

Article

# Spatial Statistical Models: An Overview under the Bayesian Approach

Francisco Louzada <sup>1</sup>, Diego Carvalho do Nascimento <sup>2,\*</sup> and Osafu Augustine Egbon <sup>1</sup>

<sup>1</sup> Institute of Mathematical Science and Computing, University of Sao Paulo, Sao Carlos 13566-590, Brazil; louzada@icmc.usp.br (F.L.); osafuagustine.egbon@usp.br (O.A.E.)

<sup>2</sup> Departamento de Matemática, Facultad de Ingeniería, Universidad de Atacama, Copiapó 1530000, Chile

\* Correspondence: diego.nascimento@uda.cl

**Abstract:** Spatial documentation is exponentially increasing given the availability of *Big Data in the Internet of Things*, enabled by device miniaturization and data storage capacity. Bayesian spatial statistics is a useful statistical tool to determine the dependence structure and hidden patterns in space through prior knowledge and data likelihood. However, this class of modeling is not yet well explored when compared to adopting classification and regression in machine-learning models, in which the assumption of the spatiotemporal independence of the data is often made, that is an inexistent or very weak dependence. Thus, this systematic review aims to address the main models presented in the literature over the past 20 years, identifying the gaps and research opportunities. Elements such as random fields, spatial domains, prior specification, the covariance function, and numerical approximations are discussed. This work explores the two subclasses of spatial smoothing: *global* and *local*.

**Keywords:** Bayesian spatial models; Bayesian inference; probability and statistical methods



**Citation:** Louzada, F.; Nascimento, D.C.d.; Egbon, O.A. Spatial Statistical Models: An Overview under the Bayesian Approach. *Axioms* **2021**, *10*, 307. <https://doi.org/10.3390/axioms10040307>

Academic Editor: Hari Mohan Srivastava

Received: 15 September 2021  
Accepted: 15 November 2021  
Published: 17 November 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Digital transformation technologies have generated massive amounts of data over the past few decades, which is the concept known as *Big Data*, in which data storage grows exponentially and requires an advanced analytical tool to explore and answer research questions. The technical advance has opened the door to inferential models of complex phenomena, such as spatial trends and heterogeneity in information conditioned on space and time [1,2]. Thus, adopting spatial methods in daily observed data can lead to more intelligent cities and to urban information to identify and prevent high-risk regions using the *Internet of Things* (IoT) [3]. Spatial dependencies have long been identified as a component that could hinder model precision and increase bias. Subsequent efforts to account for these errors have a research line in spatial statistics.

Space-based observation is an essential resource for IoT data, as it helps data prediction and analysis. Applications vary in complexity and are frequently carried out in risk surface detection, healthcare, agriculture, urban planning, economics, engineering, and rarely, smart appliances that learn based on location. This complex structure is accommodated in a flexible class of models related to observed data and spatial dependencies. Frequentist (classical) and Bayesian analytical methods have been used to analyze spatially varying phenomena. However, the Bayesian method is a better choice because it presents the characteristic of accommodating information from different sources. In the Bayesian framework, questions are answered through an estimation procedure by combining multiple sources of information, such as previous knowledge (prior) and the acquired information in the data (likelihood) [4].

In the field of neuroscience, the area of neurorehabilitation and the technological advances made in neuro-navigation have grown, thus personalizing the definition of transcranial accuracy [5,6]. Recently, there has been a growing interest in using Bayesian spatial

models in the meta-analysis of brain imagery to locate regions of consistent activation in the brain for diagnosis and treatment [7,8]. Another example is in epidemiology, in which patterns across geographical spaces are used to identify the areas of potentially elevated risk and create disease maps to quantify the underlying risk surface [9]. The same is true in the fields of geostatistics, solid and fluid mechanics, materials science, agriculture, Earth sciences, environmental science, etc. For instance, the Bayesian approach has been adopted to identify and map fish species richness and abundant hotspots for management and conservation of fish stocks [10]; for the calibration of computer models to quantify model uncertainty and evaluate model adequacy [11,12]; in the analysis of the performance of electrical capacitance tomography for the exploration of solid flow [13]; for multiple output separable Gaussian process emulation for a dynamic simulator to tackle problems in uncertainty and sensitivity analysis and the calibration of dynamic models [14]; for the identification of elastoplastic material parameters and incorporating model uncertainty through a Gaussian random process with a stationary covariance function [15]. Moreover, the medical literature provides detailed motivations and descriptions of spatial smoothing methods by explaining the concepts, defining the technical terms, and demonstrating various visualizations of spatial models.

The basic idea of Bayesian spatial statistics is an extension of the generalized linear model, including a spatial latent field, commonly referred to as a random field, which accounts for spatial dependencies across a spatial domain  $\mathcal{D}$ . A random field is a stochastic process indexed in space  $\mathcal{D}$  that takes values from a high-dimensional space and is defined over a parameter space of a dimensionality of at least one. The spatial latent field is modeled as a stochastic process whose uncertainty is represented by a random field probability density function,  $f_{spat}$ , which accounts for the spatial correlation or heterogeneity over a region (not necessarily delimited). The most commonly used stochastic process is the Gaussian field. In the context of spatial statistics, the Gaussian field is a stochastic process indexed by the space in which any finite collection of random variables has a multivariate normal distribution, whose covariance matrix is a function of the distance between these locations in space. Thus, a multivariate normal distribution can be referred to as a Gaussian random field.

For the interpolation task, the estimate of  $f_{spat}$  is nonlinearly interpolated across the unsampled spatial region with a specified resolution to identify the hotspot regions and provide intuition into the event chain. Kriging is one of the most frequently used techniques for spatial interpolation [16] governed by prior covariance, which gives unbiased projections on the unsampled region using a linear combination of neighboring values. A recently developed technique for spatial interpolation is the stochastic partial differential equation [17]. It is increasingly used as it is easily incorporated into a generalized hierarchical mixed-effect modeling framework and uses some bases function to make projections into unsampled regions. However, the approach is different from the frequently used spatial econometric models, which estimate spatial dependencies as a global correlation parameter across a spatial domain [18].

In the kernel of each random field model, there is a correlation matrix that quantifies the dependencies between random variables at different spatial locations and determines the full covariance structure of the stochastic process. The correlation matrix, a function of the kernel, has an enormous impact on spatial smoothing, and the challenge is how to choose its specification. A frequently used technique in epidemiology [19] and econometrics [20] to derive the spatial dependence through the covariance matrix governing the spatial field is the weight matrix with a first-order rook, bishop, or queen kind of binary neighborhood structure, which assigns one or zero depending on whether the spatial locations are immediate neighbors or not, which simplifies the complex spatial pattern. However, in many fields, a geostatistical model is often adopted, which considers the domain of exploration as a continuous spatial domain. In this manner, the uncertainty in the stochastic process is completely determined by a covariance function, such as the Matern or Gaussian covariance function. The model complexity can be approximated and

transcribed through a covariance structure. A highly dense covariance matrix implies a highly correlated spatial field and a more complex model.

Hernández-Lemus [21] discussed the historical origin and association between random fields and the joint probability function representation. As a special case, Markov random fields are visualized as a graphical model such as Markov networks or Bayesian networks (undirected and directed graphs).

### 1.1. Aim

A Bayesian spatial statistic aims to quantify spatial patterns and provide insight into the process of generating a pattern. Given the technological advances and, precisely, the storage capacity, georeferenced data acquisition is commonly present in domain sets currently. The Bayesian method is typically used to analyze these sets and to identify spatial patterns. Consequently, spatial statistics has received considerable attention in recent years, and numerous spatial models have been proposed and applied in diverse research fields. The systematic review of these models has received little attention and has specifically been conducted in epidemiology [22–24].

This systematic review focuses on the progressive development and content analysis of the Bayesian spatial models and aims to bridge the discontinuities in the literature. It aims to provide an overview and a basic knowledge of the concepts and the improvements over the last 20 years and identify the key research directions and areas of opportunity in the Bayesian spatial methodology.

### 1.2. Outline

R. A. Fisher identified the implication of spatial dependence in statistical analysis [25]. He introduced blocking in a completely randomized design to mitigate the error induced by spatial dependencies. For several years, there have not been many changes to the basic idea of Fisher's characterization of spatial dependencies, that is nearby locations/regions have a similar tendency without much interrelation complexity. Thus, we aimed to clarify this vital topic towards the popularization and development of the spatial modeling field.

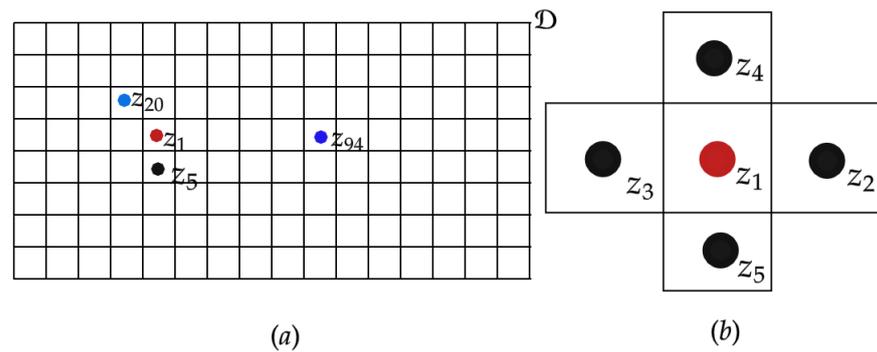
This systematic review is structured into three main parts. In the section entitled *Survey Methodology*, we describe the guidelines adopted in this work. The section called *Conceptual Scheme for Spatial Models* describes the main spatial models found in the Bayesian spatial literature. Then, the *Analyses* Section shows the empirical results obtained from the meta-analysis in papers published over the last 20 years. Finally, the conclusions are drawn in the last section.

### 1.3. Spatial Modeling Overview

Let us consider a spatial domain of an  $8 \times 16$  rectangular lattice and a variable  $z_i$ ,  $i = 1, 2, \dots, n$ , and  $n = 2^7$ . Each  $z_i$  is a latent variable in a unit square area of the spatial domain  $\mathcal{D}$ , as shown in Figure 1a. Let the latent variable  $z_i \in \mathbb{R}$  be modeled as:

$$z_i = \mathbf{x}_i^T \boldsymbol{\beta} + \gamma_i, \quad (1)$$

in which  $\mathbf{x}_i$  is a  $p \times 1$  spatial covariate predicting  $z_i$ ,  $\boldsymbol{\beta} \in \mathbb{R}^p$  is an unknown regression coefficient,  $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_n)^T \in \mathbb{R}^n$  is a Gaussian Markov Random Field (GMRF) denoted as  $\boldsymbol{\gamma} \sim N(\mathbf{0}, \mathbf{Q}^{-1}(\alpha))$ , and  $\mathbf{Q}^{-1}(\alpha) \in \mathbb{R}^{n \times n}$  is a covariance matrix with hyperparameter  $\alpha \in \mathbb{R}$ . Figure 1b shows the rook neighborhood type.



