effects not captured by the other components. Our main interest is in the trend of the process, which summarizes the permanent dynamics of the process, and from the analysis of this component we can verify if the application of Law Nº11.241/2002 led to a control of the number of fires in the São Paulo state, whose fundamental cause is the burning of cane straw. In order to adequately recover these permanent effects, we have incorporated controls for transient effects, in particular the seasonality in the number of fires related to the periods of cane harvest, and other possible transient effects but with dependence on time, which are captured by a component of cycle, which is modeled as a second-order autoregressive process. We also incorporate covariates in this analysis, to control for the impact of climate variables on the occurrence of fires (maximum temperature and accumulated rainfall in the period), as well as a fixed climate effect using the Köppen climate classification, and other effects directly related to the suitability of the land for sugarcane production, such as zoning limitations, soil classification and land slope, as discussed in the following section of the paper.

2.3 Results and Discussion

We use daily data of fire occurrences in the São Paulo state from MODIS Thermal Anomalies/Fires between January 2003 and December 2016 with confidence detection higher than 50%.

The dataset, provided by NASA[1], includes information on fire location and date, and also on the criteria for the fire classification and detection confidence. We use a quarterly aggregation of the daily observed data.

We include some covariates that could be important in the fire observations, since our dataset include fire occurrences of different causes, such as human sources (deliberately or accidentally), natural causes and also covariates that are directly linked to the potential of sugarcane production in the analyzed region. As explanatory variables we use Köppen Climate Classification system (ALVARES et al., 2013) to control climate effects and NASA MODIS Land Cover Classification (FRIEDL et al., 2002), as a control dummy for possible use of soil in agricultural activities. The classes of Köppen classification and MODIS land cover classification can be found in Additional information. We also include spatially continuous projections of maximum temperature and precipitation for each period as explanatory variables (LAURINI, 2019), based on Brazilian meteorological data provided by the National Institute of Meteorology (INMET)[2]. These covariates control for possible natural causes in the occurrence of fires, caused by the combination of high temperatures and periods of low precipitation.

Due to slope restrictions related to mechanical harvesting, we include a soil slope variable, expressed in degrees, from AMBDATA[3]. In our empirical analysis, the soil slope variable is set as a dummy, which is equal 0 in areas with slope less or equal 12% and 1 otherwise. In addition, we also include monthly global future price of sugar (U.S. cents per pound) provided by International Monetary Fund (IMF). Finally, we include as explanatory variable the Sugarcane Agroecological Zoning for the production of ethanol and sugar, provided by the Secretariat for the Environment of the São Paulo state jointly with the Secretariat of Agriculture of the São Paulo state. The zoning study evaluate the potential of the land for the production of sugarcane based on some characteristics as the physical, chemical and mineralogical characteristics of the soil, the climate risk and environmental regulations. It was established four areas with different potential of the land for the production of sugarcane: appropriate areas, appropriate areas with environmental limitations, appropriate areas with environmental restrictions and inappropriate areas (COSTA et al., 2018).

Note that covariates related to land use, zoning and climate are important to control the fact that we only observe the occurrence of fires using satellite data, and we do not know directly whether this burning is directly caused by the burning of sugarcane straw. although this is one of the fundamental causes of fires in the São Paulo state before the implementation of Law Nº11.241/2002. By incorporating these covariates, we are controlling for fires caused by climatic causes, such as high temperatures and drought periods, and also incorporating in the intensity function the location effects that would indicate the most appropriate soils for the cultivation of sugarcane, and also the restrictions linked to the very implementation of the law, especially the slope of the terrain.

To apply the inference procedures, we first define a triangulation mesh to use the SPDE approximation in the spatial domain. The mesh used in this study (see Fig. 2.9) has 117 triangles covering all the São Paulo state, and allows an approximation with sufficient precision

[1] Available at https://modis.gsfc.nasa.gov/data/dataprod/mod14.php.
[2] Available at http://www.inmet.gov.br.
[3] Available at http://www.dpi.inpe.br/Ambdata/index.php

for the LGCP. As exhaustively discussed in the literature (e.g., for a theoretical discussion see Lindgren, Rue and Lindström (2011) and for applied aspects Cameletti et al. (2013)), Bayesian inference of hierarchical spatio-temporal models face computational issues due to the big dense covariance matrices related to spatio-temporal datasets. As discussed by Simpson et al. (2016), the use of the continuous approximation for LGCP avoids the construction of very fine lattices over observation window in the process approximation, and thus allows us to estimate the parameters and random effects in a computationally efficient way. This important characteristic is desirable in spatio-temporal models, where the computational representation of the process is very intensive in memory. In this case, the choice of the number of vertices n in the triangulation was made not only considering our computational limitation but also the fact that the SPDE approach does not requires such fine grids to reach good approximations. For readers who are interested in the implementation of the SPDE approach, a detailed discussion of criteria for the choice of the mesh size in LGCP models can be found in Simpson et al. (2016). For the time index, we use a time mesh for the observed quarters in our sample, 56. With this space-time grid, we find to which polygon in time and space belongs each observed fire, defining the count process in this discretization.

As presented in Section 2.2, the model decomposes the intensity function of a stochastic intensity Poisson process in time series common components, covariate effects and spatial random effects. The time series components are given by an additive structure of trend, seasonality and cycle components (LAURINI, 2019). The trend is modeled as a first-order random walk, the seasonal component by the sum of random intercepts with restriction that the effects most sum to zero in each year and the cycle is represented by a second-order autoregressive process with possible complex roots, which allows to capture the periodic and stationary patterns. The spatial covariance function is given by the SPDE approximation to the spatially Matérn covariance (Equation 2.2). The parameters to be estimated are the those related to the covariates (β), the precisions (inverse of variance) of the trend ($1/\eta_\mu$), seasonality ($1/\eta_s$) and cycle ($1/\eta_c$) components. The second-order parameters of the cycle component are reparametrized as partial autocorrelations (PACF1 and PACF2), and the spatial covariance parameters are represented by $\log\tau$ and $\log\kappa$, assuming the smoothness parameter ν equal to 2, following the recommendation of Lindgren, Rue and Lindström (2011).

Regarding the identifiability of the model parameters, there is no significant discussion in the literature about the joint identifiability of the parameters in a model of structural time series decomposition and a spatial component continuously observed and time varying in the LGCP model used in this study. Thus, to impose the identification of this model, we impose individual restrictions on the latent factors of the model. A required constraint to the identifiability of the model parameters is to assume that the innovation terms to each component (trend, cycle, seasonality and spatial) are independent, which is imposed by the model in two ways: we assume independent priors for the parameters related to random effects, and also, we impose restrictions in the model estimation, imposing a zero-mean constraint for seasonality, cycle and spatial components, consistent with the expect identification constraints on these factors. We do not impose this restriction on the trend, but in this case the model does not include an intercept, which allows the identification of this component. Additionally, another way to verify the identifiability of the model parameters is through the precision matrix associated with the

estimations. In models estimated within the INLA framework, the identification problems lead to problems in the maximization of the posterior mode used in the Laplace approximation, and thus, in this case, a diagnosis for identification problems in the model parameters is the presence of singularities or quasi-singularities in the precision matrix of the GMRF representation.

The results of the estimation of the fixed effects (see Table 2.1) indicate a positive relation between maximum temperatures and the intensity of fire occurrences in São Paulo, and also a positive relation with soil slope, as expected, due to difficulties related to mechanical harvesting and in properties with high slope. In addition, although all regions classified as humid subtropical climate with dry winter and hot summer (Köppen Cwa), tropical climate without dry season (Köppen Af), humid subtropical oceanic climate, without dry season and hot summer (Köppen Cfa), humid subtropical climate with dry winter and temperate summer (Köppen Cwb), humid subtropical oceanic climate, without dry season and temperate summer (Köppen Cfb) and tropical climate with dry winter (Köppen Aw) show a positive relation with fire intensity, the results indicate that under humid subtropical climate with dry winter and hot summer, the intensity of fire occurrences is higher than other regions.

The negative relation between sugar price and the intensity of the fire occurrences can be related productivity effect. Indeed, given some aspects, the mechanical harvesting may be considered one of the causes of the decreasing in the sugarcane productivity given the difficulties to appropriately cut stems close to the ground, reducing the useful life of the sugarcane plantations, and also the significant amount of straw left in the field with the mechanical harvesting, increasing the incidence of pests and diseases in the sugarcane plantations (BACCARIN; OLIVEIRA; MARDEGAN, 2020). As a result, the small contribution of productivity effects for total growth in the sugarcane production in the state of São Paulo (CALDARELLI; GILIO, 2018) may be responsible of the increase in the final price of sugar. The dummy for cropland (Land Cover - Cropland) also indicates a positive impact on fire intensity, as expected.

In addition, the Sugarcane Agroecological Zoning presents a negative relation with fires intensity at inappropriate areas. In summary, the posterior distribution of the fixed effects is consistent with general fire patterns observed in the São Paulo state. Additionally, the results of the estimated random effects indicate a high precision for trend and seasonal components, whereas the precision for the components of cycle are relatively lower. In addition, a group autoregressive persistence equals to 0.70 suggests the high persistence of the spatial random effects.

In order to better understand the estimated components, we plot (see Fig. 2.1) the estimated trend, seasonality and cycle component (posterior mean of the estimated latent components along with 95% Bayesian credibility interval). The posterior mean of the estimated trend component reflects the reduction in the fire occurrences, which has been clearly intensified after 2010. In particular, this intense reduction in the fire occurrence after 2010 may be linked to public policies to encourage adherence to the "Green Protocol", which has established more responsibilities to each municipality to fulfill (or not) the goal of anticipating the decreasing of the pre-harvest burnings in the sugarcane production to crop year 2010/11, still considering the difficulties associated with slopes higher than 12% (AGUIAR et al., 2011). Therefore, the estimated long-term component suggests the efficiency of the sugarcane burning law (Law Nº11.241/2002), and the "Green Protocol", aligned with previous studies which have found a

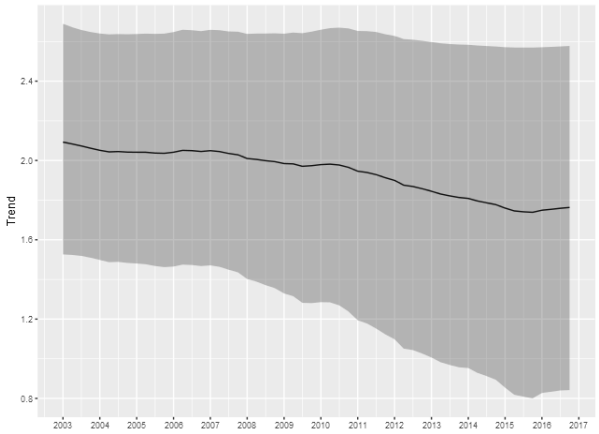
Table 2.1 – Estimated Parameters

	Mean	SD	0.025q	0.5q	0.975q	Mode
Fixed Effects						
Cwa (Köppen)	2.05	0.19	1.69	2.05	2.42	2.05
Af (Köppen)	0.27	0.18	-0.08	0.27	0.63	0.27
Cfa (Köppen)	1.20	0.17	0.86	1.20	1.54	1.20
Cwb (Köppen)	1.54	0.20	1.16	1.54	1.93	1.54
Cfb (Köppen)	0.85	0.18	0.50	0.85	1.19	0.85
Aw (Köppen)	1.53	0.19	1.17	1.53	1.90	1.53
Approp. zone with limitations	0.22	0.07	0.09	0.22	0.35	0.22
Approp. zone with restrictions	0.13	0.07	-0.01	0.13	0.28	0.13
Inappropriate zone	-0.09	0.07	-0.24	-0.09	0.05	-0.09
Land Cover - Cropland	0.16	0.08	-0.00	0.16	0.33	0.16
Sugar Price	-0.01	0.02	-0.05	-0.01	0.02	-0.01
Soil Slope (dummy)	0.27	0.08	0.11	0.27	0.43	0.27
Max. Temperature	0.02	0.00	0.02	0.02	0.03	0.02
Rainfall	-0.00	0.00	-0.00	-0.00	0.00	-0.00
Random Effects						
Precision for Trend	2379.24	8138.08	60.66	722.71	14788.14	141.72
Precision for Seas.	26933.38	23395.89	3237.08	20538.46	88612.96	9204.33
Precision for Cycle	7.57	2.41	3.62	7.34	12.96	6.85
PACF1 for Cycle	0.34	0.16	0.01	0.35	0.65	0.35
PACF2 for Cycle	-0.15	0.16	-0.45	-0.16	0.18	-0.17
Log τ	-2.32	0.04	-2.40	-2.32	-2.23	-2.33
Log κ	1.33	0.03	1.26	1.33	1.39	1.33
Group Θ	0.70	0.01	0.67	0.70	0.73	0.71
waic	103478.4					

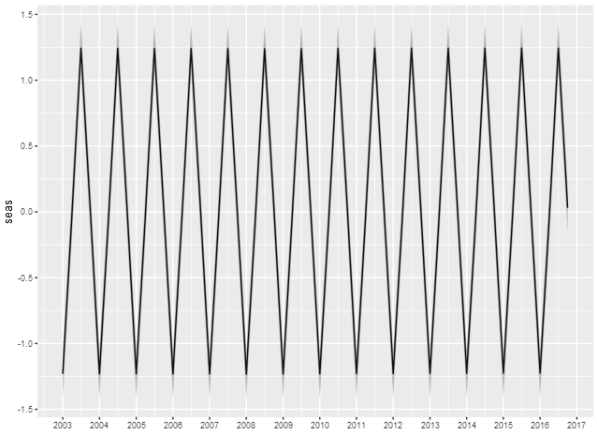
significant reduction in the burned sugarcane area of the total harvested area (RUDORFF et al., 2010), and a reduction of annual gas emissions in areas of mechanically-harvested sugarcane (ARRAES et al., 2010), both based on remote sensing data.

Also, it is important to highlight that 2009 economic crisis may have exerted some effects in decreasing the number of fire occurrences, i.e., from 2008/2009 to 2011/2012 was a period of great caution and instability, which caused reduction in the international investment in sugar-alcohol sector, thus affecting the sugarcane production (CALDARELLI; GILIO, 2018). After 2015, the estimated trend component becomes more stable, with a slight increase in the number of fire occurrences in the end of the analyzed sample. In addition, the seasonal component exhibits a very stable behavior over time, as expected. The cycle component corresponds to an AR(2) model with complex roots: 0.392 and -0.150, which is consistent with a periodic (cyclic) pattern. Plus, this result corresponds to a cycle period of 6.04 quarters.

Through the estimated spatial random effects (see Fig. 2.2 and 2.3) it is possible to observe significant variation in the temporal evolution of spatial random effects while the intensity of the fires fitted by the model (see Fig. 2.4 and Fig. 2.5; in all figures columns indicate the quarter whereas rows indicate the years) suggests that the model has a good fit. In order to facilitate the visualization, the figures were separated in two blocks, the first being from 2003 to 2009 and



(a) Trend



(b) Seasonal

(c) Cycle

Figure 2.1 – Trend, Seasonal and Cycle decomposition of fire occurrences in São Paulo

the second being from 2010 to 2016.

2.4 Non-separable spatio-temporal model

The spatio-temporal model proposed in Section 2 is based on a separability structure between spatial and temporal effects, using a Kronecker product structure to obtain the spatio-temporal representation. Although the form used is quite flexible, an alternative form can be formulated using a non-separable structure for the interaction between spatial and temporal effects, allowing for a more complex dependency structure. As discussed in Bakka et al. (2020), often the choice of a separable structure is given only by computational limitations, being more appropriate to use a non-separable structure. As an additional robustness analysis for our model, we estimate an alternative form of the model using the non-separable representation for the spatiotemporal random effect using the generalization of Matérn's spatiotemporal covariance matrix proposed in Bakka et al. (2020). In this representation the random field representing the structure of space-time random effects is given by a diffusion-based extension of the Matérn field (DEMF). The random field $u(s,t)$ is given by the expression:

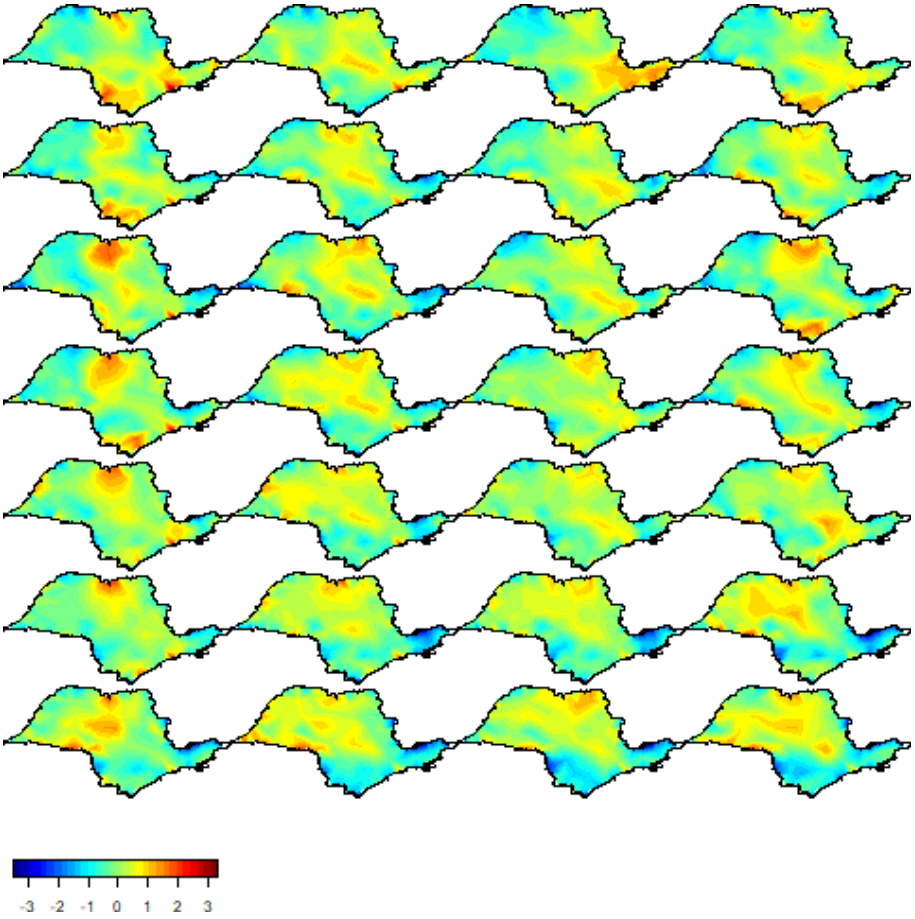


Figure 2.2 – Spatial Random Effects 2003 - 2009. **Note:** Columns indicate quarter whereas rows indicate the year, from 2003 to 2009.

$$\left(\gamma_t \frac{d}{dt} + L^{\alpha_s/2}\right)^{\alpha_t} u(s,t) = \mathcal{E}_Q(s,t) \tag{2.10}$$

with $L = \gamma_s^2 - \Delta$, $\mathcal{E}_Q(s,t)$ is Gaussian noise that is white in time but correlated, with precision operator $Q(\gamma_s, \gamma_\epsilon, \alpha_\epsilon) = \gamma_\epsilon^2 L^{\alpha_\epsilon}$, with $(\gamma_t, \gamma_s, \gamma_e)$ fixed scaling parameters, and $(\alpha_t, \alpha_s, \alpha_e)$ parameters in the model.

Defining $\alpha = \alpha_e + \alpha_s(\alpha_t - 1/2)$., and assuming that $\alpha_t, \alpha_s, \alpha_e$ satisfy $\alpha > 1$ the solution $u(s,t)$ has marginal spatial covariance function given by:

$$C(u(t,s_1), u(t,s_2)) = \frac{\sigma^2}{\Gamma(\nu_s)2^{\nu-1}} (\gamma_s \|s_1 - s_2\|)^{\nu_s} K_{\nu_s}(\gamma_s \|s_1 - s_2\|), \tag{2.11}$$

where $\nu_s = \alpha - 1$ and

$$\sigma^2 = \frac{\Gamma(\alpha_t - 1/2)\Gamma(\alpha - 1)}{\Gamma(\alpha_t)\Gamma(\alpha) 8\pi^{3/2} \gamma_e^2 \gamma_t \gamma_s^{2(\alpha-1)}}. \tag{2.12}$$

We estimate a non-separable version of the LGCP model, replacing the separable structure $\xi(s,t)$ of the original model with the random field $u(s,t)$ defined above, having the representation now given by:

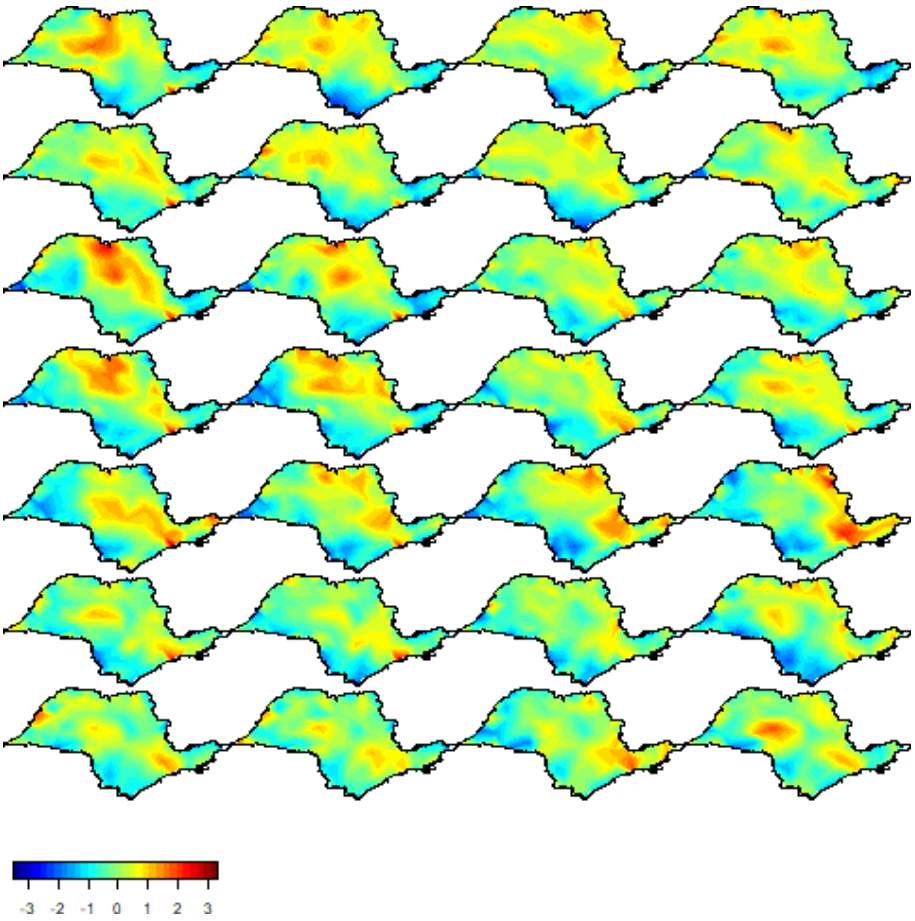


Figure 2.3 – Spatial Random Effects 2010 - 2016. **Note:** Columns indicate quarter whereas rows indicate the year, from 2010 to 2016.

$$
\begin{aligned}
Y(s,t) &= Poisson(|e(s,t)|exp(\lambda(s,t)) \\
\log\lambda(s,t) &= \mu_t + s_t + c_t + z(s,t)\beta + u(s,t) \\
\mu_t &= \mu_{t-1} + \eta_\mu \\
s_t &= s_{t-1} + s_{t-2} + \ldots + s_{t-m} + \eta_s \\
c_t &= \theta_1 c_{t-1} + \theta_2 c_{t-2} + \eta_c
\end{aligned}
\tag{2.13}
$$

(2.14)

Following the discussion in Bakka et al. (2020), we assume that the scale parameters $(\gamma_t, \gamma_s, \gamma_e)$ are given by (1,1,5), and we use the reparametrization of the parameters defining σ, r_s, r_t as:

$$c_1 = \frac{\Gamma(\alpha_t - 1/2)\Gamma(\alpha - 1)}{\Gamma(\alpha_t)\Gamma(\alpha)4\sqrt{\pi}} \tag{2.15}$$

$$\sigma = \gamma_\epsilon^{-1} c_1^{1/2} \gamma_t^{-1/2} \gamma_s^{-(\alpha-1)} \tag{2.16}$$

$$r_s = \gamma_s^{-1}\sqrt{8\nu_s} \tag{2.17}$$

$$r_t = \gamma_t\sqrt{8(\alpha_t - 1/2)}\gamma_s^{-\alpha_s}, \tag{2.18}$$

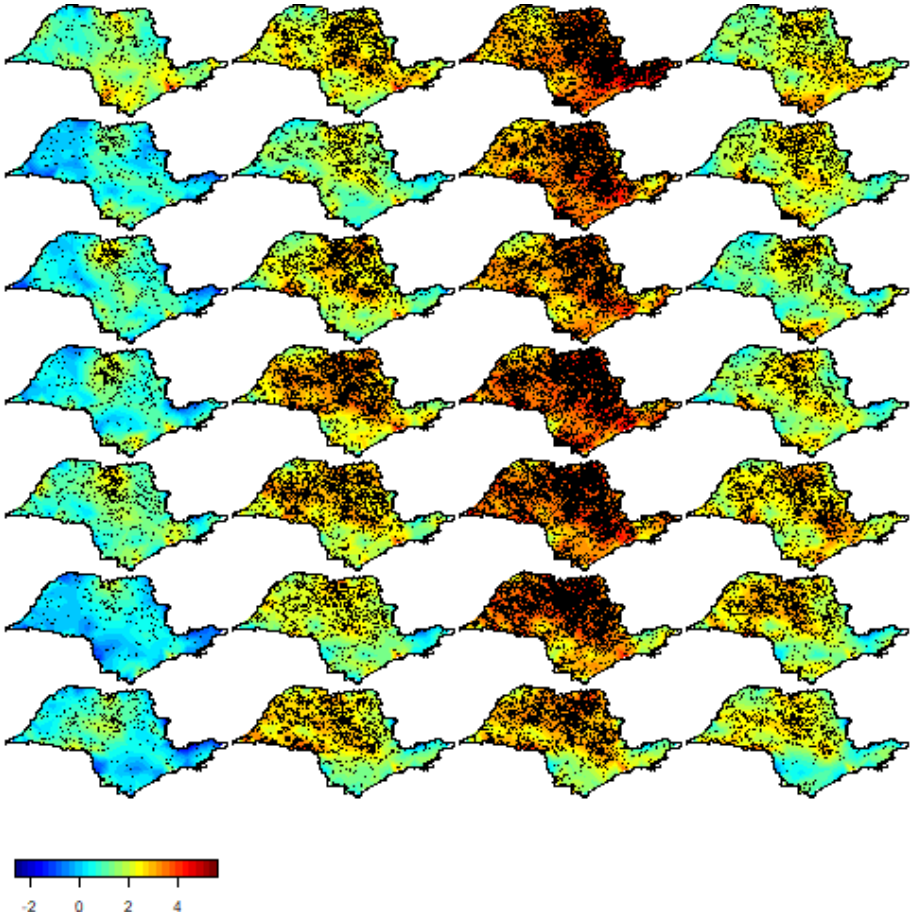


Figure 2.4 – Fitted Intensity and Fires 2003 - 2009. **Note:** Columns indicate quarter whereas rows indicate the year, from 2010 to 2016. Black dots indicate observed fires in each period.

and the model is estimated using as parameters $\log\sigma$, $\log r_s$, $\log r_t$. The model is estimated using the same mesh structure and the prior structure of the separable model, again using the INLA approximations to perform the Bayesian inference procedures. The estimated posterior distribution of parameters in the non-separable spatio-temporal LGCP model is presented in the Table 2.2, the estimated latent components of trend, seasonal and cycle components in the Figure 2.6 and the non-separable spatio-temporal random effect is presented in the Figures 2.7 and 2.8.

The results of the non-separable model indicate a more relevant variation in relation to fixed effects related to land type and use, and similar effects for climate parameters, suggesting that the non-separable structure indicates an interaction behavior with these covariates. In this specification the spatiotemporal random effect component also shows a wider range of values compared to that estimated by the separable model, as can be seen in Figures 2.7 and 2.8. We can also observe that the non-separable model has the best fit for the observed data (see Table 2.3 in the Appendix) by comparing models using WAIC (WATANABE, 2013), indicating a possible fit gain with the non-separable structure. However, when observing the distribution obtained for the latent components (Figure 2.6), we can observe that the qualitative patterns are similar to those obtained by the model with the separable structure (Figure 2.1), and so the general interpretation of the trend, seasonality and cycle components are similar between the

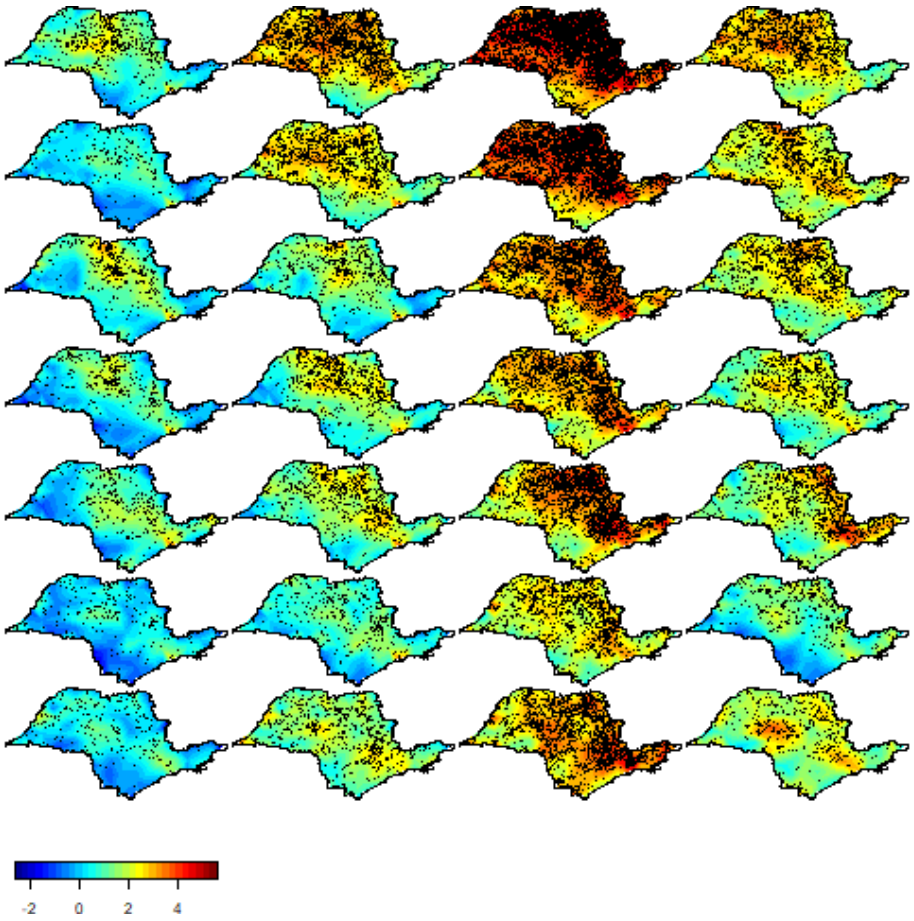


Figure 2.5 – Fitted Intensity and Fires 2010 - 2016. **Note:** Columns indicate quarter whereas rows indicate the year, from 2010 to 2016. Black dots indicate observed fires in each period.

two models, also given support to the reduction in the posterior mean of the estimated trend component, and indicating the success of Law Nº11.241/2002 in the control of the number of observed fire occurrences related to the burning of sugarcane straw.

2.5 Conclusion

Brazil is a major sugarcane producer, and the São Paulo state represents more than half of national production. Nevertheless, the adverse effects of the large-scale sugarcane production have led to environmental policies to reduce the problems associated to pre-harvest sugarcane burning. Due to environmental and economic reasons the São Paulo state implemented a legislation (Law Nº11.241/2002) to banish the burning practice gradually by 2031, following a calendar that takes into account the terrain slope. In 2007 the so-called "Green Protocol" anticipates the end of pre-harvest burning, setting 2014 as a target year for properties with slope less or equal 12% and 2017 for non-mechanizable areas.

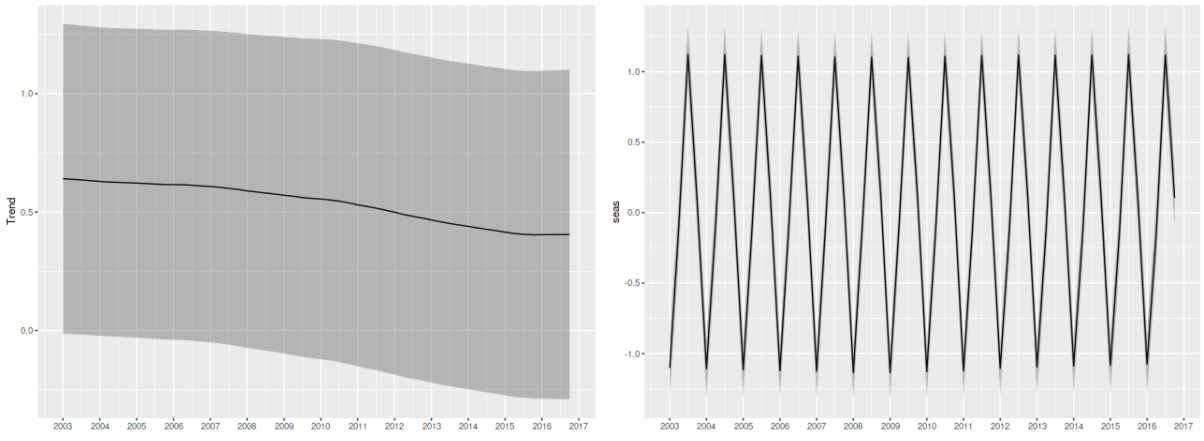
The objective of this work is to analyze the impact of the regulatory change given by the ban on burning sugarcane straw in the state of São Paulo, introduced by Law Nº11.241/2002. Our approach for this analysis was through a spatio-temporal decomposition of the occurrence of fire outbreaks, using satellite measurements, using a model based on the Log Gaussian Cox

Table 2.2 – Estimated Parameters - Non-separable spatio-temporal LGCP model

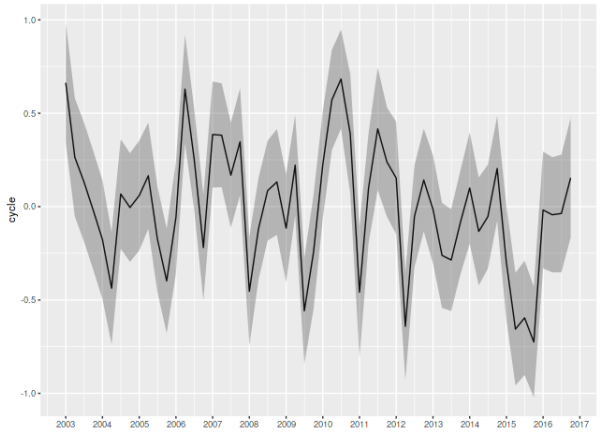
	Mean	SD	0.025q	0.5q	0.975q	Mode
Fixed Effects						
Cwa (Köppen)	3.43	0.28	2.89	3.42	3.99	3.41
Am (Köppen)	2.83	0.40	2.04	2.83	3.63	2.83
Af (Köppen)	2.66	0.28	2.12	2.66	3.24	2.65
Cfa (Köppen)	3.17	0.27	2.64	3.16	3.73	3.15
Cwb (Köppen)	3.03	0.28	2.49	3.03	3.61	3.01
Cfb (Köppen)	3.17	0.27	2.64	3.16	3.73	3.15
Aw (Köppen)	3.37	0.28	2.83	3.36	3.93	3.35
Approp. zone with limitations	0.15	0.03	0.09	0.15	0.21	0.15
Approp. zone with restrictions	0.24	0.03	0.17	0.24	0.30	0.24
Inappropriate zone	0.10	0.04	0.01	0.10	0.18	0.10
Land Cover - Cropland	0.27	0.02	0.22	0.27	0.33	0.27
Sugar Price	-0.01	0.01	-0.03	-0.01	0.01	-0.01
Soil Slope (dummy)	0.06	0.04	-0.03	0.06	0.16	0.06
Max. Temperature	0.02	0.00	0.01	0.02	0.02	0.02
Rainfall	0.00	0.00	0.00	0.00	0.00	0.00
Random Effects						
$\log \sigma$	2.20	0.02	2.15	2.20	2.25	2.20
$\log r_s$	-0.65	0.01	-0.67	-0.65	-0.63	-0.65
$\log r_t$	-0.10	0.00	-0.12	-0.10	-0.08	-0.10
Precision for Trend	2300.48	620.64	1260.58	2246.76	3674.66	2142.17
Precision for Seas.	1217.83	186.32	837.52	1229.49	1546.11	1282.09
Precision for Cycle	7.05	2.07	3.28	7.01	11.11	6.87
PACF1 for Cycle	0.18	0.07	0.03	0.18	0.32	0.18
PACF2 for Cycle	-0.13	0.15	-0.45	-0.12	0.13	-0.07
waic	103294.84					

Process structure, and including as controls related to possible land use, climate, zoning, sugar price and soil slope as covariates. The analysis of the impact of the Law No. 11.241/2002 on the intensity of fires is not trivial, since the regulation was introduced gradually, and additionally there was a self-regulation of the producers' market in this sector, leading to an acceleration of the implementation of the law. In this sense, this regulation has a very different characteristic from traditional forms of regulation that clearly delimit a period before and after, which prevents the use of traditional forms of causal impact analysis as differences-in-differences methods or control and treatment groups.

Since there is no simple temporal or spatial delimitation of the validity of the law, to overcome the problem, we build a statistical model that allowed us to analyze the general trends in the intensity of fire occurrences, controlling for the covariates related to the possible use of the soil in sugarcane planting before and after the implementation of the law. To perform inference procedures we adopt a dynamic representation of Log-Gaussian Cox process, where the intensity function is modelled as a sum of components of trend, seasonality, cycles, covariates and time varying spatial effects. We adopt the SPDE approach to approximate the initial Gaussian Field to a Gaussian Markov Random Field, which is defined by sparse matrices and fits naturally within Bayesian hierarchical modelling framework, and allows to use Integrated Nested Laplace



(a) Trend (b) Seasonal



(c) Cycle

Figure 2.6 – Trend, Seasonal and Cycle decomposition of fire occurrences in São Paulo - Non-separable model

Approximations approach to perform the inference procedures.

By controlling for the covariates of land use in sugarcane planting, and also for climatic variables, which are mostly related to natural-induce fires, generated by the combination of high temperatures and low rainfall, we can analyze the impact general of the regulation by the trend estimated by the model. This component captures the general behavior of the intensity of fire outbreaks not controlled by the climate and by the covariates that control the areas affected by the regulation introduced in Law Nº11.241/2002.

Thus, our main focus was to verify the general behavior of the trend that captures the persistent patterns in the fire outbreaks, which would be linked to possible burning of remaining sugarcane straw. As we do not have a clear delimitation of treatment and control effect in time and space, the results of our analysis do not have a strict causal interpretation, but can be seen as an observational/descriptive analysis using the possible controls for the regions under influence from the law (producing regions or with potential for the production of sugar cane with mechanized harvesting) and also for the other important source of fire occurrence, which are the effects of a combination of dry weather and high temperatures. The results obtained indicate that there was no increase in the estimated trend for fire occurrence intensity, and a reduction

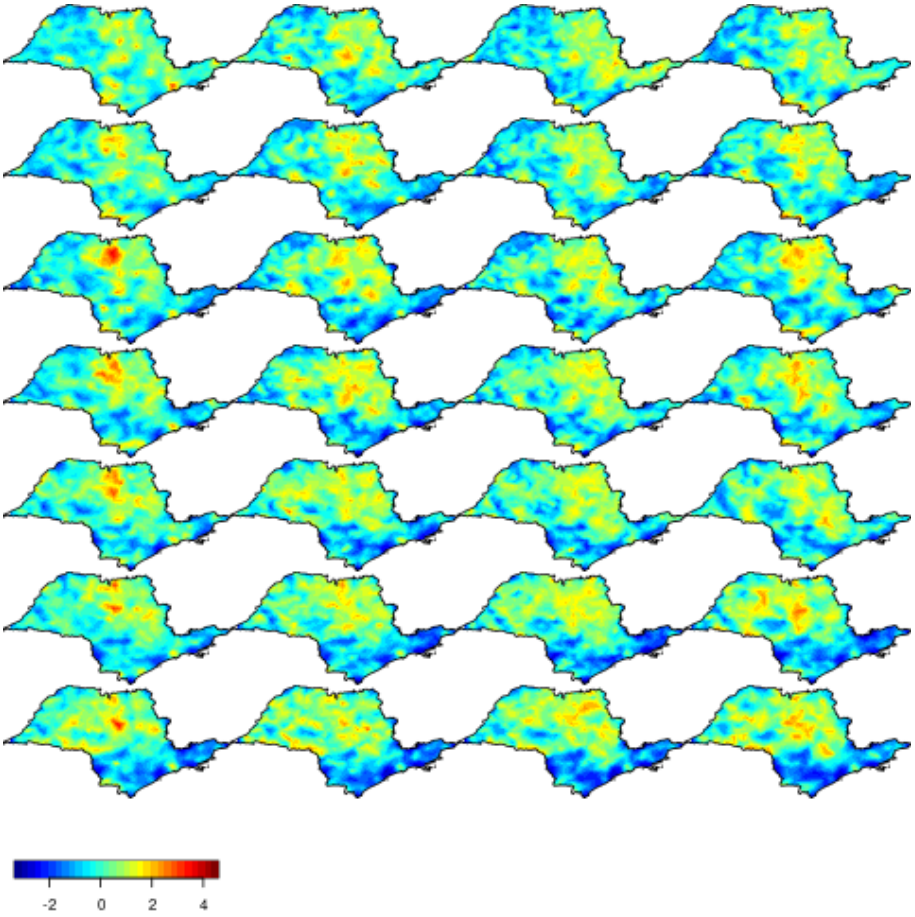


Figure 2.7 – Spatial Random Effects 2003 - 2009 - Non-separable model. **Note:** Columns indicate quarter whereas rows indicate the year, from 2003 to 2009.

if we observe the point estimate for the trend. Note that in the period under analysis there was a large increase in sugarcane planted area in the state of São Paulo, and several periods combining abnormal temperatures and drought. Thus, our empirical results are consistent with a positive impact of Nº11.241/2002 on the burning of sugarcane, reducing the environmental and ecological impact of this sector.

2.6 Appendix

2.6.1 Köppen Climate Classification for Brazil

Köppen Climate Classification for Brazil (ALVARES et al., 2013):

1. Cwa: (C) Humid subtropical (w) With dry winter (a) and hot summer
2. Am: (A) Tropical (m) monsoon
3. Af: (A) Tropical (f) without dry season
4. Cfa: (C) Humid subtropical (f) Oceanic climate, without dry season (a) and hot summer
5. Cwb: (C) Humid subtropical (w) With dry winter (b) and temperate summer
6. Csb: (C) Humid subtropical (s) With dry summer (b) and temperate summer

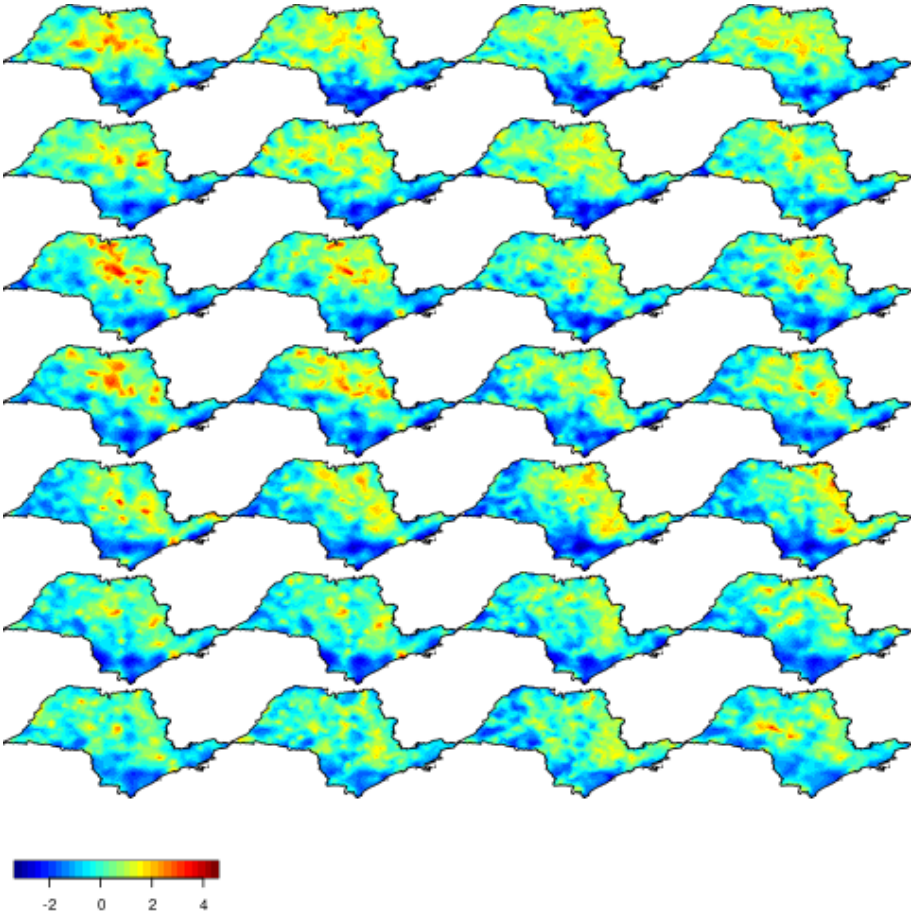


Figure 2.8 – Spatial Random Effects 2010 - 2016 - Non-separable model. **Note:** Columns indicate quarter whereas rows indicate the year, from 2010 to 2016.

7. Csa: (C) Humid subtropical (s) With dry summer (a) and hot summer
8. Cfb: (C) Humid subtropical (f) Oceanic climate, without dry season (b) and temperate summer
9. BSh: (B) Dry (S) Semi-arid (h) low latitude and altitude
10. As: (A) Tropical (s) with dry summer
11. Cwc: (C) Humid subtropical (w) With dry winter (c) short and cool summer
12. Aw: (A) Tropical (w) with dry winter

MODIS Land Cover Classification (FRIEDL et al., 2002):

1. Evergreen Needleleaf Forests: dominated by evergreen conifer trees (canopy >2m). Tree cover >60%.
2. Evergreen Broadleaf Forests: dominated by evergreen broadleaf and palmate trees (canopy >2m). Tree cover >60%.
3. Deciduous Needleleaf Forests: dominated by deciduous needleleaf (larch) trees (canopy >2m). Tree cover >60%.

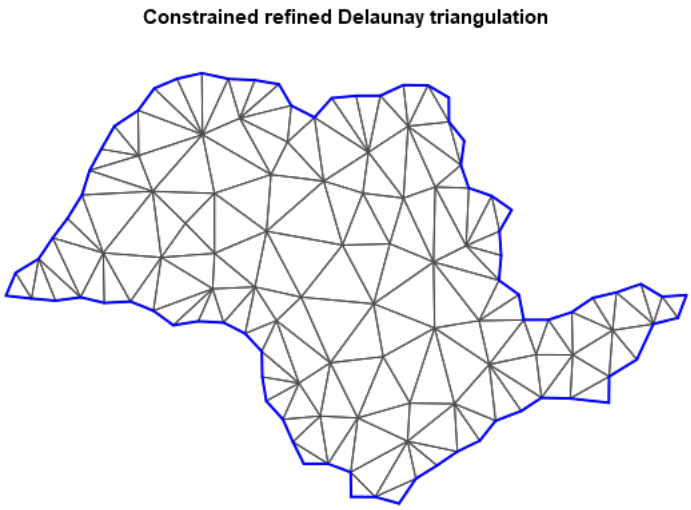


Figure 2.9 – Triangulated Mesh

4. Deciduous Broadleaf Forests: dominated by deciduous broadleaf trees (canopy >2m). Tree cover >60%.

5. Mixed Forests: dominated by neither deciduous nor evergreen (40-60% of each) tree type (canopy >2m).Tree cover >606 Closed Shrublands: dominated by woody perennials (1-2m height) >60% cover.

7. Open Shrublands: dominated by woody perennials (1-2m height) 10-60% cover.

8. Woody Savannas: tree cover 30-60% (canopy >2m). 9 fb13 Savannas: tree cover 10-30% (canopy >2m).

9. Grasslands: dominated by herbaceous annuals (<2m).

10. Permanent Wetlands: permanently inundated lands with 30-60% water cover and >10% vegetated cover.

11. Croplands: at least 60% of area is cultivated cropland.

12. Urban and Built-up Lands: at least 30% impervious surface area including building materials, asphalt and vehicles.

13. Cropland/Natural Vegetation Mosaics: mosaics of small-scale cultivation 40-60% with natural tree, shrub,or herbaceous vegetation.

14. Permanent Snow and Ice: at least 60% of area is covered by snow and ice for at least 10 months of the year.

15. Barren: at least 60% of area is non-vegetated barren (sand, rock, soil) areas with less than 10% vegetation.

16. Water Bodies: at least 60% of area is covered by permanent water bodies.

2.6.2 Model's comparison using WAIC

The Table 2.3 reports the widely applicable information criterion (WAIC), also known as Watanabe–Akaike information criterion (WATANABE, 2013), for several specifications of the models used in our work, imposing restrictions on the components included and also on the specification of spatial random effects used in the article. The two best specifications by WAIC are the model with covariates, all components of trend, seasonality and cycle, and time-varying spatial random effects, whose results are reported in Table 2.1, with a WAIC value of 103478.4, and the model with non-separable random effects, which also include all covariates and latent components of trend, seasonality, and cycle (Table 2.2), with a WAIC value of 103294.8. For this criterion, the best specification is given by the model with the non-separable structure of space-time random effects.

Table 2.3 – Model's comparison using WAIC

Model	WAIC
Model 1: including all the latent components (trend, seasonality, cycle), covariates, and time varying spatial effect	103478.4
Model 2: including all the latent components (trend, seasonality, cycle), covariates, and fixed spatial effect	146354.3
Model 3: including all the latent components (trend seasonality, cycle) and covariates	182148.4
Model 4: including the trend and seasonality components, and covariates	182149.0
Model 5: including trend component and covariates	182150.2
Model 6: including fixed intercept and covariates	235982.2
Model 7: including all the latent components (trend, seasonality, cycle) and time varying spatial effect	104065.4
Model 8: including all the latent components and spatial effect	146437.9
Model 9: including all the latent components (trend, seasonality, cycle), covariates, and non-separable random effect	103294.8

3 SPATIO-TEMPORAL ANALYSIS OF FIRE OCCURRENCE IN AUSTRALIA

Fernanda Valente Márcio Laurini

ABSTRACT

Fire is one of the most notorious hazards in Australia, with important economic impacts and damage to ecosystems. There is a concern of worsening fire conditions under climate variability, but there is little understanding of the variability in fire occurrence related to climate patterns. We present a statistical decomposition for spatio-temporal analysis of changes in fire occurrence in Australia and its association with climate factors. We found evidence of variability in the trend results for fire occurrence, and also some evidence that this variation is related to climate patterns. Our approach has applicability to other climate-related issues, providing a useful tool to identify possible changes in the intensity of occurrence over time, capturing long-term changes, and also seasonal and cyclical effects.

Keywords: Australia; Climate Variability; Fire occurrence ; Spatio-temporal models; Trend-Cycle decomposition

VALENTE, Fernanda; LAURINI, Márcio. Spatio-temporal analysis of fire occurrence in Australia. **Stochastic Environmental Research and Risk Assessment**, v. 35, n. 9, p. 1759-1770, 2021.

3.1 Introduction

There are evidence that climate and weather are important drivers of wildfire events, among different regions of the world, such as North America (e.g., Gillett et al. (2004), Westerling et al. (2006), Chen (2007), Goff, Flannigan and Bergeron (2009), Wotton, Nock and Flannigan (2010), Gedalof (2011)), Europe (e.g., Reinhard, Rebetez and Schlaepfer (2005), Lozano et al. (2017)), as well as in Oceania (e.g., Williams, Karoly and Tapper (2001), Pitman, Narisma and McAneney (2007), Clarke, Lucas and Smith (2013)). In the most systems, it is climate that controls the amount of fuel available to burning, and also determines the flammability of the available fuel and the continuity of the fire. Anthropogenic factors may also exhibit influences on fire, directly by starting and managing fires or indirectly through anthropologically driven climate changes (ALDERSLEY; MURRAY; CORNELL, 2011). In Australian ecosystems, fire plays an important role, influencing and determining the vegetation, due to factors such as floral composition, topography, and climate (RAHMAN et al., 2018). In the savannahs of northern Australia, intense fires dominate whereas massive fires in the arid zone occur after periods with above-average rainfall, with relatively less frequency. On the other hand, in the temperate forests of the south, large and intense fires occur, but is less extensive and also less regular (MORITZ et al., 2014).

The relationship between fire occurrence and the climate factors in Australia has been explored in the literature (VERDON; KIEM; FRANKS, 2004; RUSSELL-SMITH et al., 2007; MARIANI et al., 2016; DOWDY, 2018). In particular, the most important Australian climate drivers are the El Niño Southern Oscillation (ENSO), Indian Ocean Dipole (IOD) and Southern Annular Mode (SAM), causing spatio-temporal variations of temperature and rainfall (HENDON; THOMPSON; WHEELER, 2007; RISBEY et al., 2009), and therefore affecting the Australian fire behaviour. Additionally, some studies have identified some trends in the variables underlying fire indices. It has been observed an increase in the temperature extremes in Australia, with a particular increase in the number of record warm days while the number of record cold days has decreased (ALEXANDER et al., 2007). Some studies have also found that Australian rainfall patterns have changed, with a significant decrease in rainfall in the southwest of Australia and a significant increase in the proportion of total precipitation from extreme events in eastern Australia (TIMBAL; ARBLASTER; POWER, 2006; GALLANT; HENNESSY; RISBEY, 2007).

As a consequence of climate patterns, current and potential future changes in fire activity might pose threats to ecosystems and human health (ABATZOGLOU; WILLIAMS, 2016), and understanding the patterns of fire occurrence is important to avoid loss and facilitate management decisions. The fire weather index is a common methodology used to assess the fire danger and to explore the effects of climatic indices in the Australian fire danger. In particular, several studies have adopted the McArthur Forest Fire Danger Index (FFDI) to find evidences of the variability of fire weather in Australia, and also to assess the linkages between climate drivers and FFDI (VERDON; KIEM; FRANKS, 2004; CLARKE; LUCAS; SMITH, 2013; HARRIS; LUCAS, 2019). However, there is an important drawback related to the fire weather index methods. Usually these indexes are calculated based on weather monitoring station data, which may limit the analysis due to the fact that this kind of data may not be ideal for understanding the aspects of the spatial variability (DOWDY, 2018). In order to be complementary to previous studies (DOWDY, 2018) have proposed to assess the long-term variations in fire weather conditions based on gridded data, and have found that changes in fire weather conditions in southern Australia are related to anthropogenic climate change. However, while there is an increased concern of worsening fire conditions under climate change and variability, there is still little understanding of the relationship between the spatio-temporal changes of Australian fire occurrence and climate factors.

In order to contribute to this literature, we propose an alternative way to verify the existence of changes in the patterns of the fire activity, through the estimation of long-term and periodic components, using statistical tools to decompose the observed data into trend, seasonal, and cycle components. In addition, since fire occurrence can be associated with their spatial coordinates and temporal instant, to take into account the spatial heterogeneity of climate effects, we propose to combine elements of structural time series decomposition with spatio-temporal models with continuous spatial random effects, which can be thought as a process of decomposing geostatistical time series into a sum of persistent and mean-reverting components (LAURINI, 2019; VALENTE; LAURINI, 2020).

Therefore, in this study we will analyze the variability in the patterns of the fire occurrence in Australia within spatio-temporal point process framework, through a structural decomposi-

tion (e.g., Harvey (1990)) in spatio-temporal point pattern data. In particular, we extent the trend-cycle decomposition in spatio-temporal models to spatio-temporal point pattern data, by proposing to use a dynamic representation of a Log Gaussian Cox process (LGCP) where the intensity function is modeled through the decomposition of components into trend, seasonality, cycles, covariates and spatial effects (LAURINI, 2019; VALENTE; LAURINI, 2020). This is a useful formulation to identify possible changes in the intensity of occurrence over time, being capable to capture seasonal and cyclical effects, and to identify long-term changes in the fire events, that may be associated with climate variability.

We present here, the results of analyzing data for fires in Australia, from 2003 to 2019, using two different model specifications, with and without covariate effects, in order to assess the relationship of the variability in the patterns of fire events and climate factors. In addition to the above, we were also interested in assessing the possible variability in the maximum temperature and rainfall in Australia, evidencing the relationship between the changes in patterns of fire events and the seasonal, internannual and longer-term climate variability. In summary, results indicate an increase in the trend component of the fire occurrence, when the covariate effects are not included. On the other hand, when we included explanatory variables to control the main fixed effect related to climate patterns, the trend component remains relatively stable, which may suggests that the variability in the fire occurrence is attributable at least in part to climate factors. In addition, the results also give support to the increase in the trend component of the observed maximum temperature series.

3.2 Materials and Methods

3.2.1 Data

We used daily data of fire occurrence in Australia from MODIS (Moderate Resolution Imaging Spectroradiometer) Thermal Anomalies/Fires product between January 2003 and December 2019 (GIGLIO; JUSTICE, 2015). The dataset, provided by NASA, includes information like fire occurrences (day/night), fire location (geographic coordinates), the criteria for the fire detection which are based on the apparent temperature of the fire pixel and the difference between the fire pixel and its background temperature, the detection confidence value which ranges from 0% and 100%, and other layers describing fire pixel attributes (GIGLIO; JUSTICE, 2015). In addition, in order to facilitate the visualization of the results, we used a quarterly aggregation of the daily data. To illustrate, we plot the number of fire occurrences in Australia by quarter (see Figure 4.1) and the spatial distribution by quarter (see Figure 3.2). In general, fires are observed in all quarters in Australia, although fire frequency is higher experienced during the last quarter of the year (summer/dry seasons), especially in Western Australia, Northern, and Queensland. The consistency of fire occurrence throughout the year in wealthier countries such as the United States and Australia is related to fuel management practices occurring during the cooler/wetter nonfire season, especially in highly populated areas (EARL; SIMMONDS, 2018).

We also included some covariates that could be important in the fire observations since our data set includes fire occurrences of different causes, such as human sources (deliberately or accidentally), and natural causes, that can be influenced by climate variables. As expla-

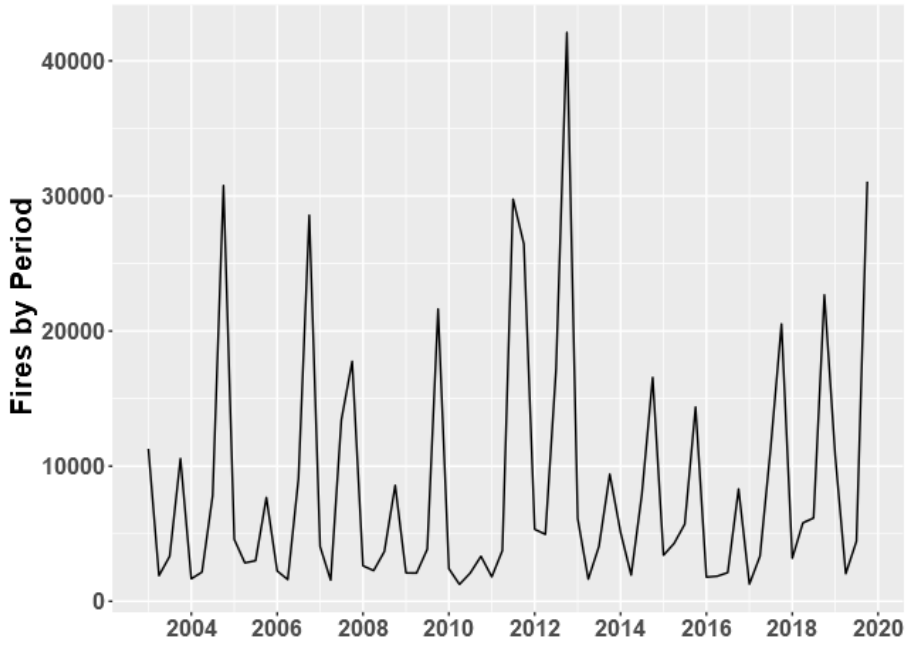


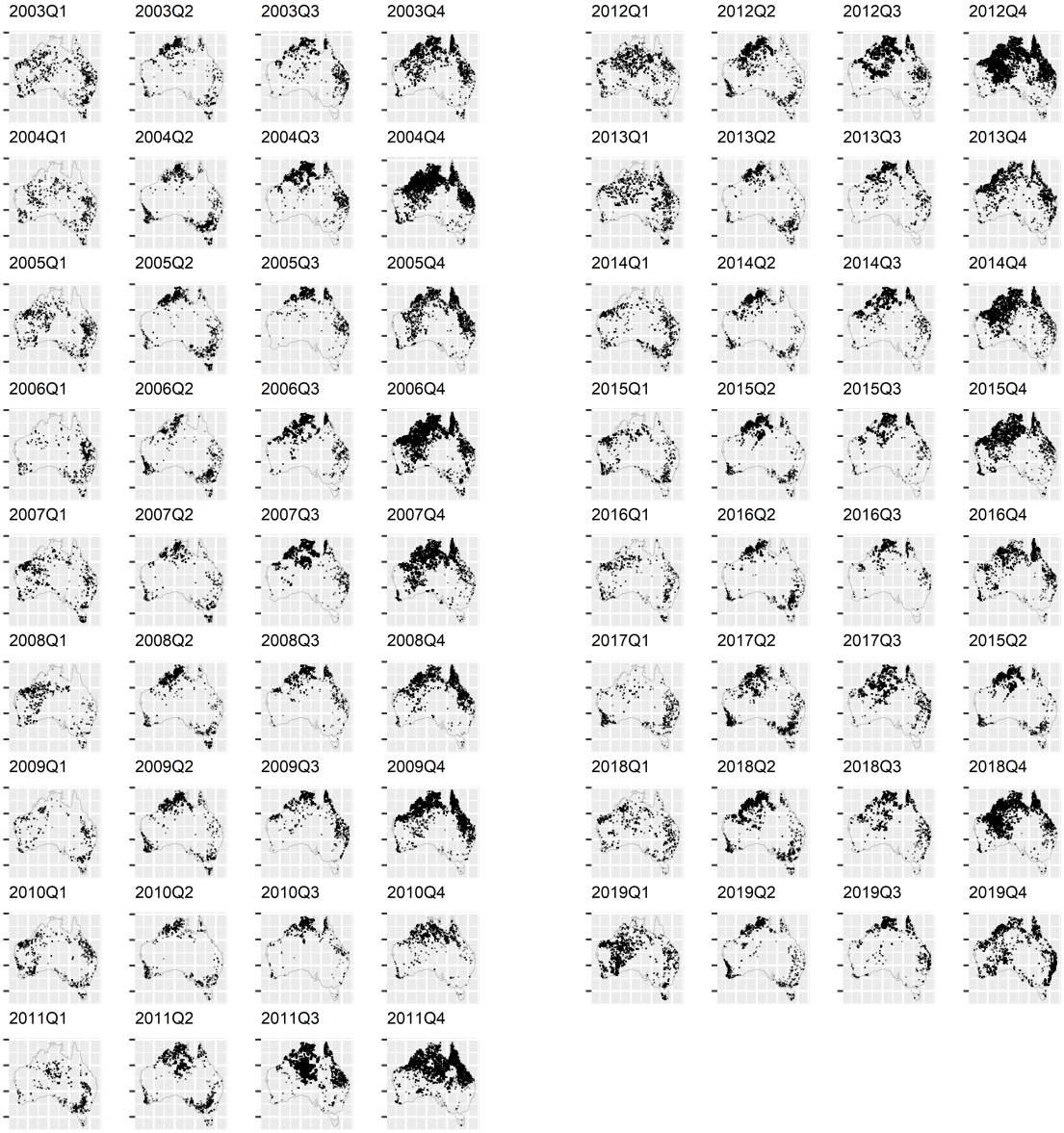
Figure 3.1 – Fires in Australia by quarter - From 2003 to 2019

natory variables, we used monthly remote sensing data of maximum temperature and rainfall from AusCover data portal (http://www.auscover.org.au/) of Terrestrial Ecosystem Research Network (TERN), the distance from the geographic positioning of the sealed roads from National Topographic Database of Australia (GEODATA TOPO 250K series 3, available at https://data.gov.au/data/dataset/a0650f18-518a-4b99-a553-44f82f28bb5f), and the geographical latitude. Temperature, rainfall and latitude are direct explanatory factors for the natural occurrence of fires, while the distance to sealed roads is a control for human influence in the occurrence of these events.

3.2.2 Spatio-temporal Log Gaussian Cox Process

One way to deal with spatio-temporal point pattern data comes from spatio-temporal point processes. The Poisson process is a common structure used to model point process. However, this structure is limited even if one assumes a inhomogeneous distribution in space through a function of deterministic intensity. The limitations are related to the lack of possible sources of uncertainty and the fact that the Poisson process is conditionally independent. An alternative is to allow the dependency function to be a stochastic function, known as Cox process. In this paper, we used the structure of Log Gaussian Cox Process, which is a particular case of the Cox process, where the log-intensity function is a Gaussian random field. Additionally, to identify long term variability, and cyclical and seasonal effects, we adopted a decomposition of the intensity function into trend, seasonality, and cycle components along with spatial random effects.

Therefore, the model used in this work is a spatio-temporal formulation of point processes with stochastic intensity, using a decomposition of the intensity function into components that vary over time and space. The proposed model (VALENTE; LAURINI, 2020) can be written as



(a) 2003 - 2011 (b) 2012 - 2019

Figure 3.2 – Spatial distribution of fires in Australia - From 2003 to 2019

follows:

$$
\begin{aligned}
Y(s,t) &= Poisson(|e(s,t)|exp(\lambda(s,t)),\\
\lambda(s,t) &= \alpha + \mu_t + s_t + c_t + z(s,t)\beta + \xi(s,t)\\
\mu_t &= 2\mu_{t-1} - \mu_{t-2} + \eta_\mu\\
s_t &= s_{t-1} + s_{t-2} + \ldots + s_{t-m} + \eta_s\\
c_t &= \theta_1 c_{t-1} + \theta_2 c_{t-2} + \eta_c\\
\xi(s,t) &= \Phi\xi(s,t-1) + \omega(s,t)
\end{aligned}
\tag{3.1}
$$

where $Y(s,t)$ is the number of occurrences in a region s in time t, $e(s,t)$ is the exposure offset for the region s, α is the intercept, μ_t is the long term trend modeled as a second-order random walk (RW2), which imposes a smoothness structure that is able to identify the trend component.

In addition, the RW2 structure can be thought as a non-parametric trend structure since it can be related to spline models, which allows to identify in a more adequate way the persistent patterns of long-term variability. The s_t represents the seasonal components, c_t is a cycle component represented by an second-order autoregressive process with possibly complex roots. This component allows the reproduction of patterns with periodic (cyclic) components, which are appropriate for effects that are repeated over time (irregular periodicity) but eventually dissipate. In the problem in question, the cycle component is interesting since it allows to reproduce the effect of climatic variables that generate periodic patterns that last for more than a year, and thus climatic effects that are beyond pure seasonal components. The $z(s,t)$ is a set of covariates observed in the location s and period t, η_μ, η_c and η_s are nonspatial independent innovations with $\eta_\mu \sim N(0, \sigma^2_{\eta_\mu})$, $\eta_c \sim N(0, \sigma^2_{\eta_c})$ and $\eta_s \sim N(0, \sigma^2_{\eta_s})$. The $\xi(s,t)$ are the spatial random effects represented by the Gaussian process $\omega(s,t)$ continuously projected in space and given by

$$Cov(\omega(s,t)\omega_(s',t')) = \begin{cases} 0 & \text{if } t \neq t' \\ \sigma^2 C(h) & \text{if } t = t' \end{cases} \qquad \text{for } s \neq s'$$

where $C(h)$ is a covariance function of the Matérn class and σ^2 is the marginal variance. More detailed discussion about the method are available in Supplementary Information.

As the LCGP likelihood is analytically intractable, it is necessary to approximate the likelihood. To do this, one may use the SPDE approach (LINDGREN; RUE; LINDSTRÖM, 2011), by using the approximation of SPDE solution as follows:

$$\omega(s,t) \approx \tilde{\omega}(s,t) = \sum_{j=1}^{n} w_j \varphi_j(s,t) \tag{3.2}$$

where n is the number of vertices of the triangulation, $\{w_j\}_{j=1}^n$ are the weights with Gaussian distribution and $\{\varphi_j\}_{j=1}^n$ are the basis functions defined for each node on the mesh. The idea is to calculate the weights $\{w_j\}$, which determine the values of the field at the vertices, while the values inside the triangles are determined by linear interpolation (LINDGREN; RUE; LINDSTRÖM, 2011).

Replacing the Gaussian Field (GF) $\omega(s,t)$ by the Gaussian Markov Random Field (GMRF) approximation $\tilde{\omega}(s,t)$ in Eq. (3.1), and approximating the integral in the LGCP likelihood by a quadrature rule, it results that the approximate likelihood consists of $(n + n_t)T$ independent Poisson random variables, where n is the number of vertices and n_t is the number of observed point processes (see Appendix for details). In addition, according to Simpson et al. (2016), the LGCP formulation fits naturally within the Bayesian hierarchical modeling framework and are latent Gaussian models, therefore, it may be fitted using the Integrated Nested Laplace Approximations (INLA) approach of (RUE; MARTINO; CHOPIN, 2009).

3.3 Results

Our analysis is based on a statistical model to decompose the temporal and spatial patterns of fire occurrence into long-term changes and transient effects. The model consists of a trend component, identifying persistent changes in fire patterns, transient components capturing cyclical and seasonal effects, and a spatial component capturing the territorial heterogeneity in

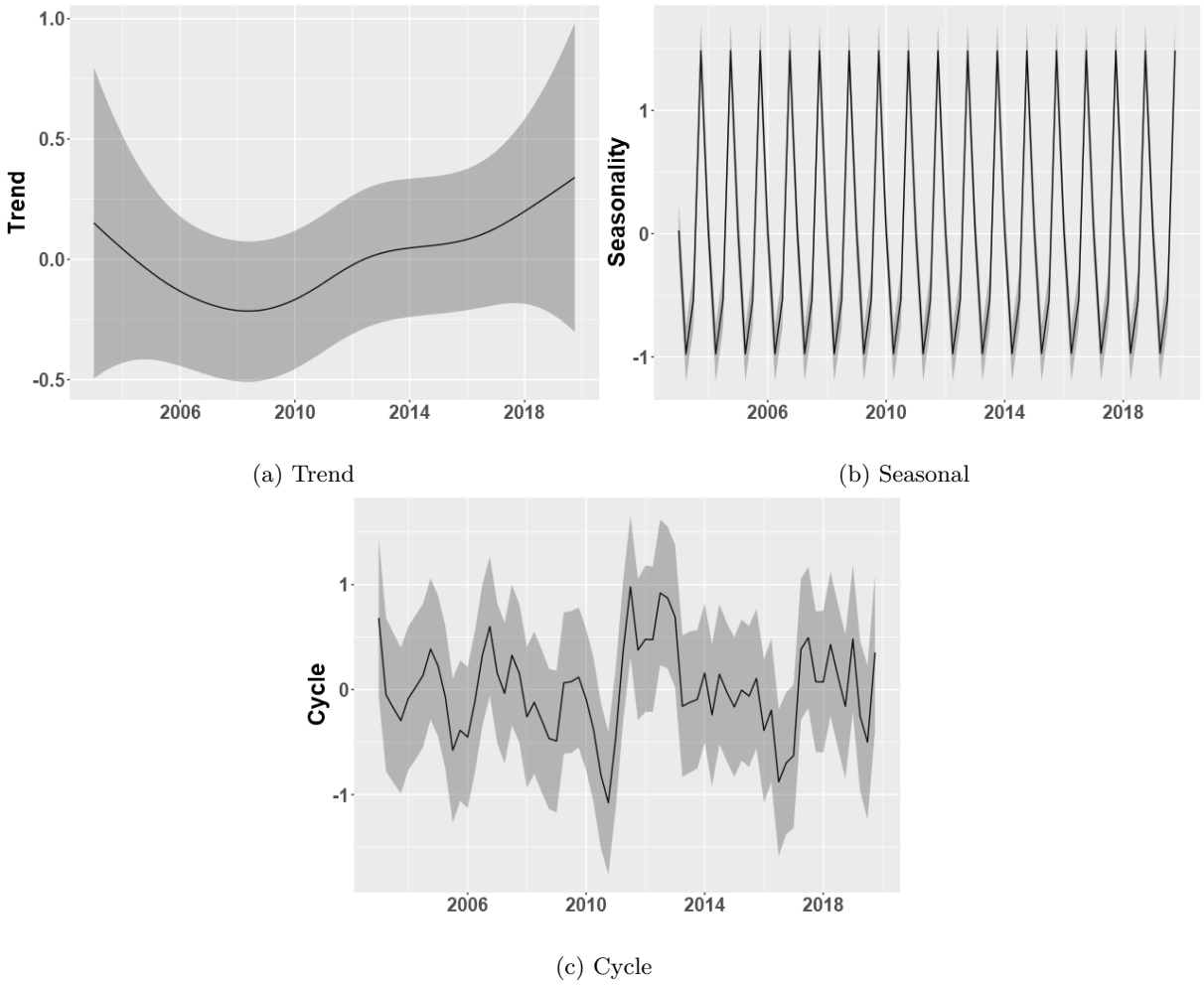
the occurrence of these events. We performed the estimation of the parameters based on two different specifications, with and without covariates effects, which we call model M2 and M1, respectively. It is important to note that the formulation of model M1 is able to explain the spatio-temporal patterns of observed fire occurrence. However, we were also interested in to assess if the changes in the patterns of fire events can be related to climate variability. While maximum temperature and rainfall are related to climate factors, the location of the sealed roads provides evidence of intentional fire, since proximity to highways indicates accessibility and may provide evidence of human-induced fires. We use alternative formulations using unsealed roads, and the combination of sealed and unsealed roads. The specification with sealed roads had a slightly superior performance in terms of model fit, and so it is the form maintained in the model. We also included latitude as a covariate in this model, since there is evidence pointing to the dependence of fires regimes on this variable (MURPHY et al., 2013; WILLIAMSON et al., 2016).

