UNIVERSIDADE FEDERAL DE MINAS GERAIS

Instituto de Ciências Exatas

Programa de Pós-Graduação em Estatística

Guilherme Aparecido Santos Aguilar

Beyond Gaussian processes: Flexible Bayesian modeling and inference for geostatistical processes

Belo Horizonte

2022

Guilherme Aparecido Santos Aguilar

Beyond Gaussian processes: Flexible Bayesian modeling and inference for geostatistical processes

Versão Final

Tese apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Doutor em Estatística.

Orientador: Prof. Dr. Flávio Bambirra Gonçalves
Coorientador: Prof. Dr. Marcos Oliveira Prates

Belo Horizonte

2022

:

Aguilar, Guilherme Aparecido Santos.

A283b Beyond Gaussian processes [manuscrito] : flexible Bayesian modeling and inference for geostatistical processes / Guilherme Aparecido Santos Aguilar. – 2022.
69 f. il.

Orientador: Flávio Bambirra Gonçalves.
Coorientador: Marcos Oliveira Prates.
Tese (doutorado) - Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Estatística.
Referências: f. 46-50

1. Estatística – Teses. 2. Distribuições de caudas pesadas – Teses. 3. Algoritmo MCMC – Teses. 4. Poisson, Distribuição de – Teses. I. Gonçalves, Flávio Bambirra. II. Prates, Marcos Oliveira. III. Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Estatística. IV.Título.

CDU 519.2(043)

Ficha catalográfica elaborada pela bibliotecária Belkiz Inez Rezende Costa
CRB 6/1510 Universidade Federal de Minas Gerais - ICEx

UNIVERSIDADE FEDERAL DE MINAS GERAIS

PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

FOLHA DE APROVAÇÃO

"Beyond Gaussian processes: Flexible Bayesian modeling and inference for geostatistical processes"

GUILHERME APARECIDO SANTOS AGUILAR

Tese submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ESTATÍSTICA, como requisito para obtenção do grau de Doutor em ESTATÍSTICA, área de concentração ESTATÍSTICA E PROBABILIDADE.

Aprovada em 25 de março de 2022, pela banca constituída pelos membros:

Prof(a). Flávio Bambirra Gonçalves - Orientador
DEST/UFMG

Prof(a). Marcos Oliveira Prates - Coorientador
DEST/UFMG

Prof(a). Vinícius Diniz Mayrink
DEST/UFMG

Prof(a). Dani Gamerman
DME/UFRJ

Prof(a). Thais Cristina Oliveira da Fonseca
DME/UFRJ

Prof(a). Fernanda De Bastiani
CCEN/UFPE

Belo Horizonte, 25 de março de 2022.

Resumo

Este trabalho propõe uma nova família de modelos geoestatísticos que possuem características que não podem ser adequadamente acomodadas por processos gaussianos tradicionais. A família é especificada hierarquicamente e combina a dinâmica dimensional infinita dos processos gaussianos com a de qualquer distribuição contínua multivariada. Esta combinação é definida estocasticamente através de um processo de Poisson latente e a nova família é denominada Processo de Mistura Poisson-Gaussiana - POGAMP. Enquanto a tentativa de definir um processo geoestatístico designando algumas distribuições contínuas arbitrárias como distribuições de dimensão finita geralmente leva a processos não válidos, o POGAMP pode ter suas distribuições de dimensão finita arbitrariamente próximas a qualquer distribuição contínua e ainda ser uma processo válido. São fornecidos resultados formais para estabelecer sua existência e outras propriedades importantes, como continuidade absoluta em relação a uma medida de processo gaussiana. Além disso, um algoritmo MCMC é cuidadosamente desenvolvido para realizar inferência Bayesiana quando o POGAMP é observado discretamente em algum domínio do espaço. Simulações são realizadas para investigar empiricamente as propriedades de modelagem do POGAMP e a eficiência do algoritmo MCMC. Finalmente, um conjunto de dados real é analisado para ilustrar a aplicabilidade da metodologia proposta.

Palavras-chave: Caudas pesadas; MCMC; Assimetria; Processo de Poisson.

Abstract

This work proposes a novel family of geostatistical models to account for features that cannot be properly accommodated by traditional Gaussian processes. The family is specified hierarchically and combines the infinite dimensional dynamics of Gaussian processes to that of any multivariate continuous distribution. This combination is stochastically defined through a latent Poisson process and the new family is called the Poisson-Gaussian Mixture Process - POGAMP. Whilst the attempt of defining a geostatistical process by assigning some arbitrary continuous distributions to be the finite-dimension distributions usually leads to non-valid processes, the POGAMP can have its finite-dimensional distributions to be arbitrarily close to any continuous distribution and still be a valid process. Formal results to establish its existence and other important properties, such as absolute continuity with respect to a Gaussian process measure are provided. Also, a MCMC algorithm is carefully devised to perform Bayesian inference when the POGAMP is discretely observed in some space domain. Simulations are performed to empirically investigate the modelling properties of the POGAMP and the efficiency of the MCMC algorithm. Finally, a real dataset is analysed to illustrate the applicability of the proposed methodology.

Keywords: Heavy tails; MCMC; Skewness; Poisson process.

List of Figures

List of Tables

Contents

Chapter 1

Introduction

Continuous spatial statistical modelling, referred to as geostatistics, is an appealing approach to explain the probabilistic dynamics of some response variable observed over a continuous region (Zareifard and Khaledi, 2013). Spatial prediction is an important problem in fields such as petroleum engineering, civil engineering, mining, geography, geology, environmental, hydrology and climate studies (Dubrule, 1989; Hohn, 1998; Kim and Mallick, 2004; Cressie, 2015; Bevilacqua et al., 2021). Geostatistical problems consider a partial realization of an underlying random field indexed by locations that vary continuously through a fixed region in space and focus on predicting values at unobserved locations or regions. The underlying random field is often assumed to be a Gaussian processes (GP) (Palacios and Steel, 2006). This assumption facilitates prediction and provides some justification for the use of the spatial prediction. In particular, a GP is completely characterized by its mean and covariance functions. A GP measure implies that the joint distribution at any finite collection of locations is a multivariate normal. The stability of the multivariate normal distribution under summation and conditioning offers tractability and simplicity (Alodat and Al-Rawwash, 2009), besides having closure under marginal and conditional distributions (Alodat and Al-Momani, 2014; Alodat and AL-Rawwash, 2014; Zareifard et al., 2018).

The GP inherits, through its finite dimensional distributions, the aforementioned properties of the normal distribution, which serves as a strong and convenient candidate to solve numerous problems and case studies in real life. Nevertheless, traditional methodologies that assume normality sometimes fail. In practice, the normality assumption might not be appropriate to fit real data, especially in the presence of heavy tailed or skewed

behaviors. A common way to model this type of data is assuming that the random field of interest is the result of an unknown nonlinear transformation of a Gaussian random field. Nevertheless, this approach has some potential weaknesses. It may not be suitable due to the fact that the Gaussianity and the correlation in data are not invariant under transformation. Thus, considering more general models, which include many of the existing models as special cases, is a remedy for skewed data. The problem of modeling skewed data has been a subject of interest in recent years. Therefore, the question whether multivariate skewed distributions can be consistently extended to random field modeling is of interest for many spatial and spatio-temporal applications. The skew-normal distribution family extends the widely employed family of normal distributions by introducing a skewing function, and shares many statistical properties of the normal distribution family. Different families of multivariate skewed distributions have been proposed in the literature, for more information see Alodat and AL-Rawwash (2014), Azzalini and Valle (1996) and Mahmoudian (2017).

Another alternative to handle non-Gaussian geostatistical processes is the use of copulas (Nelsen, 2007). These separate the marginal distributions from the dependence structure, making the construction of non-Gaussian fields somewhat trivial. Prates et al. (2015) and Hughes (2015) explored the use of copulas for real data, while Bárdossy (2006) first used copulas to study groundwater quality from a geostatistical perspective. Later, other works explored this direction (Bárdossy and Li, 2008; Kazianka and Pilz, 2010, 2011), however, they all relied upon a restricted set of copula families to model the spatial process, because many copula computations are impracticable in high dimensions. Further, selecting an appropriate copula is not a trivial task. Therefore, the pair-copula construction was suggested to allow for more copula families in geostatistical processes (Gräler et al., 2010; Gräler and Pebesma, 2011). Nevertheless, this approach does not allow for a full likelihood inference.

The aim of this work is to propose a flexible family of geostatistical models that are able to properly accommodate features such as positiveness, skewness, and heavy tails. That is achieved by specifying a hierarchical model, through an augmented Poisson process (PP), that assumes any continuous joint distribution (in a wide class of distribution) for the process in a random finite collection of locations. Another advantage of this model is that the conditional measure of its infinite-dimensional remainder is a GP itself. This

feature ought to facilitate the computation in simulation and inference contexts. Inspired by its particular hierarchical structure, the proposed family is called the Poisson-Gaussian Mixture Process - POGAMP.

Another major advantage of the POGAMP is that its existence is guaranteed, despite the distribution chosen for the process at the random finite collection of locations. Its construction allows the use of scalable Gaussian methods to make inference for massive datasets feasiable.

Attempts to naively specify a process by directly assigning arbitrary distributions to the finite-dimensional distributions often lead to misspecified process, i.e., it does not define a probability measure for the infinite-dimensional process. For example, Mahmoudian (2018) showed that several Skew Gaussian Random Fields considered in the literature are ill-defined according to the Kolmogorov existence theorem (KET).

Genton and Zhang (2012) showed that a spatial random field with a multivariate skew-normal joint distribution as defined by Azzalini and Valle (1996) cannot be identified correctly with probability 1 by a single realization, even if the number of spatial locations increases to infinity. They propose a simple remedy considering the modification presented by Zhang and El-Shaarawi (2010). However, the finite-dimensional distribution of this process is not the multivariate skew-normal distribution of Azzalini and Valle (1996). Uribe-Opazo et al. (2021) discuss aspects of identification and robustness of maximum likelihood estimators.

Proposing a valid process is not a trivial task. Furthermore, strategies to establish the existence of the process may compromise its modelling capabilities. Palacios and Steel (2006) proposed a stochastic process satisfying Kolmogorov's conditions, called Gaussian-log-Gaussian model, which is based on a log-Gaussian mixing process, where the t-process is a special case. In this spatial model, however, the scale mixing introduces a potential problem with the continuity of the resulting random field. This discontinuity essentially derives from the fact that two separate locations, no matter how close, are assigned independent mixing variables.

Allard and Naveau (2007) proposed a process through the Closed-Skew Normal distribution, $CSN_{n,n}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{I}_n, \mathbf{0}, \boldsymbol{\Sigma})$, but the marginal distribution from this process depends on the scale matrix $\boldsymbol{\Sigma}$ and dimension n and is, therefore, ill-defined. Hosseini et al. (2011) developed a new process through $CSN_{n,1}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{1}_n\boldsymbol{\Sigma}^{-1/2}, \mathbf{0}, 1)$ but, even with parameter

constraints, the process is ill-defined, since it leads to a dimensional dependent marginal. In order to make it a valid process, Mahmoudian (2018) proposed a reparametrization by setting $CSN_{n,m}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \frac{\lambda}{\sqrt{1+\lambda^2}}\boldsymbol{\Sigma}^{-1/2}, \mathbf{0}, \frac{\lambda}{1+\lambda^2}\boldsymbol{I}_n\right)$. Note that several restrictions have been applied in order to make the process valid, such as setting the penultimate parameter to zero.

Zareifard and Khaledi (2013) used a similar approach to define a valid process using the Unified skew-normal (SUN) distribution, $SUN_{n,m}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta}, \boldsymbol{\nu}, \boldsymbol{\Gamma})$ (Arellano-Valle and Azzalini, 2006). The authors considered a δ-representation of the SUN distribution, such that $SUN_{n,m}(\mathbf{0}, \omega^2\boldsymbol{H}, \omega\delta\boldsymbol{H}, \mathbf{0}, \boldsymbol{H})$, but Mahmoudian (2018) showed that the finite-dimensional distributions of this process depends on the dimension and unrelated parameters. As result, the process does not satisfy the marginal consistency condition. In order to establish its validity, parameter constraints are considered, leading to $SUN_{n,m}(\boldsymbol{\mu}, \omega^2\boldsymbol{H}, \delta\boldsymbol{H}^{1/2}, \mathbf{0}, \boldsymbol{I}_n)$, where $\bar{\boldsymbol{\Sigma}} = \omega^2\bar{\boldsymbol{H}}$. Also, Mahmoudian (2017) proposed a Skew-Gaussian Process for which the existence depends on specifying a Skew-Gaussian distribution family that is closed under marginalization and has symmetry under permutation and consistency.

Another non-Gaussian process was proposed by Bevilacqua et al. (2021). The authors formulated a stationary process with Student-t marginals obtained through scale mixing of a Gaussian process with an inverse square root process having Gamma marginals. They propose the method of weighted pairwise likelihood for the t process estimation, since they only have the closed form for the Student-t distribution in the bivariate case. Furthermore, the t process is well-defined only for degrees of freedom parameter ν in $\{3, 4, \dots\}$ and for $\nu > 2$ under non-infinite divisibility of G^2 (standard GP). The authors also propose a Skew-t process, but they do not perform any inference study since the bivariate distribution in this case is complicated. Tagle et al. (2020) proposed a spatial Skew-t model that relies on a partition of the spatial domain, where a common latent process across each region of the partition is assumed and estimated, through approximated inference. Their model has discontinuities at the boundaries and assumes conditional independence among points for different regions.

Recently, new directions continue to appear as alternatives to model non-Gaussian data. For example, Wang et al. (2019) overcame the traditional linearity assumption of kriging by using neural networks, while Zheng et al. (2021) combined copulas with scalable

Gaussian processes to handle massive non-Gaussian data.

The literature review on non-Gaussian geostatistical processes presented above motivates the development of a process that can depart from Gaussianity and still have considerable freedom to define the desirable finite-dimensional distributions. This is the main contribution offered by the POGAMP from a modelling perspective. Its unified approach allows the definition of a valid process whose finite-dimensional distributions can be arbitrarily similar to any valid continuous distribution. Some appealing choices are: Skew-normal, Student-t, Skew-t, Gamma.

From an inference perspective, this thesis offers contribution to the literature by proposing a not trivial MCMC algorithm to perform Bayesian inference, based on the observation of a POGAMP in a finite collection of locations, in a compact domain S, typically in $\mathbb{R}^2$. The algorithm consists of an infinite-dimensional Markov chain that converges to the posterior distribution of all the unknown quantities of the model. The infinite-dimensionality of the chain is due to the same property of the POGAMP and requires the use of non-standard simulation techniques and MCMC updates to devise a valid, exact and efficient MCMC algorithm. The term 'exact' refers to the property that the limiting distribution of the MCMC chain is the exact (infinite-dimensional) target posterior distribution. In particular, exact simulation of infinite-dimensional Markov chains is achieved by a neat simulation technique called retrospective sampling.

This thesis is organized as follows. Chapter 2 presents the POGAMP, establishes its existence and provides results related to some of its important properties. Chapter 3 presents the MCMC algorithm to perform Bayesian inference for discretely observed POGAMPs. Simulated studies is performed in order to investigate the behavior of the models, its flexibility and several properties; the efficiency of the MCMC algorithm is investigated in simulated examples; and the applicability of the proposed methodology through the analysis of a real datasets are presented in Chapter 4. Finally, some final remarks and directions for future work are presented in Chapter 5.

Chapter 2

A novel family of geostatistical models

2.1 The Poisson-Gaussian Mixture Process

The Poisson-Gaussian Mixture Process is hierarchically defined through conditional and marginal measures. This not only allows for a tractable way to define the model, but also gives good intuition about its dynamics and a clear interpretation of its components. The definition of a joint probability measure using a decomposition into conditional and marginal measures, however, does not guarantee the existence of the former. The existence of the joint measure needs to be established in order to yield the existence of the marginal geostatistical process of interest. The formal definition of the POGAMP is provided below in Definition 2.2 and its existence is established in Theorem 2.1 and Corollary 2.1. Before defining the POGAMP, we present the definition of Gaussian processes.

A Gaussian process (GP) is an uncountable collection of r.v.'s for which the joint distribution of any finite collection of those variable is a multivariate Normal distribution.

Definition 2.1. *A Gaussian process Y is a stochastic process in some space $S \in R^d$ such that, $\forall\ n > 1$ and $(s_1, \ldots, s_n) \in S^n$, we have that*

$$(Y(s_1), \ldots, Y(s_n)) \sim N_n(\mu(s_1, \ldots, s_n), \Sigma(s_1, \ldots, s_n)),$$

where $\mu(s_1, \ldots, s_n) \in \mathbb{R}^d$ and $\Sigma(s_1, \ldots, s_n)$ is a valid covariance matrix.
If μ is a constant vector and $cov(Y(s_i), Y(s_i)) = \sigma^2$, $\forall\ i$, and $cov(Y(s_i), Y(s_j)) = \sigma^2\rho(|s_i - s_j|)$, $\forall\ i \neq j$, the Gaussian process is called stationary and isotropic.

The main contribution of this thesis lies on the proposal of the POGAMP process defined in Definition 2.2 below and on the establishment of its existence. In the course of

this chapter, some important properties of the POGAMP are presented and the theoretical and practical implications of those properties are discussed.

Definition 2.2. ***The Poisson-Gaussian Mixture Process - POGAMP***. *Let* (Y, N), *for* $Y = \{Y(s), s \in S \subset \mathbb{R}^2\}$, *be a coordinate process on* $(\mathcal{Y} \times \mathcal{N}, \mathcal{B}(\mathcal{Y}) \otimes \mathcal{B}(\mathcal{N}))$, *where* $\otimes$ *is the tensor product of algebras,* $\mathcal{Y}$ *is the Banach space of continuous functions in* S *(a compact set) and* $\mathcal{B}(\mathcal{Y})$ *is the Borel* σ*-algebra generated by the open sets of* $\mathcal{Y}$ *in the strong topology.* $\mathcal{N}$ *is the locally compact separable metric space of point patterns in* S, *such that no two points are in the same location and the number of points is finite, with* σ*-algebra* $\mathcal{B}(\mathcal{N})$. *Define* $\mathcal{P}$ *as a candidate probability measure on* $(\mathcal{Y} \times \mathcal{N}, \mathcal{B}(\mathcal{Y}) \otimes \mathcal{B}(\mathcal{N}))$ *under which the coordinate process* (Y, N) *is such that:*
i) N *is a Poisson process on* S *with non-negative intensity function* $\lambda := \{\lambda(s), s \in S\}$ *and define its* $|N|$ *events as* $S_N := \{\dot{s}_1, \ldots, \dot{s}_{|N|}\}$;
ii) conditional on N, $Y_N := (Y(\dot{s}_1), \ldots, Y(\dot{s}_{|N|}))$ *has Lebesgue probability density function (pdf)* f_N, *such that* f_N *is continuous on* $(S_N, Y_N) \in (S^{|N|} \times \mathcal{Y}_N)$, *where* $\mathcal{Y}_N$ *is the support of* f_N;
iii) conditional on (N, Y_N), Y *is a Gaussian Process with mean* μ *and covariance function* Σ.

The spatial process Y is defined hierarchically through the augmented Poisson process N. Conditional on its existence, the former is considerably general and flexible, given the flexibility to specify the density f_N and the intensity function λ. The Gaussian process in *iii* is called the base GP.

We shall define each model in this family based on the choice of the class of distributions f that defines f_N for each possible value of N. For example, one may consider f to be a multivariate Skew-t distribution with isotropic covariance function.

The components N and f have clear purposes in the definition of the POGAMP. The distribution f should have the features desired for the finite dimensional distributions of the Y process that, in particular, are not featured by the normal distribution; for example, skewness and heavy tails. Process N stochastically defines how close the finite dimensional distributions (fdd) of Y should be to f and how this similarity should vary across the region S. Intuitively, the fdd's of Y are some sort of mixture between f and a normal distribution and the intensity function λ regulates the "weight" of each of the two distributions in the mixture, which may vary continuously across S. This means that, for

a smaller λ, Y gets closer to the GP characteristics, however, as λ increases, Y gets closer to the f characteristics.

Finally, note that the definition of the POGAMP can be extended for $S \subset \mathbb{R}^d$, $d \in \mathbb{N}$. Nevertheless we will focus on the case $d = 2$, as this is the case in most of the geostatistical applications.

Theorem 2.1. ***Existence of the POGAMP***. *Suppose that N is a homogeneous Poisson process with rate $\lambda > 0$. Then, $\mathcal{P}$, as defined in Definition 2.2, is a probability measure on $(\mathcal{Y} \times \mathcal{N}, \mathcal{B}(\mathcal{Y}) \otimes \mathcal{B}(\mathcal{N}))$, which implies that Y (under $\mathcal{P}$) is a valid stochastic process.*

The existence of the process Y when N is a non-homogeneous Poisson process is established by the following corollary.

Corollary 2.1. *$\mathcal{P}$ as defined in Definition 2.2 is a probability measure on $(\mathcal{Y} \times \mathcal{N}, \mathcal{B}(\mathcal{Y}) \otimes \mathcal{B}(\mathcal{N}))$.*

In order to prove Theorem 2.1, we present the following Lemma.

Lemma 2.1. *Let $(\Omega_1, \mathcal{F}_1, \mu_1)$ be a probability space such that, for each $\omega_1 \in \Omega_1$, there exists a probability measure μ_{2,ω_1} on $(\Omega_2, \mathcal{F}_2)$. Consider the joint measurable space $(\Omega, \sigma(\mathcal{F}))$, where $\Omega = \Omega_1 \times \Omega_2$ and $\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2$, and suppose that*

$$g_A(\omega_1) = \int_{\Omega_2} \mathbb{1}_{[(\omega_1,\omega_2)\in A]} d\mu_{2,\omega_1}(\omega_2), \text{ is } \mathcal{F}_1\text{-measurable, } \forall A \in \mathcal{F}.$$

Then, there exists a probability measure μ on $(\Omega, \sigma(\mathcal{F}))$ that satisfies

$$\mu(A) = \int_{\Omega_1} \int_{\Omega_2} \mathbb{1}_{[(\omega_1,\omega_2)\in A]} d\mu_{2,\omega_1}(\omega_2) d\mu_1(\omega_1), \quad \forall A \in \mathcal{F}, \quad \omega_1 \in \Omega_1 \quad \text{and} \quad \omega_2 \in \Omega_2,$$

and μ is called the joint measure.