**Figure 1.** (a) A 2-dimensional topology divided into an  $8 \times 16$  regular rectangular spatial domain  $\mathcal{D}$ . Each square box is represented by a latent variable  $z$ . At these locations, we intend to determine  $z_i, i = 1, 2, \dots, (8 \times 16)$ , which explains the overall pattern in the spatial domain. To do so, we observe noisy samples  $y_i$  at each spatial site. The observed samples are denoised and used to predict  $z$ . Rectangular lattice data of this nature are common in agriculture science, environmental science, neuroimaging, etc. (b) A 1st-order rook type of spatial dependence for the field  $z_1$ . The illustrative diagram shows the relationship between  $z_1$  and its four neighbors ( $z_2, z_3, z_4, z_5$ ). In the neighborhood structure, the  $j$ th entry of the binary neighborhood matrix is set to 1,  $j \in \{2, 3, 4, 5\}$ . The total number of neighbors is four  $\forall i$ , except those at the edges, which have either two or three neighbors.

Let us suppose we observe an outcome  $y_i$  at each spatial site  $i, i = 1, 2, \dots, n$  with additive noise  $e_i$  independent of  $\gamma_i$ , then:

$$y_i = z_i + e_i, \tag{2}$$

and  $e_i \sim N(0, \tau^{-2})$  for all  $i$ . By this model specification, the density of  $y$  is a Gaussian distribution with mean  $\mathbb{E}(y_i | \beta, \gamma_i) = \mathbf{x}_i^T \beta + \gamma_i$ , and for  $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$ , the covariance matrix  $\text{var}(\mathbf{y} | \beta, \gamma) = \tau^{-2} I$ . In a simple case, the GMRF precision matrix  $\mathbf{Q}$  can be specified assuming the rook neighborhood spatial structure in Figure 1b. In this case,  $\mathbf{Q}(\alpha) = \frac{1}{\alpha} (\mathbf{D} - \mathbf{W})$ , in which  $\mathbf{W}$  is an  $n \times n$  matrix of entries  $w_{ij} = 1$  if site  $i$  and  $j$  are neighbors and zero otherwise. The diagonal elements of  $\mathbf{W}$  are set to zero as each site is not considered a neighbor of itself.  $\mathbf{D}$  is an  $n \times n$  diagonal matrix with entry  $d_{ii}$  (the number of neighbors of spatial site  $i$ ). In this example,  $d_{ii} = 4, \forall i$  except those  $i$  ( $d_{ii} = 2$  or  $d_{ii} = 3$ ) at the edges of  $\mathcal{D}$ , and  $\alpha \in \mathbb{R}^+$ .

The main interest is to estimate the uncertainties about  $\beta$  and  $\gamma$  and make predictions of the latent variable  $z$ . In a Bayesian framework, for the prior distribution on  $\beta, \tau, \sigma^2$ , and  $\alpha$  are either elicited depending on the available information or are assigned vague priors, allowing the data to decide. In this example, suppose  $\beta \sim N(\mathbf{0}, \sigma^2 I)$  and  $\phi = \{\alpha, \sigma^2, \tau\}$ . Usually, the elements of  $\phi$  or reparameterized  $\phi$  are assigned an appropriate prior distribution. For instance,  $\alpha$  and  $\sigma^2$  can be assigned an inverse gamma distribution and  $\tau$  a gamma distribution eliciting its hyperparameters from expert knowledge or data.

However, for this example, assume that it is known (fixed), for simplicity. Thus, the derivation of the conditional posterior distributions from the joint posterior is trivial. The joint posterior distribution and posterior conditional distribution for  $\beta$  and  $\gamma$  are respectively given as:

$$\begin{aligned} f(\beta, \gamma | \mathbf{y}, \phi) &\propto N(\mathbf{X}^T \beta + \gamma, \tau^{-2} I) * N(\mathbf{0}, \mathbf{Q}^{-1}(\alpha)) * N(0, \sigma^2 I), \\ \beta | \mathbf{y}, \gamma, \phi &\sim N\left(\tau^2 I (\mathbf{X} \tau^2 I \mathbf{X}^T + \sigma^{-2} I)^{-1} \mathbf{X} (\mathbf{y} - \gamma), (\mathbf{X} \tau^2 I \mathbf{X}^T + \sigma^{-2} I)^{-1}\right), \\ \gamma | \mathbf{y}, \beta, \phi &\sim N\left(\tau^2 I (\tau^2 I + \mathbf{Q}(\alpha))^{-1} (\mathbf{y} - \mathbf{X}^T \beta), (\tau^2 I + \mathbf{Q}(\alpha))^{-1}\right), \end{aligned} \tag{3}$$

in which  $\mathbf{X}$  is a  $p \times n$  matrix. The marginal posterior distribution is easily obtained using the Gibbs sampling algorithm.

In some applications, the process could have a multiplicative noise ( $y_i = z_i e_i$ ), binary outcome  $y$ , non-Gaussian error  $e_i$ , non-Markov Gaussian field  $\gamma$ , non-Gaussian field  $\gamma$ , or irregular or continuous spatial domain  $\mathcal{D}$ . In these cases, the modeling framework becomes more complicated than the specification given in the example.

In general, let  $y_i$  be the observed outcome having a probability distribution function  $f(\cdot|\theta_i)$ , with parameter  $\theta_i = (\theta_i^1, \dots, \theta_i^q) \in \mathbb{R}^q, \forall i$ . Suppose that  $\theta_i^j \in \mathcal{X}$ , a component of  $\theta_i$ , defined by  $(\beta_i, \gamma_i)$ , is of interest and the remaining components  $(\theta_i^{-j})$  are hyperparameters (i.e.,  $\phi_i$  and the others that will be discussed in the sequence). The structural linear predictor is formulated as:

$$g(\theta_i^j) = \mathbf{x}_i^T \beta + \gamma_i, i = 1, 2, \dots, n, \tag{4}$$

in which  $g(\cdot): \mathcal{X} \rightarrow \mathbb{R}$  is a link function and  $\mathcal{X}$  is the parameter space of  $\theta_i^j, \forall i$ .  $\mathbf{x}$  is the fixed-effect design vector.  $\beta$  is the unknown fixed-effect vector.  $\gamma$  is a random field, whose uncertainty is to be quantified. The choice of the probability distribution assumed for  $f(y|\theta)$  guides the researcher on the choice of  $g(\cdot)$ . Examples of commonly used canonical link functions include the identity ( $g(\theta) = \theta$ ), logit ( $g(\theta) = \log(\theta/(1 - \theta))$ ), and log ( $g(\theta) = \log(\theta)$ ).

The spatial effect is assigned a random field model  $f_{spat}(\gamma|\psi, \mathbf{Q}(\alpha))$  over  $\mathcal{D}$ , in which  $\psi$  is the vector of hyperparameters accounting for dispersion, and the spatial information is coded in the correlation matrix  $\mathbf{Q}(\alpha)$ , which accounts for the spatial dependencies of  $\gamma$  across a spatial domain with hyperparameter  $\alpha$ . In the Bayesian framework,  $f_{spat}(\cdot, \cdot)$  is usually multidimensional, representing the prior knowledge on the spatial latent field, and assumes a probability density function (not necessarily an exponential family) such as Gaussian; thus,  $\gamma$  is referred to as the Gaussian random field. Other common probability distributions include asymmetric Laplace, Student-t, and log-Gamma. The choice of the density of  $\gamma$  is informed by the level of spatial detail and the process complexity. Thus, it is not straightforward to choose the right model to learn the spatial process. For instance, in Figure 1a, an increase in the grid resolution might give a minimal model uncertainty as more samples would be drawn, but leading to a higher dimension of  $\gamma$ , thus increasing the model complexity. The choice of the prior distribution is not general, but informed by the objective of the study and the stochastic process governing  $\gamma$ .  $\alpha$  is a vector of hyperparameters that controls the spatial range and smoothness of the covariance function defining the covariance matrix of the spatial model. It is a global smoothing if the same set of  $\alpha$  smooths the entire spatial region, whereas it is a local smoothing if different sets smooth the spatial region. Additionally, the spatial process can be modeled in a semiparametric framework, such as the spatial mixture model, and the Dirichlet process.

In the frequentist paradigm, the model parameters are fixed and inferences are made by maximizing the joint log-likelihood over the parameter space. However, inference in the Bayesian framework is based on the Bayes theorem, which allows the uncertainty representation of the model parameters by a probability distribution and is updated as more information becomes available. Thus, the model parameters are considered random. Suppose that  $\beta | \Sigma \sim f(\cdot|\Sigma)$  is a centered distribution with dispersion matrix  $\Sigma$  representing the uncertainty about covariables/features as fixed effects ( $\beta$ ) and  $f(\gamma | \cdot)$  is the spatial effect model, i.e., spatial prior for  $\gamma$  before observing the data (based on the prior distribution or expert prior knowledge). Let  $\phi = \{\theta^{-j}, \psi, \Sigma\}$  and  $\alpha$  be assigned a joint prior distribution  $f(\phi)$  and  $f(\alpha)$ , respectively. The Bayes update is given as:

$$f(\beta, \gamma, \phi, \alpha|y) = \frac{f(\alpha)f(\phi)f(\beta|\Sigma)f(\gamma|\psi, \Sigma(\alpha))f(y|\theta_i^{-j}\beta, \gamma)}{\int f(\alpha)f(\phi)f(\beta|\Sigma)f(\gamma|\psi, \Sigma(\alpha))f(y|\theta_i^{-j}\beta, \gamma)dy}, \tag{5}$$

Inferences about the spatial component  $\gamma$  are derived from the marginal posterior distribution:

$$\begin{aligned} f(\gamma|y) &= \int f(\beta, \gamma, \phi, \alpha|y) d\beta d\phi d\alpha, \\ f(\beta|y) &= \int f(\beta, \gamma, \phi, \alpha|y) d\gamma d\phi d\alpha. \end{aligned} \quad (6)$$

The hyperparameters of the prior distributions are elicited from the expert opinion. The elicitation allows experts, practitioners, and nonstatisticians to contribute with their judgments about the spatial process, which are scientifically incorporated into a probability distribution. For example, experts could be asked to provide summary statistics (mean, mode, and quantiles) of the process, which are translated into a complete probability distribution. The elicitation technique minimizes the uncertainty of the parameters and leads to better inference. In addition, the power prior is a standard technique to construct an informative prior distribution from historical data and has been useful in clinical research [26]. The elicitation technique for a high-dimensional spatial process is not a trivial task. The level of difficulty increases for nonstationary stochastic processes, and caution is required as the wrong elicitation could lead to misleading inference. Different choices of priors are the weakly informative priors and objective priors, in which the former uses minimal subjective information to encode a prior distribution and the latter is not subjectively elicited.