First, we estimated the parameters described in Equation (3.1) without the effects of the covariates (model M1). In this case, the estimated parameters are the intercept (α), the precision of the trend component ($1/\eta_\mu$), seasonal component ($1/\eta_s$), and cycle component ($1/\eta_c$), the parameters of the second-order autoregressive process of the cycle (PACF1 and PACF2), the parameters of spatial covariance ($\log\tau$ and $\log\kappa$), and the parameter of spatial time dependence (Φ). The parameters $\log\tau$ and $\log\kappa$ are due to the parameterization proposed by Lindgren, Rue and Lindström (2011), which are better defined in Appendix. In the second model specification (model M2) were included four explanatory variables namely, maximum temperature, rainfall, latitude, and sealed roads. In addition to the above, in this case, the estimated parameters include the parameters associated with the set of observed covariates (β).

The estimated precision parameters of trend, seasonal and cycle components under model M1 (see Supplementary Table 3.2) show a high precision associated with the seasonal component (estimated posterior mean equals 14267.626) as well as the trend component (estimated posterior mean equals 8379.048), whereas the cycle component shows a relatively minor precision (estimated posterior mean equals 2.228). The partial correlation parameters are related to the autoregressive parameters in the AR(2) representation of the cycle. The estimated parameters (0.194 and -0.044) indicate the presence of a cyclic component with the estimated period for the cycle component being equal 5.88 quarters.

Based on the estimated trend, seasonality, and cycle components of model M1 (posterior mean and 95% Bayesian credibility interval; Figure 3.3), the most notable result is the trend component, which shows that there was a decrease in Australian fire occurrence from 2003 to 2010. This was followed by an upsurge in 2011, previously discussed in the literature (e.g., Giglio, Randerson and Werf (2013), Dutta, Das and Aryal (2016), Earl and Simmonds (2017)). In addition, from 2011 to 2020, the fire levels exhibit a growth pattern. The seasonal component is stable with very tight credible intervals, which is consistent with the estimated precision parameter.

In order to assess if the variability in the patterns of fires can be associated with climate factors, we performed the estimation of model M2, which were included four explanatory variables. The estimated posterior means (see Supplementary Table 3.1) indicate a negative relation between fire events and rainfall (-0.002), and the distance from sealed roads (-0.007), as ex-



(a) Trend

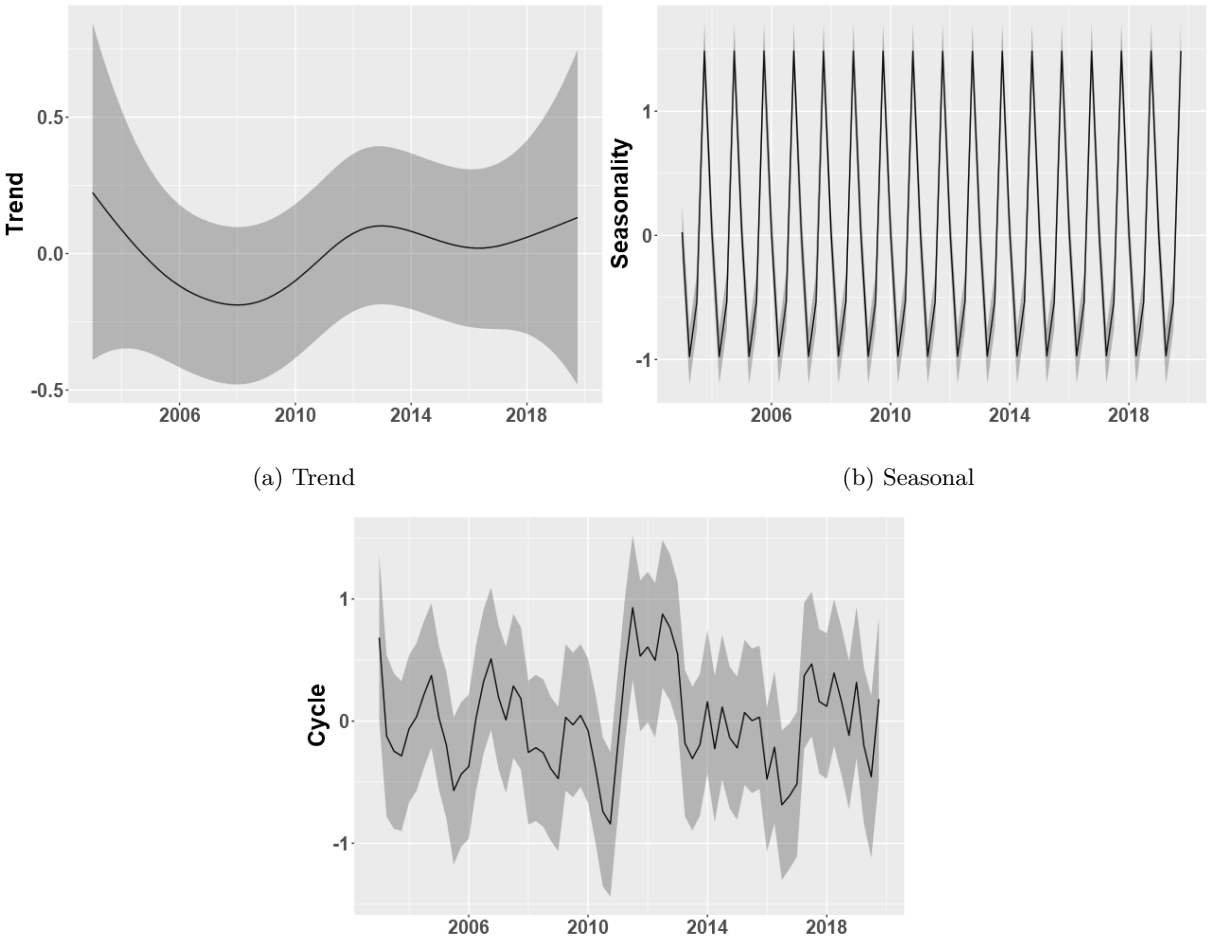
(b) Seasonal

(c) Cycle

Figure 3.3 – Trend, Seasonal and Cycle decomposition of fire occurrence in Australia - Model without covariates (M1). The shaded areas in the graph represents the 95% Bayesian credibility interval.

pected, and a positive relation between maximum temperature and fire occurrences (0.149). In addition, there are evidences of the influence of the latitudinal gradient in the fire activity, which is reflected in the weather conditions during the fire events Williamson et al. (2016), Murphy et al. (2013). Based on these discussions, we included the geographical latitude as a covariate, which exerts a positive effect on the fire occurrence. It is possible to note that, under model M2 specification, the trend component is relatively stable after the upsurge in 2011 (Figure 3.4).

The spatial heterogeneity of the fire occurrences in Australia can be seen through the estimated spatial random effects under models M1 (see Supplementary Figure 3.9) and M2 (Figures 3.5 and 3.6). In addition, to show the importance of the trend, seasonal and cycle components in the analysis of fire occurrence, we plotted the observed total fire count and the predicted value of fire count in each year given by the sum of the estimated trend, seasonal, cycle and intercept components of the models M1 (see Supplementary Figure 3.10) and M2 (Figure 3.7). The results provide evidence that the latent components explain the most part of the variability observed in the total fires count since the prediction of the total count mostly lies inside the 95% credibility interval for the whole period. Also, the estimated intensity function and the observed fire occurrence (black dots) for models M1 (see Supplementary Figure 3.11) and M2 (Supplemen-



(a) Trend

(b) Seasonal

(c) Cycle

Figure 3.4 – Trend, Seasonal and Cycle decomposition of fire occurrence in Australia - Model with covariates (M2). The shaded areas in the graph represents the 95% Bayesian credibility interval.

tary Figure 3.12) shows that the estimated log intensity function explains the spatio-temporal variation observed in the fire count, which suggests that the proposed model has a good fit. Supplementary Table 3.3 shows some fit measures for the two models (mean error (ME), root mean squared error (RMSE), mean absolute error (MAE) and mean percentage error (MPE)). In general, the models present a good fit in these measures, with a small negative bias in both models, but of very low magnitude, with mean percentual errors of -0.332% and -0.327% for models M1 and M2, respectively.

In order to support the evidence that changes in fire occurrences have been related to climate factors, we performed a similar method for decomposition of trend, cycle and seasonal components in spatio-temporal models to investigate the existence of variability in the patterns of the maximum temperature and rainfall from 2003 to 2019 in Australia. In summary, the central idea of the method is to decompose the time series in a similar way to a time series structural decomposition, with the innovation process in each location that contains an error component projected in the spatial continuum (LAURINI, 2019). For the observed maximum temperature our analysis show the presence of a tendency from 2010 to 2019 (see Figure 3.8), evidencing the

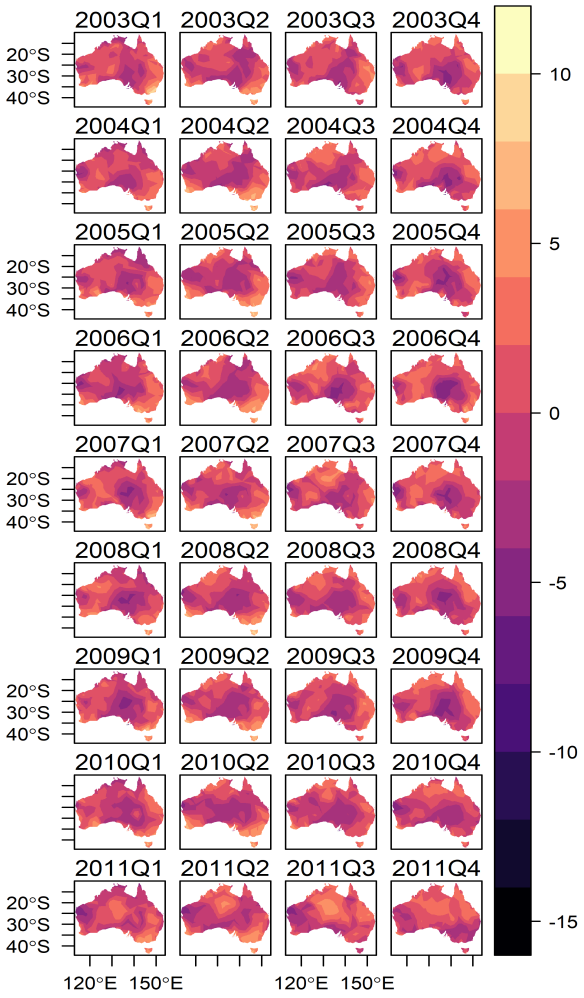


Figure 3.5 – Spatial Random Effects - Model with covariates (M2) 2003-2011

relationship between climate factors and fire occurrence in Australia, in agreement with previous results (e.g., Hughes (2003), Griffiths et al. (2005), Alexander et al. (2007)). On the other hand, the results obtained through the modeling carried out in this article do not indicate the presence of relevant changes in the trend of rainfall series (see Supplementary Figure 3.13). In addition, the proposed model was not able to capture a significant cycle component.

3.4 Discussion

Before moving to discuss our findings, it is worth spending a few words on what we consider the main limitations of this study. There is a meaningful limitation related to the selected covariates in our analysis. Since the proposed model performs a spatio-temporal analysis for the occurrences of a process observed continuously in space, the covariates must to be available at every location of the interest region within the observation window. Due to this methodological constraint, the number of available covariates are limited. In particular, in our paper, we were able to include only information in climatic patterns and measures of human presence.

The second limitation is more significant and difficult to overcome. As previously stated, our

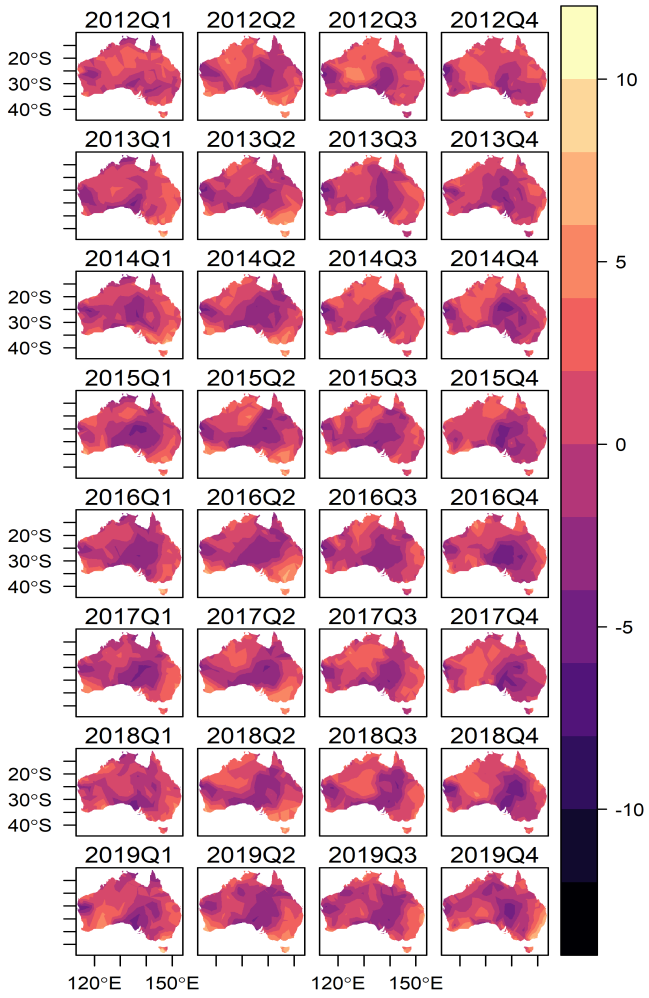


Figure 3.6 – Spatial Random Effects - Model with covariates (M2) 2012-2019

database contains only 16 years of data from fire occurrences in Australia and due to the limited data sample the results demand our attention. In other words, given data limitations, our results may be sensitive to uncertainty and caution is required in its interpretation. In fact, strongest results to provide baselines for assessing the long-term changes in the pattern of fire occurrence in Australia require longer time series. Although, we believe that the problem addressed in our paper is important and timely, and the proposed method can give some new insights to this subject, significantly contributing to the literature of statistical analysis of climate variability through spatio-temporal models. In addition, to highlight the potential of our approach, we discuss here about the validity of using statistical modeling for the analysis of relevant climate-related issues under data availability limitations, and also how the model can provide compelling evidence (although not conclusive) of the impact of climate patterns in the spatio-temporal variability of Australian fire activity.

In fact, the most relevant issue is the separation between long-term and transient effects of the fire occurrences, which is central in the interpretation of the model results. The trend component plays a crucial role, incorporating the persistent changes in the fire occurrence. In our model, the use of a random walk model imposes an statistical identification that ensures that the

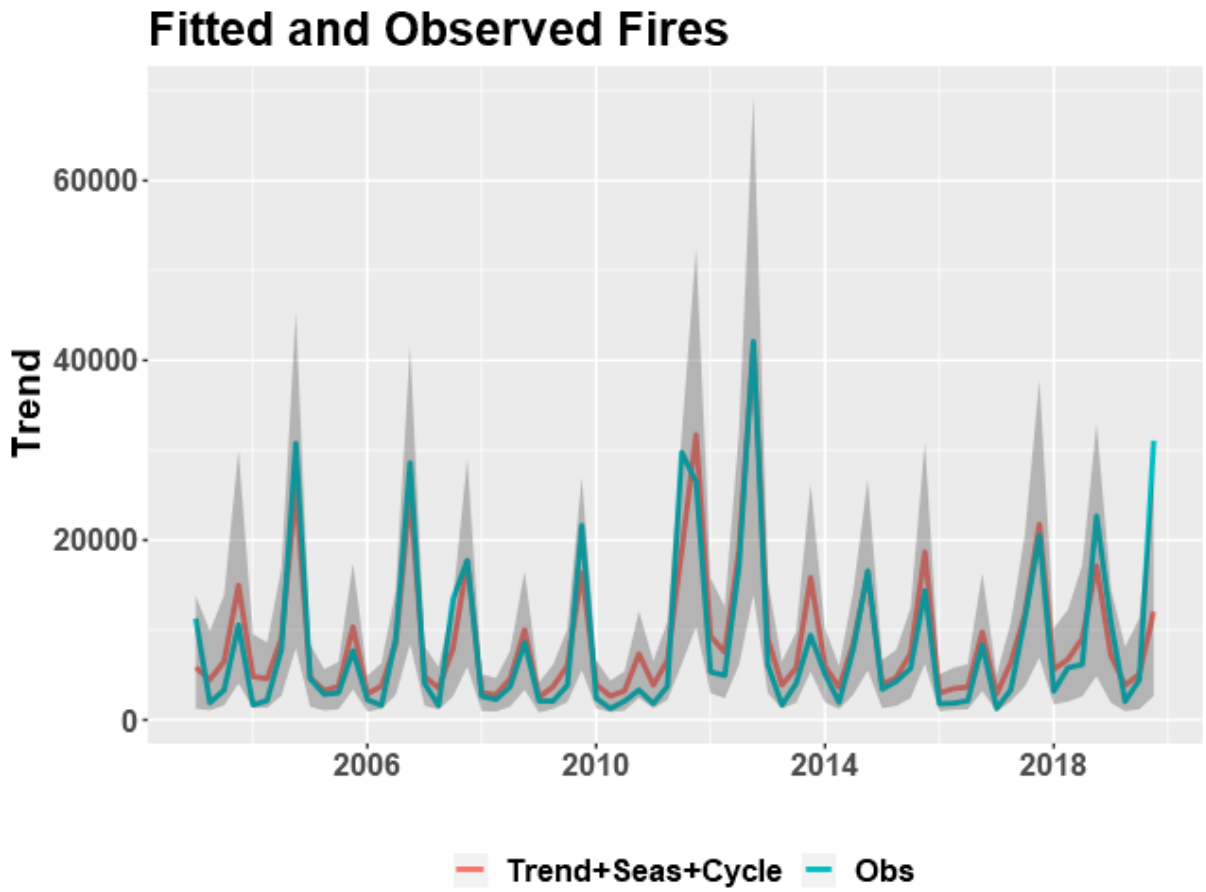
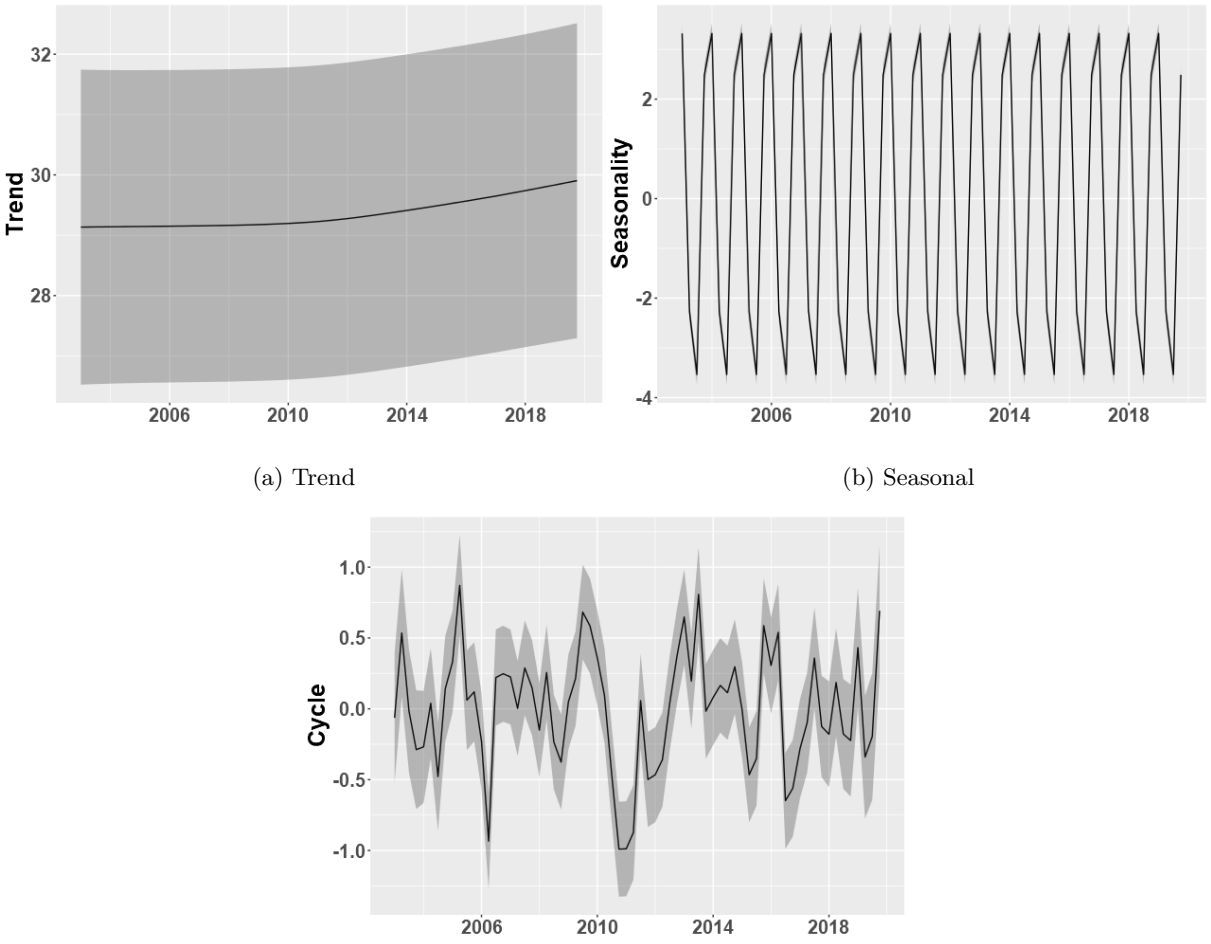


Figure 3.7 – Predicted fires by the sum of trend, seas and cycle components and observed fires - Model with covariates (M2). Shaded areas in the graph represent the 95% Bayesian credibility interval.

trend component only captures long-term changes, isolating the effects of short-term changes which are captured by the cycle and seasonality components, and also the spatial patterns. Thus, we are using interpretation features that are common to other statistical models that try to identify possible long-term movements in climate-related issues through non-stationary latent components, which comprises a rich literature of statistical models to analyze climate changes (e.g. Bloomfield (1992), Estrada, Perron and Martínez-López (2013), Laurini (2019), Valente and Laurini (2020)). After all, the whole idea is precisely imposing an identification structure that encompasses all the persistent changes in a common component, which aggregates all changes with relevant long-term effects, which is in turn a way of estimating patterns of climate variability (or climate change, considering longer observed time series) using temporal and spatio-temporal models. In particular, in the presence of limited data it is necessary to impose restrictions in order to be able to separate long term and transient effects, which is also necessary due to the non-stationary nature of the long-term climate variability processes.

With these caveats in mind we move on to discuss the findings. Our evidence suggests that there were an increasing trend of the intensity of fire occurrence in Australia since 2010. Yet, as previously discussed, mathematically, the estimated long-term component can be seen as the accumulation of all shocks that occurred in the past with non-transitory effects, and this is the reason why the level shift would correspond to persistent changes. However, despite the



(a) Trend

(b) Seasonal

(c) Cycle

Figure 3.8 – Trend, Seasonal and Cycle decomposition of maximum temperature

evidence of long-term changes in fire activity, given data limitation in this particular case, we cannot state with confidence that the observed variability in the estimated trend component corresponds to long-term changes. Indeed, since the decadal climate variability and climate change overlap, based on a short observed time series, it is difficult to distinguish between the two effects. Changes in fire weather over longer time scales have been widely discussed in the literature, and it have been associated with anthropogenic climate changes (DOWDY; FIELD; SPESSA, 2016; HARRIS; LUCAS, 2019), but also with climate variability (e.g., Interdecadal Pacific Oscillation) (VERDON; KIEM; FRANKS, 2004). On the other hand, it is worth noting that, for longer time series, our proposed model could be considered as an important tool to identify the distinct effects from climate change and climate variability, given the model ability to capture persistent and mean-reverting (seasonal and components with irregular periodicity) terms, taken into account the effects of covariates and the spatial heterogeneity.

Given the mean-reverting and irregular periodicity nature of the cycle component, it was possible to capture the effects of interannual and/or multi-year climate variability in the fire activity. As an example, based on the results, it is possible to observe that the estimated cycle component was able to capture the considerable decrease in the fire activity in 2010-11, which

coincides with weak to moderate and moderate to strong La Niña events[1], being one of the wettest calendar years on record for Australia, and also the lagged response from the same La Niña, which drove up fuel loads in central Australia, which consequently increase the fire activity.

By including covariates in the model we were able to investigate the link between the increases in fire activity and anthropological activities and climate factors over time. Our findings suggest that climate patterns and human activities are underlying factors that have driven the upward trend of the fire occurrence. Fire requires sufficient biomass, biomass available to burn, ambient conditions conductive to spread, and ignitions (BRADSTOCK, 2010). These factors influence the spatio-temporal fire activity, and is strongly dependent on the meteorological conditions and how these conditions interact with the vegetation, or fuel, resulting in fundamentally different fire regimes across Australia. The anthropological influence may further complicate the influence of biophysical drivers, through land-use modification, human ignitions, and fire suppression (ANDELA et al., 2017). Supporting the evidence that the variability in the fire occurrence have been related to climate variability, our findings suggest a growth pattern of the trend of maximum temperature. This result is in agreement with some other studies, which have reported that the variability in fire occurrence conditions are consistent with observed changes in temperature and rainfall throughout Australia (RUSSELL-SMITH et al., 2007; CLARKE; LUCAS; SMITH, 2013; DOWDY, 2018). On the other hand, the lack of meaningful results related to changes in the trend of rainfall series may be related to different patterns of precipitation throughout Australia. Many studies have analyzed Australian rainfall trends and have shown more regionally dependent variations (HUGHES, 2003). Thus, since we assessed the rainfall trend over a large area, it is too hard to capture the presence of a tendency for rainfall.

In addition, by including the geographical latitude as a covariate, our results corroborates previous analyses of spatio-temporal variability in fire activity over the entire country, pointing out the importance of the latitudinal information in shaping temporal patterns of fire activity (MURPHY et al., 2013; WILLIAMSON et al., 2016). Our findings suggest that the latitude has a positive effect on fire intensity, in that by increasing the geographical latitude the fire frequency is more intense. It is worth noting that as the entire analyzed region is below the equator, the latitudes are negative, indicating that the higher fire frequency is concentrated along with the northern Australia. Indeed, as discussed by Murphy et al. (2013), the variation in Australian fire regimes is related to latitudinal gradient in season rainfall, driven by summer monsoon activity in the north and winter frontal activity in the south. The very reliable dry-wet cycle in the north allows the high fire frequencies to occur annually (mostly in the dry season, from April to November), and thereby limiting maximum fire intensities. On the other hand, in the south, as the influence of the summer monsoon rainfall diminishes, fire frequency becomes strongly constrained by the fire weather and fuel moisture (BRADSTOCK, 2010), and the coincidence of extreme fire conditions and abundant fine fuels might only occurs every 5-10 years, which in turn increase maximum fire frequencies, occurring mostly during the summer and autumn. In the arid zones, the fire activity is constrained by the lack of continuous fuels or slow vegetation grows, and is characterized by intermittent periods of fire activity, which occurs mainly after

[1] See <http://www.bom.gov.au/climate/enso/lnlist/index.shtml>.

periods of high-rainfall, often associated with La Niña events(MURPHY et al., 2013).

From the spatial perspective, there is evidence of the temporal variability of fire activity across different regions of the country. Such a knowledge is important for management and planning applications that requires to consider potential threats to human life and economic losses under natural hazards. Based on our analysis it is possible to observe that fire is most frequent in northern Australia, where tropical monsoon climate dominates. On the other hand, fire occurrence is less frequent in the arid (central) and temperate (southern) zones. Indeed, it is possible to observe an increase in the fire activity in the arid central zones in 2011, which reflects the lagged response of the La Niña event between 2010 and 2011, that caused periods of high-rainfall. The temperate southern Australia is characterized by infrequent but intense fires, which is associated with severe drought. In particular, based on the spatial random effects, it is possible to see that our model was able to capture the intense fire activity in southern Australia between 2003 and 2011, when the "Millennium Drought" (DIJK et al., 2013) was broken by the above average rainfall in 2010 and 2011. It is worth noting that the results of this spatio-temporal variability of the fire activity are also useful to highlight the influence of the climate variability in this kind of event. In particular, is possible to observe that spatio-temporal results are consistent with the expected variations under Australian climate influences, like the monsoon and the east to southeasterly winds in the northern areas of Australia, and also the frontal systems and blocking highs in the southern Australia. These kind of climate drivers are capable to affect weather variables, such as temperature and rainfall, influencing the fire activity (DOWDY, 2018). Therefore, the spatio-temporal analyses of the variability of the fire events over Australia is a useful tool to understand the general patterns and temporal variability of the fire activity, and how climate drivers can influence it over space and time.

Evidence that climate patterns are responsible for the variability in fire occurrence in Australia has been previously reported in the literature, however our approach provides a structural time series decomposition of fire occurrence in Australia into a sum of trend, seasonal, and cycle components plus the effect of additional covariates, taking into account the spatial heterogeneity. Our method reveals the fire occurrence behavior and its association with climate factors avoiding some problems usually faced by inference procedures on climate-related issues, such as the dimensionality of the data and the difficulty to include spatial information of climatic effects. Furthermore, our results enable a more comprehensive understanding of the variability of fire occurrences in Australia under climate variability and can better inform the management and policy decisions.

3.5 Conclusion

As a contribution to the understanding of Australian patterns of fire occurrence, we propose to use a dynamic representation of a Log Gaussian Cox process where the intensity function is modeled through a decomposition of components into trend, seasonality, cycles, covariates and spatial effects, which is useful to identify persistent changes in the intensity of occurrences over time, and to capture seasonal and cyclical effects, taking into account the spatial heterogeneity. Within this framework, our findings suggested the existence of the variability in the trend component of Australian fire activity, suggesting that this variation may be associated with

anthropological activities and climate factors over time. Furthermore, we find a growth pattern of the trend of maximum temperature, evidencing the relationship between the variability in the fire occurrence have been related to climate patterns.

3.6 Supplementary Information

3.6.1 Methods

The covariance function of the Matérn class is given by

$$C(h) = \frac{2^{1-\nu}}{\Gamma(\nu)}(\kappa||h||)^{\nu}K_{\nu}(\kappa||h||) \tag{3.3}$$

where $h = ||s - s'||$ is the Euclidean distance between locations s and s', $\kappa > 0$ is a spatial scale parameter, $\nu > 0$ is the smoothness parameter and K_{ν} is a modified Bessel function.

The marginal variance σ^2 is defined by:

$$\sigma^2 = \frac{\Gamma(\nu)}{4\pi\kappa^{2\nu}\tau^2\Gamma(\nu + \frac{d}{2})} \tag{3.4}$$

where τ is a scaling parameter and d is the space dimension. To reduce the number of estimated parameters, we adopted a parameterization in terms of $\log\tau$ and $\log\kappa$ for the covariance function as follows

$$\begin{aligned} \log\tau &= \frac{1}{2}\log\left(\frac{\Gamma(\nu)}{\Gamma(\alpha)(4\pi)^{d/2}}\right) - \log\sigma - \nu\log\rho \\ \log\kappa &= \frac{\log(8\nu)}{2} - \log\rho \end{aligned} \tag{3.5}$$

where $\rho = \frac{(8\nu)^{1/2}}{\kappa}$.

Considering a bounded region $\Omega \in \mathbb{R}^2$, then the likelihood for an LGCP associated with data $Y = \{s_i \in \Omega : i = 1, \ldots, n; t = 1, \ldots, T\}$ is of the form

$$\pi(Y|\lambda) = \exp\left(|\Omega| - \int_{\Omega}\lambda(s,t)ds\right)\prod_{t=1}^{T}\prod_{i=1}^{n_t}\lambda(s_i,t). \tag{3.6}$$

In order to approximate the likelihood in Equation (3.6), we adopted the SPDE approach (LINDGREN; RUE; LINDSTRÖM, 2011). The main idea is to approximate the initial GF to a GMRF, and this approximation is possible due the fact that the GF x(s) with the Matérn covariance function is a stationary solution to the linear fractional SPDE (LINDGREN; RUE; LINDSTRÖM, 2011)

$$(\kappa - \Delta)^{\alpha/2}x(s) = W(s), \quad s \in \mathbb{R}^d, \quad \alpha = \nu + d/2, \quad \kappa > 0, \quad \nu > 0 \tag{3.7}$$

where $\Delta = \sum_{i=1}^{d}\frac{\partial^2}{\partial s_i^2}$ is the Laplacian operator and $W(s)$ is a spatial white noise. Therefore, in order to find a GMRF approximation of a GF, we need to find the stochastic weak solution of SPDE, through FEM, in which the approximation of SPDE solution is given by equation

$$\omega(s,t) \approx \tilde{\omega}(s,t) = \sum_{j=1}^{n} w_j\varphi_j(s,t). \tag{3.8}$$

Additionally, the basis function was chosen to be piecewise linear on each triangle, and the points where the weights $\{w_j\}$ are evaluated are given by the function:

$$\varphi_l(s,t) = \begin{cases} 1 & \text{at vertex } l \\ 0 & \text{elsewhere} \end{cases} \tag{3.9}$$

Through the replacement of GF $\omega(s,t)$ by the GMRF approximation $\tilde{\omega}(s,t)$ in equation

$$\xi(s,t) = \Phi\xi(s,t-1) + \omega(s,t) \tag{3.10}$$

and approximating the integral in Equation (3.6) by a quadrature rule, it is possible to obtain the LGCP likelihood approximation and, to perform inference procedures through the INLA algorithm, which provides accurate and efficient approximations on Bayesian hierarchical models that can be represented as latent Gaussian models.

3.6.2 Supplementary Results

Table 3.1 – Estimated Parameters - Model without Covariates (M1)

	Mean	SD	0.025quant	0.5quant	0.975quant	Mode
Fixed effects						
Intercept	-0.237	0.042	-0.32	-0.237	-0.154	-0.237
Random Effects						
Precision for trend	8379.048	1779.985	4798.899	8487.590	11414.706	8942.509
Precision for seasonality	14267.626	1525.812	11518.750	14182.542	17503.051	14010.271
Precision for cycle	2.228	0.198	1.852	2.225	2.627	2.226
PACF1 for cycle	0.194	0.057	0.070	0.200	0.289	0.220
PACF2 for cycle	-0.044	0.049	-0.140	-0.044	0.053	-0.043
Log τ	-1.638	0.018	-1.675	-1.637	-1.606	-1.632
Log κ	-0.393	0.016	-0.423	-0.393	-0.361	-0.395
Group Φ	0.481	0.025	0.444	0.477	0.538	0.458
Deviance Information Criterion (DIC)	62259.83					
Marginal log-Likelihood						

Table 3.2 – Estimated Parameters - Model with Covariates (M2)

	Mean	SD	0.025quant	0.5quant	0.975quant	Mode
Fixed effects						
Intercept	-1.201	0.344	-1.878	-1.201	-0.527	-1.200
Rainfall	-0.002	0.000	-0.003	-0.002	-0.002	-0.002
Max. Temp.	0.149	0.014	0.122	0.149	0.176	0.149
Distance to roads	-0.002	0.001	-0.003	-0.002	-0.001	-0.002
Latitude	0.076	0.014	0.048	0.076	0.103	0.076
Random Effects						
Precision for trend	7820.833	2097.167	3963.801	7795.822	11947.168	7750.662
Precision for seasonality	13165.237	3099.644	7565.813	13061.425	19543.288	12918.825
Precision for cycle	5.253	1.083	3.609	5.082	7.819	4.718
PACF1 for cycle	0.355	0.081	0.197	0.355	0.514	0.349
PACF2 for cycle	-0.351	0.060	-0.464	-0.353	-0.226	-0.358
Log τ	-1.738	0.018	-1.776	-1.737	-1.705	-1.734
Log κ	-0.270	0.015	-0.299	-0.270	-0.238	-0.272
Group Φ	0.503	0.013	0.480	0.501	0.531	0.497
Deviance Information Criterion (DIC)	62027.57					
Marginal log-Likelihood	-43955.13					

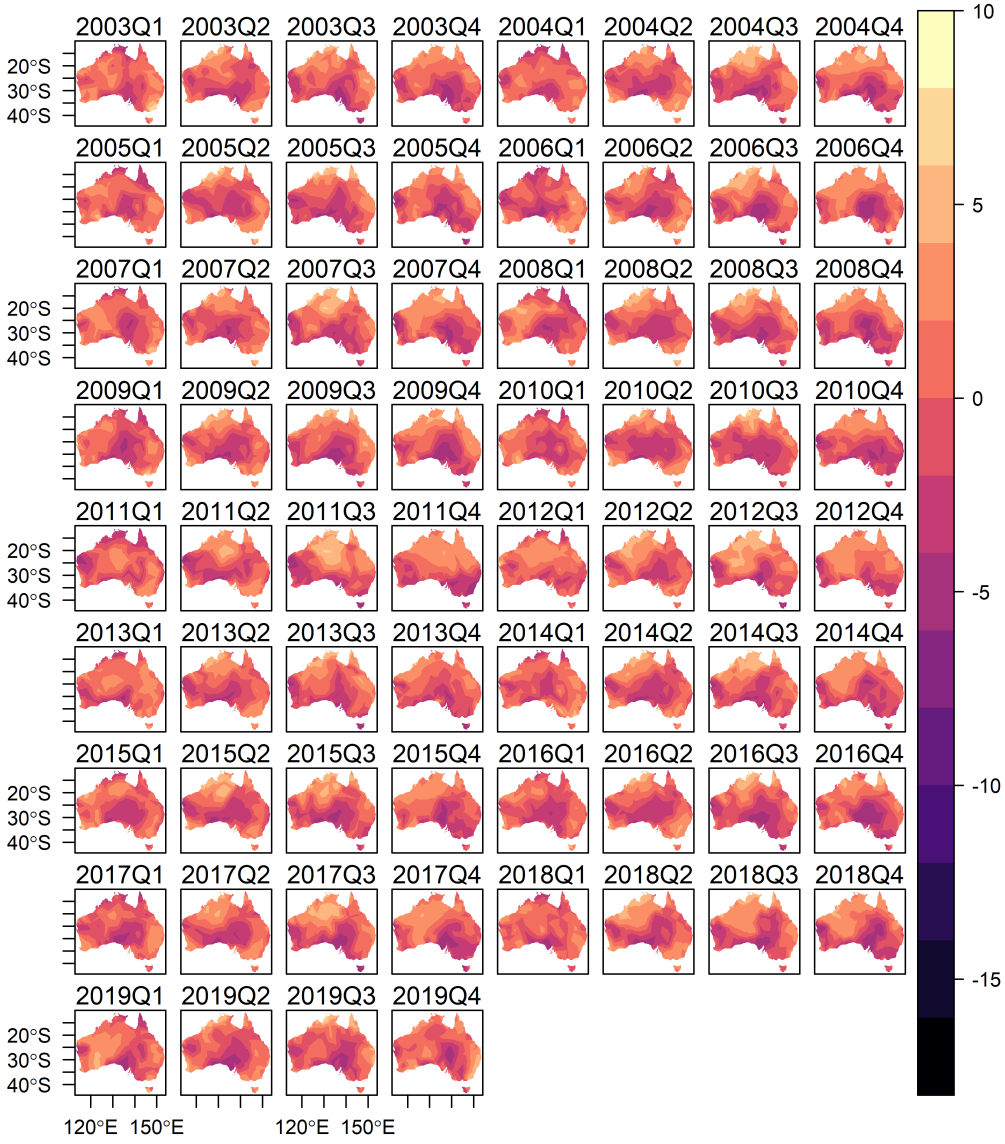


Figure 3.9 – Spatial Random Effects - Model without covariates (M1)

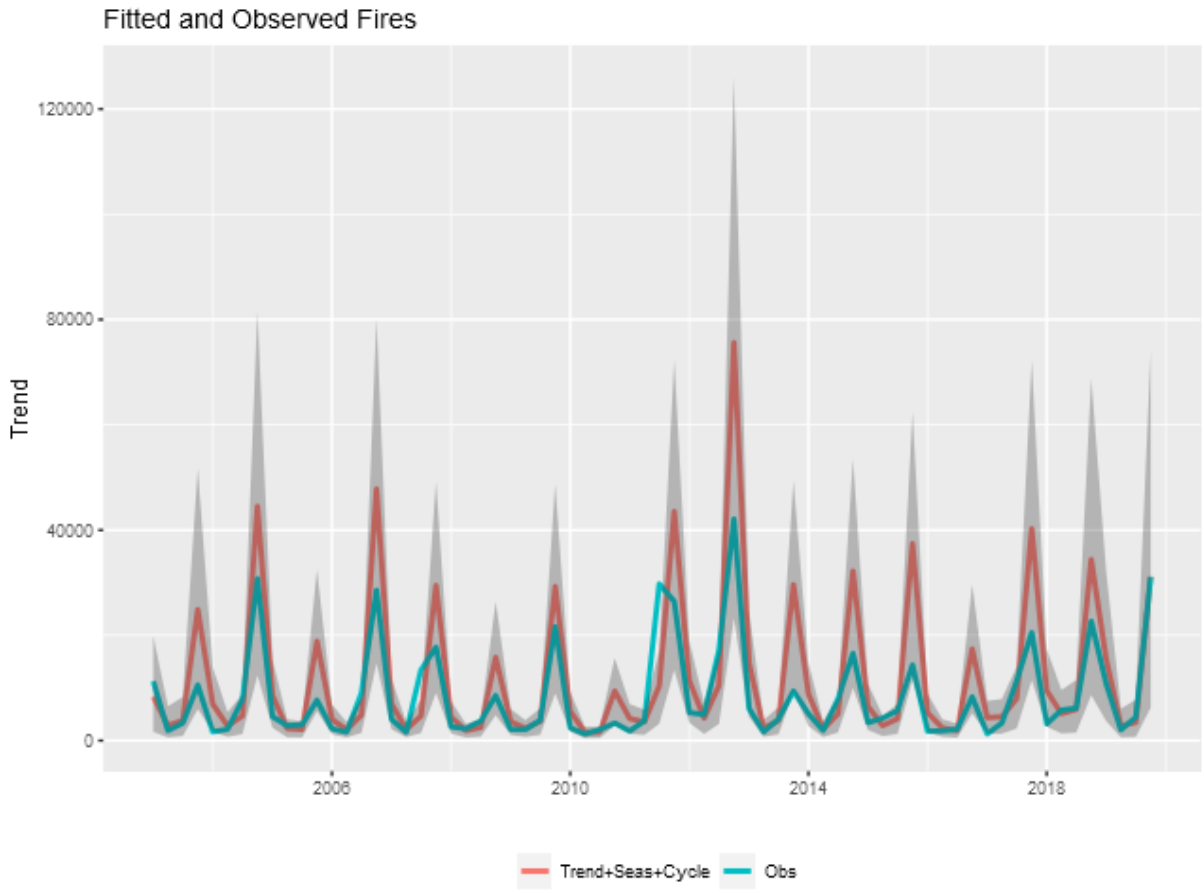


Figure 3.10 – Fitted Intensity and Observed Fires - Model without covariates (M1). Shaded areas in the graph represent the 95% Bayesian credibility interval.

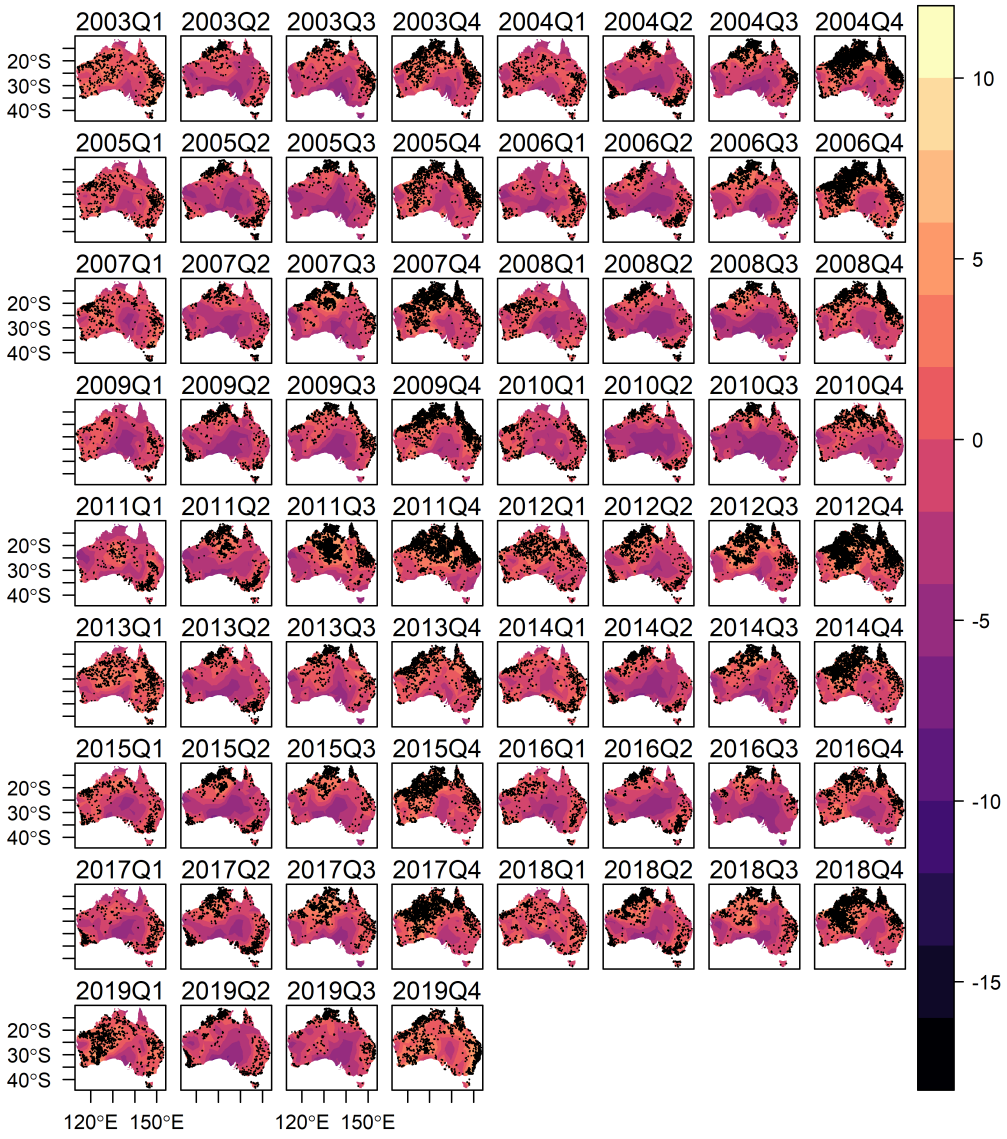


Figure 3.11 – Fitted and Observed Fires - Model without covariates (M1)

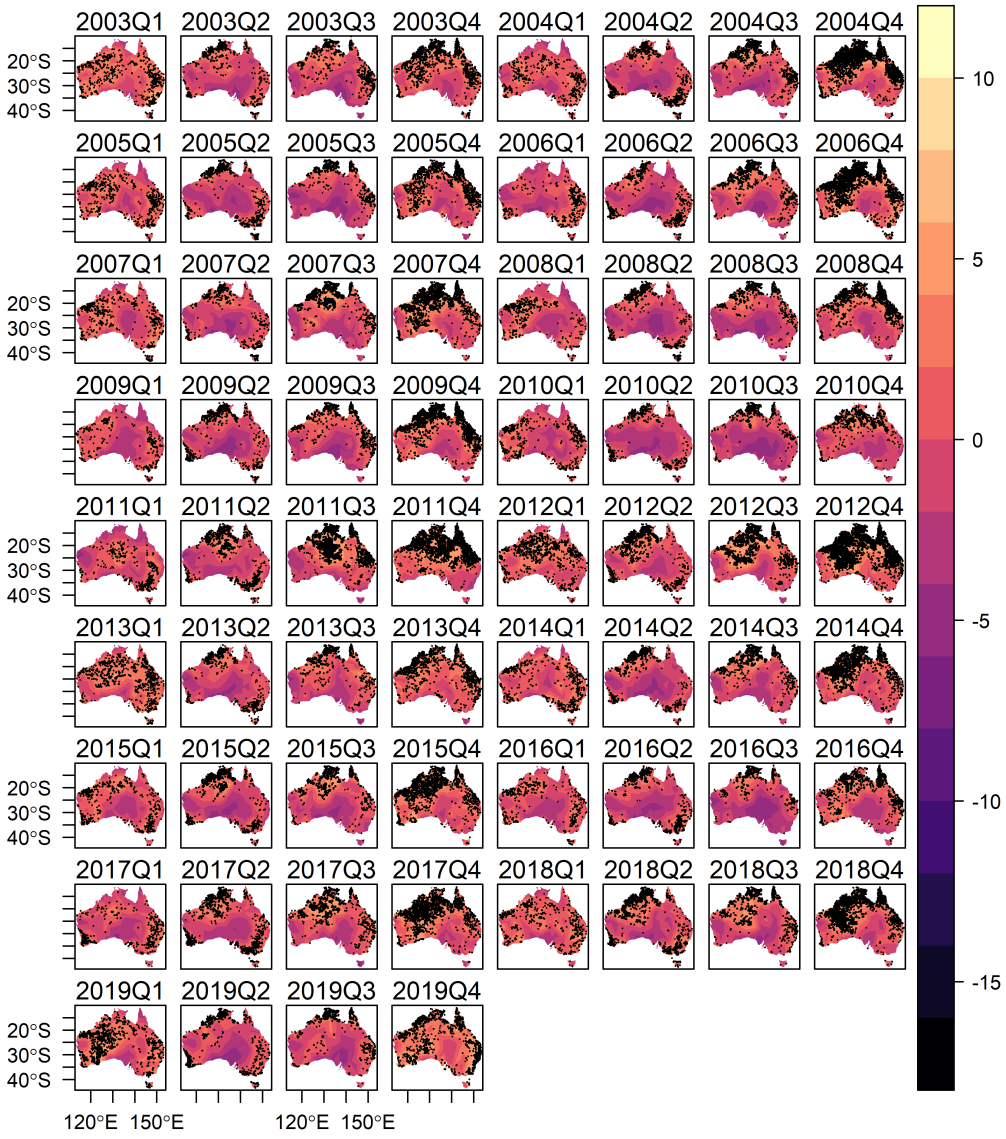
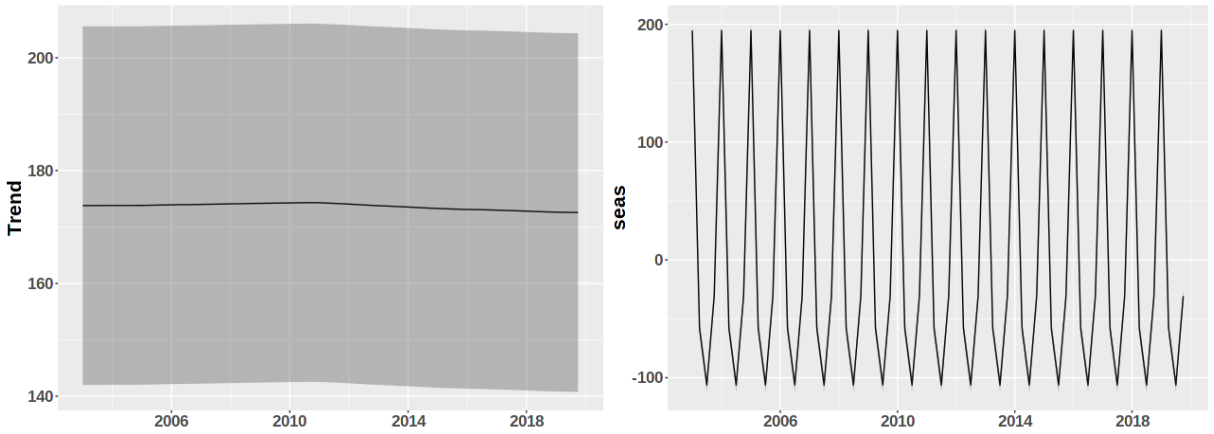


Figure 3.12 – Fitted and Observed Fires - Model with covariates (M2)

Table 3.3 – Model Fit Measures

Model	ME	RMSE	MAE	MPE
M1	-11.888	12.038	11.888	-0.332
M2	-11.783	12.007	11.783	-0.327

Note: ME - Mean Error, RMSE - Root Mean Squared Error, MAE - Mean Absolute Error, MPE - Mean Percentual Error. M1 - Model without covariates. M2 - Model with rainfall, max. temperature, distance to sealed roads and latitute covariates.



(a) Trend (b) Seasonal

Figure 3.13 – Trend and Seasonal decomposition of rainfall

4 A SPATIO-TEMPORAL APPROACH TO ESTIMATE POSSIBLE CHANGES IN THE PATTERNS OF FIRE OCCURRENCE IN THE BRAZILIAN AMAZON

Fernanda Valente Márcio Laurini

ABSTRACT

Fire occurrences are relevant disturbances in the Amazon region, with effects on the atmosphere composition, forest structure and composition, and the cycle of carbon. Understanding the patterns of fire occurrence is important to avoid loss and facilitate management decisions. As a contribution, we analyze the changes in the patterns of fire occurrence in the Brazilian Amazon, within the spatio-temporal point process framework. To perform inference procedures, we propose a novel methodology to extent the trend-cycle decomposition in spatio-temporal models to spatio-temporal point pattern data, by proposing to use a dynamic representation of a Log Gaussian Cox process where the intensity function is modeled through the decomposition of components into trend, seasonality, cycles, covariates and spatial effects. Overall, our results show that long-term movements of fire occurrence exhibits a marked decrease between the beginning of the sample and 2012, followed by an increase that extents to the end of the sample, which can be related to governance measures and market mechanisms. Additionally, our model was able to capture the spatial variability in the Brazilian Amazon, higher in regions where the climate has dry seasons.

Keywords: Legal Amazon, Fire Occurrence, Spatio-temporal models, Structural Decomposition, Point Process

VALENTE, Fernanda; LAURINI, Márcio. A spatio-temporal analysis of fire occurrence patterns in the Brazilian Amazon. **Scientific Reports**, v. 13,12727 (2023).

4.1 Introduction

The Amazon biome is one of Earth's greatest biological treasures, containing more than half of the world's rainforests and a quarter of all terrestrial species (MALHI et al., 2008). The Amazon rainforest also provides an environmental service by storing carbon, in both biomass and soils, and thus reducing the global warming (FEARNSIDE, 2012). Plus, evaporation and precipitation over Amazonia play important roles in the global atmospheric circulation, with effects on the climate across South America and North Hemisphere (GEDNEY; VALDES, 2000; WERTH; AVISSAR, 2002). Although the Amazon rainforest is shared by nine countries, about 60 percent of the Amazon Basin is in Brazil, where the political-administrative area called Legal Amazon encompasses nine Brazilian states, corresponding to 61% of the national territory.

Fire occurrences, related to both natural and anthropogenic activity, are relevant disturbances in the Amazon region, affecting the atmosphere composition (CRUTZEN; ANDREAE, 1990; LONGO et al., 2009), forest structure and composition (COCHRANE; SCHULZE, 1999),

and the cycle of carbon. In general, fire occurrence in the Amazon rainforest is related to land use, land cover, and climate patterns. In terms of total forest loss and fire occurrence, most of the changes in the land cover and human activities are concentrated along with the southern and eastern extent of the Brazilian Amazon region, called "arc of deforestation" (MORTON et al., 2006), which is related to the presence of roads and human accessibility (SIEGERT et al., 2001; SERRA et al., 2014). The expansion of roads and agriculture in the Legal Amazon began in the early 1970s when the Transamazon Highway was built. The construction of roads was accompanied by high rates of deforestation. For instance, between 1980 and 1990, the rates of deforestation in the Legal Amazon increased considerably, where approximately 225000km^2 of forest were cleared. In the same period, the extension of paved roads increased by more than 100% and unpaved roads increased by approximately 460% (FERRAZ, 2001).

From an economic standpoint, the fire occurrence in the Amazon region generates a great variety of costs with private and social consequences. In rural properties, the main losses occur when burning gets out of control and spread into pasture and forest areas. Also, losses related to fire occurrence may reach social proportions, through the release of carbon into the atmosphere, affecting global climate patterns, and provoking adverse health outcomes which impose direct and indirect costs on society, such as medical costs, labor loss, and utility loss (MENDONÇA et al., 2004; MENDONÇA; SACHSIDA; LOUREIRO, 2006).

Many factors may change the patterns of fire occurrence in the Amazon region. Previous studies have reported the impact of dry conditions on forest fire risk (NEPSTAD et al., 2004; ARAGÃO et al., 2007), and the effects of deforestation and fragmentation on regional climate, which have shown a significant increase in the mean surface temperature, and a decrease in the annual evapotranspiration and precipitation, which can further increase fire danger (NOBRE; SELLERS; SHUKLA, 1991; COSTA et al., 2007). Changes in the patterns of forest fire has also been reported to be related to agricultural expansion (MORTON et al., 2008), since the fire is a common and inexpensive tool used by Brazilian farmers to expand agricultural frontiers and to maintain and renew pastures. Also, due to climatic change, circulation shifts and increased anomalies such as El Niño events exacerbate extreme dry seasons in Amazonia (MARENGO, 2004; LI; FU; DICKINSON, 2006; MARENGO et al., 2008), changing the vegetation structure, potentially transforming the forest from highly resistant to fire ignition to extensively flammable (COCHRANE; BARBER, 2009), eventually leading to future increases burning frequency. Furthermore, there is evidence that forest fires create positive feedbacks in fire susceptibility, fuel loading, and fire intensity whereby recurrent fires become more likely and severe (COCHRANE et al., 1999; SIEGERT et al., 2001).

One way to verify the existence of changes in the patterns of the climate-related events, such as fire occurrence, is through the estimation of permanent and periodic components (BLOOMFIELD, 1992; PROIETTI; HILLEBRAND, 2017; LAURINI, 2019), using statistical tools to decompose the observed temporal variability into trend, seasonal and cycle components. However, the existing methods used to extract trends, seasonal, and cycle components face some problems to perform inference procedures on climate-related issues. First, these models are not fully adapted to the dimensionality of data sources used in climatology. Also, it does not take into account the spatial heterogeneity of climate effects. An alternative way to circumvent the aforementioned problems is a method that combines elements of structural time series decompo-

sition with spatio-temporal models with continuous spatial random effects, which can be thought as a process of decomposing geostatistical time series into a sum of trend, seasonal and cycle components and the effect of additional covariates (LAURINI, 2019; VALENTE; LAURINI, 2020).

The purpose of this paper is to analyze the existence of changes in the patterns of the fire occurrence in the Legal Amazon, within the spatio-temporal point process framework. To do this, we propose a methodology to extent the trend-cycle decomposition in spatio-temporal models to spatio-temporal point pattern data, by proposing to use a dynamic representation of a Log Gaussian Cox process (LGCP) where the intensity function is modeled through the decomposition of components into trend, seasonality, cycles, covariates and spatial effects (LAURINI, 2019; VALENTE; LAURINI, 2020). This is a useful formulation to identify possible changes in the intensity of occurrence over time, such as permanent changes in the fire occurrence, and to capture seasonal and cyclical effects.

The LGCP is a particular case of the Cox process, where the log-intensity function is a Gaussian random field. Due to the stochastic property of the LGCP, fitting this model is often computationally expensive. In this sense, to perform the estimation in a computationally effective way, we use the stochastic partial differential equation (SPDE) approach (LINDGREN; RUE; LINDSTRÖM, 2011) to transform the initial Gaussian random field (GRF) to a Gaussian Markov Random Field (GMRF), which is defined by sparse matrices. Furthermore, the resulting Bayesian hierarchical model fits within the integrated nested Laplace approximations (INLA) framework (RUE; MARTINO; CHOPIN, 2009), also providing significant computational improvements.

We present here the results of analyzing data for fire occurrence in the Legal Amazon, from January 2002 to December 2022. Our database contains daily fire reports from Moderate-Resolution Imaging Spectroradiometer (MODIS), with information such as spatial coordinates and temporal instant of fire events. Also, we include explanatory variables to control the main fixed effect related to climatic conditions and the use of the soil. Our results show that long-term movements of fire occurrence exhibits a marked decrease between the beginning of the sample and 2012, followed by an increase that extents to the end of the sample, which can be related to governance actions and market mechanisms. Our model also was able to capture the variability in the Legal Amazon, especially in the regions classified as wet tropical (Am), which is characterized by a dry season, that occurs between August and November (third and fourth quarters), and tropical with dry season (Aw). On the other hand, in western Amazon, where the climate is predominantly tropical without dry season (Af), the variability is low.

Our paper proceeds as follows. In section 2 we present the proposed method and data. In section 3 we present the results. In section 4 we discuss the results. We conclude in Section 5.

4.2 Material and Methods

4.2.1 Methods

Among models for the spatial point process, the Poisson process is the most fundamental structure. However, its application is limited due to its simplistic nature (TENG; NATHOO;

JOHNSON, 2017), even if one assumes a non-homogeneous distribution in space through a function of deterministic intensity. The limitations are related to the lack of possible sources of uncertainty and the fact that the Poisson process is conditionally independent. A related, but more flexible structure, is the Log Gaussian Cox process, a hierarchical structure where at the first level the process is assumed Poisson conditional on the intensity function, and at the second level, the log of the intensity function is assumed to be a Gaussian field (TENG; NATHOO; JOHNSON, 2017).

Given the doubly-stochastic property of the LGCP, fitting this model is a computational challenge. Within the Bayesian framework, the conditional autoregressive approach is a possible alternative to perform inference procedure and may be fitted using the INLA (ILLIAN et al., 2010). However, this approach is based on regular lattices over the observation window (SIMPSON et al., 2016), which could be highly inefficient since it requires to construct a much fine grid. For spatial models that combine a GRF with a Matérn correlation structure, the stochastic partial differential equations approach is a way to bypass the problem of inefficiency in the estimation under INLA method. The key of the approach is to use the fact that a GRF with Matérn covariance function is a solution to a SPDE and then the SPDE representation is used in conjunction with basis representation to construct a discrete approximation of the continuous field over the vertices of a 2-dimensional mesh covering the spatial domain (SIMPSON et al., 2016), i.e., the idea of the SPDE approach is to approximate the initial Gaussian field to a Gaussian Markov random field. One of the main advantages of this approach is the fact that GMRFs are defined by sparse matrices, allowing computationally effective methods.

In this paper we propose a spatio-temporal formulation of point processes with stochastic intensity, using a decomposition of the intensity function into components that vary in time and space. Specifically, we propose to use a LGCP structure, where the intensity function is decomposed into trend, seasonal, and cycle components together with spatial random effects, which allows us to identify permanent changes, and cyclical and seasonal effects. To perform inference procedure, we use the SPDE approach, allowing the use of Bayesian inference procedures based on INLA.

We first give a brief description of the SPDE approach, and a detailed discussion can be found in (LINDGREN; RUE; LINDSTRÖM, 2011; SIMPSON et al., 2016). Spatio-temporal data can be represented as realizations of a stochastic process indexed by a space and a time dimension

$$Y(s,t) = \{y(s,t) \mid (s,t) \in D \times T \in \mathbb{R}^2 \times \mathbb{R}\} \tag{4.1}$$

where D is a subset of $\mathbb{R}^2$, T is a subset of $\mathbb{R}$, s denotes a spatial coordinate and t denotes a time index. Using this structure, we can represent a spatio-temporal LGCP modelled as

$$\begin{aligned} Y(s,t) &= Poisson(|e(s,t)|exp(\lambda(s,t)), \\ \lambda(s,t) &= z(s,t)\beta + \xi(s,t) \\ \xi(s,t) &= \Phi\xi(s,t-1) + \omega(s,t) \end{aligned} \tag{4.2}$$

where $Y(s,t)$ is the number of occurrences in a region s and in time t, $e(s,t)$ is the exposure offset for the region s, $z(s,t)$ is a set of covariates observed in the location s and period t, and $\xi(s,t)$ are the spatial random effects represented by the Gaussian process $\omega(s,t)$ continuously

projected in space and given by

$$Cov(\omega(s,t)\omega_(s',t')) = \begin{cases} 0 & \text{if } t \neq t' \\ \sigma^2 C(h) & \text{if } t = t' \end{cases} \qquad \text{for } s \neq s' \tag{4.3}$$

where $C(h)$ is a covariance function of the Matérn class, which can be written as

$$C(h) = \frac{2^{1-\nu}}{\Gamma(\nu)}(\kappa||h||)^{\nu} K_{\nu}(\kappa||h||) \tag{4.4}$$

where $h = ||s - s'||$ is the Euclidean distance between locations s and s', $\kappa > 0$ is a spatial scale parameter, $\nu > 0$ is the smoothness parameter and K_{ν} is a modified Bessel function. The marginal variance σ^2 is defined by:

$$\sigma^2 = \frac{\Gamma(\nu)}{4\pi\kappa^{2\nu}\tau^2\Gamma(\nu + \frac{d}{2})} \tag{4.5}$$

where τ is a scaling parameter and d is the space dimension. Additionally, we adopt a parameterization in terms of $\log\tau$ and $\log\kappa$ for the covariance function (LINDGREN; RUE; LINDSTRÖM, 2011):

$$\begin{aligned} \log\tau &= \frac{1}{2}\log\left(\frac{\Gamma(\nu)}{\Gamma(\alpha)(4\pi)^{d/2}}\right) - \log\sigma - \nu\log\rho \\ \log\kappa &= \frac{\log(8\nu)}{2} - \log\rho \end{aligned} \tag{4.6}$$

where $\rho = \frac{(8\nu)^{1/2}}{\kappa}$. This representation is advantageous since, conditional on the value of ν, it is necessary to estimate only two parameters.

Considering a bounded region $\Omega \in \mathbb{R}^2$, it follows that the likelihood for an LGCP associated with data $Y = \{s_i \in \Omega : i = 1, \ldots, n; t = 1, \ldots, T\}$ is of the form

$$\pi(Y|\lambda) = \exp\left(|\Omega| - \int_{\Omega} \lambda(s,t)ds\right) \prod_{t=1}^{T}\prod_{i=1}^{n_t} \lambda(s_i,t). \tag{4.7}$$

Due to the doubly-stochastic property of the intensity function, the likelihood in (4.7) is analytically intractable. Since the term $\omega(s,t)$ corresponds to a GF with Matérn covariance, it is possible to use the SPDE approach to approximate the initial GF to a GMRF. The first main important result for the SPDE approach, is the fact that a GF x(s) with the Matérn covariance function is a stationary solution to the linear fractional SPDE (WHITTLE, 1954; LINDGREN; RUE; LINDSTRÖM, 2011)

$$(\kappa - \Delta)^{\alpha/2} x(s) = W(s), \quad s \in \mathbb{R}^d, \quad \alpha = \nu + d/2, \quad \kappa > 0, \quad \nu > 0 \tag{4.8}$$

where $\Delta = \sum_{i=1}^{d} \frac{\partial^2}{\partial s_i^2}$ is the Laplacian operator and $W(s)$ is a spatial white noise. Therefore, to find a GMRF approximation of a GF, it is necessary to find the stochastic weak solution of a SPDE, which can be constructed through Finite Elements Method (FEM) (LINDGREN; RUE; LINDSTRÖM, 2011). Thus, the approximation of SPDE solution is given by

$$\omega(s,t) \approx \tilde{\omega}(s,t) = \sum_{j=1}^{n} w_j \varphi_j(s,t) \tag{4.9}$$

where n is the number of vertices of the triangulation, $\{w_j\}_{j=1}^n$ are the weights with Gaussian distribution and $\{\varphi_j\}_{j=1}^n$ are the basis functions defined for each node on the mesh. In summary, the idea is to calculate the weights $\{w_j\}$, which determine the values of the field at the vertices, while the values inside the triangles are determined by linear interpolation (LINDGREN; RUE; LINDSTRÖM, 2011). Here, the basis functions are chosen to be piecewise linear on each triangle:

$$\varphi_l(s,t) = \begin{cases} 1 & \text{at vertex } l \\ 0 & \text{elsewhere} \end{cases} \tag{4.10}$$

The stochastic weak solution of (4.8) is found by requiring

$$\{\langle \phi, (\kappa^2 - \Delta)^{\alpha/2}\omega\rangle\}_\Omega \stackrel{d}{=} \{\langle \phi, W\rangle\}_\Omega, \tag{4.11}$$

where $\{\phi_i(s), i = 1, \ldots, m\}$ are test functions and "$\stackrel{d}{=}$" denotes equality in distribution. Replacing (4.9) in (4.11) gives us

$$\{\langle \phi_i, (\kappa^2 - \Delta)^{\alpha/2}\varphi_j\rangle\}_\Omega \mathbf{w} \stackrel{d}{=} \{\langle \phi_i, W\rangle\}_\Omega, \tag{4.12}$$

for $i = 1, \ldots, m$, where m is the number of test functions. The finite dimensional solution is obtained by finding the distribution for the Gaussian weights in equation (4.9) that fulfils (4.12) for only a specific set of test functions, with $m = n$. When $\phi_k = (\kappa^2 = \Delta)^{1/2}\varphi_k$ for $\alpha = 1$ and $\phi_k = \varphi_k$ for $\alpha = 2$, these two approximations are denoted the *least squares* and the *Galerkin* solution, respectively. Choosing $\alpha = 2$ and $\phi_k = \varphi_k$ yields

$$\left(\kappa^2\{\langle \varphi_i, \varphi_j\rangle\} + \{\langle \varphi_i, -\Delta\varphi_j\rangle\}\right) \mathbf{w} \stackrel{d}{=} \{\langle \varphi_i, W\rangle\}. \tag{4.13}$$

Define the $n \times n$ matrices, $\mathbf{C}$ and $\mathbf{G}$ as

$$\begin{aligned} C_{ij} &= \langle \varphi_i, \varphi_j\rangle \\ G_{ij} &= \langle \nabla\varphi_i, \nabla\varphi_j\rangle, \end{aligned} \tag{4.14}$$

then a weak solution to (4.8) is given by (4.9), where

$$(\kappa^2\mathbf{C} + \mathbf{G})\mathbf{w} \sim N(0, \mathbf{C}) \tag{4.15}$$

and the precision of the weights, $\mathbf{w}$, is

$$\mathbf{Q}_{\alpha=\mathbf{2}} = (\kappa^2\mathbf{C} + \mathbf{G})^T\mathbf{C}^{-\mathbf{1}}(\kappa^2\mathbf{C} + \mathbf{G}). \tag{4.16}$$

Although G_{ij} and C_{ij} are sparse matrices, $\mathbf{C}^{-1}$ is not sparse. The solution is to replace $C_{ij} = \langle \varphi_i, \varphi_j\rangle$ by the diagonal matrix $C_{ii} = \langle \varphi_i, 1\rangle$, that yields a Markov approximation. Therefore, $\mathbf{w}$ is a GMRF with precision matrix defined by (4.16).

Replacing the GF $\omega(s,t)$ by the GMRF approximation $\tilde{\omega}(s,t)$ in equation (4.17), and approximating the integral in (4.7) by a quadrature rule, results that the approximate likelihood consists of $(n + n_t)T$ independent Poisson random variables, where n is the number of vertices and n_t is the number of observed point processes. By obtaining the LGCP likelihood approximation, it is possible to perform inference procedures through the INLA algorithm, which provides accurate and efficient approximations on Bayesian hierarchical models that can be represented

as latent Gaussian models. For reasons of space, we do not detail the INLA method here, which can be found in (RUE; MARTINO; CHOPIN, 2009).

The dynamic formulation proposed in this paper is a generalization of the formulation given in Equation (4.17). In this case, we include the components μ_t and $s_t$as follows:

$$\begin{aligned} Y(s,t) &= Poisson(|e(s,t)|exp(\lambda(s,t)), \\ \lambda(s,t) &= \mu_t + s_t + z(s,t)\beta + \xi(s,t) \\ \Delta^2\mu_t &= \mu_t - 2\mu_{t+1} + \mu_{t+2} \\ s_t &= s_{t-4} + \eta_s \\ c_t &= \theta_1 c_{t-1} + \theta_2 c_{t-2} + \eta_c \\ \xi(s,t) &= \Phi\xi(s,t-1) + \omega(s,t) \end{aligned} \tag{4.17}$$

where μ_t is the long term trend modeled as a second-order random walk (RW2), also known as the local-level model. The s_t represents the seasonal components, which is based on a seasonal autoregressive model. The c_t is a cycle component represented by an second-order autoregressive process with possible complex roots, which allows to capture cyclic patterns if the roots are complex numbers (LAURINI, 2019). The η_μ, η_c and η_s are nonspatial independent innovations with $\eta_\mu \sim N(0, \sigma^2_{\eta_\mu})$, $\eta_c \sim N(0, \sigma^2_{\eta_c})$ and $\eta_s \sim N(0, \sigma^2_{\eta_s})$. In all estimation procedures, we use default priors for the SPDE model in the R-INLA package implementation, which is available upon request from the authors.

4.2.2 Data

In this paper we use daily data of fire occurrence in the Legal Amazon from MODIS Thermal Anomalies/Fires between January 2002 and December 2022, which provides information such as fire occurrences (day/night), fire location, the logical criteria for the fire selection, and detection confidence. In order to provide better interpretations of the results, we use a quarterly aggregation of the daily data. In addition, from the computational aspect, the use of a very high frequency could lead to numerical problems in the estimation and inference processes since the dimension of the spatio-temporal covariance matrix is given by the Kronecker product between the time and spatial dimensions.

To illustrate, Figure 4.1 provides the number of fire events over time in the Legal Amazon, while Figure 4.2 shows a graphical distribution of the fires over time and space. From July to October 2005 large areas of the Amazon region experienced one of the most strong drought of the past 100 years (MARENGO et al., 2008). The event in 2005 was driven by elevated tropical North Atlantic sea surface temperatures associated with a weaker cold anomaly in the South Atlantic (MARENGO et al., 2008; COX et al., 2008), and caused intense forest fire. After the peak in 2005, the fire occurrence in the Legal Amazon decreased until 2012, whereas from 2013 to 2022 forest fires increased (see Figure 4.1). The spatial distribution of fire occurrence shows that forest fires are more concentrated in the region called “arc of deforestation”, an area that extends from Maranhão to Acre, but with a pattern of increasing toward central areas. Additionally, it is possible to note that most of the fire events occur during the third and fourth quarter, the dry season (May to October).