Lemma 2.1 provides the conditions required for the pair composed of a marginal and a conditional probability measures to define a valid joint probability measure. The Lemma considers a general framework, unlike the versions found in the literature which only consider finite-dimensional real measurable spaces. The proofs of Lemma 2.1, Theorem 2.1 and Corollary 2.1 are presented in Appendix A.

Let $\mathcal{G}$ be the probability measure on $(\mathcal{Y} \times \mathcal{N}, \mathcal{B}(\mathcal{Y}) \otimes \mathcal{B}(\mathcal{N}))$ that differs from $\mathcal{P}$ on the distribution of N and $(Y_N|N)$ such that, under $\mathcal{G}$, N is a unit rate homogeneous PP on S and $(Y_N|N)$ has the normal distribution induced by the same Gaussian process that

defines the measure of $Y|Y_N$ under $\mathcal{P}$. We shall refer to the process (Y, N) under $\mathcal{G}$ as the augmented GP, since the induced marginal measure of Y is the base GP from Definition 2.2.

Corollary 2.2. *Consider the probability measures $\mathcal{P}$ and $\mathcal{G}$ as previously defined in Definition 2.2. Then $\mathcal{P}$ is absolutely continuous with respect to (w.r.t.) $\mathcal{G}$ with Radon-Nikodym derivative given by $\frac{f}{g}(Y_N)$, where g is the respective density under $\mathcal{G}$.*

The Corollary 2.2 has important consequences, such as its base GP almost surely (a.s.) convergence implies POGAMP a.s. convergence. Therefore, interesting features of the GP are inherited, such as the continuity and differentiability of the generated surfaces. Another important consequence of this Corollary is its implications for the derivation and validity of the MCMC algorithm to be proposed in the next chapter.

Corollary 2.3. *The Kullback Lieber divergence between the POGAMP and the augmented GP, when the N is a homogeneous PP, is given by*

$$D_{KL}(\mathcal{P} \parallel \mathcal{G}) = \sum_{N=0}^{\infty} \frac{\mathrm{e}^{-\lambda\mu(\mathrm{S})}\lambda^{\mathrm{N}}}{N!} \int_{S_N} \int_{Y_N} \log \left[\frac{f}{g}(Y_N) \right] f(Y_N)\, dY_N\, dS_N$$

The proofs of the corollaries 2.2 and 2.3 are given in Appendix A.

As mentioned before, one of the main advantages of the POGAMP is that its finite-dimensional distributions inherit the characteristics of some chosen class of multivariate distribution f and is a valid process under some very mild conditions. This idea is formally described in the following two results.

Theorem 2.2. *Consider a monotonic sequence of positive real numbers $\{\lambda_n\}_{n=1}^{\infty}$ such that $\lambda_n \uparrow \infty$ and a set of any r locations in S, for $r \in \mathbb{N}$. Let $Y_{n,r}$ be a sequence of the POGAMPs at those r locations such that, for each $n \in \mathbb{N}$, the rate of the PP N of the n-th POGAMP is λ_n. Then, $Y_{n,r} \xrightarrow{d} Y_{f,r}$, where $Y_{f,r}$ is a r-dimensional random variable with distribution f at the r specified locations.*

Corollary 2.4. *Suppose that, for a given class of multivariate distributions f, the stochastic process Y with finite-dimensional distributions defined by f exists. Then, the POGAMP Y_n, with rate $\lambda_n \uparrow \infty$, is such that $Y_n \xrightarrow{d} Y$. Conversely, if f does not define a valid process, there exists no limit for the sequence Y_n.*

Note that, in the cases that f defines a valid process, the POGAMP can get arbitrarily close to this by increasing the rate λ, with the advantage of having the tractability of the Gaussian measure. Furthermore, even if f does not define a valid process, the POGAMP can still have f.d.d.'s arbitrarily close f. In particular, the POGAMP allows us to perform valid likelihood-based inference, given that its existence has been proved.

Example 1. *Suppose that we are interested in proposing a process that has the characteristics of the Skew-t distribution, that is, we want any finite-dimensional distribution to be similar to a multivariate skew-t. Therefore, one way to obtain such a process taking advantage of the POGAMP would be to define f as the desired Skew-t. Consider, for example, the following two scenarios.*

- *If the PP has a considerably small λ, which results in a small number of points of $|N|$, the resulting process will be similar to the base GP.*
- *However, if we define a sufficiently large λ, which results in a large number of points of $|N|$, the finite-dimensional distributions of the resulting POGAMP will be similar to f.*

The Skew-t case will be considered as an example in Chapter 4 to illustrate the definition of the POGAMP. In particular, we shall consider different values of λ.

2.2 The marginal POGAMP Y

We now explore the marginal distribution of Y at a finite collection of locations $\mathbf{s} = \{s_1, \ldots, s_r\}$, defined as Y_r. Under the POGAMP measure, we have that $(Y_r \mid Y_N, S_N)$ has the multivariate normal distribution induced by the base GP.

Defining $\lambda_S = \int_S \lambda(s)\mathrm{d}s$, the marginal Lebesgue density of Y_r is given by

$$
\begin{aligned}
p(Y_r) &= \int p\left(N, Y_N, Y_r\right) \mathrm{d}\mathcal{P}\left(N, Y_N\right) = \int p\left(|N|, S_N, Y_N, Y_r\right) \mathrm{d}\mathcal{P}\left(|N|, S_N, Y_N\right) \\
&= \int p(|N|)p(S_N \mid |N|)f(Y_N \mid S_N, |N|)p(Y_r|Y_N, S_N)\mathrm{d}\mathcal{P}\left(|N|, S_N, Y_N\right) \qquad (2.1) \\
&= \sum_{|N|=0}^{\infty} \frac{\mathrm{e}^{-\lambda_S}\left[\lambda_S\right]^{|N|}}{|N|!} \int_{S^{|N|}} p(S_N \mid |N|) \int_{\mathbb{R}^{|N|}} f(Y_N \mid S_N, |N|)p(Y_r \mid Y_N, S_N)\mathrm{d}Y_N\mathrm{d}S_N \\
&\propto \sum_{|N|=0}^{\infty} \frac{\left[\lambda_S\right]^{|N|}}{|N|!} \int_{S^{|N|}} p(S_N \mid |N|) \int_{\mathbb{R}^{|N|}} f(Y_N \mid S_N, |N|)p(Y_r \mid Y_N, S_N)\mathrm{d}Y_N\mathrm{d}S_N \\
&= \sum_{|N|=0}^{\infty} \frac{\left[\lambda_S\right]^{|N|}}{|N|!} c_{|N|} g^*(Y_r; |N|), \qquad (2.2)
\end{aligned}
$$

where

$$
g^*(Y_r; |N|) = \frac{g(Y_r; |N|)}{\int_{\mathbb{R}^n} g(Y_r; |N|)\mathrm{d}Y_r} = \frac{g(Y_r; |N|)}{c_{|N|}},
$$

$$
g(Y_r; |N|) = \int_{S^{|N|}} p(S_N \mid |N|) \int_{\mathbb{R}^{|N|}} f(Y_N \mid S_N, N)p(Y_r \mid Y_N, S_N)\mathrm{d}Y_N\mathrm{d}S_N.
$$

Note that $p(Y_r \mid Y_N, S_N)$ is a normal distribution induced by the base GP. If N is a homogeneous PP, we have that $\lambda_S = \lambda\mu(S)$, where λ is the rate and $\mu(S)$ is the area of S, and $(S_N \mid |N|)$ are $|N|$ iid Uniform rv's on S. If N is an inhomogeneous PP, $p(S_N \mid |N|) = \prod_{j=1}^{|N|} \frac{\lambda(\dot{s}_j)}{\int_S \lambda(s)\mathrm{d}s}$. Finally, note that the marginal density of Y_r is typically intractable.

2.3 Symmetry of the POGAMP

We now state an interesting property regarding symmetry of the POGAMP. First, we consider the following definitions.

Definition 2.3. Symmetry w.r.t. S. *Suppose that S is symmetric and let* $\mathbf{s}$ *and* $\mathbf{s}'$ *be two sets of k locations in S, for any $k \in \mathbb{N}$. We say that* $\mathbf{s}$ *and* $\mathbf{s}'$ *are symmetric w.r.t. S if there exists a rotation of $\mathbb{R}^2$ that does not change S and under which* $\mathbf{s}$ *in the rotated space is equal to* $\mathbf{s}'$ *in the non-rotated one.*

Definition 2.4. Symmetry w.r.t. N. *Suppose that S is symmetric and let* $\mathbf{s}$ *and* $\mathbf{s}'$ *be two sets of k locations in S, for any $k \in \mathbb{N}$. We say that* $\mathbf{s}$ *and* $\mathbf{s}'$ *are symmetric w.r.t. the Poisson process N if $p(\mathbf{s}) = p(\mathbf{s}')$, under the probability measure of N, where p is the density of N w.r.t. some dominating measure.*

We shall refer to sets that are symmetric w.r.t S and N as doubly symmetric.

Proposition 2.1. *Suppose that f is stationary and S is symmetric. Then, for any two sets $\mathbf{s}$ and $\mathbf{s}'$ in S that are doubly symmetric, we have that $Y(\mathbf{s}) \stackrel{d}{=} Y(\mathbf{s}')$, under the POGAMP measure.*

The proof of the Proposition 2.1 is presented in Appendix *A*.

We now present three examples of the symmetry property of the POGAMP. In Figure 2.1a, N is a homogeneous PP and the four sets of points defined by the different colours are doubly symmetric between themselves, note that the space is symmetrical since the four datasets have identical distances, furthermore the four sets are under the same rate of PP λ. In Figure 2.1b, the PP N is piecewise constant. The blue and red sets are doubly symmetric (under the PP rate λ_1) and so are the black and green ones (under the PP rate λ_2). In Figure 2.1c, the PP rate is proportional to a bivariate normal density function with symmetry at the centroid. In this case, the four sets of points defined by the different colours are doubly symmetric between themselves, as in the first case, since all sets are under the same value of λ (λ_5).

2.4 Covariance function of Y

We now express the covariance function of Y for the general POGAMP and for the particular (and appealing) case where the base GP and the f distribution have the same covariance function.

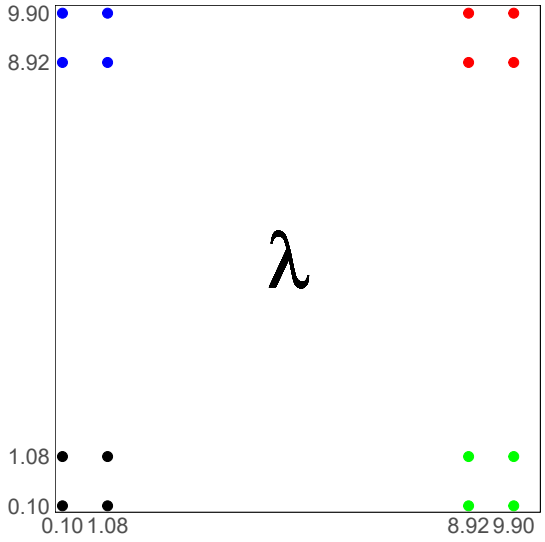
Proposition 2.2. *Let s_1 and s_2 be two arbitrary locations in S and define $\dot{s}_i$ as the i-th location from N. Now let Σ_{iN} be the row vector of the covariances between $Y(s_i)$ and $Y(\dot{s}_j)$ under the base GP, for $j = 1, \ldots, N$, for $i = 1, 2$, and let $\Sigma_{NN,f}$ be the covariance matrix of Y_N under the f distribution. Then,*

$$Cov(Y_1, Y_2) = \mathrm{E}[Cov(Y_1, Y_2|Y_N)] + \mathrm{E}[\Sigma_{1N}\Sigma_{NN}^{-1}\Sigma_{NN,f}(\Sigma_{2N}\Sigma_{NN}^{-1})^T]. \tag{2.3}$$

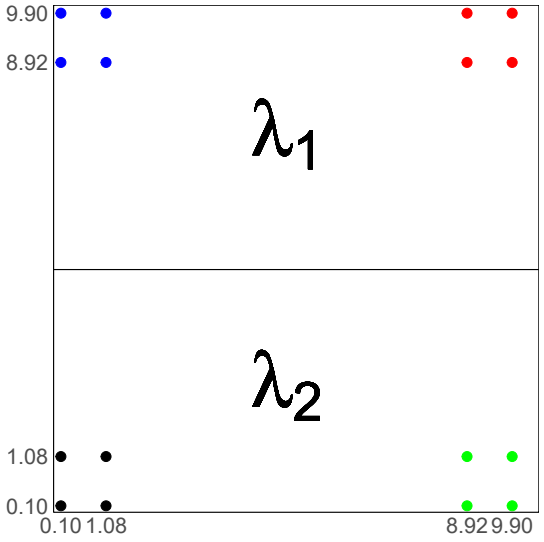
If, additionally, the covariance function is the same under the base GP and f, we have that

$$Cov(Y_1, Y_2) = \rho(s_1, s_2) \tag{2.4}$$

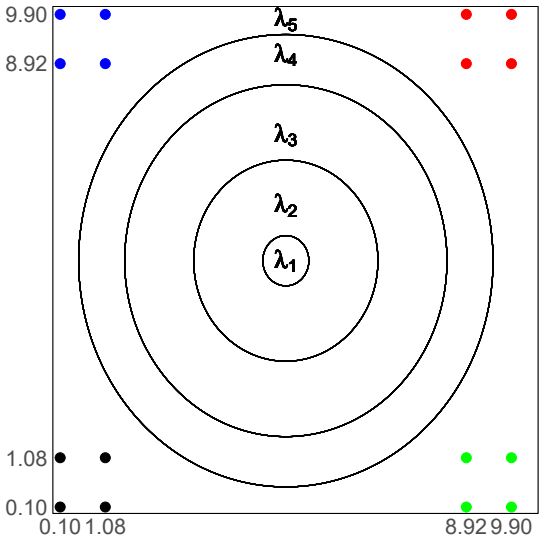
where $\rho(s_1, s_2)$ is the covariance of $(Y(s_1), Y(s_2))$ under the base GP.



(a)



(b)



(c)

Figure 2.1: Symmetry examples of the POGAMP for different λ's

The proof of the Proposition 2.2 is presented in Appendix *A*.

The second result in Proposition 2 has important practical implications as it provides a tractably analytical representation of the covariance function of the POGAMP, allowing for a clear interpretation of this.

This chapter presented a detailed presentation of the POGAMP theoretical results. The results that were presented and proved are of extreme importance to the work, as they guarantee the existence of a highly flexible process, allowing the user to choose any pdf that has the desired characteristics for the process. That is, if one wants to model data that have heavy tail characteristics, one can define a student-t as the f (with the desired degree of freedom) for the POGAMP, and thus have a process where any finite collection of points observed will have the characteristics of this chosen f. Note that the flexibility presented here is not limited to the heavy tail or skewness, but to any desired characteristic, depending only on the choice of f and how much POGAMP will approach it, depending on the intensity of λ of the PP .

The following chapter will present the MCMC developed to perform inference on POGAMP model.

Chapter 3

MCMC

Bayesian methods provide a complete paradigm for statistical inference under uncertainty. The Bayesian paradigm is based on an interpretation of probability as a rational conditional measure of uncertainty, there are several references that address this topic more widely (see Robert, 2007).

The development of Bayesian Statistics is centered around the development of computational methods, required to perform inference under the Bayesian paradigm. In particular, Markov chain Monte Carlo (MCMC) has become the workhorse of Bayesian inference in most cases where analytical tractability is unfeasible. MCMC algorithms are based on the construction of Markov chains which have a posterior distribution as its limiting distribution. The most popular MCMC algorithms are the Gibbs Samplings and the Metropolis-Hastings. These are quite general and flexible and provide straightforward efficient solutions in many cases.

The Gibbs sampling, introduced by Geman and Geman (1993) and Gelfand and Smith (1990) in a statistical context, is an extremely useful tool in solving problems involving the estimation of more than one parameter. It is based on a Markov chain that is updated using the full conditional distributions of the coordinates of the chains. The algorithm offers great flexibility in terms of coordinates blocking and update schemes. For details, see Gamerman and Lopes (2006).

When it is not possible to sample directly from the full conditional distribution of a block in the Gibbs sampling, the Metropolis-Hastings algorithm is a good alternative. The algorithm proposes a move from a suitably chosen proposal distribution and accepts it with a probability that preserves the detailed balance property. Gaussian random

walks are generally suitable to be used as proposal distributions. Optimal tuning results (Roberts et al., 1997) suggest that the variance of the random walk proposal should vary from around 0.44 (for unidimensional updates) to around 0.234 (for high dimensions - $\geqslant 5$).

Performing statistical inference for infinite-dimensional models - when the dimension of the unknown quantities of the model (parameters and latent variables) is infinitely uncountable, is a highly complex problem. For many years, solutions had to rely on discrete (finite-dimensional) approximations of those quantities. This naturally represents a significant source of error in the analysis and this error is typically hard to be measured and/or controlled. The advances in computational methods, specially Monte Carlo under a Bayesian approach, have brought a new perspective to deal with infinite-dimensional problems by allowing for an analysis where no discretization error is involved, only Monte Carlo one. The latter is much simpler to quantify and control, resulting in more precise and computationally efficient analyses.

Exact inference solutions for infinite-dimensional problems are possible mainly due to a neat simulation technique called retrospective sampling. It basically allows to deal with infinite-dimensional random variables by unveiling only a finite-dimensional representation of this which has two main properties: i) it is enough to know its value in order to execute the steps of the algorithm in context, for example, MCMC; ii) any finite-dimensional part of the infinite-dimensional remainder of that r.v. can be simulated conditional on it. The idea of retrospective sampling in the context of simulation of infinite-dimensional r.v.'s was introduced in Beskos and Roberts (2005) to perform exact simulation of diffusion paths. It was later used in a statistical context in several works (see, for example, Beskos et al. (2006) and Gonçalves and Gamerman (2018)).

This thesis proposes an infinite-dimensional MCMC algorithm that relies on retrospective sampling to perform exact Bayesian inference for discretely-observed POGAMPs. In particular, the proposed algorithm consists of a Gibbs sampling with Metropolis-Hastings steps. Each step of the algorithm is carefully designed to be computationally efficient and provide feasible and reasonable solutions to the inference problem at hands.

Suppose that a POGAMP Y is observed at some finite collection of locations $s_1, \ldots, s_n$ in S. The vector of unknown quantities to be estimated is given by $\psi = (N, Y_N, Y_u, \theta_G, \theta_f, \lambda)$, where $Y_u = Y \backslash (Y_N, Y_o)$ and $Y_o = (Y_{s_1}, \ldots, Y_{s_n})$. Vectors θ_G and θ_f are the sets of param-

eters indexing the base GP and f, respectively.

Under the Bayesian paradigm, inference about ψ is based on the posterior distribution of $(\psi \mid Y_o)$, which has a density, w.r.t. a suitable dominating measure, proportional to

$$\pi(Y_o, \psi) = \pi_{GP}(Y_u \mid Y_N, Y_o, \theta_G)\pi_{GP}(Y_o \mid Y_N, \theta_G)\pi_f(Y_N \mid \theta_f, N)\pi(N \mid \lambda) \times \pi(\lambda, \theta_f, \theta_G). \tag{3.1}$$

We devise a Gibbs sampling MCMC with the following blocks

$$N, \quad Y_N, \quad \lambda, \quad \theta_G, \quad \theta_f, \quad Y_u,$$

and standard probability theory implies that all the full conditional densities are proportional to (3.1).

The infinite-dimensionality of the algorithm is due to the same property of the component Y_u. As it was mentioned before, retrospective sampling is employed so that the Markov chain can be simulated without any approximations by unveiling Y_u only at a finite (though random) collection of locations at each iteration of the Gibbs sampler. Finally, blocks Y_N and θ_G are sample via collapsed Gibbs sampling (see Liu, 1994) by integrating out Y_u.

3.1 Sampling Y_N

By integrating out Y_u, we have that

$$\pi(Y_N \mid \cdot) \propto \pi_{GP}(Y_o \mid Y_N, N, \theta_G)\pi_f(Y_N \mid N, \theta_f). \tag{3.2}$$

We cannot sample directly from the density above, so Y_N is sampled in a Metropolis-Hastings step. For a proposal distribution $q(Y_N^*; Y_N)$, the acceptance probability of a move $Y_N \to Y_N^*$ is given by

$$\alpha(Y_N, Y_N^*) = 1 \wedge \frac{\pi_{GP}(Y_o \mid Y_N^*, N, \theta_G)\pi_f(Y_N^* \mid N, \theta_f)q(Y_N|Y_N^*)}{\pi_{GP}(Y_o \mid Y_N, N, \theta_G)\pi_f(Y_N \mid N, \theta_f)q(Y_N^*)|Y_N}. \tag{3.3}$$

We propose two options for the proposal distribution q. The first one consists of the base Gaussian process conditional on the observations Y_o, i.e., $q(Y_N^*|Y_N)q(Y_N^*) = \pi_{GP}(|Y_o, N, \theta_G)$. Simplifications in (3.3) lead to the following acceptance probability:

$$\alpha(Y_N, Y_N^*) = 1 \wedge \frac{\pi_f(Y_N^* \mid N, \theta_f)}{\pi_{GP}(Y_N^* \mid N, \theta_G)} \frac{\pi_{GP}(Y_N \mid N, \theta_G)}{\pi_f(Y_N \mid N, \theta_f)}. \tag{3.4}$$

The second option for the proposal distribution is a Gaussian random walk, properly tuned to have an acceptance rate of approximately 0.234 (see Roberts et al., 1997), assuming that the dimension $|N|$ is greater than 5. Simplifications in (3.3) lead to the following acceptance probability:

$$\alpha(Y_N, Y_N^*) = 1 \wedge \frac{\pi_f(Y_N^* \mid N, \theta_f)\pi_{GP}(Y_o \mid Y_N^*, N, \theta_G)}{\pi_f(Y_N \mid N, \theta_f)\pi_{GP}(Y_o \mid Y_N, N, \theta_G)}. \tag{3.5}$$

Since the locations N change along the MCMC chain, it is not possible to use an empirical covariance matrix of the chain in the proposal random walk. Instead, we update this matrix at some fixed iterations of chain and up to a certain point by adopting a multiple of the correlation matrix under f, $(\boldsymbol{\Sigma_f})$, for values of θ_f given by the its average over the last few hundred iterations. The multiplication factor is chosen so to tune the acceptance rate accordingly.