In some cases, the parameter  $\alpha$  of the correlation function defined for the spatial model is fixed and  $f(\alpha)$  is excluded from Equation (5). “Fixed” means that they are first estimated from the data before performing Bayesian inference. This is done to minimize the model complexity. For example, the maximum likelihood estimate of the Matern covariance function parameters can be obtained from the data to determine the process’s spatial range and smoothness ([27], p. 58). These estimates are used as hyperparameters  $\Sigma(\alpha)$  in Bayesian inference.

## 2. Research Methodology

The data collected in this study aimed to determine the field in which Bayesian spatial statistics was most applied, as well as the current stage of development of these spatial models, identifying their trends and contributions to the Bayesian spatial literature. The collection and reporting methods were based on the guidelines of the Preferred Reporting Item for Systematic Review and Meta-Analysis (PRISMA) [28,29]. This procedure includes an electronic search strategy, a clear objective to define the inclusion and exclusion criterion, and an appropriate method for reporting the findings.

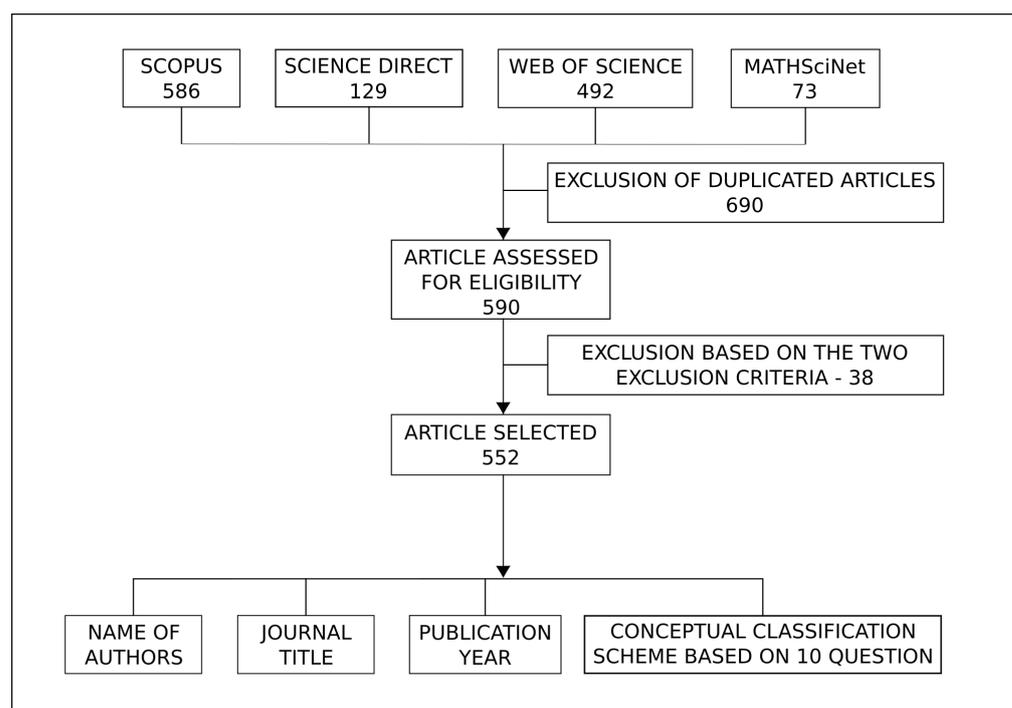
An online electronic search was conducted on 10 June 2020, in the following four databases: Elsevier’s Scopus, Science Direct, Thompson Reuters Web of Science, and the American Mathematical Society’s MathSciNet database. Queries of the word “Bayesian Spatial” and “Bayesian spatial” were performed, using the Boolean operator “OR”, throughout 2001–2020. Title, abstract, and keywords were used in Scopus and Science Direct, the topic (which entails title, abstract, and keywords) in Web of Science, and “Anywhere” in MathSciNet.

The time frame was chosen to capture the diversification of the application of Bayesian spatial models in various fields due to the improvement in technology for data collection and computation. The Mendeley Windows application was used to remove duplicate articles. The resulting set was further examined manually looking for more duplicates not identified by Mendeley’s application. The titles and abstracts of the articles included (after removing duplicates) were first screened for the Bayesian spatial methodology before applying the following inclusion criteria:

- Search results that were written in English and articles published in peer-reviewed journals available online. We excluded books, dissertations/theses, conference proceedings, and reviews (or any other form that was not an article);

- Articles that specifically implemented Bayesian spatial models, excluding the ones that only mentioned Bayesian spatial models.

Articles that did not meet the two inclusion criteria were excluded from the review. The search flowchart is presented in Figure 2. Using the search keywords mentioned earlier, 586 articles were retrieved from Scopus, 129 from Science direct, 492 from Web of Science, and 73 from MathSciNet. After excluding duplicated ones, 590 articles were assessed for eligibility, and 38 were further excluded based on the two exclusion criteria, leaving 552 articles selected for conceptual classification.



**Figure 2.** Flowchart of the systematic review search procedure in the Scopus, Science Direct, Web of Science, and MathSciNet databases. From 1280 articles, based on the queries words, 728 articles were removed (duplicate papers, non-English written, not peer-reviewed, nor Bayesian spatial modeling), and 552 remained to be analyzed. Then, information such as the authors' names, journal titles, publication year, and the conceptual classification scheme were explored.

As a structure of the dataset, 552 articles that met the eligibility criteria were classified into the following categories:

- Names of all authors;
- Publication year;
- Journal title;
- Response to the ten items of the conceptual classification scheme on Bayesian spatial models.

This survey was divided into two parts: theoretical models and empirical analyses of the published articles. These analyses scrutinized the results from the last 20 years, as given in the next section. In the first part, we discuss the different approaches under the statistical innovation and their differences, divided into eight topics: fields of application, spatial domains, spatial priors, response variables, statistical models, prior specification, computation techniques, simulation, and validation. We present several applications that adopted spatial models using the systematic methodology presented in Section 2.

### 3. Conceptual Scheme for Spatial Models

This research focused on content analysis in a Bayesian spatial model to systematically assess the content of a large volume of recorded information in this field. It aimed to

provide an in-depth insight into the contributions and identify key research directions and opportunities in the Bayesian spatial methodology. To accomplish this, we applied a conventional approach to content analysis [30] by scrutinizing samples of the articles to clearly define the characteristics that better explain the scope and richness of the literature, identifying the key concepts and patterns. The first step was to cluster similar characteristics. We analyzed all the articles, always following the idea of subsequent updates in the clustering, that is new classes were added whenever new data that did not fit the defined characteristics were found. This approach made room for the literature to be classified without an a priori presumption.

Every Bayesian spatial analysis aims to estimate the spatial pattern over an extended geographical region to identify regions with extreme realization. In Bayesian hierarchical models, the spatial pattern is represented with a component that uses the same set of smoothing parameters across the entire study region. This type of smoothing is referred to as global smoothing. In some geographical settings, global smoothing may be inappropriate due to the complexity of the geographical settings, and the spatial pattern is likely to exhibit a localized behavior. Thus, localized regions are smoothed using different parameters, and this is referred to as local smoothing [31].

In the review process, as a research methodology, we first classified each article into one of the two disjoint classes, “theoretical” and “applied”. The theoretical methods involve investigating the fundamental principles and reasons for the occurrence of an event, random phenomenon, or process. On the contrary, applied research involves solving a particular problem with known or accepted theories and principles.

### 3.1. Spatial Statistics Fields of Application

Bayesian spatial statistics is a useful tool to determine the dependence structure and hidden patterns in space, through the prior knowledge and data likelihood. In some cases, the hypotheses of interest of a random phenomenon do not directly relate to the effect of spatial dependencies. However, it is crucial to adjust for spatial variation [32]. Adjustment for spatial patterns in modeling random occurrence has been practiced across various fields such as agriculture, medicine, biology, epidemiology, geography, geology, economics, climatology, and ecology, among others [33]. Moreover, spatial dependence in agriculture has long received consideration. Ronald A. Fisher identified spatial variations and used them to establish (random) blocks in the experiments in order to mitigate the effect of spatial dependencies in a randomized experimental design [25,34].

In many biological and medical experiments, such as gene classification and brain mapping, the randomized blocking technique may not be a viable alternative. Moreover, in demography, disease mapping, image analysis, remote sensing, manufacturing engineering, and species detection, the variation due to spatial proximity cannot be neglected. It may result in bias and inconsistent estimates. Responses at a close range tend to have a similar behavior and variation. The homogeneity of the variation depreciates with the increased distance. An efficient procedure to tackle the effect of spatial proximity is to consider random field statistical models. Random field statistical models, known as spatial models, describe the distribution of a random phenomenon over a spatial domain.

Spatial models have long been applied in various fields. In 1949, Isard described the general theory of the spatial formation of economic activities focusing on the geographic distribution of the costs, prices, and location of industries [35]. Spatial statistics applied to economics, often referred to as spatial econometrics, have gained more attention in recent years to analyze economic data over a wide range of spatial domains [36]. Similarly, in 1950, D.A. Krige took advantage of nearby variations to pursue the spatial prediction of the gold distribution in South Africa, basing the predictions practically on lognormal de Wijsian spatial models [37]. In epidemiology and public health, spatial statistics have gained increasing importance in predicting disease outbreaks [38–42]. The problems that arose in these fields motivated several intuitions to improve the existing spatial models in

the literature. Some of the most advanced models identify, for example, spatial risk factors, disease surveillance, and spatial predictive models [43].

In this research, the application fields were classified into five major groups: 1. biological and medicine: these include research in biology, medicine, epidemiology, and public health; 2. economics and humanity: these include economics, demography, criminology, accident analysis; 3. physical science and engineering; 4. agricultural and environmental science; 5. sports.