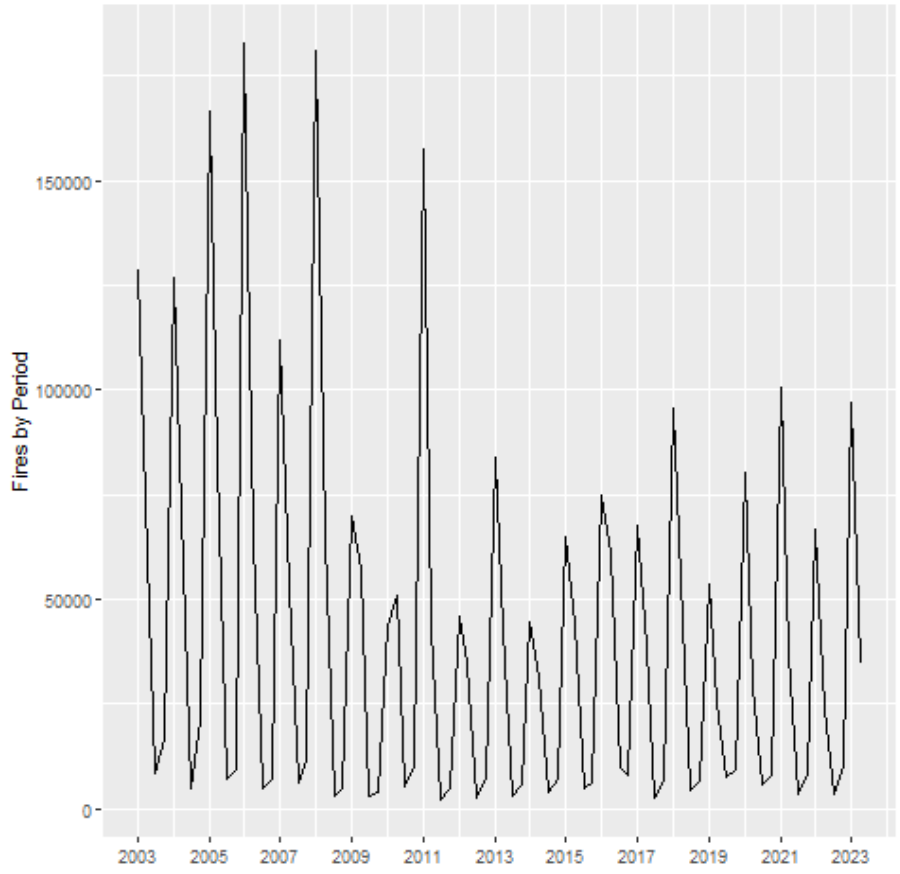


Figure 4.1 – Fires in Legal Amazon by quarter between 2002 and 2022

Since our data base includes fire occurrence of different causes, such as human sources (deliberately or accidentally), and natural causes, it is important to include explanatory variables in the analysis to control the main fixed effects related to climatic conditions and to control for possible use of the soil in agricultural and livestock activities. Thus, to reach our goal, we include explanatory variables, as the Koppen climate Classification, which classifies the climate in a certain region by types (see Appendix) that are characterized by two or three characters, where the first indicate the climate zone defined by the temperature and rainfall, the second is defined by the rainfall distribution, and the third considers the sea seasonal temperature variation (ALVARES et al., 2013). According to Köppen Classification, the climate in the Legal Amazon is mostly wet tropical (Am) in the central areas, tropical with dry winter (Aw) in the South Eastern Amazon and tropical without dry season (Af) in Western Amazon.

We also include MapBiomas Collection 7 for Amazon biome that contains annual land use and land cover maps (LULC), that classifies the Amazon Biome into 7 different classes of land cover/land use including forest formation, savanna formation, wetland, grassland formation, pasture, agriculture, other non-vegetated area, non observed, and water bodies[1].

Evidence of intentional fire can be seen through the proximity of fire outbreaks and highways, as proximity to highways implies in human accessibility and lower transportation costs for agricultural and livestock production. Therefore, as explanatory variable, we also include the dis-

[1] Available at <https://mapbiomas.org/>

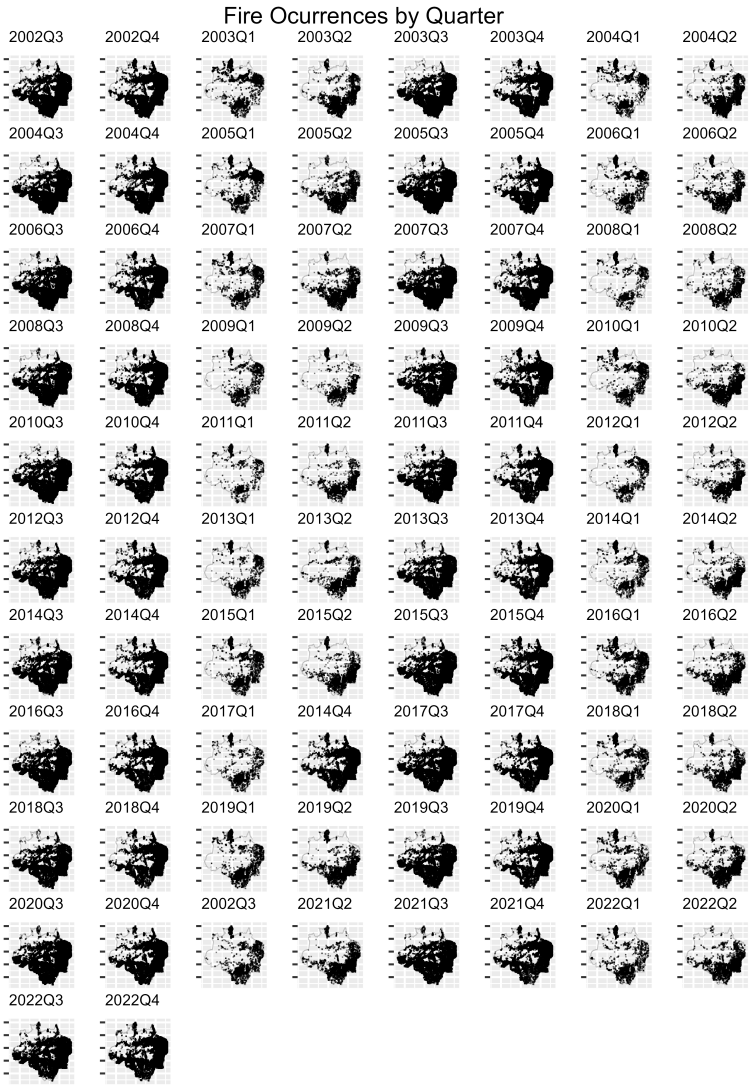


Figure 4.2 – Spatial distribution of fires in Legal Amazon between 2002 and 2022

tance of fire occurrence from federal and state highways. The data base containing the location of federal and state highways is provided by *Departamento Nacional de Infraestrutura de Transportes* (DNIT) and *Empresa de Planejamento e Logística* (EPL).

Lastly, we include rainfall and maximum temperature information which were constructed using the time series of the monitoring stations provided by *Agência Nacional de Águas* and *Instituto Nacional de Meteorologia* (INMET), whereas the maximum temperature data were obtained based on the information provided by the INMET. In both cases, we used the spatially continuous projections for each period in the sample, which were calculated based on the methodology proposed by Laurini (2017).

4.3 Results

We perform inference procedure based on the the specification described in Equation (4.17). Thus, the estimated parameters are the precision of the trend component ($1/\eta_\mu$), seasonal component ($1/\eta_s$), and cycle component ($1/\eta_c$), the parameters of the second-order autoregressive process of the cycle (PACF1 and PACF2), the parameters associated with the set of observed co-

variates (β), the parameters of spatial covariance ($\log \tau$ and $\log \kappa$), and the parameter of spatial time dependence (Φ).

Table 4.1 reports the estimated parameters. As might be expected, the results indicate a negative relation between the distance to roads and the fire occurrence. The importance of the highways as a prime driver of fire occurrence and deforestation at local scales has been discussed in the literature, showing that the roads play important roles facilitating transformation of land-use practices, creating fresh access to new settlements in frontier regions, and reducing transportation costs in earlier settled areas (FERRAZ, 2001; FEARNSIDE, 2006).

Regarding the rainfall and temperature covariates, the results indicate a negative relationship between rainfall and the intensity of fire occurrences and, on the other hand, higher temperatures are related with higher incidence of fires, according to the results.

Table 4.1 – Estimated Parameters for the Legal Amazon

	Mean	SD	0.025quant	0.5quant	0.975quant	Mode
Fixed effects						
Distance Highways	-0.012	0.001	-0.014	-0.012	-0.010	-0.012
Temperature	0.155	0.011	0.132	0.155	0.177	0.155
Rainfall	-0.006	0.001	-0.007	-0.006	-0.005	-0.006
Koppen 1 (Cwa)	0.012	0.248	-0.474	0.012	0.497	0.012
Köppen 2 (Am)	0.252	0.061	0.134	0.252	0.371	0.252
Köppen 3 (Af)	0.219	0.094	0.035	0.219	0.404	0.219
Köppen 4 (Cfa)	0.153	0.254	-0.344	0.153	0.650	0.153
Köppen 10 (As)	0.265	0.159	-0.045	0.265	0.576	0.265
Köppen 12 (Aw)	0.444	0.143	0.165	0.444	0.724	0.444
Forest Formation	-0.024	0.055	-0.131	-0.024	0.084	-0.024
Savanna Formation	0.056	0.084	-0.109	0.056	0.220	0.056
Mangrove	0.133	0.152	-0.166	0.133	0.432	0.133
Wetland	0.205	0.112	-0.015	0.205	0.426	0.205
Grassland	0.151	0.078	-0.003	0.151	0.304	0.151
Pasture	0.076	0.079	-0.078	0.076	0.230	0.076
Mosaic of Uses	0.078	0.128	-0.173	0.078	0.328	0.078
Beach, Dune and Sand Spot	0.542	0.248	0.055	0.542	1.029	0.542
Other non Vegetated Areas	-0.541	0.177	-0.888	-0.541	-0.195	-0.541
River, Lake and Ocean	0.154	0.066	0.024	0.154	0.284	0.154
Soybean	0.088	0.177	-0.258	0.088	0.435	0.088
Other Temporary Crops	0.116	0.169	-0.214	0.116	0.447	0.116
Random Effects						
Precision for trend	5.270	0.185	4.897	5.271	5.633	5.284
Precision for seasonality	1.059	0.082	0.933	1.049	1.251	1.010
PACF4 for seasonality	0.136	0.039	0.074	0.132	0.224	0.112
Precision for cycle	4.673	0.161	4.344	4.676	4.983	4.694
PACF1 for cycle	0.308	0.020	0.264	0.309	0.344	0.314
PACF2 for cycle	-0.366	0.017	-0.403	-0.365	-0.336	-0.360
Log τ	-2.135	0.007	-2.148	-2.136	-2.120	-2.137
Log κ	-0.035	0.008	-0.052	-0.034	-0.022	-0.031
Group Φ	0.842	0.003	0.837	0.842	0.847	0.841

As described earlier, the climate of the Legal Amazon, according to Köpper classification, are mostly wet climate, occurring precipitation in all months of the year (Af), monsoon, with a mean annual total precipitation > 1500mm and a dry season occurring between August and November (Am), and tropical with dry season (Aw). As expected, obtained results suggest that the types of climate with dry season (Am, As, and Aw) have higher influence on fire occurrence than those without dry season (Cwa, Af, and Cfa). Our analysis associated to land cover classifications show a positive relation between fire occurrence and savanna formation, mangrove, wetland, grassland, pasture, mosaic of uses, sand spot, water bodies, soybean, and other temporary crops. On the other hand, the estimated parameters indicate a negative relation between fire occurrence and forest formation and other non vegetated areas.

Regarding the random effects, the precision parameters represent the variability associated with the trend, seasonal and cycle components, where high values indicate low variability. Based on the results reported in the Table 4.1, it is possible to note a high precision associated with the cycle component as well as the trend component, whereas the seasonality component shows a relatively minor precision.

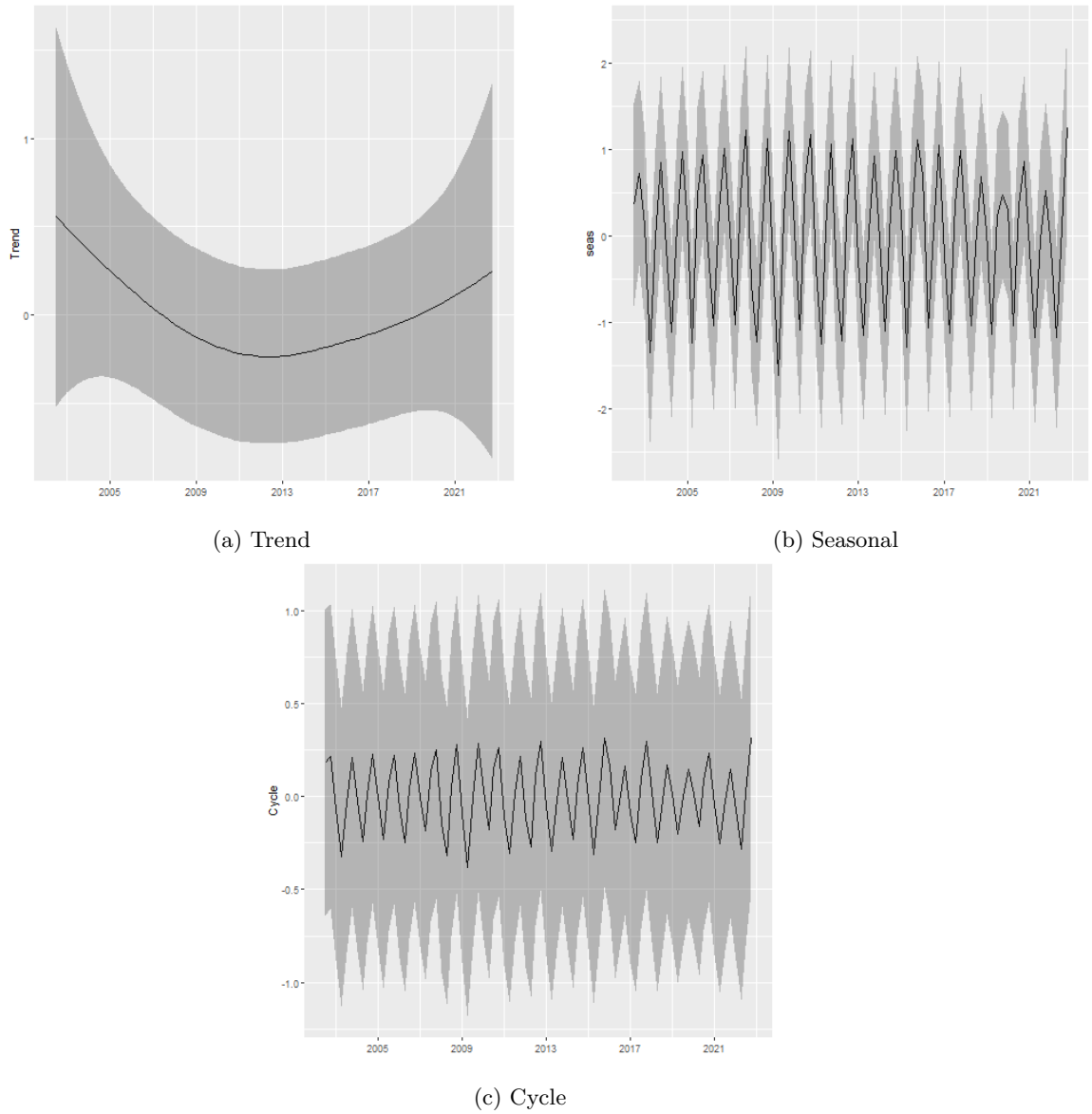
A primary empirical motivation for the present study was to assess the existence of changes in the patterns of fire occurrence in the Legal Amazon. To better understanding the results, we plotted the estimated trend, seasonal and cycle components (posterior mean and 95% Bayesian credibility interval; see Figure 4.3). The trend component exhibits a marked decrease between the beginning of the sample and 2012, followed by an increase that extents to the end of the sample. Regarding the cycle and seasonal components, based on Figure 4.3, it is possible to note that both are quite stable, and the model does not indicate relevant changes in those components.

The spatial heterogeneity of the fire occurrence in the Legal Amazon can be better seen through the estimated spatial random effect (posterior mean of estimated spatial random effect; see Figure 4.4).

In order to show the model's ability to fit the fire occurrence, we plotted the estimated log intensity function and the observed fire occurrence (black dots; see Figure 4.5), which shows that the estimated log intensity function explains the spatio-temporal variation observed in the fire count in the Legal Amazon, suggesting that the model has a good fit. Additionally, to show the importance of the trend, seasonal and cycle components in the analysis of fire occurrence in the Legal Amazon, we plotted the observed total fire count and the predicted value of fire count in each year given by the sum of the estimated trend, seasonal, cycle and intercept components (see Figure 4.6)

4.3.1 Amazon Biome

The region known as Legal Amazon in Brazil comprises nine Brazilian states, containing three different biomes: Amazon, Cerrado and Pantanal. These biomes differ from each other not only in vegetation and fauna, but in the way they provide ecosystem services. As a complementary analysis, we did the same previous analysis, but now considering the Amazon biome, in order to uncover possible changes in the patterns only in this biome. The results (see Table 4.2) obtained with the analysis of the Amazon biome differ from the previous one mostly in terms



(a) Trend

(b) Seasonal

(c) Cycle

Figure 4.3 – Trend, Seasonal and Cycle decomposition of fire occurrences in the Legal Amazon

of land cover. While we consider the three biomes in the analysis, the relationship between fire occurrence and wetland, grassland, pasture and soybean were positive. On the other hand, when we consider only the Amazon biome, these relationships become negative.

In the case of wetlands and grasslands, this result can be attributed to the fact that the Pantanal biome is the major wetland ecosystem of the world, characterized by the well-defined dry and wet seasons. Also, the Panantal and Cerrado are biomes in which fire-dependent ecosystems (savanna and grassland) predominates, i.e., in these type of formation fires are typically mild and frequent, often occurring in the transitional months between seasons, mostly during dry seasons, and providing benefits to the fauna and flora (PIVELLO, 2011; PIVELLO et al., 2021). Differently from the Pantanal and Cerrado, the Amazon biome is covered predominantly by dense forest formation, which is considered fire-sensitive. As a consequence, in the absense of Pantanal and Cerrado biomes in the analysis, when we consider only the Amazon biome, due to its features, the relationship between the intensity of fire activity and cerrado, savanna, and

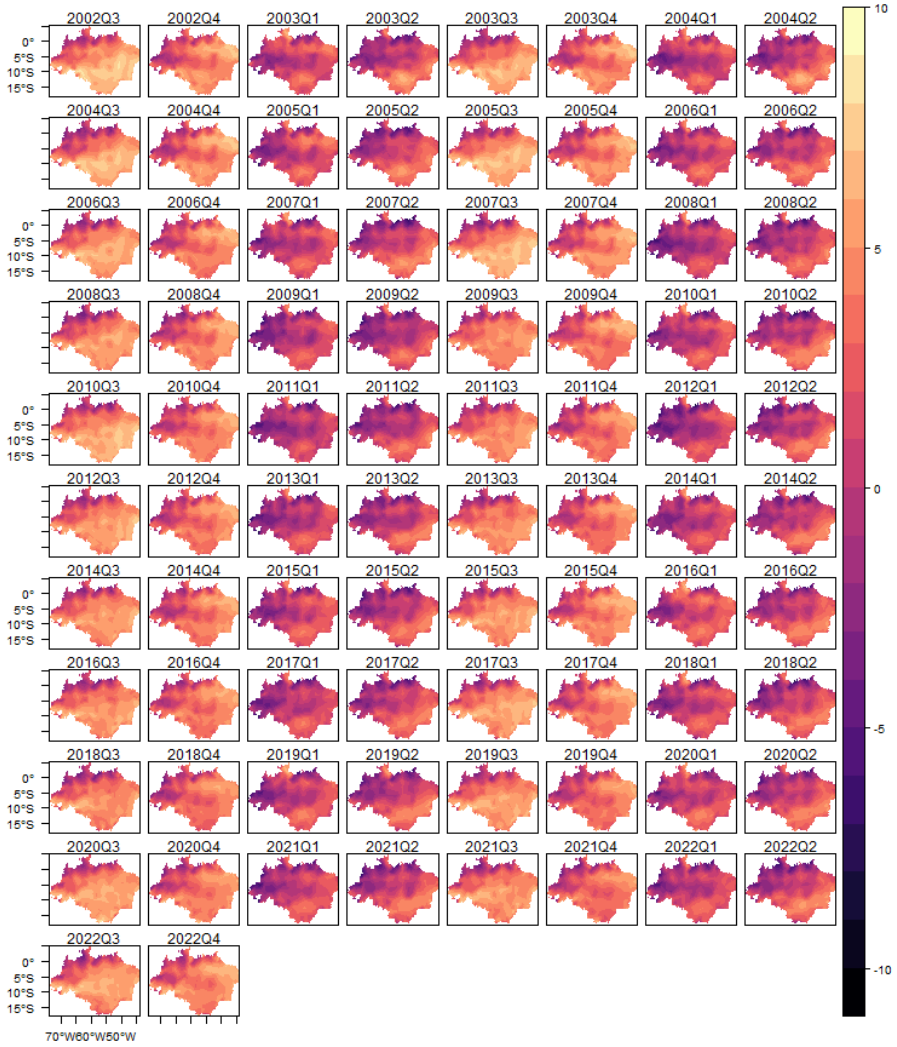


Figure 4.4 – Spatial Random Effects - Legal Amazon

wetland become negative.

Regarding the pasture and soybean fields, according to Oliveira et al. (2022), in the Cerrado and Pantanal, the climate is the major determinant of fire activity, while human action is the main driver factor in the Amazon biome. In this case, the incidence of accidental fires in pastures and agricultural areas caused by climate variables is higher in the Panantal and Cerrado than in the Amazon biome. As a consequence, when we analyze only the Amazon biome, the relationship between the intensity of fire activity and pasture and soybean fields become negative.

Figure 4.7 shows the posterior mean and 95% Bayesian credibility interval for the estimated trend, seasonal and cycle components considering the data for the Amazon biome. As the previous result, the trend component also exhibits a marked decrease between the beginning of the sample and 2012, followed by an increase that extents to the end of the sample. Similarly, the cycle and seasonal components are also quite stable, and the model does not indicate relevant changes in those components.

The spatial heterogeneity of the fire occurrence in the Amazon biome (see Figure 4.9 in Appendix) is also very similar to the previous analyze, as well as the estimated log intensity

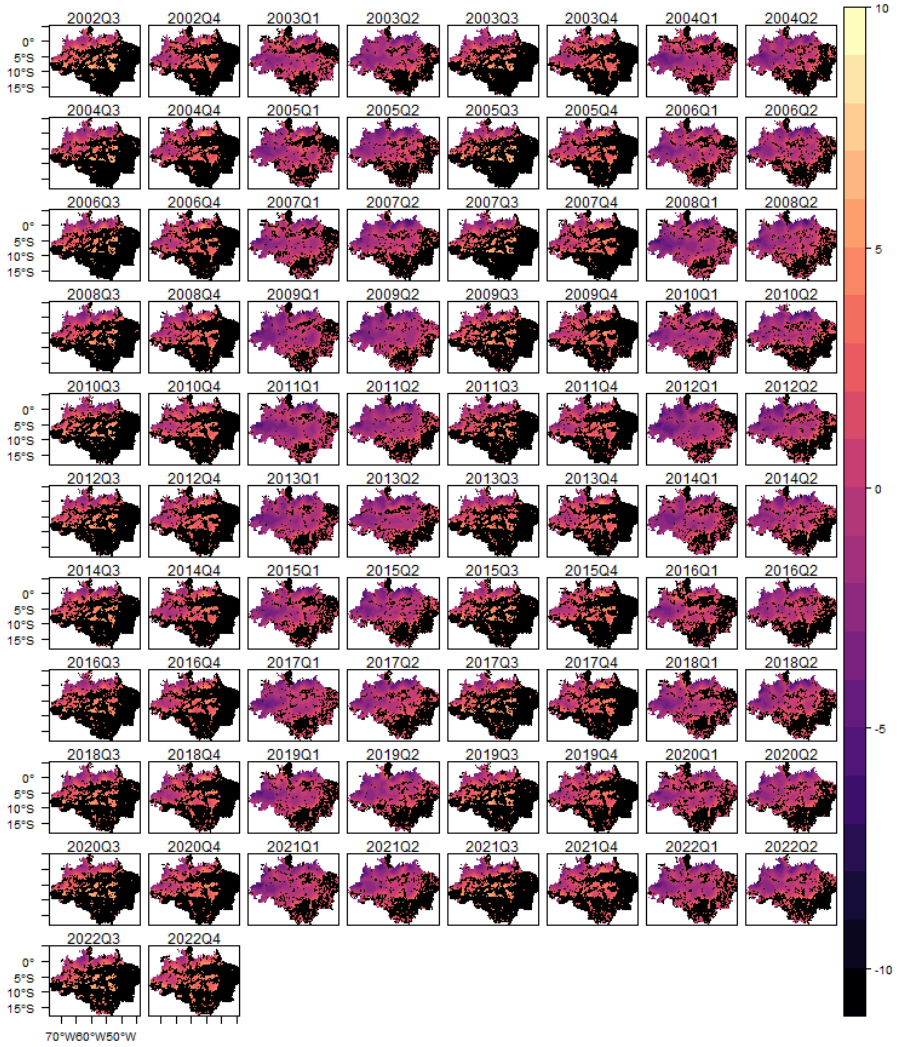


Figure 4.5 – Estimated log-intensity function and observed fire occurrence - Legal Amazon

function and the observed fire occurrence (black dots; see Figure 4.10 in Appendix), and the observed total fire count and the predicted value of fire count in each year given by the sum of the estimated trend, seasonal, cycle and intercept components (see Figure 4.11 in Appendix).

4.3.2 Monthly Data

In order to consolidate our results, we also provide a monthly analysis of the changes in the patterns of fire intensity in the Legal Amazon, and the results are presented in Table 4.4 in Appendix. With respect to the estimated trend, seasonality, and cycle components, the monthly analysis revealed the same patterns observed in the previous analysis, however, as expected, with less uncertainty. Lastly, considering the observed total fire count and the predicted value of fire count in each year given by the sum of the estimated trend, seasonal, cycle and intercept components (see Figure 4.12 in Appendix), it also shows the importance of the trend, seasonal and cycle components in the analysis of fire occurrence in the Legal Amazon.

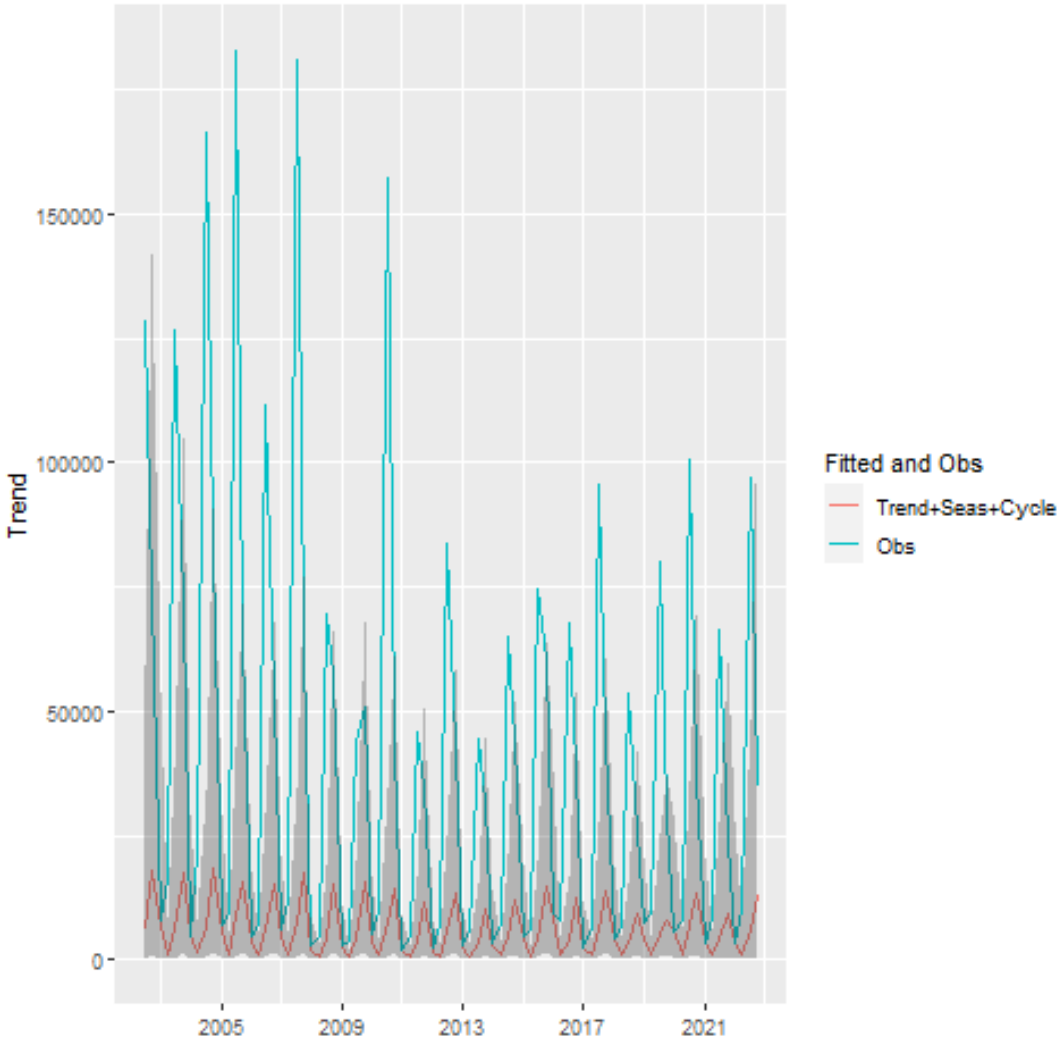


Figure 4.6 – Predicted fires given by the sum of trend, seasonality and cycle components and observed fires. Shaded areas in the graph represent the 95% Bayesian credibility interval.

4.4 Discussion

Our evidence suggests that there is a variability in the intensity of fire occurrence in both the Legal Amazon (considering Amazon, Cerrado and Pantanal biomes) and only in the Amazon biome. This result corroborates with the results presented by Libonati et al. (2021), which brings an analysis of the relationships between deforestation, fire, and droughts in the Brazilian Amazon.

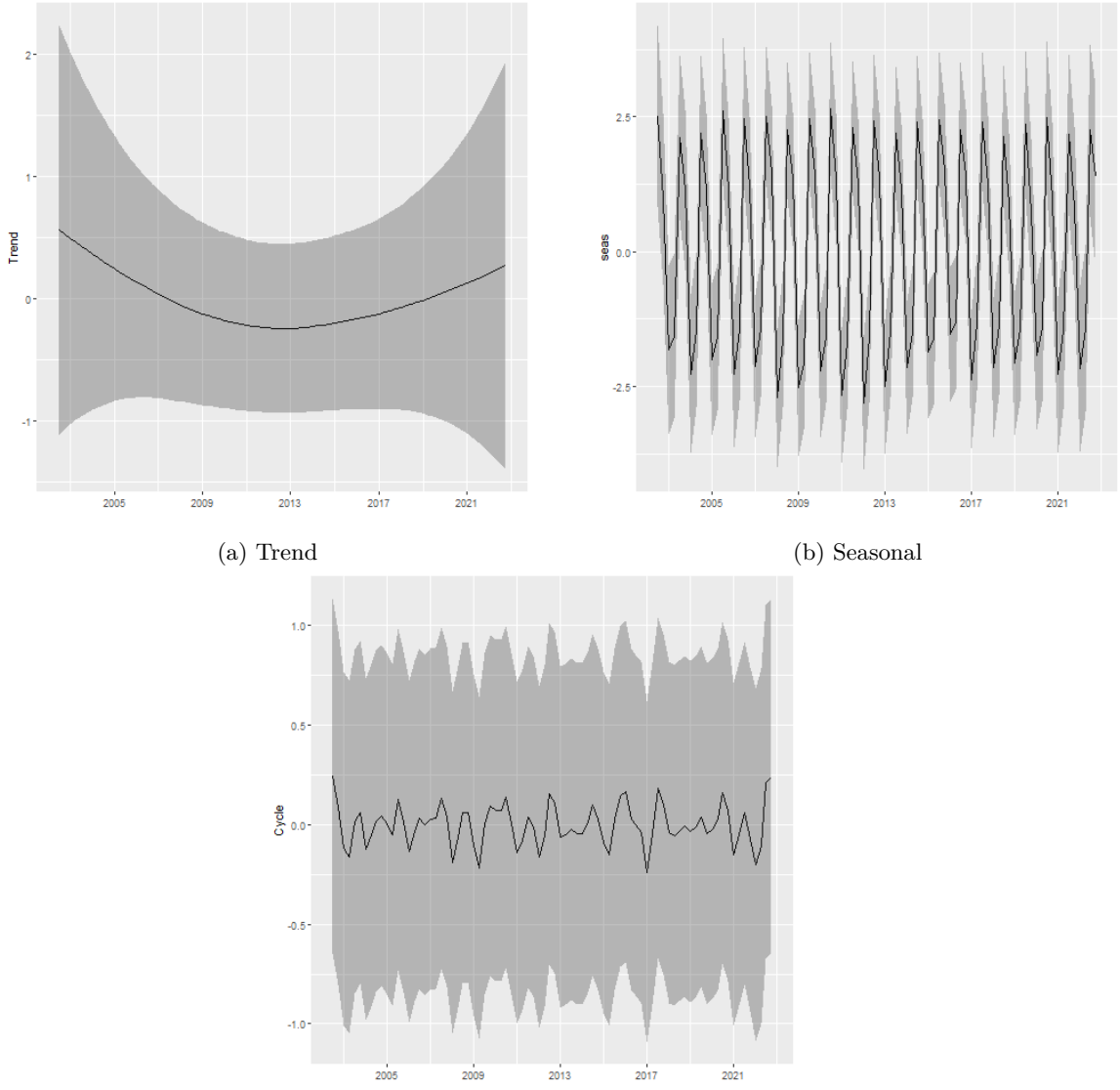
Our analysis of the trend component revealed a decrease in the fire activity between 2002 and 2012, which can be primarily attributed to governance measures along with market mechanisms. Indeed, the set of policies adopted to reduce the deforested area underwent significant revisions during 2000s, introducing innovative procedures for monitoring, environmental control, and territorial management, such as the Action Plan for the Prevention and Control of Deforestation in the Legal Amazon (PPCDAm), which launched its first and more successful phase in 2004. Additionally, novel policy measures were implemented in 2008, targeting municipalities with

Table 4.2 – Estimated Parameters for the Amazon biome

	Mean	SD	0.025quant	0.5quant	0.975quant	Mode
Fixed effects						
Distance Highways	-0.011	0.001	-0.014	-0.011	-0.009	-0.011
Rainfall	0.003	0.001	0.002	0.003	0.004	0.003
Koppen 1 (Cwa)	-0.088	0.431	-0.932	-0.088	0.756	-0.088
Köppen 2 (Am)	0.325	0.065	0.197	0.325	0.454	0.325
Köppen 3 (Af)	0.357	0.087	0.186	0.357	0.529	0.357
Köppen 4 (Cfa)	0.534	0.209	0.123	0.534	0.945	0.534
Köppen 10 (As)	-0.391	0.158	-0.701	-0.391	-0.082	-0.391
Köppen 12 (Aw)	0.238	0.136	-0.028	0.238	0.503	0.238
Forest Formation	-0.094	0.158	-0.404	-0.094	0.215	-0.094
Savanna Formation	0.001	0.169	-0.331	0.001	0.333	0.001
Mangrove	0.298	0.252	-0.196	0.298	0.792	0.298
Forest Plantation	0.530	0.561	-0.569	0.530	1.629	0.530
Wetland	-0.418	0.195	-0.799	-0.418	-0.037	-0.418
Grassland	-0.105	0.165	-0.430	-0.105	0.219	-0.105
Pasture	-0.071	0.159	-0.383	-0.071	0.241	-0.071
Beach, Dune and Sand Spot	0.064	0.234	-0.394	0.064	0.522	0.064
Urban Area	0.532	0.571	-0.587	0.532	1.651	0.532
Salt Flat	-0.146	0.358	-0.848	-0.146	0.556	-0.146
River, Lake and Ocean	0.101	0.156	-0.204	0.101	0.407	0.101
Soybean	-0.053	0.181	-0.408	-0.053	0.302	-0.053
Other Temporary Crops	0.042	0.189	-0.328	0.042	0.413	0.042
Random Effects						
Precision for trend	5.622	0.453	4.817	5.591	6.604	5.506
Precision for seasonality	0.421	0.034	0.361	0.418	0.493	0.411
PACF4 for seasonality	0.769	0.016	0.739	0.768	0.801	0.765
Precision for cycle	4.033	0.408	3.270	4.020	4.875	4.007
PACF1 for cycle	0.149	0.048	0.055	0.148	0.246	0.144
PACF2 for cycle	-0.379	0.057	-0.478	-0.383	-0.255	-0.398
Log τ	-2.157	0.014	-2.182	-2.157	-2.129	-2.158
Log κ	-0.125	0.012	-0.148	-0.126	-0.099	-0.129
Group Φ	0.873	0.004	0.863	0.873	0.881	0.874

critical rates of deforestation and constraining rural credit (ASSUNÇÃO et al., 2015). The effectiveness of the policies and how the market mechanisms have impacted the deforestation in the Legal Amazon has been widely discussed in the literature, showing that, in general, the conservation policies, the decreases in agricultural prices, and the availability of rural credit has curbed deforestation (e.g., Assunçãoa et al. (2013), Hargrave and Kis-Katos (2013), Lapola et al. (2014), Nepstad et al. (2014), Assunção et al. (2015), Marle et al. (2017)).

We also showed that from 2013 onwards there is an increasing pattern in the long term component, which appears to reflect the increasing number in the international markets opened for Brazilian beef and soy during this period, increasing the pressure on forests. Additionally, the Brazil's Forest Code (Law 4771/1965) was replaced by Law 12651/2012, reducing restrictions and pardoning areas of illegal clearing done by 2008, causing significant environmental and social



(a) Trend

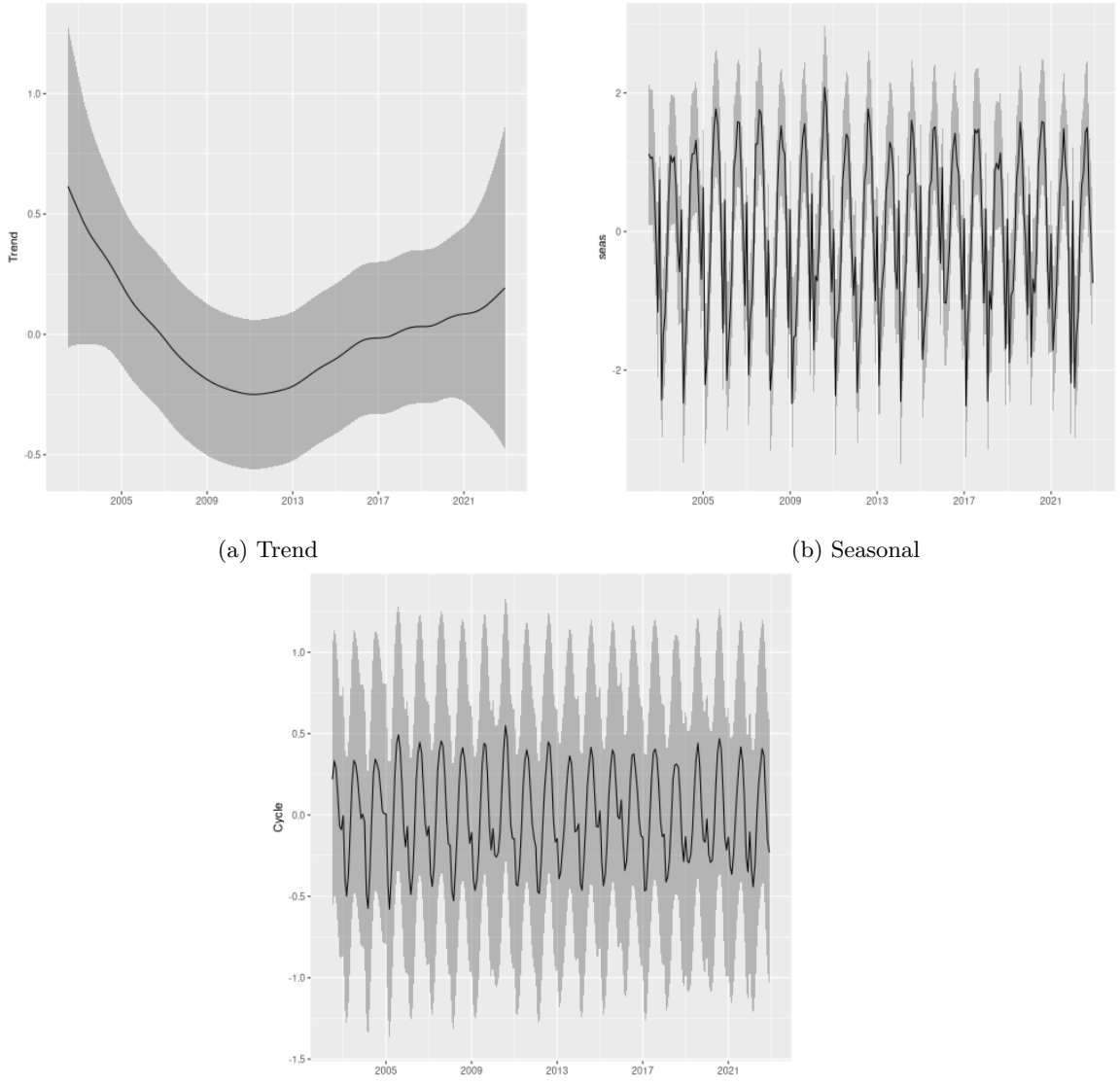
(b) Seasonal

(c) Cycle

Figure 4.7 – Trend, Seasonal and Cycle decomposition of fire occurrences in the Amazon biome.

issues (FEARNSIDE, 2017).

By including covariates in the model we were able to investigate the relationships between the intensity of fire occurrences, human activities, and climate variables over time. Essentially, four factors are required for fire to occur: sufficient biomass, biomass available to burn, ambient conditions conductive to spread, and ignitions (BRADSTOCK, 2010). These factors rely on meteorological conditions and how these conditions interact with different kinds of vegetation. Our findings suggest that climate patterns and human activities are underlying factors that have driven the trend of the fire occurrence. Supporting the evidence that the intensity in the fire occurrence have been related to climate variability, our model provides evidence of the positive relationship between temperature and fire activity, and a negative relationship between rainfall and fire occurrence. Additionally, the human influence may further complicate the influence of biophysical drivers, through land-use modification, human ignitions, and fire suppression (ANDELA et al., 2017). In this context, our results provided evidence that pasture, mosaic



(a) Trend

(b) Seasonal

(c) Cycle

Figure 4.8 – Trend, Seasonal and Cycle decomposition of fire occurrences in the Legal Amazon.

of uses, soybean and other temporary crops are positive related with fire intensity, whereas other non vegetated areas are negatively related with fire occurrence. Indeed, humans have manipulated fire regimes for several thousand years, suppressing wildfires to protect lives and properties, and creating landscapes that inhibit large-scale fire spread. As a consequence, such anthropogenic influences result in fire regimes that differ in terms of frequency, severity, and seasonality from how ecosystems would burn in the absence of humans (RABIN et al., 2015; SYPHARD et al., 2017).

From the spatial perspective, it is possible to observe that the spatial random effects capture the variability in the Legal Amazon, especially in the regions classified as wet tropical (Am), which is characterized by a dry season, that occurs between August and November (third and fourth quarters), and tropical with dry season (Aw). On the other hand, in western Amazon, where the climate is predominantly tropical without dry season (Af), the variability is low.

4.5 Conclusion

The Amazon biome plays an important role in the climate system, with relevance at regional and global scales. Fire occurrences, related to both natural and anthropogenic activities, are relevant disturbances in the Legal Amazon, with significant effects. Changes in the patterns of fire occurrence in the Amazon region have been widely reported in the literature and are related with a variety of factors, including dry conditions, deforestation, agricultural expansion, climate changes, and climatic anomalies such as El Ninõ events.

The purpose of this paper was to analyze the existence of changes in the patterns of the fire occurrence in the Legal Amazon, within the spatio-temporal point process framework. To perform inference procedures, we proposed a structural decomposition to spatio-temporal point pattern data. In particular, we proposed to use a dynamic representation of a Log Gaussian Cox process where the intensity function was modeled through the decomposition of components into trend, seasonality, cycles, covariates and spatial effects. This useful formulation was able to capture permanent changes in the fire occurrence and also, to identify seasonal and cyclic effects. Plus, the resulting Bayesian hierarchical structure allowed us to perform inference in a computationally effective way within the integrated nested Laplace approximation framework.

We presented here, the results of analyzing data for fire occurrence in the Legal Amazon reported by MODIS, from January 2002 to December 2022. Also, we included explanatory variables to control the main fixed effect related to climatic conditions and the use of the soil in agricultural activities. Our results show that the estimated trend component of fire occurrence exhibits a marked decrease from 2002 to 2012, followed by an increase that extents to the end of the sample, which can be related to governance actions and market mechanisms. Furthermore, our model also was able to capture the variability in the Legal Amazon, especially in the regions classified as wet tropical (Am), which is characterized by a dry season, that occurs between August and November (third and fourth quarters), and tropical with dry season (Aw). On the other hand, in western Amazon, where the climate is predominantly tropical without dry season (Af), the variability is low.

4.6 Appendix

4.6.1 Climate classification

4.6.2 Amazon biome - Additional results

4.6.3 Monthly data - Additional results

Table 4.3 – Köppen Climate Classification

Characters	**Description**
Cwa	(C) Humid subtropical (w) With dry winter (a) and hot summer.
Am	(A) Tropical (m) monsoon.
Af	(A) Tropical (f) without dry season.
Cfa	(C) Humid subtropical (f) Oceanic climate, without dry season (a) and hot summer.
Cwb	(C) Humid subtropical (w) With dry winter (b) and temperate summer.
Csb	(C) Humid subtropical (s) With dry summer (b) and temperate summer.
Csa	(C) Humid subtropical (s) With dry summer (a) and hot summer.
Cfb	(C) Humid subtropical (f) Oceanic climate, without dry season (b) and temperate summer.
BSh	(B) Dry (S) Semi-arid (h) low latitude and altitude.
As	(A) Tropical (s) with dry summer.
Cwc	(C) Humid subtropical (w) With dry winter and (c) short and cool summer.
Aw	(A) Tropical (w) with dry winter.

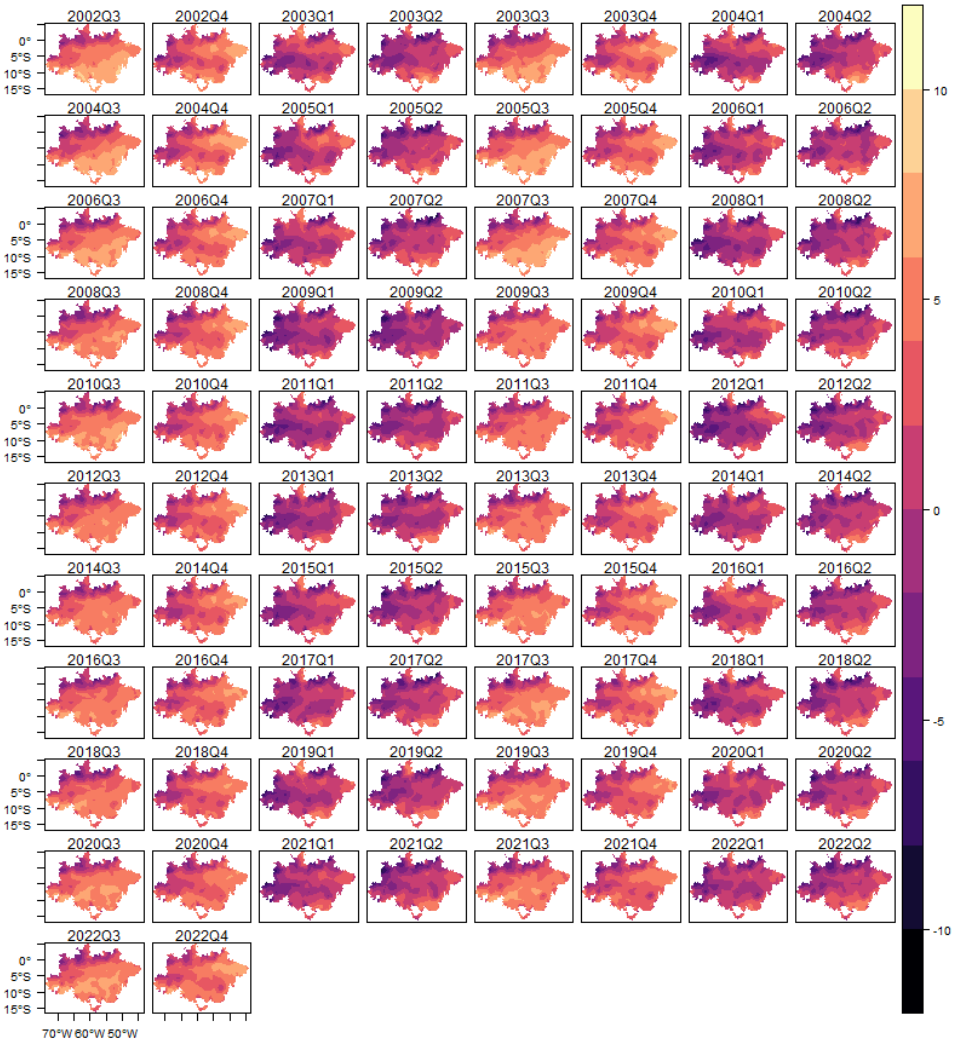


Figure 4.9 – Spatial Random Effects - Amazon biome

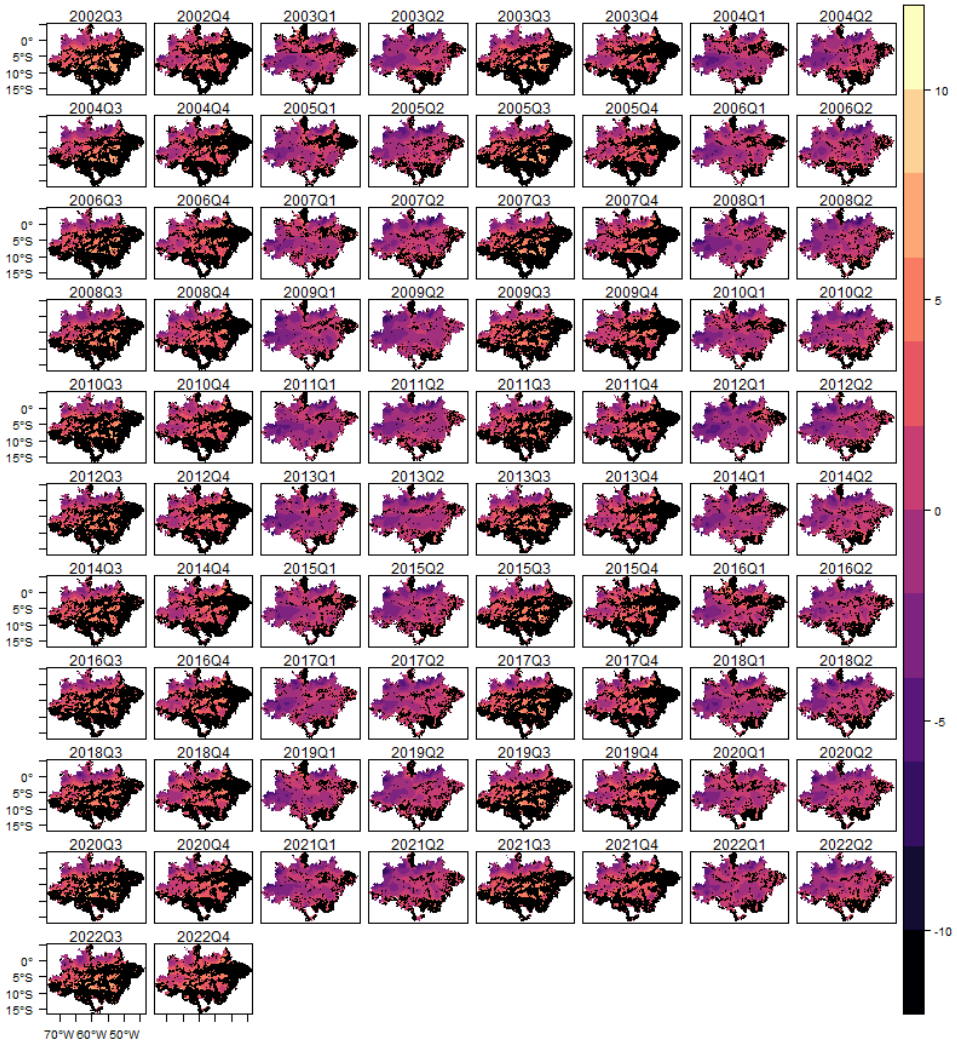


Figure 4.10 – Estimated log-intensity function and observed fire occurrence - Amazon biome

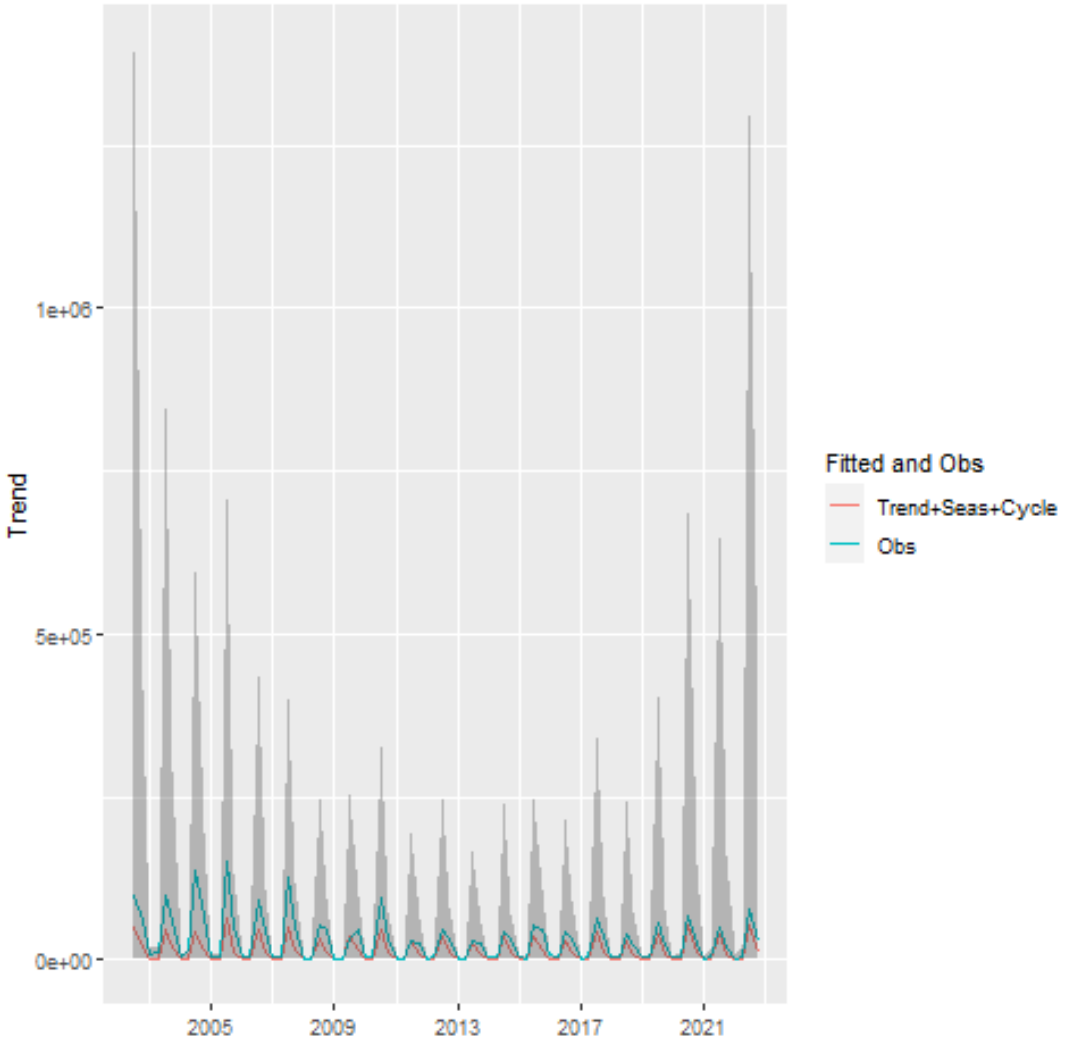


Figure 4.11 – Predicted fires given by the sum of trend, seasonality and cycle components and observed fires. Shaded areas in the graph represent the 95% Bayesian credibility interval.

Table 4.4 – Estimated Parameters for the Legal Amazon - Monthly data

	Mean	SD	0.025quant	0.5quant	0.975quant	Mode
Fixed effects						
Distance Highways	-0.013	0.001	-0.015	-0.013	-0.012	-0.013
Temperature	0.086	0.005	0.076	0.086	0.096	0.086
Rainfall	-0.014	0.001	-0.016	-0.014	-0.012	-0.014
Koppen 1 (Cwa)	-0.061	0.181	-0.415	-0.061	0.294	-0.061
Köppen 2 (Am)	0.306	0.044	0.220	0.306	0.391	0.306
Köppen 3 (Af)	0.030	0.065	-0.098	0.030	0.158	0.030
Köppen 4 (Cfa)	0.237	0.206	-0.167	0.237	0.642	0.237
Köppen 10 (As)	0.312	0.113	0.091	0.312	0.533	0.312
Köppen 12 (Aw)	0.736	0.090	0.559	0.736	0.913	0.736
Forest Formation	0.037	0.041	-0.043	0.037	0.117	0.037
Savanna Formation	0.140	0.062	0.019	0.140	0.261	0.140
Mangrove	0.085	0.111	-0.133	0.085	0.303	0.085
Wetland	0.243	0.090	0.065	0.243	0.420	0.243
Grassland	0.264	0.058	0.150	0.264	0.378	0.264
Pasture	0.211	0.059	0.097	0.211	0.326	0.211
Mosaic of Uses	0.302	0.108	0.091	0.302	0.513	0.302
Beach, Dune and Sand Spot	0.651	0.184	0.291	0.651	1.011	0.651
Other non Vegetated Areas	-0.617	0.157	-0.924	-0.617	-0.310	-0.617
River, Lake and Ocean	0.235	0.049	0.140	0.235	0.330	0.235
Soybean	0.372	0.135	0.107	0.372	0.638	0.372
Other Temporary Crops	0.251	0.144	-0.031	0.251	0.533	0.251
Random Effects						
Precision for trend	5.402	0.115	5.144	5.411	5.588	5.468
Precision for seasonality	0.908	0.013	0.883	0.908	0.936	0.906
PACF4 for seasonality	0.067	0.009	0.048	0.068	0.083	0.070
Precision for cycle	4.862	0.092	4.689	4.858	5.057	4.844
PACF1 for cycle	0.341	0.009	0.321	0.342	0.356	0.345
PACF2 for cycle	-0.349	0.008	-0.366	-0.349	-0.334	-0.347
Log τ	-2.250	0.006	-2.265	-2.249	-2.240	-2.245
Log κ	0.286	0.004	0.280	0.285	0.294	0.284
Group Φ	0.853	0.002	0.850	0.852	0.856	0.851

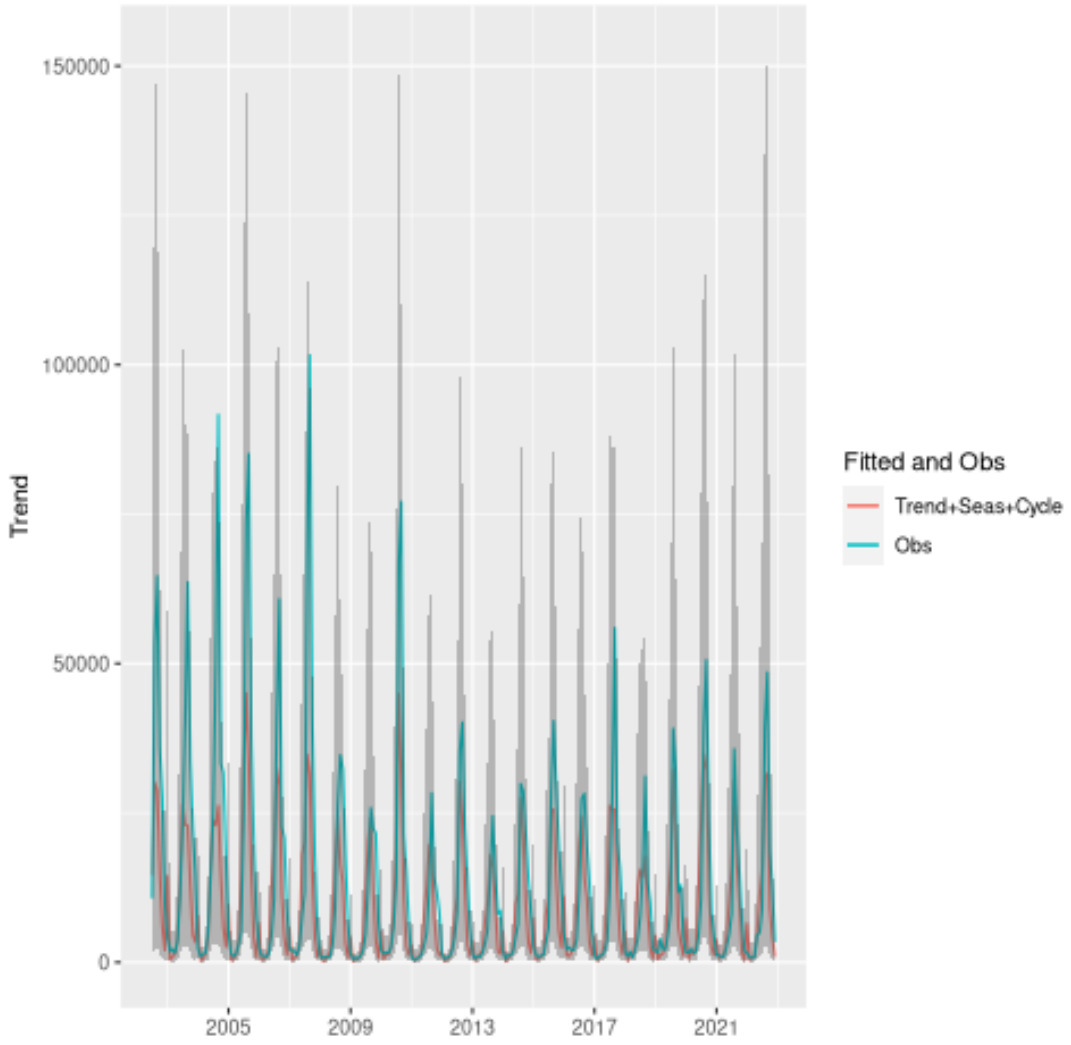


Figure 4.12 – Predicted fires given by the sum of trend, seasonality and cycle components and observed fires. Shaded areas in the graph represent the 95% Bayesian credibility interval.

5 THE DYNAMICS OF FIRE ACTIVITY IN THE BRAZILIAN PANTANAL

Fernanda Valente Márcio Laurini

ABSTRACT

We analyze the changes in the patterns of fire occurrence in the Brazilian Pantanal based on remote sensing data resources. To estimate the spatio-temporal dynamics of fire occurrence we use a structural representation in a log-Gaussian Cox process, decomposing the intensity function in components of trend, seasonality, cycles, covariates and spatial effects. The results obtained indicate that there are relevant variations in the trends of fires occurrence in the Brazilian Pantanal, the intensity of occurrence of fires is statistically higher in natural vegetation, and that a relevant part of the record of fires observed in the first three quarters of 2020 cannot be explained by climatic factors alone, possibly being caused by intentional human actions.

Keywords: Fires, Spatio-Temporal Point Process, LGCP, Brazilian Pantanal

5.1 Introduction

The Pantanal biome is the major wetland ecosystem of the world, located in the Upper Paraguay River Basin (UPRB) between the Cerrado and Amazon biomes. The Brazilian Pantanal is located in the southwest region, mostly in the state of Mato Grosso do Sul (65%), but also in the state of Mato Grosso (35%) (SILVA; ABDON, 1998). The biome is mainly characterized by the well-defined dry and wet seasons, producing periodic fluctuations in the water level (flood pulse), shaping the scope of terrestrial and aquatic places on the lowland, and influencing the fauna and flora. The vegetation in the Pantanal is heterogenous, with several vegetation classes identified, and also serving as the habitat for substantial populations of animals, including threatened species (TOMAS et al., 2019).

This high Pantanal's biological diversity has attracted great attention, making this biome more susceptible to anthropological threats. In the last few decades, the Pantanal has experienced a rapid evolution of the agricultural and livestock systems, replacing areas of natural vegetation by production zones (HARRIS et al., 2005; SILVA et al., 2011; ARVOR et al., 2012). The monitoring activities of 2012-2014 to assess the environmental impact in the UPRB have identified that 58% of the original vegetation in the plateau areas were converted to anthropic uses, whereas in the lowlands this conversion corresponds to 42%. In addition, the report also have found that 99% of all converted areas have been used as pastureland, while 0.6% for agriculture, and 0.4% for mining and urban areas (PANTANAL; WWF-BRASIL, 2015).

Related to agriculture and cattle ranching, the inclusion of exotic grass species and the burning practice are important threats in the region. As a consequence of replacing the original vegetation by cultivated pastures and the uncontrolled fires, the severe erosion has led to changes in the hydrological regimes and the patterns of water flow (HARRIS et al., 2005). As discussed by Ivory et al. (2019), changes in the vegetation productivity in these landscape are likely linked

to changes in rainfall and the flood-pulse, with different responses based on position relative to inundated areas. As a consequence, these changes in the vegetation cycle can lead to changes in the fire activity. Although occasional fire plays an important ecological role in the wetland, with positive impacts on some vegetation structure and nutrient recycling (OLIVEIRA et al., 2014; ARRUDA et al., 2016), fire is one of the most important environmental disturbances, affecting the variations of flood and dry periods, and, as a consequence, changing the required time of plants and animals to recover after the dry periods (JUNK et al., 2006).

Extensive and more frequent fire events has been reported in the Brazilian Pantanal in the last few years. For instance, in 2020, more than 22 thousand fire outbreaks were registered in the Pantanal, with a burned area of 33.000km², exceeding by 176% the historical record of fire outbreaks registered in 2005 since the beginning of the monitoring by the *Instituto Nacional de Pesquisas Espaciais* (INPE) in 1998. In addition, the area burnt in 2019 in the Brazilian Pantanal was surprisingly 996% higher than in 2018, which is particularly high when compared with the neighbor biomes (Cerrado and Amazon) that have recorded an increase around 40% and 65% of the burned area from 2018 to 2019, respectively. In addition, it is worth noting that 95.72% of the fire events in the Pantanal have occurred in native vegetation, whereas only 4.28% was in anthropized areas (MapBiomas, 2020). One reason for this increasing may be related to land use and climate changes, which has the potential to affect the rainfall intensity and the dry period, favoring the frequency of fire events, mostly human-induced (accidentally or deliberately), which tends to start in grasslands and then move to woodlands (ARRUDA et al., 2016). In the absence of specific legislation requiring that landowners restrict a minimum percentage of native vegetation cover in the farms, it is expected an average vegetation loss around 10% for the plateau and 3% for the lowland by 2050 (GUERRA et al., 2020b; GUERRA et al., 2020a).

Given the widely-expected trend of agriculture and livestock expansion, and the importance of the Brazilian Pantanal to provide ecosystem services and the economic valuation of the region (COSTANZA et al., 1997; SEIDL; MORAES, 2000; COSTANZA et al., 2014), there is an urgency to evaluate the possible changes in the patterns of fire occurrence in the region, in order to find solutions to minimize the impacts. Despite the existent long-term studies evaluating the variations in rainfall patterns in the Pantanal and the related consequences for fire frequency, the study of climate landscape dynamics is underdeveloped in the Pantanal (SCHULZ et al., 2019). In particular, for fire occurrences, there is a gap in research that our study aims to help fill. Additionally, the development of remote sensing technology have became an important tool to assessing fire occurrences and monitoring the possible changes in the patterns of these events. In this sense, to monitor the patterns of fire occurrence in the Brazilian Pantanal, we propose to model the counting events geographically located (fire spots) based on remote sensing data resources through trend-cycle decomposition for spatio-temporal point process models. In particular, we use a dynamic representation of a log Gaussian Cox process (LGCP), where the intensity function is modeled through decomposition of components in trend, seasonality, cycles, covariates and spatial effects (LAURINI, 2019; VALENTE; LAURINI, 2020; VALENTE; LAURINI, 2021a; VALENTE; LAURINI, 2021c), assuming that spatial effects are time varying, based on an autoregressive functional structure.

5.2 Data and Methods

Our main goal is to analyze the changes in the permanent and transitory patterns of fire occurrence in the Brazilian Pantanal biome, which was delimited based on The Map of Biomes and Coastal-Marine System of Brazil from *Instituto Brasileiro de Geografia e Estatística* (IBGE)[1]. For that, we use daily data of fire spots in the Brazilian Pantanal from July-1998 to December-2022, provided by the Programa Queimadas[2], from Brazilian National Institute of Spatial Research (*Instituto Nacional de Pesquisa Espacial*-INPE). We also included covariates that could be important in the fire observations since our data set includes fire occurrences of different causes, such as human sources and natural causes, that can be influenced by climate variables. In particular, we included information on maximum temperature, rainfall, and land use/land cover (LULC) from 1998 to 2022. A detailed description can be found in the Appendix.

To properly deal with point pattern data we use a spatio-temporal decomposition based on a LGCP, a doubly stochastic version of the Poisson process, where the log of the intensity function is given by a Gaussian random field. This framework is a flexible way to overcome the limited structure of the Poisson process, by allowing to introduce more complex stochastic structures in the intensity function, controlling for general processes of spatial dependence. However, inference procedures on the LGCP are difficult given the fact that the likelihood of theses processes is analytically intractable. To bypass this problem, Simpson et al. (2016) proposed to approximate the LGCP likelihood through the stochastic partial differential equation (SPDE) approach (LINDGREN; RUE; LINDSTRÖM, 2011), which is a computationally effective way to deal with spatio-temporal models in the context of point pattern data. In addition, as proposed by Laurini (2017) and Valente and Laurini (2020), the LCGP structure allows us to estimate long-term changes and transient components through a trend-cycle decomposition (à la Harvey (1990)) of the intensity function. In particular, we proposed to decompose the intensity function into latent factors of trend, seasonality, and cycle, along with covariates and spatial effects.

In the analysis of the temporal pattern, the main object is the trend component, which shows the evolution of the average level of occurrences over time, and thus shows the persistent patterns of fire occurrence in the Pantanal. Variations in this component may indicate variations that are possibly related to changes in the patterns of land use management in agricultural activities, such as the use of burning for the removal of native vegetation and later use in pastures and plantations. This interpretation is possible by controlling the climatic effects through covariates, and also by controlling other possible non-permanent effects by including the seasonality and cycle components. In addition to the inclusion of these common temporal and covariate components, the model used in the analysis includes a structure of time-varying spatial random effects, which allow the capture of the remaining spatial patterns, allowing to analyze also if there are other effects in the fire patterns that have spatial dependence.

To provide a clearer idea of the method employed to reach our goals, we present a brief description of the SPDE approach in the Appendix, following the notation adopted by Lindgren, Rue and Lindström (2011). Herein we provide some details about the model structure used in

[1] Available at <https://www.ibge.gov.br/geociencias/cartas-e-mapas/informacoes-ambientais/15842-biomas.html?=&t=o-que-e>.

[2] Data and more information available at http://queimadas.dgi.inpe.br/queimadas/portal

our analysis. Consider a dynamic version of a spatial LGCP where the intensity function is modelled as follows:

$$\begin{aligned} Y(s,t) &= Poisson(|e(s,t)|exp(\lambda(s,t)) \\ \log\lambda(s,t) &= \mu_t + s_t + c_t + z(s,t)\beta + \xi(s,t) \\ \mu_t &= \mu_{t-1} + \eta_\mu \\ s_t &= s_{t-1} + s_{t-2} + \ldots + s_{t-m} + \eta_s \\ c_t &= \theta_1 c_{t-1} + \theta_2 c_{t-2} + \eta_c \\ \xi(s,t) &= \Theta\xi(s,t-1) + \omega(s,t) \end{aligned} \tag{5.1}$$

where $Y(s,t)$ is the number of occurrences in a region s in time t, $e(s,t)$ is the exposure offset for the region s, μ_t is the long term trend, s_t represents the seasonal components, c_t is a cycle component represented by an second-order autoregressive process with complex roots, $z(s,t)$ is a set of covariates observed in the location s and period t, and $\xi(s,t)$ are the spatial random effects represented by the Gaussian process $\omega(s,t)$ continuously projected in space and given by

$$Cov[\omega(s,t),\omega(s',t')] = \begin{cases} 0 & \text{if } t \neq t' \\ \sigma^2 C(h) & \text{if } t = t' \end{cases} \qquad \text{for } s \neq s' \tag{5.2}$$

In this structure, the trend component is modeled as a first order random walk, which is a way widely used to model persistent components in time series models. The seasonality component is given by components that add up to zero within the year, which incorporate seasonal deviations from the series average in each period. The cyclic component is modeled with a second-order stationary autoregressive process, which is a parsimonious way of recovering periodic patterns in time series.

The spatial component is defined by a spatially continuous covariance function. By assumption, $C(h)$ is a covariance function of the Matérn class, which can be written as

$$C(h) = \frac{2^{1-\nu}}{\Gamma(\nu)}(\kappa||h||)^\nu K_\nu(\kappa||h||) \tag{5.3}$$

where $h = ||s - s'||$ is the Euclidean distance between locations s and s', $\kappa > 0$ is a spatial scale parameter, $\nu > 0$ is the smoothness parameter and K_ν is a modified Bessel function.

To sum up, our goal is to obtain the intensity function of a counting event, which is decomposed into components of trend, seasonality, cycle, plus the effect of covariates and the so-called spatial random effect, which captures the local effects not captured by the other components. By using this structure, we are able to assess possible changes in the patterns of fire occurrence in the Brazilian Pantanal. In addition, the reason to incorporate covariates in the analysis is to control the impact of climate variables and other effects related to the land use and land cover.

5.3 Results and Discussion

In this section, we discuss the results of the final specification for the analyzed model, where the statistically significant covariates were rainfall, given by the rainfall accumulated in the quarter, and land use/cover variables in each location analyzed. Regarding the maximum

temperature variable, it did not provide a statistically significant pattern to explain the variation in the fire events, even though there was a long-term increase in maximum temperature trends in the Pantanal region. Additionally, from the obtained results (see Table 5.2), it is possible to observe a negative relation between rainfall and the intensity of fire events, i.e., the lower precipitation leads to higher fire intensity. On the other hand, natural forest such as forest formation, savanna, and natural grassland has a positive effect on fire intensity. It is important to emphasize that other types of land use/cover were not statistically significant in terms of the intensity of the process, especially land use in agricultural activities.