3.2 Sampling N

From 3.1, we have that the full conditional density of N is proportional to

$$\pi(N \mid \cdot) \propto \pi_{GP}(Y_u \mid Y_N, Y_o, \theta_G)\pi_{GP}(Y_o \mid Y_N, N, \theta_G)\pi_f(Y_N \mid \theta_f, N)\pi(N \mid \lambda). \tag{3.6}$$

We adopt a proposal distribution that is invariant w.r.t. $\pi(N \mid \lambda)$ and proposes one of the two types of move:

$$\begin{aligned} &\text{w.p. } \frac{\lambda\mu(S)}{\lambda\mu(S) + |N| + 1}, \text{ add one location uniformly distributed in } S; \\ &\text{w.p. } \frac{|N|}{\lambda\mu(S) + |N|}, \text{ for } N > 0, \text{ remove one uniformly chosen location;} \end{aligned} \tag{3.7}$$

For large values of λ, this proposal is bound to lead to a poor mixing of the chain. In order to mitigate this problem, we divide S into K regular squares and use that same proposal in each of those. The value of K is chosen empirically in terms of the mixing properties and computational cost of the algorithm.

The acceptance probability of a move $N_k \to N_k^*$ in each sub-region k is given by

$$\alpha(|N|_k, |N|_k + 1) = 1 \wedge \frac{\pi_f(Y_{N_k^* \setminus N_k} \mid Y_{N_{\setminus k}}, Y_{N_k}, \theta_f)}{\pi_{GP}(Y_{N_k^* \setminus N_k} \mid Y_{N_{\setminus k}}, Y_{N_k}, \theta_G)}, \tag{3.8}$$

$$\alpha(|N|_k, |N|_k - 1) = 1 \wedge \frac{\pi_{GP}(Y_{N_k \setminus N_k^*} \mid Y_{N_{\setminus k}}, Y_{N_k^*}, \theta_G)}{\pi_f(Y_{N_k \setminus N_k^*} \mid Y_{N_{\setminus k}}, Y_{N_k^*}, \theta_f)}, \tag{3.9}$$

where $Y_{N_{\setminus k}}$ is Y_N in all the other $K-1$ sub-regions and N_k is N restricted to sub-region k and has $|N|_k$ points. The expressions above are obtained by noticing that the prior of N_k is a $PP(\lambda)$ in sub-region k and the proposal distribution is invariant w.r.t. this prior. More details about the construction of the acceptance probabilities (3.8) and (3.9) can be found in **??**.

Whenever a move of the type $|N|_k \to |N|_k + 1$ is proposed, the value of Y at the new location s_k is sampled retrospectively from $\pi_{GP}(Y(s_k)|Y_N, Y_o, \theta_G)$. On the other hand, whenever a move of the type $|N|_k \to |N|_k - 1$ is proposed and accepted, a virtual step is performed to simply erase the value of Y at the removed location. Theoretically, this step consists of updating Y at that location, retrospectively, from its full conditional distribution $\pi_{GP}(\cdot \mid Y_N, Y_o, \theta_G)$. Note the infinite dimensionality of the MCMC chain together with the retrospective sampling of Y_u avoids the use of a (complicated) reversible jump step to update N.

Since then, this block has become the most expensive of the entire MCMC, because on each step it is possible to add or remove a point in each of the k parts of the S space, so we have two points to consider in this scenario . The first point refers to the proposal of a new point $(\pi_{GP}(Y(s_k)|Y_N, Y_o, \theta_G))$, where it is necessary to obtain the inverse of the joint covariance matrix of the locations of the observed points and N. Therefore, in each of the k parts of the space it would be necessary to calculate a new inverse of this matrix as long as a point was added or removed. And if a point from N is added or removed, we enter the second point mentioned above, since it becomes necessary to recalculate the covariance matrices of both the GP and the f in the N locations, then it is possible to perform the N acceptance probability calculations (3.8 and 3.9).

In order to overcome these issues, we have developed two approaches that allow us to update the inverse matrices when we add/remove one dimension.

When a point is added (and this added point is always in the last entry of the matrix), the inverse is updated using the Schur complement. Suppose that in the current step we have already calculated the inverse of the covariance $(\bar{\Sigma^{-1}})$ with dimension $n \times n$, then if we add a single dimension, we are adding an integer (d) in the input $(n+1) \times (n+1)$, that is, we will have a new inverse Σ^{-1^*} as follows $\Sigma^{-1^*} = \begin{bmatrix} \Sigma & B \\ B^T & d \end{bmatrix}^{-1}$, where B is the covariance between the current covariance points and the new one added. Then, using the Schur complement, we have

$$\Sigma^{-1^*} = \begin{bmatrix} \Sigma^{-1} + \Sigma^{-1} B \Upsilon B^T \Sigma^{-1} & -\Sigma^{-1} B \Upsilon \\ -\Upsilon B^T \Sigma^{-1} & \Upsilon \end{bmatrix}^{-1}, \text{ where } \Upsilon = (d - B^T \Sigma^{-1} B)^{-1}.$$

Note that Υ is a 1×1 matrix and since we already have Σ^{-1} calculated, the cost to obtain Σ^{-1^*} is modest.

When one point is removed, we apply some algebra to avoid the calculation of a new inverse of the entire matrix. More details about the algebra performed above can be found in the Appendix A.

3.3 Sampling λ, θ_G and θ_f

Standard conjugated Bayesian analysis calculations imply that, for a prior distribution $Gamma(\alpha_\lambda, \beta_\lambda)$, the full conditional distribution of λ is a $Gamma(\alpha_\lambda + |N|, \beta_\lambda + \mu(S))$.

Both θ_G and θ_f are sampled via MH steps with properly tuned adaptive Gaussian random walk proposals (see Roberts and Rosenthal, 2009). The respective acceptance probabilities are given by

$$\alpha(\theta_G, \theta_G^*) = 1 \wedge \frac{\pi_{GP}(Y_o \mid Y_N, N, \theta_G^*)\pi(\theta_G^*)}{\pi_{GP}(Y_o \mid Y_N, N, \theta_G)\pi(\theta_G)}, \tag{3.10}$$

$$\alpha(\theta_f, \theta_f^*) = 1 \wedge \frac{\pi_f(Y_N \mid \theta_f^*, N)\pi(\theta_f^*)}{\pi_f(Y_N \mid \theta_f, N)\pi(\theta_f)}, \tag{3.11}$$

for suitably chosen priors $\pi(\theta_G)$ and $\pi(\theta_f)$.

When the parameters θ_G and θ_f have the same interpretation and are chosen to be the same, i.e. $\theta_G = \theta_f := \theta$, the acceptance probability is given by

$$\alpha(\theta, \theta^*) = 1 \wedge \frac{\pi_{GP}(Y_o \mid Y_N, N, \theta^*)\pi_f(Y_N \mid \theta^*, N)\pi(\theta^*)}{\pi_{GP}(Y_o \mid Y_N, N, \theta) f(Y_N \mid \theta, N)\pi(\theta)}. \tag{3.12}$$

3.4 Prediction

Under the MCMC approach presented in the previous theoretical point of view sections to perform inference for discretely observed POGAMPs, it is straightforward to perform prediction for functions of the unobserved (infinite-dimensional) part of the process Y_u.

Let $h(Y_u)$ be some some finite-dimensional real and tractable function of Y_u. This includes, for example, the process Y at a finite collection of locations. Under the Bayesian

approach, prediction ought to be performed through the posterior predictive distribution of h, i.e., $\pi(h(Y_u)|Y_o)$. An approximate sample from this distribution can be obtained within the proposed MCMC algorithm by sampling from the respective full conditional distribution of $h(Y_u)$ at each iteration of the Gibbs sampler. From a theoretical point of view, this consist of performing the following integral via Monte Carlo.

$$\pi(h(Y_u)|Y_o) = \int \pi_{GP}(Y_u|Y_N, Y_o, \theta_G)\pi(Y_N, \theta_G|Y_o)dY_N d\theta_G. \tag{3.13}$$

More specifically, $h(Y_u)$ is simulated from is full conditional distribution by sampling Y_u at the finite collection of locations required to compute $h(Y_u)$, from the base GP, and then computing this function.

More sophisticated, still simple, simulation techniques allow for Monte Carlo estimation of some intractable functions $h(Y_u)$, for example, $h(Y_u) = \int_S g(Y_u)ds$, for tractable functions g. We define a random variable $U \sim Uniform(S)$ and an i.i.d. sample $U^{(1)}, \ldots, U^{(M)}$ of this and consider the following Monte Carlo estimator of $h(Y_u)$.

$$\hat{h}(Y_u) = \frac{1}{M}\sum_{j=1}^{M} \mu(S)g(Y(U^{(j)})^{(j)}) \tag{3.14}$$

where $\mu(S)$ is the area of S and the superscript $^(j)$ refers to the j-th values from the MCMC sample of size M. The estimator in (3.14) is justified by the following result,

$$E_U[\mu(S)g(Y(U))] = h(Y_u). \tag{3.15}$$

The estimator in (3.14) can be improved, in terms of variance reduction, by dividing S into m equal squares and applying the same idea to each of those. The final estimator is obtained by summing the m estimators.

We conclude Chapter 3, where the development of the MCMC was presented in a rigorous way, detailing the functioning of each block and some tricks used to obtain a better computational performance. In the next chapter, some results of POGAMP simulations will be presented, illustrating some important properties of the model. Simulations of the MCMC, presented in this chapter, will also be illustrated, as will a brief application of POGAMP on real data.

Chapter 4

Simulation and Application

4.1 POGAMP Simulation

We perform some simulation studies in order to explore the modelling properties of the POGAMP model. All the simulations are coded in *R* (R Core Team, 2022) and run in a 2 x Intel(R) Xeon(R) CPU E5-2690 0 2.90GHz processor with 32 Gbytes RAM, provided by *lsncs - cluster of the Numerical Simulation Laboratory of FCT/UNESP*.

We consider four classes of f distributions:

1) p-variate Skew Normal.

$$\phi_p(\mathbf{x};\boldsymbol{\xi},\boldsymbol{\Omega},\boldsymbol{\alpha}) = 2\phi_p(\mathbf{x}-\boldsymbol{\xi};\boldsymbol{\Omega})\Phi(\boldsymbol{\alpha}^\top\boldsymbol{\omega}^{-1}(\mathbf{x}-\boldsymbol{\xi})), \quad \mathbf{x}\in\mathbb{R}^p; \quad \boldsymbol{\xi}\in\mathbb{R}^p; \quad \boldsymbol{\Omega}>\mathbf{0};$$
$$\boldsymbol{\alpha}\in\mathbb{R}^p, \quad \text{where} \quad \boldsymbol{\omega} = (\boldsymbol{\Omega}\odot I_p)^{1/2}.$$

2) p-variate Gamma distribution.

$$g_p(\mathbf{x};\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\Sigma}) = c_\Phi\{G_1(x_1),\cdots,G_p(x_p)\mid\boldsymbol{\Sigma}\}\prod_{i=1}^{p} f(x_i;\alpha_i,\beta_i), \quad \mathbf{x}\in\mathbb{R}^{+^p};$$
$$\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\Sigma}>\mathbf{0}.$$

where G and g are the univariate Gamma distribution function and probability function, respectively. Also, $c_\Phi(u\mid\boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{-1/2}\exp\{-\frac{1}{2}\mathbf{q}^\top\boldsymbol{\Sigma}^{-1}\mathbf{q}+\frac{1}{2}\mathbf{q}^\top\mathbf{q}\}$ and $\mathbf{q}=(q_1,\cdots,q_p)$, with $q_i=\Phi^{-1}(u_i)$, where Φ is the standard normal c.d.f. Details on the multivariate Gamma distribution can be found in Xue-Kun Song (2000).

3) p-variate Student-t.

$$t_p(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma},\nu) = \frac{\Gamma\left[(\nu+p)/2\right]}{\Gamma(\nu/2)\nu^{p/2}\pi^{p/2}\left|\boldsymbol{\Sigma}\right|^{1/2}}\left[1+\frac{1}{\nu}(\mathbf{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]^{-(\nu+p)/2},$$

$\mathbf{x}\in\mathbb{R}^p$; $\boldsymbol{\mu}\in\mathbb{R}^p$; $\boldsymbol{\Sigma}>\mathbf{0}$; and $\nu>2$ for defined variance and mean.

4) p-variate Skew Student-t.

$$st_p(\mathbf{x};\boldsymbol{\xi},\boldsymbol{\Omega},\boldsymbol{\alpha},\nu) = 2t_p(\mathbf{x};\bar{\boldsymbol{\Omega}},\nu)\;T\left(\boldsymbol{\alpha}^\top\mathbf{x}\sqrt{\frac{\nu+p}{\nu+Q(\mathbf{x})}};\nu+p\right), \quad \mathbf{x}\in\mathbb{R}^p;$$

$\boldsymbol{\xi}\in\mathbb{R}^p$; $\boldsymbol{\Omega}>\mathbf{0}$; $\boldsymbol{\alpha}\in\mathbb{R}^p$; and $\nu>2$ for defined variance and mean.

where $Q(\mathbf{x})=\mathbf{x}^\top\bar{\boldsymbol{\Omega}}\mathbf{x}$ and $T(.;\rho)$ denotes the univariate Student-t c.d.f. with ρ degrees of freedom. Details on the Skew normal and Skew Student-t distributions can be found in Azzalini (2013).

We parametrize the Skewness of the Skew normal and Skew t distributions through the Pearson's index of Skewness γ (see Henze, 1986) such that

$$\gamma = r\delta^3(2r^2-1)(1-r^2\delta^2)^{-3/2}, \quad \gamma\in(-0.99527;0.99527), \tag{4.1}$$

where $\delta=\dfrac{\alpha}{\sqrt{1+\alpha^2}}$, $\alpha\in(-\infty,\infty)$, $\delta\in[-1,1]$, $r=\sqrt{2/\pi}$ and $s=\left(\dfrac{2}{4-\pi}\right)^{1/3}$. Furthermore, we have the following γ-parameterization

$$\delta=\frac{s\gamma^{1/3}}{r\sqrt{1+s^2\gamma^{2/3}}}, \quad \alpha=\frac{s\gamma^{1/3}}{\sqrt{r^2+s^2\gamma^{2/3}(r^2-1)}}.$$

Azzalini (1985) argues that the γ-parametrization eases likelihood-based inference procedures.

We set $S=[0,10]\times[0,10]$ and consider two values for the intensity function of the Poisson process N (0.01 and 5) combined with the following f distributions: $SN_p(\boldsymbol{\xi},\boldsymbol{\Sigma},\boldsymbol{\gamma})$, $\mathcal{G}_p(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\Sigma})$, $t_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)$ and $St_p(\boldsymbol{\xi},\boldsymbol{\Sigma},\boldsymbol{\gamma},\nu)$. The base GP and all the f distributions consider the following isotropic covariance

$$\boldsymbol{\Sigma}=\sigma^2\exp\{-\frac{1}{\zeta}||r-r'||^l\}, \quad \forall r,r'\in S,$$

and the following parameter values are adopted:

- Base GP: $\boldsymbol{\mu}=\mathbf{0}$, $\sigma^2=1$, $\zeta=10$, $l=1.5$;

- $\mathcal{G}$: $\boldsymbol{\alpha} = \mathbf{2}$, $\boldsymbol{\beta} = \mathbf{2}$, $\sigma^2 = 1$, $\zeta = 10$, $l = 1.5$;
- SN_p: $\boldsymbol{\xi} = \mathbf{0}$, $\sigma^2 = 1$, $\zeta = 10$, $\boldsymbol{\gamma} = \mathbf{0.91}$, $l = 1.5$;
- t_p: $\boldsymbol{\mu} = \mathbf{0}$, $\sigma^2 = 1$, $\zeta = 10$, $\nu = 2.1$, $l = 1.5$;
- St_p: $\boldsymbol{\xi} = \mathbf{0}$, $\sigma^2 = 1$, $\zeta = 10$, $\boldsymbol{\gamma} = \mathbf{0.91}$, $\nu = 2.1$, $l = 1.5$.

For each f, 50 thousand iid replications of the respective process are generated to compute some Monte Carlo estimates. The Y process is sampled at 121 regularly spaced locations.

4.1.1 Results

We use the results from the simulations to investigate and illustrate some properties of the POGAMP models such as marginal stationarity and isotropy.

Figure 4.1 shows the 121 fixed locations where the processes are sampled in all the 50 thousand iid replications. The colored geometric shapes in the Figure 4.1 represent the points where the analyzes were performed. Note that each point has an explicit coordinate on the axes, they will be important to better understand where each analysis is referring to.

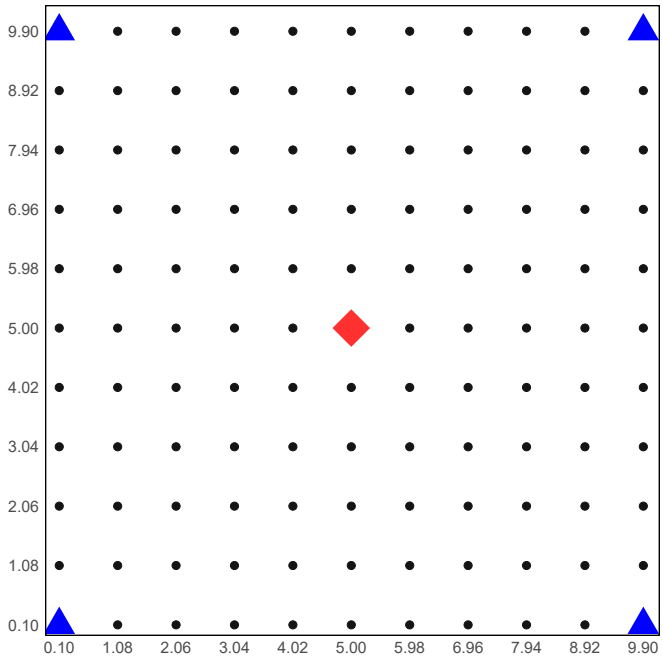


Figure 4.1: Locations used for the empirical analyses.

We compare the marginal densities at the four blue triangle locations in Figure 4.1 for each model. Figure 4.2 shows those marginal densities for each model with $\lambda = 5$. Each curve represents the density referring to the 50 thousand iid replications at each point (blue triangles). Results illustrate the symmetry result stated in Proposition 2.1, since all the lines are practically overlapping in all the four different pdf scenarios.

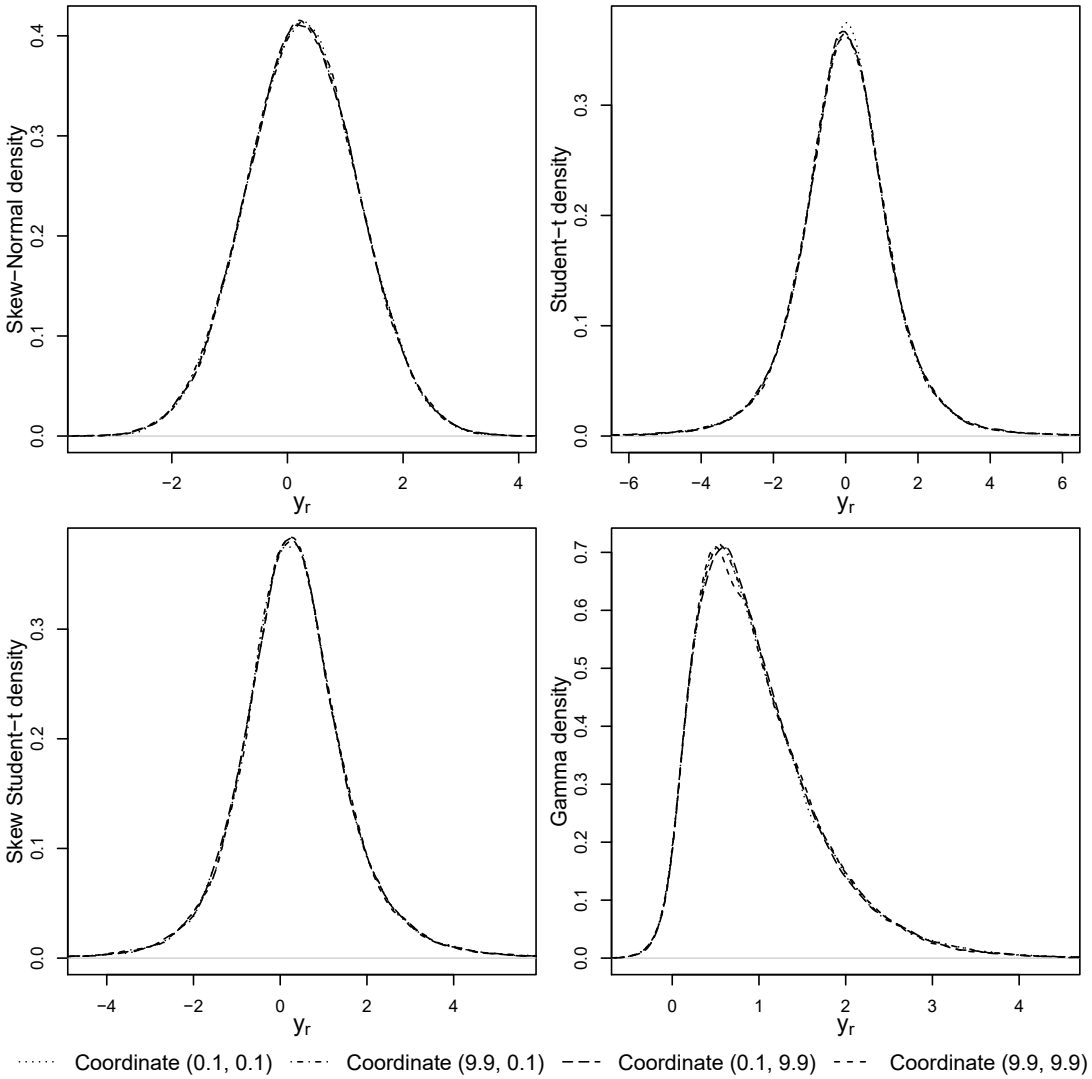


Figure 4.2: Marginal distributions at blue triangles locations

We compare the marginal density at the centroid (red losangle in Figure 4.1) for two different values of λ (0.01 and 5) for each model. Figure 4.3 shows how the marginal distribution approaches f as λ increases, illustrating the result in Theorem 2.2. Note that for a smaller λ, where consequently we will have less information from f (less points in N), the densities are closer to the base GP, while when the λ increases (more points in N), the density in any finite collection of points resulting from POGAMP is very similar

to the chosen f.