Spatial statistical models play a key role in determining the spatial pattern or quantifying the relative positions of biological components, such as DNA, involved in some biological functions. Some examples include the study of the relative positioning of primordial and growing follicles in mice to identify the likely source of some regulatory muscles [44]; determine the spatial patterns, relative position, and interaction of *Arabidopsis thaliana* heterochromatin [45]; for molecular profiling using the Bayesian hierarchical negative binomial distribution for diagnoses and treatment procedures [46]. Moreover, it is useful for disease mapping to determine the onset of an epidemic disease. A recent application includes the analysis of the mortality rate of COVID-19 in Spain and Italy using a Gaussian process to explain the spatial pattern [47,48]; the spatial mapping of schistosomiasis in Tanzania to determine the prevalence and spatial pattern [49]. It is also considered to be a powerful tool for image analysis in the medical field. An example includes the multivariate spatial model for characterizing neuroimaging data with a linear combination of multiscale basis functions to explore traits or symptoms in brain disorders [50].

Bayesian spatial statistics has been embraced in economics to identify the spatial pattern of a household's share of economic distress, to understand the formation of new business, and in studies on consumer and producer behavior. It has been used to identify the impact of economic, social, and demographic factors on the spatial variability of the household share of economic distress [51]; identify the spatial structure of the calls to the Portuguese health line, accounting for the demographics, socio-economic information, and characteristics of the health systems [52]; identify clustering in severe mobility crash risk and diagnosing of active transportation safety issues [53]. Moreover, it has been employed in the analysis of spatial patterns and hotspot detection of violent and property crimes at a small spatial scale in Toronto, Canada [54]; map the main features of fertility, such as timing, pace, and scale, and to detect spatial disparity in fertility transition in Brazil [55].

Moreover, spatial statistics has been increasingly applied in physical and environmental sciences. It has been used to provide estimates for the curvature of a railway sleeper supported on compacted ballast, through the multiple output Gaussian process to guide inference in unobserved regions [56]; to quantify the uncertainty, in spatially varying material parameters, such as polycrystalline, through a Gaussian random field [57,58]; to study material properties and spatial variability in elastostatics [59]. In environmental science, it has been used in extreme value analysis to quantify the uncertainty associated with an increased risk of flooding in Great Britain [60]; to determine the spatial pattern of the association of socioeconomic factors to Japanese encephalitis; to understand the seasonal effect and spatial variability in yield maps in farm precision in Southeastern Australia [61]; to determine the expected number of scores in a golf game [62].

Gaussian Markov random field models have a long history in image analysis [63,64]. Recently, they have been gaining attention in the machine-learning community for image processing [65,66]. For instance, Per Siden and Fredrik Lindsten established a connection between Convolutional Neural Networks (CNNs) and a Gaussian Markov random field, which the authors applied on temperature data [66]. Vemulapalli et al. [67] proposed a deep network architecture based on the Gaussian conditional random field for image denoising. Moreover, Lee et al. [68] developed an exact equivalence of infinitely wide neural networks and Gaussian processes and further linked the performance of these Gaussian processes to the theory of signal propagation in random neural networks.

### 3.2. Spatial Domains

Geographic reference data, also known as spatial data, is a collection of a stochastic process indexed by space. In other words, suppose  $Z(s)$  is a random process observed at location  $s$ , and the set  $Z(s) \equiv \{z(s), s \in \mathcal{D}\}$  is spatial data, in which  $\mathcal{D}$ , a subset of  $\mathbb{R}^d$ , is usually, but not necessarily) fixed and represents a spatial domain. According to Blangiardo [69], the spatial domains are distinguished as follows:

- Area or lattice data: This is a simple way to represent spatial data in the domain  $\mathcal{D}$ . In this type of spatial domain,  $z(s)$  is a random aggregated realization across an area  $s$  of distinct boundaries. For area data, the boundaries are irregular, such as administrative divisions, whereas for the lattice, the boundaries are a regular division of  $\mathcal{D}$ . For simplicity, it may be necessary to aggregate other types of spatial domain realizations to form area or lattice data. This process may sometimes be referred to as a discretization of  $\mathcal{D}$ ;
- Geostatistical or point-reference data:  $z(s)$  is a realization at a specific location  $s$  in a continuous spatial domain  $\mathcal{D}$ . Location  $s$  is considered to be a coordinate made up of longitudes and latitudes and sometimes includes altitudes. Location  $s$  could also be represented in Cartesian coordinates;
- Spatial point pattern: Realization  $z(s)$  represents the occurrence or nonoccurrence of an event at location  $s$ . In this case, the location itself is considered to be random. The random realization is a location indicator of the presence or absence of a phenomenon of interest in the domain  $\mathcal{D}$ . In agriculture, for example, the interest may be the distribution of a specific tree species, in which each realization is the presence or absence of the tree species in domain  $\mathcal{D}$ . In epidemiology, the realization may be the house address of a patient that has a particular disease [70,71].

For instance, Reference [60] adopted a Bayesian spatial model on lattice data to identify patterns for the risk of flooding in Britain; Reference [72] adopted the SPDE model on geostatistical data to predict the spatial occurrence of fish species; Reference [73] proposed a Bayesian technique to estimate the spatial point pattern of American sweetgum trees and Swedish pines.

### 3.3. Spatial Priors

A prior distribution needs to be elicited or vaguely specified for the posterior distribution's complete estimation in an empirical or full Bayesian approach. For example, in hierarchical models, the prior distribution assumed for a random field (spatial component)  $\gamma$  is termed a spatial model. We encountered several types of spatial models (priors) in the literature, and most were a subclass of the Gaussian Markov Random Field (GMRF) defined as a Gaussian random field with the Markov property [71,74]. The choice of spatial model is strongly informed by the study objective and available information. For example, for area spatial data, the CAR family is a good choice to account for cluster heterogeneity, whereas the SPDE is appropriate for geostatistical data for identifying hotspots.

In the literature, due to the large class of priors, we collapsed the encountered priors based on the most frequently used and seldom-used classes, which are the *Conditional Autoregressive (CAR)*, *Besag–York–Mollie (BYM)*, *Leroux CAR*, *Stochastic Partial Differential Equation (SPDE)*, *Gaussian Markov Random Fields (GMRF)*, *non-GMRF*, and *others*. See Table 1 for a summary. The details of each model are presented in Appendix A.2. The Lorex CAR class comprises priors with similar specifications, such as Dean's and Simpson's CAR models. The GMRF class consists of GMRF priors except for the CAR family and the SPDE stated earlier. The hyperparameters of these stochastic processes are either elicited, drawn from the previous study, or assigned as a weakly informative or objective prior distribution. For examples of elicitation technique, see [75,76], and for the objective prior technique, see [77]. Moreover, Simpson et al. [78] developed a framework to construct informative priors called the Penalized Complexity (PC) prior. The PC prior has a single parameter that is subjectively set to control the amount of flexibility allowed in the model. The construction of the PC prior for the BYM spatial model and Student-t process was

described in [78]. The PC prior for the Gaussian random fields model is implemented in R-INLA [79].

**Table 1.** Summary of the spatial models and their variations.

Spatial Model	Article	Spatial Smoothing		Gaussian Process		Non-Gaussian Process		
		Global	Local	GMRF	Non-GMRF	Parametric	Semiparametric	Nonparametric
CAR dissimilarity	Lee and Mitchell, 2012 [31]		✓	✓		✓		
Intrinsic CAR/BYM	Besag et al., 1991 [80]	✓		✓		✓		
Proper CAR	Besag, 1974 [64]	✓		✓		✓		
Leroux	Leroux et al., 2000 [81]	✓		✓		✓		
Geostatistical	Clements et al., 2006 [49]	✓		✓		✓		
Globalspline	Lee and Durbán (2009) [82]	✓			✓			✓
Simpson CAR	Simpson et al. [78]	✓		✓		✓		
Dean's CAR	Dean et al. [83]	✓		✓		✓		
SPDE	Lindgren, Rue and Lindström, 2011 [17]	✓		✓		✓		
Mixture Model	Green and Richardson [84]	✓			✓			✓
Spatial Partition Model	Leonhard and Raßer [18]		✓		✓			✓
Asymmetric Laplace	Kuzobowski and Pogorski [85]	✓			✓	✓		
Student-t	Fonseca [86]	✓			✓	✓		
Log-Gamma	Bradley et al. [87]	✓			✓	✓		
Dirichlet	Gelfand et al., 2005 [88]	✓			✓			✓

The non-GMRF is a large class that consists of nontrivial prior models that use a spatial correlation function to determine the covariance matrix of the spatial process in a continuous space, including the asymmetric Laplace, log-Gamma, skewed normal, Student-t process, and Dirichlet process. We created a class called *Others* to accommodate unspecified models and those that do not belong to the aforementioned classes. Each spatial model is discussed in Appendix A.2. For instance, References [89,90] adopted ICAR and the BYM model to map the spatial pattern of tuberculosis in South Africa and malnutrition in Nigeria; Reference [91] adopted a Bayesian hierarchical spatial quantile regression model with an asymmetric Laplace spatial component to determine the risk factors of the radon-222 noble gas, which arises naturally from uranium decays; Reference [72] adopted the SPDE model to predict the spatial occurrence of fish species.

Checking for prior data conflict in Bayesian spatial statistics is an important technique to measure the adequacy of the adopted prior information and the data likelihood. The check designs some statistic and compares its observed values to the reference distribution derived from the prior predictive density of the data [92]. Prior conflict occurs whenever the prior assigns most of its mass to regions of the parameter space that lie in the tail of the likelihood. In this case, the source of prior knowledge conflicts with the observed information. Nott et al. [92] proposed a technique that deals with prior expansion for prior data conflict checking. Egidi et al. [93] proposed an automatic elicited prior with a mixture of informative and uninformative components, which favors uninformative components whenever prior conflict occurs.

A response variable is a quantity used to describe a random process to relate it to a deterministic process mathematically. In statistical modeling, the most frequently used response variables are the discrete (categorical), ordinal, and continuous variables. The types of variables used in modeling a random phenomenon are intuitive from the process under study. The statistical models used to describe a random phenomenon vary depending on the quantity and parameters of interest.

The Bernoulli distribution is often used for modeling the random phenomenon of two possible outcomes. The Binomial, negative binomial, hypergeometric, and Poisson distributions are frequently used for modeling count cases such as disease occurrence, wildlife, signals, and more [46,94]. The Poisson distribution has been used to approximate binomial distributions for a large sample size [95,96]. The equality of the mean to variance restriction imposed by the Poisson distribution considers the negative binomial a better choice to model a random variable that exhibits overdispersion [97]. The multinomial distribution is often used to model a phenomenon of more than two categories usually encountered in biological experiments [98]. It is a generalization of the binomial distribution.