It is not surprising that ecosystems formed by savannas and grasslands are positively related to fire intensity since they are mostly fire-dependent ecosystems (HARDESTY; MYERS; FULKS, 2005). Indeed, fires in these type of formation are typically mild and frequent, often occurring in the transitional months between seasons and providing benefits to the fauna and flora (PIVELLO, 2011; PIVELLO et al., 2021). However, in this kind of vegetation there is also the occurrence of anthropogenic fires, which is used to clean the field, control pests, and to stimulate the regrowth of grasses to cattle, increasing the fire frequency. The natural fires are usually controlled by the rainy season, while human-induced fires usually also occur in dry season and are more intense than natural fires, spreading more easily and without rain to extinguish it. As a consequence, recurrent human-induced fires can affect the spatial pattern and intensity of the fire activity even in the fire-dependent environments. In combination with drought, these alterations can cause severe and catastrophic fires, as those recorded in 2020 in the Pantanal (PIVELLO et al., 2021). Additionally, it is important to emphasize that riparian and gallery forests along water bodies in the grassland and savanna formation are classified as fire-sensitive environments, and can be gradually reduced when in contact with recurrent fires (ARRUDA et al., 2016). However, it is important to highlight that our results do not distinguish whether the fire occurrence is natural or caused by human actions, thus additional studies are needed to better elucidate the role of human influences in the fire activity in this kind of vegetation.

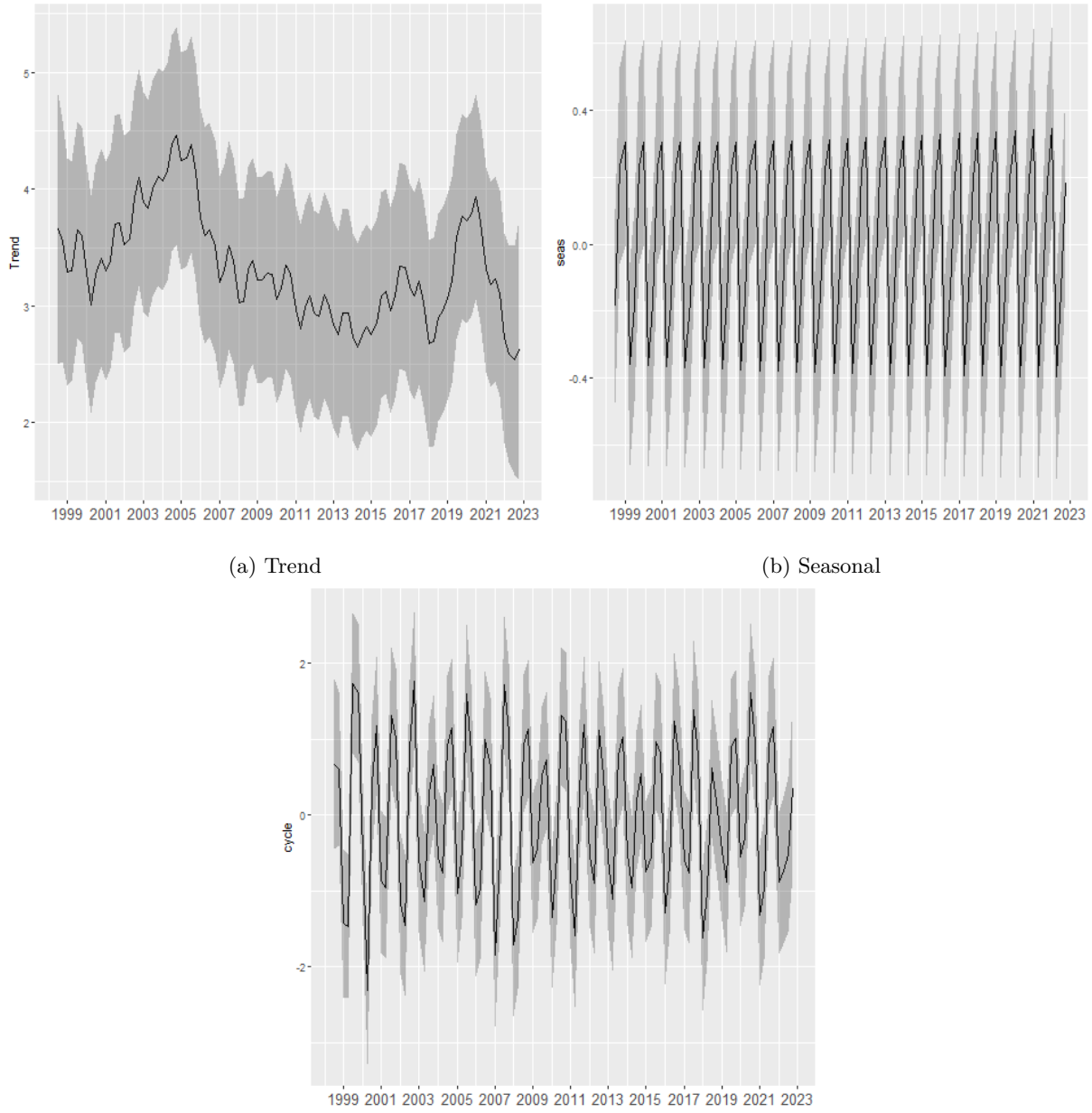
On the other hand, forest formation is considered fire-sensitive, i.e., are not capable to adapt to natural fire regimes and where fire disrupts the ecosystem (HARDESTY; MYERS; FULKS, 2005). However, according to our results, this formation is positively related with the intensity of fire activity in the Pantanal, which may be due to a combination of factors, such as recurrent droughts, human activities, and the lack of environmental policies to cope with illegal fires (PIVELLO et al., 2021).

The precision parameters of the seasonality, cycle, and trend cannot be directly interpreted, and they are better interpreted through the time series of posterior distribution of these components shown in Figure 5.1. Thus, we decided to focus on the interpretation of the figures with the time series containing the posterior averages (solid lines) and the estimated 95% Bayesian credibility intervals (shaded areas) for each component. These components are interpreted as the contribution to the log-intensity function of the process. The mapping for the number of occurrences of the event in each cell of the mesh is given by the exponential of the value of this component, multiplied by the area of each cell in the mesh.

Regarding the trend component (sub-figure (a) of Figure 5.1), we can observe a constant growth between 1998 and 2005, a rapid reduction between 2005 and 2008, with a stabilization in the values of the trend between 2008 and 2014. From 2015 until 2023, we can see two peaks,

with the most significant between 2018 and 2021. Finally, at the end of the observed period, the trend reverts to the lower average level between 1998 and 2023.

The estimated seasonal component (sub-figure (b) of Figure 5.1) is quite stable, and the model does not indicate relevant changes in this pattern. It is also important to note that the estimated cyclical component (sub-figure (c) of Figure 5.1) has a stable behavior over time.



(a) Trend

(b) Seasonal

(c) Cycle

Figure 5.1 – Trend, Seasonal and Cycle decomposition of fire occurrences in the Brazilian Pantanal

The quarterly spatial heterogeneity of the fire events over 1998 and 2022 in the Brazilian Pantanal is clearer through the estimated spatial random effects (see Figure 5.2), which capture the observed variability of dry (with higher risk of fire occurrence) and wet periods in the region, with higher variability in the most fire-susceptible areas, i.e., mainly in the south and central areas, but also in the northeast and north portion of the Pantanal.

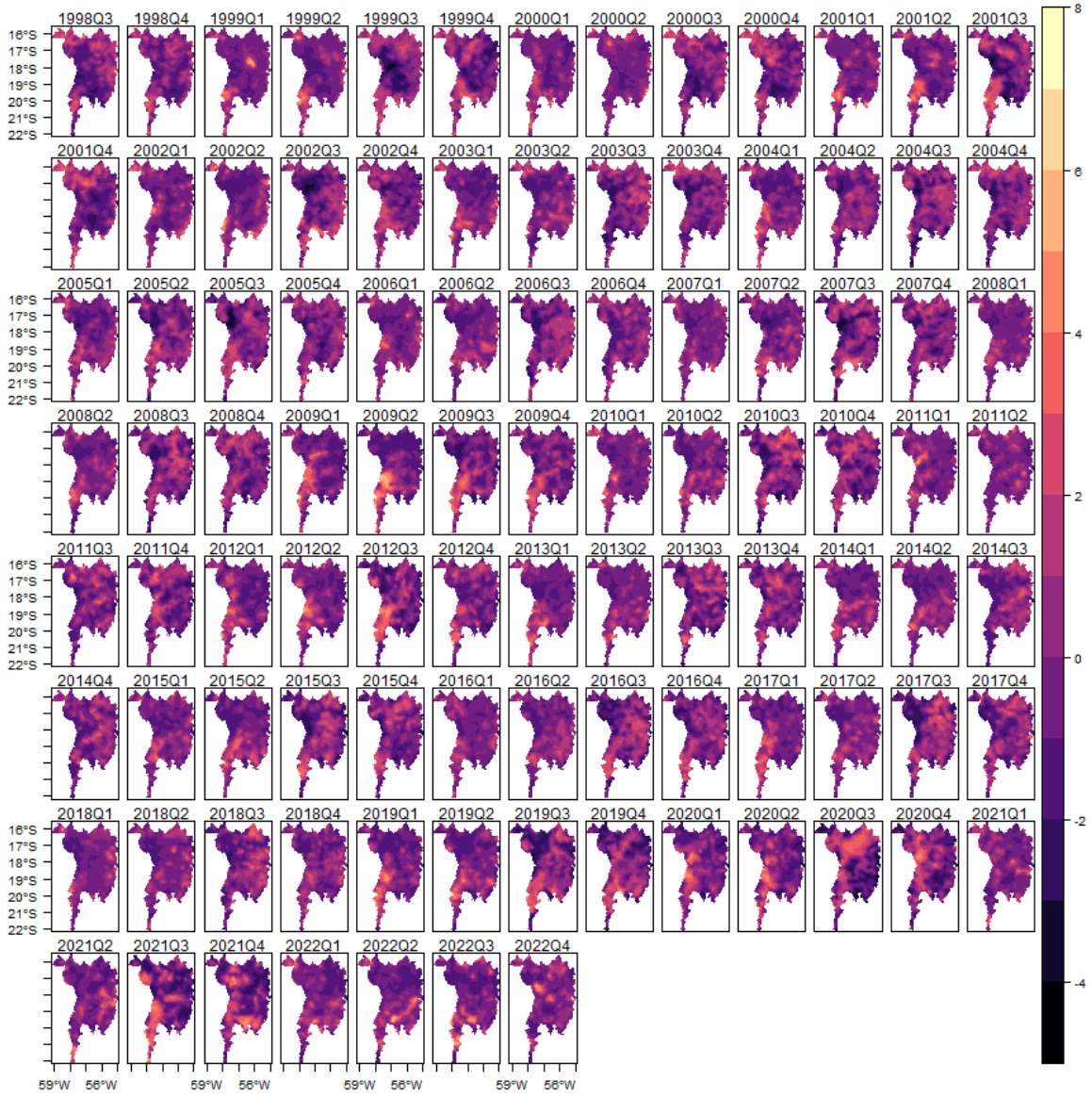


Figure 5.2 – Spatial Random Effects

Figure 5.3 shows the predicted values for the model for the log-intensity of the LGCP process, between 1998 and 2022. We can observe that this function adequately explains the variations observed in the occurrences of fires in each quarter of the sample, in particular the periods with the record of occurrences of events in 2005Q3, 2007Q3 and 2020Q3.

As discussed by Marengo et al. (2021), the Pantanal region has suffered a prolonged drought since 2019. During both the summer of 2019 and 2020, there were significant changes in upper, middle and lower-level circulation and moisture transport in South America, causing rainfall anomalies, which has caused a decrease in the hydrometric levels throughout the Paraguay River. These unusually drought and warm conditions favoring the fire spread, causing unprecedented wildfires in the region in the year 2020 (LIBONATI et al., 2020). However, by analysing the results of the estimated model and, in particular the growing trend, the hypothesis that the record of fires is explained only by the low accumulated rainfall is not supported. Low rainfall

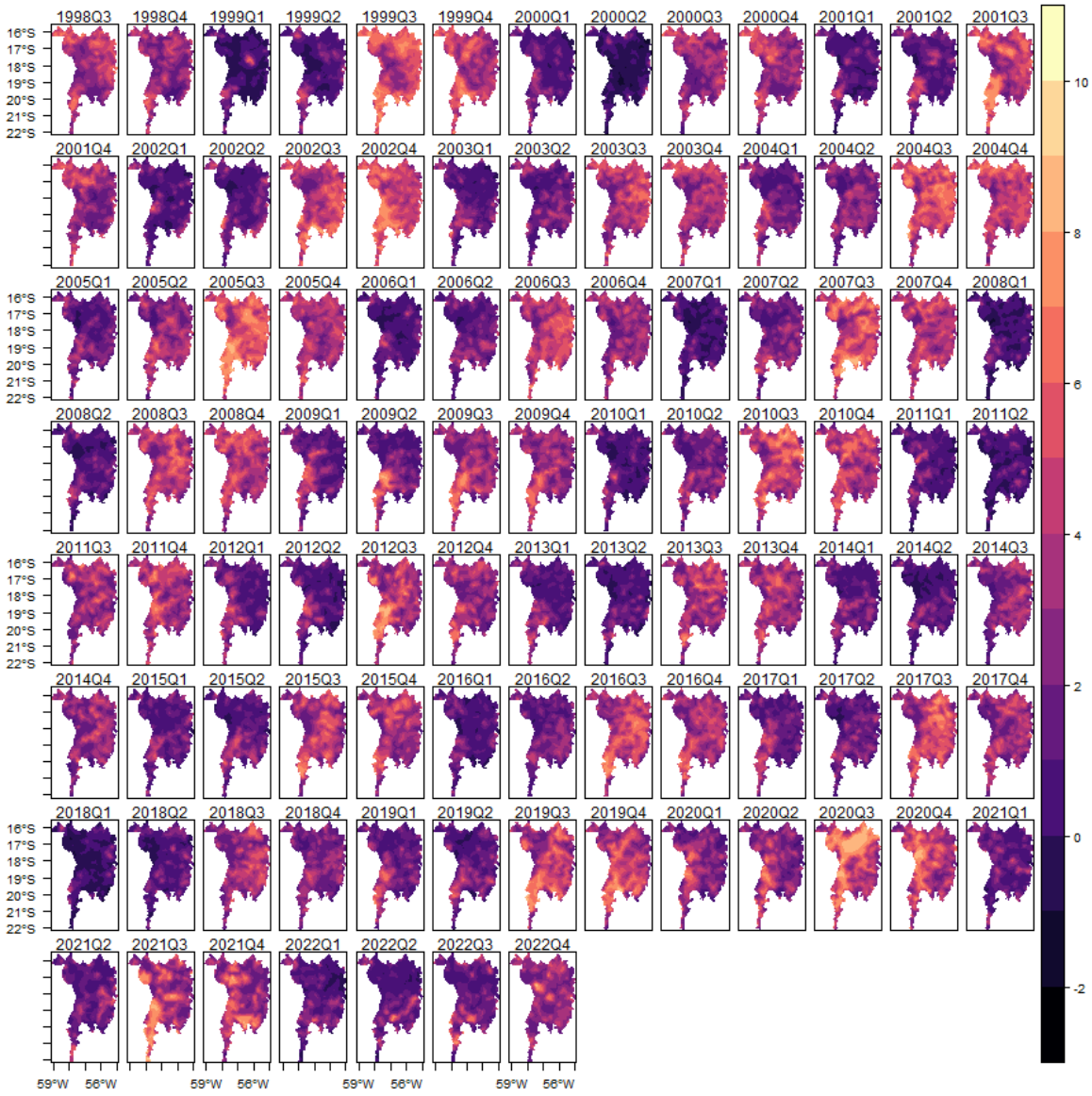


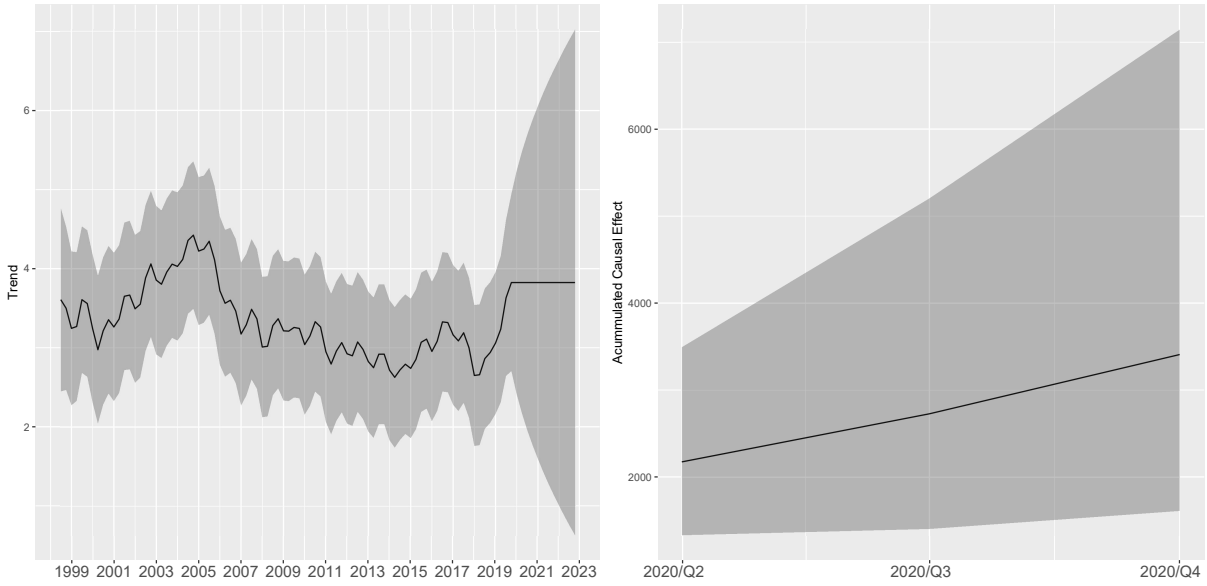
Figure 5.3 – Fitted Log Intensity

plays an important role in the rainfall record, but there is also a permanent change in the average pattern that is captured by the trend component.

The recent observed increase in the number of fire outbreaks in the Brazilian Pantanal may be linked to the recent Brazilian environmental policies. Indeed, we do not have a singular government intervention that could affect the fire occurrence, but rather a set of interventions that could affects the Brazil's environmental system. Although the weakening of Brazilian environmental laws are not entirely new, the election of Brazil's former president, Jair Bolsonaro, has weakened environmental protect measures. Supported by the ruralist groups, Bolsonaro has introduced several measures that encourage the expansion of agriculture and livestock, such as drastic reduction in funds for controlling and monitoring biomes and freer use of agrochemicals and pesticides, leading to substantial environmental damage (ABESSA; FAMÁ; BURUAEM, 2019).

Therefore, considering this scenario where along with the relaxation of environmental legislation, we have an historical record of fire outbreaks registered in 2020, we aimed to estimate the difference of the observed and predicted trajectory of the trend of fire events in the Brazilian Pantanal. One way to analyze this issue is through a counterfactual decomposition, which can also be interpreted as a form of causal effect given by some intervention. In this work we use a generalization of the causal decomposition mechanism for time series proposed in Brodersen et al. (2015). The idea is to use a state space model with the inclusion of covariates to perform the analysis of a possible causal effect given by intervention, controlling for the effect of covariates that are not affected by the intervention. More details about the method used to construct the counterfactual can be found in the Appendix.

From this decomposition, we map the difference between the estimated and predicted trend values for the event occurrence scale, which is the variable of interest in our counterfactual decomposition. In this analysis the interpretation is what is the difference in the number of fires with the trend estimated with the entire sample in relation to the trend forecast before the intervention. The Figure 5.4 shows the (a) posterior predictive expectation of the counterfactual (solid line) and the estimated 95% Bayesian credibility interval (shaded area), and (b) the cumulative difference between estimated trend based on the intensity function decomposition and the counterfactual prediction (solid line) and the estimated 95% Bayesian credibility interval (shaded area). It is important to note that this mechanism is considering the effect of covariates on the model, especially the rainfall observed in the year 2020. Thus, this decomposition gives the counterfactual effect on the number of fires if there had not been a change in the behavior of the model trend.



(a) Trend Forecast (b) Cumulative Difference

Figure 5.4 – Counterfactual Analysis

The posterior mean of the counterfactual difference in the number of occurrences is 2176.375 for the first quarter of 2020, 2729.177 for the second quarter and 3410,083 for the third quarter of 2020, totaling an estimated value of 8315.635 fire spots given by the difference in the trend

estimated by the model, for a total of 18259 observed fires. This result indicates that about 45% of fires occurrences cannot be attributed to climatic factors (rainfall) or other transient factors, and are permanent changes in the pattern of fires in the Pantanal. This result supports the interpretation that changes in the occurrence patterns of fires in the Pantanal are partly permanent changes, and cannot be explained by climatic factors and other non-permanent effects, supporting the hypothesis of an increase in the number of intentional fires due to human actions.

Additionally, in the Appendix, we provide a non-separable version of our spatio-temporal model, which allows us to analyse a more complex dependency structure, despite the computational cost involved in this type of analysis.

5.4 Conclusion

As a contribution to assess the changes in the patterns of fire occurrences in the Brazilian Pantanal, we propose to use a dynamic version of a Log Gaussian Cox process where the intensity function is decomposed into latent components such as trend, seasonality, cycles, and covariates and spatial effects, allowing us to identify long-term changes in the intensity of occurrences over time, and also to capture mean-reverting effects, taking into account the spatial heterogeneity. Within this framework, our findings suggest the existence of a variability in the trend component, which exhibits a growth pattern between 1998 and 2005, and after 2019, whereas it remained relatively stable between 2008 and 2019, and also a statistically significant major incidence of fires in natural vegetation areas. By analyzing the historical record on fire spots in the first three quarters of 2020, our results suggest that it cannot be totally attributed to climate variables, providing evidence of human-induced events.

5.5 Appendix

5.5.1 The spatial covariance function and model details

In this section we provide a brief description of the SPDE approach proposed by lindgren2011explicit. The spatial structure of the model is given by the Matérn family, as discussed in the section Data and Methods. The marginal variance of the covariance function σ^2 is given by:

$$\sigma^2 = \frac{\Gamma(\nu)}{4\pi\kappa^{2\nu}\tau^2\Gamma(\nu+\frac{d}{2})} \tag{5.4}$$

where τ is a scaling parameter and d is the space dimension. In order to easier obtain the results, we adopt a parameterization in terms of $\log\tau$ and $\log\kappa$ for the covariance function, following Lindgren, Rue and Lindström (2011):

$$\begin{aligned}\log\tau &= \frac{1}{2}\log\left(\frac{\Gamma(\nu)}{\Gamma(\alpha)(4\pi)^{d/2}}\right) - \log\sigma - \nu\log\rho \\ \log\kappa &= \frac{\log(8\nu)}{2} - \log\rho\end{aligned} \tag{5.5}$$

where $\rho = \frac{(8\nu)^{1/2}}{\kappa}$. To approximate the LGCP likelihood, we adopt SPDE approach, using the fact that the term $\omega(s,t)$ corresponds to a random field with a Matérn covariance, which allows

to approximate this structure with a Gaussian Markov Random Field (GMRF). Thus, the first main result for the SPDE approach, is that the GF $\omega(s)$ with the Matérn covariance function is a stationary solution to the linear fractional SPDE (WHITTLE, 1954; LINDGREN; RUE; LINDSTRÖM, 2011):

$$(\kappa - \Delta)^{\alpha/2} x(s) = W(s), \quad s \in R^d, \quad \alpha = \nu + d/2, \quad \kappa > 0, \quad \nu > 0 \tag{5.6}$$

where $\Delta = \sum_{i=1}^{d} \frac{\partial^2}{\partial s_i^2}$ is the Laplacian operator and $W(s)$ is a spatial white noise. Therefore, in order to find a GMRF approximation of a GF, we need to find the stochastic weak solution of SPDE (5.6). Using Finite Elements Method (FEM), it is possible to construct an approximated solution of SPDE lindgren2011explicit, which is given by

$$x(s,t) \approx \tilde{x}(s,t) = \sum_{j=1}^{n} w_j \varphi_j(s,t) \tag{5.7}$$

where n is the number of vertices of the triangulation, $\{w_j\}_{j=1}^{n}$ are the weights with Gaussian distribution and $\{\varphi_j\}_{j=1}^{n}$ are the basis functions defined for each node on the mesh. In summary, the idea is to calculate the weights $\{w_j\}$, which determine the values of the field at the vertices, while the values inside the triangles are determined by linear interpolation (LINDGREN; RUE; LINDSTRÖM, 2011), and the equation (5.7) represents a link between the GF and GMRF, where $\{\omega_j\}$ has a Markovian structure (LINDGREN; RUE; LINDSTRÖM, 2011). By replacing the GF by the GMRF approximation, we obtain an approximation of the LGCP likelihood, which consists of $(n + n_t)T$ independent Poisson random variables, where n is the number of vertices and n_t is the number of observed fires (SIMPSON et al., 2016). Under the GMRF structure, it is possible to estimate the model within the Bayesian framework using the Integrated Nested Laplace Approximation (INLA) framework, which allows the use of deterministic approximations to perform the estimation of latent parameters and components in models with an additive structure. More detailed description of the INLA method can be found in Rue, Martino and Chopin (2009). In all analyzes we use the standard reference prior structure described in Martino and Rue (2020). Details can be obtained from the authors.

5.5.2 Data details

The daily data of fire spots used in the paper is provided by the Programa Queimadas[3], from Brazilian National Institute of Spatial Research (*Instituto Nacional de Pesquisa Espacial*-INPE), which uses two different sensors as the main source of information, namely Moderate Resolution Imaging Spectroradiometer (MODIS) Aqua and Terra products, and the Advanced Very High Resolution Radiometer (AVHRR) from National Oceanic and Atmospheric Administration (NOAA). In particular, the NOAA/AVHRR was used as reference satellite for INPE from 1998 to 2002, when it was replaced by MODIS sensor, causing inhomogeneities in the observed time series, which can be considered an important limitation of our analysis.

However, despite the discontinuity in the observed data, it is worth noting that even indicating a fraction of the actual number of fires and forest fires, for using the same detection method and collecting images at close times over the years, the results obtained from both reference

[3] Data and other information available at http://queimadas.dgi.inpe.br/queimadas/portal

satellite allows us to analyze the spatial and temporal trends of the fires. In this sense, to reach our goal, we use the data from the MODIS/NASA and AVHRR/NOAA satellites, with the data validation carried out by the Queimadas system. The data set provides geographical information, and time and period of fire spots within Brazilian Pantanal, which is located within the states of Mato Grosso do Sul and Mato Grosso, in the southwest region. In addition, in order to provide a clearer interpretation of the results obtained, we used a quarterly aggregation of the daily data, which is the sum of the observed fire events in each quarter of the year.

We have plotted the number of quarterly observed of fire spots (see Figure 5.5), from 1998 to 2020, where it is possible to see the unprecedented fire outbreak in 2020. To emphasize the large number of fire outbreaks detected in 2020, in Figure 5.6 we show the observed fire events (black dots) in the Brazilian Pantanal in the third quarters of 2018 and 2020 (on the top of the figure). Also, in the bottom of Figure 5.6, it is possible to observe the non-parametric kernel density estimate of the intensity function (BADDELEY; RUBAK; TURNER, 2015) of the occurrence process. We choose to compare the fire outbreak in 2020 with 2018 since the latter presented fire patterns close to the average in comparison to the past 20 years. Based on Figure 5.6, we can see a notable increase in the intensity and spatial distribution of fire outbreaks in 2020 compared to 2018.

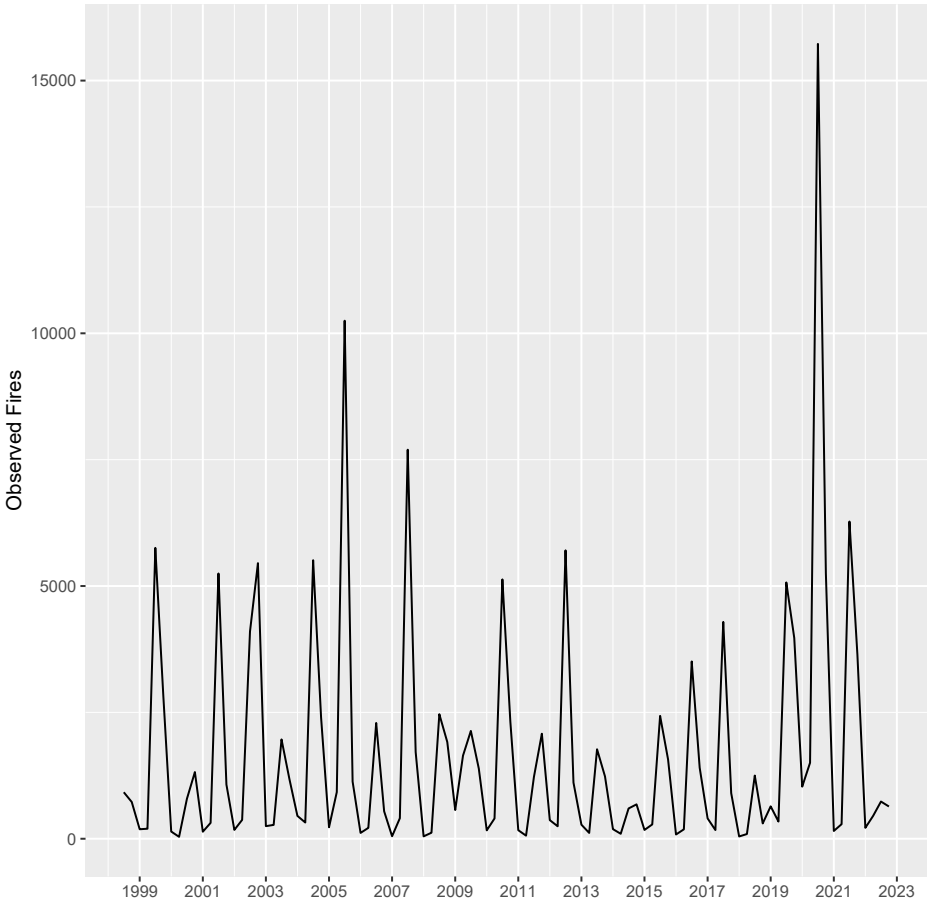


Figure 5.5 – Observed fire events by quarter in Brazilian Pantanal

Regarding the covariates, it is worth discussing about a meaningful limitation related to the selected covariates in our analysis. Since the proposed model performs a spatio-temporal

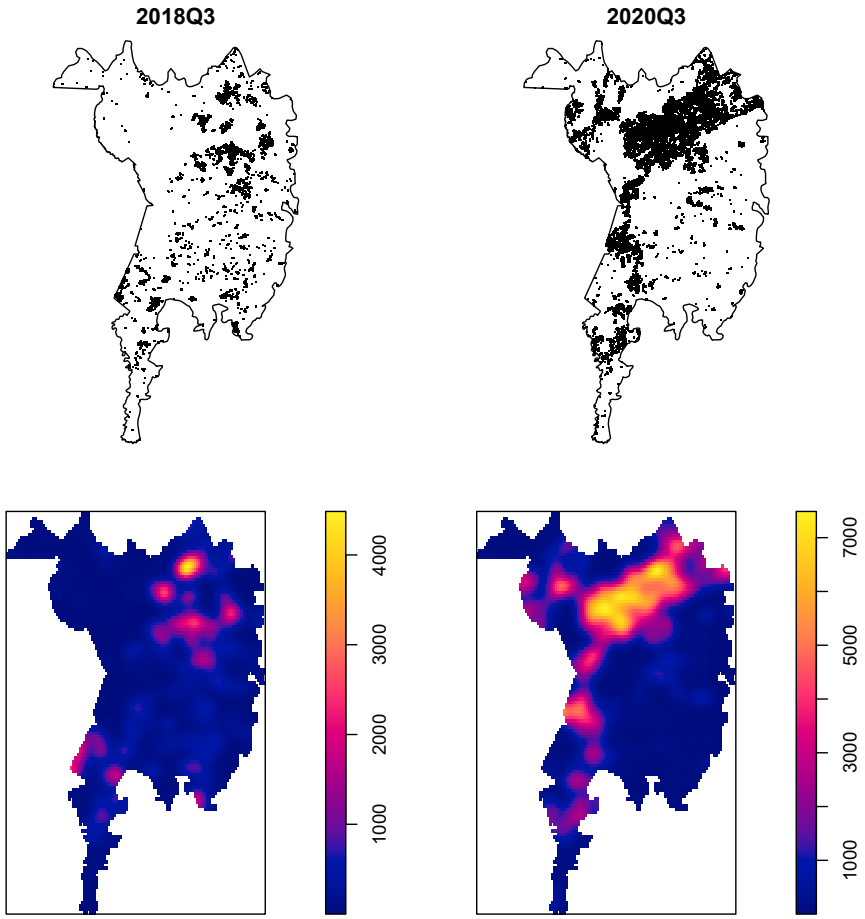


Figure 5.6 – Observed fires (top left and top right) and nonparametric intensity estimation (bottom left and botom right) - 2018Q3 (top and bottom left) and 2020Q3 (top and bottom right)

analysis for the occurrences of a process observed continuously in space, the covariates must to be available at every location of the interest region within the observation window. Due to this methodological constraint, the number of available covariates are limited, and we were able to include limited information in climatic patterns and land use. In particular, regarding the rainfall and maximum temperature data, we calculated the spatially continuous projections from weather station data, following the Laurini (2017) methodology, as discussed below.

The rainfall data were constructed using the time series of the monitoring stations provided by Brazilian National Agency of Waters (*Agência Nacional de Águas*-ANA) and National Institute of Meteorology (*Instituto Nacional de Metereologia* - INMET), whereas the maximum temperature data were obtained based on the information provided by the INMET. For both data, we calculated the spatially continuous projections for each period in the sample, following the methodology proposed by Laurini (2017). By adopting this methodology, we were able to avoid some common problems faced for the analysis of the data sources used in climatology, namely, the dimensionality of the spatio-temporal dataset, the importance of the spatial features, and missing data. In particular, by combining a structure of trend-cycle decomposition with the continuous spatial formulation, the approach allows us not only to estimate the patterns throughout the spatial continuum and how it propagates throughout the area of interest, but

also provides a way to solve the missing data problems by adding the latent components with the prediction obtained for the spatial effect in the geographic position of the weather station using the continuous projection of the spatial effect, without the necessity of additional treatments for missing data or interpolation methods. In summary, this methodology allows us to control for possible changes in weather patterns, and is also based on possible changes in trends, seasonality and cycles in climate data.

In addition, we included, as categorical variables, yearly information on LULC provided by the Landsat-based MapBiomas project (Collection 5). The database includes annual historical maps of each biome, which contains a hierarchical system of classification of land use/land cover following the Food and Agriculture Organization (FAO) and *Instituto Brasileiro de Geografia e Estatística* (IBGE) classification systems. The first level contains six classes, namely, forest, non-forest formation, farming, non vegetated area, water, and non-observed. Forest constitutes natural forest and forest plantation, whereas the non forest natural formation includes wetland, grassland, salt flat, rocky outcrop, and other non forest formations. The farming class includes pasture and agricultural land and mosaic. The non vegetated area are defined by beach and dune, urban area, mining e and other non vegetated areas. Finally, water class includes river, lake, ocean, and aquaculture. The accuracy statistics varies according to the level and biome. In particular, considering the Pantanal biome, the first level has 81.6% of overall accuracy with 12.9% of allocation mismatch and 5.6% of quantity mismatch. For the second and third levels, the overall accuracy is 73.5%, whereas the allocation and quantity mismatch are 17.5% and 9%, respectively. The methodology overview of the MapBiomas project is available at <https://mapbiomas.org/>, whereas the accuracy assessment for the Brazilian biomes is available in the MapBiomas accuracy statistics web page[4]. A detailed description of the MapBiomas land use/land cover classification can be found in Table 5.1, while Figure 5.7 shows a map with the Mapbiomas classifications for the year 2019.

5.5.3 Additional results of the spatial-temporal model

As discussed in the Data and Methods section, the estimation of the spatio-temporal model is based on the construction of a mesh, which represents a discretization of the continuous space for the evaluation of the likelihood function of the Log Gaussian Cox Process. In this work, we use a mesh with 804 triangles, as shown in Figure 5.8. Through the INLA method, we estimated the posterior distribution of the parameters described in the Equation (5.1). Thus, the estimated parameters (see Table **??**) are the precision of the trend component ($1/\eta_\mu$), seasonal component ($1/\eta_s$), and cycle component ($1/\eta_c$), the parameters of the second-order autoregressive process of the cycle (PACF1 and PACF2), the parameters associated with the set of observed covariates (β), the parameters of spatial covariance ($\log\tau$ and $\log\kappa$), and the parameter of spatial time dependence (Φ).

The estimated cycle parameters indicates low persistence in this component, with low values of partial autocorrelations, but with a relevant variance over time, as shown in Figure 5.1. The parameters Log τ , Log κ and Group Φ are linked to the representation of the Matérn spatial covariance matrix used in the representation of the model, and which are also better interpreted

[4] <https://mapbiomas.org/en/estatistica-de-acuracia?cama_set_language=en>

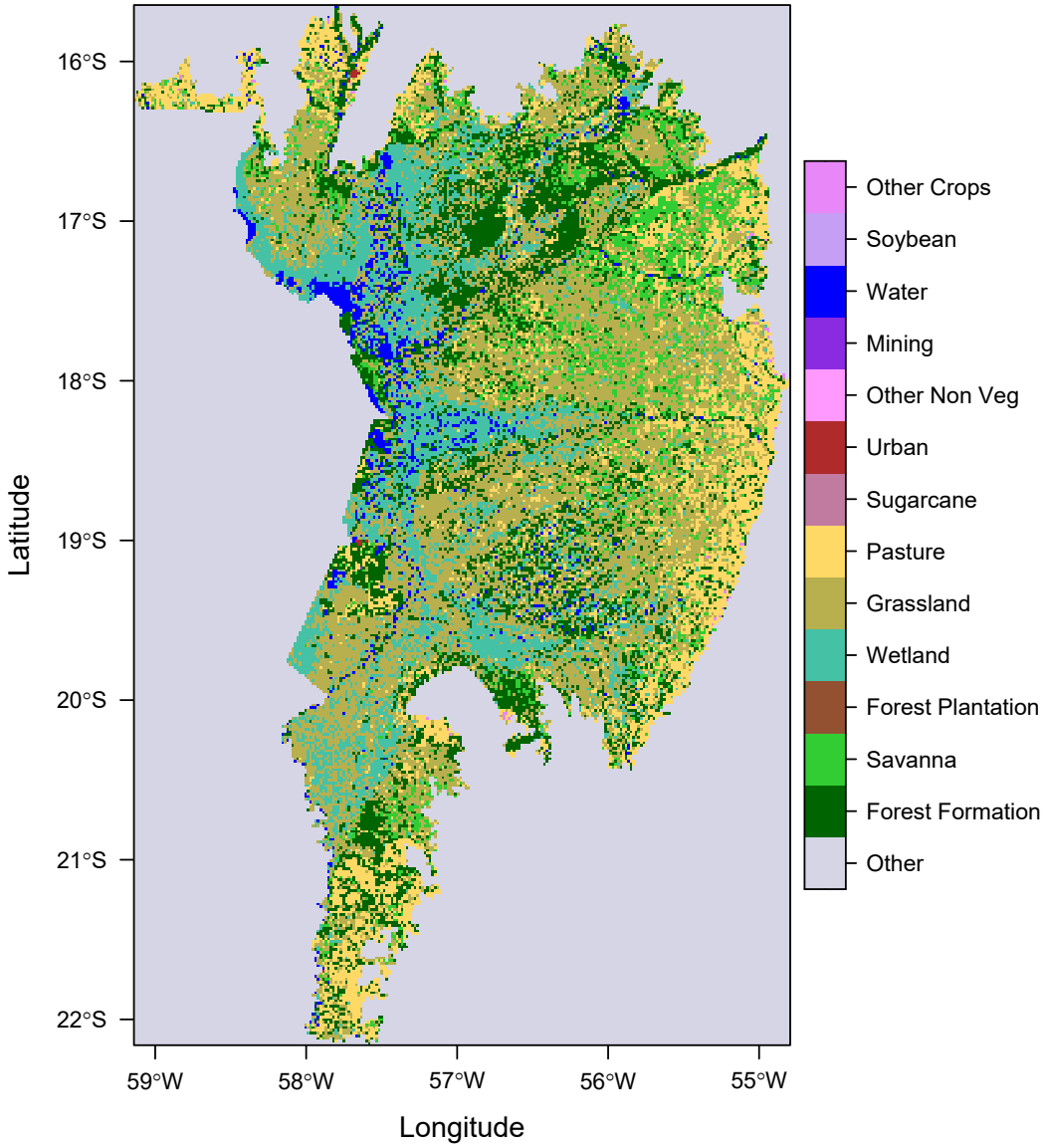


Figure 5.7 – Land use/Land cover in the Brazilian Pantanal from MapBiomas - 2019

by the posterior distribution of the spatial random effect (Figure 5.2) . It is important to note that the temporal persistence of the spatial effect is relatively low, with Group Φ parameter estimated with a posterior mean value of 0.327.

5.5.4 Counterfactual inference

From the decomposition of the intensity function, we propose to use the Brodersen et al. (2015) approach to obtain the posterior predictive expectation of the counterfactual, which is estimated on the basis of a diffusion-regression state-space model that predicts the counterfactual response in a synthetic control that would have occurred without any intervention. The synthetic control method (ABADIE; GARDEAZABAL, 2003; ABADIE; DIAMOND; HAINMUELLER, 2010) is a systematic way to construct the counterfactual. By constructing a synthetic control, Brodersen et al. (2015) proposes to combine three sources of information, using a state-space

Table 5.1 – MapBiomas Land Use/land cover Classification

LULC	ID
1. Forest	1
1.1. Natural Forest	2
1.1.1. Forest Formation	3
1.1.2. Savanna Formation	4
1.1.3. Mangrove	5
1.2. Forest Plantation	9
2. Non Forest Natural Formation	10
2.1. Wetland	11
2.2. Grassland Formation	12
2.3. Salt Flat	32
2.4. Rocky Outcrop	29
2.5. Other non Forest Formation	13
3. Farming	14
3.1. Pasture	15
3.2. Agriculture	18
3.2.1. Temporary Crop	19
3.2.1.1. Soybean	39
3.2.1.2. Sugar cane	20
3.2.1.3. Other Temporary Crops	41
3.2.2. Perennial Crop	36
3.3. Mosaic of Agriculture and Pasture	21
4. Non vegetated area	22
4.1. Beach and Dune	23
4.2. Urban Infrastructure	24
4.3. Mining	30
4.4. Other non Vegetated Areas	25
5. Water	26
5.1. River, Lake and Ocean	33
5.2. Aquaculture	31
6. Non Observed	27

time-series model: (1) the time series behavior of the response itself, prior to the intervention; (2) the behavior of other time series that were predictive of the target series prior to the intervention; (3) prior knowledge about the model parameters.

In the structure proposed in Brodersen et al. (2015), the permanent causal effect is given by the difference of the values adjusted by the model using all available information, with the values predicted by the model using the trend component forecast using the data before the intervention. Thus, the causal effect is given by the difference in the forecast using the trend observed using the entire sample, with the model using a trend forecast with the information before the intervention occurred. In this form, the model captures the causal effect as the permanent change given by the intervention, controlling for all other transient and covariate effects included in the model. Our generalization is based on the generalization of this model of time series to a spatio-temporal context, given by the inclusion of random spatial effects.

Through Bayesian approach, the inference procedure start by specifying a prior distribution, $p(\theta)$, on the model parameters as well as a distribution on the initial state values, $p(\delta_0|\theta)$.

Figure 5.8 – Spatial mesh of the Brazilian Pantanal

According to Brodersen et al. (2015), the posterior inference can be broken down into three pieces: simulate draws of the model parameters θ and the state vector δ given the observed data $\mu_{1:n}$, using MCMC; from posterior simulations, simulate from the posterior predictive distribution $p(\tilde{\mu}_{n+1:m}|\mu_{1:n})$ over the counterfactual time series $\tilde{\mu}_{n+1:m}$ given the observed pre-intervention $\mu_{1:n}$; use the posterior predictive samples to compute the posterior distribution of the (pointwise) impact, which is defined as

$$\phi_t^{(\tau)} := \mu_t - \tilde{\mu}_t^{(\tau)} \tag{5.8}$$

for each draw τ and each time point $t = n+1, \ldots, m$. From the same samples it is possible to obtain posterior distribution of cumulative impact, estimated by

$$\sum_{t'=n+1}^{t} \phi_{t'}^{(\tau)} \quad \forall t = n+1, \ldots, m. \tag{5.9}$$

5.5.5 Non-separable spatio-temporal model

The model proposed in this paper assumes a separable structure between the spatial and temporal effects, assuming a Kronecker product between the spatial and temporal covariances to obtain the spatio-temporal representation, which is advantageous due to its flexibility. Despite the computational cost, by assuming a non-separable structure for time and space, it is possible to analyse a more complex dependency structure. In order to provide an additional robustness

Table 5.2 – Estimated Parameters

	Mean	SD	0.025quant	0.5quant	0.975quant	Mode
Fixed effects						
Rainfall	-0.00	0.000	-0.003	-0.002	-0.001	-0.002
Forest formation	0.093	0.036	0.023	0.093	0.163	0.093
Savanna formation	0.083	0.038	0.009	0.083	0.156	0.083
Grassland formation	0.173	0.034	0.106	0.173	0.239	0.173
Random effects						
Precision for trend	5.341	0.120	5.099	5.339	5.592	5.336
Precision for seasonality	13385.045	396.383	12693.202	13365.335	14273.920	13247.130
Precision for cycle	1.346	0.015	1.318	1.345	1.377	1.342
PACF1 for cycle	0.143	0.013	0.120	0.142	0.170	0.139
PACF2 for cycle	-0.101	0.012	-0.128	-0.100	-0.080	-0.095
Log τ	-4.029	0.007	-4.045	-4.029	-4.015	-4.028
Log κ	2.166	0.008	2.151	2.165	2.184	2.162
Group Φ	0.355	0.005	0.345	0.355	0.364	0.355

analysis, we estimated a non-separable version of the proposed model presented in the previous section, following Bakka et al. (2020), which provides a non-separable representation for the spatio-temporal random effects using the generalization of the Matérn covariance structure. In particular, in this representation the structure of the spatio-temporal random effects is given by a diffusion-based extension of the Matérn field, i.e., the random field $u(s,t)$ can be written as

$$\left(\gamma_t \frac{d}{dt} + L^{\alpha_s/2}\right)^{\alpha_t} u(s,t) = \mathcal{E}_Q(s,t) \tag{5.10}$$

with $L = \gamma_s^2 - \Delta$, $\mathcal{E}_Q(s,t)$ is Gaussian noise that is white in time but correlated, with precision operator $Q(\gamma_s, \gamma_\epsilon, \alpha_\epsilon) = \gamma_\epsilon^2 L^{\alpha_\epsilon}$, with $(\gamma_t, \gamma_s, \gamma_e)$ being a fixed scaling parameters, and $(\alpha_t, \alpha_s, \alpha_e)$ parameters in the model.

Defining $\alpha = \alpha_e + \alpha_s(\alpha_t - 1/2)$, and assuming that $\alpha_t, \alpha_s, \alpha_e$ satisfy $\alpha > 1$ the solution $u(s,t)$ has marginal spatial covariance function given by:

$$C(u(t,s_1), u(t,s_2)) = \frac{\sigma^2}{\Gamma(\nu_s)2^{\nu-1}} (\gamma_s \|s_1 - s_2\|)^{\nu_s} K_{\nu_s}(\gamma_s \|s_1 - s_2\|), \tag{5.11}$$

where $\nu_s = \alpha - 1$ and

$$\sigma^2 = \frac{\Gamma(\alpha_t - 1/2)\Gamma(\alpha - 1)}{\Gamma(\alpha_t)\Gamma(\alpha) 8\pi^{3/2} \gamma_e^2 \gamma_t \gamma_s^{2(\alpha-1)}}. \tag{5.12}$$

In order to carry on our analysis, we estimate a non-separable version of our LGCP model, replacing the random field structure, $\xi(s,t)$, of the model defined in 5.1, with the random field $u(s,t)$ previously defined. In this case, the representation of the model 5.1 can be written now as

$$
\begin{aligned}
Y(s,t) &= Poisson(|e(s,t)|exp(\lambda(s,t)) \\
\log\lambda(s,t) &= \mu_t + s_t + c_t + z(s,t)\beta + u(s,t) \\
\mu_t &= \mu_{t-1} + \eta_\mu \\
s_t &= s_{t-1} + s_{t-2} + \ldots + s_{t-m} + \eta_s \\
c_t &= \theta_1 c_{t-1} + \theta_2 c_{t-2} + \eta_c
\end{aligned}
\tag{5.13}
$$

We also assume the parametrization of the parameters defining σ, r_s, r_t as:

$$
\begin{aligned}
c_1 &= \frac{\Gamma(\alpha_t - 1/2)\Gamma(\alpha - 1)}{\Gamma(\alpha_t)\Gamma(\alpha)4\sqrt{\pi}} \\
\sigma &= \gamma_\epsilon^{-1} c_1^{1/2} \gamma_t^{-1/2} \gamma_s^{-(\alpha-1)} \\
r_s &= \gamma_s^{-1}\sqrt{8\nu_s} \\
r_t &= \gamma_t\sqrt{8(\alpha_t - 1/2)}\gamma_s^{-\alpha_s},
\end{aligned}
\tag{5.14}
$$

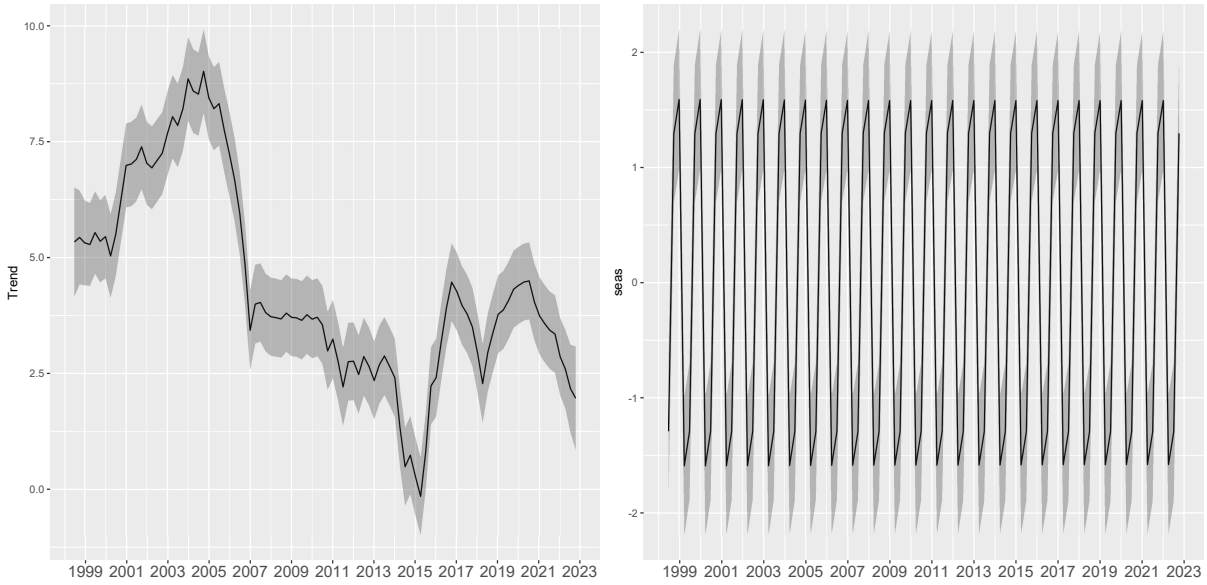
and the model is estimated using as parameters $\log\sigma$, $\log r_s$, $\log r_t$. Due to the higher computational cost and memory limitations in representing the model with the non-separable structure, we estimate the model with an alternative mesh with a lower resolution than the separable model, shown in Figure 5.9, with the same prior previously defined for the estimation of the model 5.1, and also adopting the INLA approximations to perform the Bayesian inference procedures. The estimated posterior distribution of parameters in the non-separable spatio-temporal LGCP model is presented in the Table 5.3, the estimated latent components of trend, seasonal and cycle components in the Figure 5.10 and the non-separable spatio-temporal random effect is presented in the Figures 5.11.

Figure 5.9 – Spatial mesh of the Brazilian Pantanal - Non-separable spatio-temporal model

The results of the estimated non-separable spatio-temporal model indicate similar effects for the covariates in relation to those obtained in the model with the separable spatial random effects. Additionally, this model was able to capture a wider range of values for the random effects when compared to those estimated by the separable model, as can be seen in Figure 5.11. Regarding the estimated components, in Figure 5.10, it is possible to observe that the non-separable model was able to capture an increase in the trend component during 2019 and 2020, when the historical fires occurred in the Pantanal biome. Despite that, the estimated cycle and seasonal components presented similar patterns to those obtained by the model with the separable structure.

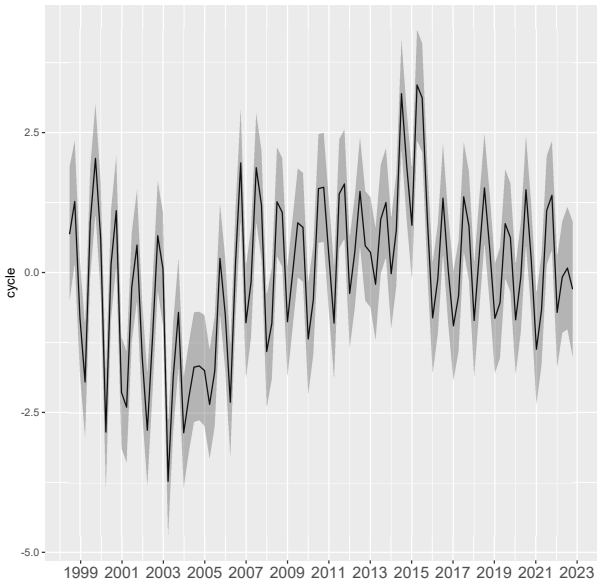
Table 5.3 – Estimated Parameters - Non-separable spatio-temporal model

	Mean	SD	0.025quant	0.5quant	0.975quant	Mode
Fixed effects						
Rainfall	-0.006	0.000	-0.007	-0.006	-0.006	-0.006
Forest formation	0.417	0.035	0.348	0.417	0.485	0.417
Savanna formation	0.606	0.042	0.524	0.606	0.689	0.606
Grassland formation	0.594	0.032	0.531	0.594	0.657	0.594
Random effects						
Precision for trend	3.999	0.125	3.728	4.002	4.225	4.037
Precision for seasonality	8535.374	408.976	7826.038	8512.086	9459.335	8390.847
Precision for cycle	0.672	0.042	0.608	0.668	0.771	0.647
PACF1 for cycle	0.062	0.018	0.028	0.061	0.100	0.058
PACF2 for cycle	-0.719	0.009	-0.736	-0.719	-0.698	-0.721
$\log \sigma$	1.237	0.013	1.211	1.237	1.263	1.237
$\log r_s$	-0.922	0.006	-0.933	-0.922	-0.910	-0.922
$\log r_t$	0.507	0.004	0.498	0.507	0.515	0.508



(a) Trend

(b) Seasonal



(c) Cycle

Figure 5.10 – Trend, Seasonal and Cycle decomposition of fire occurrences in the Brazilian Pantanal - Non-separable spatio-temporal model

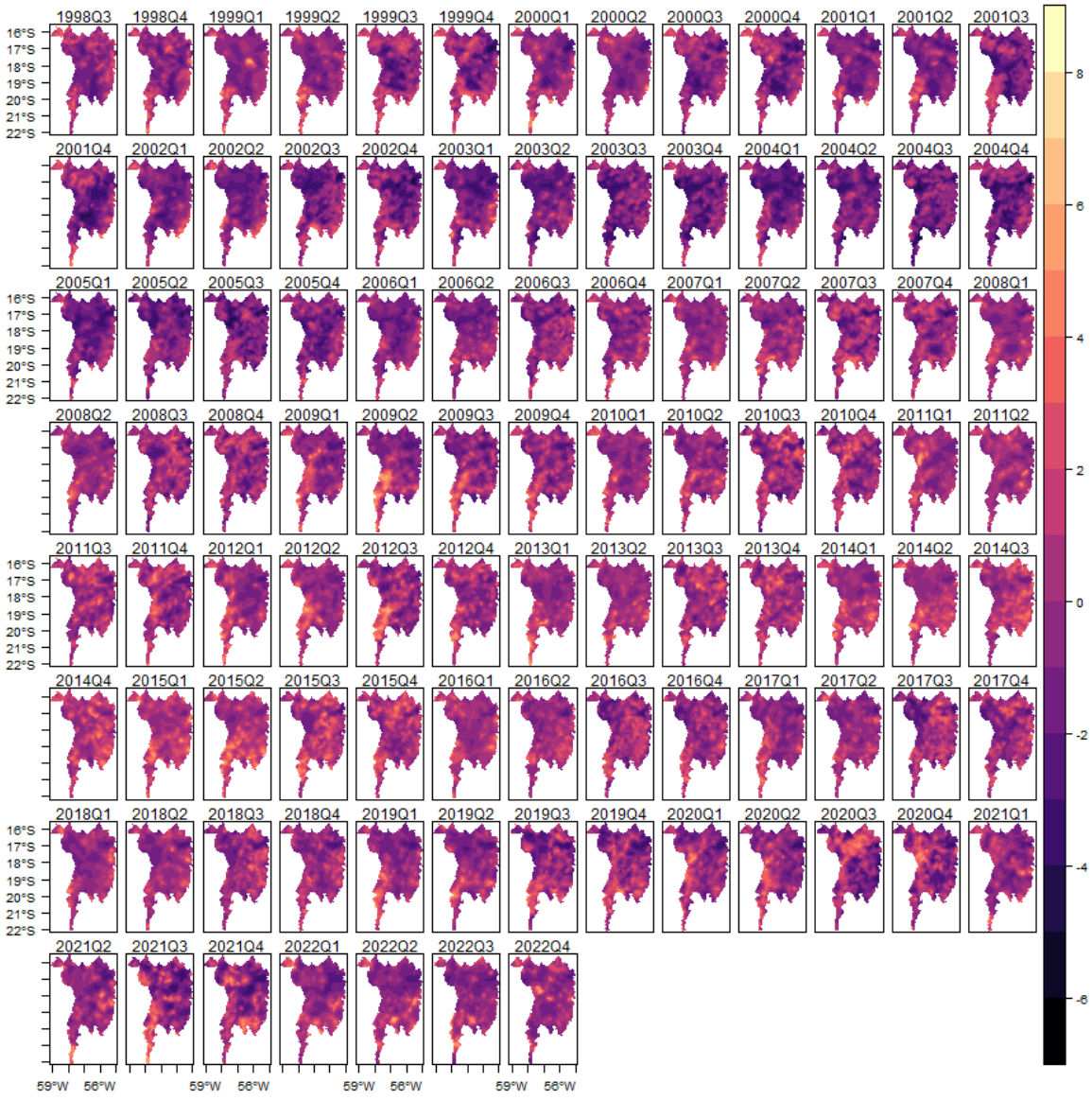


Figure 5.11 – Spatial Random Effects - Non-separable spatio-temporal model

6 URBAN CLIMATE CHANGE: A STATISTICAL ANALYSIS FOR SÃO PAULO

Fernanda Valente Márcio Laurini

ABSTRACT

In this paper we propose novel statistical decompositions of climatic time series into long-term trend, seasonal and cycle components to analyze the existence of changes in rainfall and temperature patterns. The main contribution is the generalization of these decompositions to extreme values, probability of rain and the duration of dry days. We apply these methods for the metropolitan region of São Paulo. This region is characterized by a rapid urbanization process and a high population density, which is related to increased danger and vulnerability to hydrometeorological hazards. We analyzed daily series of temperature and precipitation between 1933 and 2019 from Institute of Astronomy, Geophysics and Atmospheric Sciences of University of São Paulo (IAG-USP) and the results indicate a significant increase in the trend component of daily temperature and in the rain patterns. In addition, the analysis for annual temperature records and annual records of daily rainfall show support for the hypothesis of permanent changes in observed climatic patterns. On the other hand, there is no significant evidence of changes in the pattern of duration of dry days nor changes in the trend component for the annual maximum of dry days, indicating the stability of this component.

Keywords: Climate change; extremes; metropolitan area of São Paulo ; trend-cycle decomposition

VALENTE, Fernanda; LAURINI, Márcio. Urban climate change: A statistical analysis for São Paulo. **Urban Climate**, v. 41, p. 101077, 2022.

6.1 Introduction

The climate change patterns, induced by human processes or natural activities, have led to changes in the frequency and the intensity of extreme weather and climate events (ALLEN et al., 2014), which have significant effects in the city dynamics and their citizen's life. The effects of climate-related disasters are often intensified in cities due to cultural, demographic and economic characteristics of urban residents, city governments, built environment, and human-induced stresses, such as overexploitation of resources and environmental degradation (GENCER; FOLORUNCHO; LINKIN, 2018).

One way to verify the existence of changes in climate patterns is through the estimation of permanent and transitory components (BLOOMFIELD, 1992; VELARDE; MIGON; PEREIRA, 2004; PROIETTI; HILLEBRAND, 2017; LAURINI, 2019). The main idea is to decompose the temporal variability observed through statistical tools for extraction of trend, seasonality and cycle components, which allows to identify possible climate change, since it allows the estimation of variations in the long term behavior of time series. A survey on the use of trend analysis in

climate research can be found in Mudelsee (2019), the use of stochastic trends in Kaufmann, Kauppi and Stock (2010), and general methods of statistical analysis in climate research in Storch and Zwiers (2002).

In the paper herein, we propose novel decompositions of climatic time series into trend, seasonal and cycle components to analyze the existence of changes in rainfall and temperature patterns in the Metropolitan Area of São Paulo (MASP). The main method contribution of this article is to use the decomposition in a more general context than the traditionally used form based on a Gaussian likelihood. The decomposition is applied to extreme value processes, modeling annual maximum rainfall and temperature, Bernoulli-Gamma composite processes for the occurrence of rain and the amount of rain observed, and in a dynamic model for the duration of dry days. In addition, it is important to note that our proposed method contributes to the literature by extending the trend-cycle decomposition for process in climate modeling framework.

In our approach the persistent trend component is modeled as first and second order random walk whereas the seasonality process is modeled by a stochastic process with restriction that the effects must sum to zero, and the cyclic component is represented by a second-order autoregressive process with possible complex roots, which allows to capture the cyclic patterns. The proposed formulation allows to estimate permanent changes in the rainfall and temperature over time and also to capture the impact of non-permanent shocks on temporal patterns. In addition, this formulation also fits in the Bayesian framework which allows us to perform computationally efficient inference procedures through Integrated Nested Laplace Approximations (INLA) algorithm (RUE; MARTINO; CHOPIN, 2009).

We analyze daily series of temperature and precipitation between 1933 and 2019 from meteorological station located in the Institute of Astronomy, Geophysics and Atmospheric Sciences of University of São Paulo (IAG-USP). In Brazil, natural disasters such as floods, which are associated with extreme events like storms and heavy rainfall, have become recurrent in some regions, mainly in large metropolis, where the major problem is the occupation of inappropriate and risky areas such as slopes and banks of watercourses, which are subjected to inundation, flooding and landslides (JACOBI et al., 2013). Between 1991 and 2012, there were almost 39,000 natural disasters in Brazil, in which flash foods were responsible for 58% of the total fatalities and mudslide for 16%, according to The National Atlas of Natural Disasters of Brazil. Since 2010, natural disasters in Brazil have caused 31 floods and 2 intense mudslides. These events, together, were responsible for 1559 deaths and affected almost 3 million people, with financial damage in the order of $3,365,500 all over the country (MARENGO et al., 2020).

The intensity and the probability of extreme weather and climate event can be partly due to urbanization process, as discussed by Lima, Lombardo and Magaña (2018), changes in land use due to the process of urbanization growth over years may result changes in the temperature and precipitation. Indeed, there is evidence that since 1970s, both precipitation and temperature data show a tendency to increase the frequency of days with extreme values (LIMA; LOMBARDO; MAGAÑA, 2018). The increased growth in the urban population has generated excessive demand for housing and a deficit in the supply of adequate shelter (JACOBI et al., 2013). This scenario has led many people in Brazilian cities live in risky and inappropriate areas. Once the urbanization process moves to more peripheral areas, the situation becomes worse due

to the weaknesses of the state regarding land use and planning. In Brazil, according to the *Instituto Brasileiro de Geografia e Estatística* (IBGE), almost 85% of the population lives in urban areas. The MASP, which encompasses 39 municipalities in the state of São Paulo with intense process of conurbation, represents the largest urban concentration in South America, and one of the 10th largest in the world (MARENGO et al., 2020). According to IBGE, the MASP has more than 21 million of inhabitants, in which almost 60% live in the São Paulo city, which is responsible for 12% of the national Brazilian GDP. However, in some cities in the MASP it is possible to find one of the highest levels of poverty and inequality (MARENGO et al., 2020).

In MASP, floods and droughts are the most critical climate-related issues. In the variability of rainfall across MASP, it can be observed the mean annual value varying between 1300mm and 2500mm. Areas of maximum values (2400mm) are observed along the southern side of the MASP which can be related to the southeastern slope of the Serra do Mar mountain range. Also, the Serra do Mar, southeastern region of MASP, concentrate areas with high numbers of wet days (more than 150 days) whereas in the rest of the domain are observed less than 120 wet days (OBREGÓN; MARENGO; NOBRE, 2014).

Previous studies have identified a positive trend in both, annual and seasonal rainfall in MASP (OBREGÓN; MARENGO; NOBRE, 2014; DIAS et al., 2013; MARENGO; VALVERDE; OBREGON, 2013), which are related with local factors and climatic indices. At seasonal scale, the rainfall variability over the MASP may be modulated by the Southern Annular Mode (SAM), Pacific Decadal Oscillation (PDO) and El Niño Southern Oscillation (ENSO), which alter the normal progression of the seasonal rainfall cycle (MARENGO et al., 2020). In the interannual and decadal scales, the frequency and intensity of precipitation in the south and southeast of Brazil can be influenced by the surface temperatures of the Pacific Ocean (LIMA; LOMBARDO; MAGAÑA, 2018).

Regarding the variation in the total amount of rainy-season precipitation, it can be analyzed based on the number of wet days or heavy precipitations, or both (SONG; ACHBERGER; LINDERHOLM, 2011). Recent works have identified an increase in both rainy days and extreme daily precipitation (LIMA; LOMBARDO; MAGAÑA, 2018; MARENGO; VALVERDE; OBREGON, 2013; NUNES et al., 2019; ZILLI et al., 2017) and decreases in light rain (MARENGO; VALVERDE; OBREGON, 2013; ZILLI et al., 2017), probably due to natural climate variability associated with some signals of the urbanization effect. Plus, these studies have also noted that the number of consecutive dry days has increased, indicating that intense precipitation is becoming concentrated on a few days with longer dry periods occurring in between.

However, as discussed by Marengo et al. (2020), from a statistical perspective, it is challenging to deal with databases as large as climate databases, and despite the great number of studies, there is a lot of work to do. In this sense, our model provides a computationally feasible way to improve our understanding of possible changes in transient and long-term patterns of climate-related issues. In particular, from IAG station database, our proposed model allows us to analyze the patterns of the rainfall and temperature in the MASP, which is a region of unique socioeconomic complexity. Additionally, it is worth noting that our proposed model provides a way to analyze the presence of trends in climatic time series, avoiding the common issues of the Mann-Kendall non-parametric test, which is the most traditional way of testing the existence of trends in this type of time series.

This article is organized as follow. Section 2 contains a description of the statistical approach. Section 3 presents the data. Section 4 shows the results with discussion. Section 5 concludes.

6.2 Material and Methods

6.2.1 Trend-Cycle Decomposition in Climatic Models

This section briefly describes the trend-cycle decomposition proposed in this work, whereas the Generalized Extreme Value (GEV) distribution and Survival Models are described in the Appendix of the paper. We can represent a climatic model by the hierarchical representation using the follow dynamic formulation:

$$\begin{aligned} y_t &= \mu_t + s_t + c_t + \epsilon_t \\ \mu_t &= \mu_{t-1} + \eta_\mu \\ s_t &= s_{t-1} + s_{t-2} + \ldots + s_{t-m} + \eta_s \\ c_t &= \theta_1 c_{t-1} + \theta_2 c_{t-2} + \eta_c \end{aligned} \tag{6.1}$$

where y_t represents the observation y in period t, μ_t is the long term trend, s_t represents the seasonal component, c_t is a cycle component represented by a second-order autoregressive process with complex roots and, ϵ_t are innovations with $\epsilon_t \sim N(0, \sigma_\epsilon^2)$. One advantage of this formulation is it fits naturally within the Bayesian hierarchical modeling framework and are latent Gaussian models, therefore, it may be fitted using the Integrated Nested Laplace Approximations (INLA) (RUE; MARTINO; CHOPIN, 2009). The prior structure used in the estimation is shown in the Appendix.

The most traditional way of analyzing for the presence of trends in climatic time series is the Mann-Kendall non-parametric test (MANN, 1945; KENDALL; STUART, 1967). This test is based on the distribution of the sample ranks, and has the advantage of not explicitly depending on the distribution of the data. However, this test has some important limitations in detecting trends. The first is in the formulation of the null and alternative hypothesis. The null hypothesis is that the data is an independent and identically distributed sequence, against an alternative hypothesis of a monotonous trend. Assuming the presence of a monotonous trend in the data under alternative hypothesis may be problematic since is not suitable if the trend has non-linear patterns, a result predicted by several climatic models. Second, this formulation can be problematic since the series under analysis may not have the presence of a trend, thought of as a permanent component of the series, rather than some form of stochastic process with persistence, but still stationary. In our formulation this dependency would be captured by the cycle component. This cycle can capture transient climatic effects, which do not permanently alter the series' patterns. In the Mann-Kendall test, the presence of this structure would be evidence against the null hypothesis of independence, even if the process is not a trend towards a permanent component of the series. Some modifications have been proposed to modify the Mann-Kendall test for robustness to contamination in this way, but with some important drawbacks (BLAIN, 2013).

The decomposition proposed in this work circumvents these limitations existing in the Mann-Kendall test, since it allows to directly structure the trend component and the possible com-

ponents of stationary persistence, such as the cycle and seasonality components, and directly analyze the changes in each component. The model is not limited to monotonous trends, and can be adapted to different data distributions. For example, in the rainfall data we use a Gamma distribution, which directly incorporates censorship in non-positive values, and in the analysis of the maximum temperatures and rain, the distribution used is a Generalized Extreme Value (GEV) distribution.

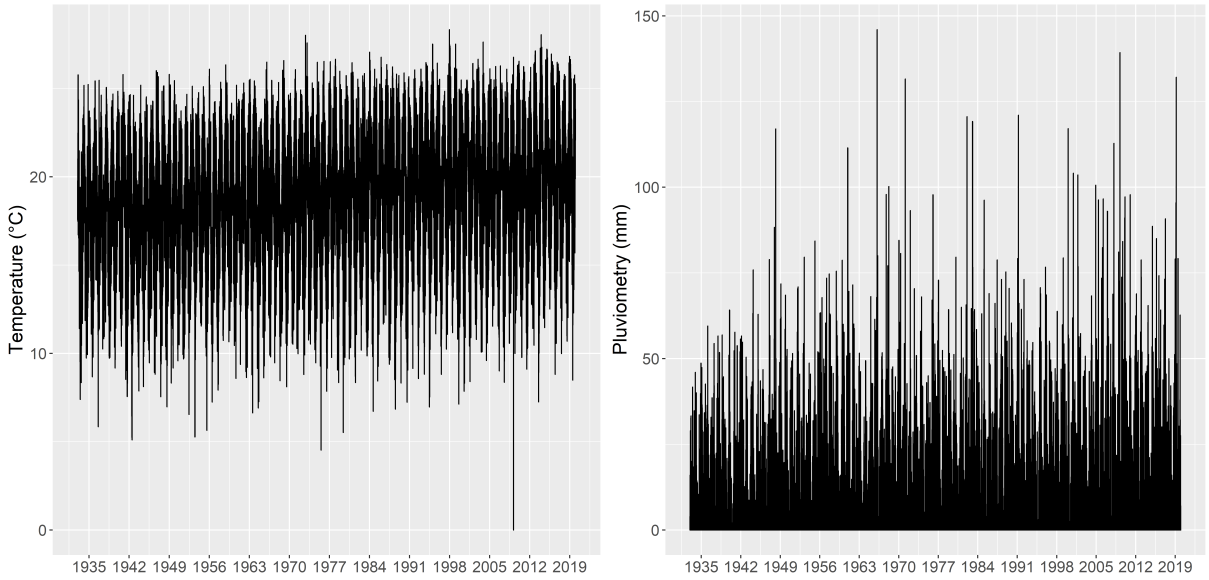
The proposed methodology is based on a decomposition into components of trend, cycle and seasonality for the processes of interest, and thus it is at the same time quite generic, as it can be applied to general processes that present persistent and transient components, and as it is fundamentally a time-varying parameter model has high flexibility and a good fit for general time series, as discussed in Harvey and Shepard (1993). In the form used in our work the models can be placed in the general context of additive processes with random effects structures based on Gaussian Markov Random Fields representations, and as discussed in Rue and Held (2005b) and Rue, Martino and Chopin (2009) this representation allows to model a wide range of latent processes, including time series for Gaussian and non-Gaussian variables, which includes all specifications used in our article, with model for binary variables (rainfall occurrence), positive variables (amount of observed rainfall) , duration (number of days without rain) and extreme values (temperature and rain records). Good fit properties and flexibility of the models based on this representation are discussed in Rue, Martino and Chopin (2009) and Bakka et al. (2018). In the particular case of the structure used in our article, the identification of trend, seasonality and cycle components fundamentally depends on two assumptions - the first is that these components enter into the latent process in an additive way, and the second is that these components are orthogonal, that is, there are no relevant interactions between the components included in the model. To force this structure, we impose some additional restrictions on the model parameterization, imposing that the cycle and seasonality components have zero sum, the trend component variation is not restricted allowing to capture possible non-stationarities in the long-term component, and we assume that the components are independent. These restrictions allow the statistical identification of the decomposition used in our work, and thus respecting the necessary restrictions for estimating unobserved component models, as discussed in Harvey and Shepard (1993)

6.2.2 Data

We used in this work the data from the IAG-USP weather station, which contains observations since 1933, being the longest series of observations available for the São Paulo metropolitan region, and widely used in other relevant works (e.g., Marengo et al. (2020), Dias et al. (2013), Obregón, Marengo and Nobre (2014)). We used daily mean temperature and rainfall data from 01/01/1933 to 12/31/2019, corresponding to 31,776 observations. The temperature corresponds to the daily mean temperature whereas the rainfall data is based on the daily total amount of rainfall. Although there are another rainfall series available for MASP (e.g., data from Mirante de Santana station) we decided to use only the data from the IAG-USP since it is the longest time series available for MASP.