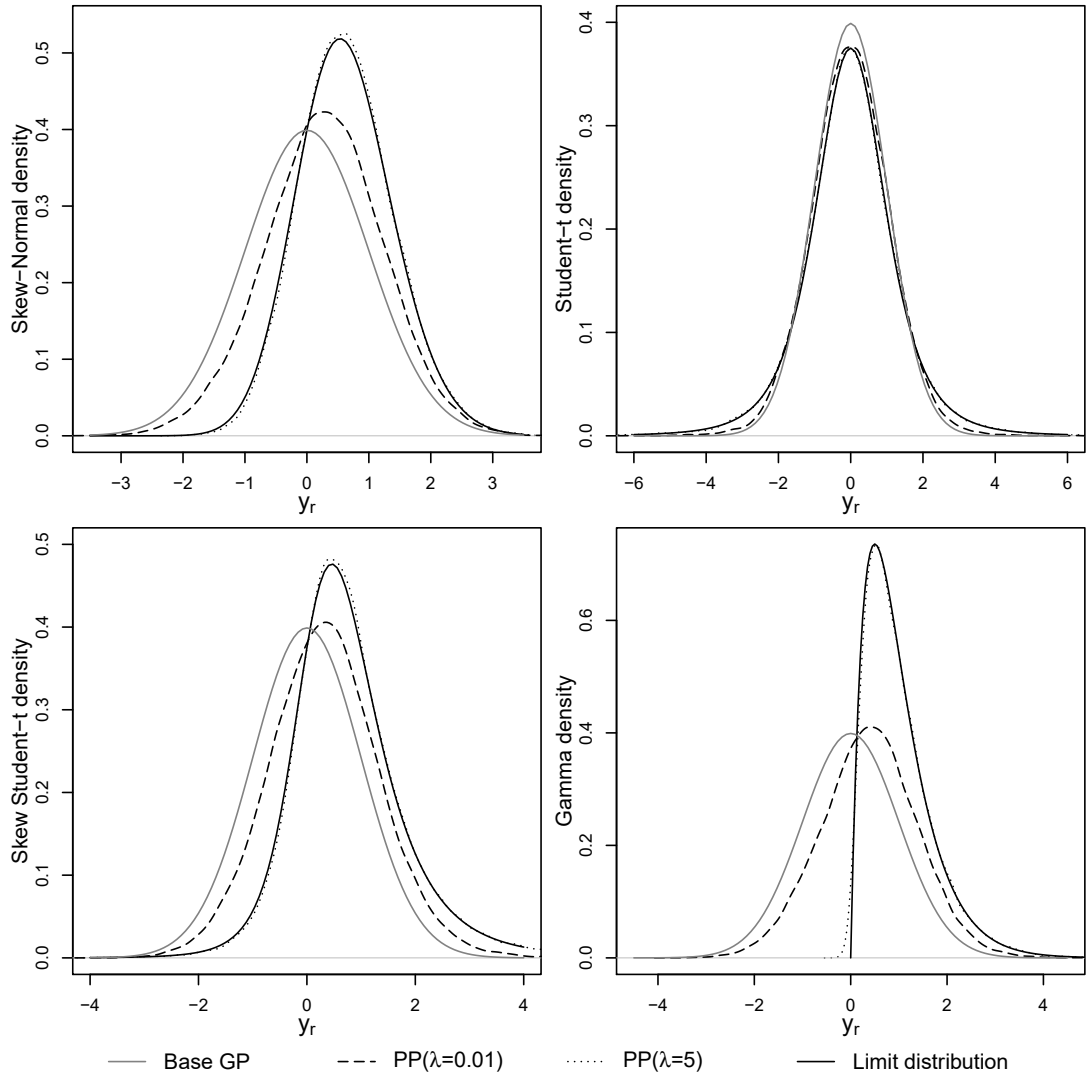


Figure 4.3: Marginal distributions at the centroid for different λ for each model

4.2 MCMC Simulation

In this chapter, we investigate the efficiency of the proposed methodology in a collection of simulated examples. We consider a variety of scenarios in terms of the choices for the f distribution, the values of parameter λ and the sample size n. We simulated 10 datasets (replications) for each of the 10 scenarios considered. In all scenarios, 10 observations were removed to make predictions at the end of this subsection.

1. Copula $SN(\mu_{SN}, \Sigma_{SN}, \gamma = 0.75)$ with PP($\lambda = 1$) and $n = 210$;

2. Copula $SN(\mu_{SN}, \Sigma_{SN}, \gamma = 0.95)$ with PP($\lambda = 1$) and $n = 210$;

3. Copula $SN(\mu_{SN}, \Sigma_{SN}, \gamma = 0.95)$ with PP($\lambda = 2$) and $n = 210$;
4. Copula $SN(\mu_{SN}, \Sigma_{SN}, \gamma = 0.95)$ with PP($\lambda = 1$) and $n = 410$;
5. Copula $SN(\mu_{SN}, \Sigma_{SN}, \gamma = 0.95)$ with PP($\lambda = 2$) and $n = 410$;
6. $t(\mu_t, \Sigma_t, \nu = 4.1)$ with PP($\lambda = 1$) and $n = 210$;
7. $t(\mu_t, \Sigma_t, \nu = 8)$ with PP($\lambda = 1$) and $n = 210$;
8. Copula $St(\mu_{St}, \Sigma_{St}, \nu = 7, \gamma = 0.75)$ with PP($\lambda = 2$) and $n = 210$;
9. Copula $St(\mu_{St}, \Sigma_{St}, \nu = 7, \gamma = 0.75)$ with PP($\lambda = 1$) and $n = 410$;
10. Copula $St(\mu_{St}, \Sigma_{St}, \nu = 7, \gamma = 0.75)$ with PP($\lambda = 2$) and $n = 410$;

We impose a restriction to the parameter space of the POGAMPs so that the stationary mean and variance of the base GP and the (conceptual) f process are the same. In particular, we set

$$\sigma^2_{SN} = \sqrt{\frac{\sigma^2_{GP}}{1 - \frac{2\delta^2}{\pi}}}; \qquad \mu_{SN} = \mu_{GP} - \sigma^2_{SN}\delta\sqrt{\frac{2}{\pi}};$$

$$\sigma^2_{St} = \sqrt{\frac{\sigma^2_{GP}}{\frac{\nu}{\nu-2} - (b_\nu\delta)^2}}; \qquad \mu_{St} = \mu_{GP} - \sigma^2_{St} b_\nu \delta;$$

$$\sigma^2_t = \frac{(\nu-2)\sigma^2_{GP}}{\nu}; \qquad \mu_t = \mu_{GP},$$

where $\delta = \frac{\alpha}{\sqrt{1+\alpha^2}}$ and $b_\nu = \frac{\sqrt{\nu}\Gamma(\frac{\nu-1}{2})}{\sqrt{\pi}\Gamma(\frac{\nu}{2})}$.

We set $\mu_{GP} = 0$, $\sigma^2_{GP} = 1$ and $\zeta = 2$ for all the scenarios. Furthermore, in order to avoid identifiability problems due to the result in Theorem 2.2, we adopt a $Exp(0.5)$ prior distribution for parameter λ. Note that this is a low informative distribution (based on calculations of a conjugate Bayesian analysis for Poisson data) and is meant only to avoid the "explosion" of the estimated λ.

The choice to specify the f distribution using copulas for the Skew distributions allows for more model flexibility. The traditional multivariate versions of the Skew normal and

t distributions (see Azzalini, 2013) have some constraints in the parameter space that precludes the elicitation of distributions with highly asymmetric marginals. In particular, the copula construction allows for a direct specification of the Skewness parameter of the respective marginal distributions, given by γ, and the correlation of the respective finite-dimensional distributions. On the other hand, for the traditional multivariate versions, both the Skewness of marginal and correlation of the finite-dimensional distributions depend on both γ and Σ_*.

Convergence diagnostic analyses strongly suggest the convergence of the MCMC algorithm in all the scenarios. Some plots are presented in Figures A.1, A.2 and A.3 - A.6, in Appendix A.2.

Figure 4.4 shows, for scenario 1, the empirical posterior density of all the parameters, for all the 10 replications. Where in the vertical line in red is the real value of the parameters and in black the empirical posterior densities for each replication, being possible to observe a greater concentration in the majority of the densities, concentrated around the vertical line.

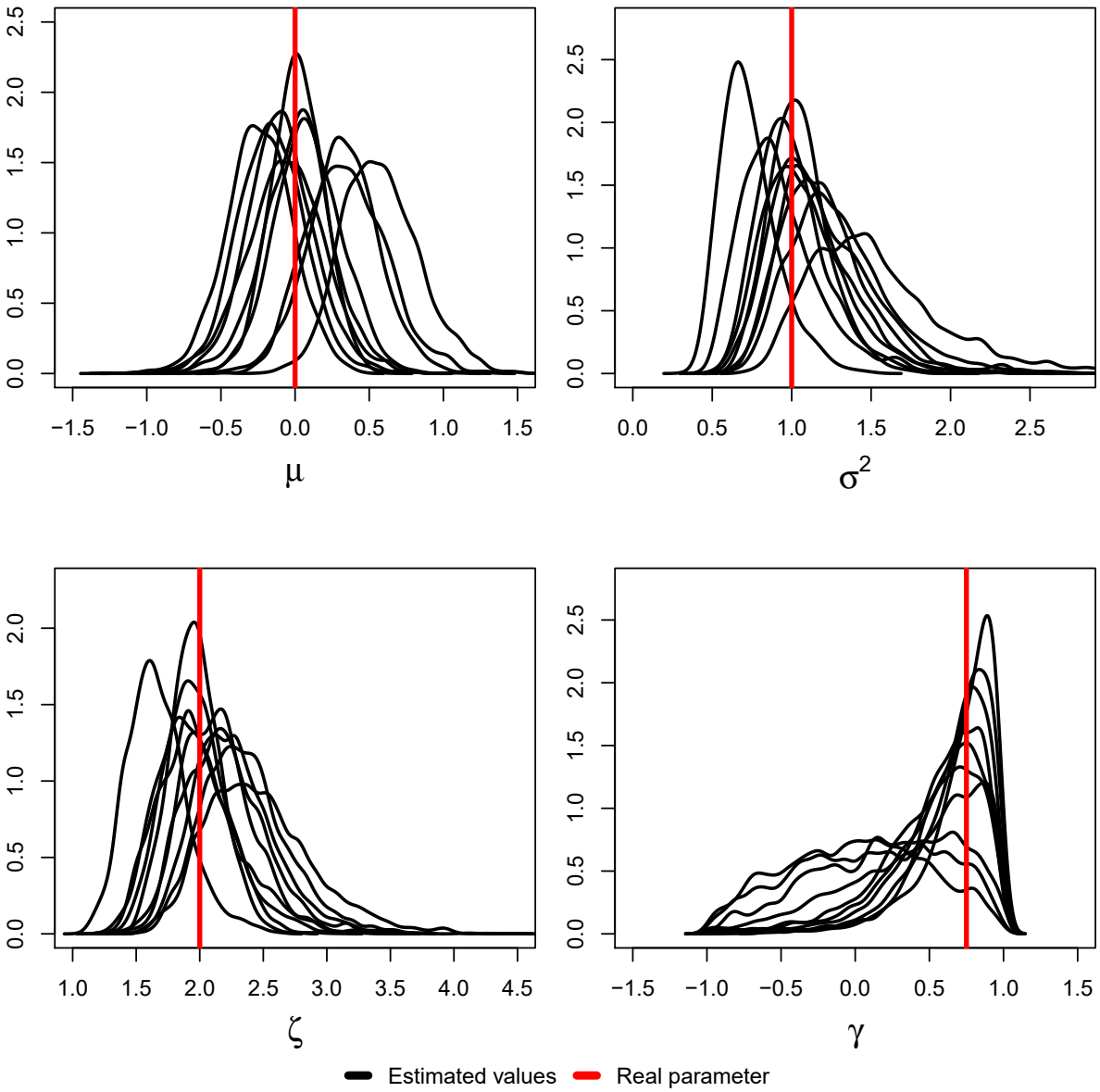


Figure 4.4: Empirical posterior density for all replications of the model parameters; $SN_{\gamma=0.75}; \lambda = 1; n = 200$

We present the following statistics bellow to evaluate the predictions

$$D_{abs} = \frac{1}{10}\sum_{i=1}^{10} |\bar{\boldsymbol{y}}_{\boldsymbol{i}} - \boldsymbol{Y}| \qquad EQM = \frac{1}{10}\sum_{i=1}^{10} (\bar{\boldsymbol{y}}_{\boldsymbol{i}} - \boldsymbol{Y})^2 ,$$

where $\boldsymbol{Y}$ is the vector of the real values and $\bar{\boldsymbol{y}}_{\boldsymbol{i}}$ is the posterior predictive.

The Table 4.1 show the mean and standard deviation of the data simulated prediction, where one replication of each scenario is chosen to perform such tests mentioned above. Overall, the best results were satisfactory and better when we have more observations. It is worth mentioning that the Student-t had good results even with fewer observations,

possibly due to the greater similarity between this distribution and the GP.

Table 4.1: MCMC simulation posterior predictive

Scenario	Mean (SD) of D_{abs}	Mean (SD) of EQM
$SN_{\gamma=0.75}$; $\lambda = 1$; n = 200	0.224 (0.17)	2.472 (1.268)
$SN_{\gamma=0.95}$; $\lambda = 1$; n = 200	0.136 (0.127)	1.431 (1.142)
$SN_{\gamma=0.95}$; $\lambda = 2$; n = 200	0.204 (0.149)	0.932 (0.414)
$SN_{\gamma=0.95}$; $\lambda = 1$; n = 400	0.058 (0.053)	1.258 (0.69)
$SN_{\gamma=0.95}$; $\lambda = 2$; n = 400	0.103 (0.089)	1.033 (0.418)
$t_{\nu=4.1}$; $\lambda = 1$; n = 200	0.053 (0.032)	1.187 (0.517)
$t_{\nu=8}$; $\lambda = 1$; n = 200	0.053 (0.032)	1.263 (0.554)
$St_{\gamma=0.75;\nu=7}$; $\lambda = 2$; n = 200	0.075 (0.05)	2.481 (1.708)
$St_{\gamma=0.75;\nu=7}$; $\lambda = 1$; n = 400	0.057 (0.051)	1.113 (0.644)
$St_{\gamma=0.75;\nu=7}$; $\lambda = 2$; n = 400	0.041 (0.039)	1.137 (0.488)

4.3 Application

In this section, we fit the POGAMP for different choices of f to a dataset of maximum temperature data observed in Australia. We consider a subset of a global dataset of merged maximum daily temperature measurements from the 'Global Surface Summary of Day' data with 'European Climate Assessment & Dataset' data in July 2011. The dataset is available in the *R* package *meteo*. The data consist of the maximum temperature observed on July 5 at 449 location sites, in the region with longitude $[110; 154]$ and latitude $[39; 12]$. We transform the coordinates to Cartesian and rescale this to obtain $S = (0, 10) \times (0, 6.7057)$. We remove 10 randomly chosen observations which are predicted by the POGAMP and used to compare the different choices of f's. The same dataset, with similar locations, was used in Bevilacqua et al. (2021).

The POGAMP is fit with the following f's: 1) Copula SN; 2) Copula St; and 3) Copula Student-t. In order to choose the initial values of θ in the MCMC, we fit a GP with the same covariance function as the base GP of the POGAMP and estimate its parameters via maximum likelihood. For the remaining parameters, we set initial values $\gamma = 0.7$ and $\nu = 8$. It is worth mentioning that the prior used for the λ parameter was an Exp. with rate $= 0.9$, then λ does not explode, therefore we are able to control the number of points

and consequently the major cost of the algorithm.

Figure 4.5 shows the observed locations.

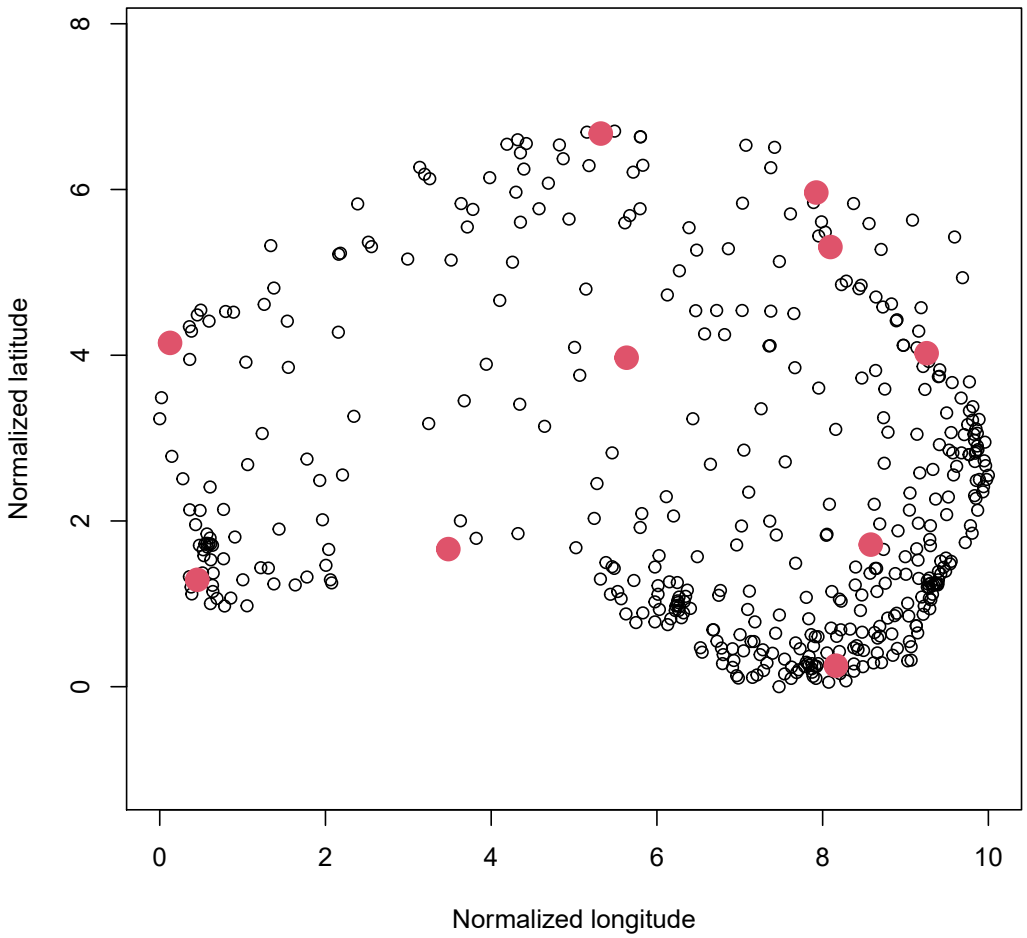


Figure 4.5: Locations of the Australia temperature example. The red dots are the 10 locations removed to perform prediction.

We impose a restriction to the parameter space of the POGAMPs so that the stationary mean and variance of the base GP and of the (conceptual) f process are the same.

Convergence diagnostics strongly suggest the convergence of the MCMC algorithm for all 3 models - see Figures A.7 - A.12, in Appendix A.3.

Table 4.2 shows some posterior statistics of the parameters. As it is expected, the estimates of the common parameters are very similar among the different models. Figure 4.6 shows the prediction at the 10 locations removed from the data.

Results suggest that the data do not present a significant symmetry departure but does present heavy tails, indicating the Student-t model to be the best choice. Nevertheless, the prediction for the 10 removed locations is very similar for all the 3 models, as we can see in Figure 4.6. This is probably due to the very low estimate of the range parameter (1.13).

Table 4.2: Posterior mean, median and s.d. of parameters.