In the continuous case, a large class of distributions of the exponential family is used, such as Gaussian, exponential, Student-t, Weibull, Gamma, and more. However, according to the central limit theorem, the Gaussian distribution is used to approximate both discrete and continuous distributions for large sample sizes [99].

An analyst's interest is to quantify the association of a random phenomenon and explanatory processes, describing a random phenomenon according to a set of explanatory variables. In the literature, the statistical models encountered are the generalized linear mixed model, the hierarchical model, the survival model, and spatial econometrics models. The details of each model are presented in Appendix A. Additionally, we created the proposed, *unspecified* and *others* classes to accommodate the proposed and validated models, as well as unspecified models. The class of *others* accommodates statistical models outside the above-listed classes. For instance, Reference [91] adopted a generalized hierarchical mixed model to determine the risk factors of the radon-222 noble gas; Reference [90] used a generalized hierarchical mixed model to determine the impact of carbon (IV) oxide on the prevalence of malnutrition; Reference [100] adopted a survival statistical model to map the prevalence of hospitalization due to Dengue in Wahidin Hospital in Makassar, Indonesia; Reference [52] adopted the spatial econometrics model (lag-model) to estimate the global spatial correlation of the calls to the Portuguese national health line.

An appropriate prior distribution specification in a Bayesian inference continues to be a challenge in various fields of application. A prior distribution is associated with the representation of the uncertainty of the parameters of interest before data are observed. The elicitation of an appropriate prior distribution is a nontrivial task [101], and such challenges are accumulated in spatial models due to the large number of associated parameters involved. In our review, we came across four main approaches. One way to set a prior distribution is to assume ignorance about the appropriate model, that is  $\mathbb{P}(\theta \propto 1)$ , and allow the data model to carry all the information. Such an approach is not always advantageous because inference on the parameters can be improved by performing prior elicitation based on identified characteristics or expert opinion. The elicitation procedure is termed the elicited prior, which leads to another method, known as prior elicitation. In elicited priors, convenient prior distributions are sometimes a choice and have spread across the literature and been set as default priors in most simulation packages. As a result, subsequent authors used such prior distributions verbatim. However, several authors did not explicitly state the type of prior used and were classified as not available. For instance, Reference [8] elicited prior information from expert opinion, from a meta-analysis of neuroimaging data and a parapsychologist to perform a Bayesian spatial point process to provide an interpretable model for brain imagining studies, and Reference [102] elicited prior knowledge from experience for predicting a particular matter.

### 3.4. Computational Techniques

In Bayesian inference, the prior information expressed through the prior distribution  $\mathbb{P}(\theta)$  and the data likelihood  $\mathbb{P}(y|\theta)$  is used for inferences. In simple or convenient cases, the given posterior distribution is represented by Equation (5). However, in practice, assessing the posterior distribution to make inferences is not a trivial task, because it usually contains compound integrands with complicated and analytically nonintegrable support [103]. Thus, authors have explored different approaches to make inferences. In the literature, we encountered several computation techniques and classified them into the *Markov Chain Monte Carlo (MCMC)*, *Integrated Nested Laplace Approximation (INLA)*, *Expectation–Maximization (EM)*, and *Maximum (Penalized quasi-) likelihood method classes*. Details of these techniques are presented in Appendix A. The MCMC class comprises all numerical approximation that uses the Monte Carlo method. Moreover, the *unspecified* class was added to accommodate articles that neither discussed nor stated the approach used in the estimation procedure. The *others* class comprises computation techniques that do not fit into the defined classes. For instance, Reference [90] utilized the INLA interface for estimating the parameters of a Gaussian latent field model; Reference [91] utilized the MCMC approach to estimate quantile regression parameters; Reference [104] utilized the EM algorithm to perform a Bayesian interpolation of nitrogen dioxide, ozone, sulfur dioxide, and surface iron in public health units in Ontario; Reference [105] compared the performance of maximum likelihood estimation with different computation techniques for spatial scan statistics.

### 3.5. Simulation Study and Validation

A simulation study is a systematic and scientific computer procedure that involves fixing model parameters to generate data by pseudo-random sampling [106]. It comprises two main steps: data generation and estimation. In the first step, a set of parameters is fixed and used to generate pseudo-random data. In the second step, the generated data are fed back to the model to estimate the “unknown” parameters and check for bias and model error.

A simulation study is usually carried out for proposed models and methods. The papers reviewed were classified into two: “yes”, if the paper contained statistical simulation studies, and “no” if it did not.

In addition to the simulation studies, we also investigated how Bayesian spatial models were validated using real data. This is a procedure to check overfitting or underfitting. A model overfits if it performs well in the training set and badly in the test set, whereas it underfits if it performs poorly in the training set. A classical approach to cross-validation is to form a disjoint subset of the whole data into training and testing sets. The model is fit into the former and tested on the later set. Doing this process  $k$  times until all observations in the dataset participate in training and testing once is called  $k$ -fold cross-validation. The whole data are split into  $k$  disjoint subsets, in which the combined  $k - 1$  sets serve as the training set and the remaining set of size  $n_k = n/k$  serves as the test set, where  $n$  is the data size. A particular case of the  $k$ -fold cross-validation is the leave-one-out cross-validation, in which one observation serves as the test set and the remainder  $n - 1$  serves as the training set. After going through all subsets, the validation measures are statistically combined to make a valid conclusion.

Since the spatial models are frequently modeled in a Bayesian framework, we included the *posterior predictive check* [107] class. In a predictive posterior check, a statistical test is chosen and computed for the observed data process. The same statistic is computed for replicated posterior predictions of the process. The model is said to present a good fit if the posterior prediction average is close to the statistic test for the observed data [103]. We added the *none or not applicable* class to accommodate papers that did not conduct cross-validation and *others* to accommodate validation methods not mentioned above. For instance, Reference [104] adopted leave-one-spatial-location-out to validate the performance of the proposed model for multivariate interpolation of air pollutant gases in

the health unit of Ontario; Reference [108] utilized the k-fold cross-validation method to examine the robustness of the adopted model to determine the geographical inequalities and nutritional status of women and children in Afghanistan; Reference [8] assessed a Bayesian point process model's performance to provide an interpretable model for brain imaging studies.

#### 4. Analyses

As described in the search procedure section, a total of 552 articles were selected after applying the exclusion criteria (duplication and context). After careful analysis, the papers were categorized into the first class, labeled as being only an applied paper, theoretical, or both, in which 4 (1%) of the papers showed no application (only theoretical with synthetic data), 188 (34%) showed an improvement in the field with real-world application, and 360 (65%) only applied the existing methodologies.

The papers were subdivided into five classes of application fields: *agricultural and environmental science*, *economics and humanities*, *medical science*, *physical science and engineering*, and *other*. Three fields held the majority of the publications, which were *agricultural and environmental science* (30.1%), *economics and humanities* (30.6%), and *medical science*, which includes epidemiology, (33.7%). Moreover, the spatial domain used was also taken into account: *area/lattice*, *geostatistical*, and *spatial point patterns*.

The *area/lattice* occurred in 65.6% and *geostatistical* in 31.2% of the reviewed papers, and in combination, they held 95.8% of the publications. It is important to note that more than one spatial domain could be used in an article, such as the 1% observed in this review. This procedure is common when a continuous spatial domain is discretized in order to lower the computational burden.

The core part of this review is the spatial priors (models) used in the literature. Whereas the literature contains numerous spatial priors applied in various problems across various research fields, this systematic review only presents models encountered during the review. Gelfand et al. [109] pointed out the first incorporation of a spatial prior and the gain in the structure of this modeling.

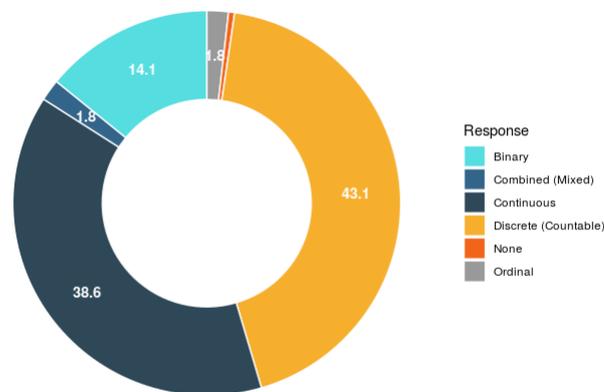
Table 1 summarizes the main spatial models found, which are classified into groups. Additionally, Table 2 presents the frequency distribution of the spatial models adopted in the literature versus the statistical modeling, showing that the statistical model most often adopted is the Generalized Linear Mixed Model (GLMM) and, as for the spatial priors, the CAR variation.

**Table 2.** Crosstab spatial priors used versus the statistical model adopted. The GLMM with a CAR spatial prior family for the spatial component is the most frequently used modeling structure in the literature, though some alternatives have been growing in the past decade such as the GLMM framework combined with non-GMRF, GLMM with SPDE, and spatial autoregressive model define dependence matrices.