Figure 6.1 shows the temperature and rainfall series analyzed. It is possible to observe in the

temperature series a gradual pattern of increase over time. When observing the series of daily rainfall, we have evidence of an increase in the observed variability of rain, especially in the greater number of extreme events observed, consistent with the results reported in the literature (e.g., Marengo et al. (2020)). Figures showing seasonal and annual patterns of temperature and rain can be seen in the Figures 6.11 and 6.12 in the Appendix.



(a) Daily temperature

(b) Daily rainfall

Figure 6.1 – Average daily temperature and total daily rainfall - IAG Station 1933-2019. Temperature measured in degrees Celsius and rainfall in millimeters.

Before moving to our findings, it is worth spending a few words on what we consider the main limitation of our work. The IAG station database is one of the longest climatic time series in Brazil and, despite the fact that it is great to analyze climatic issues, there is a possibility that this dataset may have been affected by discontinuities and contaminated by some artifacts over time. Regarding the equipment, there was a replacement of instruments that has affected the observations in 1958, with no measurable consequences. Additionally, as discussed by Sugahara et al. (2012), the maximum temperature quantities also present discontinuities around 1990, which may have been caused by urbanization since the observed warming trend in the IAG station was significantly greater than that corresponding to regional climate. On the other hand, according to Sugahara et al. (2012), the precipitation time series does not exhibit any inhomogeneity. Beyond the discontinuities related to the equipment, it is important to highlight that were changes in the time of the day that the samples were collected, between 1933 and 1950. From 1950 until present, the samples are collected every day hourly between 7am and 12pm[1]. Despite the significant limitation, we believe that the problem addressed in our paper is important and timely, and the proposed method can give some new insights to climatic analysis, significantly contributing to the literature of statistical analysis of climate-related issues. In particular, the obtained results provide important evidence about the changes in the pattern of the temperature and rainfall in the MASP.

[1] More details can be found at <http://www.estacao.iag.usp.br/instrumentos.php>

6.3 Results

In this section we describe the application of the methods defined in the Section 6.2.1 based on the data set presented in the Section 6.2.2, which is estimated within the INLA framework. We present herein the obtained results which, in summary, indicate a significant increase in the trend component of daily temperature. In terms of seasonality, the results indicate an increase in the range of seasonal variation of temperature from 1980 onwards, more than doubling the seasonal variability in relation to the beginning of the sample whereas the estimated cyclical component seems to be more stable over time. The analysis for annual temperature records support for the hypothesis of permanent changes in observed climatic patterns, indicating an increase of about 2°C in the trend between the beginning and the end of the sample.

In addition, the results seem to indicate relevant changes in the rain patterns observed in the analyzed sample, both in the probability of occurrence of rains and in the amount of daily rain observed. There is a notable change in patterns regarding the probability of rain, although they are not as straightforward in interpretation as the changes observed in the temperature series. The seasonal pattern with the probability of rain also undergoes important changes, with a reduction in the seasonal variance between 1968 and 1984 and after an increase in the variance of the seasonal effect, also indicating possible changes in weather patterns. The analysis of annual records of daily rainfall shows a pattern of growth over time, consistent with the increase in observed maximum rainfall in the analyzed sample, and thus suggesting a process of climate change consistent with the occurrence of more intense rainfall. On the other hand, from another perspective, there is no ample evidence of changes in the pattern of duration of dry days. Plus, the results indicate that there is no relevant pattern of change in the trend component for the annual maximum of dry days, indicating the stability of this component.

It is worth noting that we believe that our results are quite relevant and helpful in order to understand the dynamics of some climatic variables in the MASP since we provide important analysis of different perspectives about the possible changes in the pattern of rainfall and temperature, considering not only short-term variations, but also the permanent changes. However, it is important to highlight that it is necessary to go further in order to make inference about the relationship between the changes in the patterns of the rainfall and temperature and the urbanization process.

6.3.1 Daily Temperature

We started the analysis with an application of the model described in Section 6.2.1 for daily temperature data. We used a structure of mixed frequencies to perform statistical modeling of the process. The data are observed on a daily frequency, but we formulated the components of trend, seasonality and cycle on a monthly frequency. This formulation is important due to the separation between short-term effects, related to the daily variation of weather, to the patterns of climate change that is the main interest of our work. The use of a monthly frequency allows to recover the patterns of climate change, without the contamination of weather effects. Additionally, this structure allows a better visualization of long-term patterns of climate change, without the noise introduced by possible daily variations related to weather effects. This structure assumes a Gaussian likelihood for temperature observations.

As they are easier to interpret, we focus our attention on the graphs of the estimated components, but tables with the posterior distribution of the estimated parameters are available in the Appendix of the article. The results for the daily average temperature data are shown in Table 6.1 (see Appendix), with the estimated parameters, and in Figure 6.2, which shows the trend, cycle and seasonality components estimated by the model, and the associated 95% credibility interval of the estimation. The precision parameter related to the Gaussian component captures the precision of the unstructured component of our model, which can be thought as all the variability not explained by the trend, cycle, and seasonal components. Thus, it can be interpreted as the daily variation of the temperature not explained by the systematic components of climate variation, and thus capturing the short-term weather effects. The estimated accuracy is relatively low, showing that there is an important variability related to daily weather variations.

Parameters and components associated with trend, cycle and seasonality show a behavior consistent with the existence of long-term changes in weather patterns, especially for the components of trend and seasonality. We can see in subfigure (a) of Figure 6.2 that the trend component shows a pattern of statistically significant growth over time. In particular, the results related to the estimated trend component present an increase of more than 2°C in 2019, when compared with the beginning of the observed sample. While in 1935 the daily mean temperature was slightly above 17.5°C, in 2019 the same measure was almost 20°C. This pattern can be associated with global patterns of climate change, but also more directly with the process of population growth in the municipality of São Paulo. The Figure 6.13 in the Appendix shows the number of inhabitants of the city of São Paulo as reported by the IBGE population census. We can observe a strong growth pattern, which can be related to the movement of urbanization and rural-urban migration observed in the period under analysis. This movement led to a major change in the geography of the metropolitan region, impacting the observed climatic patterns, as discussed by Lima, Lombardo and Magaña (2018). The average annual temperature in the inner city is higher than in the surrounding, due to Urban Heat Island (UHI) Freitas et al. (2007). Under UHI condition[2], the relative heating of the MASP in comparison with surroundings can promote relatively warm air above the urban center and colder convergent air of non-urbanized areas (DRUMOND; AMBRIZZI, 2005).

Regarding the urban land cover, it is important noting that the MASP is characterized by a strong urbanization with a heterogeneous urban structure caused mainly by the rapid growth without proper planning, which can be observed through the contrast of high-rise office towers side by side the poor settlements. Most of the vegetated areas are concentrated in the wealthier neighborhoods while the peripheral areas lack of vegetation. Additionally, the denser vegetation areas, which mitigate the effects of the urban climate, are mainly concentrated outside of urban spot (LIMA; LOMBARDO; MAGAÑA, 2018). The urban area of the MASP is located in

[2] The UHI is based on the definition used in other works (e.g., Memon, Leung and Liu (2009)) given by the differences in the mean hourly temperature between urban/suburban areas and non-urban/vegetated areas. A general characterization of the UHI in MASP can be found in Ferreira et al. (2011) which indicates that the UHI has a predominant daytime character with the highest intensity between 2 pm - 16 pm and a minimum between 7 am - 8 am, and the annual variability ranged from a maximum of 2.6°C in July (4 pm) to 5.5°C in September (3 pm), and the minimum effect of -0.26°C in June (9 am) to 0.94°C in November. However, within the MASP there is variability in the effects of UHI depending on local effects such as urbanization, local circulation, vegetation cover and urban breeze circulation. A more detailed discussion of the spatial and temporal variation of the UHI in MASP can be found in Silva et al. (2017).

a plateau, divided in the crystalline edge and the sedimentary basin. Despite the fact that the crystalline edge suffers from erosion processes without vegetation, this region has recorded the highest population growth in the last decades. The sedimentary basin is characterized by low topographic differences, favoring the human occupation (LIMA; LOMBARDO; MAGAÑA, 2018). Studies assessing the effects of the urbanization on climate in the MASP have provided evidence that urbanization of areas previously covered with natural vegetation can lead to increase the temperature (FREITAS; DIAS, 2005; FREITAS et al., 2007). The changes in the land use reduce the evopotranspiration and increase the heat storage process, causing the UHI phenomenon. This process of warming in the lower part of the atmosphere has impacts on atmospheric humidity, increasing the incidence of days with heavy rain (MARENGO et al., 2010; MARENGO; VALVERDE; OBREGON, 2013).

The change in temperature patterns not only affected the trend, but also the seasonal pattern of temperatures. Figure 6.2 shows the estimated seasonal patterns, where we separate the results into four sub-periods to facilitate visualization. We can see that the model captures important variations in seasonal temperature patterns, not only in the intra-annual pattern, but also in the amplitude of the seasonal effect. We can observe that from 1980 onward an increase in the range of seasonal variation of temperature, more than doubling the seasonal variability in relation to the beginning of the sample. The estimated cyclical component seems to be more stable over time. The cycle period, calculated from the partial correlation coefficients presented in Table 6.1 (see Appendix), was estimated at 10.77 months, consistent with medium-term climatic effects.

6.3.2 Annual Temperature Records

The results obtained in the previous section show important changes in long-term trends and seasonal patterns in the average temperature. The estimated trend shows a significant pattern of growth for average temperatures. But a relevant question is whether the observed temperature extremes follow a similar pattern. To analyze this issue, we used a modified version of the decomposition of latent components to model the annual maximum temperatures observed in the series. Thus, we used a time series of maximum daily temperatures, and extracted the maximum of each year. The time series resulting from this process can be seen in Figure 6.3.

Since our interest were to model the maximums of each year, we made some changes in the model used. The main one is the use of a distribution of Generalized Extreme Values for the process, as described in the Appendix. This modification allows the estimation of the maximum limit distribution in each block (block-maxima), a common procedure adopted to estimate extreme values (COLES et al., 2001). In our case, the distribution is characterized by a time-varying location parameter, given by the sum of the trend and cycle components, a precision parameter and a shape parameter. Since we used the maximum of each year, the seasonal component was eliminated from the model.

The second change was the dependency structure imposed on the trend component. We used a second order random walk (RW2) formulation to represent the trend process. This modification was important to be able to identify the trend component, imposing a smoothness structure for this component. An RW2 structure can be related to a non-parametric spline structure for the trend, and it allows to identify in a more adequate way the persistent patterns

of long-term change in the annual maximums. The RW2 component is given by:

$$\Delta^2 u_t = u_t - 2\ u_{t+1} + u_{t+2}, \quad \mu_t \sim N(0, \eta_\mu). \tag{6.2}$$

Another modification was to use an AR(1) component to capture the cyclic components, replacing the AR(2) structure previously used to capture cycles. Our estimates show that this formulation is more suitable for modeling the dependency structure for the observed annual maximums. The results obtained with this formulation show that the estimation with the RW2-AR(1) structure induces independent residues, which is essential for the GEV process estimation structure used. The parameters estimated in this formulation are shown in Table 6.2 in the Appendix.

The precision parameters for the estimated trend and cycle are shown in Figure 6.4. We can observe that the trend component is consistent with the increase in temperature records observed in the historical series, indicating an increase of about 2°C in the trend between the beginning and the end of the sample which is compatible with other finding previously reported in the literature (e.g., Drumond and Ambrizzi (2005)). The AR(1) component is also very relevant, being very persistent, with a posterior average of the estimated coefficient equals 0.76, and a range of values between approximately -2 and 2 degrees. The analysis for annual temperature records again provides evidence for the hypothesis of permanent changes in observed climatic patterns.

6.3.3 Daily Pluviometry

To carry out the analysis of rainfall patterns, we built a model that allows the simultaneous analysis of the probability of rain occurrence and the amount of rain on a certain day, given that rain has occurred. This model is composed of two likelihood structures, the first being a Bernoulli for the occurrence of rain, denoted by B_t, and a Gamma likelihood for the amount of rain in the day, denoted by Q_t, conditional on the occurrence of rain in the Bernoulli process. This structure is also known as Hurdle model. Similar to the model used in subsection 6.3.1, we used a latent factor structure of trend, seasonality and cycle to capture possible changes in both the probability of rain and the rainfall intensity.

In this structure, we model the probability of rain occurring using a structure that allows the probability parameter to vary over time. This probability is given by the logistic transformation of the sum of latent factors of trend, seasonality and cycle, whose sum corresponds to a linear predictor for the probability of rain. The second part of the model corresponds to a distribution for the amount of rain observed on the day, conditional on the rain on that day. In this case the average of the Gamma distribution is given by the exponential of the linear predictor given by the sum of the components of trend, seasonality and cycle. The exponential transformation ensures that the predictor is always positive for the amount of rain observed each day.

The model specification used is presented in Equation 6.3. We assume that the latent components are independent between the specifications of the probability of rainfall and the amount of rainfall. We tested other specifications by imposing restrictions and common factors, but the specification assuming independence showed the best adjustment results.

$$
\begin{aligned}
Prob(B_t = 1) &= p_t^B(1 - p_t^B) \\
p_t^B &= \frac{exp(\upsilon_t^B)}{1 + exp(\upsilon_t^B)} \\
\upsilon_t^B &= \mu_t^B + s_t^B + c_t^B \\
\mu_t^B &= \mu_{t-1}^B + \eta_\mu^B \\
s_t^B &= s_{t-1}^B + s_{t-2}^B + \ldots + s_{t-m}^B + \eta_s^B \\
c_t^B &= \theta_1^B c_{t-1} + \theta_2^B c_{t-2} + \eta_c^B \\
f(Q_t|B_t = 1) &= \frac{b^a}{\Gamma(a)} y^{a-1} \exp -bQ_t \\
E(Q_t|B_t = 1) &= a/b = \exp(\upsilon_t^Q) \\
Var(Q_t|B_t = 1) &= 1/\tau = a/b^2 \\
\upsilon_t^Q &= \mu_t^Q + s_t^Q + c_t^Q \\
\mu_t^Q &= \mu_{t-1}^Q + \eta_\mu^Q \\
s_t^Q &= s_{t-1}^Q + s_{t-2}^Q + \ldots + s_{t-m}^Q + \eta_s^Q \\
c_t^Q &= \theta_1^Q c_{t-1} + \theta_2^Q c_{t-2} + \eta_c^Q. \qquad (6.3)
\end{aligned}
$$

The results obtained with this model are presented in Table 6.3 in the Appendix, and are separated in two blocks, the first being the parameters associated with the probability of rain, and the second block related with the parameters referring to the amount of rain. Note that the parameters refer to the linear predictors of the Bernoulli and Gamma models. Since the interpretation is more evident for the estimated latent components, we will focus on analyzing these results.

Figure 6.5 shows the estimated trend, cycle and seasonal components for the rain probability with the seasonal component separated into sub-periods to facilitate the visualization and interpretation of the results. We can observe that there is a great variation in the trend component associated with the probability of rain, and also a cyclical component with relevant variation over time, which has an estimated period of 11.99 months. There is a notable change in the patterns regarding the probability of rain, although they are not as straightforward interpreted as the changes observed in the temperature series. Moreover, we can also observe in Figure 6.5 that the seasonal pattern with the probability of rain undergoes important changes, with a decrease in the seasonal variance between 1968 and 1984 and after an increase in the variance of the seasonal effect, also indicating possible changes in weather patterns. In order to facilitate the interpretation of the results, the Figure 6.14 (see Appendix) shows the reverse logistic transformation that allows to directly interpret the trend, cycle and seasonal components in rain probability.

The cyclical component captures the common effect of all events in the rainfall series that are neither associated with the long-term trend component nor with seasonal effects. In this aspect, it can estimate in an aggregated way the impact of weather events such as ENSO on the rainfall pattern. Years with higher values observed in the cycle component for accumulated rainfall may be associated with periods with higher El Niño–Southern Oscillation (ENSO) in-

tensity, as discussed for example in Grimm and Tedeschi (2009) and Ropelewski and Bell (2008), which indicate that years with higher El Niño intensity are associated with higher volumes of accumulated daily rainfall, and years with higher La Niña intensity or neutral years with lower accumulated daily volumes. As pointed out by Shifts seasons associated with El Niña years presents fewer days with rainfall less than 5 mm and more days with rainfall greater than 20 mm, compared with the seasons during La Niña years. This pattern can be observed by the positive values in the cycle component in the years 1957-1959, 1990-1995 and 2015-2016, associated with more intense occurrences of El Nino, with accumulated values consistent with the results obtained in Ropelewski and Bell (2008). However, it is important to note that this interpretation needs further study, as our cyclical component aggregates all sources of systematic variation not explained by trend and seasonality components, and thus the composition of various climate effects.

The components estimated for the amount of rain are shown in Figure 6.6, with the trend, cycle and seasonal components. Using the posterior average of the estimated component, the trend component indicates an amount of rain of around $\exp(1.92) = 5.05$mm of rain at the beginning of the sample, reaching about $\exp(1.96) = 7.09$mm of rain at the end of the sample, which is a very relevant change in long-term patterns. The cyclical component is also quite relevant, with a relatively high magnitude and an estimated period of 9.667 months. The seasonal pattern shows some variation over time, but it seems to be more stable.

In agreement with previous studies, the results provide evidence of relevant changes in the rain patterns observed in the analyzed sample, both in the probability of occurrence and in the amount of daily rain observed (e.g., Nobre et al. (2010), Obregón, Marengo and Nobre (2014), Raimundo, Sansigolo and Molion (2014), Lima and Rueda (2018), Lima, Lombardo and Magaña (2018)).

6.3.4 Annual records - Daily rainfall

Similar to the analysis of annual temperature records performed in the subsection 6.3.2, we built a GEV model with latent trend and cycle components to analyze possible changes in the annual rainfall maximums observed in the sample. This analysis is particularly important since large amounts of rainfall over short periods may cause serious social and economic problems, such as floods, landslides and other related problems (MARENGO et al., 2020).

Figure 6.7 shows the annual maximum daily rainfall observed in the analyzed sample. We can see that there is an apparent increase in the pattern of maximum rainfall until early 1970s, a stabilization until mid-1990s and after an apparent growth in the trend, although this is quite irregular.

To analyze this question, we used a model that is composed of a trend process using an RW2 structure and a temporary shock process given by an AR(1) process in a GEV likelihood structure. Similar to the results for the maximum temperatures, this specification obtained the best results for the specification of the observed maximum. The estimated parameters for this model are shown in Table 6.4 (see Appendix).

The shape parameter estimated for GEV was estimated with a posterior mean of -0.155, which corresponds to the values in the attraction domain of a Weibull distribution of extremes.

The parameter associated with the first order autoregressive coefficient was estimated with a posterior mean of 0.344, which implies some persistence between annual rainfall maximums, but less than the value observed for persistence in temperature maximums.

The results for the observed trend and cycle can be seen in Figure 6.8. The model estimates a trend that shows a pattern of growth over time, consistent with the increase in observed maximum rainfall in the analyzed sample, and thus indicating a process of climate change consistent with the occurrence of more intense rainfall. The AR(1) component has a very large variability, with values between approximately -40mm and 60mm, which indicate the importance of this component in the rainfall structure. These results are consistent with previous studies (DIAS et al., 2013; MARENGO; VALVERDE; OBREGON, 2013; DUFEK; AMBRIZZI, 2008) which have found evidence of an increase in total precipitation from more intense rainfall and daily rainfall extremes.

Our results for daily pluviometry and precipitation extremes are consistent with other works that have analyzed the relationship between urbanization and changes in precipitation patterns, such as Huff and Changnon (1973), Zhu et al. (2019), Zhang et al. (2019) and Singh et al. (2020), and the effects of urban heat island (e.g., Lin et al. (2011), Gu and Li (2018)) in these processes, indicating a sustained increase in the trend component and changes in patterns seasonal variations for the daily amount of rainfall and a relevant increase in the trend component for observed annual rainfall extremes. In particular, our statistical methodology indicates that the growth in the trend component is a permanent effect of the urbanization process on the rainfall pattern.

6.3.5 Dry days duration

In the previous two sections we analyzed the probability of rainfall, the amount observed and the maximum rainfall for the data in our sample. It is possible to analyze this issue from another perspective, analyzing the consecutive number of days without rain, which is a measure associated with periods of drought. This is a very relevant point, since long periods without rain are related to problems of water availability, water supply and also to health problems related to low humidity.

For this analysis we defined a day without rain as a day when the accumulated precipitation was less than 1mm. From this definition, we calculate the duration in days of absence of rain following these definitions. Figure 6.9 shows the duration of days without rain observed in our sample.

To perform the analysis, we performed a modification in the decomposition of trend, seasonality and cycle using a modified survival analysis structure for a dynamic structure. The main modification is to use an appropriate distribution for censored data. We report below the results obtained using a lognormal distribution of survival, defined by:

$$f(y_t) = \frac{1}{y_t\sqrt{2\pi}}\sqrt{\tau}\exp\left(-\frac{1}{2}\tau(\log y_t - \upsilon_t)^2\right. \tag{6.4}$$

The decomposition is applied to the linear predictor for the log duration, denoted by υ_t, in the form:

$$
\begin{aligned}
\upsilon_t &= \mu_t + s_t + c_t \\
\mu_t &= \mu_{t-1} + \eta_\mu \\
s_t &= s_{t-1} + s_{t-2} + \ldots + s_{t-m} + \eta_s \\
c_t &= \theta_1 c_{t-1} + \theta_2 c_{t-2} + \eta_c
\end{aligned} \tag{6.5}
$$

Note that the duration is defined only on rainy days, and on the others, it is treated as not observed. The duration is determined on the day of rain occurrence, which allows to use the entire sample available in the estimation and thus the use of components in the frequency of choice of analysis. The latent components μ_t, s_t and c_t are defined on a monthly basis, in a similar way to that used in the models in sections 4.1 and 4.3.

Specifying a lognormal distribution for the duration allows a more direct interpretation of the effect of the υ_t linear predictor on the duration. An increase in the value of υ_t represents an increase in the expected duration, whereas decreases in υ_t indicate reductions in the expected duration. We tested other specifications for the duration process, such as the Weibull distribution, but the best results were obtained with the lognormal distribution.

The estimated parameters for the model of duration of days without rain are presented in Table 6.5 in the Appendix. The precision for the lognormal distribution is relatively low, consistent with the great variability observed in duration. The estimated seasonal component is quite regular, indicating that there were no relevant changes in the seasonal effect on the pattern of rainfall. The cyclic component has a low magnitude, which indicates that the model cannot capture a relevant pattern in this component.

We can see in Figure 6.10 the components estimated by the model. The trend component shows some variation over time, but this variation is not relevant. For example, looking at the posterior average of this component, we have at the beginning of the sample an expected duration given by the trend component of approximately $\exp(.86) = 2.32$ days, while at its lowest value in the early 1980s, the duration indicated by the trend would be approximately $\exp(.775) = 2.17$ days, indicating a minor variation, and so there is no empirical evidence of changes in the regular patterns of dry days.

We present an analysis of extremes of dry days duration in the Appendix.

6.4 Conclusions

The changes in the likelihood of the occurrence and/or strength of extreme weather and climate event can be partly due to natural climate variability, human-induced climate changes and/or urbanization process. The effects of climate-related disasters are often intensified in cities due to cultural, demographic and economic characteristics. In the metropolitan area of São Paulo, which encompasses 39 municipalities in the state of São Paulo with intense process of conurbation and represents the largest urban concentration in South America, floods and droughts are the most critical climate-related issues. Previous studies have identified a positive trend in both, annual and seasonal rainfall whereas other works have identified an increase in both rainy days and extreme daily precipitation, decrease in light rain, and increase in the

number of consecutive dry days, indicating that intense precipitation is becoming concentrated on a few days with longer dry periods occurring in between.

This article contributes to this literature by proposing a novel statistical decomposition of climatic time series into long-term trend, seasonal and cycle components to analyze the existence of changes in rainfall and temperature patterns. We applied this method to analyze the existence of changes in rainfall and temperature patterns in the MASP, which allows to identify statistically permanent changes in the rainfall and temperature over time and also to capture the impact of non-permanent shocks on temporal patterns. We analyze daily series of temperature and rainfall between 1933 and 2019 from Institute of Astronomy, Geophysics and Atmospheric Sciences of University of São Paulo (IAG-USP), being the longest series of observations available for the MASP. The results indicate a significant increase in the trend component of daily temperature. The seasonality also presents changes, in which the results indicate an increase in the range of seasonal variation of temperature from 1980 onwards, more than doubling the seasonal variability in relation to the beginning of the sample.

The results also suggest relevant changes in the rain patterns observed in the analyzed sample, both in the probability of occurrence of rains and in the amount of daily rain observed. There is a notable change in patterns regarding the probability of rain, although they are not as straightforward in interpretation as the changes observed in the temperature series. The seasonal pattern with the probability of rain also undergoes important changes, with a reduction in the seasonal variance between 1968 and 1984 and after an increase in the variance of the seasonal effect, also indicating possible changes in weather patterns. In addition, the analysis for annual temperature records and annual records of daily rainfall show support for the hypothesis of permanent changes in observed climatic patterns. However, there is no ample evidence of changes in the pattern of duration of dry days. Plus, the results indicate that there is no relevant pattern of change in the trend component for the annual maximum of dry days, indicating the stability of this component.

6.5 Appendix

6.5.1 Characteristics of Temperature and Rain

Figures 6.11 and 6.12 present boxplots representing the distributions of average daily temperature and accumulated rainfall, respectively. Each boxplot presents the first and third quartiles and the median of the observed data. Points in the figure represent outliers, defined as observations greater than or less than 1.5 times the interquartile range between the first and third quartiles. In Figure 6.11 one may observe the temperature series by year and months, in which is possible to note the variability among the months and an increase in the daily temperature from 1933 to 2018. Similarly, Figure 6.12 shows the rainfall series by month and year, which is possible to note the heterogeneity among the months and years. In Figure 6.12 we can observe that in the rainfall data there is a large number of outliers, and to facilitate visualization in Figure 6.12, subfigures (c) and (d) show the boxplots without the points corresponding to the outliers.

6.5.2 Census population

Figure 6.13 shows the number of inhabitants of the city of São Paulo from 1920 to 2010, reported by the *Instituto Brasileiro de Geografia e Estatistica* (IBGE) census.

6.5.3 Generalized Extreme Value (GEV) Distribution

This section briefly introduces the Generalized Extreme Value (GEV) distribution and for a more detailed discussion see Coles et al. (2001). Define $M_n = \max(X_1, \ldots, X_n)$ as the maximum over a block (block-maxima) of n values, where $X_1, \ldots, X_n$ is a sequence of independent and identically-distributed random variables. Under certain conditions, the extremal types theorem states that M_n converges to a specific distribution, known as generalized extreme value (GEV) distribution, which has the following form

$$G(z) = \exp\left\{-\left[1 + \xi(\frac{z-\mu}{\sigma})\right]^{-1/\xi}\right\} \tag{6.6}$$

where $-\infty < \mu < \infty$ is a location parameter, $\sigma > 0$ represents a scale parameter and $-\infty < \xi < \infty$ is a shape parameter.

6.5.4 Bernoulli-Gamma (Hurdle) model - Transformed probabilities

Figure 6.14 shows the reverse logistic transformation of the estimated trend, seasonality and cycle components of Hurdle model, in which can be observed a quite irregular behavior of the trend component.

6.5.5 Survival Models

Survival models (SM) are interesting in the context where the interest is to model the time until a certain event happens (GÓMEZ-RUBIO, 2020). In this paper, we aim to model the number of dry days in the MASP, adopting a SM using the trend-cycle decomposition for climatic model (LAURINI, 2019).

Considering T as the random variable that measures time to event, the survival function $S(t)$ can be defined as

$$S(t) = P(T > t), \tag{6.7}$$

considering the implicit assumptions that $S(0) = 1$ and $\lim_{t\to+\infty} S(t) = 0$. The survival function is related to a distribution function

$$F(t) = P(T < t) = 1 - S(t) \tag{6.8}$$

with density function $f(t)$. In addition, the hazard function $h(t)$ measures the instantaneous risk that the event of interest occurs at a small instant of time and can be expressed as $h(t) = f(t)/S(t)$. The cumulative hazard $H(t)$ can be defined as

$$H(t) = \int_0^t h(u)du. \tag{6.9}$$

Hence, the survival function can be expressed in terms of the cumulative function as

$$S(t) = \exp(-H(t)), \tag{6.10}$$

and the likelihood of this model can be expressed as

$$\sum_{i=1}^{n} f(t_i)^{\delta_i} S(t_i)^{1-\delta_i} \tag{6.11}$$

where t_i is time of event i and δ_i indicates whether the event was observed or not.

6.5.6 Extremes of dry days duration

The decomposition of the components of trend, cycle and seasonality did not indicate relevant changes in the pattern of duration of days without rain. However, a relevant issue is to verify the possibility of changes in the behavior of extreme duration, i.e., long periods without rain. To analyze this issue, we again use a Generalized Extreme Value distribution structure, similar to that used in sections 4.2 and 4.4. Our variable under analysis is the maximum duration of days without rain in each year (maximum annual block). Figure 6.15 shows the maximum duration series of days without rain.

The best specification for the annual maximum process for days without rain is presented in Table 6.6. The results obtained in the specification analysis indicate that there is no cyclic component in this process, and the best specification for the trend is a RW1 model. The precision of the GEV component is quite low, compatible with the high variability observed in the series of maximum duration, and the shape parameter was estimated with mean equals 0.147, indicating a distribution in the domain of attraction of a Frechet family, which suggests a distribution with more extreme values.

The estimated trend component is shown in Figure 6.16, which indicates that there is no relevant pattern of change in the long-term component, suggesting the stability of this component.

6.5.7 Posterior distribution of estimated parameters

The posterior distribution of the latent components and parameters are obtained using Bayesian estimation methods through the use of integrated nested Laplace approximations, described in Rue, Martino and Chopin (2009). We use the default prior structure of the r-inla package, and the results are robust to this specification. In particular, we use Gamma priors for the precision parameters for the Gaussian GEV and lognormal distributions, and Gaussian priors for the GEV shape parameter and the autoregressive parameters of the cycle components. The autoregressive parameters are represented by partial autocorrelations in the estimation process. Specific values of the hyperparameters used are available under request from the authors.

Table 6.1 – Posterior distribution of estimated parameters - Daily temperature. Table reports the mean, standard deviation and the quantiles of posterior distribution of estimated parameters using the INLA approximation.

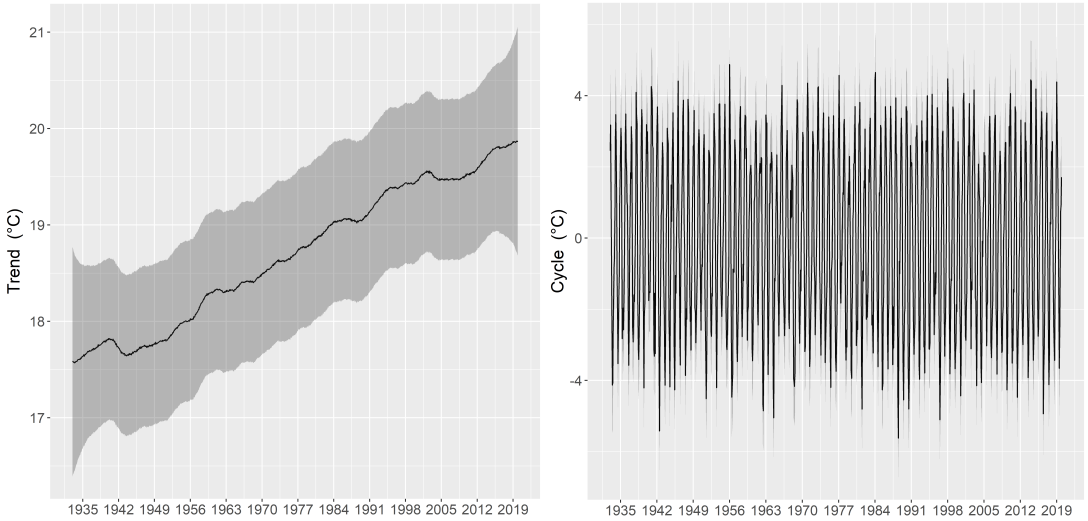
	Mean	SD	0.025q	0.5q	0.975q	Mode
Precision for Gaussian obs	0.187	0.002	0.184	0.187	0.190	0.187
Precision for trend	89.935	9.786	71.222	89.861	109.548	90.218
Precision for seasonal	7750.445	1201.137	5346.370	7802.845	9953.298	8054.902
Precision for cycle	0.156	0.010	0.138	0.155	0.178	0.154
PACF1 for cycle	0.810	0.010	0.789	0.811	0.828	0.813
PACF2 for cycle	-0.614	0.022	-0.656	-0.614	-0.569	-0.615
log-Likelihood	-73011.82					

Table 6.2 – Posterior distribution of estimated parameters - Annual maximum temperatures. Table reports the mean, standard deviation and the quantiles of posterior distribution of estimated parameters using the INLA approximation.

	Mean	SD	0.025quant	0.5quant	0.975quant	Mode
Precision for GEV obs.	5.46e+01	0.000	5.46e+01	5.46e+01	5.46e+01	5.46e+01
Shape for GEV obs.	-2.56e-01	0.000	-2.56e-01	-2.56e-01	-2.56e-01	-2.56e-01
Precision for trend	1.67e+05	0.373	1.67e+05	1.67e+05	1.67e+05	1.67e+05
Precision for cycle	5.51e+01	0.015	5.51e+01	5.51e+01	5.52e+01	5.51e+01
AR(1) for cycle	7.61e-01	0.000	7.60e-01	7.61e-01	7.62e-01	7.62e-01
log-Likelihood	-15488.63					

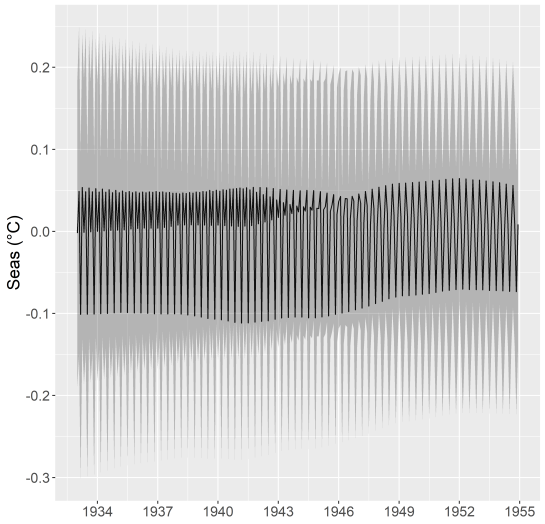
Table 6.3 – Posterior distribution of estimated parameters - Bernoulli-Gamma (Hurdle) model for probability and quantity of rainfall. Table reports the mean, standard deviation and the quantiles of posterior distribution of estimated parameters using the INLA approximation.

	Mean	SD	0.025quant	0.5quant	0.975quant	Mode
Probility of Rain						
Precision for trend	24.606	5.85e+00	16.230	23.505	3.89e+01	21.198
Precision for seasonal	10722.848	7.17e+03	3239.238	8764.001	2.96e+04	6250.269
Precision for cycle	2.993	1.62e+00	0.724	2.721	6.86e+00	1.961
PACF1 for cycle	0.866	1.00e-03	0.864	0.866	8.68e-01	0.866
PACF2 for cycle	-0.999	1.00e-03	-1.000	-0.999	-9.97e-01	-0.999
Quantity Rain						
Precision for the Gamma obs.	0.476	5.00e-03	0.465	0.476	4.85e-01	0.478
Precision for trend	12507.421	9.61e+03	3174.863	9784.494	3.78e+04	6521.456
Precision for seasonal	42016.481	3.23e+04	9897.723	32952.985	1.23e+05	21585.470
Precision for cycle	3.771	2.83e-01	3.186	3.787	4.29e+00	3.859
PACF1 for cycle	0.675	2.20e-02	0.631	0.675	7.17e-01	0.674
PACF2 for cycle	-0.307	3.90e-02	-0.386	-0.306	-2.32e-01	-0.302
log-Likelihood	-64549.85					

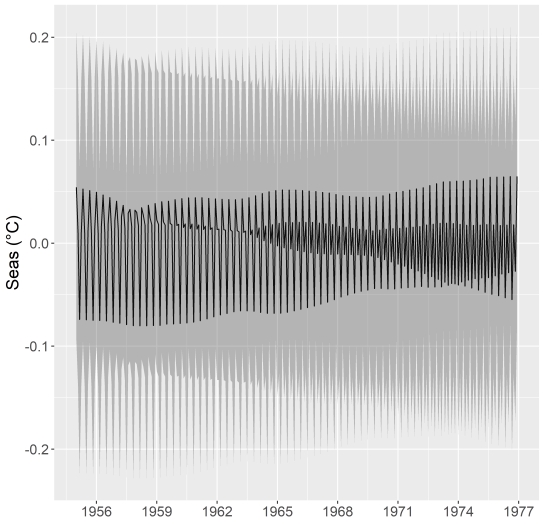


(a) Trend

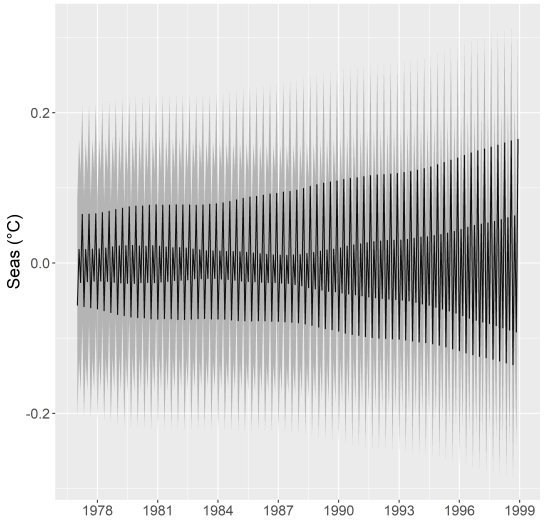
(b) Cycle



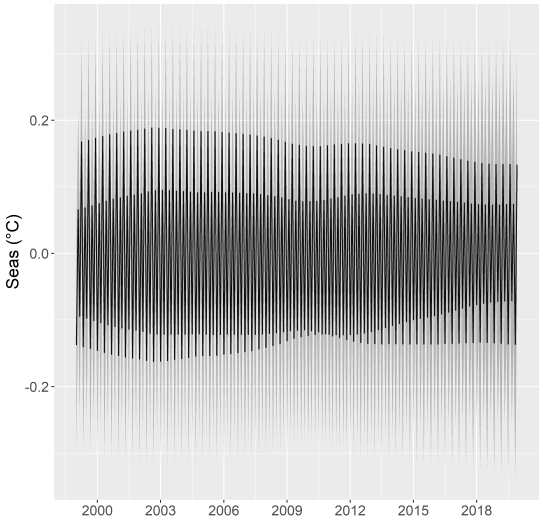
(c) Seasonal effect - 1933-1944



(d) Seasonal effect - 1955-1976



(e) Seasonal effect - 1977-1998



(f) Seasonal effect - 1999-2019

Figure 6.2 – Estimated Trend, Cycle and Seasonal Components for Average Daily Temperatures at IAG Station - 1933-2019. The graphs show the posterior mean (solid line) of the estimated component and the credibility interval of 95% (shaded ribbon). Measurements in degrees Celsius.

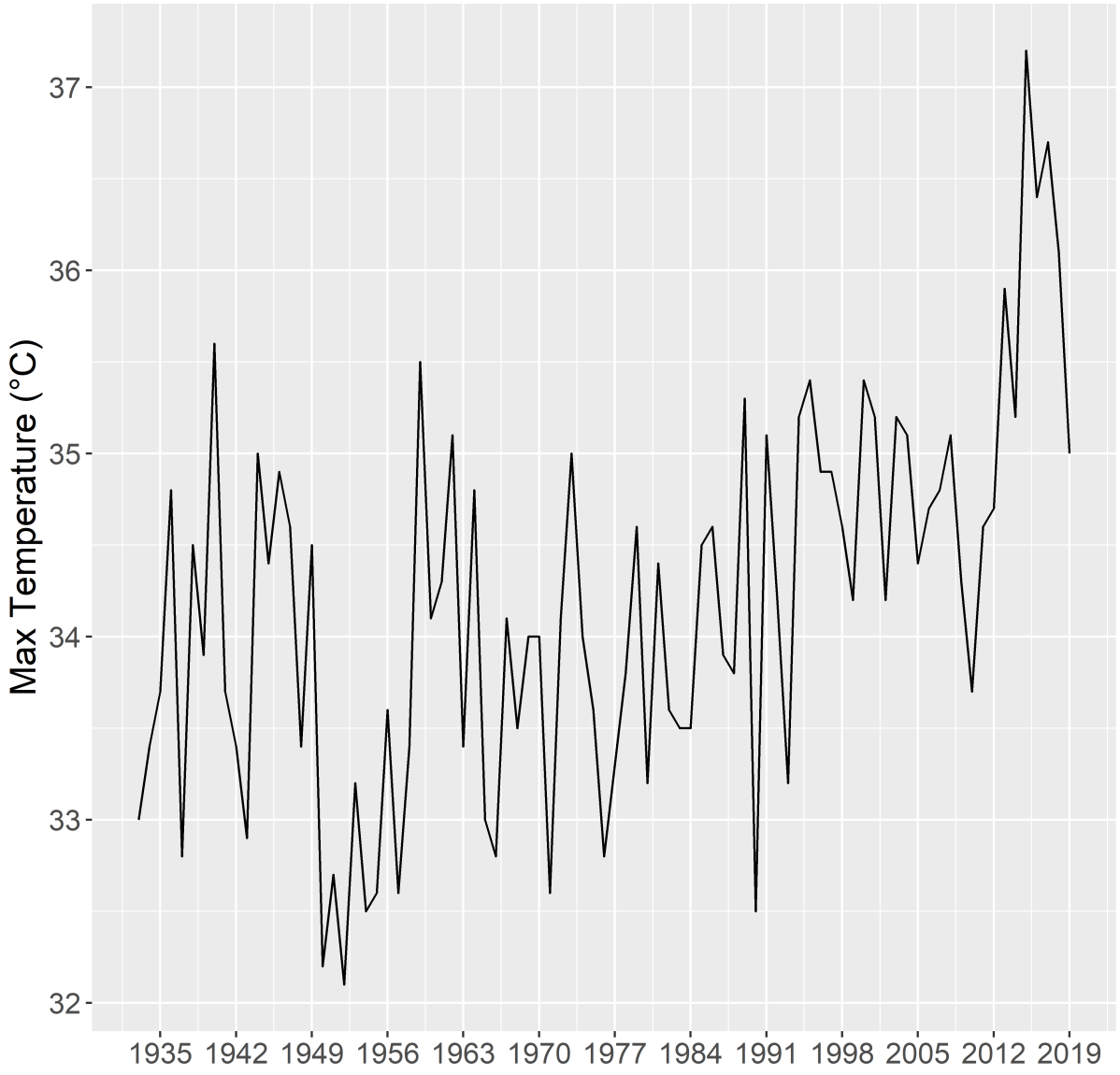
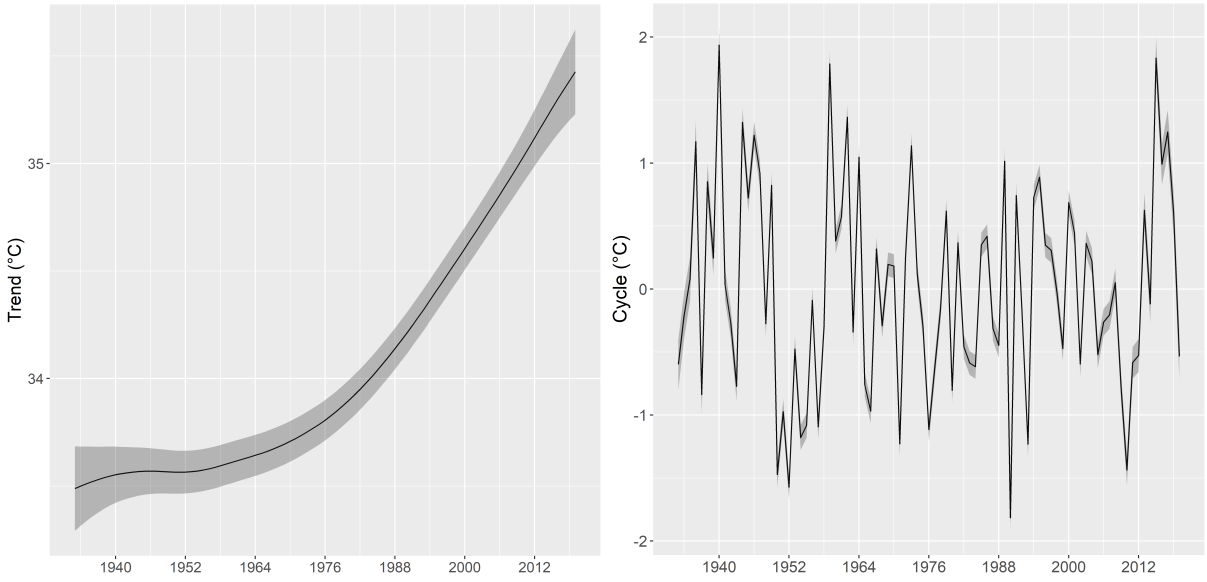


Figure 6.3 – Annual maximum temperatures at IAG Station - 1933-2019. The graph shows the annual temperature record for each year (annual block maxima), measured in degrees Celsius.

Table 6.4 – Posterior distribution of estimated parameters - Annual maximum daily rainfall. Table reports the mean, standard deviation and the quantiles of posterior distribution of estimated parameters using the INLA approximation.

	Mean	SD	0.025quant	0.5quant	0.975quant	Mode
Precision for GEV obs.	219.575	57.380	109.743	220.916	323.915	225.333
Shape for GEV obs.	-0.155	0.002	-0.159	-0.155	-0.152	-0.155
Precision for trend	2.882	0.467	1.983	2.886	3.792	2.927
Precision for cycle	0.002	0.000	0.001	0.002	0.002	0.002
AR(1) for cycle	0.344	0.055	0.237	0.344	0.452	0.341
log-Likelihood	-434.73					



(a) Trend

(b) Cycle

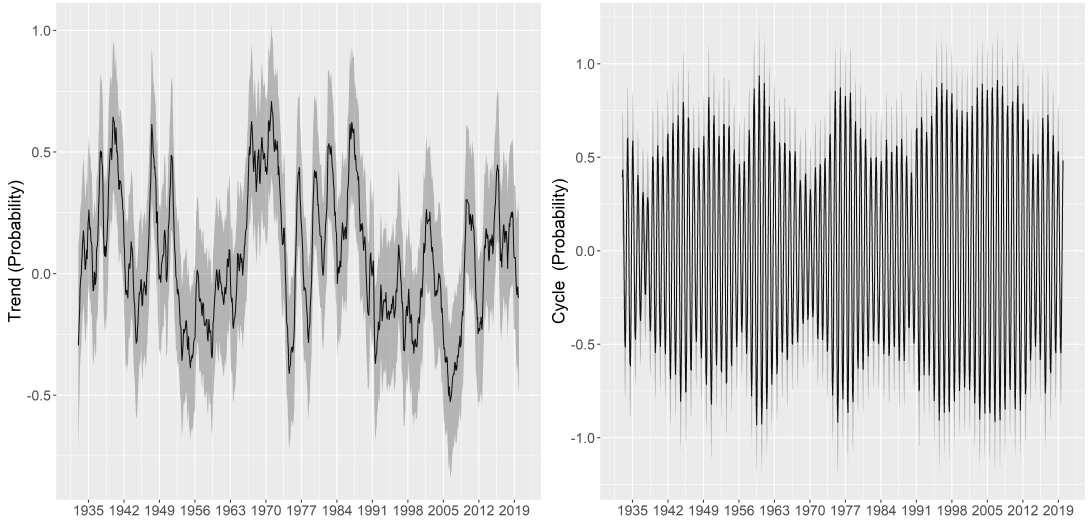
Figure 6.4 – Trend-Cycle decomposition for annual maximum temperatures at IAG Station 1933-2019. The graphs show the posterior mean (solid line) of the estimated components and the credibility intervals of 95% (shaded ribbon). Measurements in degrees Celsius.

Table 6.5 – Posterior distribution of estimated parameters - Dry days duration. Table reports the mean, standard deviation and the quantiles of posterior distribution of estimated parameters using the INLA approximation.

	Mean	SD	0.025quant	0.5quant	0.975quant	Mode
Precision for the lognormsurv obs.	1.404	2.00e-02	1.365	1.404	1.45e+00	1.403
Precision for trend	42139.008	1.13e+04	22054.639	41680.366	6.55e+04	40743.639
Precision for seasonal	71231.197	5.42e+04	20124.558	55394.099	2.15e+05	37324.778
Precision for cycle	358.205	1.24e+02	143.796	352.438	6.10e+02	329.995
PACF1 for cycle	0.376	1.19e-01	0.153	0.372	6.11e-01	0.349
PACF2 for cycle	-0.350	1.63e-01	-0.679	-0.339	-6.40e-02	-0.238
log-Likelihood	-18845.15					

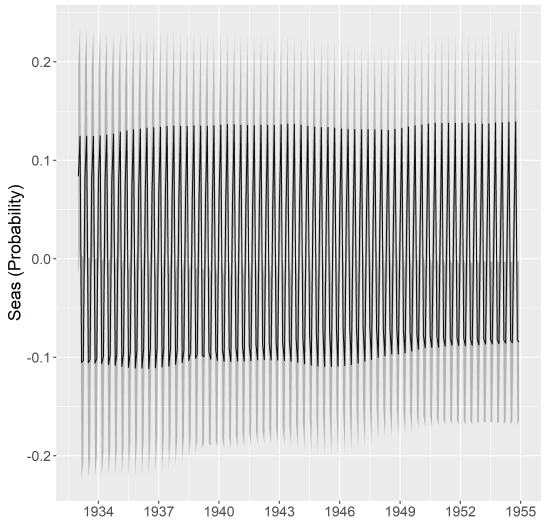
Table 6.6 – Posterior distribution of estimated parameters - Annual maximum - Dry days duration. Table reports the mean, standard deviation and the quantiles of posterior distribution of estimated parameters using the INLA approximation.

	Mean	SD	0.025quant	0.5quant	0.975quant	Mode
Precision for GEV obs.	0.005	0.000	0.005	0.005	0.006	0.005
Shape for GEV obs.	0.147	0.002	0.144	0.147	0.152	0.145
Precision for trend	51.097	6.872	36.426	51.816	62.059	55.141
log-Likelihood	-365.59					

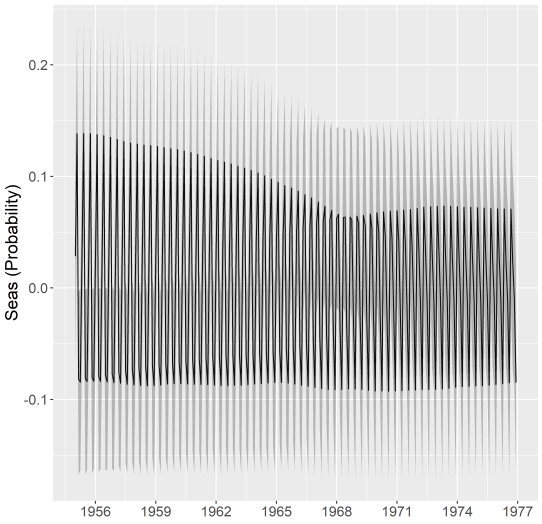


(a) Trend

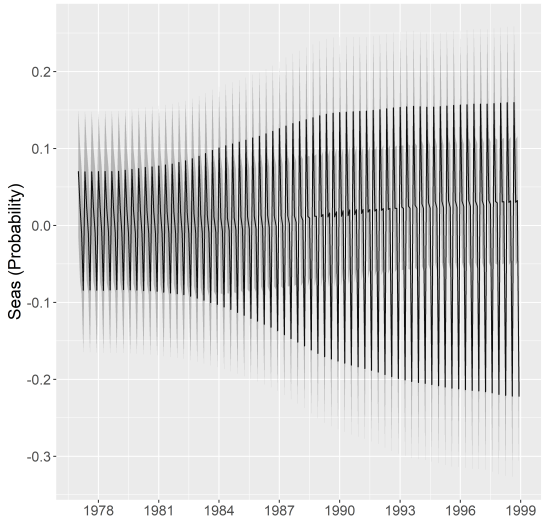
(b) Cycle



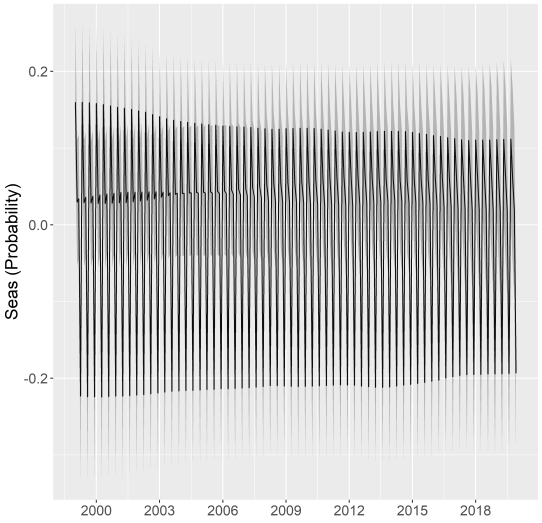
(c) Seasonal effect - 1933-1944



(d) Seasonal effect - 1955-1976

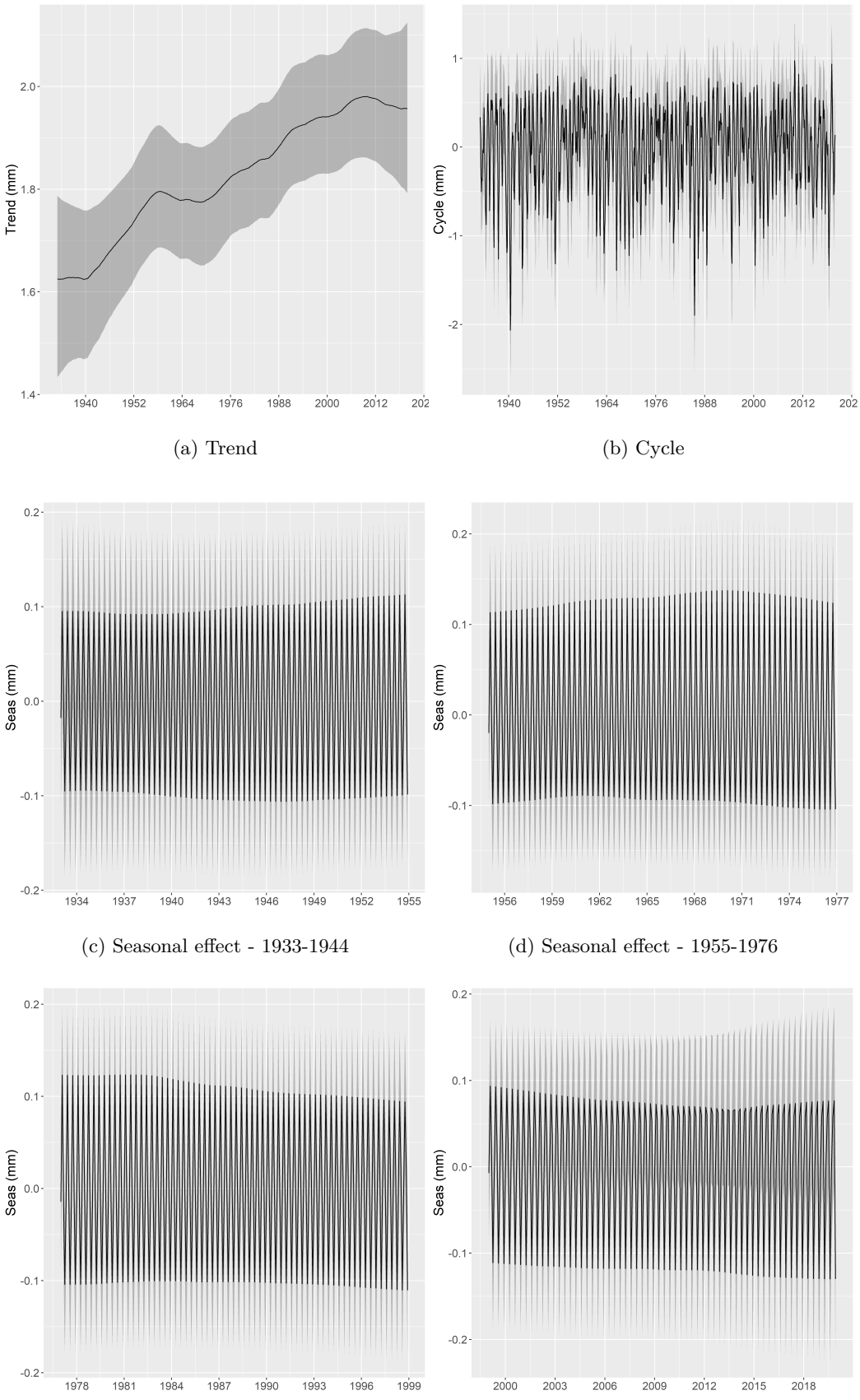


(e) Seasonal effect - 1977-1998



(f) Seasonal effect - 1999-2019

Figure 6.5 – Trend, Cycle and Seasonal decomposition for the Bernoulli-Gamma (Hurdle) model. The graphs show the posterior mean (solid line) of the estimated components and the credibility intervals of 95% (shaded ribbon) for the probability of rain in each day.



(a) Trend

(b) Cycle

(c) Seasonal effect - 1933-1944

(d) Seasonal effect - 1955-1976

(e) Seasonal effect - 1977-1998

(f) Seasonal effect - 1999-2019

Figure 6.6 – Trend, Cycle and Seasonal decomposition for the Bernoulli-Gamma (Hurdle) model. The graphs show the posterior mean (solid line) of the estimated components and the credibility intervals of 95% (shaded ribbon) for the daily accumulated rainfall quantity (measured in millimeters).

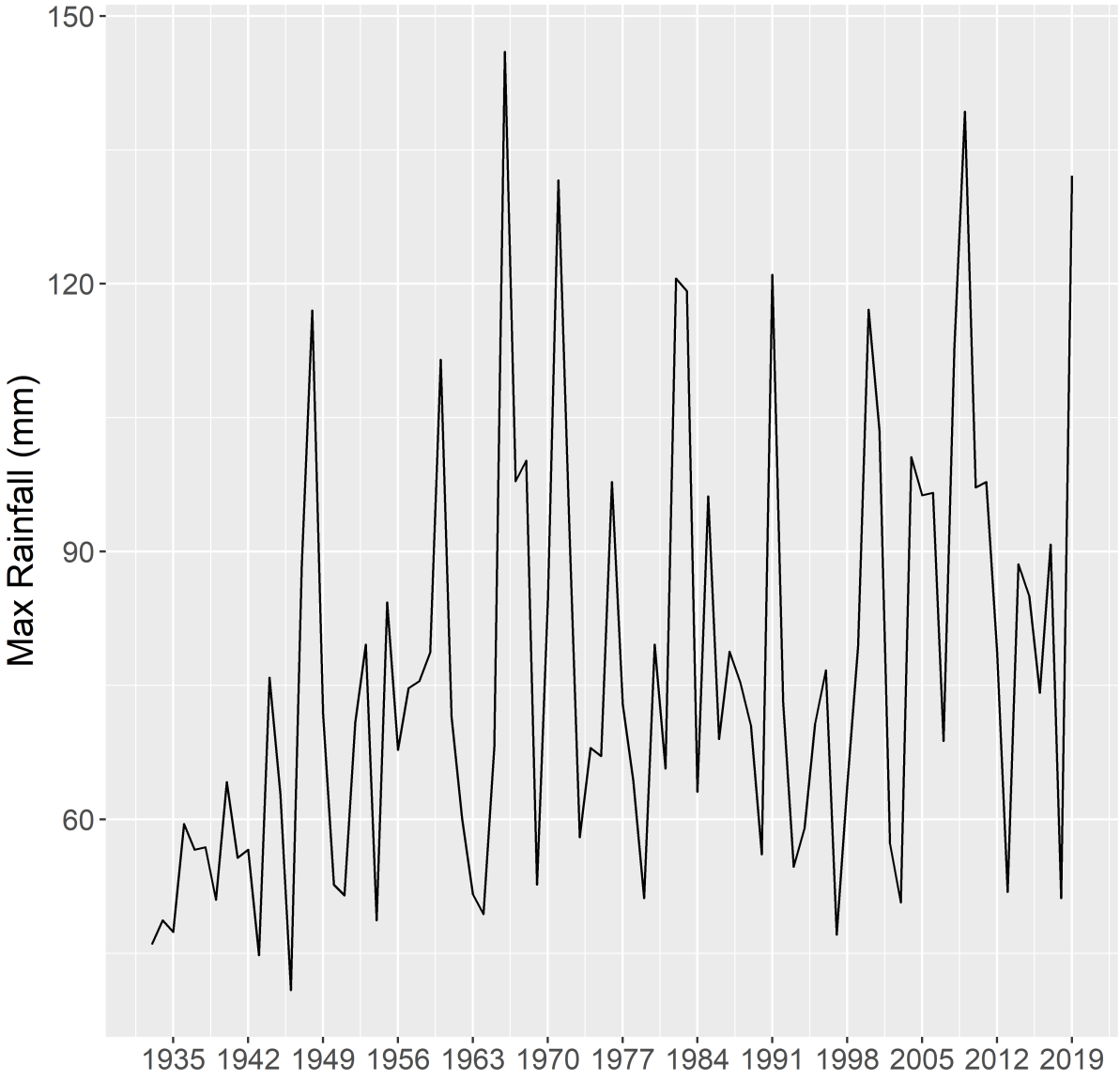
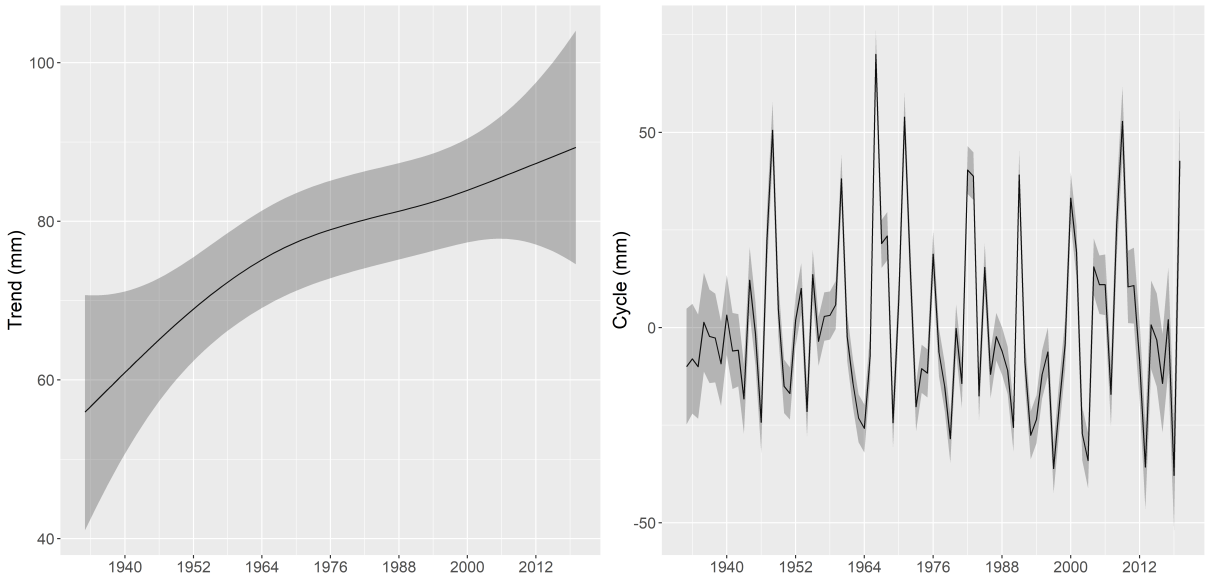


Figure 6.7 – Annual maximum daily rainfall at IAG Station - 1933-2019. The graph shows the annual rainfall record for each year (annual block maxima), measured in millimeters.



(a) Trend

(b) Cycle

Figure 6.8 – Trend-Cycle decomposition for annual maximum daily rainfall at IAG Station 1933-2019. The graphs show the posterior mean (solid line) of the estimated components and the credibility intervals of 95% (shaded ribbon). Measurements in millimeters.

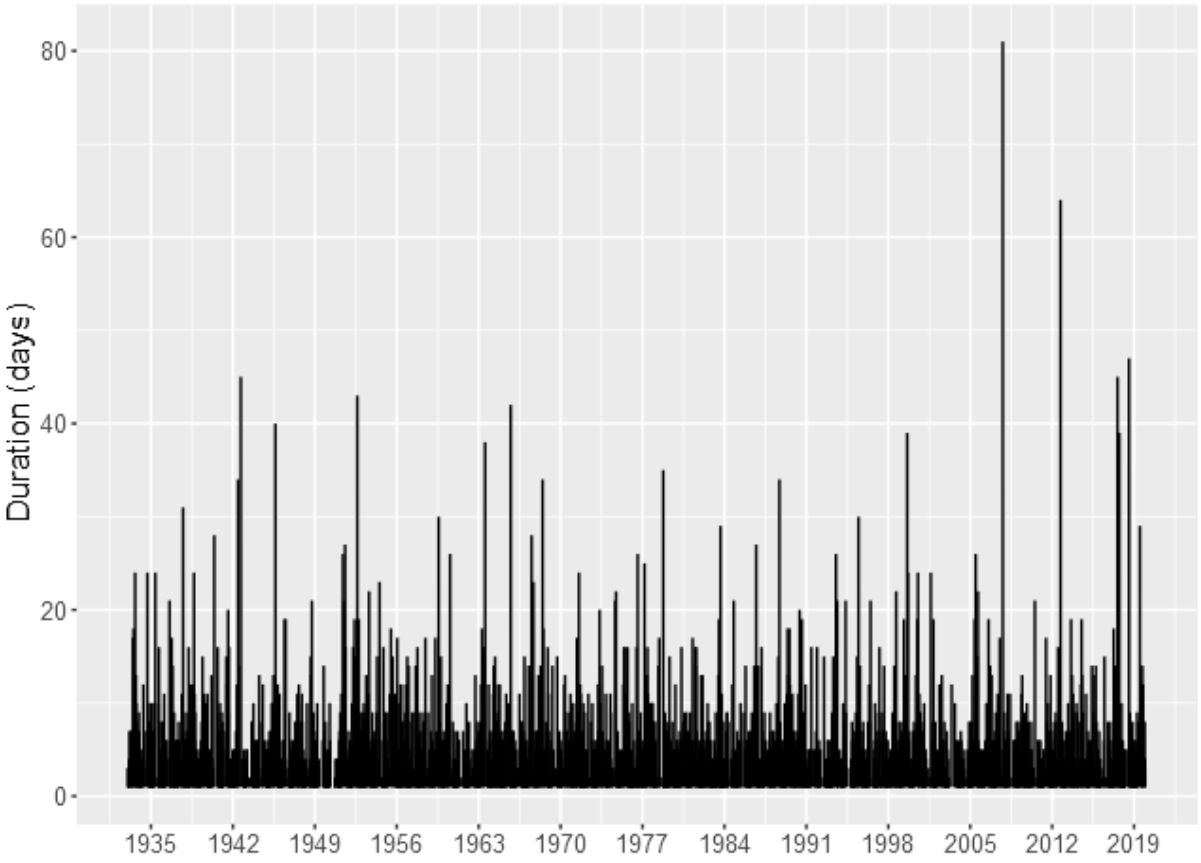
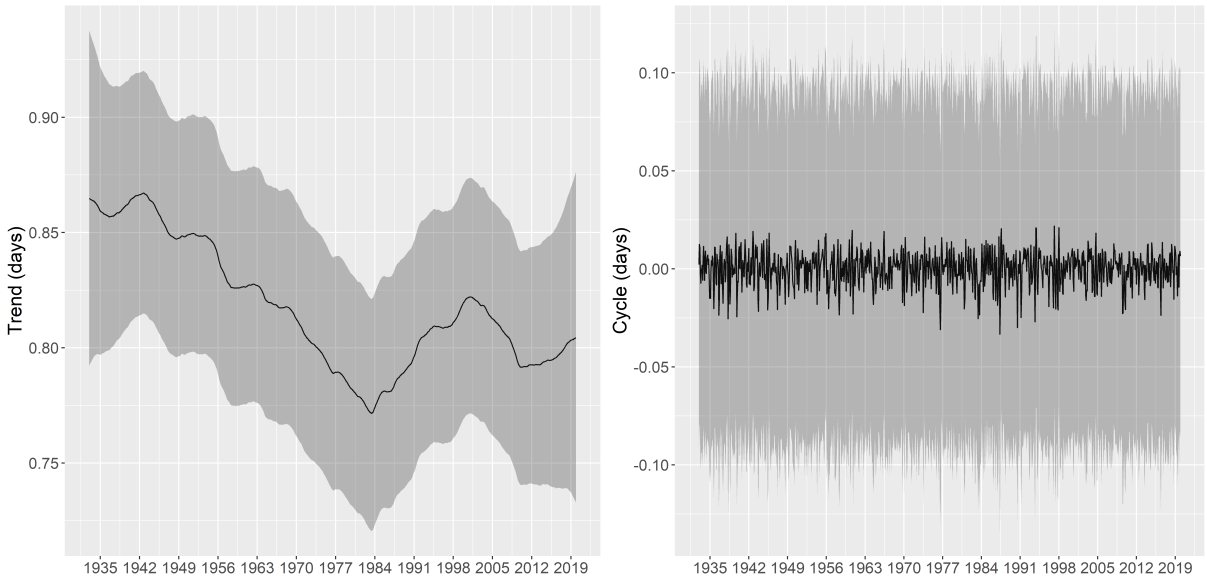
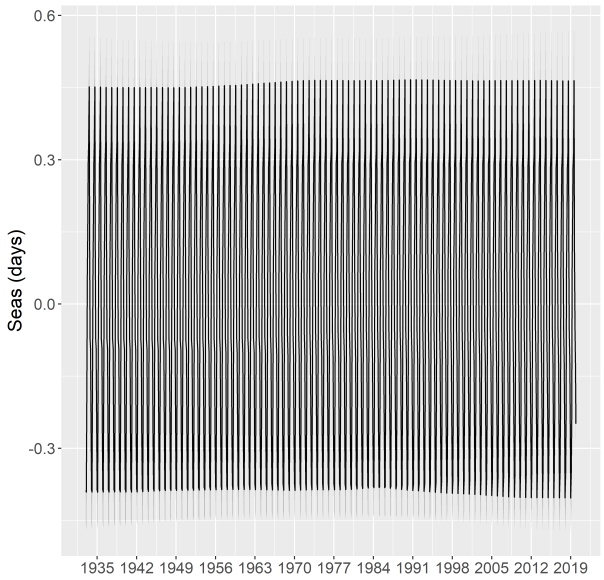


Figure 6.9 – Dry days duration - IAG Station 1933-2019. The measure is defined as the consecutive number of days without rain, defines as the days with observed precipitation less than 1mm.



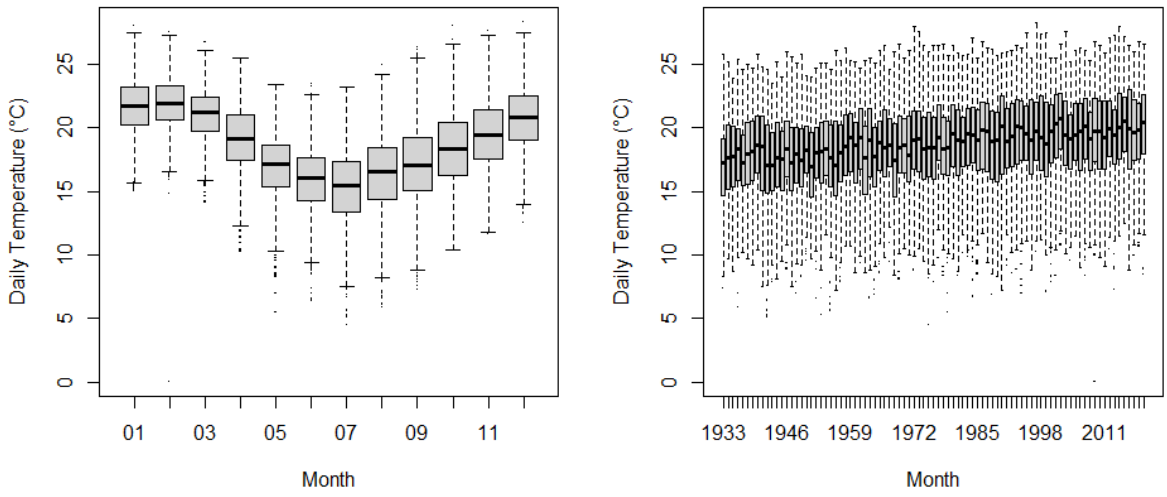
(a) Trend

(b) Cycle



(c) Seasonal

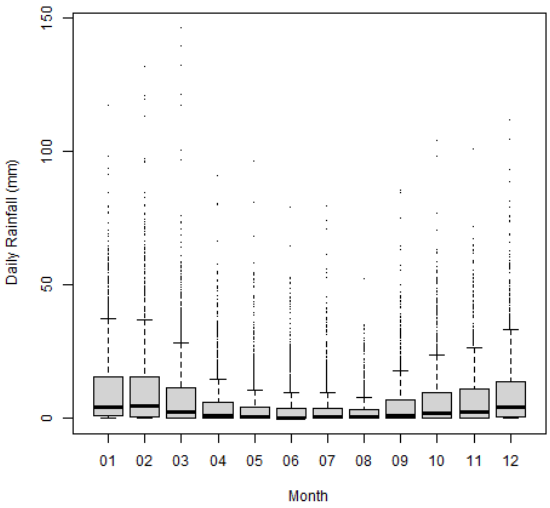
Figure 6.10 – Trend, Cycle and Seasonal decomposition for the dry days duration model. The graphs show the posterior mean (solid line) of the estimated components and the credibility intervals of 95% (shaded ribbon) for the duration model, measured in days.