	SN	Student-t	St
Parameters	Mean / Median (SD)	Mean / Median (SD)	Mean / Median (SD)
μ	21.3 /21.24 (2.34)	23.16 /23.25 (1.61)	24.1 /24.08 (1.65)
σ^2	19.33 /19.11 (2.23)	16.04 /15.74 (2.16)	16.81 /16.58 (2.2))
ζ	1.167 /1.152 (0.153)	1.136 /1.13 (0.163)	1.115 /1.099 (0.155)
γ	-0.105 /-0.149 (0.542)	–	-0.391 /-0.357 (0.251)
ν	–	6.026 /5.829 (1.527)	7.194 /6.873 (2.081)
N	44.78 /42 (21.63)	214.53 /190 (134.88)	86.39 /82 (30.25)
λ	0.67 /0.63 (0.33)	3.17 /2.78 (2)	1.29 /1.23 (0.47)

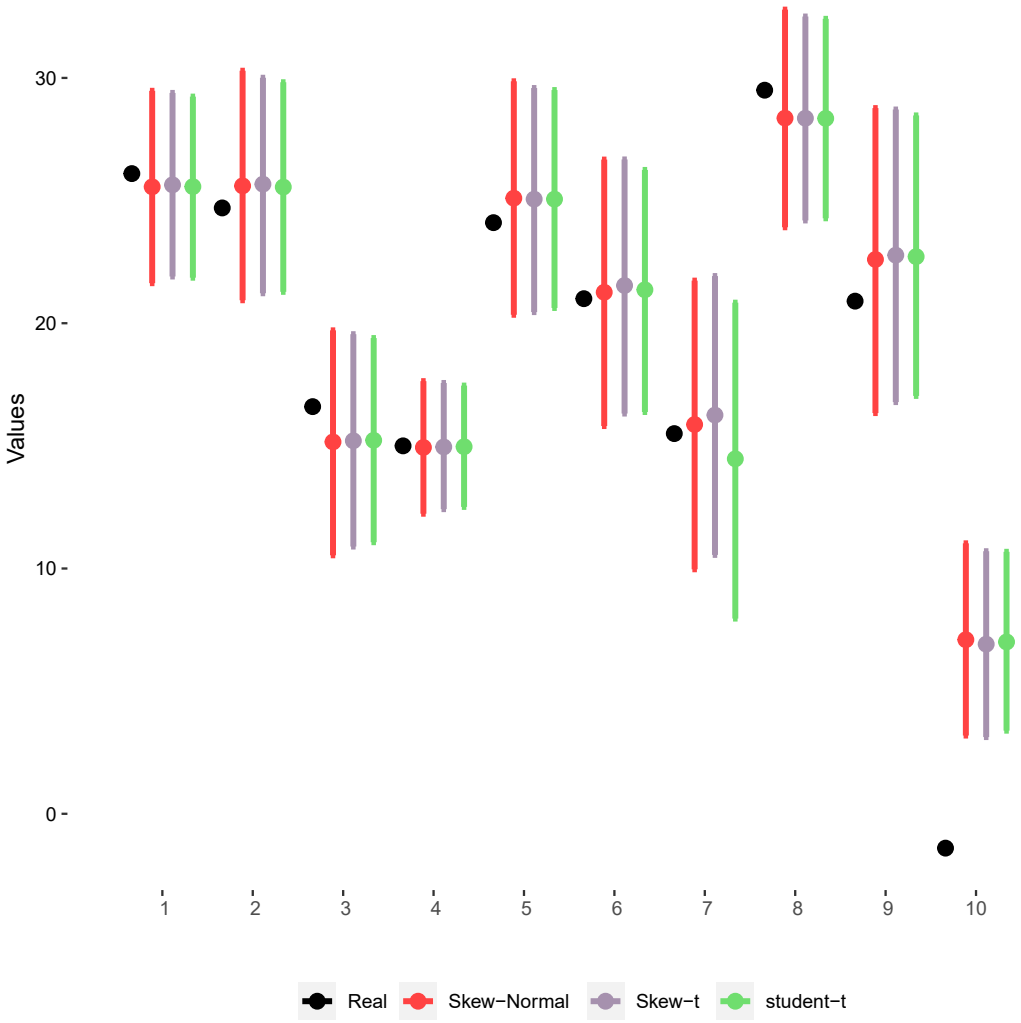


Figure 4.6: Posterior mean and Credible interval of predicted values.

Chapter 5

Final remarks

5.1 Final remarks

This thesis proposes a novel geostatistical process, called the POGAMP, aiming at providing a valid, flexible and efficient solution to fit and predict geostatiscal phenomena. The development of the POGAMP is motivated by the need to consider more flexible properties than those offered by Gaussian process but, at the same time, retain a reasonable level of computational complexity. The POGAMP offers a flexible way to define geostatistical processes based on any finite-dimensional multivariate distribution.

The existence of the POGAMP is established in a collection of theoretical results which are rigorously proved in the thesis. Moreover, a reasonably efficient infinite-dimensional MCMC algorithm is proposed to carry out inference for discretely observed POGAMPs. The proposed algorithm is exact in the sense of not resorting to discrete approximations of the process and having the exact target posterior distribution as its invariant distribution. The key idea to achieved the exactness of the method is that of retrospective sampling, where only a finite-dimensional representation of the unobserved part of the POGAMP needs to be unveiled to perform the steps of the MCMC algorithm.

A collection of simulated examples is presented to illustrate and investigate important properties of the POGAMP and of the proposed MCMC algorithm. An application to temperature data is also presented and analysed with 3 different models.

We believe that this thesis provides highly relevant contributions to the literature of geostatistics and may give rise to further developments in the area. Nevertheless, we also acknowledge that the proposed methodology urges further exploration, specially in the

lines described in the previous two sections.

5.2 An NNGP approach to deal with large datasets

The computational bottleneck of proposed methodology is the the cost related to the Gaussian process, in particular, the computation of inverses and Choleski decompositions of covariance matrices. This is a practical limitation of the methodology that prevents its use with large datasets ($\mathcal{O}(n^3)$).

One possible solution is to replace the base Gaussian process in the definition of Y by a Nearest Neighbor Gaussian Process (NNGP) approximation. Although the NNGP is an approximation to the original GP, it does define a valid Gaussian process measure. This means that the resulting process is a valid process, but for which the base GP is of the NNGP type. Moreover, since the finite-dimensional distributions of the resulting process aim at resembling the f distribution, it is reasonable to interpret this process not as an approximation for the originally proposed one but simply as an alternative to fulfill the same modelling properties.

An NNGP Z is a valid Gaussian process, devised from a parent $GP(\mu, \Sigma(\sigma^2, \tau^2))$ by imposing some conditional independence structure that leads to a sparse covariance structure. For a reference set $\mathcal{S} = \{\mathfrak{s}_1, \ldots, \mathfrak{s}_r\}$ and a maximum number m of neighbors, the NNGP factorises the distribution of Z (conditional on parameters) as follows:

$$\begin{aligned}
\pi(Z) &= \pi(Z_{\mathcal{S}})\pi(Z_{S\backslash\mathcal{S}}|Z_{\mathcal{S}}), \\
\pi(Z_{\mathcal{S}}) &= \pi_{PG}(Z_{\mathfrak{s}_1})\pi_{PG}(Z_{\mathfrak{s}_2}|Z_{\mathfrak{s}_1})\pi_{PG}(Z_{\mathfrak{s}_3}|Z_{\mathfrak{s}_1}, Z_{\mathfrak{s}_2}) \ldots \pi_{PG}(Z_{\mathfrak{s}_{m+1}}|Z_{\mathfrak{s}_1}, \ldots, Z_{\mathfrak{s}_m}) \\
&\quad \pi_{PG}(Z_{\mathfrak{s}_{m+2}}|Z_{\mathcal{N}(\mathfrak{s}_{m+2})}) \ldots \pi_{PG}(Z_{\mathfrak{s}_r}|Z_{\mathcal{N}(\mathfrak{s}_r)}), \\
\pi_{PG}(Z_{S_0}|Z_{\mathcal{S}}) &= \prod_{i=1}^{I} \pi_{PG}(Zs_i|Z_{\mathcal{N}(s_i)}), \text{ for any finite set } S_0 = \{s_1, \ldots, s_I\} \subset S\backslash\mathcal{S},
\end{aligned}$$

where π_{PG} is the respective density under the parent GP measure, $\mathcal{N}(\mathfrak{s}_i)$ is the set of the m closest neighbors of $\mathfrak{s}_i$ in $\{\mathfrak{s}_1, \ldots, \mathfrak{s}_{i-1}\}$, for $i \geqslant m + 2$, and $\mathcal{N}(s_i)$ is the set of the m closest neighbors of s_i in $\mathcal{S}$.

In our case, the parent process is the GP on $S\backslash S_N$ defined by the conditional measure of $Y_{S\backslash S_N}$ given Y_N, under the GP measure with mean μ and covariance function $\Sigma(\sigma^2, \tau^2)$.

This implies that

$$\begin{aligned}
\pi(Y_{\mathcal{S}\backslash S_N}|Y_N) &= \pi_{PG}(Y_{\mathfrak{s}_1}|Y_N)\pi_{PG}(Y_{\mathfrak{s}_2}|Y_{\mathfrak{s}_1},Y_N)\pi_{PG}\ldots\pi_{PG}(Y_{\mathfrak{s}_{m+1}}|Y_{\mathfrak{s}_1},\ldots,Y_{\mathfrak{s}_m,Y_N}) \\
&\quad \pi_{PG}(Y_{\mathfrak{s}_{m+2}}|Y_{\mathcal{N}(\mathfrak{s}_{m+2})},Y_N)\ldots\pi_{PG}(Y_{\mathfrak{s}_r}|Y_{\mathcal{N}(\mathfrak{s}_r)},Y_N), \\
\pi(Y_{S_0}|Y_{\mathcal{S}}) &= \prod_{i=1}^{I}\pi_{PG}(Y s_i|Y_{\mathcal{N}(s_i)},Y_N), \\
&\quad \text{for any finite set } S_0=\{s_1,\ldots,s_I\}\subset S\backslash(\mathcal{S}\bigcup S_N),
\end{aligned}$$

It is common to set $\mathcal{S}$ to be the set of observed locations. In our case, we set $\mathcal{S}$ to be a regular mesh in S, which allows us to use computational strategies to optimise the sampling steps of Y in $S\backslash(\mathcal{S}\bigcup S_N)$. Based on the results in Gonçalves and Dias (2020) and results of several simulated examples with our model, we could set $r = 2500$ and $m = 16$.

The computational gains from using the NNGP are quite significant and allow the use of our methodology with large datasets. Whilst the computational cost to generate Y at a given location conditional on (Y_N, Y_0) is $\mathcal{O}((N+n)^3)$, the same task has a cost $\mathcal{O}((N+m)^3)$ under the NNGP approach, where m is fixed and $m \ll n$.

5.3 Simulations and applications

It is essential to further explore the methodology proposed in this thesis trough simulated and real examples. Simulations ought to include f distributions with new parameter configurations and data sizes as well as more replications from each model. This also includes examples in which the NNGP approach proposed in the previous section in employed. Several prediction exercises will be performed for all the scenarios considered.

Other classes of the proposed family of models will also be explored. In particular, models in which the latent Poisson process N is inhomogeneous. This structure allows us to consider, for example, non-isotropic processes, by setting f to be Normal with a different covariance function than the one from the base GP, whilst still working with an isotropic structure.

Finally, new applications should vary in terms of data size and behavior in terms of asymmetry and heavy tail.

Bibliography

Allard, D. and Naveau, P. (2007). A new spatial skew-normal random field model. *Communications in Statistics—Theory and Methods*, 36(9):1821–1834.

Alodat, M. and Al-Momani, E. (2014). Skew gaussian process for nonlinear regression. *Communications in Statistics-Theory and Methods*, 43(23):4936–4961.

Alodat, M. and Al-Rawwash, M. (2009). Skew-gaussian random field. *Journal of Computational and Applied Mathematics*, 232(2):496–504.

Alodat, M. and AL-Rawwash, M. (2014). The extended skew gaussian process for regression. *Metron*, 72(3):317–330.

Arellano-Valle, R. B. and Azzalini, A. (2006). On the unification of families of skew-normal distributions. *Scandinavian Journal of Statistics*, 33(3):561–574.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12(2):171–178.

Azzalini, A. (2013). *The skew-normal and related families*, volume 3. Cambridge University Press, New York.

Azzalini, A. and Valle, A. D. (1996). The multivariate skew-normal distribution. *Biometrika*, 83(4):715–726.

Bárdossy, A. (2006). Copula-based geostatistical models for groundwater quality parameters. *Water Resources Research*, 42(11):1–12.

Bárdossy, A. and Li, J. (2008). Geostatistical interpolation using copulas. *Water resources research*, 44(7):1–15.

Beskos, A., Papaspiliopoulos, O., Roberts, G. O., and Fearnhead, P. (2006). Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):333–382.

Beskos, A. and Roberts, G. O. (2005). Exact simulation of diffusions. *The Annals of Applied Probability*, 15(4):2422–2444.

Bevilacqua, M., Caamaño-Carrillo, C., Arellano-Valle, R. B., and Morales-Oñate, V. (2021). Non-gaussian geostatistical modeling using (skew) t processes. *Scandinavian Journal of Statistics*, 48(1):212–245.

Cressie, N. (2015). *Statistics for spatial data*. John Wiley & Sons, Hoboken.

Dubrule, O. (1989). A review of stochastic models for petroleum reservoirs. *GeoStatistics*, pages 493–506.

Gamerman, D. and Lopes, H. F. (2006). *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. CRC Press, New York, 2nd edition.

Gelfand, A. E. and Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410):398–409.

Geman, S. and Geman, D. (1993). Stochastic relaxation, gibbs distributions and the bayesian restoration of images. *Journal of Applied Statistics*, 20(5-6):25–62.

Genton, M. G. and Zhang, H. (2012). Identifiability problems in some non-gaussian spatial random fields. *Chilean Journal of Statistics*, 3(2):171–179.

Gonçalves, F. B. and Dias, B. C. (2020). Exact bayesian inference for level-set cox processes. *arXiv preprint arXiv:2012.05764*.

Gonçalves, F. B. and Gamerman, D. (2018). Exact bayesian inference in spatiotemporal cox processes driven by multivariate gaussian processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(1):157–175.

Gräler, B., Kazianka, H., and de Espindola, G. M. (2010). Copulas, a novel approach to model spatial and spatio-temporal dependence. In *GIScience for Environmental Change Symposium Proceedings*, volume 40, pages 49–54.

Gräler, B. and Pebesma, E. (2011). The pair-copula construction for spatial data: a new approach to model spatial dependency. *Procedia Environmental Sciences*, 7:206–211.

Henze, N. (1986). A probabilistic representation of the skew-normal distribution. *Scandinavian Journal of Statistics*, 13(4):271–275.

Hohn, M. (1998). *GeoStatistics and petroleum geology*. Springer Science & Business Media.

Hosseini, F., Eidsvik, J., and Mohammadzadeh, M. (2011). Approximate bayesian inference in spatial glmm with skew normal latent variables. *Computational Statistics & Data Analysis*, 55(4):1791–1806.

Hughes, J. (2015). copcar: A flexible regression model for areal data. *Journal of Computational and Graphical Statistics*, 24(3):733–755.

Kazianka, H. and Pilz, J. (2010). Copula-based geostatistical modeling of continuous and discrete data including covariates. *Stochastic environmental research and risk assessment*, 24(5):661–673.

Kazianka, H. and Pilz, J. (2011). Bayesian spatial modeling and interpolation using copulas. *Computers & Geosciences*, 37(3):310–319.

Kim, H.-M. and Mallick, B. K. (2004). A bayesian prediction using the skew gaussian distribution. *Journal of Statistical Planning and Inference*, 120(1-2):85–101.

Liu, J. S. (1994). The collapsed gibbs sampler in bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association*, 89(427):958–966.

Mahmoudian, B. (2017). A skewed and heavy-tailed latent random field model for spatial extremes. *Journal of Computational and Graphical Statistics*, 26(3):658–670.

Mahmoudian, B. (2018). On the existence of some skew-gaussian random field models. *Statistics & Probability Letters*, 137:331–335.

Nelsen, R. B. (2007). *An introduction to copulas*. Springer Science & Business Media, New York, 2nd edition.

Palacios, M. B. and Steel, M. F. J. (2006). Non-gaussian bayesian geostatistical modeling. *Journal of the American Statistical Association*, 101(474):604–618.

Prates, M. O., Dey, D. K., Willig, M. R., and Yan, J. (2015). Transformed gaussian markov random fields and spatial modeling of species abundance. *Spatial Statistics*, 14:382–399.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Resnick, S. I. (2013). *A probability path*. Springer Science & Business Media, Boston.

Robert, C. (2007). *The Bayesian Choice*. Springer, New York, 2nd edition.

Roberts, G. O., Gelman, A., Gilks, W. R., et al. (1997). Weak convergence and optimal scaling of random walk metropolis algorithms. *The Annals of Applied Probability*, 7(1):110–120.

Roberts, G. O. and Rosenthal, J. S. (2009). Examples of adaptive mcmc. *Journal of Computational and Graphical Statistics*, 18(2):349–367.

Shao, J. (2003). *Mathematical Statistics*. Springer Texts in Statistics. Springer, New York.

Tagle, F., Castruccio, S., and Genton, M. G. (2020). A hierarchical bi-resolution spatial skew-t model. *Spatial Statistics*, 35:100398.

Uribe-Opazo, M. A., De Bastiani, F., Galea, M., Schemmer, R. C., and Assumpcao, R. A. B. (2021). Influence diagnostics on a reparameterized t-student spatial linear model. *Spatial Statistics*, 41:100481.

Wang, H., Guan, Y., and Reich, B. (2019). Nearest-neighbor neural networks for geostatistics. In *2019 international conference on data mining workshops (ICDMW)*, pages 196–205. IEEE.

Xue-Kun Song, P. (2000). Multivariate dispersion models generated from gaussian copula. *Scandinavian Journal of Statistics*, 27(2):305–320.

Zareifard, H. and Khaledi, M. J. (2013). Non-gaussian modeling of spatial data using scale mixing of a unified skew gaussian process. *Journal of Multivariate Analysis*, 114:16–28.

Zareifard, H., Khaledi, M. J., Rivaz, F., Vahidi-Asl, M. Q., et al. (2018). Modeling skewed spatial data using a convolution of gaussian and log-gaussian processes. *Bayesian Analysis*, 13(2):531–557.

Zhang, H. and El-Shaarawi, A. (2010). On spatial skew-gaussian processes and applications. *Environmetrics: The official Journal of the International Environmetrics Society*, 21(1):33–47.

Zheng, X., Kottas, A., and Sansó, B. (2021). Nearest-neighbor geostatistical models for non-gaussian data. *arXiv preprint arXiv:2107.07736*.

Appendix A

Appendix

A.1 Proofs

Proof of Lemma 2.1. In order to show the existence of the probability measure μ, we will prove that it satisfies the Kolmogorov axioms.

(**i**) $\mu(A) \in [0,1], \forall A \in \mathcal{F}$. We have that

$$\mu(A) = \int_{\Omega_1} \int_{\Omega_2} \mathbb{1}_{[(\omega_1,\omega_2)\in A]} d\mu_{2,\omega_1}(\omega_2) d\mu_1(\omega_1).$$

Define $g_A \colon \Omega_1 \to \mathbb{R}$

$$g_A(\omega_1) = \int_{\Omega_2} \mathbb{1}_{[(\omega_1,\omega_2)\in A]} d\mu_{2,\omega_1}(\omega_2).$$

The fact that μ_{2,ω_1} is a probability measure on $(\Omega_2, \mathcal{F}_2)$ implies that $0 \leqslant g_A(\omega_1) \leqslant 1$.

$$0 = \int_{\Omega_1} 0 d\mu_1(\omega_1) \leqslant \int_{\Omega_1} g_A(\omega_1) d\mu_1(\omega_1) \leqslant \int_{\Omega_1} 1 d\mu_1(\omega_1) = 1.$$

(**ii**) For each $\omega_1 \in \Omega_1$, we have that

$$\int_{\Omega_2} \mathbb{1}_{[(\omega_1,\omega_2)\in \Omega]} d\mu_{2,\omega_1}(\omega_2) = 1,$$

hence

$$\mu(\Omega) = \int_{\Omega_1} \int_{\Omega_2} \mathbb{1}_{[(\omega_1,\omega_2)\in \Omega]} d\mu_{2,\omega_1}(\omega_2) d\mu_1(\omega_1) = \int_{\Omega_1} 1 d\mu_1(\omega_1) = 1.$$

(**iii**) (σ-additivity). $\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$, for all countable sequence of disjoint events A_i's. Defining $A = \bigcup_{i=1}^{\infty} A_i$ and noting that we have that $\mathbb{1}_{[(\omega_1,\omega_2)\in A]} = \sum_{i=1}^{\infty} \mathbb{1}_{[(\omega_1,\omega_2)\in A_i]}$ and

$$\mu(\bigcup_{i=1}^{\infty} A_i) = \mu(A) = \int_{\Omega_1}\int_{\Omega_2} \mathbb{1}_{[(\omega_1,\omega_2)\in A]} d\mu_{2,\omega_1}(\omega_2) d\mu_1(\omega_1) =$$
$$\int_{\Omega_1}\int_{\Omega_2} \sum_{i=1}^{\infty} \mathbb{1}_{[(\omega_1,\omega_2)\in A_i]} d\mu_{2,\omega_1}(\omega_2) d\mu_1(\omega_1),$$

since $d\mu_{2,\omega_1}(\omega_2)$ is a probability measure, we have

$$\mu(A) = \int_{\Omega_1} \sum_{i=1}^{\infty} \left[\int_{\Omega_2} \mathbb{1}_{[(\omega_1,\omega_2)\in A_i]} d\mu_{2,\omega_1}(\omega_2) \right] d\mu_1(\omega_1),$$

since $g_i(\omega_1) = \int_{\Omega_2} \mathbb{1}_{[(\omega_1,\omega_2)\in A_i]} d\mu_{2,\omega_1}(\omega_2)$ is $\mathcal{F}_1$-measurable we have that

$$\mu(A) = \int_{\Omega_1} \sum_{i=1}^{\infty} g_i(\omega_1) d\mu_1(\omega_1) = \sum_{i=1}^{\infty} \int_{\Omega_1} g_i(\omega_1) d\mu_1(\omega_1)$$
$$= \sum_{i=1}^{\infty} \int_{\Omega_1}\int_{\Omega_2} \mathbb{1}_{[(\omega_1,\omega_2)\in A_i]} d\mu_{2,\omega_1}(\omega_2) d\mu_1(\omega_1) = \sum_{i=1}^{\infty} \mu(A_i).$$

Therefore μ is a probability measure. □

Proof of Theorem 2.1. (a)First we will consider the process with fixed N. We have that $\mathcal{G}$ is the measure of the augmented Gaussian process and define $\mathcal{G}_N$ as the conditional measure of Y_N given N, under $\mathcal{G}$, ν as the Lebesgue measure and $g = \frac{d\mathcal{G}_N}{d\nu}$. Analogously, $\mathcal{P}_N$ is defined as the conditional measure of Y_N given N, under $\mathcal{P}$.