	GLMM	Nonparametric	Spatial Autoregressive	Proposed	Survival Models	Not Stated	Other	Total
Conditional Autoregressive models (CARs)	227	1	1	0	9	4	3	245
Non-Gaussian Markov Random Field (non-GMRF)	101	0	49	5	1	1	16	173
Gaussian Markov Random Field (GMRF)	19	0	0	0	0	0	3	22
Stochastic Partial Differential Equations (SPDE)	20	0	0	0	0	0	0	20
Nonparametric	4	0	0	0	0	1	2	7
Not Stated	30	0	2	0	0	9	4	45
New Methodology	17	0	0	0	0	0	23	40
Total	418	1	52	5	10	15	51	552

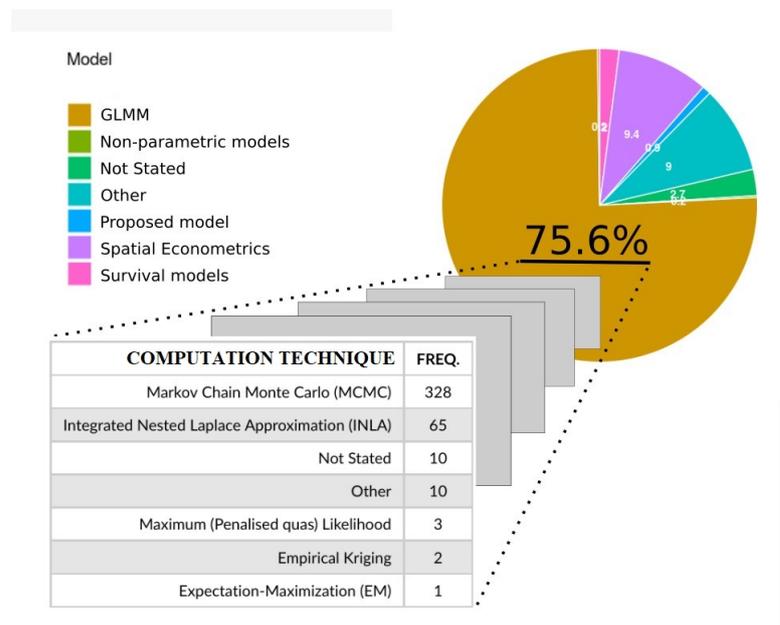
Among these spatial models are the *CAR family*, generally used for spatial area/network domains, *Stochastic Partial Differential Equations (SPDEs)*, *Gaussian Markov Random Field (GMRF)*, except for the *CAR family*, models with an author-specific defined covariance structure, which are often used for a continuous spatial domain, and finally, nonparametric methods. The *CAR family* appeared in 44.2% of the published articles reviewed. From this percentage, the *CAR* and the *BYM* model appeared in 96.3%. The *SPDE* appeared in about 3.9%, and the nonparametric procedure appeared in 1.3% of the articles reviewed. The *GMRF*, exempting the *CAR family* and the *SPDE*, appeared in the literature in 4.0% of them. Consequently, several spatial model applications are specified through the type of covariance structure adopted, based on prior knowledge of the interactions among the phenomena of interest, which appeared in 31.3% of the reviewed articles. Only a single documentation of the application of spatial models on robotic technology was encountered [110]. Other models that did not fit into any of these groups appeared in 7.2%, and 9.3% of the articles did not describe or state the type of model adopted.

The observation or response variable's nature dictates the statistical model class to be adopted to make inferences. In our search, as shown in Figure 3, discrete (countable) was the most used, followed by the continuous and binary response variable types.



**Figure 3.** Distribution of the response variable type. Most published studies presented discrete (countable) response variables. The discrete response variable is frequently used in disease prevalence, wildlife population studies, accident analysis, crime analysis, etc.

In addition, the most widely adopted statistical model for spatial analysis was *GLMM*, considering the Bayesian paradigm class. Due to the integrated complexity of the posterior marginal distribution, the *MCMC* estimation method is the most frequently adopted numerical integration, as depicted in Figure 4.



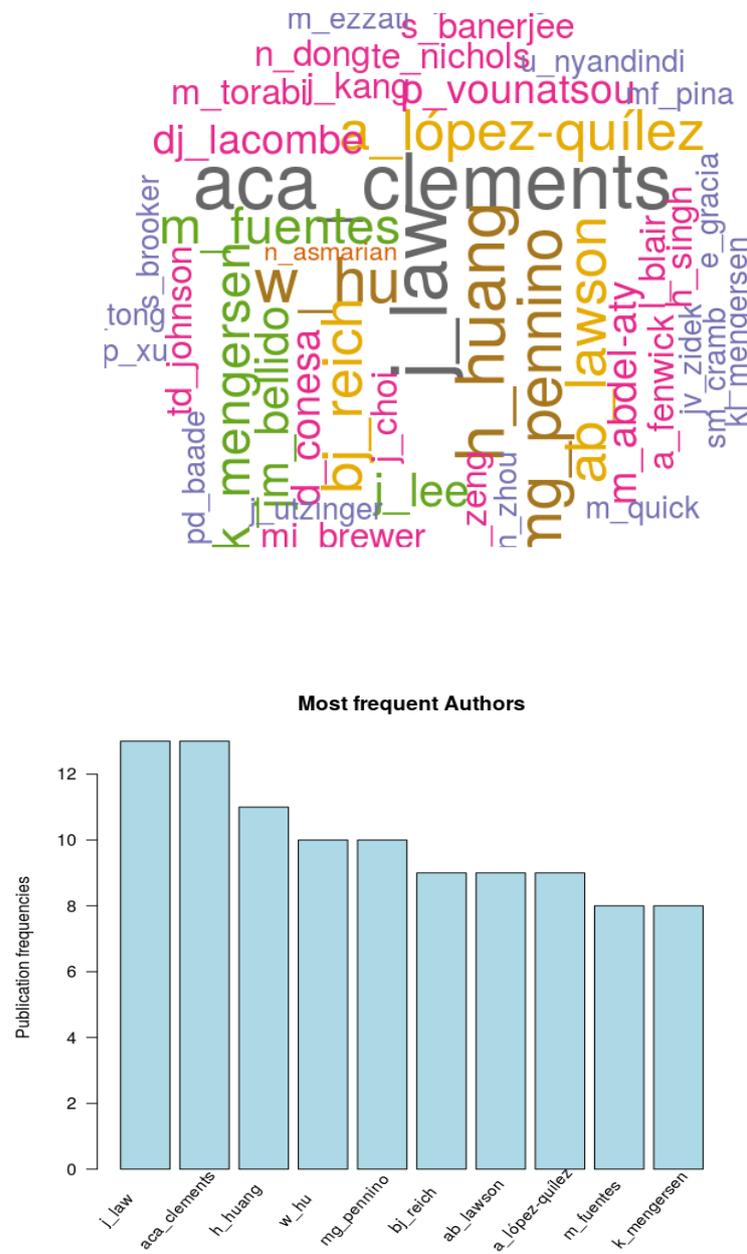
**Figure 4.** Class of models often using spatial modeling and its numerical estimation method distribution. Given the class of GLMM/hierarchical models, Markov Chain Monte Carlo (MCMC) is the most used intensive computation technique.

Maybe the most critical question regarding spatial models is related to specifying the spatial prior. Elucidating events in an area, sometimes even under a few frequencies, is desirable through direct probabilistic statements that may unravel hidden patterns [9]. The spatial interdependence can be analyzed through spatial correlation, and using the adoption between spatial fields and the Bayesian approach allows the knowledge of the information to be allocated in the modeling along with the information contained in the acquired data. Thus, the results obtained in this systematic review showed that the expert’s knowledge was used in conjunction with the data information (30.43%), as shown in Table 3, although this can be better explored.

**Table 3.** Model prior specified.

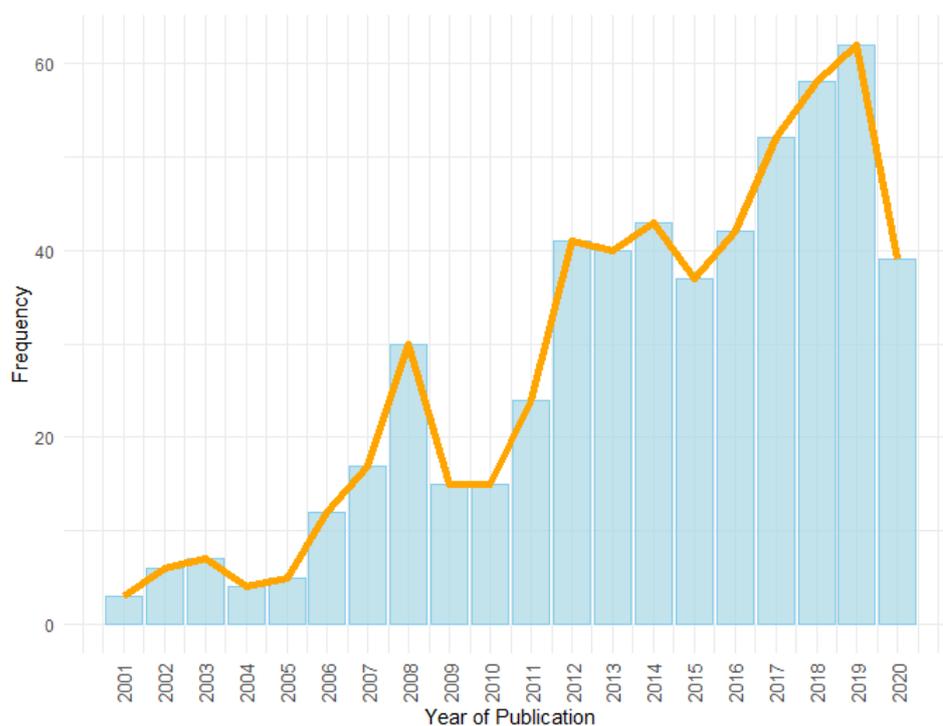
Prior Specified	Freq.
Elicited from experts or from the problem	168
No explicit use or reference/not applicable	101
Used verbatim from the literature	166
Vague prior (noninformative)	119

The authors who were the most recurrent in the literature over these past 20 years were James Law, A.C.A. Clements, and Hei Huang. Figure 5 shows the relevant authors in the Bayesian spatial models’ publication field.



**Figure 5.** Most frequent authors on the Bayesian spatial models. The top graph is a tag cloud for the 50 most frequent authors over the past 20 years. The bottom graph is a bar plot displaying the Top 10 authors and their relative frequencies. The most frequently appearing authors are Jane Law and Archie C. A. Clements. These authors are followed, in order, by Wenbiao Hu, M. Grazia Pennino, Brian J. Reich, Andrew B. Lawson, Antonio López-Quílez, Montserrat Fuentes, and Kerrie Mengersen.

The top five journals that contained the most papers related to the subject under study (combining the theoretical methodology with publications of real-world applications) were Spatial and Spatio-temporal Epidemiology (# 15), Accident Analysis and Prevention (# 14), PLoS ONE (# 14), Spatial Statistics (# 11), and Environmentrics (# 10), whereas, across time, the spatial modeling publication rate using the Bayesian approach proliferated, as shown in Figure 6. The year 2020 covers only the first half of the year.



**Figure 6.** Growth in scientific publications related to topics in Bayesian spatial models from 2001 to June 2020. There was a positive growth over the 20 years considered. The growth could be associated with the improvement in computational tools and data collection.

## 5. Concluding Remarks

Spatial statistics has gained tremendous attention in recent years due to the efficiency in collecting spatial dependence data. Neglecting such dependencies may result in bias and, consequently, lead to the wrong inference. The Bayesian approach for analyzing spatial data often outperforms the frequentist approach, given that the prior information is taken into account. In a Bayesian framework, spatial priors play a significant role in accounting for space dependencies. With the consistency in the improvement of data collection and computational tools in analyzing spatial data, Bayesian spatial statistics will further penetrate numerous fields and become one of the leading tools for analyzing data.