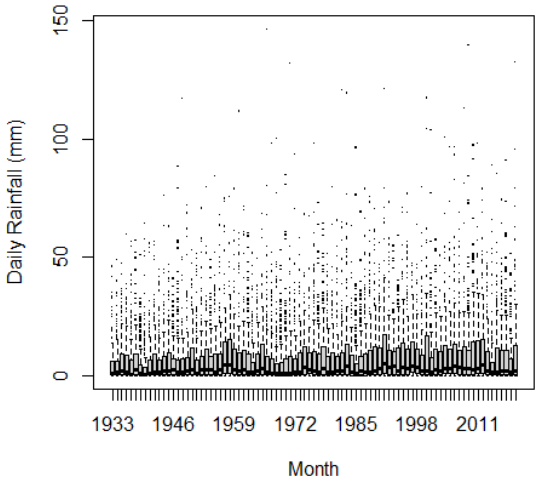
(a) Average Daily Temperature - Boxplot by month

(b) Average Daily Temperature - Boxplot by year

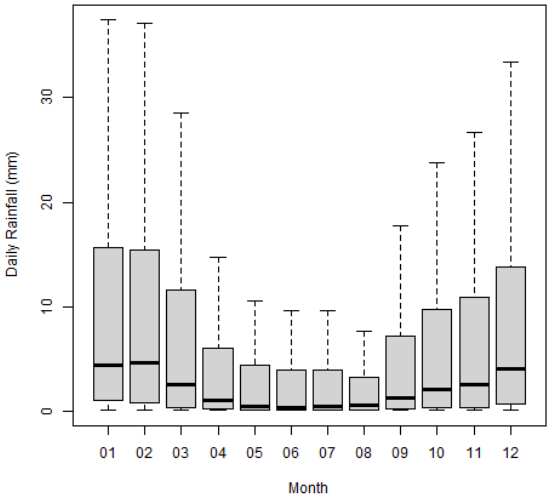
Figure 6.11 – Boxplots - Temperature (measured in degrees Celsius) by monthly and year at IAG Station - 1933-2019



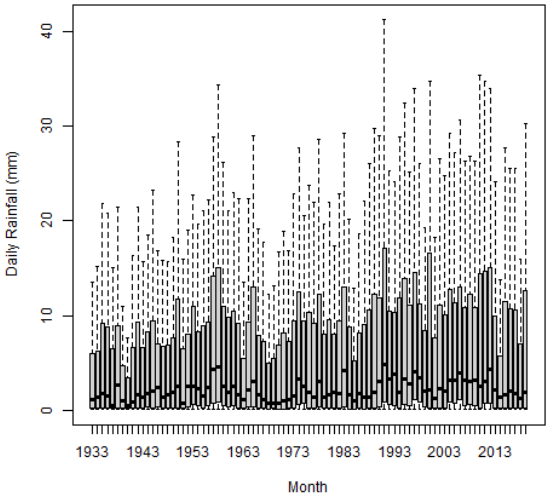
(a) Daily Rainfall - Boxplot by month



(b) Daily Rainfall - Boxplot by year



(c) Daily Rainfall - Boxplot by month without outliers



(d) Daily Rainfall - Boxplot by year without outliers

Figure 6.12 – Boxplots - Daily Rainfall (measured in millimeters) by monthly and year at IAG Station 1933-2019

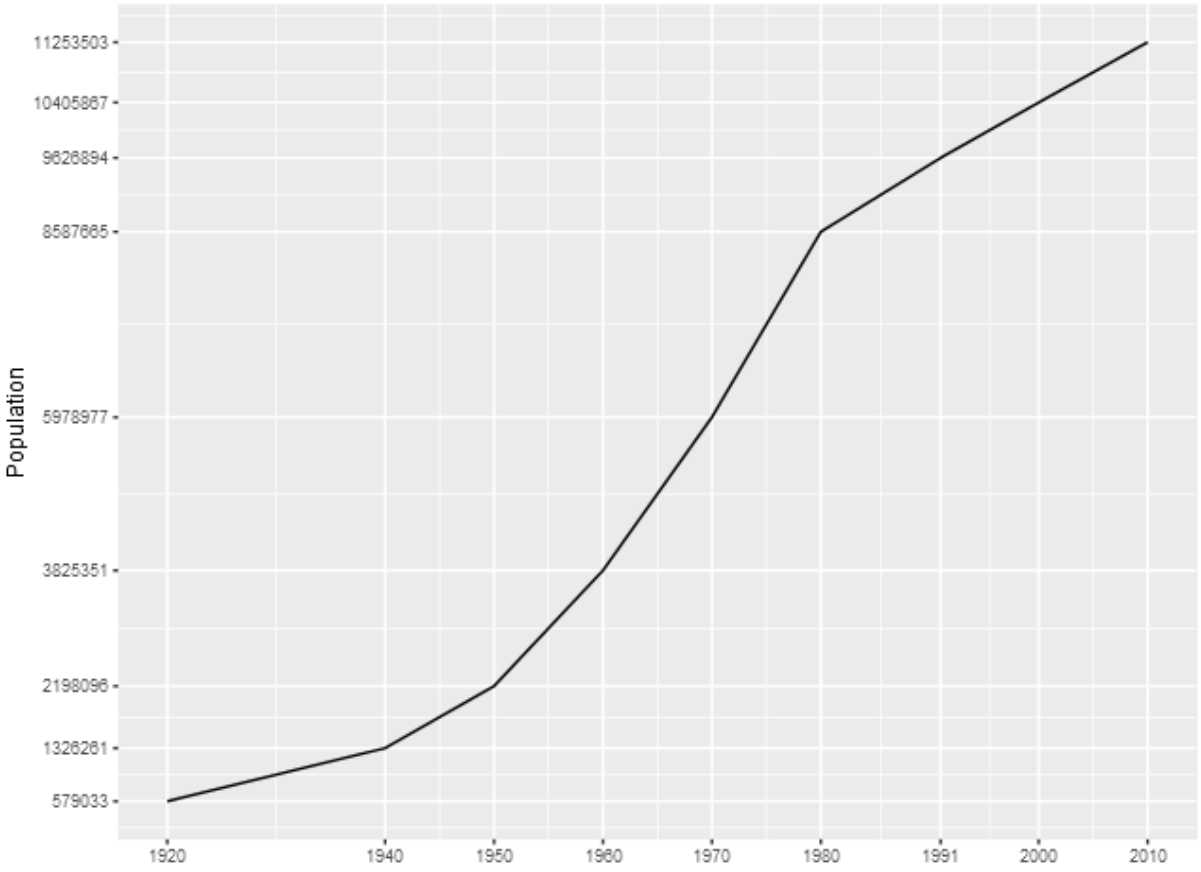
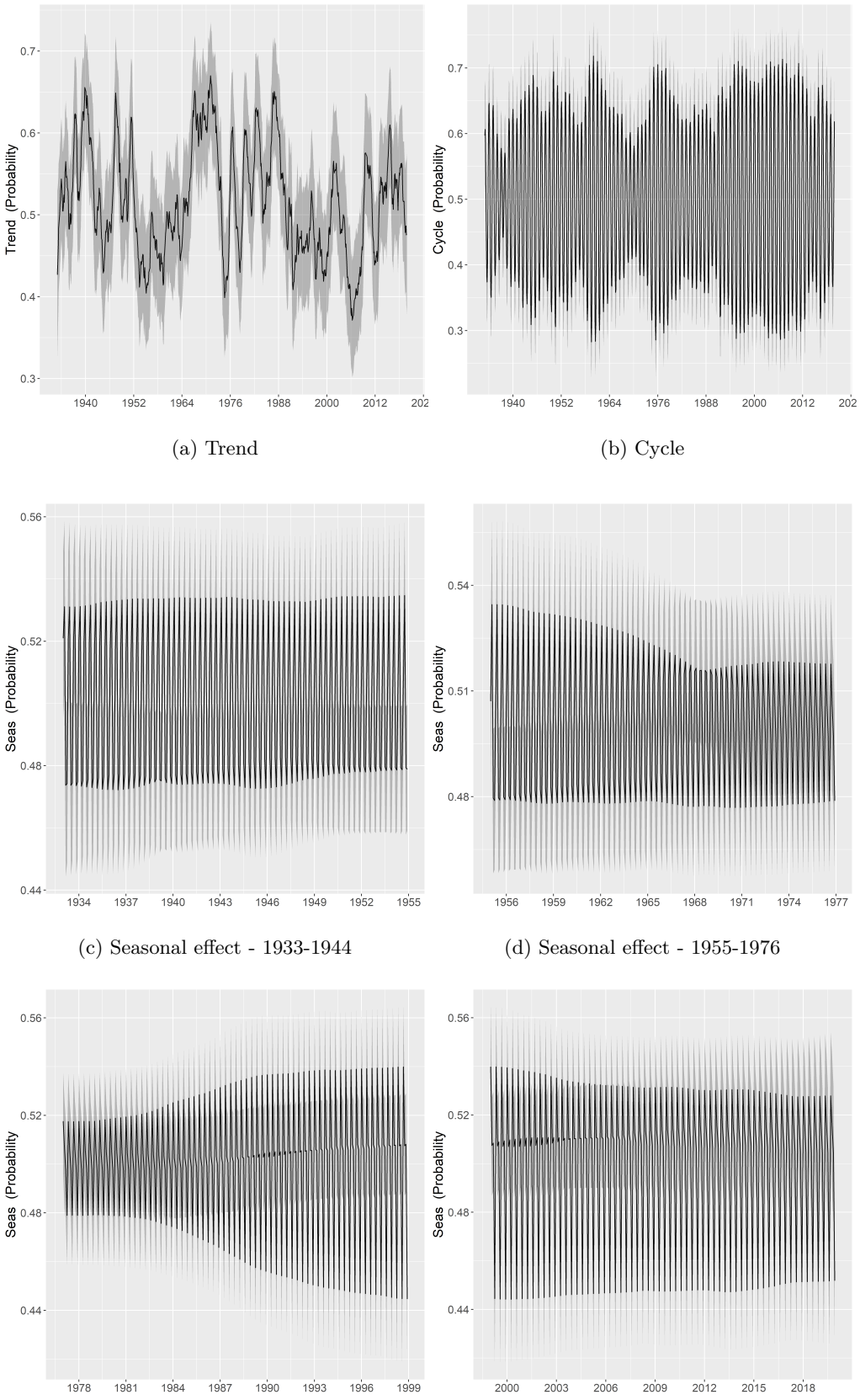


Figure 6.13 – Census population - Estimated number of inhabitants per year of the city of São Paulo - Source - *Instituto Brasileiro de Geografia e Estatistica* (IBGE)



(a) Trend

(b) Cycle

(c) Seasonal effect - 1933-1944

(d) Seasonal effect - 1955-1976

(e) Seasonal effect - 1977-1998

(f) Seasonal effect - 1999-2019

Figure 6.14 – Inverse-logistic transformation of Trend, Cycle and Seasonal decomposition for the Bernoulli-Gamma (Hurdle) model. The graphs show the posterior mean (solid line) of the estimated components and the credibility intervals of 95% (shaded ribbon) for the probability of rain in each day.

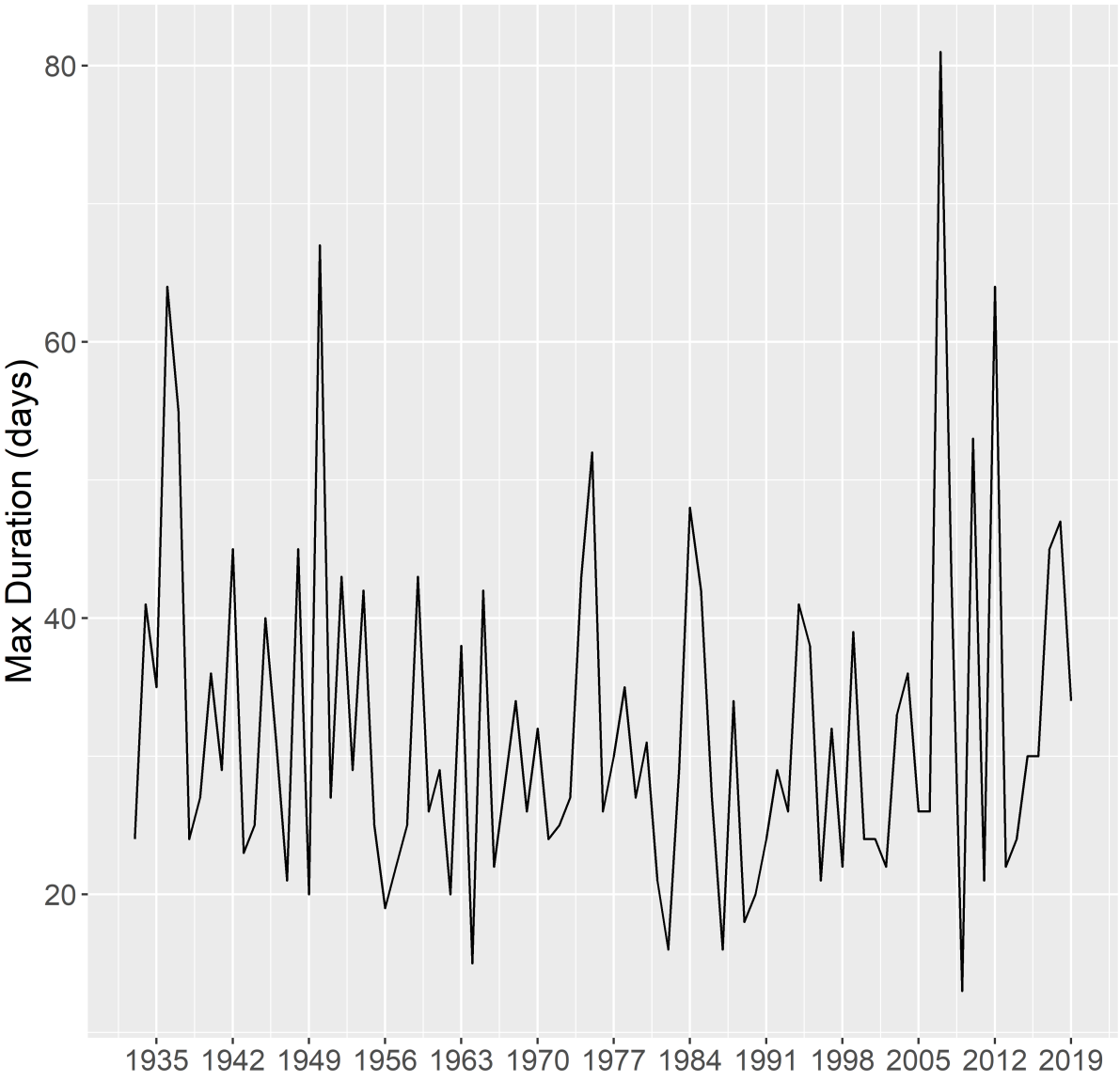


Figure 6.15 – Annual maximum number of dry days at IAG Station - 1933-2019. The graph shows the annual record for the number of dry days duration (annual block maxima). The measure is defined as the consecutive number of days without rain, defines as the days with observed precipitation less than 1mm.

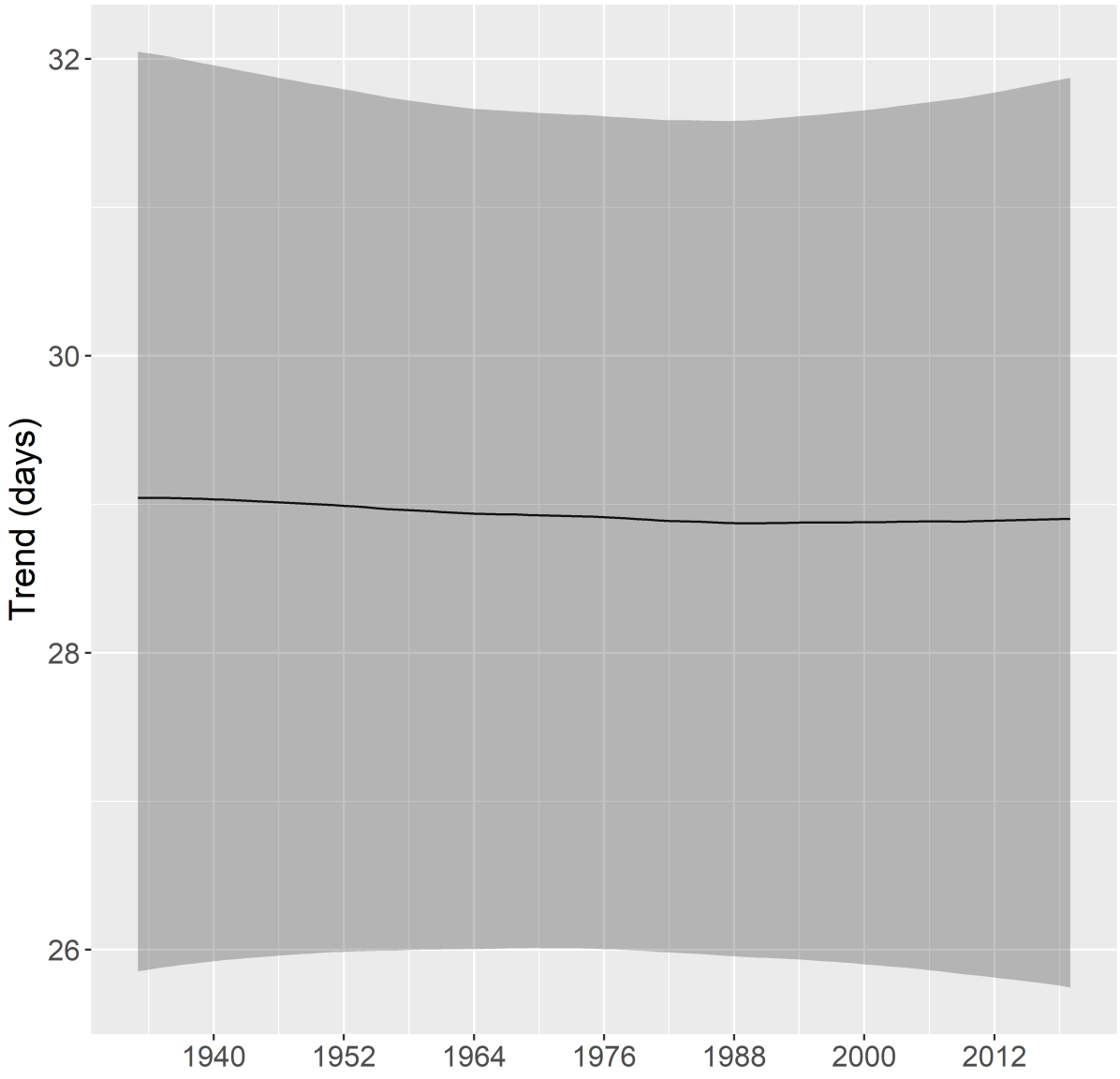


Figure 6.16 – Trend component for annual maximum of dry days duration at IAG Station 1933-2019. The graphs show the posterior mean (solid line) of the estimated component and the credibility interval of 95% (shaded ribbon). Measurements in days.

7 ROBUST TREND ESTIMATION FOR COVID-19 DATA

Fernanda Valente Márcio Laurini

ABSTRACT

Estimating patterns of occurrence of cases and deaths related to the COVID-19 pandemic is a complex problem. The incidence of cases presents a great spatial and temporal heterogeneity, and the mechanisms of accounting for occurrences adopted by health departments induce a process of measurement error that alters the dependence structure of the process. In this work we propose methods to estimate the trend in the cases of COVID-19, controlling for the presence of measurement error. This decomposition is presented in Bayesian time series and spatio-temporal models for counting processes with latent components, and compared to the empirical analysis based on moving averages. We applied time series decompositions for the total number of deaths in Brazil and for the states of São Paulo and Amazonas, and a spatio-temporal analysis for all occurrences of deaths at the state level in Brazil, using two alternative specifications with global and regional components.

Keywords: Epidemic Model; Time Series Decomposition; Spatio-Temporal Count Process

VALENTE, Fernanda; LAURINI, Márcio P. Robust trend estimation for COVID-19 in Brazil. **Spatial and Spatio-temporal Epidemiology**, v. 39, p. 100455, 2021.

7.1 Introduction

On March 11th, 2020, the World Health Organization (WHO) declared the international public health emergency caused by the novel coronavirus (SARS-CoV-2) as a global pandemic. By December 11th, 2020, there were 69,664,639 infected people and 1,583,242 deaths in 216 countries, according to Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. As the disease propagates, healthcare systems are on the verge of collapse, and despite the unprecedented global research effort, to the present date there is no effective pharmaceutical treatments available to deal with the coronavirus disease-19 (COVID-19) (DONG; HU; GAO, 2020; SANDERS et al., 2020). In order to reduce the transmission, non-pharmaceutical interventions have been proposed by many countries, e.g., social distancing, self-quarantine and lockdown. However, preventing the transmission and management decisions depends on how well we can assess the real number of infected people. World Health Organization has recommended massively testing of the population (ORGANIZATION, 2020) and thus, has caused a great demand for diagnostic test all over the world, but the limited availability and low number of applications has increasing the number of underreporting cases (PEDERSEN; MENEGHINI, 2020; VAID; CAKAN; BHANDARI, 2020; LAU et al., 2020; PRADO et al., 2020; RUSSELL et al., 2020).

In addition to the underreporting number of COVID-19 cases and deaths, it is worth noting that there are other problems related to COVID-19 data. First, a delay between the onset of

symptoms and accurate diagnosis is commonly observed, which varies from country to country, depending on the local government's strategy (CONTRERAS et al., 2020). Plus, there are also days of delay in the report of new deaths which are counted on the day that are included in the system, instead of the actual day of the death (RUSSELL et al., 2020). This creates a relevant measurement error problem, changing the series' dependency structure. The delays are also related to the lack of personnel available to report cases on weekends, which creates a seasonality structure in the cases and death reports, generating an additional aggregation problem on the time series (SKIERA et al., 2020). Plus, the delays are related to problems in the system of accounting for confirmed cases and deaths, which suffers from some instability issues. In some days the system is not available, so deaths and cases are included late.

Given the aforementioned discussion, it is possible to note that estimating the trend in COVID-19 cases and deaths it is not a trivial problem. The estimation of long-term movements is of vital importance to draw effective strategies to reduce the transmission of COVID-19. There are several different methods for trend estimation, which differ in their complexity and interpretability, where the most prevalent trend estimation methods are model-based trend extraction, nonparametric filtering, singular spectrum analysis and wavelets (ALEXANDROV et al., 2012).

Although there are many studies in the literature to estimate the trend of COVID-19 for different countries (e.g., Li, Feng and Quan (2020), Gupta and Pal (2020), Ceylan (2020), Perone (2020)), there are little understanding about the long-term movements of COVID-19 in Brazil. Nowadays the country draws a lot of attention since it has the world's second-most cases and deaths of COVID-19 (behind just the United States), and with probably substantial underestimates since the number of tested people are relatively low. For instance, up to August 12th, about 62 thousand tests per million inhabitants have been applied in Brazil, while the United States and United Kingdom has applied more than 200 thousand tests per million inhabitants. The difficulties in estimating the amount of infected people in Brazil are related to the absence of adequate laboratory infrastructure and qualified people, difficulty in buying tests due to the high international demand, and logistical distribution of tests in a country of continental dimensions such as Brazil (RIBEIRO; BERNARDES, 2020).

Notwithstanding the supposed underreporting of COVID-19 cases, the country also faces problems caused by contradictory and inaccurate data presented by official public portals. This occurs due to the polarized opposition between the federal government, and the state and municipal governments. While Brazilian government and, especially the President Jair Bolsonaro, are continually disqualify publicly both risks and the adoption of scientifically based prevention measures, most state and municipal governments have imposed social distancing along with other public health measures to control the spread of the virus (ORTEGA; ORSINI, 2020). Several Brazilian official control bodies (municipal, state and federal) have created official internet portals to report the number of cases. However, there are many discrepancies in these sources, presenting contradictory data on the impact of the disease (SILVA et al., 2020).

In a response to conflicting numbers provided by Brazilian Health Ministry, some data collection initiatives have collected the data by each municipality, providing more accurate information to COVID-19 research, in a parallel work to the federal government. The most notable work is provided by competing Brazilian mainstream media outlets, which have established a work in a

collaborative way to gather and upload necessary COVID-19 information in the 26 states and the Federal District. In addition to the daily data, the consortium also provides the so-called moving average, in order to give a better view of the evolution of confirmed cases and deaths of COVID-19 in Brazil since this is an understandable and easy to execute method to compute trend values. The moving average filter is an example of low-pass filter because it eliminates the lower or slower frequencies from the time series by means of moving average. However, this is a limited tool for some reasons. First, it cannot be used to forecasting since the trend path does not belong to any mathematical function. Second, the extent of moving average is determined ad-hoc and may carry the effect of human judgement. Also, the method assumes that the trend is always linear.

An alternative way to estimate the long-term movements is through structural decomposition (e.g., Harvey (1990)). In this sense, we propose to decompose the temporal variability observed in the time series into trend, seasonal and cycle components, which allows us to identify long-term movements, and cyclical and seasonal effects, in the presence of measurement errors. Since we are dealing with point process, we introduce a method that allows decomposition of time series in a count data framework, based on the Poisson distribution. In particular, we propose to use a Poisson process where the intensity function is decomposed into trend, seasonal and cycle components. The inference is performed following a Bayesian approach, which is able to capture the uncertainty associated with the latent factors via Bayesian credibility interval. We also present a spatio-temporal generalization of this methodology, using a formulation of Conditional Autoregressive (CAR) models with time varying spatial random effects. In particular, the resulting Bayesian hierarchical model fits within the integrated nested Laplace approximations (INLA) framework, providing an estimation in a computationally effective way (RUE; MARTINO; CHOPIN, 2009).

Regarding this context, our main goal is to estimate the patterns of the deaths by COVID-19 in Brazil through the trend-cycle decomposition. The contribution of this paper is to explore a Bayesian version of the structural decomposition in combination with count distributions, in the task of estimating the trend of deaths of COVID-19 in Brazil, and compare it with averages approach, which are not robust to the most common problems related to COVID-19 data. In particular, we performed inference procedures for deaths reported in Brazil, and also for the states of São Paulo and Amazonas. We choose to analyze these two states in different regions (São Paulo in Southeast and Amazonas in North) since Brazil is a country with continental dimensions and is characterized by a great heterogeneity in socioeconomic and cultural context among regions. For instance, in the Southeast region, the social distance and hygiene measures to reduce the transmission of COVID-19 are not feasible in subnormal agglomerates (also known as “favelas”) and peripheries, where the problem is intensified by issues in sanitation and access to water. However, it is important to highlight the difference of this problem among Brazilian regions: while about 92% of the population in Southeast are supplied with treated water, in the North, this number is just 57%. Furthermore, the North region faces a lack of heath infrastructure. As an example, the state of Amazonas concentrates all the hospital structures able to deal with COVID-19 in a single municipality, Manaus. It represents the lack of health infrastructure for the other 61 municipalities within a state whose territorial extension is larger than those of United Kingdom, Italy and France combined (FREITAS; CIDADE et al., 2020).

Additionally, we also extend our proposed model by including spatial information in the estimation process. The use of spatial information is important in estimating the trend and cycle components, due to the nature of transmission in an epidemic process, with spatial spread dynamics. Thus, the number of cases in a region has an important impact on neighboring regions, and the incorporation of this information as a prior information for the number of cases allows a more precise recovery of the trend of occurrences. We formulate two spatio-temporal models through a conditional autoregressive structure with time-varying spatial effects. Allowing spatial effects to change over time is essential in modeling an epidemic process, as this structure allows to identify the spatial pattern of spread of the infection. The spatio-temporal models differ in terms of the components of trend, seasonality and cycle. The first model assumes components common to all regions of the country, while in the second model we introduce specific trends, seasonality and cycles for each region of the country, allowing an identification of the regional patterns of the COVID-19 epidemic.

This article is organized as follow. Section 2 contains a description of the temporal and spatio-temporal models and presents the data. Section 3 shows the results with discussion. Section 4 concludes.

7.2 Data and Methods

7.2.1 Data

To perform inference procedures in all analysis we use a data set provided by BRASIL.IO (<https://brasil.io/dataset/covid19/caso_full/>), where the data collection is done by the reports of the number of cases and deaths from the official epidemiological bulletins of each municipality, by a task force of 40 volunteers. This daily data is available at municipal level, containing the temporal evolution for the number of reported deaths for each municipality and state. In this work, we analyze the sample of deaths from 02/25/2020 to 12/06/2020 for Brazil and the states of São Paulo and Amazonas.

7.2.2 Univariate Time Series Models

The model used in this work is based on a structural decomposition suitable for modelling counting series, using a Poisson distribution structure. In this distribution, the log intensity varies over time, and is given by the sum of components of trend, seasonality and cycle. As we are using a structure of Poisson processes with stochastic intensity, the characterization of the process is given by a Cox process. In the spatio-temporal analyses, we use a structure of a dynamic version of Besag-York-Mollié model (BESAG et al., 1991), where the log of the intensity function is given by a Gaussian Markov Random Field (RUE; HELD, 2005c; ILLIAN; SORBYE; RUE, 2012).

The trend component is formulated using a second order random walk structure (RUE; HELD, 2005c), while the seasonal component is composed of stochastic effects in the order of periodicity of the series, in this case since we use daily data. The cyclic component is based on a structure of a second-order autoregressive process, which aims to capture the sum of stationary effects with mean reversion, including the effects induced by measurement errors in the series

of deaths. This component is essential for the correct identification of trend and seasonality patterns, as we will discuss below. The time series model can be written as:

$$\begin{aligned} Y_t &= Poisson(exp(\lambda_t)), \\ \lambda_t &= \mu_t + s_t + c_t \\ \Delta^2\mu_t &= \eta_\mu \\ s_t &= s_{t-1} + s_{t-2} + \ldots + s_{t-m} + \eta_s \\ c_t &= \theta_1 c_{t-1} + \theta_2 c_{t-2} + \eta_c \end{aligned} \tag{7.1}$$

where Y_t is the number of occurrences in time t, μ_t is the long term trend, which can be seen as the accumulation of all shocks that occurred in the past with non-transitory effects, and is modeled as a second-order random walk (RW2). The s_t represents the seasonal components, c_t is a cycle component represented by a second-order autoregressive process with possibly complex roots,η_μ, η_c, and η_s are nonspatial independent innovations with $\eta_\mu \sim N(0, \sigma^2_{\eta_\mu})$, $\eta_c \sim N(0, \sigma^2_{\eta_c})$, and $\eta_s \sim N(0, \sigma^2_{\eta_s})$. The resulting Bayesian hierarchical structure allows us to perform inference procedure within the INLA framework, which provides accurate and efficient approximations on Bayesian hierarchical models that can be represented as latent Gaussian models. For reasons of space, we do not detail the INLA method here, but we show the fundamental aspects of the method in the Appendix.

The use of an RW2 structure for the modeling of COVID cases, replacing traditional local level processes, can be justified in several aspects. The first one is the nature of extremely fast case growth in an epidemic process, which generates a dependency structure that is best approximated by an integrated second order process (I(2)). In the stages of accelerated growth of cases, we have the presence of a trend component with a non-stationary growth rate, inducing a process that needs at least two differences to induce stationarity. The RW2 process is a parsimonious way of capturing processes with this dependency structure. A second characteristic of this process is that in these situations it also imposes a smoothness structure in the process trend, since this process can be related to formulations of smoothing splines models, as discussed for example in Green and Silverman (1994) and Lindgren and Rue (2008). A similar discussion in econometrics is in the relationship between the so-called Hodrick-Prescott filter and its formulation in state space, which corresponds to a spline model with an estimated parameter of ratio between the variances of the observation and state equations. Discussions between these properties and problems using the HP filter can be found at Harvey and Trimbur (2008) and Hamilton (2018).

7.2.3 Spatio-Temporal Analysis

The analyzes carried out so far used only local information to carry out the inference procedures on the trend and the other components. A possible extension is the use of information on cases or deaths that occurred in some definition of neighborhood in the region. This information is important since the number of occurrences in the neighborhood can be used as prior information for the estimation in the Bayesian inference procedure, which is particularly

important in the estimation of a high speed epidemic process and whose specific characteristics of transmission, latency and mortality are unknown.

In order to perform a spatio-temporal analysis of the patterns of occurrence of deaths related to COVID-19, we formulated a generalization of the decomposition of trend, seasonality and cycle components incorporating a time-varying spatial component, using a version of a model of random spatial effects with a Conditional Autoregressive (CAR) (BESAG, 1974; BESAG et al., 1991) structure with time dependence. This model allows to incorporate the existing information in the number of cases of the neighbor as prior information for the number of cases in the region of interest.

In relation to the model proposed in Equation (7.1), the spatio-temporal version adds two modifications. The first is the use of an offset E_i to control the Exposure, which in this case is the number of inhabitants of region i. This component allows to use common components of trend, seasonality and cycle for regions with different population sizes. The second modification is the $\xi_{(i,t)}$ spatial component, which adds an additional random effect for each region i , allowing to incorporate spatial variability in the occurrence rate. The CAR structure used takes as prior for the spatial effect in the region i a Normal distribution with mean given by the average of the values of the spatial effects for regions j that are neighborhood to region i, and variance controlled by a τ precision parameter multiplied by number of neighbors. We assume that this pattern varies over time, reflecting the dynamics in the spatial distribution of the number of occurrences of COVID-19. The dynamic formulation is built through an autoregressive structure in time for this component, an additional parameter Φ controlling the time dependence for this process. The model structure can be represented as:

$$
\begin{aligned}
Y_{(i,t)} &= Poisson(exp(\lambda_{(i,t)}E), \\
\lambda_{(i,t)} &= \mu_t + s_t + c_t + \xi_{(i,t)} \\
\Delta^2\mu_t &= \eta_\mu \\
s_t &= s_{t-1} + s_{t-2} + \ldots + s_{t-m} + \eta_s \\
c_t &= \theta_1 c_{t-1} + \theta_2 c_{t-2} + \eta_c \\
\xi_{(i,t)} \mid \xi_{(j,t),i\neq j,\tau} &\sim \mathcal{N}\left(\frac{1}{n_i}\sum_{i\sim j}\xi_{(j,t)}, \frac{1}{\tau n_i}\right) \\
\xi_{(j,t)} &= \Phi\xi_{(j,t-1)}
\end{aligned}
\tag{7.2}
$$

To estimate the above spatio-temporal model, we used COVID-19 data at the state level as a spatial unit definition. Although it is possible to work with data at the municipal level, we note that there are important problems with the detailing of cases at this level. The data are only reported in detail for the largest municipalities in each state, with data for small municipalities being aggregated in terms of regions or else as a residue in the total sum of cases within each state. Thus, the state aggregation offers greater reliability in the analysis of spatial data patterns. We used the neighborhood definition as a neighborhood structure based on queen-type contiguity using the territorial division of Brazilian states. As in the first model, the spatio-temporal model is also estimated using Integrated Nested Laplace Approximations,

where the prior structure for the time series components is the same used to estimate univariate models, and we also adopted a prior log-gamma with values (1.5e-5) for the precision of this component. The prior for the coefficient Φ is a penalized complexity prior (SIMPSON et al., 2017) for the correlation parameter, with values (3,0.01).

7.2.4 Monte Carlo study for SEIR model

The seasonality and cycle components are introduced in order to control the effects related to the case accounting mechanism, which, as discussed in the introduction, introduces a structure of periodicity in the data, generating a seasonal pattern. It is also important to note that the dependency pattern generated by the epidemic process and this form of measurement error generates a complex contamination structure in the series of new cases or deaths, which reduces the effectiveness of simple methods of eliminating seasonality or estimating trends, in particular the use of simple methods such as a 7-day moving average in cases, which is the form used to summarize the trend of COVID-19 cases in Brazil, provided by the Brazilian mainstream media outlets.

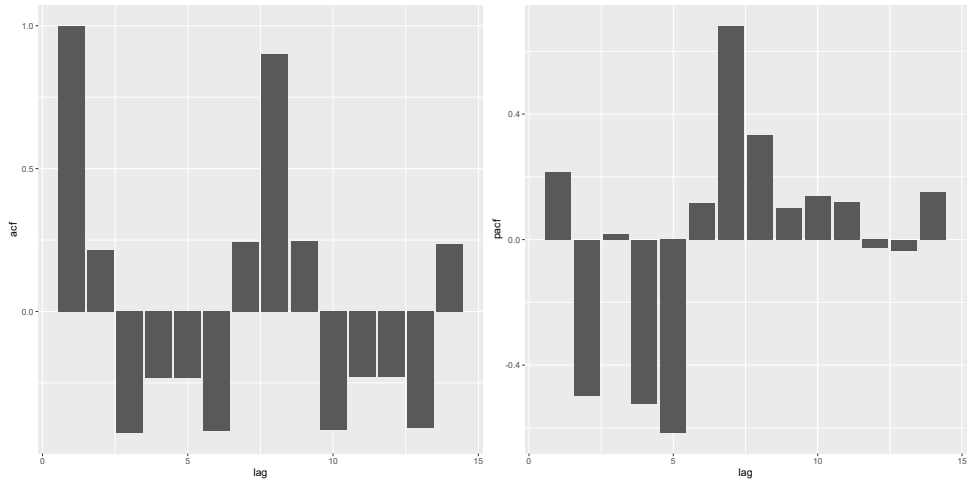
To show the complexity of this contamination pattern and the limitation of simple trend extraction methods, we will resort to simulations of an epidemic process with a contamination structure similar to that existing for COVID-19 data, as well as some applications of simple methods of trend extraction. We performed a Monte Carlo experiment with 10,000 replications, simulating a deterministic Susceptible-Exposed-Infectious-Recovered (SEIR) epidemic process, with a stochastic structure for the number of new deaths resulting from this model. The general details of this model are presented in Appendix.

The key idea is to obtain the solution of a SEIR process, which generates trajectories of the number of Susceptible, Exposed, Infectious, Recovered and Deaths over time. In our simulation, we replicated a mechanism for reporting new death cases similar to that used by health departments in Brazil. In particular, we define that each observation corresponds to a daily data in a seven-day week, and if deaths occur on Saturday or Sunday, there is a proportion of deaths that will not be counted on the same day, but only on the following Monday and Tuesday. Also, we consider that all deaths are effectively accounted for, and set that the proportion of deaths with delayed disclosure is given by a uniform distribution, with parameters (.5, .8). Indeed, this structure is a very simplified and unrealistic version of the reality due to underreporting of deaths, but it is useful to exemplify the effects of measurement error on the dynamic structure of the time series of new deaths.

As a summary of the contamination effect, we show in Figure 7.1 the average values of the autocorrelation and partial autocorrelation functions for the time series of new daily deaths. We can see that the contamination mechanism induces a complex structure of temporal dependence in the series of new deaths in the SEIR process. Although there is a seasonal pattern induced by contamination, the dependency structure is more complex than a pure seasonal pattern, even in this simple contamination experiment.

The structure used in our model uses a seasonal component to control for the periodic pattern induced in the series, but also a second-order autoregressive component, to capture the remaining dynamic effects of the measurement error structure in the series. The second-order

component is a parsimonious way of capturing dependency patterns and cycles with reversion to the mean, and in our analyzes it has been shown to be effective in allowing the recovery of the trend component in the presence of seasonality and other dynamics in the series.



(a) ACF (b) PACF

Figure 7.1 – Simulated ACF and PACF functions of new daily deaths under the reporting contamination mechanism. Mean values from 10,000 replications of the SEIR model with measurement error.

Note that in the presence of these contamination mechanisms, simple seasonal filtering methods, such as the use of 7-day moving averages, will not be sufficient to extract the trend component in the series. To illustrate this problem, we will use an example with real data, in particular, the same dataset provided by BRASIL.IO, which will be also used in this work for the temporal and spatio-temporal analyzes.

We used as an example the extraction of a trend measure for COVID-related deaths for the state of São Paulo. In this example, we compare the trend obtained by a simple 7-day moving average, a moving average with the number of days determined using the modified Akaike criterion with correction for small samples (AICc) (see Svetunkov and Petropoulos (2018) for details in this procedure), and also the trend extraction using the trend filter proposed by Hamilton (2018), assuming a periodicity of 7 days. We also used a version of the Hamilton filter based on a generalized linear model assuming that the data follow a Poisson process.

Figure 7.2 shows the time series of COVID-related deaths and the trends extracted by aforementioned methods. First, an important result is that the optimal criterion for the number of lags in the moving averages procedure selected by the AICc criterion points out to 8 lags, confirming that the dependence induced by the death accounting process generates a more complex dependency structure than a seasonality of 7 days. Additionally, another interesting result is that the Hamilton filter, a trend extraction tool based on regressions against past lags and with optimally properties in general contexts, cannot effectively separate the trend component from the seasonal pattern, probably due to the stochastic and non-stationary component of the seasonality component induced by the measurement error process.

These results show that simple trend extraction methods, such as the use of simple moving averages, may not be adequate in the presence of measurement error structures such as the

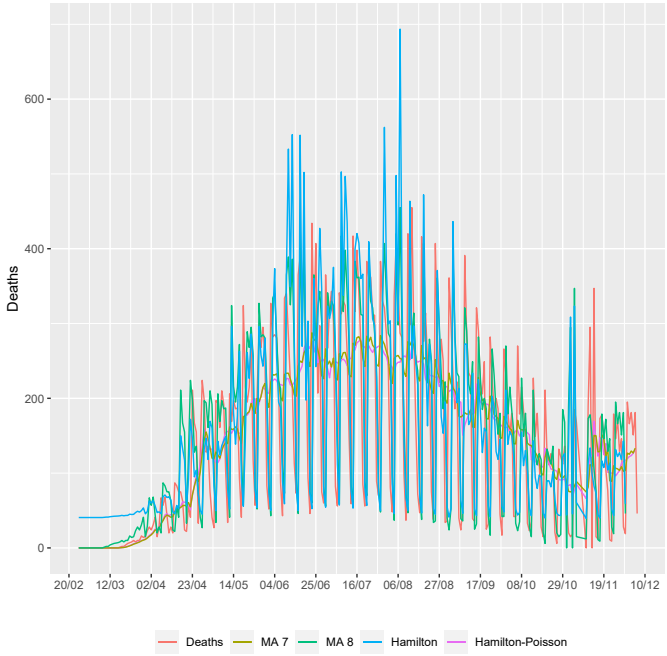


Figure 7.2 – Trend extraction using Moving Average and Hamilton Filters - COVID-19 related deaths for São Paulo State - 02/25/2020 to 12/06/2020.

COVID-19 data present here. In the following sections we show the results obtained with the component extraction methods proposed in this work, using a formulation of time series applied to univariate data (Section 7.3.1), and also a spatio-temporal version for data at the state level in Brazil (Section 7.3.3). In this work we focus on the analysis of data on deaths related to COVID-19 in Brazil, since these data are less affected by the underreporting problem. Data from confirmed cases are difficult to analyze and especially to compare among regions, since testing procedures and recommendations are quite heterogeneous, reducing the reliability of the data and the resulting analyzes.

7.3 Results

In this section we report and discuss, first, the results based on the univariate analysis, and the comparison between the estimated trend component and moving averages approach. After, we present the results based on the spatio-temporal model. Our objective is to show what are the advantages and limitations of each model used in the work, and how it can be useful to understand the dynamics of evolution of deaths by COVID-19. The pure time series model has the advantage of only depending on the data from the region itself, but it does not capture the spatial dependence of the process. The two versions of the spatio-temporal models, with global and regional components, illustrate the gains and limitations of the use of global and regional components in spatio-temporal modeling.

7.3.1 Univariate Time Series Models

We performed inference procedures based on the model described in the Section 7.2, thus, in this case, the estimated parameters for each location are the precision of the trend compo-

nent ($1/\eta_\mu$), seasonal component ($1/\eta_s$), and cycle component ($1/\eta_c$), and the parameters of the second-order autoregressive process of the cycle, parameterized as first and second order partial correlation coefficients (PACF1 and PACF2). The precision parameters represent the variability associated with the trend, seasonal and cycle components, where high values indicate low variability, and the parameters of the second-order autoregressive process of the cycle are related to the autoregressive parameters in the AR(2) representation of the cycle.

We use a structure of log-gamma priors for the precision components, with values (1,5e-05) for the trend and seasonal components. For the AR(2) component we use a penalized complexity prior for the precision, with values (3,0.01), and for the first and second order autocorrelations parameters penalized complexity priors with values (0.5, 0.5) and (0.5, 0.4). The motivations for the use of penalized complexity priors is discussed in Simpson et al. (2017). These priors are invariant to reparameterizations and have excellent robustness properties, and in the assumed values they are not informative, which is adequate for a situation of a new epidemic with unknown previous behavior. The tables with the estimated parameters for Brasil, São Paulo and Amazonas states are presented in the Appendix of the article (respectively, Tables 7.3, 7.4 and 7.5).

The dissemination of the COVID-19 among Brazilian states vary due to the heterogeneity in socioeconomic and cultural context among regions. In order to assess the strong regional differences in the spread of infections, we performed inference procedures for the states of São Paulo, localized in the Southeast region of Brazil, and for the state of Amazonas, in the North region. The Southeast region includes the three largest metropolitan areas in Brazil, São Paulo, Rio de Janeiro, and Belo Horizonte. Also, it is the main industrialized area and concentrates the biggest population of the five Brazilian regions, with a high population density. It is not surprising that the Southeast region presents the highest number of cases and deaths by COVID-19 in Brazil, and were the first region to diagnoses COVID-19 cases and to step up social distancing policies to slow down the spread of coronavirus. On the other hand, the North region is the largest region of Brazil, and concentrates a large number of indigenous people and their descendants, who are part of the COVID-19 risk group, which makes the region particularly sensitive to coronavirus disease (FERRANTE; FEARNSIDE, 2020).

A primary motivation for the present study was to estimate the long-term movements for COVID-19 data in Brazil. Thus, to better understanding and discuss the results, we plotted the estimated trend, seasonal and cycle components for death cases registered in Brazil (posterior mean and 95% Bayesian credibility interval; see Figure 7.3). The first case of COVID-19 in Brazil was diagnosed on February 26th, in the state of São Paulo. On March 17th, the first death in Brazil was registered in the state of Rio de Janeiro, and on March 20th the community transmission of the disease was announced by the Brazilian government. Thus, the most important result is related to the trend component, where it can be seen that since registering the first case of death by COVID-19 in Brazil, it took a steep movement upwards until the first week of April, where the death trend slowly reveals a turning point, i.e., the pace of the trend has slowed down, reaching the peak in the last week of April.

Despite the evidence of the effectiveness of the measures implemented by many countries to slow down the transmission of the COVID-19 (KRAEMER et al., 2020; GATTO et al., 2020; SAEZ et al., 2020), the persistence in a peak plateau stage varies across countries, depending on

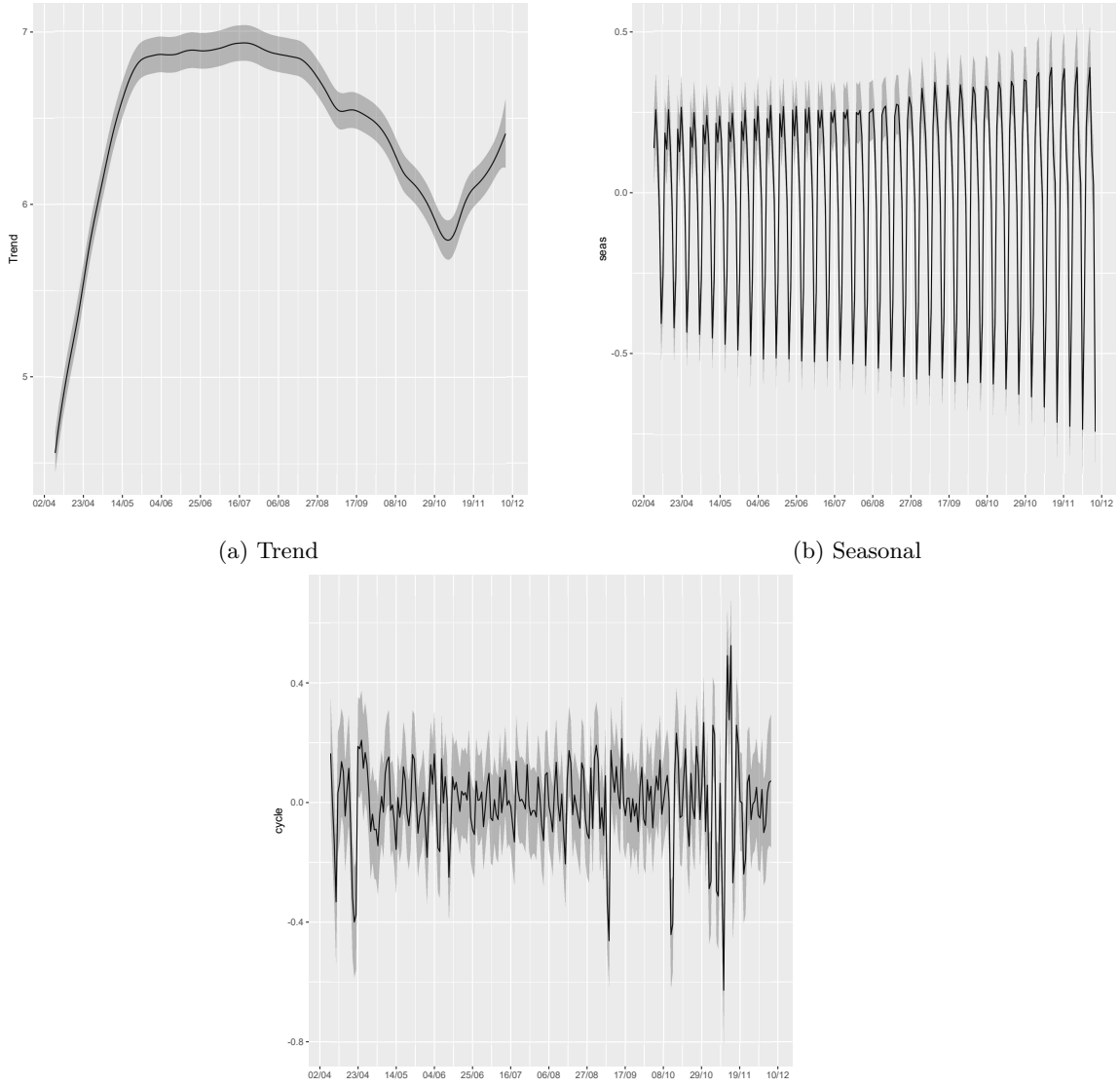
how quickly the policies were implemented and another factors such as population density, and health systems structure (HSIANG et al., 2020; DEB et al., 2020). In particular, the persistence in the high plateau stage can be explained by the containment measures introduced late, during the evolution of the new cases and deaths, which requires a long time to revert the death trend into a decreasing one.

In the case of Brazil, the trend component remained relatively stable in a high peak plateau from May to August, with a slight decrease just around mid-August. In particular, the sustained high plateau observed in the estimated trend component of deaths by COVID-19 in Brazil may be related to the absence of political actions at federal level and the early relaxation of the isolation measures (CANABARRO et al., 2020). In addition, Figure 7.4 shows the observed number of death cases by COVID-19 and the growth rate for Brazil, where it is possible to see that the days with the highest rates occurred until mid-May.

A more accelerated downward trend is observed from the end of August, with a significant reduction in the trend until the beginning of November. From this point on, we observed a further acceleration of the trend, going to the end of the analyzed sample. This acceleration was interpreted as a reflection of the accelerated reduction in social distance measures taken in various locations, and may also be related to agglomerations caused by the electoral campaign for the election of mayors and councilors in this period. This acceleration is being called a "second wave", although there was in fact no control over the pandemic in the period before this acceleration in the number of deaths. The estimated parameters for Brazil are shown in Table 7.3 in Appendix, where it is possible to note a high precision associated with the trend, seasonality and cycle components.

Figure 7.5 presents the posterior mean of the estimated trend, seasonality and cycle components, and the associated 95% Bayesian credibility interval for deaths by COVID-19 in the state of São Paulo, and Table 7.4, in the Appendix, reports the estimated parameters. The first case of COVID-19 in the state of São Paulo was on February 26th and the first death was registered on March 18th. The state government imposed the first isolation measures between March 17th and March 22th (Decree 64.881), which established the mandatory closure of non-emergency services, as well as educational institutions. The estimated trend of death (see Figure 7.5) presents a rapidly increase from the beginning of the sample until the end of March, where a turning point is slowly revealed. This change in the rising trend can be attributed to the first isolation measures adopted by the state of São Paulo authorities, as previously discussed in the literature (e.g., Cruz (2020)). However, the abrupt change in the evolution of the death trend observed in the mid-April suggests that the level of adherence of the citizens to isolation measures dropped down over time, or the measures taken by state government were not enough to cope with the outbreak. After reaching the peak on mid-May, the death trend remained in a long persistence peak plateau stage. Also, Figure 7.6 shows that the highest growth rate of deaths occurred around mid-April, whereas the lowest growth rate was registered around mid-August, suggesting a decreasing trend from mid-August. Similar to the pattern observed in the data from Brazil, in the state of São Paulo we observed a significant drop in the trend from June onwards, and a strong acceleration in the cases from November.

The estimated trend, seasonality and cycle components, and the associated 95% Bayesian credibility interval for the state of Amazonas are presented in Figure 7.7, and Table 7.5, in



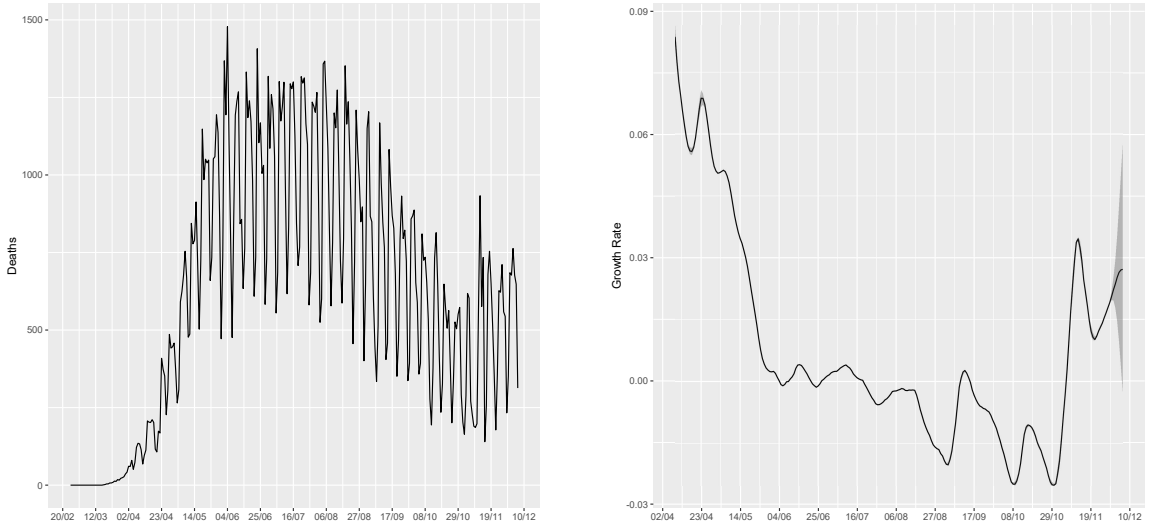
(a) Trend

(b) Seasonal

(c) Cycle

Figure 7.3 – Posterior mean and 95% credibility intervals of Trend, Seasonal and Cycle decomposition of deaths in Brazil - 02/25/2020 to 12/06/2020.

the Appendix, reports the estimated parameters. In the state of Amazonas, the first death by COVID-19 was recorded on March 24th, on same day that isolation measures were imposed by local authorities. Despite the North region being the last Brazilian region to register the first case of COVID-19, the state of Amazonas was one of the first to collapse the health system. Indeed, through the Figure 7.7 it is possible to see that the death trend in the state of Amazonas reached the peak faster than the state of São Paulo, where the social measures were able at least in part to flatten the epidemic curve and postpone the peak of death cases. This result may be explained by some reasons. First, the state capital, Manaus, concentrates all the intensive care units able to deal with the coronavirus disease in the state. With few roads and transport mainly by rivers, many patients were not able to access proper health care, especially indigenous people, who are particularly sensitive to COVID-19 and usually live in isolated areas, far from doctors and access to medicines. In addition, the state of Amazonas presents a high social vulnerability index, and one of the worst rates of hospital beds per population over country



(a) Deaths (b) Growth Rate

Figure 7.4 – Number of reported deaths and the estimated growth rate of the Trend in Brazil - 02/25/2020 to 12/06/2020.

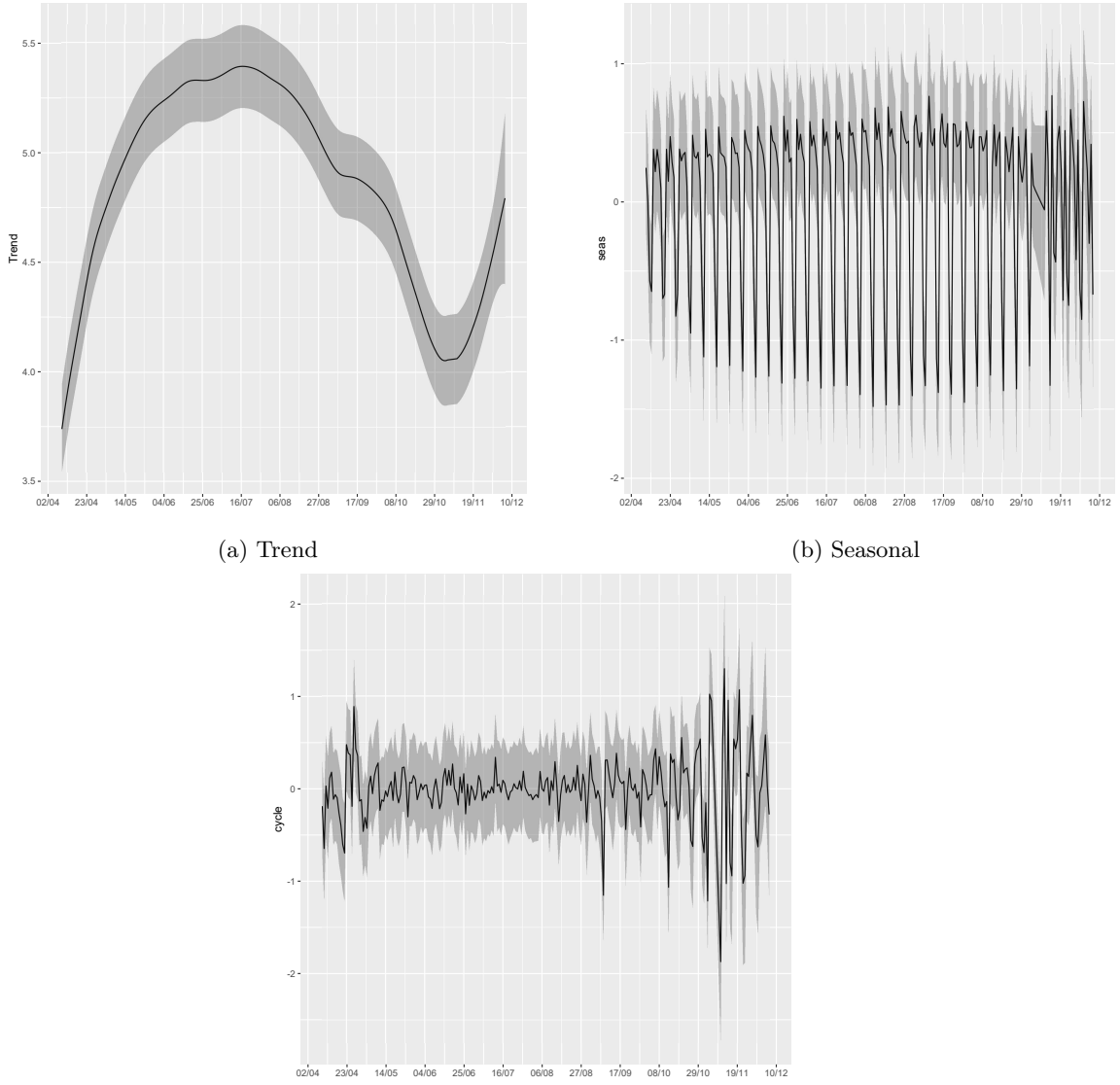
(GUERRA-SHINOHARA et al., 2020). Since the beginning of June, the state capital, and most populous city in the state, established some relaxation strategies in the containment measures, including reopening schools, which could have led to the observed increasing pattern in the estimated trend component in the end of July (FERRANTE et al., 2020). A second local peak is observed in August, followed by a reduction in the trend until mid-October, and thereafter a further acceleration that persists until the end of October, and after that period a further reduction in the trend, showing a behavior different from the pattern observed for Brazil.

The state of Amazonas is particularly interesting as it shows the difficulty in controlling the covid epidemic. According to the estimates presented in Buss et al. (2020), about 76% of the population of Manaus, the capital of Amazonas, had already been infected until October 2020, and even so the epidemic had not been fully controlled, showing that the proportion number of infected people required the existence of a herd immunity is quite high. The general trend of declines observed since November may indicate the start of herd immunity, since the trend contrasts with the acceleration pattern in the rest of Brazil. In addition, in Figure 7.8, it is possible to observe the great heterogeneity in the rates of infection growth in this state, showing the complexity in the transmission patterns of COVID-19.

In order to show the model's ability to fit the death cases, we plotted the observed number of deaths by COVID-19 and the predicted value of death count in each day given by the sum of the estimated trend, seasonal, and cycle components, along with the 95% credibility interval of this sum (see Figure 7.9), for Brazil and the states of São Paulo and Amazonas. The results visually suggest that the model has a good fit.

7.3.2 Comparison between moving averages approach and the proposed model

We compared the performance of our proposed model in monitoring the long-term movements of mortality of COVID-19 with the so-called moving average filter. The importance of the results interpretation is evident, and this is why we need to properly highlight an important point about



(a) Trend

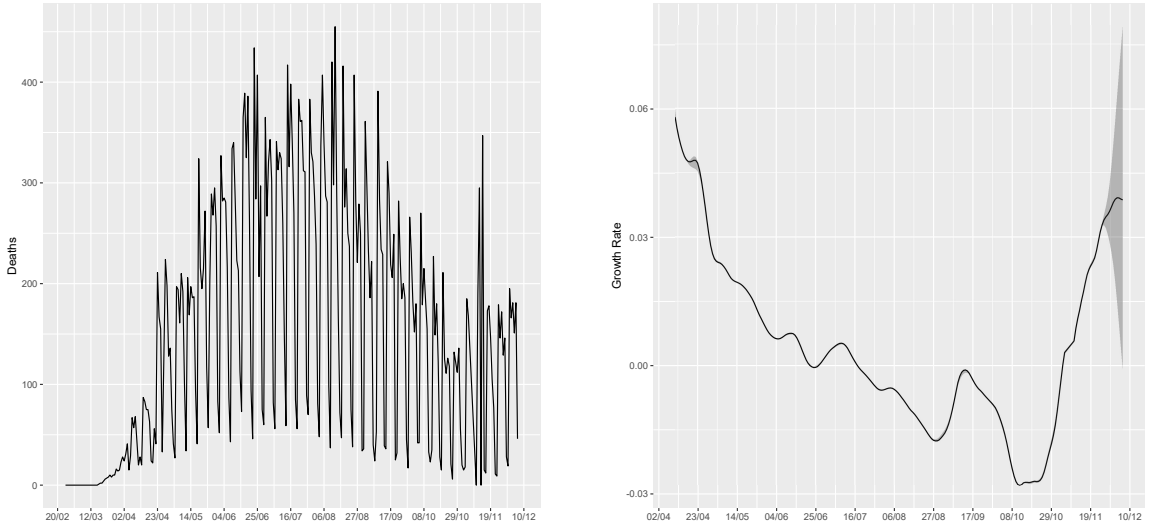
(b) Seasonal

(c) Cycle

Figure 7.5 – Posterior mean and 95% credibility intervals of Trend, Seasonal and Cycle decomposition of deaths in the state of São Paulo - 02/25/2020 to 12/06/2020.

the estimated values for the trend component. It is worth noting that the moving average filter is a way to extract a trend from time series count data. On the other hand, in our model the trend component can be interpreted as the mean of the log-transformed count data or the mean of the log intensity. Due to the non-linearity of our model, the anti-log (exponential) transformation is not suitable to obtain the original data. Therefore, in order to provide comparable results, we plotted (see Figure 7.10) the moving average of the log-transformed data (green line) along with the estimated trend component (red line; posterior mean and 95% Bayesian credibility interval) and the observed log-transformed count data (blue line).

As discussed above, one of the most advantages of our proposed model is the ability to extract a smoother trend component with reduced residual autocorrelations due to the inclusion of the seasonal and cycle components. Based on the comparison between the moving average filter and the estimated trend component from our proposed model, it is possible to see why is important to add the seasonality and cycle components in estimating a trend that really reflects



(a) Deaths (b) Growth Rate

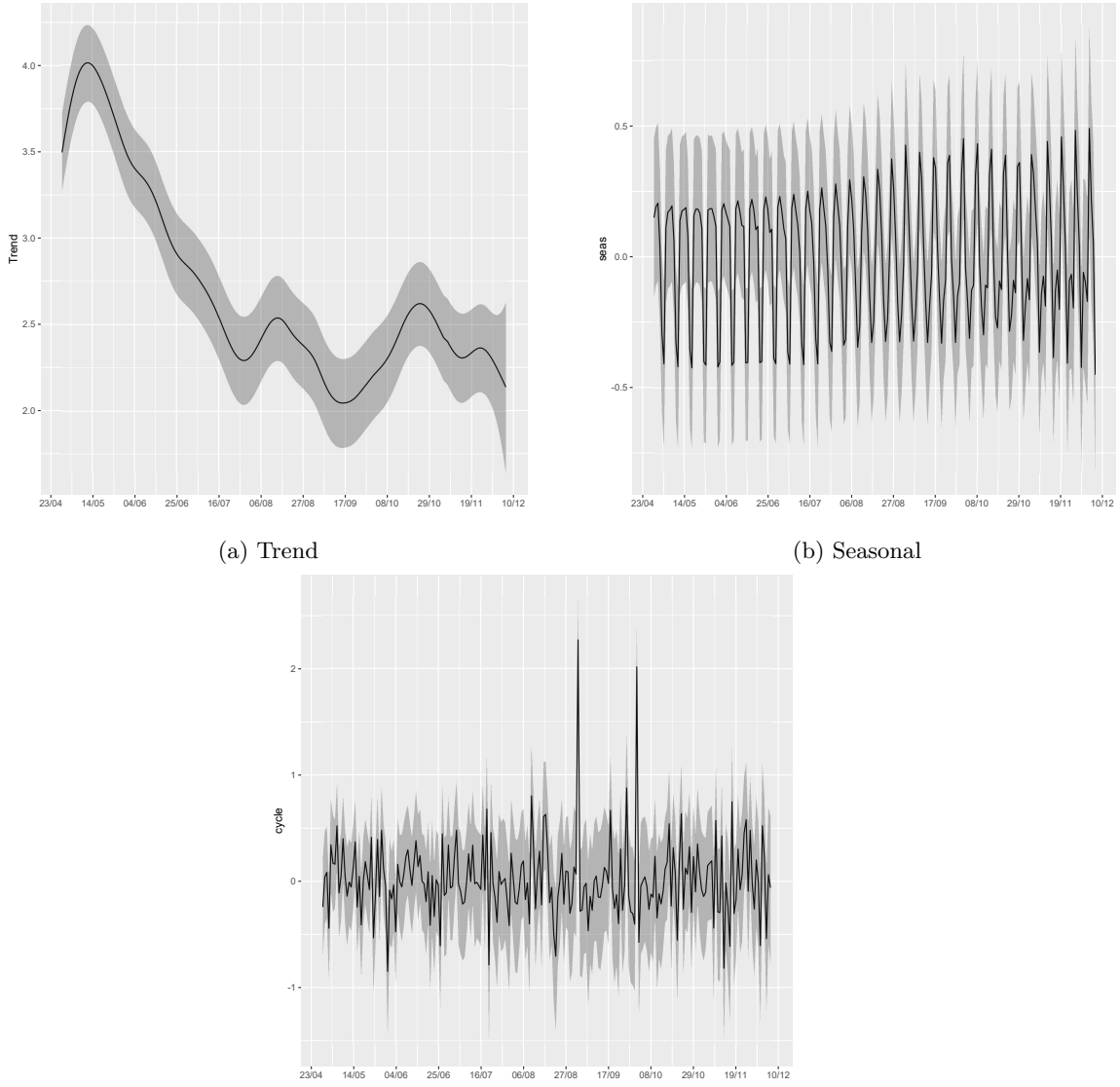
Figure 7.6 – Number of reported deaths and the estimated growth rate of the Trend in the state of São Paulo - 02/25/2020 to 12/06/2020.

the permanent patterns in COVID-related deaths. The cycle component, which corresponds to a second-order autogressive structure, is a parsimonious way of incorporating the transient effects induced by the contamination mechanism generated by the case accounting structure. Also, it is useful to capture local, non-persistent patterns in the data, isolating long-term effects. These two components allow a more robust estimation of the trend component, avoiding the greatest fluctuation observed in the trend estimation given by the moving average.

Additionally, it is important to note that the cycle and seasonal components play important roles by adding additional robustness to days with large atypical patterns in the reported data, which normally correspond to the date when adjustments are made for unreported deaths in past periods. For example, looking at Figure 7.8 containing deaths for the state of Amazonas, we can see two dates with aberrant numbers of deaths (09/02/2020 with 158 deaths and 01/10/2020 with 117 deaths). These dates correspond to adjustments where all unreported deaths in the previous months were added, and thus are observations from much earlier periods that do not correspond to current death patterns. The model captures these atypical days through the cycle component, without changing the trend estimation, as opposed to what happens with the moving average estimator, which is quite sensitive to this form of contamination, as seen in subfigure (c) of Figure 7.10. This highlights the robustness properties of the model proposed in this work to recover the trend in death patterns related to COVID-19.

7.3.3 Spatio-Temporal Analysis

For the spatio-temporal analysis we used the same data source and the same sample period as Section 7.3.1, using a state level aggregation of the death data related to COVID-19. The results of the estimation of this model are shown in Table 7.6 in the Appendix, and the estimated trend, seasonality and cycle components are shown in Figure 7.11. Note that the components are now estimated estimated with an adjustment for the size of the population in each region, which corresponds to exposure in the Poisson process. To recover the specific effect of the



(a) Trend

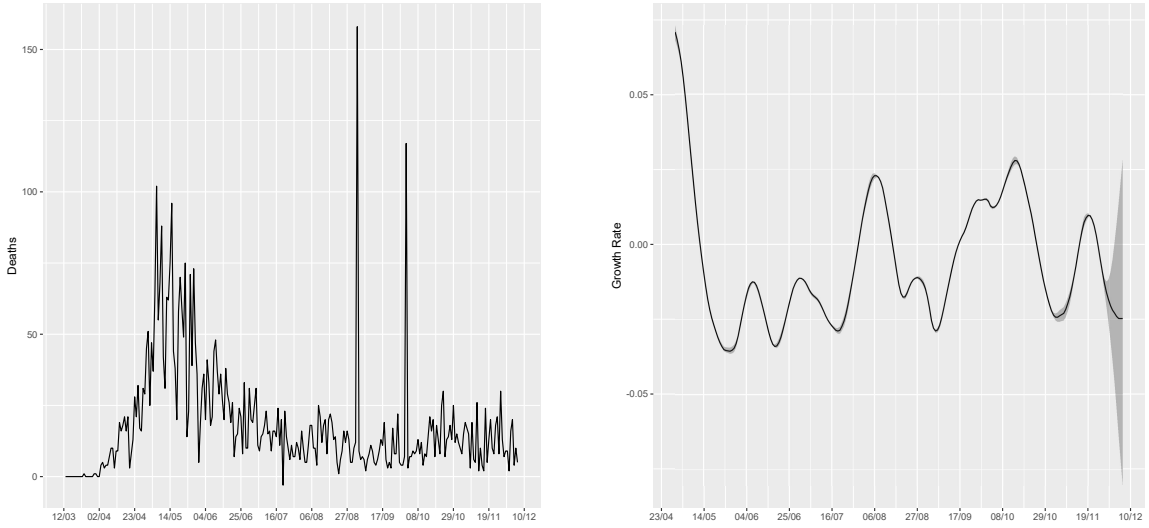
(b) Seasonal

(c) Cycle

Figure 7.7 – osterior mean and 95% credibility intervals of Trend, Seasonal and Cycle decomposition of deaths in the state of Amazonas - 02/25/2020 to 12/06/2020.

component for each state, it is necessary to multiply by the size of the population. We can observe that the spatio-temporal model allows to recover a smoother trend component compared to those obtained in univariate models, indicating that the spatial component captures part of the irregularity observed in the data on death by COVID-19. The trend estimated by this model indicate a general peak about the end of July, and also indicates the existence of a new acceleration (the "second wave") from the second week of November. The seasonality and cycle components are consistent with the patterns observed in the aggregated data for the whole country.

Although the spatio-temporal model presented above allows a useful recovery of the general patterns in the dynamics of deaths related to COVID-19 in Brazil, the assumption of common components for all Brazilian states is a very restricted assumption. The Brazilian regions have different dates of onset of contagion, and also very different dynamics over time. To avoid these restrictions, and still maintain a relatively parsimonious model, we performed a new estimation



(a) Deaths

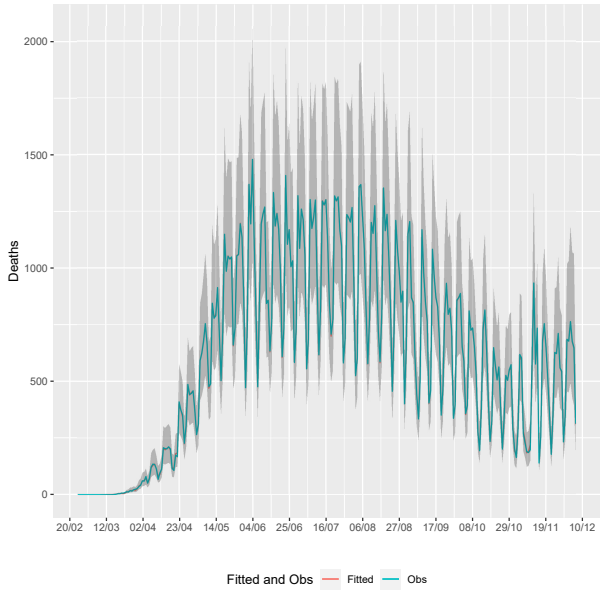
(b) Growth Rate

Figure 7.8 – Number of reported deaths and the estimated growth rate of the Trend in the state of Amazonas - 02/25/2020 to 12/06/2020.