Define the function $h\colon \mathcal{Y} \to \mathbb{R}^+$ such that $h(y) = \frac{f}{g}(y_N)$, where $y_N = h_1(y)$, $\forall y \in \mathcal{Y}$, for $h_1\colon \mathcal{Y} \to \mathbb{R}^N$ and $\frac{f}{g} : \mathbb{R}^N \to \mathbb{R}^+$ $\forall y_N \in Y_N$.

We have that h is $\mathcal{B}(\mathcal{Y})$-measurable, since Y_N is measurable and $\frac{f}{g}$ is continuous. Since h is measurable and non-negative, by the Radon-Nikodym Theorem (RNT) we have that there exists a measure $\mathcal{P}(A) = \int_A h(y) d\mathcal{G}(y)$, $y \in \mathcal{Y}$ and $\forall A \in \mathcal{B}(\mathcal{Y})$ such that

$$\frac{\mathrm{d}\mathcal{P}}{\mathrm{d}\mathcal{G}}(y) = h(y) = \frac{f}{g}(y_N).$$

We now show that $\mathcal{P}$ is a probability measure.

(**i**) $0 \leqslant \mathcal{P}(A) \leqslant 1$, $\forall A \in \mathcal{B}(\mathcal{Y})$.

Let h a non-negative measurable function and $\mathcal{G}$ a measure, then

$$\mathcal{P}(A) = \int_A h(y) d\mathcal{G}(y) \geqslant 0, \quad \text{furthermore}$$
$$\mathcal{P}(A) \leqslant \int_{\mathcal{Y}} h(y) d\mathcal{G}(y) = \int_{\mathbb{R}^N} \frac{f}{g}(y_N) d\mathcal{G}_N(y_N) = \int_{\mathbb{R}^N} \frac{f}{g}(y_N) g(y_N) d\nu$$
$$= \int_{\mathbb{R}^N} f(y_N) d\nu = 1.$$

(**ii**) $\mathcal{P}(\mathcal{Y}) = 1$. See previous equation.

(**iii**) (σ-finite).

$\mathcal{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathcal{P}(A_i)$, for all countable sequence of disjoint events A_i in $\mathcal{B}(\mathcal{Y})$. Defining $A = \bigcup_{i=1}^{\infty} A_i$ we have that $\mathbb{1}_{[y\in A]} = \sum_{i=1}^{\infty} \mathbb{1}_{[y\in A_i]}$ and

$$\begin{aligned}\mathcal{P}(A) = \mathcal{P}(\bigcup_{i=1}^{\infty} A_i) &= \int_{\mathcal{Y}} \sum_{i=1}^{\infty} \mathbb{1}_{[y\in A_i]} h(y) d\mathcal{G}(y) \\ &= \sum_{i=1}^{\infty} \int_{\mathcal{Y}} \mathbb{1}_{[y\in A_i]} h(y) d\mathcal{G}(y) = \sum_{i=1}^{\infty} \int_{A_i} h(y) d\mathcal{G}(y) = \sum_{i=1}^{\infty} \mathcal{P}(A_i).\end{aligned}$$

Therefore $\mathcal{P}$ is a probability measure.

We now show that $\mathcal{P}$ satisfies Definition 2.2 for fixed N. Let $Y = (Y_{-N}, Y_N)$ and consider $\tilde{\mathcal{P}}_N$ to be the conditional measure of $(Y_{-N}|Y_N, N)$ under $\mathcal{P}$. Analogously, consider $\tilde{\mathcal{G}}_N$ to be the conditional measure of $(Y_{-N}|Y_N, N)$ under $\mathcal{G}$. Since $\mathcal{P}_N \ll \mathcal{G}_N \ll \nu$, we have that

$$\frac{f}{g}(y_N) = \frac{\mathrm{d}\mathcal{P}}{\mathrm{d}\mathcal{G}}(y) = \frac{\mathrm{d}\mathcal{P}_N}{\mathrm{d}\mathcal{G}_N}(y_N)\frac{\mathrm{d}\tilde{\mathcal{P}}_N}{\mathrm{d}\tilde{\mathcal{G}}_N}(y_{-N}) = \frac{\frac{\mathrm{d}\mathcal{P}_N}{\mathrm{d}\nu}}{\frac{\mathrm{d}\mathcal{G}_N}{\mathrm{d}\nu}}(y_N)\frac{\mathrm{d}\tilde{\mathcal{P}}_N}{\mathrm{d}\tilde{\mathcal{G}}_N}(y_{-N}) \tag{A.1}$$

$$= \frac{\phi}{g}(y_N)\frac{\mathrm{d}\tilde{\mathcal{P}}_N}{\mathrm{d}\tilde{\mathcal{G}}_N}(y_{-N}) \quad \mathcal{G}\text{-a.s.} \tag{A.2}$$

Since the RN derivative above does not depend on Y_{-N}, we have that there exists $c \in \mathbb{R}^+$ such that

$$\frac{\mathrm{d}\mathcal{P}}{\mathrm{d}\mathcal{G}}(y) = \frac{\phi}{g}(y_N)c \quad \mathcal{G}\text{-a.s.} \tag{A.3}$$

Therefore, $\phi = \frac{f}{c}$ $\mathcal{G}$-a.s and, since $\int_{\mathbb{R}^N} \phi(y_N)d\nu = \int_{\mathbb{R}^N} f(y_N)d\nu = 1$, we have that $\phi = f$ and $\frac{\mathrm{d}\tilde{\mathcal{P}}_N}{\mathrm{d}\tilde{\mathcal{G}}_N}(y_{-N}) = 1$, $\mathcal{G}$-a.s. This proves that $\mathcal{P}$ is a probability measure and satisfies Definition 2.2 for fixed N.

(b) We now show the existence of the stochastic process (Y, N) defined in Definition 2.2. We decompose the PP N into $(|N|, S_N)$, where $|N|$ is the number of events from N and S_N is their respective locations. We have that $|N|$ is defined on $(\mathbb{N}, \mathcal{B}(\mathbb{N}))$ and S_N is defined on $(\mathbb{R}^{|N|}, \mathcal{B}(\mathbb{R}^{|N|}))$. We first show the the existence of the process (Y, S_N), for all fixed $|N|$.

If the integral

$$\int_{\mathcal{Y}} \mathbb{1}_{[(y,S_N)\in A]} \mathrm{d}\mathcal{P}_N \tag{A.4}$$

is $\mathcal{B}(\mathbb{R}^{|N|})$-mensurable, $\forall A \in \mathcal{B}(\mathbb{R}^{|N|}) \otimes \mathcal{B}(\mathcal{Y})$, then Lemma 2.1 implies in the existence of the joint measure of (Y, S_N) for all fixed $|N|$. In order to establish the mensurability of A.4, we show that it is a continuous function of S_N.

Consider $A = A_y \times A_{S_N} \in \mathcal{B}(\mathcal{Y}) \otimes \mathcal{B}(\mathbb{R}^{|N|})$. Hence

$$\begin{aligned}\int_{\mathcal{Y}} \mathbb{1}_{[(y,S_N)\in A]} \mathrm{d}\mathcal{P}_N &= \int_{\mathcal{Y}} \mathbb{1}_{[S_N \in A_{S_N}]} \mathbb{1}_{[y\in A_Y]} \mathrm{d}\mathcal{P}_N \\ &= \mathbb{1}_{[S_N\in A_{S_N}]} \int_{\mathcal{Y}} \mathbb{1}_{[y\in A_Y]} \frac{f}{g}(y_N)\mathrm{d}\mathcal{G} \\ &\leqslant \mathbb{1}_{[S_N\in A_{S_N}]} \int_{\mathcal{Y}} \frac{f}{g}(y_N)\mathrm{d}\mathcal{G} = \mathbb{1}_{[S_N\in A_{S_N}]}.\end{aligned}$$

This implies that there exists $\mathcal{G}$-integrable $m(y_N)$ such that $\mathbb{1}_{[y\in A_Y]}\frac{f}{g}(y_N) \leqslant m(y_N)$. Furthermore, by hypothesis, $\frac{f}{g}(y_N(S_N))$ is continuous in S_N. Therefore, by the Dominated Convergence Theorem and the continuity of $Y(s)$ in s, we have that

$$\begin{aligned}\lim_{n\to\infty} \int_{\mathcal{Y}} \mathbb{1}_{[y\in A_Y]} \frac{f}{g}(y_N(s_n))d\mathcal{G} &= \int_{\mathcal{Y}} \lim_{n\to\infty} \mathbb{1}_{[y\in A_Y]} \frac{f}{g}(y_N(s_n))d\mathcal{G} \\ &= \int_{\mathcal{Y}} \mathbb{1}_{[y\in A_Y]} \frac{f}{g}(y_N(s))d\mathcal{G}.\end{aligned}$$

Since $\mathbb{1}_{[S_N\in A_{S_N}]}$ is $\mathcal{B}(\mathbb{R}^{|N|})$-mensurable and is continuous in S_N (therefore $\mathcal{B}(\mathbb{R}^{|N|})$-mensurable), we have that the integral in A.4 is $\mathcal{B}(\mathbb{R}^{|N|})$-mensurable, implying the desired result.

Once the existence of the process (Y, S_N) is established for each fixed $|N|$, it remains to show that the joint process (Y, N) exists. In order to do that, it is enough to show that the integral

$$\int_{\mathbb{R}^{|N|}\times\mathcal{Y}} \mathbb{1}_{[|N|\in A_{|N|}]} \mathbb{1}_{[s_N\in A_{S_N}]} \mathbb{1}_{[y\in A_Y]} \frac{1}{\mu(S)} \mathrm{d}\mathcal{P}_N \mathrm{d}S_N \tag{A.5}$$

is $\mathcal{B}(\mathbb{N})$-measurable. Since N has a discrete marginal measure, it is enough to show that the integral in A.5 is a real function of N. We have that

$$\begin{aligned}&\int_{\mathbb{R}^{|N|}\times\mathcal{Y}} \mathbb{1}_{[|N|\in A_{|N|}]} \mathbb{1}_{[s_N\in A_{S_N}]} \mathbb{1}_{[y\in A_Y]} \frac{1}{\mu(S)} \mathrm{d}\mathcal{P}_N \mathrm{d}S_N \\ &= \mathbb{1}_{[|N|\in A_{|N|}]} \int_{\mathbb{R}^{|N|}} \frac{1}{\mu(S)^{|N|}} \mathbb{1}_{[s_N\in A_{S_N}]} \int_{\mathcal{Y}} \mathbb{1}_{[y\in A_Y]} \frac{f}{g}(y_N)\mathrm{d}\mathcal{G}\mathrm{d}S_N \\ &\leqslant \mathbb{1}_{[|N|\in A_{|N|}]} \int_{\mathbb{R}^{|N|}} \frac{1}{\mu(S)^{|N|}} \mathbb{1}_{[s_N\in A_{S_N}]} \int_{\mathcal{Y}} \frac{f}{g}(y_N)\mathrm{d}\mathcal{G}\mathrm{d}S_N \\ &= \mathbb{1}_{[|N|\in A_{|N|}]} \int_{A_{S_N}} \frac{1}{\mu(S)} \mathrm{d}S_N \leqslant 1,\end{aligned}$$

which implies the desired result.

□

Proof of Corollary 2.1. The result is implied by the fact that the two uses of Lemma 2.1 in the proof of Theorem 2.1 depends only on the σ-algebra of N, which is the same for both the homogeneous and inhomogeneous cases. □

Proof of Corollary 2.2. The result follows from equations (A.1) and (A.3) in the proof of Theorem 2.1 and

$$\frac{\mathrm{d}\mathcal{P}}{\mathrm{d}\mathcal{G}}(N,Y) = \frac{\mathrm{d}\mathcal{P}}{\mathrm{d}\mathcal{G}}(N)\frac{\mathrm{d}\mathcal{P}}{\mathrm{d}\mathcal{G}}(Y|N) = 1\frac{f}{g}(Y_N) = \frac{f}{g}(Y_N).$$

□

Proof of Corollary 2.3.

$$\begin{aligned} D_{\mathrm{KL}}(\mathcal{P} \parallel \mathcal{G}) &= \int \log\left(\frac{d\mathcal{P}}{d\mathcal{G}}\right) d\mathcal{P} = \mathrm{E}_N\left\{\mathrm{E}_{S_N}\left[\mathrm{E}_{Y_N}\left(\log\left(\frac{f}{g}(Y_N)\right)\right)\right]\middle| N\right\} \\ &= \sum_{N=0}^{\infty} \frac{\mathrm{e}^{-\lambda\mu(\mathrm{S})}[\lambda\mu(\mathrm{S})]^{\mathrm{N}}}{N!} \int_{S_N}\left[\int_{Y_N} \log\left(\frac{f}{g}(Y_N)\right) f(Y_N)\mathrm{d}Y_N\right] \\ &\times \left(\frac{1}{\mu(S)}\right)^N \mathrm{d}S_N \\ &= \sum_{N=1}^{\infty} \frac{\mathrm{e}^{-\lambda\mu(\mathrm{S})}\lambda^{\mathrm{N}}}{N!} \int_{S_N}\int_{Y_N} \log\left[\frac{f}{g}(Y_N)\right] f(Y_N)\mathrm{d}Y_N\mathrm{d}S_N \end{aligned}$$

□

Proof of Theorem 2.2. For each $Y_{n,r}$, define $\dot{S}_n$ as the k-vector with the closest locations to each point in r. We have that, for all $\boldsymbol{\epsilon} > 0$, when $\lambda_n \uparrow \infty$, the probability that at least one point of $\dot{S}_n$ is inside each open ball of radius ϵ around each location of r - converges to 1. Consider, without loss of generality, that ϵ is small enough so that all the k balls are disjoint. This implies that the probability measures on the points inside each ball are independent Poisson processes with rate λ. Therefore, the probability above is given by

$$\left(1 - \frac{(\lambda\pi\epsilon^2)^0 \exp\{-\lambda\pi\epsilon^2\}}{0!}\right)^k \xrightarrow{\lambda\to\infty} 1.$$

This implies that $\dot{S}_n \xrightarrow{p} r$ and, given the continuity of the covariance function of the base GP, $Cor(Y_{n,r}, Y_{\dot{S}_n}) \xrightarrow{p} \underset{\sim}{1}$, where $\underset{\sim}{1}$ is a k by k matrix of 1's.

We also have that

$$\begin{aligned} &\lim_{n\to\infty} \mathrm{P}(|Y_{n,r} - Y_{\dot{S}_n}| \leqslant \boldsymbol{\epsilon}) \\ &= \lim_{n\to\infty} \int_{\mathbb{R}^k} \mathrm{P}(|Y_{n,r} - Y_{\dot{S}_n}| \leqslant \boldsymbol{\epsilon} \mid Y_{S_N} = y_{S_N}) f(y_{S_N})\mathrm{d}y_{S_N}, \end{aligned} \tag{A.6}$$

where, from Definition 2.2, $(Y_{n,r} \mid Y_{S_N}) \sim \mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ with $\boldsymbol{\mu} = \boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(Y_{S_n} - \boldsymbol{\mu}_1)$ and $\Sigma = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$. The fact that $Cor(Y_{n,r}, Y_{\dot{S}_n}) \xrightarrow{p} \underset{\sim}{1}$ implies that $\Sigma \xrightarrow{p} \underset{\sim}{0}$ and $\mu \xrightarrow{p} Y_{\dot{S}_n}$ entry-wise. Thus, the Dominated Convergence Theorem implies that

$$\begin{aligned}
&\lim_{n\to\infty} \mathrm{P}(|Y_{n,r} - Y_{\dot{S}_n}| \leqslant \boldsymbol{\epsilon}) \\
&= \int_{\mathbb{R}^k} \lim_{n\to\infty} \mathrm{P}(|Y_{n,r} - Y_{\dot{S}_n}| \leqslant \boldsymbol{\epsilon} \mid Y_{S_N} = y_{S_N}) f(y_{S_N}) \mathrm{d}y_{S_N} = \int_{\mathbb{R}^k} 1 f(y_{S_N}) \mathrm{d}y_{S_N} = 1.
\end{aligned}$$

Therefore $|Y_{n,r} - Y_{\dot{S}_n}| \xrightarrow{P} 0$.

Now let f_n and f_r be the pdf of $Y_{\dot{S}_n}$ and Y_u under the f distribution. Note that $f_n(y) = \int_{\mathbb{R}^k} f_n(y \mid S_N = s_N) f(s_N) \mathrm{d}s_N = \mathrm{E}_{S_N}[f_n(y \mid S_N)]$. Hence, we have that

$$f_n(y) - f(y) = \mathrm{E}_{S_N}[f_n(y \mid S_N) - f(y)] = \mathrm{E}_{S_N}[d(S_N, r)], \tag{A.7}$$

where $d(S_N, r)$ is a continuous function of (S_N, r). Since $\dot{S}_n \xrightarrow{p} r$, we have that $d(S_N, r) \xrightarrow{p} 0$. Now, because $d(S_N, r)$ is uniformly bounded, it is also uniformly integrable and, by Theorem 1.8 from Shao (2003), we have that $\mathrm{E}_{S_N}[d_{S_N,r}] \longrightarrow 0$. Therefore, $Y_{\dot{S}_n,f} \xrightarrow{d} Y_{r,f}$, where $Y_{\dot{S}_n,f}$ and $Y_{r,f}$ are the r.v.'s with pdf f at $\dot{S}_n$ and r respectively.

Finally, using all the three results above and by Slutsky's Theorem (see Resnick, 2013, Theorem 8.6.1 (a)), we have that $|Y_{n,r} - Y_{\dot{S}_n}| \xrightarrow{P} 0$ and $Y_{\dot{S}_n,f} \xrightarrow{d} Y_{r,f}$ implies $Y_{n,r} \xrightarrow{d} Y_{r,f}$. This concludes the proof. □

Proof of Proposition 2.1. We have the marginal density of a $Y_{\mathbf{s}}$, for some finite set of locations $\mathbf{s}$ is given by

$$p(Y_u \mid \Theta) = \int p(Y_u \mid Y_N, N) f(Y_N \mid N) p(N) \mathrm{d}\mathcal{P}(\mathrm{N}, \mathrm{Y_N}) \tag{A.8}$$

Now take $Y_{\mathbf{s}}$ and $Y_{\mathbf{s}'}$ for two symmetric sets $\mathbf{s}$ and $\mathbf{s}'$ w.r.t. the Poisson process N. Then, by the symmetry of the PP measure of N and the stationarity and isotropy of the base GP and f, we have that, for each $n \in \Omega_N$, there exists $n' \in \Omega_N$ such that $p(n) = p(n')$, $f(Y_n \mid n) = f(Y_{n'} \mid n')$ and $p(Y_{\mathbf{s}} \mid Y_n, n) = p(Y_{\mathbf{s}'} \mid Y_{n'}, n')$. Therefore, given the equality in A.8, we have that $p(Y_{\mathbf{s}} \mid \Theta) = p(Y_{\mathbf{s}'} \mid \Theta)$. □

Covariance of Y.

$$
\begin{aligned}
Cov(Y_1, Y_2) &= \mathrm{E}(Y_1 Y_2) - \mathrm{E}(Y_1)\mathrm{E}(Y_2) \\
&= \mathrm{E}_{Y_N}[\mathrm{E}_{Y_1 Y_2|Y_N}(Y_1 Y_2|Y_N)] - \mathrm{E}_{Y_N}[\mathrm{E}_{Y_1|Y_N}(Y_1|Y_N)]\mathrm{E}_{Y_N}[\mathrm{E}_{Y_2|Y_N}(Y_2|Y_N)] \\
&= \mathrm{E}_{Y_N}[Cov(Y_1, Y_2|Y_N)] + \mathrm{E}_{Y_N}[\mathrm{E}_{Y_1|Y_N}(Y_1|Y_N)\mathrm{E}_{Y_2|Y_N}(Y_2|Y_N)] \\
&- \mathrm{E}_{Y_N}[\mathrm{E}_{Y_1|Y_N}(Y_1|Y_N)]\mathrm{E}_{Y_N}[\mathrm{E}_{Y_2|Y_N}(Y_2|Y_N)] \\
&= \mathrm{E}_{Y_N}[Cov(Y_1, Y_2|Y_N)] + Cov[\mathrm{E}_{Y_1|Y_N}(Y_1|Y_N), \mathrm{E}_{Y_2|Y_N}(Y_2|Y_N)] \\
&= \mathrm{E}_{Y_N}[Cov(Y_1, Y_2|Y_N)] \\
&+ Cov[\boldsymbol{\mu}_1 + \Sigma_{1N}\Sigma_{NN}^{-1}(Y_N - \boldsymbol{\mu}_N), \boldsymbol{\mu}_2 + \Sigma_{2N}\Sigma_{NN}^{-1}(Y_N - \boldsymbol{\mu}_N)] \\
&= \mathrm{E}_{Y_N}[Cov(Y_1, Y_2|Y_N)] + \Sigma_{1N}\Sigma_{NN}^{-1} Cov(Y_N, Y_N)(\Sigma_{2N}\Sigma_{NN}^{-1})^\top \\
&= \mathrm{E}_{Y_N}[Cov(Y_1, Y_2|Y_N)] + \Sigma_{1N}\Sigma_{NN}^{-1}\Sigma_{NN,f}\Sigma_{NN}^{-1}\Sigma_{2N}^\top \\
&= \mathrm{E}_N[\mathrm{E}_{Y_N}[Cov(Y_1, Y_2|Y_N)|N] + \mathrm{E}_N[\Sigma_{1N}\Sigma_{NN}^{-1}\Sigma_{NN,f}\Sigma_{NN}^{-1}\Sigma_{2N}^\top|N]
\end{aligned}
$$

where $\Sigma_{NN,f}$ is the Y_N covariance function under the f's distribution and since $Y_i|Y_N \sim \mathcal{N}_{n_i}(\boldsymbol{\mu}_i + \Sigma_{iN}\Sigma_{NN}^{-1}(Y_N - \boldsymbol{\mu}_N), \Sigma_{ii} - \Sigma_{iN}\Sigma_{NN}^{-1}\Sigma_{Ni})$ for $i = 1, 2$.