Many authors account for spatial dependencies assuming a Gaussian random field. In many real data applications, the Gaussian random field may be inappropriate, especially in extreme data, skewed data, and data with spikes and heavy tails. Examining a different random field such as the Laplace, Student-t, Pareto, Nakagami-m, and more, may improve inference. A significant issue of assuming these distributions is the nontrivial method of eliciting or fixing a prior distribution for the model parameters. For instance, the prior distribution for the degrees of freedom of the Student-t distribution is not a trivial task. However, the objective prior and the PC prior were developed in [77,78], respectively. Regardless of the prior distribution, eliciting priors for the parameters is critical, and when wrongly assumed, this could lead to misleading results and inference. In order to circumvent them, which is also not a trivial task, it is essential to consider objective priors for the random field parameters and hyperparameters to improve inference.

The Bayesian spatial literature lacks sufficient information on the objective priors, such as Jeffery's prior, reference priors, matching priors, and more. These priors stand out to elicit experts' ideas that could improve the inferences of the spatial dependencies. Deriving an objective prior distribution for a spatial random field's smoothing parameter is currently an unaddressed problem that needs urgent attention.

The Markov property is a useful tool to lower the computational cost in Bayesian inferential statistics by subjecting to the immediate neighbors' spatial dependencies. However,

the occurrence of some random phenomenon exhibits strong spatial dependencies beyond the immediate neighbors, and truncating such a dependency structure would result in bias and incorrect inferences. Moreover, there is an insufficient standard approach to determine the covariance matrix structure of the spatial effects. Thus, spatial smoothing is not trivial to compare across different models.

In neuroscience, the application of Bayesian spatial statistics to brain experiments has been gaining interest [7,8,111,112]. The complexity of the brain structure has prevented the application of classical spatial models. In other words, the primary assumption of spatial contiguity in the analysis may result in incorrect inference due to the brain's complexity. Moreover, a response received at one location on the skull may be due to brain activity in the opposite location. Beyond complexity, the dynamics of the body system of the subject influences the experiments. Thus, accounting for such complex structures is a problem that requires further study.

Despite the increasing availability of Big Data in the Internet of Things, the spatial model has not received adequate attention as a classification algorithm and regression-machine-learning model. Applications of machine-learning models to spatial data may reveal hidden patterns and be applied to navigation problems efficiently by robotic technology through information based on georeferencing [110], thus creating new lines of research.

**Author Contributions:** Conceptualization, F.L., D.C.d.N. and O.A.E.; survey question formulation and screening F.L. and D.C.d.N.; data acquisition, O.A.E. and D.C.d.N.; article review, O.A.E. and D.C.d.N.; article writing, D.C.d.N. and O.A.E., reviewed by F.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** Francisco Louzada acknowledges support from the Brazilian institutions FAPESP and CNPq. Diego Nascimento acknowledges the support from the São Paulo State Research Foundation (FAPESP Process 2020/09174-5). Osafu Augustine Egbon acknowledges support from the Brazilian institution CAPES.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare that there is no conflict of interest.

## Appendix A

### Appendix A.1

The coding of the characteristics was framed according to the conceptual classification scheme developed by Hachicha and Ghorbel [113] and applied by Tiago et al. [103]. In this framework, bias is limited to research data and clarifies reporting results and findings to draw concrete conclusions. Such classification is useful for researchers as it provides an overview of the research methodology's application and can reveal research gaps in the literature.

This review used a conceptual classification scheme of 10 items presented in Table A1. For lucidity, each item in the classification scheme is discussed in details.

**Table A1.** List of Questions for Conceptual Classification Scheme.

QUESTION 1	Is it only an application?
1.1	Yes
1.2	Both
1.3	No (only method)
QUESTION 2	What is the field of application?
2.1	Medical science
2.2	Economics and humanities
2.3	Physical science and engineering
2.4	Agricultural and environmental science
2.5	Sports
QUESTION 3	What spatial domain was employed?
3.1	Area or lattice
3.2	Geostatistical data
3.3	Spatial point patterns
3.4	Area and geostatistical data
QUESTION 4	What type of spatial priors are used?
4.1	Conditional Autoregressive (CAR)
4.2	Besag–York–Mollie (BYM)
4.3	Leroux CAR
4.4	Gaussian Markov random field (other specifications)
4.5	Covariance function (Not GMRF)
4.6	Other (new methodology/proposed)
QUESTION 5	What type of response variable is used?
5.1	Discrete (countable)
5.2	Continuous
5.3	Combined (mixed)
5.4	Ordinal
QUESTION 6	What is the statistical model used?
6.1	Generalize linear (mixed) model (or hierarchical models)
6.2	Survival and longitudinal models
6.3	Nonparametric models (machine-learning models)
6.4	Spatial econometrics
6.5	Proposed
6.6	Not stated
6.7	Other
QUESTION 7	How are model prior specified?
7.1	Vague prior (noninformative)
7.2	Used verbatim from the literature
7.3	Elicited from experts or from the problem
7.4	No explicit use or reference/not applicable
QUESTION 8	What is the estimation method applied?
8.1	Markov Chain Monte Carlo (MCMC)
8.2	Integrated Nested Laplace Approximation (INLA)
8.3	Expectation–Maximization (EM)
8.4	Maximum (penalized quasi-) likelihood method
8.5	Not stated
8.6	Other
QUESTION 9	Is the model validated through simulation?
9.1	Yes
9.2	No
QUESTION 10	Is the application validated through data-driven procedures?
10.1	Cross-validation and data splitting (K-fold/holdout)
10.2	Leave-One-Out Cross-Validation (LOOCV)
10.3	Posterior predictive check
10.4	Other
10.5	None or not applicable

In the following, we will present an overview of the primary spatial statistical models explored in this systematic review.

Appendix A.2. Class of the Gaussian Markov Random Field

In standard form, the covariance matrix of a Gaussian random field is positive definite and often dense. This restriction makes it difficult to construct an appropriate covariance matrix. Furthermore, the computational cost is high due to the cost of  $O(n^3)$  to factorize the covariance matrix [17]. In order to overcome the aforementioned challenges, a sparse covariance matrix of less computational cost needs to be constructed and closely approximate the dense covariance matrix. A possible choice is to adopt a Markovian property on the covariance matrix to have a GMRF approximation. It is a discretely indexed finite random vector that has a joint Gaussian distribution with a sparse precision matrix. Following the definition given by Rue [114], the set  $Z(s) \in \mathbb{R}^d$  over a spatial domain  $s \in D$  is said to be a GMRF with respect to graph  $\mathcal{G}(V, \mathcal{E})$ , centered at  $\mu$  and precision matrix  $\mathbf{Q} > \mathbf{0}$  and density of the form:

$$\pi(\mathbf{Z}|\mu, \mathbf{Q}) = (2\pi)^{-n/2} |\mathbf{Q}|^{1/2} \exp\left(-\frac{1}{2} (z - \mu)^T \mathbf{Q} (z - \mu)\right), \tag{A1}$$

in which  $Q_{i,j} \neq 0$  if  $i$  and  $j$  are considered neighbors,  $V$  is a set of vertices, and  $\xi$  are the edges of graph  $G$  representing spatial  $D$ . Usually, the covariance function is constructed in a way that it is a function of the relative positions of location  $i$  and  $j$  in  $D$ .  $\mathbf{Q}$  is considered stationary when it is only a function of the distance between the locations in  $D$ . Let  $\delta_i = \{z : z \text{ is a neighbor of } i\}$ , the full conditionals be given as:

$$\pi(z_i|\mathbf{Z}_{-i}) = \pi(z_i|\delta_i), \tag{A2}$$

for every  $z_i \in \mathbf{Z}$ , and  $\mathbf{Z}_{-i}$  is the vector  $\mathbf{Z}$  excluding  $z_i$ . Some of the subclasses of GMRF identified in the literature for spatial smoothing are the proper and Intrinsic Conditional Autoregressive (ICAR), Besag–York–Mollie (BYM and BYM2), Leroux CAR, and Dean’s CAR.

**–Conditional Autoregressive (CAR)**

Let the spatial domain  $D$  be partitioned into  $n$  disjointed areas  $E$  such that  $D = \cup_{i=1}^n E_i$ .  $E$  could be regular or not, as in Lattice or area respectively. Let  $z_i$  be a random variable observed at  $E_i$  and the vector  $\mathbf{Z} = (z_1, \dots, z_n)^T$  with  $\mu = (\mu_1, \dots, \mu_n)$ . Let  $\mathbf{Z}_{-i}$  be a  $(n - 1)$  dimensional vector in a way that  $\mathbf{Z}_{-i} = (z_1, \dots, z_{i-1}, z_{i+1}, \dots, z_n)^T$ , the full conditionals of a proper CAR model are given as:

$$\pi(z_i|\delta_i, \sigma^2, \mu_i) \sim N\left(\mu_i + \sum_{\{j:z_j \in \delta_i\}} c_{ij}(z_j - \mu_j), \kappa_i^{-2}\right), \tag{A3}$$

in which  $\delta_i = \{z : z \in E_j \text{ is a neighbor of } E_i, j = 1, 2, \dots, n\}$ ,  $\kappa_i^{-2} > 0 \forall i = 1, \dots, n$  and  $c_{ij}$  is a function of the adjacency between  $E_i$  and  $E_j$ . To make  $\pi(\circ)$  a proper distribution,  $\mathbf{C} = (c_{ij})$  and  $\kappa$  are carefully chosen. A frequent choice for  $\mathbf{C}$  is a function of contiguity between the areas. Let  $\mathbf{A} = (a_{ij})$  be an  $n \times n$  matrix, in a way that  $a_{ii} = 0$ ,  $a_{ij} = 1$  if  $z_j \in \delta_i$  for  $i \neq j$ , and 0 otherwise. Notice that  $\mathbf{A}$  is symmetric, since if  $E_j$  is a neighbor of  $E_i$ , then  $E_i$  is a neighbor of  $E_j$ . We define  $c_{ij} = \rho \frac{a_{ij}}{d_i}$ ,  $\kappa^2 = \frac{d_i}{\sigma^2}$  and  $d_i = \sum_j a_{ij}$  [64,114,115]. The full conditionals and the joint distribution of (A3) are given by:

$$\begin{aligned} \pi(z_i|\delta_i, \sigma^2, \mu_i) &\sim N\left(\mu_i + \rho \sum_{\{j:z_j \in \delta_i\}} \frac{a_{ij}(z_j - \mu_j)}{d_i}, \frac{\sigma^2}{d_i}\right) \\ \mathbf{Z} &\sim N_n\left(\mu, \sigma^2(\mathbf{I} - \mathbf{C})^{-1}\mathbf{M}\right), \end{aligned} \tag{A4}$$

in which  $\mathbb{M} = \text{diag}(\frac{1}{d_1}, \dots, \frac{1}{d_n})$ . and  $|\rho| < 1$  is chosen in a way that  $(\mathbb{I} - \mathbf{C})^{-1} > 0$ . Besag, York, and Mollie [80] proposed the intrinsic CAR (ICAR), by setting  $\rho = 1$ , then  $c_{ij} = \frac{a_{ij}}{d_i}$ . The full conditionals of the ICAR:

$$\pi(z_i | \delta_i, \sigma^2, \mu_i) \sim N\left(\mu_i + \frac{\sum_{\{j:z_j \in \delta_i\}} a_{ij}(z_j - \mu_j)}{d_i}, \frac{\sigma^2}{d_i}\right), \tag{A5}$$

The model inferred that the conditional expectation of  $z_i$  equals the weighted deviations of its neighbors in addition to its mean.