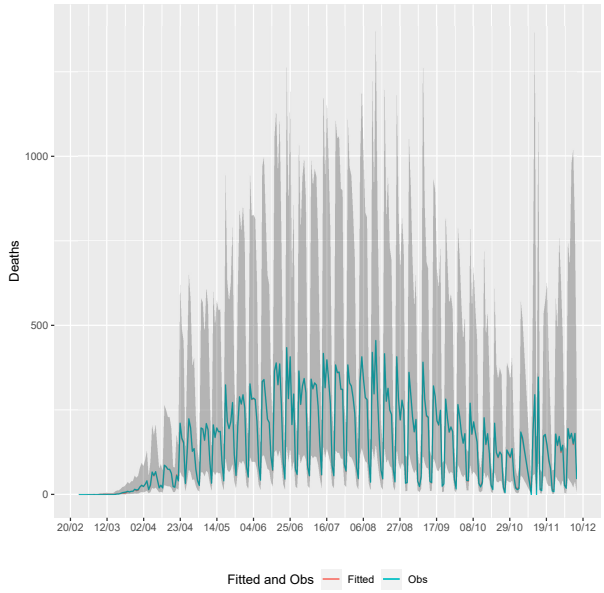
of the spatial model allowing individual components of trend, seasonality and cycle for each region of the country, that is, components for the North, Northeast, Midwest regions, Southeast and South, using the regional division of the country used by *Instituto Brasileiro de Geografia e Estatística* (IBGE).

The parameters estimated in this version with individual components for the Brazilian regions are shown in Table 7.1. Figures 7.12, 7.14 and 7.15 show the components of trend, seasonality and estimated cycles for each region of Brazil in this model. In Figure 7.16 in the Appendix we also show the posterior mean of the spatial effects for two different dates in the sample, illustrating the evolution of the spatial dynamics in deaths related to COVID. Again, the spatio-temporal model makes it possible to recover trend components with greater smoothness than univariate models, indicating the importance of the process of spatial dependence in the dynamics of COVID-19. The regional components of trend also seem to confirm the patterns observed for the number of deaths in Brazil. There seems to be a reduction in the trend of deaths for the Southeast, South and Northeast regions from the end of July, and also a new acceleration from November.

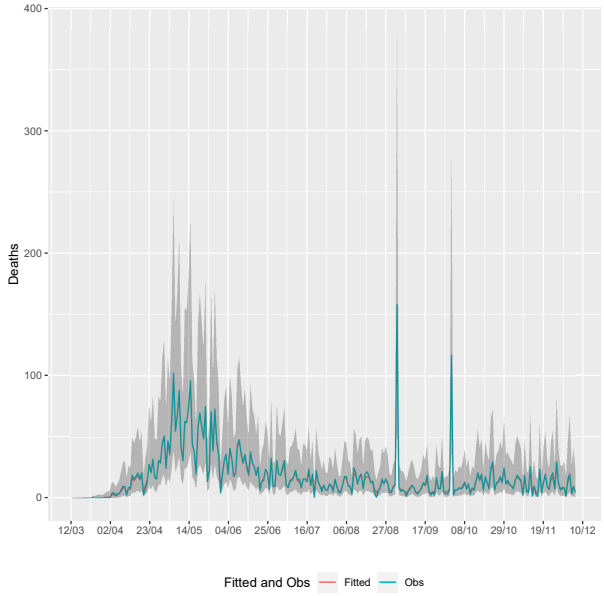
As the latent components in this formulation are formulated in terms of exposure, that is, the number of inhabitants in each region, it is not so easy to directly interpret the estimated trend value. For a more direct interpretation, we carried out a transformation of the trend towards the number of daily deaths per million inhabitants, that is, assuming an exposure of 1,000,000 for each trend. The trend in daily deaths per million inhabitants per region is shown in the Figure 7.13. We can observe the rapid peak that was reached in the North, which reflects the high number of deaths observed in Amazonas in May, and then the rapid reduction and the new acceleration observed at the end of the analyzed sample. Another notable pattern is the high rate of deaths in the Center-West region, which reflects the number and acceleration of cases and deaths in the Federal District, a region with a higher population density. For the Southeast, South and Northeast regions, the pattern is more similar, indicating a acceleration



(a) Brazil



(b) São Paulo



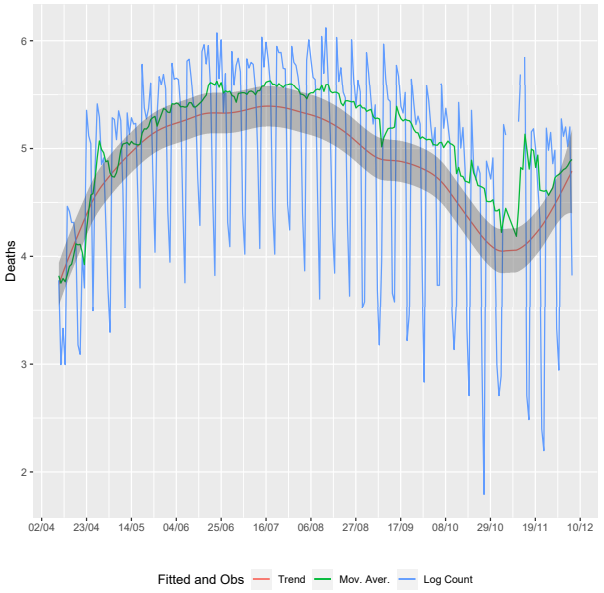
(c) Amazonas

Figure 7.9 – Predicted values and observed deaths - Brazil, São Paulo and Amazonas - 02/25/2020 to 12/06/2020.

in the death rate at the end of the analyzed sample.

The seasonal components (Figure 7.14) are quite different across regions, which is probably reflecting different patterns of occurrence and also methods of accounting for cases and deaths. The cycle (Figure 7.15) components show a little present cycle at the beginning of the sample, which is expected due to a limited number of cases, but a greater relative importance at the end of the sample, after the acceleration in the cases. We can also analyze these components by the estimated autoregressive coefficients. In Table 7.1 we have the first and second order partial correlation coefficients for each region, and we can observe a variety of patterns in these estimated coefficients. The second-order autoregressive component seems to have a clear interpretation of the cyclical component for the Southeast, North and Northeast regions, while in terms of

(a) Brazil



(b) São Paulo

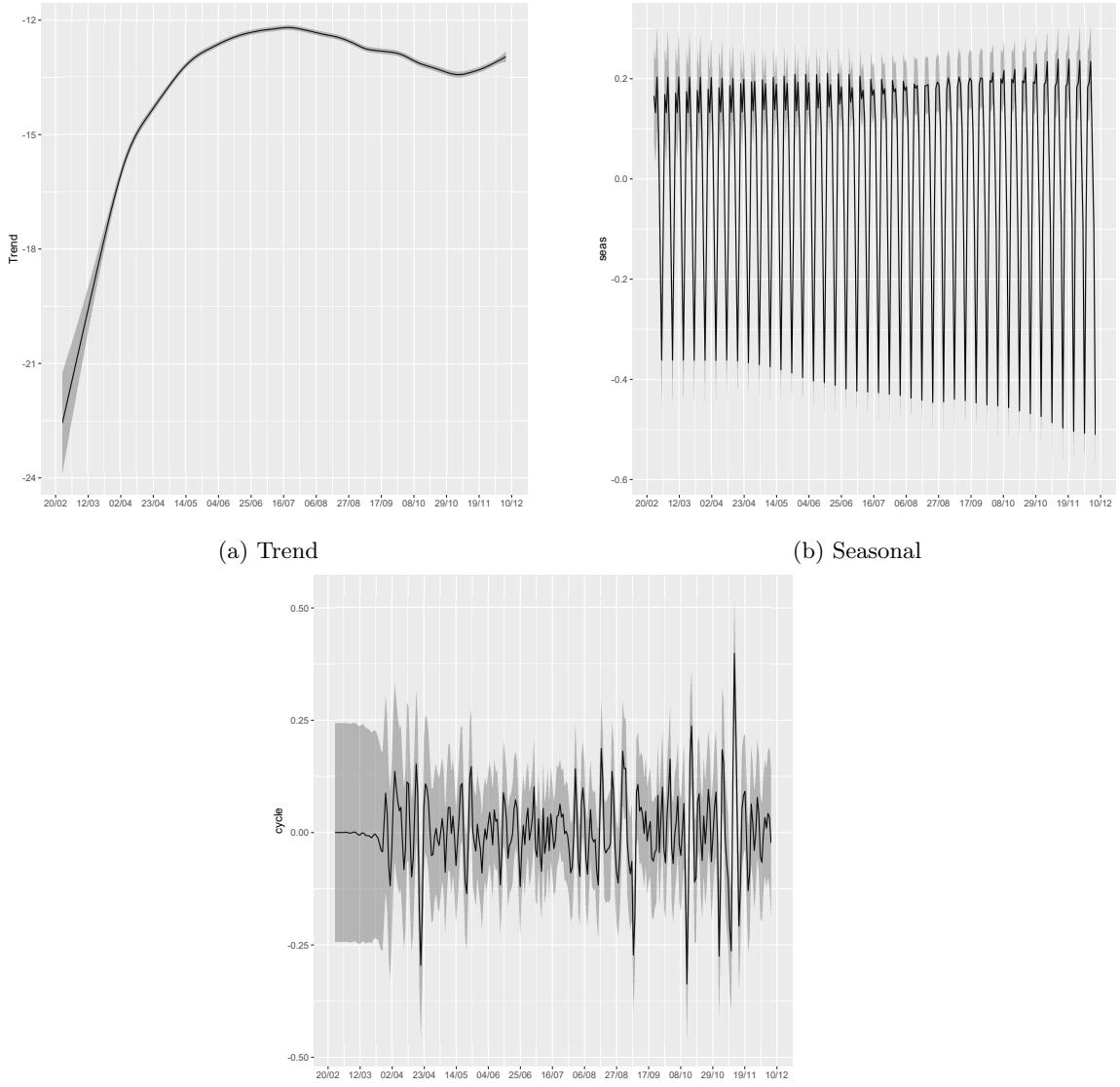
(c) Amazonas

Figure 7.10 – Predicted Values - Estimated trend and moving average filter for the log-transformed count data - - 02/25/2020 to 12/06/2020.

magnitude it seems less important for the South and Midwest regions.

We can compare the models with common components and regional components using information criteria. In Tables 7.6 and 7.1 we report two common information criteria for Bayesian estimates. Deviance Information Criterion (DIC) and Widely Applicable Information Criterion (WAIC), which is a generalized version of AIC. By the two criteria the model with regional components is selected as the most suitable model. This result indicates the existence of a relevant heterogeneity in covid-19 standards in Brazil.

We compared some model fit measures for all models analyzed in this work, which allows to analyze especially the importance of the spatial component in the parameterization of the models, as well as the effects in the adoption of specific components of trend, seasonality and



(a) Trend

(b) Seasonal

(c) Cycle

Figure 7.11 – Posterior mean and 95% credibility interval of Trend, Seasonal and Cycle decomposition - Spatio-Temporal model with common trends, seasonal and cycle components - 02/25/2020 to 12/06/2020. Note: Components estimated with the adjustment for the size of the population in each region (exposure).

cycle for each region. Table 7.2 shows the mean error (ME), root mean squared error (RMSE) and the mean absolute error (MAE) of univariate models and space-time models, using the posterior median as the point measure of model fit. We also built the specific adjustment measures for the states of São Paulo and Amazonas, obtained by the models with spatial components, for comparison with the adjustment of the time series models.

An important first result is that the fit of spatio-temporal models is clearly superior to that of univariate models of time series in all metrics. Note that this gain reflects the importance of inter-state transmission patterns in the COVID-19 dynamics. Regarding the inclusion of specific components for regions, we can see that in general these specific components lead to a general reduction in the mean error term, but are slightly worse in terms of RMSE and MAE for some regions, which indicates a possible overparameterization of the model and the presence

Table 7.1 – Estimated parameters of deaths reported in Brazil - Spatio-temporal model with region specific trend, seasonal and cycle components.

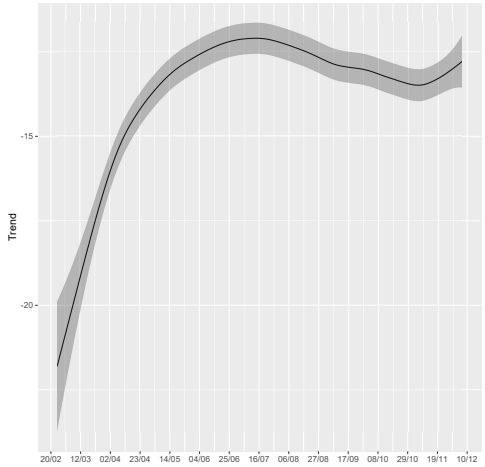
	Mean	SD	0.025quant	0.5quant	0.975quant	Mode
Precision for trend southeast	1.68e4	1257.446	1.46e4	1.67e4	1.95e4	1.65e4
Precision for trend south	4.94e3	336.875	4.29e3	4.94e3	5.61e3	4.95e3
Precision for trend north	1.65e4	1186.588	1.41e4	1.65e4	1.88e4	1.66e4
Precision for trend northeast	4.80e4	4829.782	4.10e4	4.72e4	5.95e4	4.46e4
Precision for trend center-west	3.05e4	3986.805	2.45e4	2.98e4	3.99e4	2.80e4
Precision for seas southeast	2.43e2	18.723	2.09e2	2.42e2	2.82e2	2.39e2
Precision for seas south	2.28e3	163.394	1.96e3	2.28e3	2.60e3	2.29e3
Precision for seas north	1.11e4	1284.404	8.37e3	1.12e4	1.32e4	1.18e4
Precision for seas northeast	1.27e5	9207.029	1.12e5	1.26e5	1.48e5	1.24e5
Precision for seas center-west	5.62e3	421.451	4.92e3	5.58e3	6.56e3	5.45e3
Precision for cycle southeast	3.21e1	2.927	2.65e1	3.20e1	3.80e1	3.20e1
PACF1 for cycle southeast	1.47e−1	0.037	7.30e−2	1.47e−1	2.20e−1	1.47e−1
PACF2 for cycle southeast	−1.24e−1	0.034	−1.92e−1	−1.23e−1	−6.00e−2	−1.20e−1
Precision for cycle south	3.45e1	2.481	3.02e1	3.43e1	3.99e1	3.36e1
PACF1 for cycle south	1.16e−1	0.034	5.10e−2	1.16e−1	1.84e−1	1.13e−1
PACF2 for cycle south	4.50e−2	0.036	−2.00e−2	4.30e−2	1.19e−1	3.60e−2
Precision for cycle north	1.54e1	1.449	1.31e1	1.52e1	1.88e1	1.47e1
PACF1 for cycle north	4.83e−1	0.034	4.23e−1	4.80e−1	5.55e−1	4.69e−1
PACF2 for cycle north	−3.30e−1	0.030	−3.91e−1	−3.29e−1	−2.72e−1	−3.25e−1
Precision for cycle northeast	4.07e2	28.659	3.57e2	4.05e2	4.69e2	3.99e2
PACF1 for cycle northeast	7.49e−1	0.027	7.06e−1	7.45e−1	8.07e−1	7.31e−1
PACF2 for cycle northeast	−2.70e−2	0.098	−2.32e−1	−2.00e−2	1.51e−1	5.00e−3
Precision for cycle center-west	2.37e1	2.481	1.99e1	2.33e1	2.95e1	2.22e1
PACF1 for cycle center-west	6.10e−2	0.042	−2.90e−2	6.40e−2	1.34e−1	7.60e−2
PACF2 for cycle center-west	−2.41e−1	0.035	−3.04e−1	−2.43e−1	−1.69e−1	−2.50e−1
Precision for CAR	4.81e−1	0.013	4.57e−1	4.81e−1	5.08e−1	4.79e−1
Group Φ	8.00e−1	0.006	7.88e−1	8.00e−1	8.12e−1	8.00e−1
Deviance Information Criterion (DIC)	40003.80					
Watanabe-Akaike information criterion (WAIC)	40265.38					

of possible common trends.

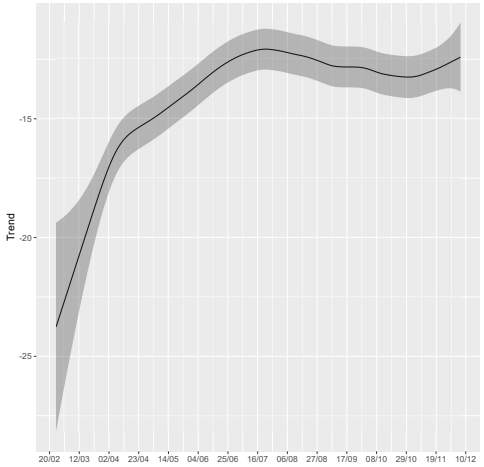
7.4 Conclusions

The methods proposed in this work are based on hierarchical formulations using Bayesian inference methods for time series and spatio-temporal processes. In particular, we proposed methods to estimate the trend in the deaths by COVID-19, through trend-cycle decomposition for counting processes with latent components, which were applied for the total number of deaths in Brazil and for the states of São Paulo and Amazonas. In addition, the estimated trend component was compared to the empirical analysis based on moving averages. In order to achieve the main goal, we used the BRASIL.IO daily data series on deaths by COVID-19 in Brazil and in the states of São Paulo and Amazonas, from 02/25/2020 to 12/06/2020.

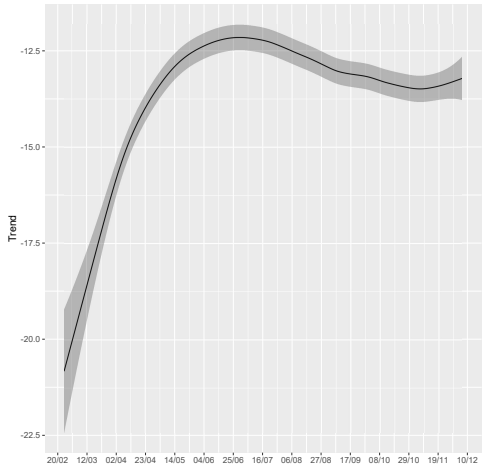
The use of Bayesian inference methods is especially useful in situations such as the spreading a new epidemic with transmission, latency and mortality characteristics that are not well known. The hierarchical formulation reflects the process of information accumulation in time and space, where the posterior distribution in a period t serves as a prior for the next period, and also the use of information from neighbors as a prior structure to make inference about the parameters in a certain region. In a situation where there is little prior information on essential aspects of the problem, the incorporation of learning in time and space through a Bayesian mechanism allows for the efficient use of the new information available, which is essential for the emergency formulation of health and prevention policies, in a general environment of information of dubious



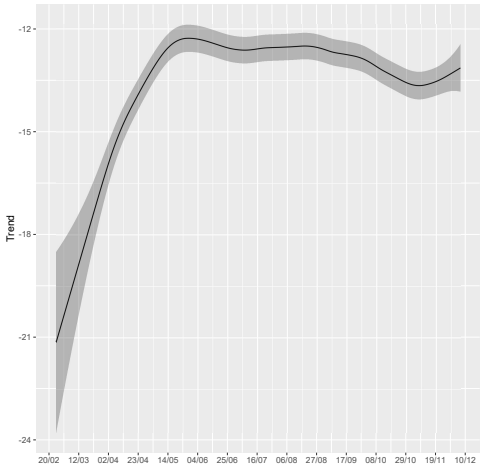
(a) Trend Southeast



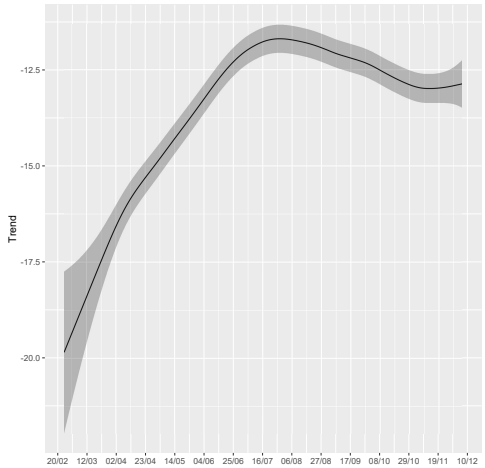
(b) Trend South



(c) Trend Northeast



(d) Trend North



(e) Trend Center-West

Figure 7.12 – Posterior mean and 95% credibility interval of Trend - Spatio-temporal model with region specific trends, seasonal and cycle components - 02/25/2020 to 12/06/2020. Note: Components estimated with the adjustment for the size of the population in each region (exposure).

quality or propagation of false information or without a scientific basis. Bayesian learning is essential in this context as a basis for evidence and data-based policies.

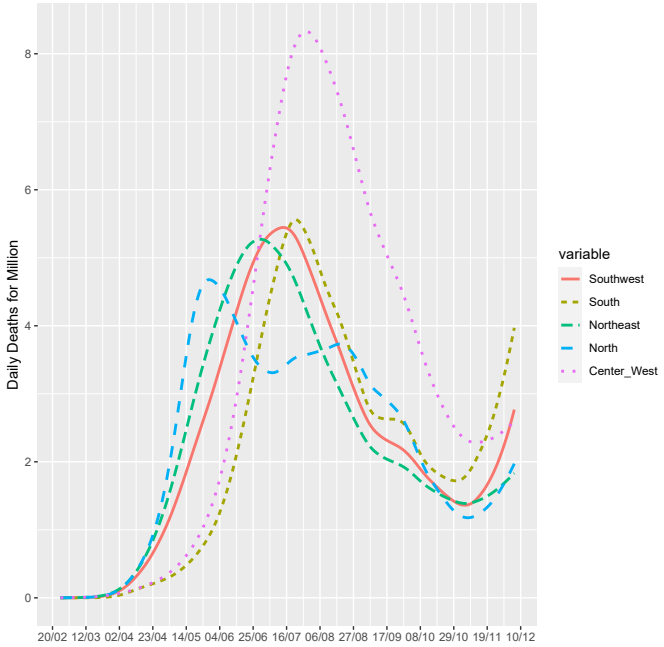
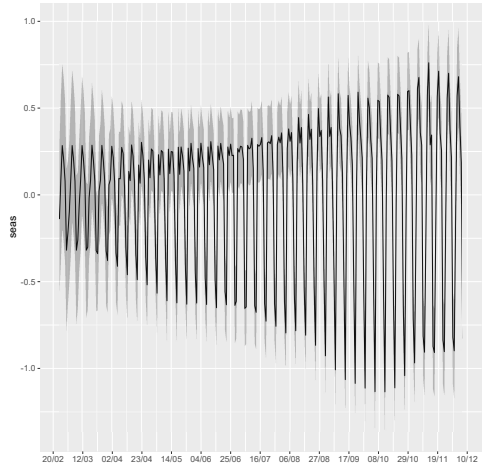


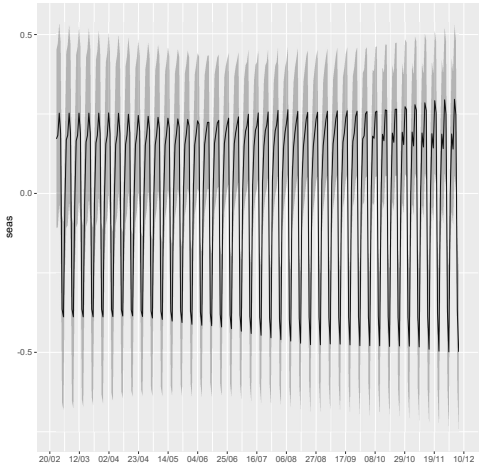
Figure 7.13 – Region-specific Trends in Deaths by day by Million inhabitants - 02/25/2020 to 12/06/2020.

The results for Brazil showed that our proposed models were able to capture the long-term movements in death cases, showing an accelerated growth pattern until early April, where the pace of the trend has slowed down, reaching the peak in July, suggesting that the isolation measures taken by authorities might have been partly effective to change the growth of the death trend and postpone the peak. However, the observed sustainable high peak between May and August provided evidence that the social distancing measures might have not been enough to slow down the spread of the COVID-19, which can be related to the absence of political actions at federal level. The model also captured a negative trend in the number of new deaths until the beginning of November 2020, and a possible "second wave" after this period, with a new accelerated growth in the trend, which may be related to the electoral period and an excessive relaxation of social distance measures by local authorities. The results of the estimated trend component for the states of São Paulo are similar to those observed for Brazil. For the state of Amazonas, the estimated trend reached the peak faster than the previous analyzed states, which may be due to the lack of health care infrastructure and the difficult of some patients to access proper heath care, concentrated in the state capital, Manaus.

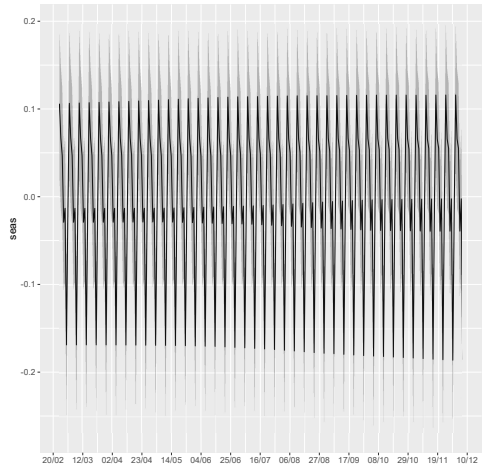
Additionally, the results obtained indicate that the trend component estimated by the univariate and spatio-temporal models is a more robust indication of the general patterns in the occurrence of deaths related to COVID-19, overcoming the existing limitations in simple smoothing measures such as the use of moving averages, which are not robust to the measurement errors introduced by the case accounting mechanisms. The models also obtained a very precise adjustment for the number of occurrences, especially in spatio-temporal models with the incorporation of spatial propagation patterns. These models also provided a smoother component of trend, when compared to univariate models, which can be explained by the greater information available in the estimation process, using all the information available to Brazilian states, which



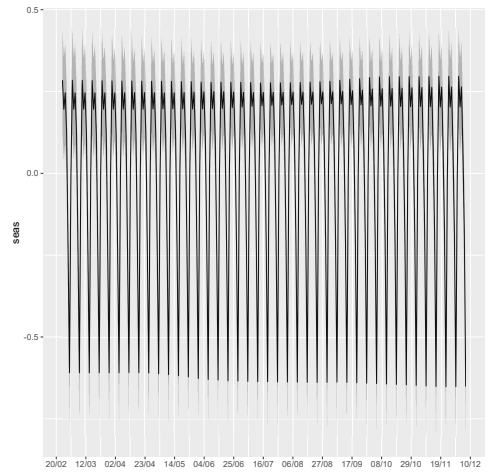
(a) Seasonal Southeast



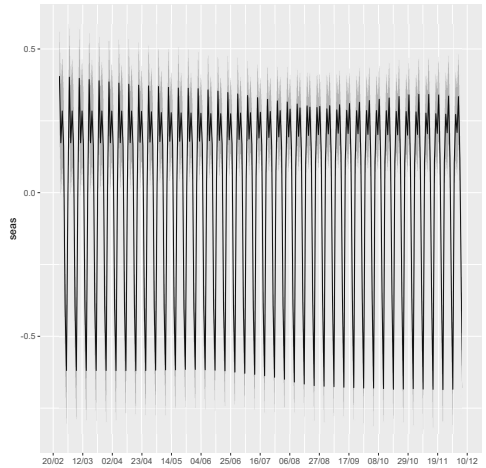
(b) Seasonal South



(c) Seasonal Northeast



(d) Seasonal North

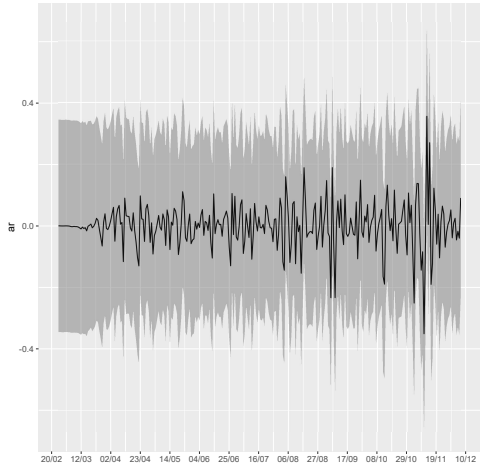


(e) Seasonal Center-West

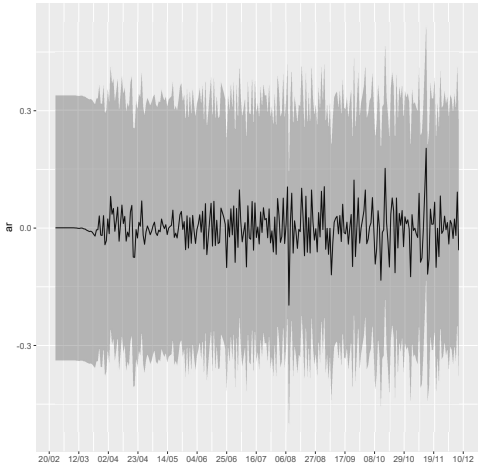
Figure 7.14 – Posterior mean and 95% credibility interval of Seasonal - Spatio-temporal model with region specific trends, seasonal and cycle components - 02/25/2020 to 12/06/2020. Note:Components estimated with the adjustment for the size of the population in each region (exposure).

helps to mitigate the measurement error problem.

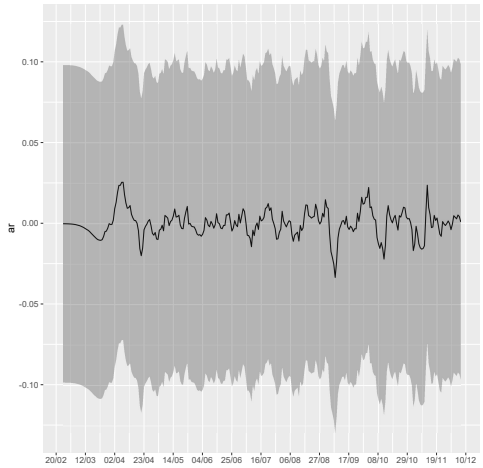
We believe that our work contributes to the enormous effort of studying and analyzing



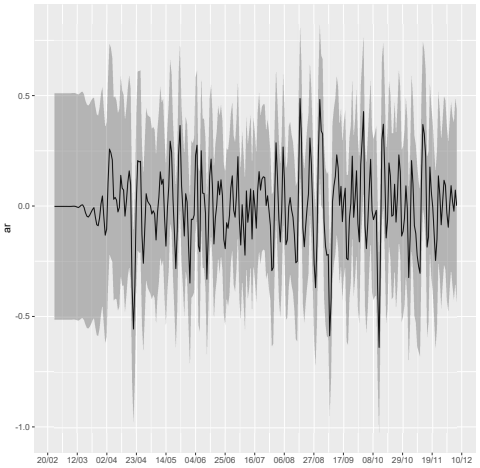
(a) Cycle Southeast



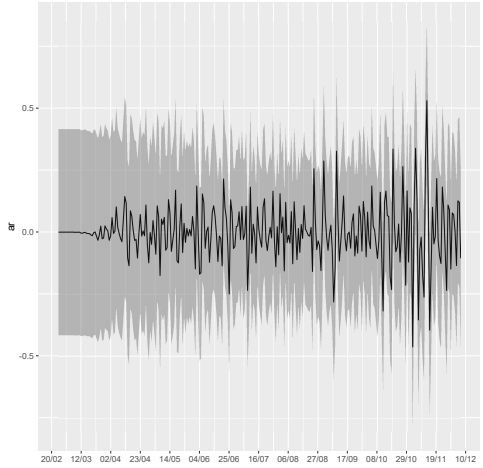
(b) Cycle South



(c) Cycle Northeast



(d) Cycle North



(e) Cycle Center-West

Figure 7.15 – Posterior mean and 95% credibility interval of Cycle - Spatio-temporal model with region specific trends, seasonal and cycle components - 02/25/2020 to 12/06/2020. Note: Components estimated with the adjustment for the size of the population in each region (exposure).

the impacts of COVID-19 in Brazil, complementing several other initiatives of epidemiological, statistical and computational modeling, such as projects Brasil.IO (<https://brasil.io/dataset/

Table 7.2 – Model Fit Statistics

	ME	RMSE	MAE
Univariate Models			
Brazil	-0.5236	5.7872	4.0368
SP	-0.6094	1.9461	1.4112
AM	-0.2586	1.7920	1.2175
Spatio-temporal Model Common Components			
Brazil	−1.18e−6	2.3204	1.4613
Southeast	-0.0004	3.5379	2.3078
South	0.0124	1.5199	1.0590
Northeast	0.0008	2.0400	1.3224
North	0.0026	1.8968	1.2172
Center-West	-0.0152	2.4802	1.6213
SP	0.0040	2.2366	1.4146
AM	0.0044	2.2926	1.4384
Spatio-temporal Region-Specific Components			
Brazil	−9.79e−7	2.4051	1.5297
Southeast	1.48e−7	3.5985	2.4151
South	−2.03e−6	1.5589	1.1067
Northeast	−1.41e−6	2.2406	1.4285
North	−6.09e−7	2.0050	1.2895
Center-West	−1.00e−6	2.3820	1.5754
SP	0.0036	2.3341	1.4881
AM	0.0047	2.3669	1.5019

covid19/caso_full/>), Covid 19 Analytics (<https://covid19analytics.com.br/>), COVID-19 Brasil (<https://ciis.fmrp.usp.br/covid19/>), MonitoraCovid-19 (<https://bigdata-covid19.icict.fiocruz.br/>) and several other initiatives of great value for Brazilian society.

7.5 Appendix

7.5.1 Estimated Posterior Distribution of Parameters

Table 7.3 – Estimated parameters of deaths reported in Brazil

	Mean	SD	0.025quant	0.5quant	0.975quant	Mode
Precision for trend	10147.060	3996.089	4447.125	9454.153	19872.796	8201.470
Precision for seasonality	2017.824	803.563	881.916	1874.032	3987.542	1618.580
Precision for cycle	38.444	4.970	29.407	38.212	48.877	37.840
PACF1 for cycle	0.290	0.067	0.156	0.291	0.421	0.290
PACF2 for cycle	-0.204	0.078	-0.351	-0.206	-0.047	-0.210

Table 7.4 – Estimated parameters of deaths reported in the state of São Paulo

	Mean	SD	0.025quant	0.5quant	0.975quant	Mode
Precision for trend	10959.712	4710.471	4245.146	10164.461	2.24e4	8622.798
Precision for seasonality	6.625	30.858	15.651	37.878	1.28e2	27.062
Precision for cycle	5.058	1.085	3.119	5.007	7.350	4.933
PACF1 for cycle	-0.002	0.086	-0.170	-0.003	1.67e−1	-0.004
PACF2 for cycle	-0.126	0.085	-0.292	-0.127	4.20e−2	-0.126

Table 7.5 – Estimated parameters of deaths reported in the state of Amazonas

	Mean	SD	0.025quant	0.5quant	0.975quant	Mode
Precision for trend	18777.970	2.28e4	3222.621	11039.803	8.43e4	5523.388
Precision for seasonality	590.342	5.47e2	150.322	421.364	2.04e3	259.600
Precision for cycle	4.812	6.52e−1	3.655	4.771	6.210	4.692
PACF1 for cycle	-0.069	1.05e−1	-0.269	-0.071	1.41e−1	-0.076
PACF2 for cycle	0.020	8.70e−2	-0.146	0.018	1.94e−1	0.010

Table 7.6 – Estimated parameters of deaths reported in Brazil - spatio-temporal model with common trend, seasonal and cycle components.

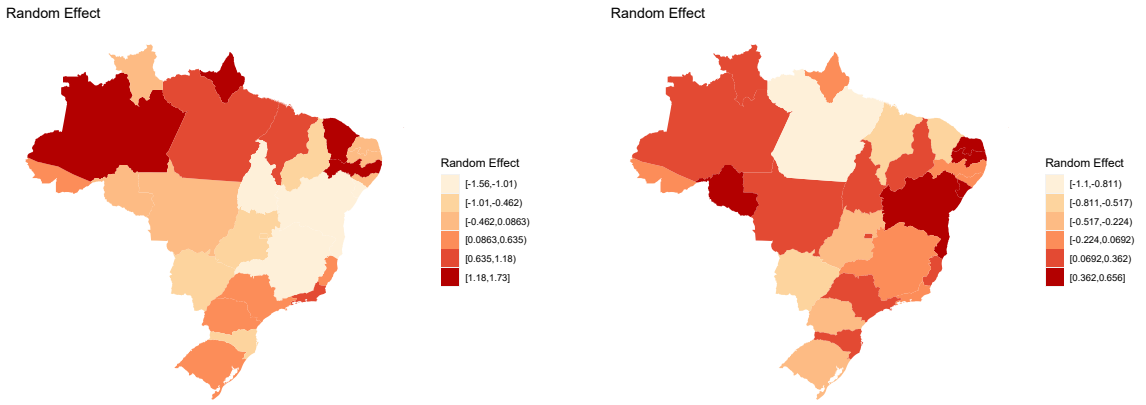
	Mean	SD	0.025quant	0.5quant	0.975quant	Mode
Precision for trend	30562.848	8148.802	16735.991	29929.161	48350.874	28689.914
Precision for seasonality	12034.725	3292.864	7366.350	11413.624	20093.088	10209.154
Precision for cycle	67.156	9.515	51.093	66.207	88.328	64.095
PACF1 for cycle	0.4308	0.052	0.334	0.439	0.537	0.440
PACF2 for cycle	-0.559	0.057	-0.657	-0.563	-0.436	-0.575
Precision for CAR	0.439	0.019	0.405	0.437	0.481	0.431
Group Φ	0.785	0.010	0.763	0.785	0.803	0.787
Deviance Information Criterion (DIC)	40015.16					
Watanabe-Akaike information criterion (WAIC)	40188.55					

7.5.2 Spatial random effects

As an illustration of the dynamics of spatial random effects, we show in Figure **??** the effects estimated by the model with the regions specific components for the days 04/26/2020 and 08/14/2020. The variation observed in the two dates shows the importance of using a dynamic structure for the spatial effects in the modeling of epidemic processes.

7.5.3 SEIR Model with measurement error

We use a Susceptible-Exposed-Infective-Resistant (SEIR) model to generate realizations of an epidemic model calibrated to reproduce some essential aspects of the COVID-19 epidemic. We use the following structure, based on Liu et al. (2020), to generate a solution for the SEIR model:



(a) 26/04/2020 (b) 26/04/2020

Figure 7.16 – Posterior mean of Spatial Random Effects - Spatio-temporal model with region specific trends, seasonal and cycle components - 26/04/2020 and 26/04/2020

$$
\begin{aligned}
dS(t)/dt &= -\beta I(t)S(t)/N \\
dE(t)/dt &= \beta S(t)I(t)/N - \delta(t)E(t-7) \\
dI(t)/dt &= \delta(t)E(t-7) - \gamma(t)I(t-10) - \eta(t)I(t-10) \\
dR(t)/dt &= \gamma(t)I(t-10) \\
dF(t)/dt &= \eta(t)I(t-10)
\end{aligned}
\tag{7.3}
$$

$S(t)$ is the number of susceptible individuals in the population in period t, $E(t)$ is the number of exposed individuals, $I(t)$ the number of infected, $R(t)$ the resistant individuals and $F(t)$ the number of fatalities. The total number of individuals is N, given by the sum of the individuals in all states. β is a parameter which controls the average number of exposed cases that are generated by one infected person. The parameter δ is the probability of an exposed individual migrate to the infected state. γ is the parameter for the probability of transition of a infected individual to the recovered state, and η the probability of the infected individual to migrate to fatality state. We assume the values 1.3629, 0.0262,0.01,0.004 for $\beta, \gamma, \delta, \eta$, and a total population size of 10 million of inhabitants. The model assumes a period of 7 days between exposure and infection, and 10 days between recovery or fatality after infection. The model is simulated for 180 periods, and we focus on the number of fatalities generated by the model. As commented in the main text, we generated a measurement error process assuming that fatalities on Saturday and Monday has a proportion, drawn from a uniform distribution with parameters (.5, .8), of data with delayed disclosure, on the following Monday and Tuesday. We generated 10,000 replications of the measurement error process, and calculated the ACF and PACF functions for the difference between the true number of fatalities and the reported daily number of fatalities, which is the seasonal effect induced by the measurement error process. For

the simulations of the SEIR model we use the deSolve package from the r-project.org software. In particular, we use the dede function, which is a solver for differential equations with delay.

7.5.4 Integrated Nested Laplace Approximations

The INLA method proposed by Rue, Martino and Chopin (2009) is a methodology based on deterministic Laplace approximations to perform accurate and efficient approximations on the class of Bayesian hierarchical models that can be represented as Gaussian Markov random fields (GMRF). See Rue and Held (2005c) for a detailed discussion on this class of models.

A latent GMRF model is a hierarchical model with the first stage/level defining a conditional distribution for the observed variable y, usually assumed to be conditionally independent given the latent factors x and some additional (hyper)-parameter θ, in the form:

$$\pi(y|x,\theta) = \prod_j \pi(y_j|x_j,\theta), \quad j \in J \tag{7.4}$$

with y_j for $j \in J$ observed values and J a subset of the latent factors; and $\pi(y|x,\theta)$ defining the likelihood function of observed variables. The latent (hyper)-parameters constitutes the second stage in the hierarchical formulation:

$$x_i = \text{Offset}_i + \sum_{k=0}^{\eta_f - 1} \omega_{ki} f_k(c_{ki}) + z_i^T \beta + \epsilon_i, \quad i = 0, \ldots, \eta_x - 1 \tag{7.5}$$

The offset term is a prior known component to be included in the linear prediction; for example, in the Poisson likelihoods the offset is the exposure effect. ω_k are known weights for each observed data point in the sample, and $f_k(c_{ki})$ represents the effect of covariates with value c_{ki} for each observation i; β are the regression parameters of linear covariates z_i. Finally, the third and last stage of the model consists of the prior distribution for the hyperparameters θ.

The INLA approach obtains accurate approximations using sequential Laplace approximations in the mode of the posterior distributions of the latent factors, written as:

$$\pi(x_i|Y) = \int \pi(x_i|\theta,Y)\pi(\Theta,Y)d\theta \tag{7.6}$$

and for the the marginal posterior distribution of (hyper)parameters:

$$\pi(\theta_j|Y) = \int \pi(\theta|Y)d\theta\_j \tag{7.7}$$

The element θ_j denotes the vector θ with its jth element omitted. The INLA method is realized in three main steps. The first is an approximation to the full posterior distribution $\pi(\theta|y)$ by a Laplace approximation in the mode of the distribution, where the mode is found using a numerical optimization algorithm. The second step is an approximation to the full conditional distributions $\pi(x_i|\theta,y)$ for specific values of θ. The last step of the approximation gets an approximation for the marginal posterior distributions in (7.6) and (7.7) by combining the two approximations in the previous steps and integrating out the irrelevant factors. This method was introduced in Rue, Martino and Chopin (2009), and extended to several classes of models. Surveys of recent developments in this methodology for spatial modelling can be

found at Bakka et al. (2018), a discussion on the use on log Gaussian Cox processes for discrete domains in Illian, Sorbye and Rue (2012), and textbook references in Blangiardo and Cameletti (2015) and Gomez-Rubio (2020).

8 ESTIMATING SPATIOTEMPORAL PATTERNS OF DEATHS BY COVID-19 OUTBREAK ON A GLOBAL SCALE

Fernanda Valente Márcio Laurini

ABSTRACT

Our main objective is to estimate the trend of deaths by COVID-19 on a global scale, considering the six continents. The study design was a retrospective observational study conducted using the secondary data provided by the Our World in Data project on a public domain. This study was conducted based on worldwide deaths by COVID-19 recorded for the Our World in Data project from February 29th, 2020 to February 17th, 2021. Estimating the trend in COVID-19 deaths is not a trivial task due to the problems associated with the COVID-19 data, such as the spatial and temporal heterogeneity, observed seasonality, and the delay between the onset of symptoms and diagnosis, indicating a relevant measurement error problem and changing the series' dependency structure. To bypass the aforementioned problems, we propose a method to estimate the components of trend, seasonality, and cycle in COVID-19 data, controlling for the presence of measurement error and considering the spatial heterogeneity. We used the proposed model to estimate the trend component of deaths by COVID-19 on a global scale. The model was able to capture the patterns in the occurrence of deaths related to COVID-19, overcoming the problems observed in COVID-19 data. We found compelling evidence that spatio-temporal models are more accurate than univariate models to estimate the patterns of the occurrence of deaths. Based on the measures of dispersion of the models' prediction in relation to observed deaths, it is possible to note that the models with spatial component are significantly superior to the univariate model. The findings suggested that the spatial dynamics have an important role in the COVID-19 epidemic process since the results provided evidence that spatio-temporal models are more accurate to estimate the general patterns of the occurrence of deaths related to COVID-19.

Keywords: Time Series Decomposition; Spatio-temporal Point Process; COVID-19

VALENTE, Fernanda; LAURINI, Marcio Poletti. Estimating spatiotemporal patterns of deaths by COVID-19 outbreak on a global scale. **BMJ open**, v. 11, n. 8, p. e047002, 2021.

8.1 Introduction

On March 11, 2020, the World Health Organization (WHO) declared the coronavirus disease (COVID-19) outbreak as a pandemic, which is currently a public health threat in more than 200 countries. Since there is no effective pharmaceutical treatments available to deal with the COVID-19 (DONG; HU; GAO, 2020; SANDERS et al., 2020), some public health measures have been proposed by countries to cope with the rapid spread of the COVID-19, such as social distancing, self-quarantine and lockdown. However, preventing the transmission depends on how well we can assess the number of infected people. Statistical modeling is helpful to obtain

the trend component and better understand the evolution of the confirmed cases and deaths of COVID-19, and also provide key information about the effects of non-pharmaceutical measures to control the transmission. Regarding on the methods for trend estimation, there are several different methods reported in the literature, which differ in their complexity and interpretability. The most prevalent trend estimation methods are model-based trend extraction, nonparametric filtering, singular spectrum analysis and wavelets (ALEXANDROV et al., 2012).

To estimate COVID-19 trends, some of aforementioned approaches have been applied to the current outbreak (LI; FENG; QUAN, 2020; GUPTA; PAL, 2020; CEYLAN, 2020; PERONE, 2020). However, it should be noted that despite the great variety of papers dedicated to study the long-movements of COVID-19, there is little understanding about the spread pattern of COVID-19 in a global scale. In addition, it is worth noting some problems related to COVID-19 data. First, is observed a delay between the onset of symptoms and accurate diagnosis, which may vary from country to country, depending on the local government's strategy (CONTRERAS et al., 2020). Also, there are unconformities in the report of new deaths which are counted on the day that are included in the system, instead of the actual day of the death (RUSSELL et al., 2020). These problems create a relevant measurement error problem, changing the series' dependency structure. In addition, the delays are also related to the lack of personnel available to report cases on weekends, which creates a seasonality structure in the cases and death reports, generating an additional aggregation problem on the time series (SKIERA et al., 2020).

To bypass these problems, we propose to estimate the long-term movements of deaths by COVID-19 in a global scale through a structural decomposition (HARVEY, 1990). The main idea is to decompose the temporal variability observed in the data into trend, seasonal and cycle components, which allows to identify permanent movements, and cyclical and seasonal effects, in the presence of measurement errors. In addition, since we are interested to assess the spread pattern of COVID-19 in a global scale, the importance of a spatial component is undeniable, since the use of this information is important in estimating the trend and cycle components, due to the nature of transmission in an epidemic process, with spatial spread dynamics. In the light of this, we include a component that considers the spatial heterogeneity among different areal unities in the study of COVID-19 evolution. We propose two different spatial formulation which relies on the definition of neighborhoods. The first is based on regions that present common borders, while the second includes information of air transport network. Regarding this context, the contribution of this paper is to explore a Bayesian version of a trend-cycle decomposition in combination with count distributions, taking into account the spatial heterogeneity (VALENTE; LAURINI, 2020), in the task of estimating the permanent movements of deaths by COVID-19 outbreak in a global scale.

8.2 Material and Methods

In this paper, we present the results of estimating the proposed model for COVID-19-related death data for countries collected by the Our World in Data project (available at https://ourworldindata.org/covid-deaths), using daily information on new deaths for the period from February 29th, 2020, until February 17th, 2021. Therefore, the study design was a retrospective observational study conducted using secondary data on a public domain. Patients

and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research. We use the countries on that basis that contained information on the total population, corresponding to a total of 205 countries. Also, we use the continental division of countries as a region definition, and thus k=6, corresponding to the continents of Asia, Europe, South America, North America, Africa, and Oceania.

8.2.1 Patient and Public Involvement Statement

No patient involved.

8.2.2 Statistical Model

The proposed model is a generalization of the Bayesian model introduced by Valente and Laurini (2020) for estimating the COVID-19 trend for data in Brazil, controlling for problems of measurement error in the construction of death statistics. The model is based on a generalized decomposition of trend, seasonality and cycle components incorporating a time-varying spatial component. The spatial component is based on the Conditional Autoregressive (CAR) struture (BESAG, 1974), based on the information of occurrences in some definition of neighborhood to formulate prior information for the number of occurrences in the region of interest. The structure of the model can be summarized by the equations:

$$
\begin{aligned}
Y_{(i,t)} &= Poisson(exp(\lambda_{(i,t)}E), \\
\lambda_{(i,t)} &= \mu_{(k,t)} + s_{(k,t)} + c_{(k,t)} + \xi_{(i,t)} \\
\Delta^2\mu_{(k,t)} &= \eta_{\mu(k,t)} \\
s_{(k,t)} &= s_{(k,t-1)} + s_{(k,t-2)} + \ldots + s_{(k,t-m)} + \eta_{s(k,t)} \\
c_{(k,t)} &= \theta_1 c_{(k,t-1)} + \theta_2 c_{(k,t-2)} + \eta_{c(k,t)} \\
\xi_{(i,t)} \mid \xi_{(j,t),i\neq j,\tau} &\sim \mathcal{N}\left(\frac{1}{n_i}\sum_{i\sim j}\xi_{(j,t)}, \frac{1}{\tau n_i}\right) \\
\xi_{(j,t)} &= \Phi\xi_{(j,t-1)}
\end{aligned}
\tag{8.1}
$$

where $Y_{(i,t)}$ is the total number of occurrences (deaths) in region i and time t. E is the exposure offset, controlling for the total population regarding the cases in $Y_{(i,t)}$. This offset allows to use common components for regions with different population size. It is important to highlight that the used database contains daily information on deaths by COVID-19 of 205 countries, thus the region i corresponds to the country level. However, to estimate the latent factors, we define the region k, which contains a certain group of individuals i. In particular, we define k corresponding to each of the six continents, namely, Asia, Europe, Oceania, Africa, North America, and South America, where each continent k contains a certain group of countries i. In this sense, the $\mu_{(k,t)}$ component captures the general trend of the process for a region k for some definition of region, containing a group of individuals i, and representing the average level of occurrences in period t for the region k. We assume a second-order random walk (RW2) structure for the trend component. The RW2 is a flexible structure suitable for epidemics with rapid spread, which is able to model the evolution of the average number of cases in fast-growing

processes, since it assumes that both the average level and the rate of growth of the level are non-stationary processes (VALENTE; LAURINI, 2020). Also, it is worth noting that this model has a non-parametric nature, since it can be formulated as a spline model, and thus suitable for a new epidemic whose parametric case structure is unknown.

The model also contains region specific seasonal ($s_{(k,t)}$) and cycle ($c_{(k,t)}$) components. These two components are used to capture periodic components in the series of deaths related to COVID-19, where we assume that these components are transient, i.e., have a sum equal to zero. As discussed by Valente and Laurini (2020), these components are a way of controlling the effects of measurement errors introduced by the mechanism for disclosing the number of cases and deaths related to COVID-19, where the total number of cases reported on a certain day corresponds to the total recorded in the day, instead of the exact date of the diagnosis confirmation or death. Using simulations of a SEIR model, the authors show that this accounting mechanism introduces patterns similar to seasonality and cycle effects in the series of cases and deaths. The $s_{(k,t)}$ component represents a sum of stochastic components, with the number of components given by the periodicity of the series and with a sum-to-zero restriction. The $c_{(k,t)}$ component is represented by a second-order autoregressive process with possibly complex roots, which permits to capture cyclical components, also assuming a restriction of zero sum. It is worth noting that the cyclic component also allows us to capture other patterns of periodic dependency in the series, such as transmission waves without permanent effects in the series.

The $\eta_{\mu(k,t)}$, $\eta_{c(k,t)}$, and $\eta_{s(k,t)}$ components are nonspatial independent innovations with zero mean Gaussian distributions with estimated precision (inverse of variances). $\xi_{(i,t)}$ is the spatial random effect component, allowing to incorporate spatial variability in the occurrence rate, n_i is the number of neighbors of region i, and $i \sim j$ indicates that the two regions i and j are neighbors. The CAR structure used defines a prior for the spatial effect in the region i given by a Gaussian distribution with mean given by the average of the values of the spatial effects for regions j that are neighborhood to region i, and variance controlled by a τ precision parameter multiplied by number of neighbors. As the spatial variability in an epidemic process varies over time, we assume that this component is also time varying. The dynamic formulation for the spatial CAR is parameterized through an autoregressive structure, via a parameter Φ controlling for the time dependence for this process.

The spatial formulation adopted in this paper depends on the definition of the neighborhood structure. We use two definitions of neighborhood. The first is the spatial contiguity, i.e., two regions are considered neighbors if there is a common border. The second formulation considers not only the spatial contiguity but also an air transport network, which allows to incorporate into the model the transmission mechanism given by the contact among passengers and air transport crew between two countries. We detail the structure used in the air transport network in the Supplementary Material.

The prior structure assumes log-gamma distribution for all precision components, with values (1,5e-05) for the trend and seasonal components, and the spatial random effect. For the cycle (AR(2) structure) component we use a penalized complexity prior (SIMPSON et al., 2017) for the precision, with values (3,0.01), and for the first and second order partial autocorrelation parameters we use penalized complexity priors with values (0.5, 0.5) and (0.5, 0.4). The prior for the coefficient Φ is also a penalized complexity prior for the correlation parameter, with

values defined as (3,0.01).

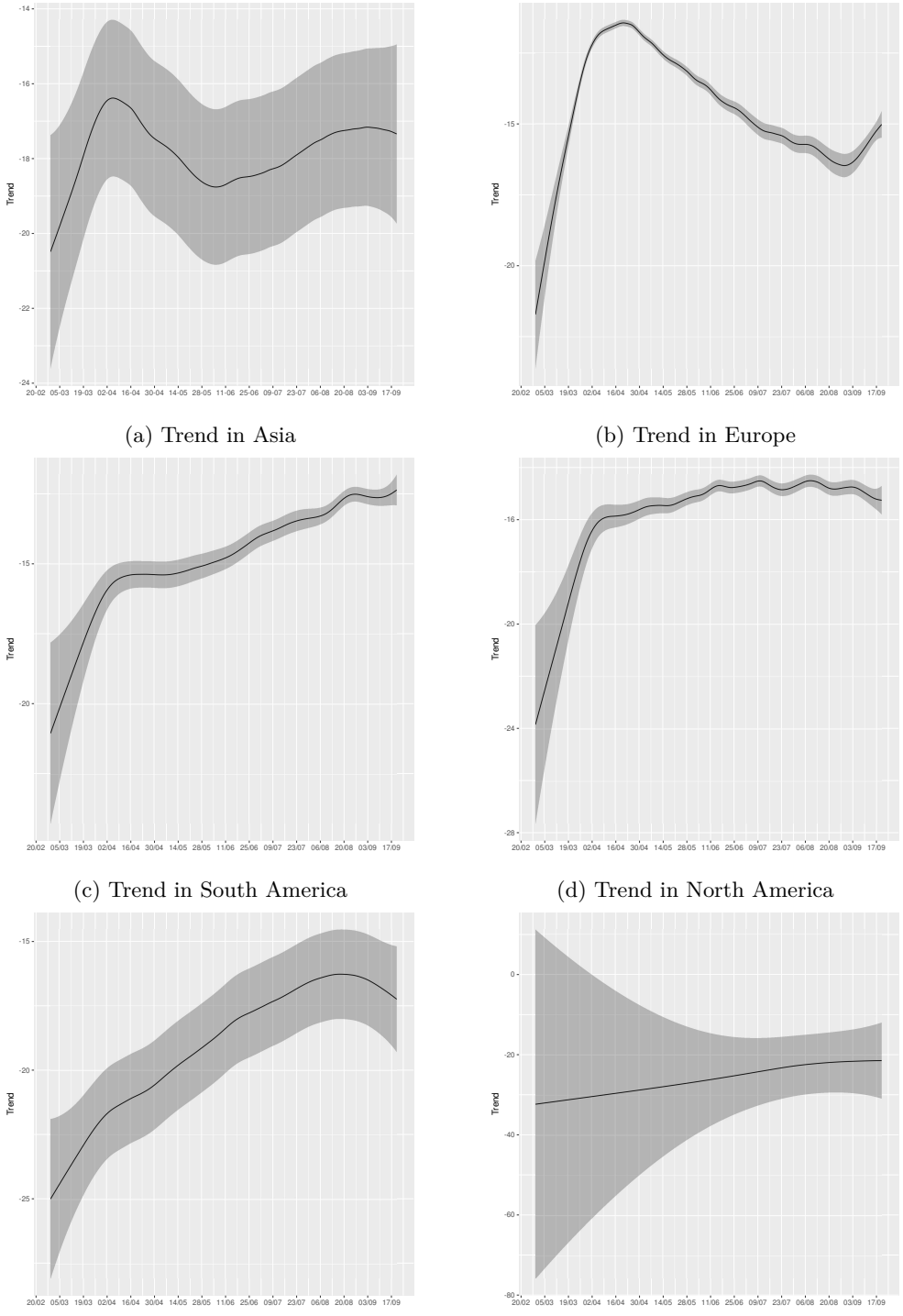
The resulting additive hierarchical structure allows us to perform a Bayesian inference procedure within the Integrated Nested Laplace Approximations (INLA) approach (RUE; MARTINO; CHOPIN, 2009), which provides accurate and efficient approximations on additive hierarchical models that can be represented as Gaussian Markov Random Fields (RUE; HELD, 2005c). For reasons of space we do not detail the INLA method here, which is widely discussed Rue, Martino and Chopin (2009) and Valente and Laurini (2020).

8.3 Results

To estimate the trend of deaths by COVID-19 outbreak in a global scale, we performed inference procedures based on the model described in Equation (8.1), considering three different formulations. First, we estimated the parameters without the effects of the spatial component (model M1, hereafter), where the estimated parameters are the precision of the trend ($1/\eta_{\mu(k,t)}$), seasonal ($1/\eta_{s(k,t)}$) and cycle ($1/\eta_{c(k,t)}$) parameters for Asia, Africa, Europe, North America, South America, and Oceania, and the parameters of the second order autoregressive process of the cycle (PACF1 and PACF2) for each continent. In the second and third formulation (M2 and M3, respectively), we consider a time-varying spatial component, modelled as a spatial random effect with a CAR structure with time dependency, where the estimated values at any given region are conditional of neighboring values. The difference between models M2 and M3 is how the neighborhood structure was defined. While in the model M2, the neighborhood structure was constructed based on contiguity, using the territorial division, in the model M3 not only contiguity was considered to define, but also information of an air transport network. In the case of models M2 and M3, the estimated parameters also include the log precision parameter τ, and parameter Φ of time dependency.

It is worth noting that the precision coefficients are not directly interpretable due to the different scales of the latent components, but a higher precision indicates less variability of the estimated latent component. A higher precision parameter indicates a smoother component or with less temporal variability. In this case, the most direct interpretation is for the estimated path of the latent component. The PACF1 and PACF2 components are the representations of the first and second order autoregressive coefficients in terms of partial correlation coefficients, which have some advantages in terms of computational representation, especially for the imposition of the stationarity of the component that is important in the definition of the cyclic component. The cycle components can be interpreted as the correlation between component in the period t and t_1 (PACF1) and t_2 (PACF2), controlling for the other correlation placed in this specification. We emphasize that the cycle component serves to capture the structures of transient dependence in the series of deaths, which are generated by the component of measurement error or by the composition of other non-permanent patterns in the series of observed deaths.

The results of the estimation of models M1, M2, and M3 are shown in Table 8.1. In order to provide a better interpretation of the results, we plotted the estimated trend (see Figure 8.1), seasonality (see Figure 8.2) and cycle (see Figure 8.3) components for all six considered continents, individually, considering the model M2. For reason of space only the estimated components obtained with model M2 are shown in this paper, and the other results are available



(a) Trend in Asia

(b) Trend in Europe

(c) Trend in South America

(d) Trend in North America

(e) Trend in Africa

(f) Trend in Oceania

Figure 8.1 – Estimated Trends - Spatio-temporal model with region specific trends, seasonal and cycle components (Model M2)

upon request from the authors.

Table 8.1 – Estimated parameters of deaths reported in worldwide – Models M1, M2, and M3

	Mean	SD	0.025quant	0.5quant	0.975quant	Mode

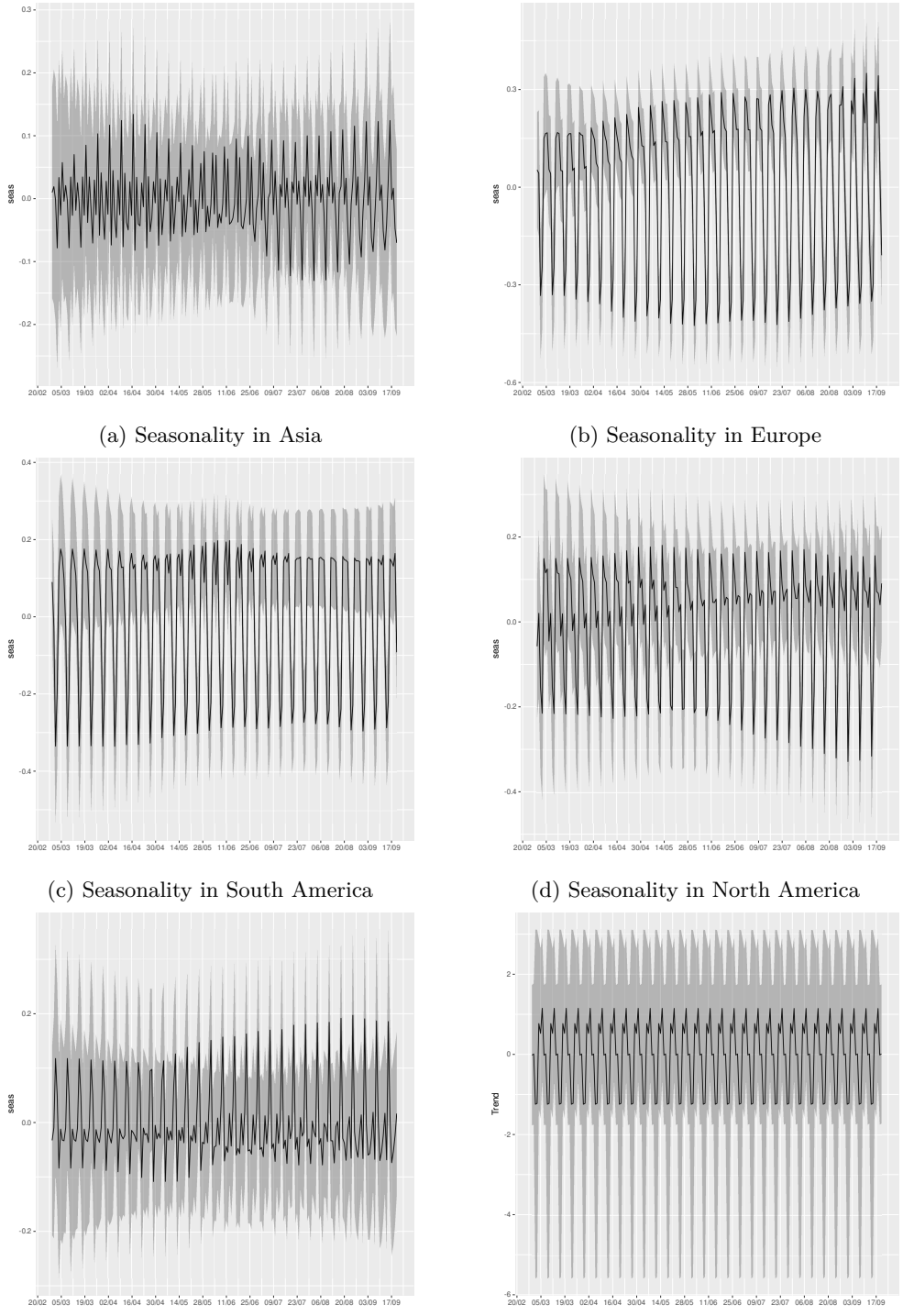
Model without spatial component (M1)						
Precision trend (Asia)	47065.94	1.71×10^4	22255.2	44171.67	8.86×10^4	38965.84
Precision trend (Europe)	18244.38	5.04×10^3	10229.5	17617.95	2.99×10^4	16431.35
Precision trend (Africa)	25340.83	9.38×10^3	11735.8	23766.84	4.80×10^4	20918.04
Precision trend (North America)	8287.84	2.39×10^3	4505.76	7995.11	1.38×10^4	7439.04
Precision trend (South America)	31740.07	1.24×10^4	13512.9	29788.63	6.14×10^4	26043.59
Precision trend (Oceania)	3631.98	1.67×10^3	1285.36	3340.43	7.72×10^3	2758.72
Precision seasonality (Asia)	37041.63	2.47×10^4	8771.08	31026.05	1.01×10^5	21071.69
Precision seasonality (Europe)	1480.08	5.50×10^2	695.52	1383.10	2.83×10^3	1211.25
Precision seasonality (Africa)	14441.12	1.08×10^4	3037.37	11564.34	4.29×10^4	7337.34
Precision seasonality (North America)	1904.25	6.96×10^2	889.36	1789.01	3.58×10^3	1579.57
Precision seasonality (South America)	2935.31	1.37×10^3	1126.92	2651.20	6.40×10^3	2174.77
Precision seasonality (Oceania)	16849.83	1.94×10^4	1802.66	11049.72	6.73×10^4	4780.55
Precision cycle (Asia)	61.37	5.57	51.09	61.14	7.30×10^1	60.71
PACF1 cycle (Asia)	0.04	7.00×10^{-2}	-0.08	0.04	1.84×10^{-1}	0.03
PACF2 cycle (Asia)	-0.00	6.10×10^{-2}	-0.12	-0.00	1.20×10^{-1}	-0.00
Precision cycle (Europe)	74.31	8.67	58.55	73.88	9.26×10^1	73.10
PACF1 cycle (Europe)	0.116	7.60×10^{-2}	-0.03	0.11	2.64×10^{-1}	0.11
PACF2 cycle (Europe)	-0.01	6.80×10^{-2}	-0.14	-0.01	1.20×10^{-1}	-0.00
Precision (Africa)	32.36	3.32	26.23	32.23	3.93×10^1	32.01
PACF1 cycle (Africa)	0.01	7.10×10^{-2}	-0.12	0.01	1.56×10^{-1}	0.01
PACF2 cycle (Africa)	0.04	7.20×10^{-2}	-0.09	0.04	1.92×10^{-1}	0.04
Precision cycle (North America)	43.92	4.60	35.43	43.74	5.35×10^1	43.45
PACF1 cycle (North America)	0.20	6.80×10^{-2}	0.07	0.20	3.37×10^{-1}	0.20
PACF2 cycle (North America)	-0.03	6.70×10^{-2}	-0.16	-0.03	1.02×10^{-1}	-0.03
Precision cycle (South America)	20.57	1.89	17.05	20.51	2.45×10^1	20.41
PACF1 cycle (South America)	-0.17	7.20×10^{-2}	-0.31	-0.17	-3.30×10^{-2}	-0.18
PACF2 cycle (South America)	0.12	6.50×10^{-2}	-0.00	0.12	2.54×10^{-1}	0.12
Precision cycle (Oceania)	6.37	1.75	3.53	6.18	1.04×10^1	5.81
PACF1 cycle (Oceania)	0.10	1.32×10^{-1}	-0.16	0.10	3.55×10^{-1}	0.11
PACF2 cycle (Oceania)	-0.36	1.84×10^{-1}	-0.68	-0.37	3.50×10^{-2}	-0.41
Spatial model with Continent-specific trend, seasonal and cycle components (M2)						
Precision trend (Asia)	1915.71	188.69	1601.40	1894.92	2337.89	1839.90
Precision trend (Europe)	2656.69	333.92	2066.33	2633.71	3375.65	2586.41
Precision trend (Africa)	6258.31	1048.8	4712.68	6076.86	8763.29	5635.08
Precision trend (North America)	2466.70	332.46	1857.12	2453.74	3160.46	2436.04
Precision trend (South America)	4711.42	698.15	3365.21	4716.52	6076.10	4777.84
Precision trend (Oceania)	2331.84	409.06	1748.52	2255.16	3321.72	2066.04
Precision seasonality (Asia)	1903.49	257.95	1460.44	1880.72	2470.37	1831.02
Precision seasonality (Europe)	1215.00	184.48	924.43	1189.27	1641.53	1128.18
Precision seasonality (Africa)	1898.14	265.92	1441.20	1874.79	2482.77	1824.65
Precision seasonality (North America)	2130.87	308.40	1566.84	2118.77	2773.61	2103.00
Precision seasonality (South America)	2463.52	466.22	1803.15	2374.75	3596.30	2162.72
Precision seasonality (Oceania)	2637.86	942.55	1603.34	2379.59	5126.21	1864.64
Precision cycle (Asia)	96.18	14.71	71.99	94.49	129.39	90.73
PACF1 cycle (Asia)	0.17	0.074	0.02	0.18	0.30	0.19

PACF2 cycle (Asia)	0.027	0.07	-0.11	0.02	0.15	0.03
Precision cycle (Europe)	78.77	11.98	57.88	77.87	104.81	76.1
PACF1 cycle (Europe)	0.052	0.09	-0.13	0.05	0.23	0.06
PACF2 cycle (Europe)	-0.23	0.10	-0.43	-0.22	-0.03	-0.21
Precision cycle (Africa)	65.88	9.91	48.99	64.97	87.78	63.03
PACF1 cycle (Africa)	0.024	0.08	-0.14	0.02	0.18	0.03
PACF2 cycle (Africa)	-0.03	0.07	-0.17	-0.02	0.11	-0.02
Precision cycle (North America)	57.17	9.50	41.70	56.03	78.82	53.53
PACF1 cycle (North America)	0.08	0.08	-0.09	0.09	0.22	0.11
PACF2 cycle (North America)	0.14	0.08	-0.00	0.13	0.33	0.10
Precision cycle (South America)	47.60	9.04	32.82	46.56	68.15	44.41
PACF1 cycle (South America)	-0.09	0.07	-0.24	-0.08	0.05	-0.07
PACF2 cycle (South America)	0.34	0.06	0.2	0.34	0.46	0.35
Precision cycle (Oceania)	45.55	6.95	32.58	45.40	59.72	45.40
PACF1 cycle (Oceania)	0.22	0.08	0.07	0.21	0.40	0.17
PACF2 cycle (Oceania)	0.00	0.08	-0.13	-0.00	0.17	-0.03
log τ	0.06	0.003	0.062	0.067	0.072	0.06
Group ϕ	0.96	0.002	0.957	0.96	0.963	0.96
Spatial model with Continent-specific trend, seasonal and cycle components - Air transport network (M3)						
Precision trend (Asia)	4225.73	229.88	3747.12	4240.22	4640.09	4301.89
Precision trend (Europe)	2464.55	112.49	2249.92	2462.48	2691.35	2459.06
Precision trend (Africa)	5411.90	318.02	4733.93	5439.88	5957.38	5559.36
Precision trend (North America)	2287.83	110.09	2067.37	2290.43	2498.60	2303.33
Precision trend (South America)	4521.49	309.20	3851.89	4554.24	5031.37	4697.28
Precision trend (Oceania)	2484.95	119.96	2271.99	2476.21	2741.47	2450.54
Precision seasonality (Asia)	1963.97	93.72	1769.15	1969.62	2133.86	1993.57
Precision seasonality (Europe)	1356.31	63.77	1235.08	1355.02	1485.31	1352.74
Precision seasonality (Africa)	2401.57	124.52	2191.46	2388.59	2676.46	2348.16
Precision seasonality (North America)	2462.17	135.86	2173.94	2473.27	2698.28	2520.30
Precision seasonality (South America)	2238.56	110.55	2043.07	2230.27	2475.86	2205.89
Precision seasonality (Oceania)	1781.24	118.39	1526.03	1793.13	1978.95	1844.69
Precision cycle (Asia)	136.51	5.60	126.18	136.25	148.16	135.54
PACF1 cycle (Asia)	0.20	0.02	0.16	0.20	0.25	0.20
PACF2 cycle (Asia)	0.07	0.02	0.03	0.07	0.12	0.07
Precision cycle (Europe)	124.94	5.53	114.67	124.71	136.40	124.11
PACF1 cycle (Europe)	0.01	0.02	-0.03	0.01	0.05	0.01
PACF2 cycle (Europe)	-0.06	0.02	-0.11	-0.06	-0.02	-0.06
Precision cycle (Africa)	104.27	4.74	94.81	104.36	113.40	104.81
PACF1 cycle (Africa)	0.09	0.02	0.04	0.09	0.14	0.08
PACF2 cycle (Africa)	0.03	0.02	-0.01	0.03	0.08	0.02
Precision cycle (North America)	95.80	4.49	86.84	95.89	104.45	96.36
PACF1 cycle (North America)	0.30	0.02	0.24	0.30	0.34	0.31
PACF2 cycle (North America)	0.02	0.02	-0.02	0.02	0.06	0.03
Precision cycle (South America)	73.15	3.72	65.43	73.35	79.95	74.24
PACF1 cycle (South America)	0.12	0.02	0.08	0.12	0.17	0.12
PACF2 cycle (South America)	0.18	0.02	0.14	0.18	0.24	0.18
Precision cycle (Oceania)	48.14	2.44	43.90	47.92	53.47	47.27

PACF1 cycle (Oceania)	0.19	0.02	0.14	0.19	0.23	0.18
PACF2 cycle (Oceania)	-0.06	0.02	-0.11	-0.06	-0.02	-0.06
log τ	0.01	0.00	0.01	0.01	0.01	0.01
Group ϕ	0.96	0.00	0.95	0.96	0.96	0.96

To provide goodness-of-fit measures to compare how well the models fit the data, we calculate the fit statistics of all the models, such as the Mean Error (ME), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE), reported in Table 8.2. These measures serve as general parameters of model adequacy. The mean error is important to verify the presence of relevant biases in the model's predictions. In the presence of high mean errors, the model may be underestimating or overestimating the number of deaths observed. The root mean squared errors (RMSE) and mean absolute errors (MAE) are measures of dispersion of the model's predictions in relation to the observed cases, and serve as a measure of the model's accuracy, using different weighting metrics (squared error for the RMSE and absolute error in the MAE). Quadratic loss is a traditional measure of predictive accuracy, linked for example to estimation by ordinary least squares, penalizing squared error and avoiding the cancellation of errors with distinct signs. The MAE has a similar interpretation, since absolute errors also avoid cancellation of positive and negative errors but has the advantage of being less sensitive to some deviant observation. In general, the results indicate an adequate adjustment for the models estimated in the article, such as almost zero mean errors indicating the absence of predictive bias, and very low RMSE and MAE measures indicating great precision in fitting the observed deaths.