If the covariance of the $'f'$ is the same as the base GP, then we have the following representation, when N is fixed.

$$
Cov(Y_1, Y_2) = \mathrm{E}_{Y_N}[Cov(Y_1, Y_2|Y_N)] + \Sigma_{1N}\Sigma_{NN}^{-1}\Sigma_{2N}^\top
$$

Now, considering the case where the covariance is isotropic.

$$
\begin{aligned}
Cov(Y_1, Y_2) &= \mathrm{E}_{Y_N}\left[Cov(Y_1, Y_2|Y_N)\right] \\
&+ \left\langle \sum_{i=1}^{N} \rho(|s_1 - \dot{s}_i|)\rho_{ij}^{-1} \right\rangle_{j=1}^{N} \Sigma_{NN,f} \left[\left\langle \sum_{i=1}^{N} \rho(|s_2 - \dot{s}_j|)\rho_{ij}^{-1} \right\rangle_{i=1}^{N} \right]^\top \\
&= \mathrm{E}_{Y_N}\left\{ \rho_{1,2} - \sum_{j=1}^{N} \left[\sum_{i=1}^{N} \rho(|s_1 - \dot{s}_i|)\rho_{ij}^{-1}\rho(|s_2 - \dot{s}_j|) \right] \right\} \\
&+ \sum_{l=1}^{N} \left[\sum_{k=1}^{N}\sum_{i=1}^{N} \rho(|s_1 - \dot{s}_i|)\rho_{ik}^{-1}\rho_{kl,f} \sum_{j=1}^{N} \rho_{lj}^{-1}\rho(|s_2 - \dot{s}_j|) \right] \\
&= \mathrm{E}_N\left\{ \rho_{1,2} - \sum_{j=1}^{N} \left[\sum_{i=1}^{N} \rho(|s_1 - \dot{s}_i|)\rho_{ij}^{-1}\rho(|s_2 - \dot{s}_j|) \right] \middle| N \right\} \\
&+ \mathrm{E}_N\left\{ \sum_{l=1}^{N} \left[\sum_{k=1}^{N}\sum_{i=1}^{N} \rho(|s_1 - \dot{s}_i|)\rho_{ik}^{-1}\rho_{kl,f} \sum_{j=1}^{N} \rho_{lj}^{-1}\rho(|s_2 - \dot{s}_j|) \right] \middle| N \right\}.
\end{aligned}
$$

□

A.2 MCMC Simulation

The following results come from the first scenario presented in the MCMC simulation, where f has the following parameters Copula $SN_m(\mu_{SN}, \Sigma_{SN}, \gamma = 0.75)$ with PP($\lambda = 1$) and $n = 200$.

Convergence diagnostic for the mean of the vector Y_N.

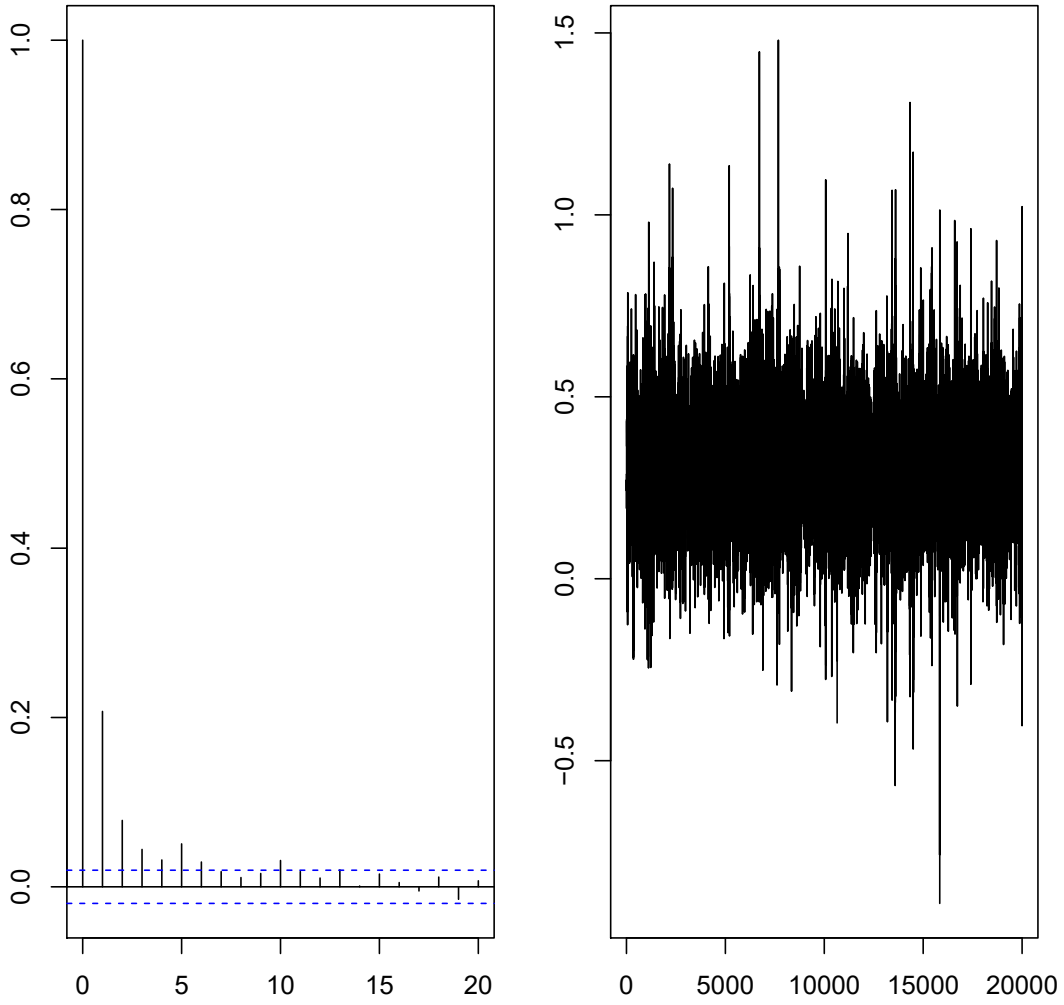


Figure A.1: Convergence diagnosis of the Y_N mean for the model $SN_{\gamma=0.75}; \lambda = 1; n = 200$ for one replication.

Convergence diagnostic for the number of the points N.

Convergence diagnostic for the block N, parameters μ, σ^2, ζ and γ.

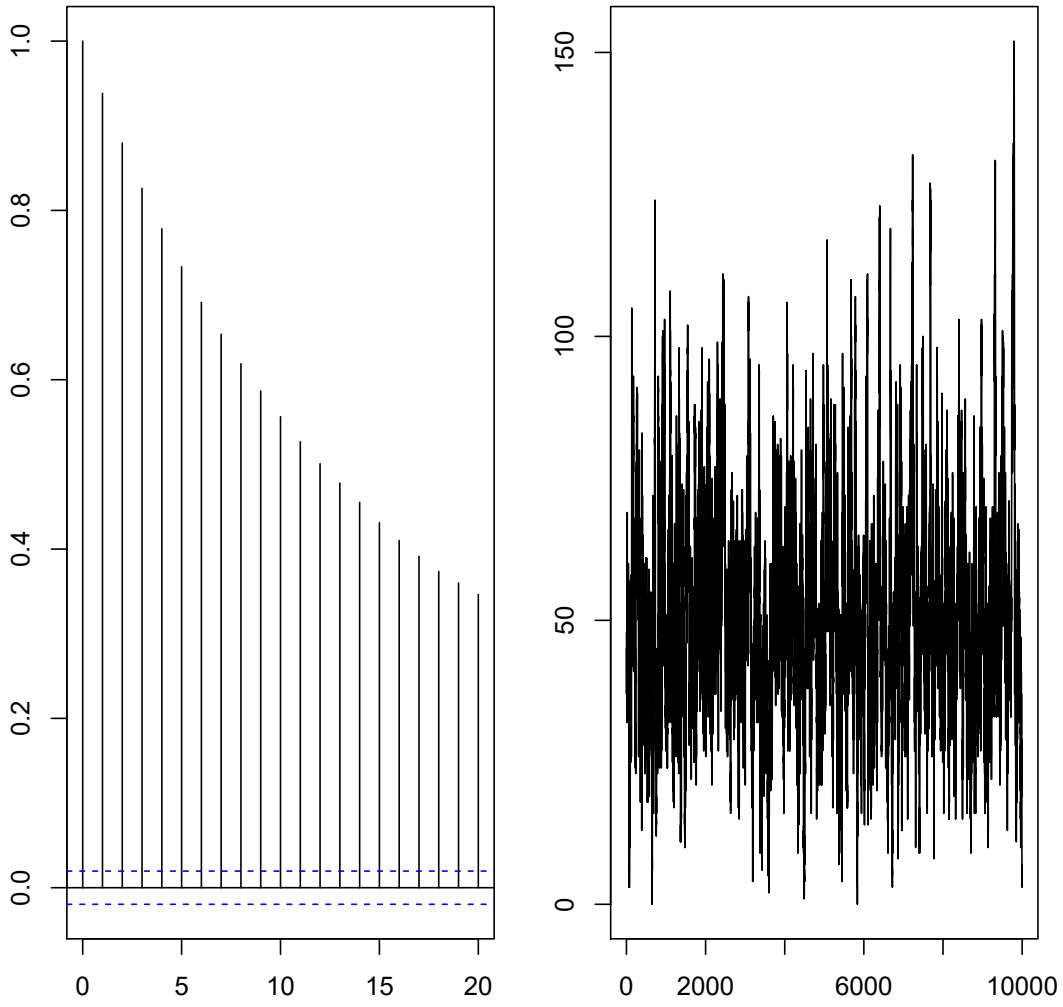


Figure A.2: Convergence diagnosis of the number of the points N for the model $SN_{\gamma=0.75}; \lambda = 1; n = 200$ for one replication.

A.3 MCMC Application

Convergence results referring to the MCMC application using 'f' as Skew-Normal.

Convergence diagnostic for the mean of the vector Y_N.

Convergence diagnostic for the number of the points N.

Convergence diagnostic for the block N, parameters μ, σ^2, ζ and γ.

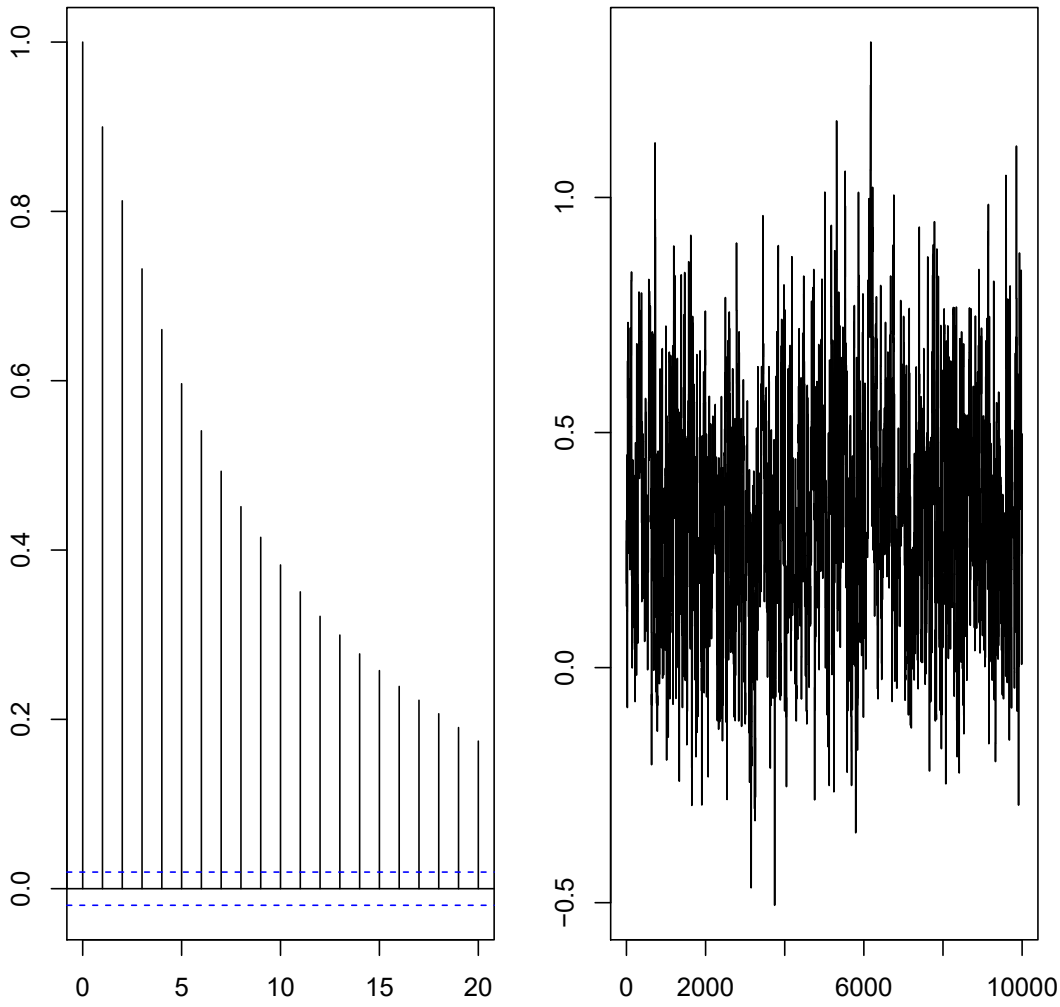


Figure A.3: Convergence diagnosis of the parameter μ for the model $SN_{\gamma=0.75}; \lambda = 1; n = 200$ for one replication.

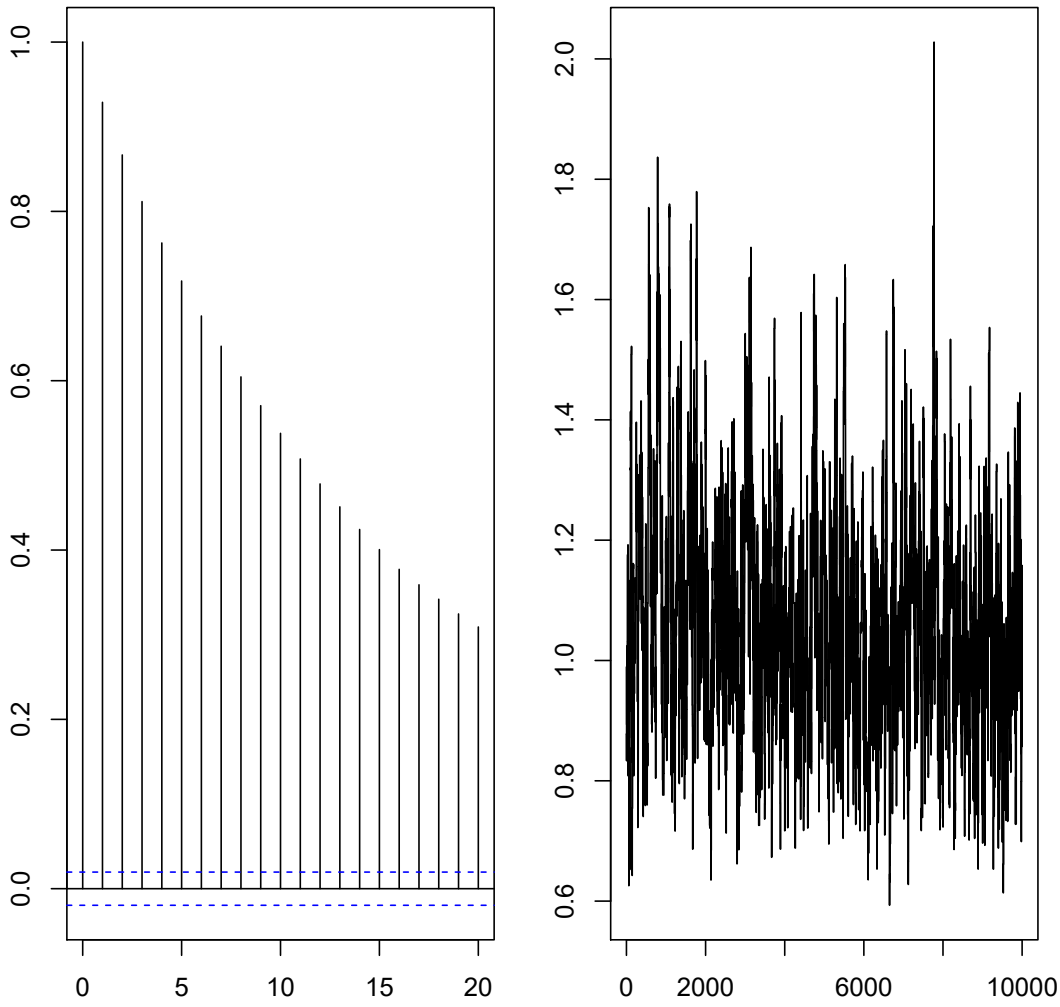


Figure A.4: Convergence diagnosis of the parameter σ^2 for the model $SN_{\gamma=0.75}; \lambda = 1; n = 200$ for one replication.

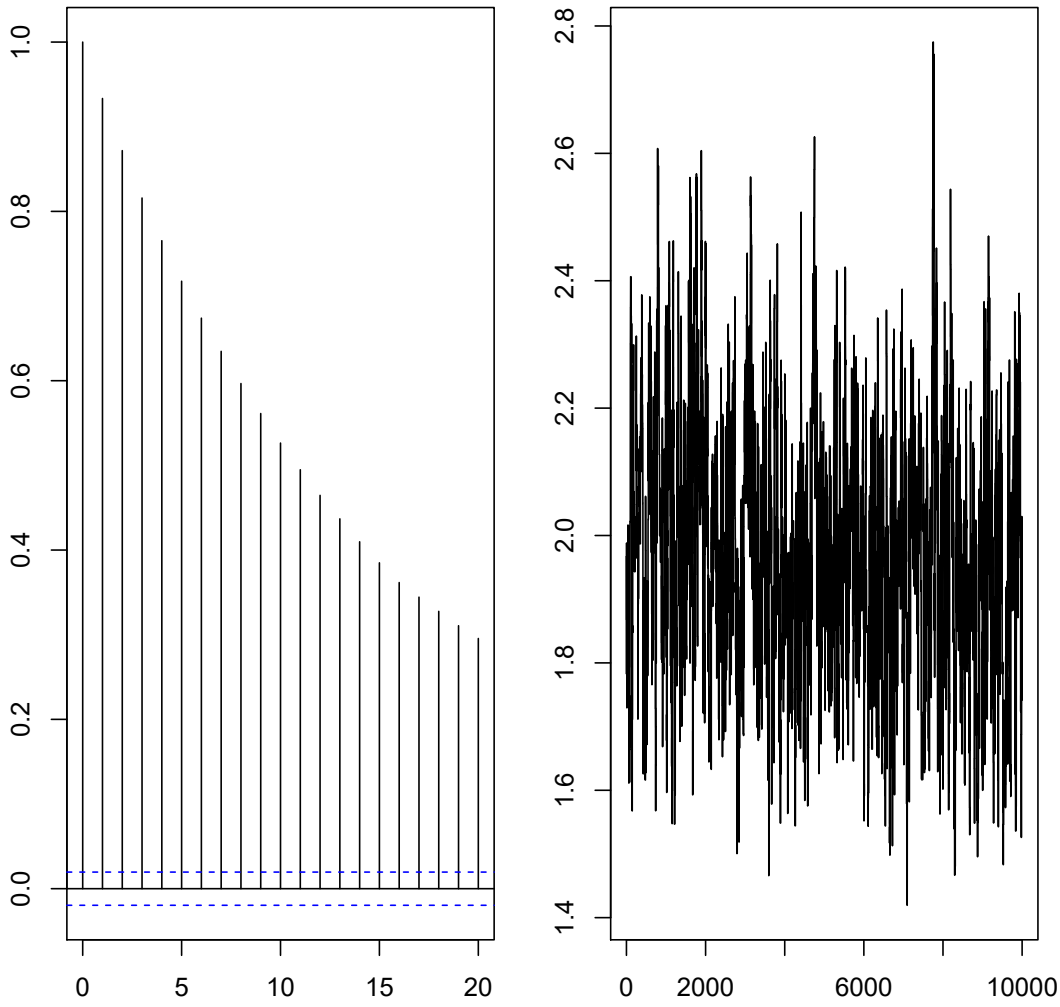


Figure A.5: Convergence diagnosis of the parameter ζ for the model $SN_{\gamma=0.75}; \lambda = 1; n = 200$ for one replication.

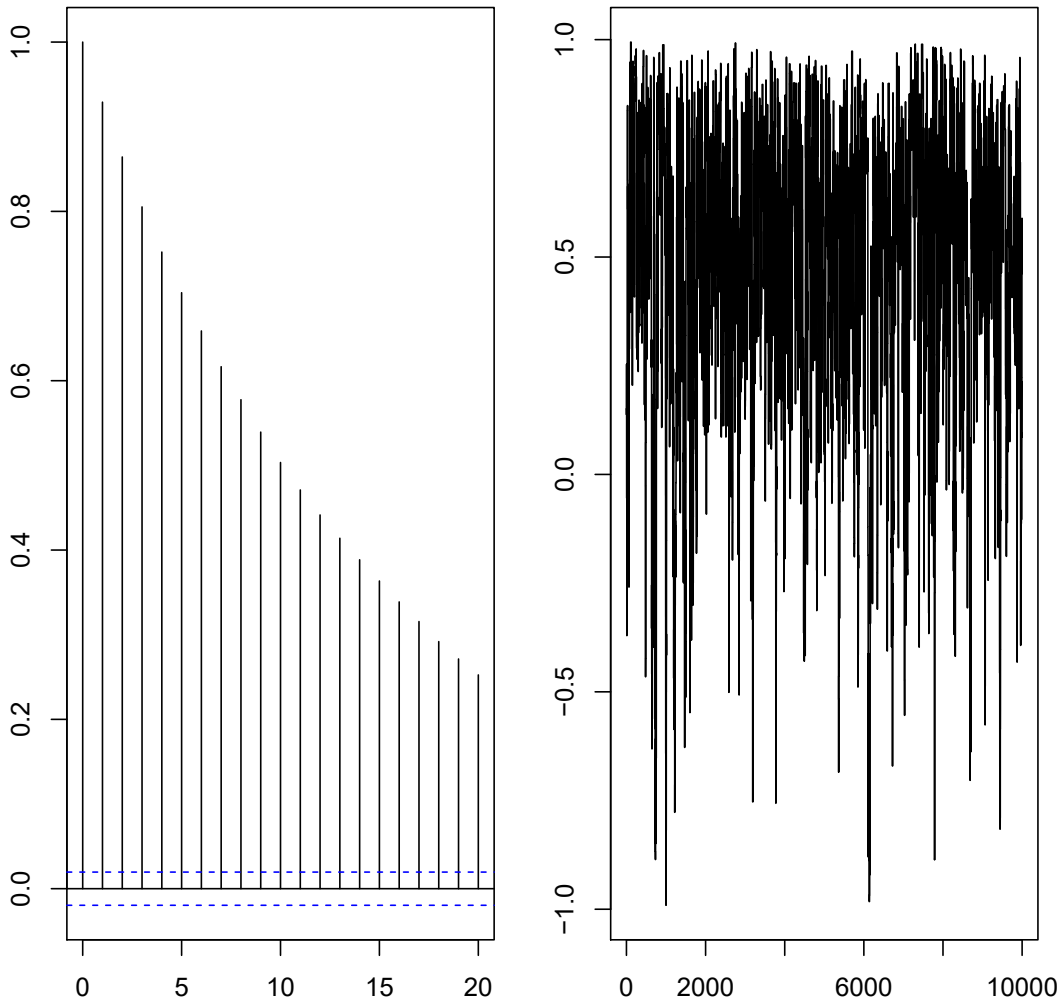


Figure A.6: Convergence diagnosis of the parameter γ for the model $SN_{\gamma=0.75}; \lambda = 1; n = 200$ for one replication.