In summary, we present the results of analyzing data for COVID-19 in 205 countries from February 29th, 2020, to February 17th, 2021, divided according to the following six continents: Asia, Europe, North America, South America, Africa, and Oceania. It is worth noting that we did not consider the Antarctica continent since the continent was only affected by the coronavirus pandemic in December 2020. The results show that our model was able to capture the long-term movements in cases that resulted in death cases for the six continents, showing an accelerated growth pattern until mid-April in Asia, and a second wave driven by the spread of the disease in the Middle East, South Asia, and Southeast Asia. The estimated trend component for Europe shows that the continent has experienced the peak of the first wave at the end of April, and a new accelerated growth phase at the end of the analyzed sample, mostly due to the relaxation of the isolation measures taken by the governments. For North and South America, the trend shows that the former presents a sustainable high peak plateau since mid-May, whereas the latter still shows a pattern of rising. For Africa, the results show that the acceleration phase is over after a relatively low number of reported deaths, despite the experts' prediction of millions of COVID-19 deaths in the continent. For Oceania, the results show that the number of deaths by COVID-19 was small and well-controlled by the measures taken by the local governments. Additionally, our results were able to show the importance of the use of spatial information in the task of estimating the trend and mean-reverting components, which are related to the spatial transmission dynamics in an epidemic process.



(a) Seasonality in Asia

(b) Seasonality in Europe

(c) Seasonality in South America

(d) Seasonality in North America

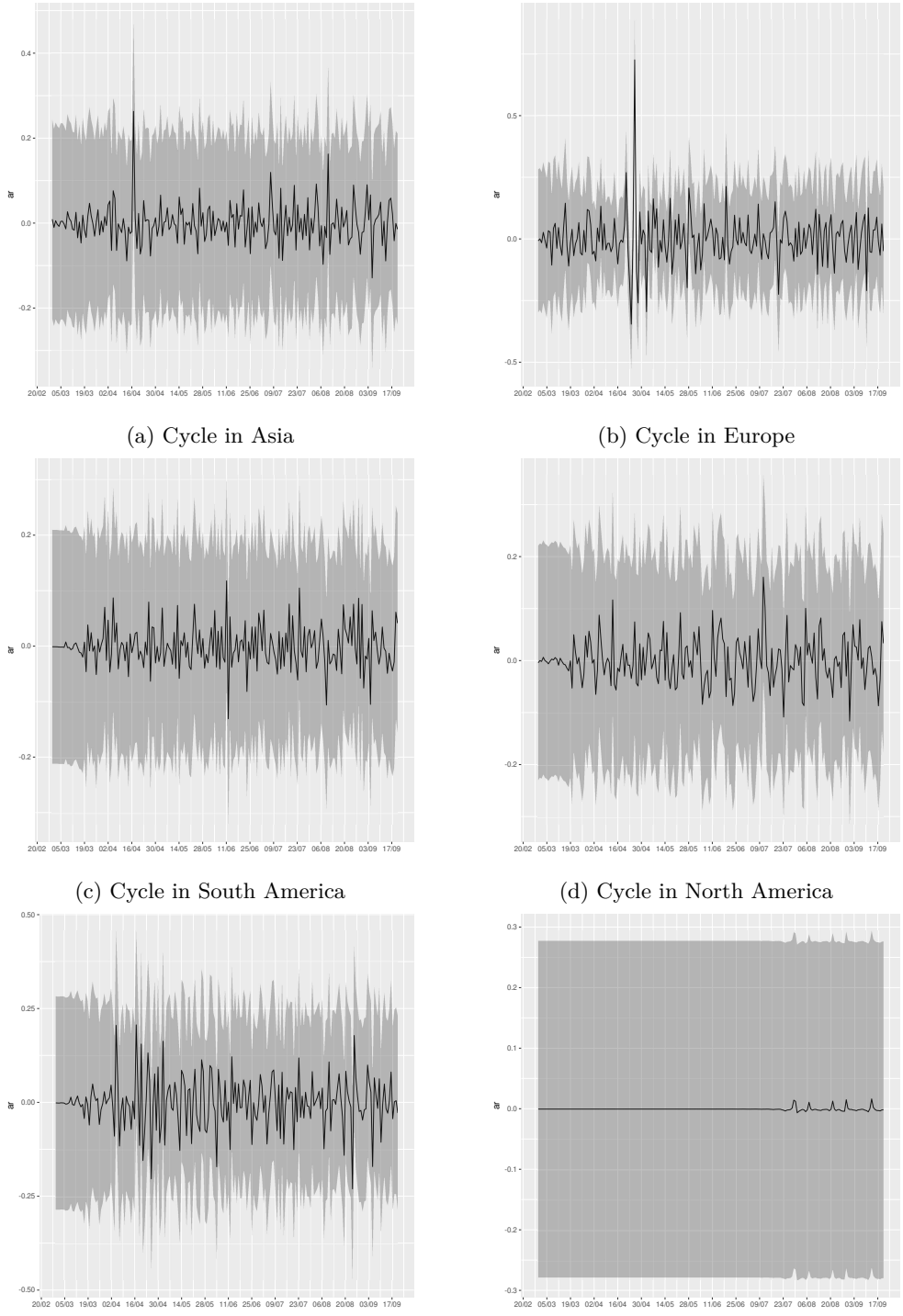
(e) Seasonality in Africa

(f) Seasonality in Oceania

Figure 8.2 – Estimated Seasonality - Spatio-temporal model with region specific trend, seasonal and cycle components (Model M2)

8.4 Discussion

The first epicenter of the disease was in East Asia, mostly in China and South Korea, which reached the epidemic peak around mid-April. The observed second peak of the COVID-19 in Asia was mostly due to the spread of the disease in the Middle East, South Asia and South East Asia. As of December 2020, deaths by COVID-19 have risen in Iraq, Indonesia, and India, which



(a) Cycle in Asia

(b) Cycle in Europe

(c) Cycle in South America

(d) Cycle in North America

(e) Cycle in Africa

(f) Cycle in Oceania

Figure 8.3 – Estimated Cycle - Spatio-temporal model with region specific trend, seasonal and cycle components (Model M2)

the latter has the world's third-most deaths per 100,000 people and are driving the numbers in Asia. After reaching the peak of the second wave in December, the number of deaths by COVID-19 in the continent has slowly decreased.

The second outbreak epicenter was in Western Europe, especially in Spain, Italy, France and the United Kingdom. The continent has experienced the peak of the first wave in the end

Table 8.2 – Model Fit Statistics

	ME	RMSE	MAE
Model without spatial component (model M1)			
World	-6.7495e-07	90.4915	28.1972
Asia	-3.1729e-08	86.3027	29.1699
Europe	-6.8675e-07	97.3253	34.6779
South America	-1.3301e-06	156.1249	55.9166
North America	-3.2541e-06	126.8047	47.4272
Africa	-9.1082e-08	32.2776	8.7439
Oceania	-6.9796e-08	1.6949	0.6356
Spatio-Temporal Model (model M2)			
World	-2.0870e-08	1.9503	0.8978
Asia	-2.1111e-07	1.3780	0.6438
Europe	-6.1200e-07	2.3485	1.2610
South America	-3.5786e-06	4.0655	2.0413
North America	4.8557e-06	1.2455	0.6629
Africa	-2.9075e-07	1.3234	0.6380
Oceania	-9.7523e-08	0.07023	0.01209
Spatio-Temporal Model - Air transport network (model M3)			
World	-1.8349e-06	1.8429	0.8790
Asia	-1.2525e-06	1.5854	0.7988
Europe	-1.8629e-06	2.7506	1.4923
South America	-3.8558e-06	2.4099	1.2811
North America	-6.1345e-06	0.9105	0.4961
Africa	-2.5206e-07	0.9226	0.4686
Oceania	-7.8468e-08	0.4489	0.1707

of April. However, after consistent drops in the number of daily recorded deaths, the trend in Europe have presented a pattern of rising since the beginning of September, mostly due to the loosening of lockdown and social distance measures. Due to the increased number of deaths most European countries induced new restriction measures to control the new peak in deaths. As of the end of February, daily deaths by COVID-19 has been failing in most European countries, but in some others, like France, Italy, and the Czech Republic, highest numbers of deaths have been recorded in February, helping to maintain the European trend in a high level. The last epicenter occurred in South and North America, mostly driven by Brazil and the United States, which have, up to October 5th, the second and first highest numbers of recorded deaths by 100,000 people in the world, respectively. While North America presents a sustainable high peak plateau since around mid-May, the South America still shows a pattern of rising. Between October and December 2020, the trend for North America showed a slight decrease, however, returning to the previous level in January 2021, when the United States has recorded the highest levels of daily deaths, and Canada has experienced a winter surge. In the of the analyzed sample, the trend component for North America gives signals that the component has fallen, which may be associated with the advancement of vaccination campaigns. The estimated trend component for South America reached the peak in early January, and gives signals of decreasing at low rates. In this continent, there is a particular concern with Brazil, which has recorded more than 265,000 deaths, the world's second highest number of death by COVID-19, and due to the circulation

of the new Brazilian variant of the virus (known as P.1 or VOC202101/02).

It is possible to note that the initial acceleration phases of the Africa and Oceania COVID-19 outbreak grew at a lower rate than those observed in Asia, Europe and Americas. Africa and the most affected by COVID-19 countries have significant differences in reliable reporting, demography, and sociocultural aspects. However, despite the experts' prediction of millions of COVID-19 deaths in Africa (MARTINEZ-ALVAREZ et al., 2020), the prevalence and mortality are low after two peaks of the disease, in mid-August 2020, and December 2020. The reason for this outcome is still unclear, however, it may be related to the limited number of tests available and the lack of high-quality data (MBOW et al., 2020). However, it is worth noting that in the end of the analyzed sample, the trend component shows a slight growth pattern. The Oceania has a natural geographical advantage, associated with effective measures, making the cumulative number of cases and deaths small and controllable (BECK; HENSHER; WEI, 2020). In particular, Australia and New Zealand have been praised for the effective lockdown measures taken in both countries, and their relative low number of deaths by COVID-19. Additionally, the estimated seasonal components (see Figure 8.2) are significantly different among continents, reflecting different patterns of the spreading of the COVID-19, and the government's strategy to reduce the transmission of the disease, which varies from country to country. In particular, it is possible to see the estimated seasonality components for Asia, Europe, South America, and Africa with an increasing amplitude over time, while in North America the increasing amplitude in the seasonality component was followed by a decrease in the end of the analyzed sample. For the Oceania, the estimated seasonality component is roughly constant over time. Regarding the cycle components, for the Oceania, it was significantly less noisy than the estimated components for the other continents.

It is worth noting that the latent components of our proposed model are formulated in terms of the exposure offset for the number of populations observed in each region, allowing the counts of COVID-19 death to be comparable. However, it is not an easy task to directly interpret the estimated trend component. Therefore, to provide better understandable results, we carried out a transformation of the estimated trend component towards the number of daily deaths per million inhabitants per region (see Figure 8.4; Supplemental Material), where it is possible to note that Europe reached the peak faster than the other continents until the end of April 2020, where the death trend reveals a turning point to a decreasing pattern, followed by a new acceleration phase in September. Also, it is possible to see that the trend in South America reached the relative highest peak in early January 2021, followed by a decreasing pattern. Despite the Asia has been the first epicenter of the COVID-19 outbreaks in early 2020, Europe and South America have had more deadly COVID-19 outbreaks. For the North America continent, the trend remains stable with a slight decrease in the end of the analyzed sample, whereas for the Asia, Oceania and Africa not too much details can be discerned, given the scale of the graph.

Regarding on the fit of the three different formulations, it is possible to observe that M2 and M3 models presented a similar adjustment result, with few gains in the use of air transport information in defining the model's neighborhood structure. Also, it is important to note that the similarity in model fit of models M2 and M3 can be related to massive travel restrictions and border control to mitigate the outbreak of the COVID-19 pandemic. As discussed in the

literature, a larger reduction in the passenger air flow led to a gradual decrease the spread of COVID-19 (LINKA et al., 2020), which makes this transmission channel less important. In addition, it is worth noting that both models with spatial component (M2 and M3) are clearly superior to the univariate model (M1) in terms of RMSE and MAE, showing the importance of spatial transmission patterns in the COVID-19 dynamics.

Within the limitations of this study, there are some concerns on the reliability of the data used. It is well known that the reporting number of the COVID-19 cases and deaths do not reflect the reality, varying these underreporting rates from country to country, depending on the number of available tests and the measures taken by the authorities. In particular, the number of available tests has changed over time and space. Despite the limitations, this study provides evidence of the effectiveness of public health measures taken and vaccines and may be used as a tool to guide the authorities to propose new restrictions to control the transmission and to assess the effects of the restrictions loosening. Previous methods to estimate the trend of COVID-19 have been useful in this context, e.g., Yang et al. (2020) which found evidence of the effectiveness of the control measures on January 2020 to reduce the COVID-19 epidemic size in China, and Post et al. (2020) which developed dynamic metrics to inform public health surveillance and to identify where and when corrective measures are necessary to cope with the spread of the COVID-19. Based on our outlined spatio-temporal analysis methodology, it is possible to estimate the trends related to COVID-19 in a more reliable way since the proposed model can control for the measurement errors induced by the case accounting processes, making it a useful tool to design strategy for preventing the spread of the disease, targeting vaccines, and to for the planning of measures of economic reopening.

8.5 Supplementary Material

8.5.1 Transformation of the estimated trend component

8.5.2 Air Transport Network

To build a definition of spatial neighborhood that controls for possible transmission routes linked to air transport, we carry out a network analysis using the air transport routes provided by the OpenSky network project (<https://opensky-network.org/>), compiled in the project *Crowdsourced air traffic data from The OpenSky Network 2020* (<https://zenodo.org/record/4088202>). This dataset contains the cleaned information on all transport routes observed by the OpenSky project since January 2019. Table 8.3 shows the list of the current officially assigned ISO 3166-1 alpha-3 codes for the 205 countries considered in our analysis, using the English short country names officially used by the ISO 3166 Maintenance Agency. We used the air transport network of April 2020, which reflects the average conditions of the air network after the pandemic was decreed.

To construct the air transport network, we used the concept of hierarchical clustering used in data analysis (KOLACZYK; CSÁRDI, 2014), and the fast greedy algorithm (CLAUSET; NEWMAN; MOORE, 2004). In hierarchical methods, the candidate partition is modified in a way that minimizes a specified measure of weighted cost. In particular, we considered herein the modularity as a cost measure, which is used for detecting community structure in networks. Let

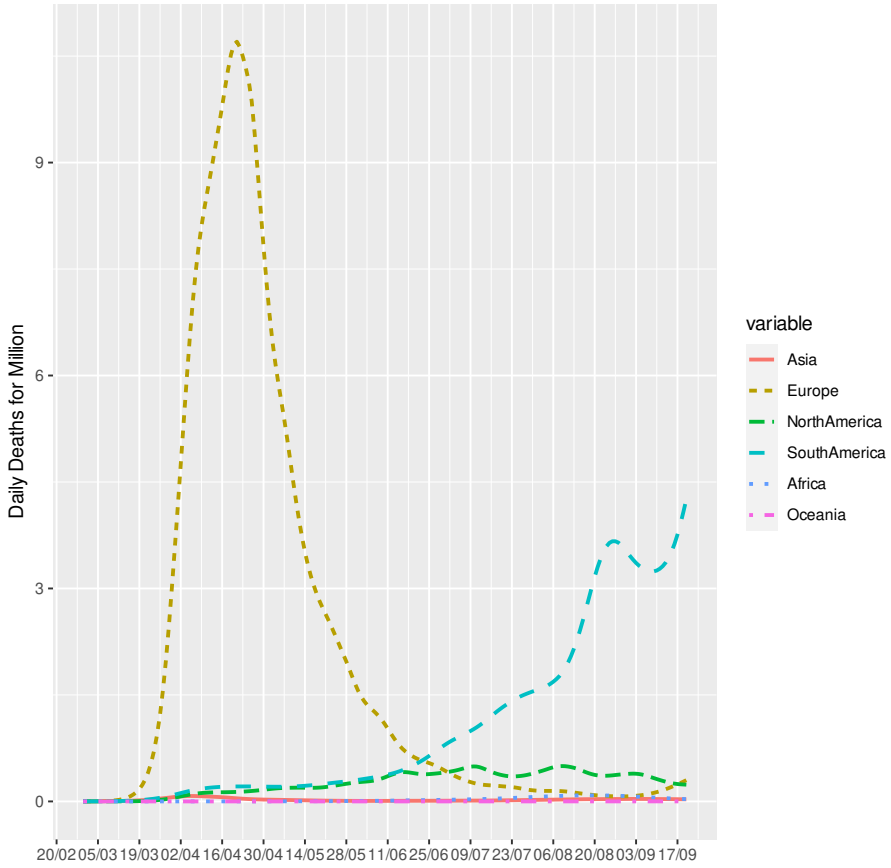


Figure 8.4 – Trends in Deaths by day by Million inhabitants

$\mathcal{C} = \{C_1, \ldots, C_K\}$ be a given candidate partition and define $f_{ij} = f_{ij}(\mathcal{C}$ to be a fraction of edges in the original network that connect vertices in C_i with vertices in C_j, thus, the modularity of $\mathcal{C}$ is defined as

$$\text{mod } (\mathcal{C}) = \sum_{k=1}^{K} [f_{kk}(\mathcal{C}) - f_{kk}^{*}]^2, \tag{8.2}$$

where f_{kk}^{*} is the expected value of f_{kk} under some model of random edge assignment. We use weighted cost measure, using the number of flights between countries as the weight function in the analysis. Table 8.4 show the eight selected communities found by the aforementioned method and used in this work. Also, the visual representation of the partitioning of the air transport network can be seen in Figure 8.5, with countries in the same color representing the same community based on air transport routes. The final result of this analysis is used in the construction of a generalized version of the spatial neighborhood matrix. Thus two countries are considered neighbors if they have a common border or if they are in the same community derived from the analysis of hierarchical clustering in a network.

AFG	ALB	DZA	AND	AGO	AIA	ATG	ARG	ARM	ABW	AUS	AUT
AZE	BHS	BHR	BGD	BRB	BLR	BEL	BLZ	BEN	BMU	BTN	BOL
BIH	BWA	BRA	VGB	BRN	BGR	BFA	BDI	KHM	CMR	CAN	CPV
CYM	CAF	TCD	CHL	CHN	COL	COM	COG	CRI	CIV	HRV	CUB
CYP	CZE	COD	DNK	DJI	DMA	DOM	ECU	EGY	SLV	GNQ	ERI
EST	ETH	FRO	FLK	FJI	FIN	FRA	PYF	GAB	GMB	GEO	DEU
GHA	GIB	GRC	GRL	GRD	GUM	GTM	GGY	GIN	GNB	GUY	HTI
HND	HKG	HUN	ISL	IND	IDN	IRN	IRQ	IRL	IMN	ISR	ITA
JAM	JPN	JEY	JOR	KAZ	KEN	KWT	KGZ	LAO	LVA	LBN	LSO
LBR	LBY	LIE	LTU	LUX	MKD	MDG	MWI	MYS	MDV	MLI	MLT
MRT	MUS	MEX	MDA	MCO	MNG	MNE	MSR	MAR	MOZ	MMR	NAM
NPL	NLD	NCL	NZL	NIC	NER	NGA	MNP	NOR	OMN	PAK	PSE
PAN	PNG	PRY	PER	PHL	POL	PRT	PRI	QAT	ROU	RUS	RWA
KNA	LCA	VCT	SMR	STP	SAU	SEN	SRB	SYC	SLE	SGP	SVK
SVN	SOM	ZAF	KOR	ESP	LKA	SDN	SUR	SWZ	SWE	CHE	SYR
TWN	TJK	TZA	THA	TLS	TGO	TTO	TUN	TUR	TCA	UGA	UKR
ARE	GBR	USA	VIR	URY	UZB	VAT	VEN	VNM	ESH	YEM	ZMB
ZWE											

Table 8.3 – Country Codes

1	GBR	IRL	SWE	ARE	ITA	ROU	SEN	PRT	DNK	GNB	PRY	UKR
	HUN	LTU	SRB	LVA	SVK	BGR	MKD	MDA	SVN	BIH	EST	HRV
	ALB	MNE	GMB	MRT	MCO	LBN	ISR	EGY	CYP	SYR	JOR	
2	FRA	POL	IND	FIN	NOR	QAT	GRC	KWT				
3	NLD	ZAF	AUT	MAR	TUN	BWA	LSO					
4	JPN	ESP	CAN	THA	HKG	RUS	SAU	PHL	CZE	BRN	LUX	ISL
	KOR	BMU	BHR	UZB	LKA	NZL	MMR	OMN	IMN	GEO	ARM	VNM
5	ARG	USA	MEX	BOL	BEL	URY	DZA	BRB	CHL	PER	GIN	BHS
	CPV	JEY	GUY	GGY	VEN	LCA	VCT	ECU	COL			
6	AUS	DEU	TWN	IDN	PYF	SGP	PSE	AZE				
7	BRA	TUR	PAK	IRN	TJK	KGZ	BLR	MLT	TTO	LBY	KHM	
8	CHN	CHE	MYS	KAZ	NPL	LAO						

Table 8.4 – Network Analysis - Selected Communities

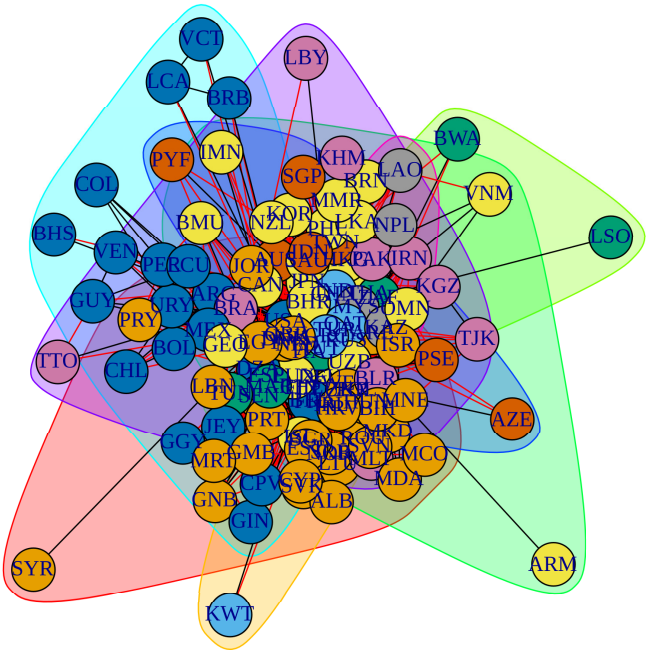


Figure 8.5 – Communities

CONCLUSIONS

This thesis comprises a collection of eight self-contained papers in the field of spatial statistics, with applications in diverse contexts, ranging from climate events to deaths by COVID-19. The first five papers focus on the application of structural decomposition for spatio-temporal analysis of climate-related events, such as fires and tornadoes. These studies yielded valuable insights, including the confirmation of constant trends in tornado occurrences in the United States and the effectiveness of regulatory changes in reducing pre-harvest sugarcane burning in the São Paulo state of Brazil. Additionally, they shed light on the relationship between climate factors and fire patterns in Australia and the Legal Amazon, as well as variations in fire occurrence trends in the Brazilian Pantanal.

Moving beyond climate-related events, the sixth paper extended the decomposition approach to analyze changes in rainfall and temperature patterns in the metropolitan area of São Paulo. While omitting the spatial dimension, this paper introduced the same decomposition structure to examine various climatic aspects, revealing increases in daily temperature trends and rainfall patterns. Furthermore, the examination of annual temperature records and annual records of daily rainfall provides additional support for the hypothesis of long term alterations in observed climatic patterns.

The seventh paper addressed the COVID-19 pandemic by proposing a method to estimate trends in death cases while accounting for measurement error. This research applied time series decomposition analysis to examine COVID-19-related deaths in Brazil, São Paulo, and Amazonas, and spatio-temporal analysis for occurrences of deaths by COVID-19 at the state level in Brazil. This approach provided valuable insights into the pandemic's dynamics at different regions.

Finally, the eighth paper extended the previous analysis to a global scale, offering a compelling demonstration of the importance of spatio-temporal models in accurately estimating COVID-19-related deaths. These models outperformed univariate models, highlighting the crucial role of spatial dynamics in understanding the pandemic's progression.

In summary, this thesis contributes significantly to the field of spatial-temporal statistics by introducing a modeling framework that incorporates the extraction of trend, cycle, and seasonality components through a structural decomposition structure, within the context of models for spatial point pattern data, extreme values, probaility of rain, and the duration of dry days. The findings presented in these papers enhance our understanding of long term and mean-reverting components in the context of climate-related events, climatic changes, and the dynamics of the COVID-19 pandemic, taking into accouting the spatial aspect.

REFERENCES

ABADIE, Alberto; DIAMOND, Alexis; HAINMUELLER, Jens. Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, Taylor & Francis, v. 105, n. 490, p. 493–505, 2010.

ABADIE, Alberto; GARDEAZABAL, Javier. The economic costs of conflict: A case study of the Basque Country. *American Economic Review*, v. 93, n. 1, p. 113–132, 2003.

ABATZOGLOU, John T; WILLIAMS, A Park. Impact of anthropogenic climate change on wildfire across western US forests. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 113, n. 42, p. 11770–11775, 2016.

ABESSA, Denis; FAMÁ, Ana; BURUAEM, Lucas. The systematic dismantling of brazilian environmental laws risks losses on all fronts. *Nature Ecology & Evolution*, Nature Publishing Group, v. 3, n. 4, p. 510–511, 2019.

AGEE, Ernest; CHILDS, Samuel. Adjustments in tornado counts, f-scale intensity, and path width for assessing significant tornado destruction. *Journal of Applied Meteorology and Climatology*, v. 53, n. 6, p. 1494–1505, 2014.

AGUIAR, Daniel Alves et al. Remote sensing images in support of environmental protocol: Monitoring the sugarcane harvest in são paulo state, brazil. *Remote Sensing*, Molecular Diversity Preservation International, v. 3, n. 12, p. 2682–2703, 2011.

ALDERSLEY, Andrew; MURRAY, Steven J; CORNELL, Sarah E. Global and regional analysis of climate and human drivers of wildfire. *Science of the Total Environment*, Elsevier, v. 409, n. 18, p. 3472–3481, 2011.

ALEXANDER, Lisa V et al. Trends in Australia's climate means and extremes: a global context. *Australian Meteorological Magazine*, v. 56, n. 1, p. 1–18, 2007.

ALEXANDROV, Theodore et al. A review of some modern approaches to the problem of trend extraction. *Econometric Reviews*, Taylor & Francis, v. 31, n. 6, p. 593–624, 2012.

ALLEN, Myles R et al. *Climate Change 2014 Synthesis Report*. [S.l.], 2014.

ALVARES, Clayton Alcarde et al. Köppen's climate classification map for brazil. *Meteorologische Zeitschrift*, E. Schweizerbart'sche Verlagsbuchhandlung, v. 22, n. 6, p. 711–728, 2013.

ANDELA, Niels et al. A human-driven decline in global burned area. *Science*, American Association for the Advancement of Science, v. 356, n. 6345, p. 1356–1362, 2017.

ARAGÃO, Luiz Eduardo OC et al. Spatial patterns and fire response of recent Amazonian droughts. *Geophysical Research Letters*, Wiley Online Library, v. 34, n. 7, 2007.

ARRAES, Christiano Luna et al. Temporal analysis of the reduction in gas emission in areas of mechanically-harvested sugarcane using satellite imagery. *International Journal of Agriculture and Natural Resources*, v. 37, n. 1, p. 113–121, 2010.

ARRUDA, Wellinton de Sa et al. Inundation and fire shape the structure of riparian forests in the Pantanal, Brazil. *PloS One*, Public Library of Science San Francisco, CA USA, v. 11, n. 6, p. e0156825, 2016.

ARVOR, Damien et al. Analyzing the agricultural transition in Mato Grosso, Brazil, using satellite-derived indices. *Applied Geography*, Elsevier, v. 32, n. 2, p. 702–713, 2012.

ASSUNÇÃO, Juliano et al. Deforestation slowdown in the Brazilian Amazon: prices or policies. *Environment and Development Economics*, Cambridge University Press, v. 20, n. 6, p. 697–722, 2015.

ASSUNÇÃOA, Juliano et al. Does credit affect deforestation? evidence from a rural credit policy in the Brazilian Amazon. *Climate Policy Initiative, Rio de Janeiro, Brasil*, 2013.

BACCARIN, José Giacomo; OLIVEIRA, Jonatan Alexandre de; MARDEGAN, Gláucia Elisa. The environmental, social and economic effects of recent technological changes in sugarcane on the state of São Paulo, Brazil. *Journal of Agrarian Change*, Wiley Online Library, v. 20, n. 4, p. 598–617, 2020.

BADDELEY, A.; RUBAK, E.; TURNER, R. *Spatial Point Patterns: Methodology and Applications with R*. [S.l.]: Chapman and Hall/CRC Press, 2015.

BAKKA, H. et al. The diffusion-based extension of the Matérn field to space-time. 2020. Available at: <arXiv:2006.04917v1>.

BAKKA, Haakon et al. Spatial modeling with R-INLA: A review. *WIREs Computational Statistics*, v. 10, n. 6, p. e1443, 2018.

BECK, Matthew J; HENSHER, David A; WEI, Edward. Slowly coming out of covid-19 restrictions in australia: Implications for working from home and commuting trips by car and public transport. *Journal of Transport Geography*, Elsevier, v. 88, p. 102846, 2020.

BESAG, J. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, Taylor & Francis, v. 36, n. 2, p. 192–236, 1974.

BESAG, J. et al. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, n. 43, p. 1–59, 1991.

BLAIN, G. C. The Mann-Kendall test: the need to consider the interaction between serial correlation and trend. *Acta Scientiarum. Agronomy*, v. 35, n. 4, p. 393–402, 2013.

BLANGIARDO, M.; CAMELETTI, M. *Spatial and Spatio-Temporal Models with R-INLA*. [S.l.]: Wiley, 2015.

BLOOMFIELD, Peter. Trends in global temperature. *Climatic change*, Springer, v. 21, n. 1, p. 1–16, 1992.

BRADSTOCK, Ross A. A biogeographic model of fire regimes in Australia: current and future implications. *Global Ecology and Biogeography*, Wiley Online Library, v. 19, n. 2, p. 145–158, 2010.

Brazilian Sugarcane Industry Association. *UNICA Reports*. [S.l.]: São Paulo, Brazil, 2018.

BRODERSEN, Kay H et al. Inferring causal impact using Bayesian structural time-series models. *The Annals of Applied Statistics*, Institute of Mathematical Statistics, v. 9, n. 1, p. 247–274, 2015.

BUSS, Lewis F. et al. Three-quarters attack rate of SARS-CoV-2 in the Brazilian Amazon during a largely unmitigated epidemic. *Science*, v. 371, n. 6526, p. 288–292, 2020.

CALDARELLI, Carlos Eduardo; GILIO, Leandro. Expansion of the sugarcane industry and its effects on land use in São Paulo: Analysis from 2000 through 2015. *Land Use Policy*, Elsevier, v. 76, p. 264–274, 2018.

CAMELETTI, Michela et al. Spatio-temporal modeling of particulate matter concentration through the spde approach. *AStA Advances in Statistical Analysis*, Springer, v. 97, n. 2, p. 109–131, 2013.

CANABARRO, Askery et al. Data-driven study of the covid-19 pandemic via age-structured modelling and prediction of the health system failure in Brazil amid diverse intervention strategies. *PLOS ONE*, Public Library of Science, v. 15, p. 1–13, 07 2020. Available at: <https://doi.org/10.1371/journal.pone.0236310>.

CANÇADO, José ED et al. The impact of sugar cane–burning emissions on the respiratory system of children and the elderly. *Environmental health perspectives*, National Institute of Environmental Health Sciences, v. 114, n. 5, p. 725–729, 2006.

CEYLAN, Zeynep. Estimation of covid-19 prevalence in Italy, Spain, and France. *Science of The Total Environment*, v. 729, p. 138817, 2020. ISSN 0048-9697.

CHEN, Xiongwen. Spatial pattern of wildfire occurrences in Alabama, USA. *International Journal of Environmental Studies*, Taylor & Francis, v. 64, n. 2, p. 229–242, 2007.

CLARKE, Hamish; LUCAS, Christopher; SMITH, Peter. Changes in Australian fire weather between 1973 and 2010. *International Journal of Climatology*, Wiley Online Library, v. 33, n. 4, p. 931–944, 2013.

CLAUSET, A.; NEWMAN, M.; MOORE, C. Finding community structure in very large networks. *Physical Review E*, v. 70, n. 6, p. 66111, 2004.

COCHRANE, Mark A et al. Positive feedbacks in the fire dynamic of closed canopy tropical forests. *Science*, American Association for the Advancement of Science, v. 284, n. 5421, p. 1832–1835, 1999.

COCHRANE, Mark A; BARBER, Christopher P. Climate change, human land use and future fires in the Amazon. *Global Change Biology*, Wiley Online Library, v. 15, n. 3, p. 601–612, 2009.

COCHRANE, Mark A; SCHULZE, Mark D. Fire as a recurrent event in tropical forests of the eastern Amazon: Effects on forest structure, biomass, and species composition 1. *Biotropica*, Wiley Online Library, v. 31, n. 1, p. 2–16, 1999.

COLES, Stuart et al. *An introduction to statistical modeling of extreme values*. [S.l.]: Springer, 2001.

CONTRERAS, Sebastián et al. Statistically-based methodology for revealing real contagion trends and correcting delay-induced errors in the assessment of covid-19 pandemic. *Chaos, Solitons & Fractals*, v. 139, p. 110087, 2020.

COOK, Ashton Robinson et al. The impact of El Niño–southern oscillation (ENSO) on winter and early spring US tornado outbreaks. *Journal of Applied Meteorology and Climatology*, v. 56, n. 9, p. 2455–2478, 2017.

COSTA, José Augusto Fontoura et al. *Energy Law and Regulation in Brazil*. [S.l.]: Springer, 2018.

COSTA, Marcos H et al. Climate change in Amazonia caused by soybean cropland expansion, as compared to caused by pastureland expansion. *Geophysical Research Letters*, Wiley Online Library, v. 34, n. 7, 2007.

COSTANZA, Robert et al. The value of the world's ecosystem services and natural capital. *Nature*, Nature Publishing Group, v. 387, n. 6630, p. 253–260, 1997.

_____. Changes in the global value of ecosystem services. *Global Environmental Change*, Elsevier, v. 26, p. 152–158, 2014.

COX, Peter M et al. Increasing risk of Amazonian drought due to decreasing aerosol pollution. *Nature*, Nature Publishing Group, v. 453, n. 7192, p. 212–215, 2008.

CRUTZEN, Paul J; ANDREAE, Meinrat O. Biomass burning in the tropics: Impact on atmospheric chemistry and biogeochemical cycles. *Science*, American Association for the Advancement of Science, v. 250, n. 4988, p. 1669–1678, 1990.

CRUZ, Carlos Henrique de Brito. Social distancing in São Paulo state: demonstrating the reduction in cases using time series analysis of deaths due to COVID-19. *Revista Brasileira de Epidemiologia*, SciELO Public Health, v. 23, p. e200056, 2020.

DEB, Pragyan et al. The effect of containment measures on the COVID-19 pandemic. CEPR Discussion Paper No. DP15086. 2020.

DIAS, Maria AF Silva et al. Changes in extreme daily rainfall for São Paulo, Brazil. *Climatic Change*, Springer, v. 116, n. 3-4, p. 705–722, 2013.

DIFFENBAUGH, Noah S; SCHERER, Martin; TRAPP, Robert J. Robust increases in severe thunderstorm environments in response to greenhouse forcing. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 110, n. 41, p. 16361–16366, 2013.

DIJK, Albert IJM Van et al. The Millennium Drought in southeast Australia (2001–2009): Natural and human causes and implications for water resources, ecosystems, economy, and society. *Water Resources Research*, Wiley Online Library, v. 49, n. 2, p. 1040–1057, 2013.

DONG, Liying; HU, Shasha; GAO, Jianjun. Discovering drugs to treat coronavirus disease 2019 (COVID-19). *Drug Discoveries & Therapeutics*, International Research and Cooperation Association for Bio & Socio-Sciences ..., v. 14, n. 1, p. 58–60, 2020.

DOWDY, Andrew J. Climatological variability of fire weather in Australia. *Journal of Applied Meteorology and Climatology*, v. 57, n. 2, p. 221–234, 2018.

DOWDY, Andrew J; FIELD, Robert D; SPESSA, Allan C. Seasonal forecasting of fire weather based on a new global fire weather database. In: FIRE, International Association of Wildland (Ed.). *Proceedings for the 5th International Fire Behaviour and Fuels Conference*. [S.l.]: International Association of Wildland Fire, 2016.

DRUMOND, Anita Rodrigues Moraes; AMBRIZZI, Tércio. The role of SST on the South American atmospheric circulation during January, February and March 2001. *Climate Dynamics*, Springer, v. 24, n. 7-8, p. 781–791, 2005.

DUFEK, AS; AMBRIZZI, Tércio. Precipitation variability in São Paulo state, brazil. *Theoretical and Applied Climatology*, Springer, v. 93, n. 3-4, p. 167–178, 2008.

DUTTA, Ritaban; DAS, Aruneema; ARYAL, Jagannath. Big data integration shows Australian bush-fire frequency is increasing significantly. *Royal Society Open Science*, The Royal Society Publishing, v. 3, n. 2, p. 150241, 2016.

EARL, Nick; SIMMONDS, Ian. Variability, trends, and drivers of regional fluctuations in Australian fire activity. *Journal of Geophysical Research: Atmospheres*, Wiley Online Library, v. 122, n. 14, p. 7445–7460, 2017.

_____. Spatial and temporal variability and trends in 2001–2016 global fire activity. *Journal of Geophysical Research: Atmospheres*, Wiley Online Library, v. 123, n. 5, p. 2524–2536, 2018.

ESTRADA, Francisco; PERRON, Pierre; MARTÍNEZ-LÓPEZ, Benjamín. Statistically derived contributions of diverse human influences to twentieth-century temperature changes. *Nature Geoscience*, Nature Publishing Group, v. 6, n. 12, p. 1050–1055, 2013.

FEARNSIDE, Phillip. Deforestation of the Brazilian Amazon. In: *Oxford Research Encyclopedia of Environmental Science*. [S.l.: s.n.], 2017.

FEARNSIDE, Philip F. Containing destruction from Brazil's amazon highways: now is the time to give weight to the environment in decision-making. *Environmental Conservation*, Cambridge University Press, v. 33, n. 3, p. 181–183, 2006.

FEARNSIDE, Philip M. Brazil's Amazon forest in mitigating global warming: unresolved controversies. *Climate Policy*, Taylor & Francis, v. 12, n. 1, p. 70–81, 2012.

FERRANTE, Lucas; FEARNSIDE, Philip M. Protect indigenous peoples from COVID-19. *Science*, American Association for the Advancement of Science, v. 368, n. 6488, p. 251–251, 2020.

FERRANTE, Lucas et al. Brazil's policies condemn Amazonia to a second wave of COVID-19. *Nature Medicine*, Nature Publishing Group, n. 26, p. 1315, 2020.

FERRAZ, Claudio. Explaining agriculture expansion and deforestation: evidence from the Brazilian Amazon-1980/98. Ipea working paper, 2001.

FERREIRA, Mauricio et al. Radiation balance at the surface in the city of são Paulo, Brazil: Diurnal and seasonal variations. *Theoretical and Applied Climatology*, v. 107, p. 229–246, 01 2011.

Food and Agriculture Organization of the United Nations. *FAOSTAT database*. [S.l.]: FAO-ESS Rome, Italy, 2017.

FREITAS, Carlos Machado de; CIDADE, Natália da Cunha et al. COVID-19 as a global disaster: Challenges to risk governance and social vulnerability in Brazil. *Ambiente & Sociedade*, SciELO Brasil, v. 23, p. e0115, 2020.

FREITAS, Edmilson D et al. Interactions of an urban heat island and sea-breeze circulations during winter over the metropolitan area of São Paulo, Brazil. *Boundary-layer meteorology*, Springer, v. 122, n. 1, p. 43–65, 2007.

FREITAS, Edmilson Dias de; DIAS, Pedro Leite da Silva. Alguns efeitos de áreas urbanas na geração de uma ilha de calor. *Revista Brasileira de Meteorologia*, v. 20, n. 3, p. 355–366, 2005.

FRIEDL, M.A et al. Global land cover mapping from MODIS: algorithms and early results. *Remote Sensing of Environment*, v. 83, n. 1, p. 287 – 302, 2002. ISSN 0034-4257. The Moderate Resolution Imaging Spectroradiometer (MODIS): a new generation of Land Surface Monitoring.

GALLANT, Ailie JE; HENNESSY, Kevin J; RISBEY, James. Trends in rainfall indices for six Australian regions: 1910-2005. *Australian Meteorological Magazine*, v. 56, n. 4, p. 223–241, 2007.

GATTO, Marino et al. Spread and dynamics of the COVID-19 epidemic in italy: Effects of emergency containment measures. *Proceedings of the National Academy of Sciences*, National Academy of Sciences, v. 117, n. 19, p. 10484–10491, 2020.

GEDALOF, Ze'ev. Climate and spatial patterns of wildfire in North America. In: *The landscape ecology of fire*. [S.l.]: Springer, 2011. p. 89–115.

GEDNEY, Nicola; VALDES, Paul J. The effect of Amazonian deforestation on the northern hemisphere circulation and climate. *Geophysical Research Letters*, Wiley Online Library, v. 27, n. 19, p. 3053–3056, 2000.

GENCER, E; FOLORUNCHO, M R.; LINKIN, M. Disasters and risk in cities. In: ROSENZWEIG, Cynthia et al. (Ed.). *Climate change and cities: Second assessment report of the urban climate change research network*. [S.l.]: Cambridge University Press, 2018. p. 61–98.

GENSINI, Vittorio A; BROOKS, Harold E. Spatial trends in United States tornado frequency. *npj Climate and Atmospheric Science*, Nature Publishing Group, v. 1, n. 1, p. 38, 2018.

GIGLIO, L.; JUSTICE, C. MOD14A2 MODIS/Terra thermal anomalies/fire 8-day L3 global 1km SIN grid v006 [data set]. *NASA EOSDIS Land Processes DAAC*, 2015.

GIGLIO, Louis; RANDERSON, James T; WERF, Guido R Van Der. Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4). *Journal of Geophysical Research: Biogeosciences*, Wiley Online Library, v. 118, n. 1, p. 317–328, 2013.

GILLETT, NP et al. Detecting the effect of climate change on Canadian forest fires. *Geophysical Research Letters*, Wiley Online Library, v. 31, n. 18, 2004.

GOFF, Héloïse Le; FLANNIGAN, Mike D; BERGERON, Yves. Potential changes in monthly fire risk in the eastern Canadian boreal forest under future climate change. *Canadian Journal of Forest Research*, NRC Research Press, v. 39, n. 12, p. 2369–2380, 2009.

GOLDEMBERG, José; COELHO, Suani Teixeira; GUARDABASSI, Patricia. The sustainability of ethanol production from sugarcane. *Energy policy*, Elsevier, v. 36, n. 6, p. 2086–2097, 2008.

GÓMEZ-RUBIO, Virgilio. *Bayesian inference with INLA*. [S.l.]: CRC Press, 2020.

GOMEZ-RUBIO, V. *Bayesian Inference with INLA*. [S.l.]: CRC Press, 2020.

GREEN, P.J.; SILVERMAN, B.W. *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. [S.l.]: CRC Press, 1994.

GRIFFITHS, G M et al. Change in mean temperature as a predictor of extreme temperature change in the Asia–Pacific region. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, Wiley Online Library, v. 25, n. 10, p. 1301–1330, 2005.

GRIMM, Alice M.; TEDESCHI, Renata G. Enso and extreme rainfall events in South America. *Journal of Climate*, American Meteorological Society, Boston MA, USA, v. 22, n. 7, p. 1589 – 1609, 2009.

GU, Yaofeng; LI, Dan. A modeling study of the sensitivity of urban heat islands to precipitation at climate scales. *Urban Climate*, v. 24, p. 982–993, 2018. ISSN 2212-0955.

GUERRA, Angélica et al. The importance of legal reserves for protecting the pantanal biome and preventing agricultural losses. *Journal of environmental management*, Elsevier, v. 260, p. 110128, 2020.

_____. Drivers and projections of vegetation loss in the pantanal and surrounding ecosystems. *Land Use Policy*, Elsevier, v. 91, p. 104388, 2020.

GUERRA-SHINOHARA, Elvira Maria et al. Overview on covid-19 outbreak indicators across brazilian federative units. *medRxiv 2020.06.02.20120220*, Cold Spring Harbor Laboratory Press, 2020.

GUPTA, Rajan; PAL, Saibal Kumar. Trend analysis and forecasting of COVID-19 outbreak in India. *medRxiv 2020.03.26.20044511*, Cold Spring Harbor Laboratory Press, 2020.

HAMILTON, James D. Why you should never use the Hodrick-Prescott Filter. *The Review of Economics and Statistics*, v. 100, n. 5, p. 831–843, 2018.

HARDESTY, Jeff; MYERS, Ron; FULKS, Wendy. Fire, ecosystems, and people: a preliminary assessment of fire as a global conservation issue. In: JSTOR. *The George Wright Forum*. [S.l.], 2005. v. 22, n. 4, p. 78–87.

HARGRAVE, Jorge; KIS-KATOS, Krisztina. Economic causes of deforestation in the Brazilian Amazon: a panel data analysis for the 2000s. *Environmental and Resource Economics*, Springer, v. 54, n. 4, p. 471–494, 2013.

HARRIS, Mônica B et al. Safeguarding the Pantanal wetlands: threats and conservation initiatives. *Conservation Biology*, Wiley Online Library, v. 19, n. 3, p. 714–720, 2005.

HARRIS, Sarah; LUCAS, Chris. Understanding the variability of Australian fire weather between 1973 and 2017. *PloS one*, Public Library of Science, v. 14, n. 9, p. e0222328, 2019.

HARVEY, A.C. *Forecasting, Structural Time Series and the Kalman Filter*. [S.l.]: Cambridge University Press, 1989.

HARVEY, A.; TRIMBUR, T. Trend estimation and the Hodrick-Prescott Filter. *Journal of Japan Statistical Society*, v. 38, n. 1, p. 41–49, 2008.

HARVEY, Andrew C. *Forecasting, structural time series models and the Kalman filter*. [S.l.]: Cambridge university press, 1990.

HARVEY, A. C.; SHEPARD, N. Structural time series models. In: MADDALA, G. S.; RAO, C. R.; VINOD, H. D. (Ed.). *Handbook of Statistics 11*. [S.l.]: Elsevier Science Publisher, 1993. p. 261–302.

HENDON, Harry H; THOMPSON, David WJ; WHEELER, Matthew C. Australian rainfall and surface temperature variations associated with the Southern Hemisphere annular mode. *Journal of Climate*, v. 20, n. 11, p. 2452–2467, 2007.

HSIANG, Solomon et al. The effect of large-scale anti-contagion policies on the COVID-19 pandemic. *Nature*, Nature Publishing Group, v. 584, n. 7820, p. 262–267, 2020.

HUFF, F. A.; CHANGNON, S. A. precipitation modification by major urban areas. *Bulletin of the American Meteorological Society*, American Meteorological Society, Boston MA, USA, v. 54, n. 12, p. 1220 – 1233, 1973.

HUGHES, Lesley. Climate change and Australia: trends, projections and impacts. *Austral Ecology*, Wiley Online Library, v. 28, n. 4, p. 423–443, 2003.

ILLIAN, JB et al. *Fitting a log Gaussian Cox process with temporally varying effects–a case study*. [S.l.], 2010. Preprint Statistics.

ILLIAN, J. B; SORBYE, S. H.; RUE, H. A toolbox for fitting complex spatial point process models using integrated nested Laplace approximation (INLA). *Annals of Applied Statistics*, v. 4, n. 12, p. 1499–1530, 2012.

IVORY, Sarah J et al. Vegetation, rainfall, and pulsing hydrology in the pantanal, the world's largest tropical wetland. *Environmental Research Letters*, IOP Publishing, v. 14, n. 12, p. 124017, 2019.

JACOBI, Pedro R et al. Water governance and natural disasters in the Metropolitan Region of são Paulo, Brazil. *International Journal of Urban Sustainable Development*, Taylor & Francis, v. 5, n. 1, p. 77–88, 2013.

JUNK, Wolfgang J et al. Biodiversity and its conservation in the Pantanal of Mato Grosso, Brazil. *Aquatic Sciences*, Springer, v. 68, n. 3, p. 278–309, 2006.

KAUFMANN, R.K.; KAUPPI, H.; STOCK, J.H. Does temperature contain a stochastic trend? evaluating conflicting statistical results. *Climatic Change*, v. 101, p. 395–405, 2010.

KENDALL, M. A.; STUART, A. *The Advanced Theory of Statistics, 2nd ed.* [S.l.]: Charles Griffin, 1967.

KOLACZYK, Eric D; CSÁRDI, Gábor. *Statistical analysis of network data with R*. [S.l.]: Springer, 2014.

KRAEMER, Moritz UG et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*, American Association for the Advancement of Science, v. 368, n. 6490, p. 493–497, 2020.

KRAINSKI, Elias T et al. *Advanced spatial modeling with stochastic partial differential equations using R and INLA*. [S.l.]: Chapman and Hall/CRC, 2018.

KUNKEL, Kenneth E et al. Monitoring and understanding trends in extreme storms: State of knowledge. *Bulletin of the American Meteorological Society*, American Meteorological Society, v. 94, n. 4, p. 499–514, 2013.

LAMBERT, Diane. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, Taylor & Francis, v. 34, n. 1, p. 1–14, 1992.

LAPOLA, David M et al. Pervasive transition of the brazilian land-use system. *Nature climate change*, Nature Publishing Group UK London, v. 4, n. 1, p. 27–35, 2014.

LAU, Hien et al. Evaluating the massive underreporting and undertesting of COVID-19 cases in multiple global epicenters. *Pulmonology*, Elsevier, v. 15, n. 2, p. 110–115, 2020.

LAURINI, MP. A spatio-temporal approach to estimate patterns of climate change. *Environmetrics*, Wiley Online Library, v. 30, n. 1, p. e2542, 2019.

LAURINI, Márcio Poletti. The spatio-temporal dynamics of ethanol/gasoline price ratio in Brazil. *Renewable and Sustainable Energy Reviews*, Elsevier, v. 70, p. 1–12, 2017.

LEE, Cameron C. Utilizing synoptic climatological methods to assess the impacts of climate change on future tornado-favorable environments. *Natural hazards*, Springer, v. 62, n. 2, p. 325–343, 2012.

LI, Qiang; FENG, Wei; QUAN, Ying-Hui. Trend and forecasting of the COVID-19 outbreak in China. *Journal of Infection*, Elsevier, v. 80, n. 4, p. 469–496, 2020.

LI, Wenhong; FU, Rong; DICKINSON, Robert E. Rainfall and its seasonality over the Amazon in the 21st century as assessed by the coupled models for the IPCC AR4. *Journal of Geophysical Research: Atmospheres*, Wiley Online Library, v. 111, n. D2, 2006.

LIBONATI, Renata et al. Rescue brazil's burning pantanal wetlands. *Nature*, Nature Publishing Group UK London, v. 588, n. 7837, p. 217–219, 2020.

LIBONATI, R et al. Twenty-first century droughts have not increasingly exacerbated fire season severity in the brazilian amazon. *Scientific reports*, Springer, v. 11, n. 1, p. 1–13, 2021.

LIMA, Gabriela Narcizo; LOMBARDO, Magda Adelaide; MAGAÑA, V. Urban water supply and the changes in the precipitation patterns in the metropolitan area of São Paulo–Brazil. *Applied Geography*, Elsevier, v. 94, p. 223–229, 2018.

LIMA, Gabriela Narcizo; RUEDA, Víctor Orlando Magaña. The urban growth of the metropolitan area of Sao Paulo and its impact on the climate. *Weather and Climate Extremes*, Elsevier, v. 21, p. 17–26, 2018.

LIN, Chuan-Yao et al. Impact of the urban heat island effect on precipitation over a complex geographic environment in Northern Taiwan. *Journal of Applied Meteorology and Climatology*, American Meteorological Society, Boston MA, USA, v. 50, n. 2, p. 339 – 353, 2011.

LINDGREN, F.; RUE, H. On the second-order random walk model for irregular locations. *Scandinavian Journal of Statistics*, v. 35, n. 4, p. 691–700, 2008.

LINDGREN, Finn; RUE, Håvard; LINDSTRÖM, Johan. An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Wiley Online Library, v. 73, n. 4, p. 423–498, 2011.

LINKA, Kevin et al. Outbreak dynamics of covid-19 in europe and the effect of travel restrictions. *Computer Methods in Biomechanics and Biomedical Engineering*, Taylor & Francis, p. 1–8, 2020.

LIU, Xiuli et al. Modeling the situation of COVID-19 and effects of different containment strategies in china with dynamic differential equations and parameters estimation. *medRxiv 2020.03.09.20033498*, Cold Spring Harbor Laboratory Press, 2020.

LONGO, KM et al. Biomass burning in Amazonia: Emissions, long-range transport of smoke and its regional and remote impacts. *Geophysical Monograph Series*, v. 186, p. 207–232, 2009.

LOZANO, Olga M et al. Assessing climate change impacts on wildfire exposure in Mediterranean areas. *Risk Analysis*, Wiley Online Library, v. 37, n. 10, p. 1898–1916, 2017.

MALHI, Yadvinder et al. Climate change, deforestation, and the fate of the Amazon. *science*, American Association for the Advancement of Science, v. 319, n. 5860, p. 169–172, 2008.

MANN, H. B. Non-parametric tests against trend. *Econometrica*, v. 13, n. 3, p. 245–259, 1945.

MARENGO, José A. Interdecadal variability and trends of rainfall across the Amazon basin. *Theoretical and Applied Climatology*, Springer, v. 78, n. 1-3, p. 79–96, 2004.

MARENGO, Jose A et al. Trends in extreme rainfall and hydrogeometeorological disasters in the Metropolitan Area of São Paulo: a review. *Annals of the New York Academy of Sciences*, Wiley Online Library, 2020.

_____. Extreme drought in the brazilian pantanal in 2019–2020: Characterization, causes, and impacts. *Frontiers in Water*, Frontiers, v. 3, p. 13, 2021.

MARENGO, José A et al. The drought of Amazonia in 2005. *Journal of Climate*, v. 21, n. 3, p. 495–516, 2008.

MARENGO, Jose A et al. An intercomparison of observed and simulated extreme rainfall and temperature events during the last half of the twentieth century: part 2: historical trends. *Climatic Change*, Springer, v. 98, n. 3, p. 509–529, 2010.

MARENGO, Jose A; VALVERDE, Maria C; OBREGON, Guillermo O. Observed and projected changes in rainfall extremes in the metropolitan area of são paulo. *Climate research*, v. 57, n. 1, p. 61–72, 2013.

MARIANI, Michela et al. ENSO controls interannual fire activity in southeast Australia. *Geophysical Research Letters*, Wiley Online Library, v. 43, n. 20, p. 10–891, 2016.

MARLE, Margreet JE van et al. Fire and deforestation dynamics in amazonia (1973–2014). *Global biogeochemical cycles*, Wiley Online Library, v. 31, n. 1, p. 24–38, 2017.

MARTINEZ-ALVAREZ, Melisa et al. Covid-19 pandemic in west africa. *The Lancet Global Health*, Elsevier, v. 8, n. 5, p. e631–e632, 2020.

MARTINO, S.; RUE, H. *Implementing Approximate Bayesian Inference using Integrated Nested Laplace Approximation: a manual for the inla program.* [S.l.], 2020. Https://inla.r-inla-download.org/r-inla.org/doc/inla-manual/inla-manual.pdf.

MBOW, Moustapha et al. Covid-19 in africa: Dampening the storm? *Science*, American Association for the Advancement of Science, v. 369, n. 6504, p. 624–626, 2020.

MEMON, Rizwan Ahmed; LEUNG, Dennis Y.C.; LIU, Chun-Ho. An investigation of urban heat island intensity (UHII) as an indicator of urban heating. *Atmospheric Research*, v. 94, n. 3, p. 491–500, 2009. ISSN 0169-8095.

MENDONÇA, Mário Jorge Cardoso De et al. The economic cost of the use of fire in the Amazon. *Ecological Economics*, Elsevier, v. 49, n. 1, p. 89–105, 2004.

MENDONÇA, Mário Jorge Cardoso De; SACHSIDA, Adolfo; LOUREIRO, Paulo RA. Estimation of damage to human health due to forest burning in the Amazon. *Journal of Population Economics*, Springer, v. 19, n. 3, p. 593–610, 2006.

MØLLER, Jesper; SYVERSVEEN, Anne Randi; WAAGEPETERSEN, Rasmus Plenge. Log Gaussian Cox processes. *Scandinavian Journal of Statistics*, Wiley Online Library, v. 25, n. 3, p. 451–482, 1998.

MOORE, Todd W. On the temporal and spatial characteristics of tornado days in the United States. *Atmospheric Research*, Elsevier, v. 184, p. 56–65, 2017.

MOORE, Todd W; DEBOER, Tiffany A. A review and analysis of possible changes to the climatology of tornadoes in the United States. *Progress in Physical Geography: Earth and Environment*, SAGE Publications Sage UK: London, England, v. 43, n. 3, p. 365–390, 2019.

MOORE, Todd W; MCGUIRE, Michael P. Using the standard deviational ellipse to document changes to the spatial dispersion of seasonal tornado activity in the United States. *npj Climate and Atmospheric Science*, Nature Publishing Group, v. 2, n. 1, p. 1–8, 2019.

MORITZ, Max A et al. Learning to coexist with wildfire. *Nature*, Nature Publishing Group, v. 515, n. 7525, p. 58–66, 2014.

MORTON, DC et al. Agricultural intensification increases deforestation fire activity in Amazonia. *Global Change Biology*, Wiley Online Library, v. 14, n. 10, p. 2262–2275, 2008.

MORTON, Douglas C et al. Cropland expansion changes deforestation dynamics in the southern brazilian amazon. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 103, n. 39, p. 14637–14641, 2006.

MUDELSEE, Manfred. Trend analysis of climate time series: A review of methods. *Earth-Science Reviews*, v. 190, p. 310 – 322, 2019. ISSN 0012-8252.

MURPHY, Brett P et al. Fire regimes of Australia: a pyrogeographic model system. *Journal of Biogeography*, Wiley Online Library, v. 40, n. 6, p. 1048–1058, 2013.

NEPSTAD, Daniel et al. Amazon drought and its implications for forest flammability and tree growth: A basin-wide analysis. *Global Change Biology*, Wiley Online Library, v. 10, n. 5, p. 704–717, 2004.

_____. Slowing amazon deforestation through public policy and interventions in beef and soy supply chains. *science*, American Association for the Advancement of Science, v. 344, n. 6188, p. 1118–1123, 2014.

NOBRE, Carlos A; SELLERS, Piers J; SHUKLA, Jagadish. Amazonian deforestation and regional climate change. *Journal of Climate*, v. 4, n. 10, p. 957–988, 1991.

NOBRE, Carlos A et al. Vulnerabilidades das megacidades brasileiras às mudanças climáticas: Região metropolitana de São Paulo. *Embaixada Reino Unido, Rede Clima e Programa FAPESP em Mudanças Climáticas*, 2010.

NUNES, Lucí Hidalgo et al. Patterns of extreme precipitation in Santos. In: *Climate Change in Santos Brazil: Projections, Impacts and Adaptation Options*. [S.l.]: Springer, 2019. p. 45–57.

OBREGÓN, Guillermo O; MARENGO, José A; NOBRE, Carlos A. Rainfall and climate variability: long-term trends in the Metropolitan Area of são Paulo in the 20th century. *Climate Research*, v. 61, n. 2, p. 93–107, 2014.

OLIVEIRA, Maiby Teodoro de et al. Regeneration of riparian forests of the brazilian Pantanal under flood and fire influence. *Forest Ecology and Management*, Elsevier, v. 331, p. 256–263, 2014.

OLIVEIRA, Ubirajara et al. Determinants of fire impact in the brazilian biomes. *Frontiers in Forests and Global Change*, Frontiers Media SA, v. 5, p. 735017, 2022.

ORGANIZATION, World Healh. *WHO Director-General's opening remarks at the media briefing on COVID-19-11 March 2020*. 2020.

ORTEGA, Francisco; ORSINI, Michael. Governing COVID-19 without government in Brazil: Ignorance, neoliberal authoritarianism, and the collapse of public health leadership. *Global Public Health*, Taylor & Francis, v. 15, n. 9, p. 1257–1277, 2020.

PANTANAL, Instituto SOS; WWF-BRASIL. *Monitoramento das alterações da cobertura vegetal e uso do Solo na Bacia do Alto Paraguai–Porção Brasileira. Período de análise: 2012-2014*. [S.l.]: WWF-Brasil Brasília, 2015.

PARAISO, Maria Leticia de Souza; GOUVEIA, Nelson. Health risks due to pre-harvesting sugarcane burning in são paulo state, brazil. *Revista Brasileira de Epidemiologia*, SciELO Public Health, v. 18, p. 691–701, 2015.

PEDERSEN, Morten Gram; MENEGHINI, Matteo. Quantifying undetected COVID-19 cases and effects of containment measures in Italy. *ResearchGate Preprint (online 21 March 2020) DOI*, v. 10, 2020.

PERONE, Gaetano. An ARIMA model to forecast the spread and the final size of COVID-2019 epidemic in Italy. *medRxiv 2020.04.27.2008153*, Cold Spring Harbor Laboratory Press, 2020.

PITMAN, AJ; NARISMA, GT; MCANENEY, J. The impact of climate change on the risk of forest and grassland fires in Australia. *Climatic Change*, Springer, v. 84, n. 3-4, p. 383–401, 2007.

PIVELLO, Vânia R. The use of fire in the cerrado and amazonian rainforests of brazil: past and present. *Fire ecology*, Springer, v. 7, n. 1, p. 24–39, 2011.

PIVELLO, Vânia R et al. Understanding brazil's catastrophic fires: Causes, consequences and policy needed to prevent future tragedies. *Perspectives in Ecology and Conservation*, Elsevier, 2021.

POST, Lori Ann et al. Dynamic public health surveillance to track and mitigate the us covid-19 epidemic: longitudinal trend analysis study. *Journal of medical Internet research*, JMIR Publications Toronto, Canada, v. 22, n. 12, p. e24286, 2020.

PRADO, Marcelo Freitas do et al. Analysis of COVID-19 under-reporting in Brazil. *Revista Brasileira de Terapia Intensiva*, Associação de Medicina Intensiva Brasileira, v. 32, n. 2, p. 224, 2020.

PROIETTI, Tommaso; HILLEBRAND, Eric. Seasonal changes in central England temperatures. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, Wiley Online Library, v. 180, n. 3, p. 769–791, 2017.

RABIN, Sam S et al. Quantifying regional, time-varying effects of cropland and pasture on vegetation fire. *Biogeosciences*, Copernicus GmbH, v. 12, n. 22, p. 6591–6604, 2015.

RAHMAN, Shahriar et al. Forest fire occurrence and modeling in Southeastern Australia. In: *Forest fire*. [S.l.]: InTechOpen, 2018.

RAIMUNDO, Clebson do Carmo; SANSIGOLO, Clóvis Angeli; MOLION, Luiz Carlos Baldicero. Tendências das classes de precipitação na região metropolitana de São Paulo. *Revista Brasileira de Meteorologia*, SciELO Brasil, v. 29, n. 3, p. 397–408, 2014.

REINHARD, Mia; REBETEZ, M; SCHLAEPFER, R. Recent climate change: rethinking drought in the context of forest fire research in Ticino, South of Switzerland. *Theoretical and Applied Climatology*, Springer, v. 82, n. 1-2, p. 17–25, 2005.

RIBEIRO, Helena. *Sugar cane burning in Brazil: respiratory health effects*. [S.l.]: FACULDADE DE SAUDE PUBLICA DA USP, 2008.

RIBEIRO, Leonardo Costa; BERNARDES, Américo Tristão. *Estimate of underreporting of COVID-19 in Brazil by Acute Respiratory Syndrome hospitalization reports*. [S.l.], 2020.

RISBEY, James S et al. On the remote drivers of rainfall variability in Australia. *Monthly Weather Review*, v. 137, n. 10, p. 3233–3253, 2009.

ROPELEWSKI, C. F.; BELL, M. A. Shifts in the statistics of daily rainfall in South America conditional on ENSO phase. *Journal of Climate*, American Meteorological Society, Boston MA, USA, v. 21, n. 5, p. 849 – 865, 2008.

RUDORFF, Bernardo Friedrich Theodor et al. Studies on the rapid expansion of sugarcane for ethanol production in São Paulo state (Brazil) using Landsat data. *Remote sensing*, Molecular Diversity Preservation International, v. 2, n. 4, p. 1057–1076, 2010.

RUE, H.; HELD, L. *Gaussian Markov Random Fields: Theory and Applications*. [S.l.]: Chapman & Hall - CRC, 2005.

_____. _____. [S.l.]: Chapman and Hall/CRC, 2005.

_____. _____. [S.l.]: CRC Press, 2005.

RUE, Håvard; MARTINO, Sara; CHOPIN, Nicolas. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series b (statistical methodology)*, Wiley Online Library, v. 71, n. 2, p. 319–392, 2009.

RUSSELL-SMITH, Jeremy et al. Bushfires 'down under': patterns and implications of contemporary Australian landscape burning. *International Journal of Wildland Fire*, CSIRO, v. 16, n. 4, p. 361–377, 2007.

RUSSELL, Timothy W et al. Using a delay-adjusted case fatality ratio to estimate under-reporting. *Centre for Mathematical Modeling of Infectious Diseases Repository*, 2020.

SAEZ, Marc et al. Effectiveness of the measures to flatten the epidemic curve of COVID-19. the case of Spain. *Science of the Total Environment*, Elsevier, p. 138761, 2020.

SANDERS, James M et al. Pharmacologic treatments for coronavirus disease 2019 (COVID-19): a review. *Jama*, American Medical Association, v. 323, n. 18, p. 1824–1836, 2020.

SCHULZ, Christopher et al. Physical, ecological and human dimensions of environmental change in brazil's pantanal wetland: Synthesis and research agenda. *Science of the Total Environment*, Elsevier, v. 687, p. 1011–1027, 2019.

SEIDL, Andrew F; MORAES, Andre Steffens. Global valuation of ecosystem services: application to the Pantanal da Nhecolandia, Brazil. *Ecological Economics*, Elsevier, v. 33, n. 1, p. 1–6, 2000.

SERRA, Laura et al. Spatio-temporal log-Gaussian Cox processes for modelling wildfire occurrence: The case of Catalonia, 1994–2008. *Environmental and Ecological Statistics*, Springer, v. 21, n. 3, p. 531–563, 2014.

SIEGERT, F et al. Increased damage from fires in logged forests during droughts caused by El Nino. *Nature*, Nature Publishing Group, v. 414, n. 6862, p. 437–440, 2001.

SILVA, Fernanda et al. Spatial and temporal variability patterns of the urban heat island in são Paulo. *Environments*, v. 4, p. 1–26, 04 2017.

SILVA, J et al. Evolution of deforestation in the Brazilian Pantanal and surroundings in the timeframe 1976-2008. *Embrapa Informática Agropecuária-Artigo em periódico indexado (ALICE)*, Geografia, Rio Claro, v. 36, p. 35-55, jun. 2011., 2011.

SILVA, João dos Santos Vila da; ABDON, Myrian de Moura. Delimitação do Pantanal brasileiro e suas sub-regiões. *Pesquisa Agropecuária Brasileira*, v. 33, n. 13, p. 1703–1711, 1998.

SILVA, Lena Veiga E et al. An analysis of COVID-19 mortality underreporting based on data available from official Brazilian government internet portals. *Journal of Medical Internet Research*, v. 22, n. 8, p. e21413, 2020.

SIMPSON, Daniel et al. Going off grid: Computationally efficient inference for log-Gaussian Cox processes. *Biometrika*, Oxford University Press, v. 103, n. 1, p. 49–70, 2016.

SIMPSON, Daniel; LINDGREN, Finn; RUE, Håvard. Think continuous: Markovian gaussian models in spatial statistics. *Spatial Statistics*, Elsevier, v. 1, p. 16–29, 2012.

SIMPSON, D. et al. Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science*, v. 32, n. 1, p. 1–28, 2017.

SINGH, Jitendra et al. Urbanization alters rainfall extremes over the contiguous United States. *Environmental Research Letters*, IOP Publishing, v. 15, n. 7, p. 074033, jul 2020.

SKIERA, Bernd et al. How to best predict the daily number of new infections of Covid-19. *arXiv preprint arXiv:2004.03937*, 2020.

SONG, Yanling; ACHBERGER, Christine; LINDERHOLM, Hans W. Rain-season trends in precipitation and their effect in different climate regions of China during 1961–2008. *Environmental Research Letters*, IOP Publishing, v. 6, n. 3, p. 034025, 2011.

STORCH, H. v.; ZWIERS, F. W. *Statistical Analysis in Climate Research*. [S.l.]: Cambridge University Press, 2002.

SUGAHARA, Shigetoshi et al. Homogeneity assessment of a station climate series (1933–2005) in the metropolitan area of são paulo: instruments change and urbanization effects. *Theoretical and Applied Climatology*, Springer, v. 107, n. 3, p. 361–374, 2012.

SVETUNKOV, Ivan; PETROPOULOS, Fotios. Old dog, new tricks: a modelling view of simple moving averages. *International Journal of Production Research*, v. 56, n. 18, p. 6034–6047, 2018.

SYPHARD, Alexandra D et al. Human presence diminishes the importance of climate in driving fire activity across the United States. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 114, n. 52, p. 13750–13755, 2017.

TENG, Ming; NATHOO, Farouk; JOHNSON, Timothy D. Bayesian computation for log-Gaussian Cox processes: A comparative analysis of methods. *Journal of Statistical Computation and Simulation*, Taylor & Francis, v. 87, n. 11, p. 2227–2252, 2017.

TIMBAL, Bertrand; ARBLASTER, Julie M; POWER, Scott. Attribution of the late-twentieth-century rainfall decline in southwest Australia. *Journal of Climate*, v. 19, n. 10, p. 2046–2062, 2006.

TIPPETT, Michael K et al. Climate and hazardous convective weather. *Current Climate Change Reports*, Springer, v. 1, n. 2, p. 60–73, 2015.

TIPPETT, Michael K; LEPORE, Chiara; COHEN, Joel E. More tornadoes in the most extreme US tornado outbreaks. *Science*, American Association for the Advancement of Science, v. 354, n. 6318, p. 1419–1423, 2016.

TOMAS, Walfrido M et al. Sustainability agenda for the Pantanal wetland: perspectives on a collaborative interface for science, policy, and decision-making. *Tropical Conservation Science*, SAGE Publications Sage CA: Los Angeles, CA, v. 12, p. 1940082919872634, 2019.

VAID, Shashank; CAKAN, Caglar; BHANDARI, Mohit. Using machine learning to estimate unobserved COVID-19 infections in North America. *The Journal of Bone and Joint Surgery. American Volume*, Wolters Kluwer Health, v. 102, n. 13, p. e70, 2020.

VALENTE, Fernanda; LAURINI, Márcio. Tornado occurrences in the United States: A spatio-temporal point process approach. *Econometrics*, Multidisciplinary Digital Publishing Institute, v. 8, n. 2, p. 25, 2020.

VALENTE, F; LAURINI, M. Pre-harvest sugarcane burning: A statistical analysis of the environmental impacts of a regulatory change in the energy sector. *Cleaner Engineering and Technology*, Elsevier, v. 4, p. 100255, 2021.

VALENTE, F.; LAURINI, M. Spatio-temporal analysis of fire occurrence in australia. *Stochastic Environmental Research and Risk Assessment*, in press, 2021.

VALENTE, Fernanda; LAURINI, Márcio. Spatio-temporal analysis of fire occurrence in australia. *Stochastic Environmental Research and Risk Assessment*, Springer, v. 35, n. 9, p. 1759–1770, 2021.

VALENTE, F.; LAURINI, M. P. Robust trend estimation for covid-19 data. Submitted. <https://www.dropbox.com/s/6dqhsy6lehwar45/COVIDpreprint.pdf?dl=0>. 2020.

VELARDE, Luis Guillermo Coca; MIGON, Hélio S; PEREIRA, Basilio de B. Space–time modeling of rainfall data. *Environmetrics*, Wiley Online Library, v. 15, n. 6, p. 561–576, 2004.

VERDON, Danielle C; KIEM, Anthony S; FRANKS, Stewart W. Multi-decadal variability of forest fire risk—Eastern Australia. *International Journal of Wildland Fire*, CSIRO, v. 13, n. 2, p. 165–171, 2004.

WATANABE, S. A widely applicable Bayesian information criterion. *Journal of Machine Learning Research*, v. 14, p. 867–897, 2013.

WERTH, David; AVISSAR, Roni. The local and global effects of Amazon deforestation. *Journal of Geophysical Research: Atmospheres*, Wiley Online Library, v. 107, n. D20, p. LBA–55, 2002.

WESTERLING, Anthony L et al. Warming and earlier spring increase western US forest wildfire activity. *science*, American Association for the Advancement of Science, v. 313, n. 5789, p. 940–943, 2006.

WHITTLE, Peter. On stationary processes in the plane. *Biometrika*, JSTOR, v. 41, n. 3/4, p. 434–449, 1954.

WILLIAMS, Allyson AJ; KAROLY, David J; TAPPER, Nigel. The sensitivity of Australian fire danger to climate change. *Climatic Change*, Springer, v. 49, n. 1-2, p. 171–191, 2001.

WILLIAMSON, Grant J et al. Measurement of inter-and intra-annual variability of landscape fire activity at a continental scale: the Australian case. *Environmental Research Letters*, IOP Publishing, v. 11, n. 3, p. 035003, 2016.

WOTTON, B Mike; NOCK, Charles A; FLANNIGAN, Mike D. Forest fire occurrence and climate change in Canada. *International Journal of Wildland Fire*, CSIRO, v. 19, n. 3, p. 253–271, 2010.

YANG, Zifeng et al. Modified seir and ai prediction of the epidemics trend of covid-19 in china under public health interventions. *Journal of thoracic disease*, AME Publications, v. 12, n. 3, p. 165, 2020.

ZHANG, Yongyong et al. Regional patterns of extreme precipitation and urban signatures in metropolitan areas. *Journal of Geophysical Research: Atmospheres*, v. 124, n. 2, p. 641–663, 2019.

ZHU, Xiudi et al. Impact of urbanization on hourly precipitation in Beijing, China: Spatio-temporal patterns and causes. *Global and Planetary Change*, v. 172, p. 307–324, 2019. ISSN 0921-8181.

ZILLI, Marcia T et al. A comprehensive analysis of trends in extreme precipitation over southeastern coast of Brazil. *International Journal of Climatology*, Wiley Online Library, v. 37, n. 5, p. 2269–2279, 2017.