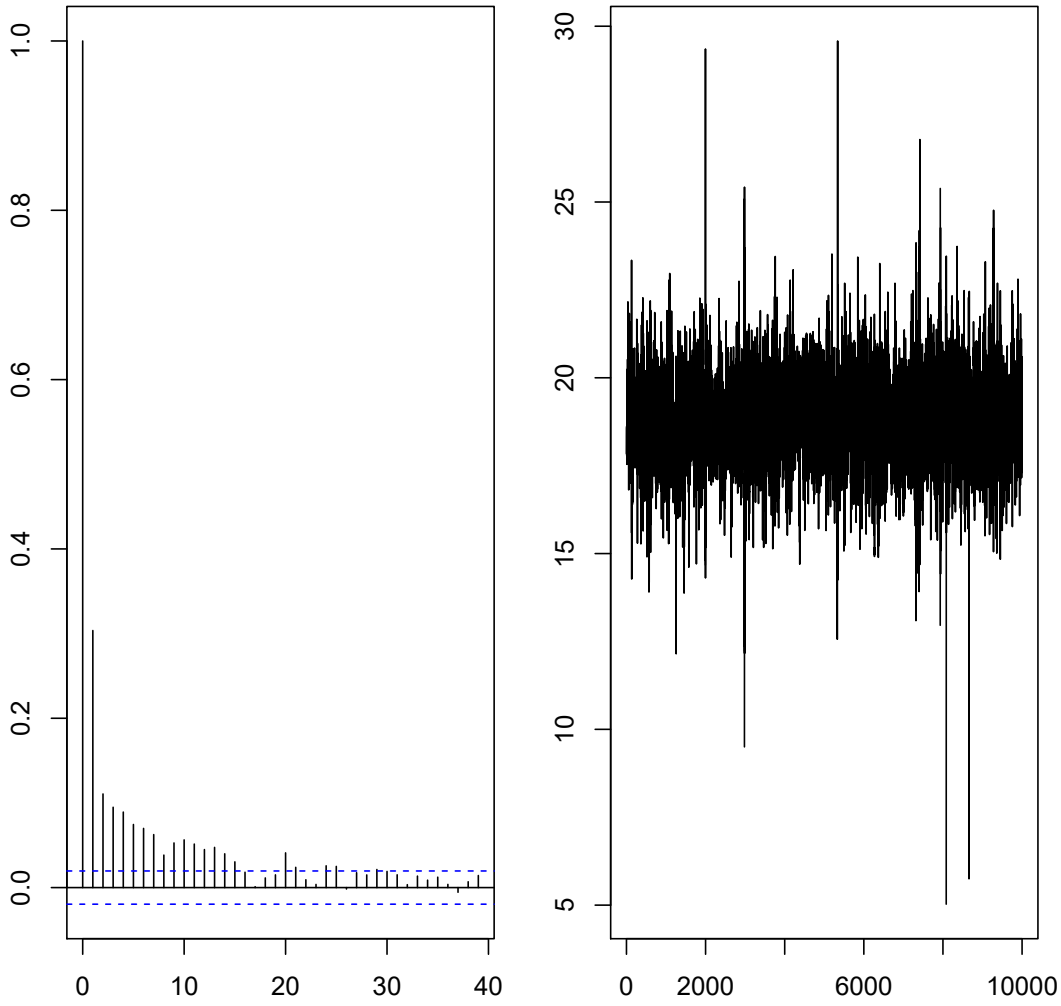


Figure A.7: Convergence diagnosis of the Y_N.

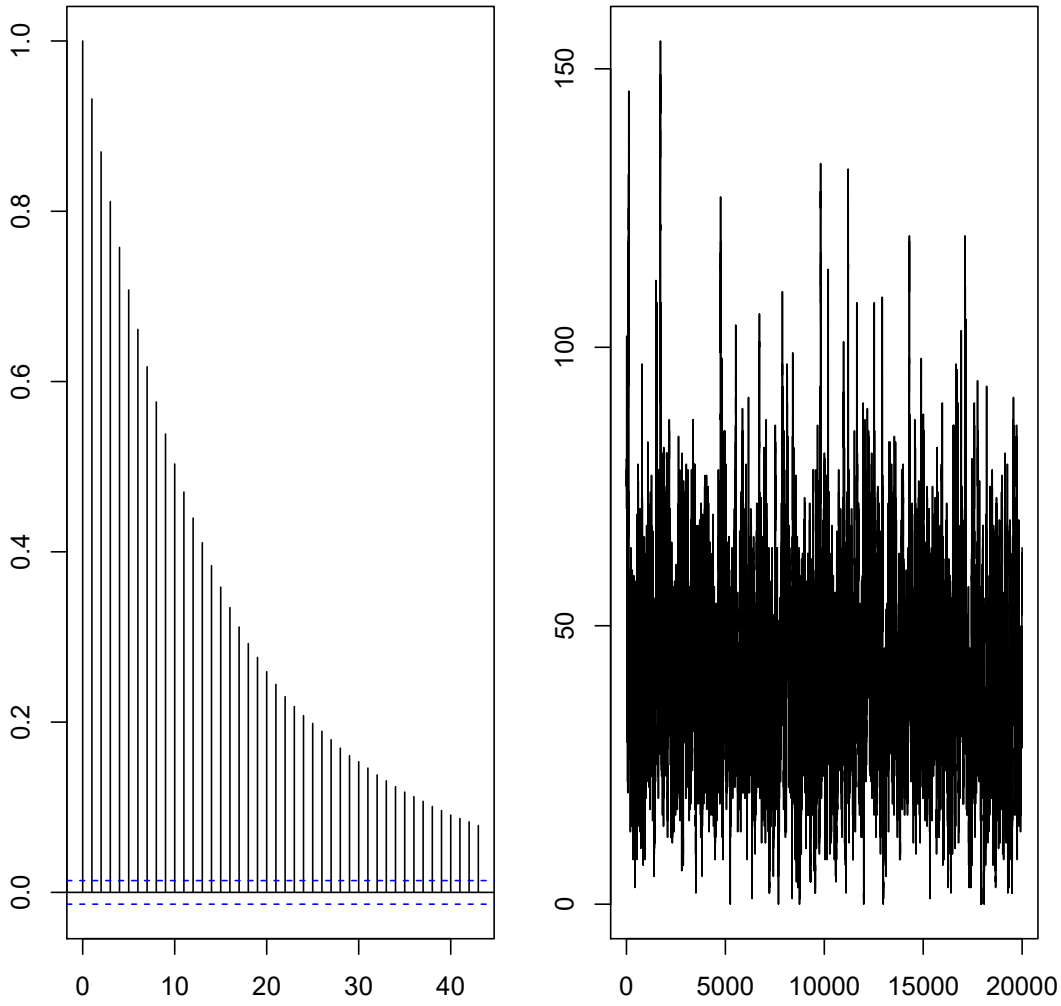


Figure A.8: Convergence diagnosis of the number of the points N.

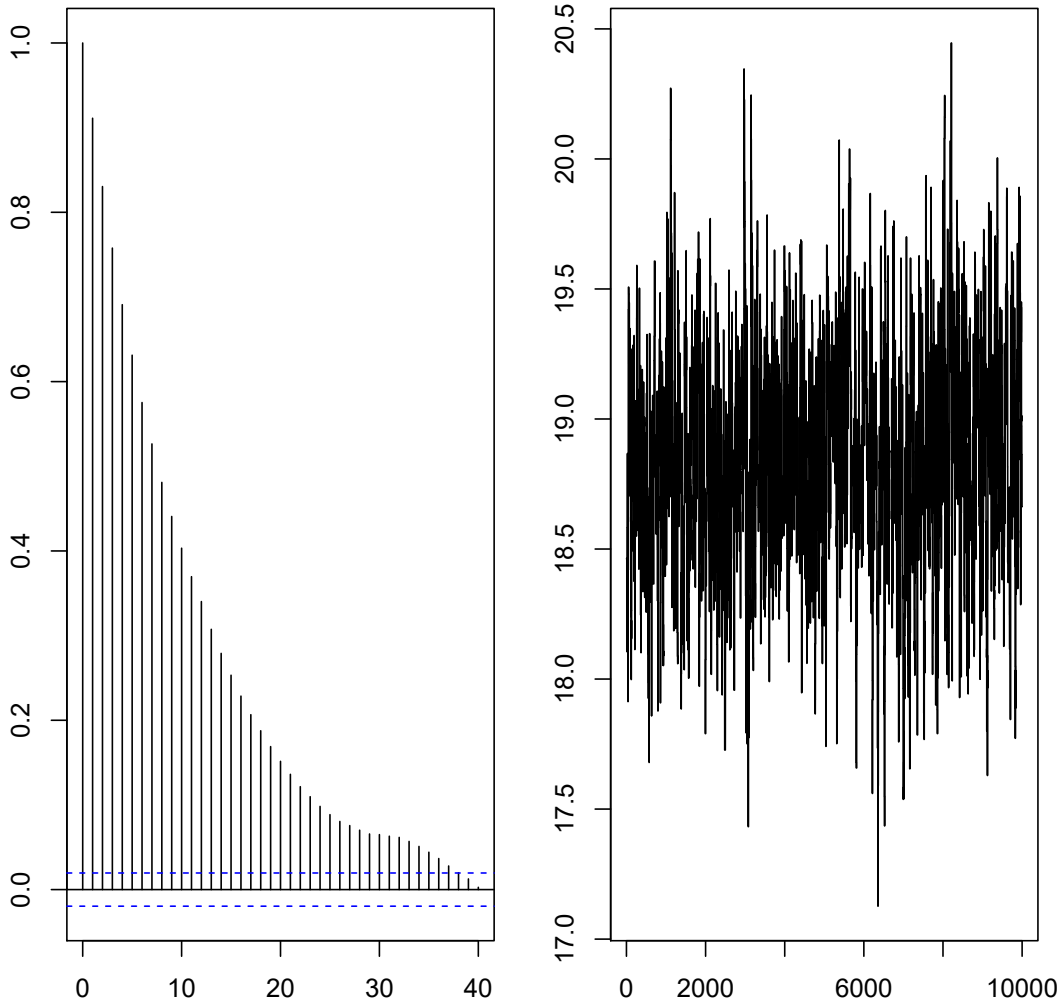


Figure A.9: Convergence diagnosis of the parameter μ.

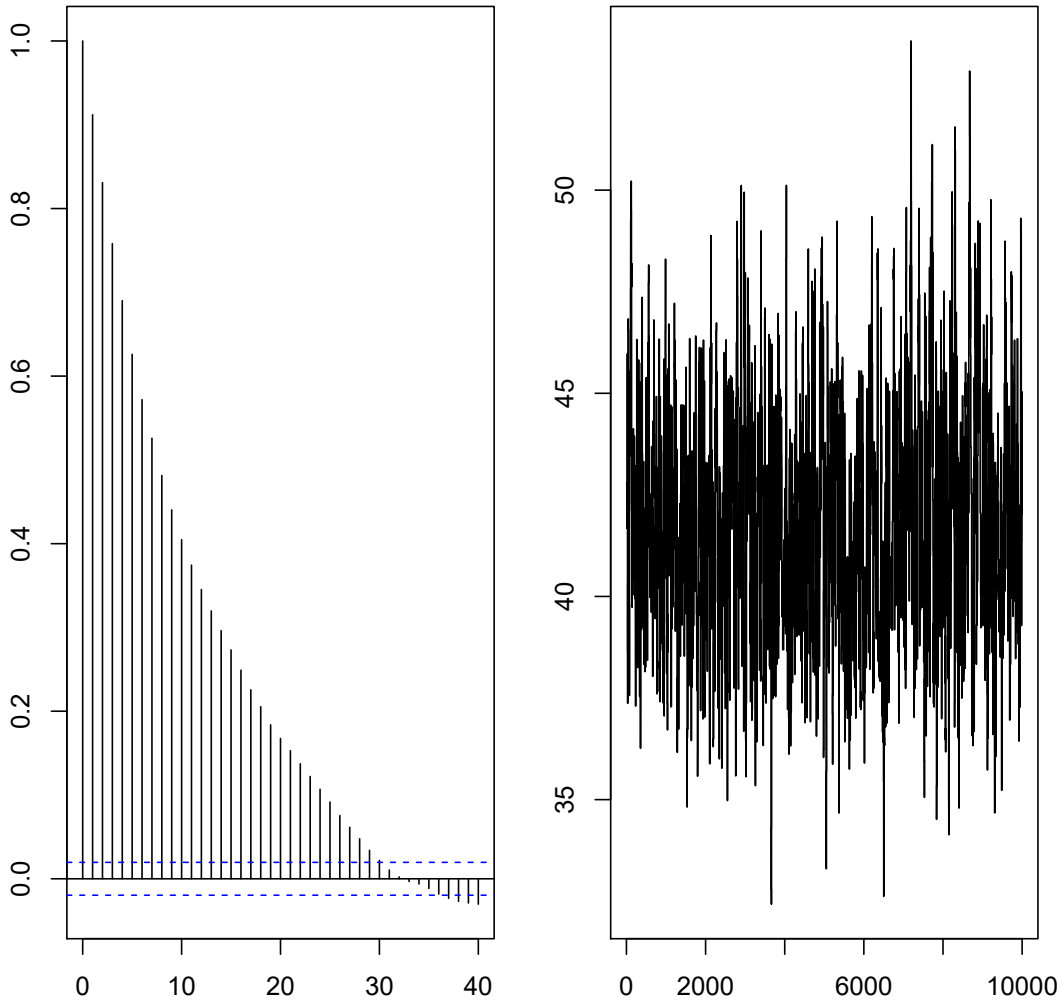


Figure A.10: Convergence diagnosis of the parameter σ^2.

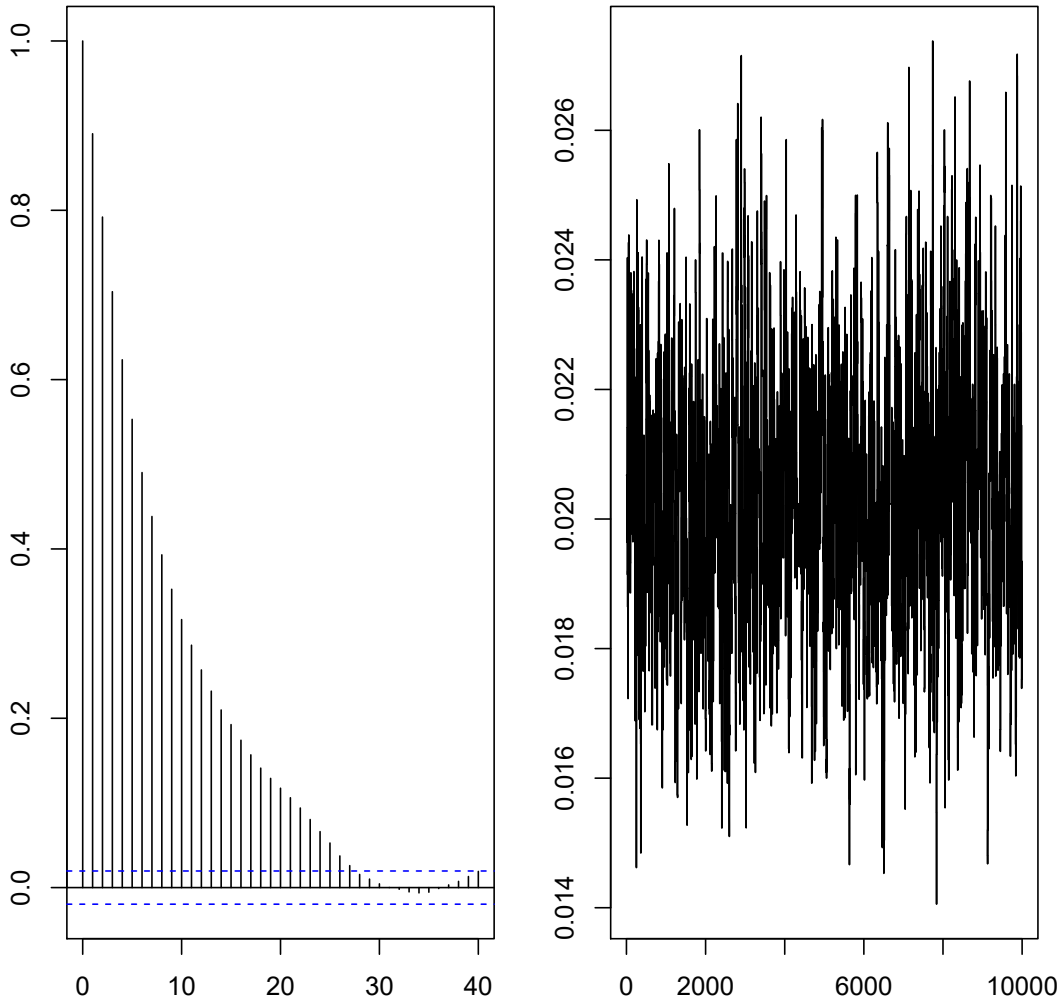


Figure A.11: Convergence diagnosis of the parameter ζ.

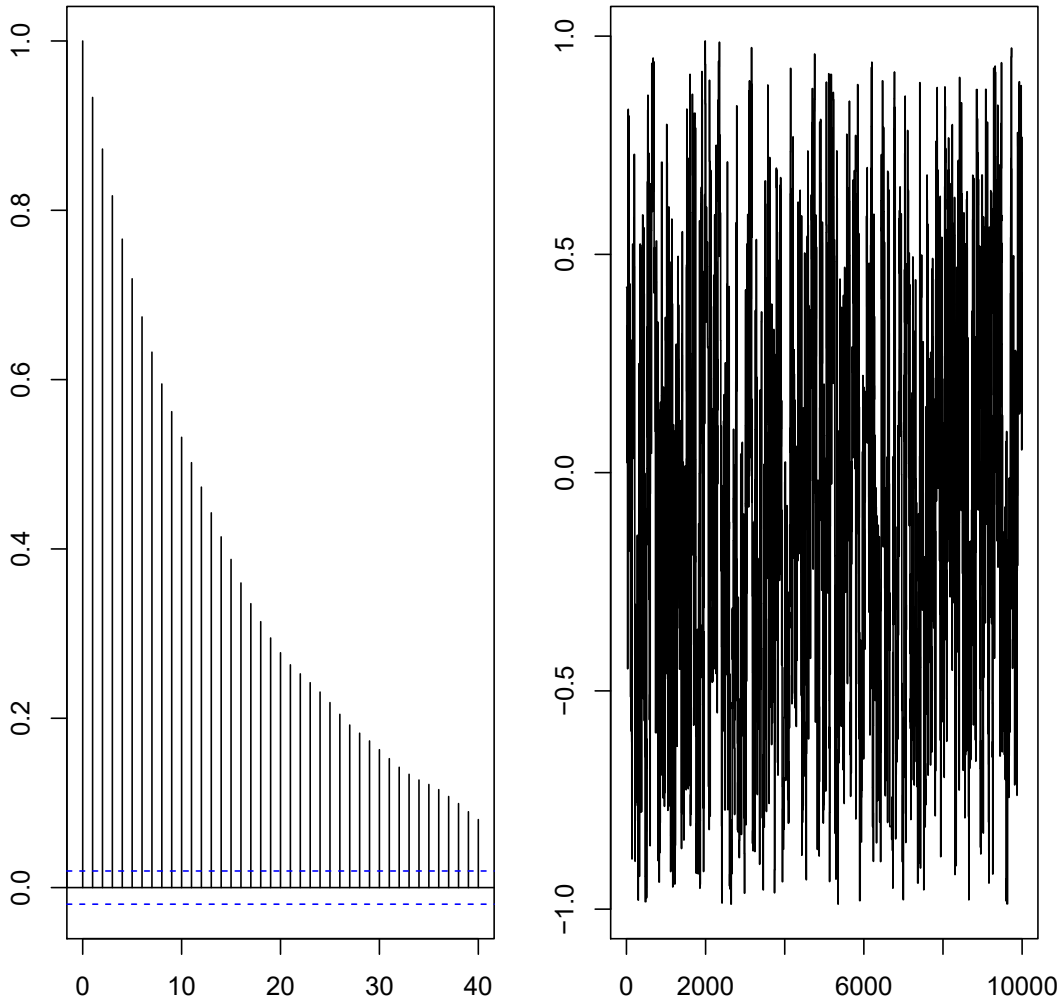


Figure A.12: Convergence diagnosis of the parameter γ.