**–Besag–York–Mollie (BYM)**

The BYM, proposed by Besag [80], is a variant of the CAR model that incorporates an additional term to control for overdispersion in spatial data. Suppose  $z$  is partitioned in a way that  $z_i = u_i + v_i$ , the unstructured term  $v$  is modeled as:

$$v_i \sim N(\mathbf{0}, \sigma_v^2), \tag{A6}$$

and the structured component is modeled as  $u_i \sim \text{ICAR}$ . Thus,

$$\text{Var}(z | \sigma_u, \sigma_v) = \sigma_v^2 \mathbb{I} + \sigma_u^2 (\mathbb{I} - \mathbf{C})^{-1} \mathbb{M}. \tag{A7}$$

The BYM poses identifiable problem in a way that each observation is represented by  $u_i$  and  $v_i$ , thus, only the sum  $u_i + v_i$  is identifiable [116]. Setting appropriate hyperparameters is challenging. However, constraining  $z_i$  to sum to zero allows the confounding problem to be avoided and both components to be fitted.

**–Dean’s Conditional Autoregressive**

A reparameterized version of the BYM proposed by Dean et al. [83] and its covariance matrix are given as:

$$\begin{aligned} \mathbf{z} &= \sigma(\sqrt{\phi} \mathbf{u} + \sqrt{1 - \phi} \mathbf{v}), \\ \text{Var}(z | \sigma_u, \sigma_v) &= \sigma^2((1 - \phi) \mathbb{I} + \phi (\mathbb{I} - \mathbf{C})^{-1} \mathbb{M}), \end{aligned} \tag{A8}$$

in which  $\sigma_u = \sigma^2 \phi$  and  $\sigma_v = \sigma^2(1 - \phi)$ .

**–Simpson Conditional Autoregressive**

For the purpose of scaling and interpretability of the hyper prior, Simpson et al. [78] proposed a modification of the BYM which avoids the problems posed by BYM model. The combined random effect and the covariance matrix are given by,

$$\begin{aligned} \mathbf{z} &= \sigma(\sqrt{\phi} \mathbf{u}_* + \sqrt{1 - \phi} \mathbf{v}), \\ \text{Var}(z | \sigma) &= \sigma^2((1 - \phi) \mathbb{I} + \phi \mathbf{Q}), \end{aligned} \tag{A9}$$

in which  $\mathbf{u}$  and  $\mathbf{v}$  are as previously defined,  $\sigma \geq 0$ ,  $\phi \in [0, 1]$  and  $\mathbf{Q}$  is the ICAR covariance matrix such that  $\text{Var}(u_i) \approx 1$ .

**–Leroux Conditional Autoregressive (Leroux CAR)**

Leroux et al. [81] proposed a variant of the CAR model which, unlike the BYM, only requires a single set of spatial component. The full conditionals is given by:

$$\pi(z_i | \delta_i, \sigma^2, \mu_i) \sim N\left(\mu_i + \frac{\rho \sum_{\{j:z_j \in \delta_i\}} a_{ij}(z_j - \mu_j) + (1 - \rho) \mu_0}{\rho d_i + 1 - \rho}, \frac{\sigma^2}{\rho d_i + 1 - \rho}\right), \tag{A10}$$

which is a limiting case of the ICAR model. The model avoids the identifiability problem in the BYM model by the specification of a single hyper parameter for random effect [117]. The specification (A10) smooths the neighboring random effect weighted by  $\rho$  and the global mean weighted by  $1 - \rho$ .

### –Conditional Autoregressive dissimilarity

The variations of the CAR model earlier discussed exhibit a global degree of spatial smoothing. In many instances, a global smoothing may be inappropriate, especially in areas that exhibit locally constrained spatial structure. Lee and Mitchell [31] proposed a CAR dissimilarity to smooth area elevated risks which depend on local spatial parameters. It is one out of many proposed local spatial smoothings, such as locally adaptive model [118], localized conditional autoregressive [119], to number a few.

#### Appendix A.3. Class of Gaussian Non-Markov Random Field Models

Markov property is known to lighten the burden of factorizing a dense covariance matrix. However, a random realization with a strong correlation structure outside its neighborhoods can be poorly accounted for. Moreover, the discretization imposed by the GMRF sometimes does not account for the presence of discontinuities in spatial domains. These could be addressed by relaxing the Markov assumption and possibly allowing a different distribution other than Gaussian.

### –Spatial Weight Matrix

In spatial Econometrics, a weight matrix and a correlation parameter are often used to specify a dependency structure among economic variables of interest observed at different spatial locations in a Spatial Lag Model (SLM) and Spatial Error Model (SER) [20]. The weight matrix is essential in the covariance matrix specification for parametric models, such as the class of GMRF prior. The spatial weight configuration specification varies with the relationships among geographical locations, such as spatial distance, interactions, nearest neighbors, and contiguity. The specification is categorized into four: adjacency-based weights, weight-based on geographical distance, the distance between covariate values, and hybrid of geographical distance and covariates [23]. In most cases, nearby neighbors to a spatial reference location receive higher weights compared to further neighbors. Although the diagonal elements of the spatial weight matrix are universally accepted to be set to zero [120], in the literature, there is no standard to define the weight matrix the specification is dominated by choice of computational convenience [20].

### –Spatial Covariance Function

In this section, we briefly discussed the frequently used stationary and isotopic correlation functions to construct a valid correlation matrix for a random field in Geostatistical models. A covariance function  $C(s_i, s_j)$  is said to be stationary when  $C(s_i, s_j) = C(s_{i+l}, s_{j+l})$  for any lag  $l$ , and isotopic when it only depends on the distance  $s_i$  and  $s_j$ . A covariance function must be positive definite and symmetric.

Consider a Gaussian random field  $Z(s)$ , such that the realizations  $\{z(s_i), s_i \in D, i = 1, \dots, n\}$ ,  $D \subseteq \mathbb{R}^2$  are of interest. Note that the information we seek can be described by the mean  $\mathbb{E}(Z(s))$  and the covariance matrix  $Cov(s_i, s_j)$ . Thus, we could describe the mean by a linear function of covariates, and the covariance matrix as  $Cov(s_i, s_j) = \sigma^2 \rho(\|s_i - s_j\|)$ , in which  $\rho(\|s_i - s_j\|)$  is the correlation function.

Let  $d(s_i, s_j)$  be Euclidean distance from site  $i$  to  $j$ . The four most commonly used correlation function [71,74] are given by,

$$\begin{aligned}
 \text{Exponential : } \rho(s_i, s_j) &= \exp\left(-3\frac{d(s_i, s_j)}{r}\right), \\
 \text{Gaussian : } \rho(s_i, s_j) &= \exp\left(-3\frac{d(s_i, s_j)^2}{r^2}\right), \\
 \text{Spherical : } \rho(s_i, s_j) &= \begin{cases} 1 - \frac{2}{\pi} \left(\frac{d(s_i, s_j)}{r}\right) \sqrt{1 - \left(\frac{d(s_i, s_j)}{r}\right)^2} + \sin^{-1}\left(\frac{d(s_i, s_j)}{r}\right), & d(s_i, s_j) \leq r, \\ 0, & d(s_i, s_j) > r \end{cases} \quad (\text{A11}) \\
 \text{Matern : } \rho(s_i, s_j) &= \frac{1}{\Gamma(v)2^{v-1}} \left(S_v \frac{d(s_i, s_j)}{r}\right)^v K_v\left(S_v \frac{d(s_i, s_j)}{r}\right),
 \end{aligned}$$

in which  $K_v$  is a modified Bessel function of the second kind of order  $v$ ,  $S_v$  is the scaling factor, and  $r > 0$  defines a significant range of correlation for exponential, Gaussian, and Matern correlation function. Whereas in spherical correlation function,  $r$  is defined as the correlation length [74], and  $v$  is the smoothness parameter, which measures the differentiability of the Gaussian random field [121].

**–Skewed Gaussian random field**

Skewed form of the Gaussian random field has been used to model a skewed air pollution data. A skewed Gaussian random field cannot be defined in the same way as a Gaussian random field due to its marginals’ dependence on its component parameter. For condensed information, see [33,122,123].

**–Geostatistical model**

The geostatistical model incorporates an exponential distance decay on the average of a Gaussian process in a geostatistical model given by:

$$\begin{aligned}
 z_i &\sim N(\mu_i, \sigma^2), \\
 \mu_i &= \exp(-(\phi d_{ij})^k),
 \end{aligned} \quad (\text{A12})$$

in which  $d_{ij}$  is the distance between points  $i$  and  $j$ ,  $\phi$  controls the decay rate and  $k$  is the smoothing parameter. The decay rate is modeled as *uniform*(0.1, 6). However, the parameter could be intuitively determined from the random process of interest [117]. Moreover, other valid spatial covariance function could replace the exponential function.

**–Stochastic Partial differential Equation (SPDE)**

SPDE, proposed by Lindgren [17], enables fitting a GRF with a continuously and smoothly decaying covariance function while gaining computation benefits from a GMRF representation. It represents a continuous random field using a discretely indexed spatial process. According to Blangiardo and Cemeletti [69], the SPDE model is defined as follows. Let  $\mathbf{z}(\mathbf{s})$  be as defined previously for continuous spatial points  $s \in D \subset R^d$ . The SPDE model is given by:

$$(k^2 - \Delta)^{\alpha/2} \tau \mathbf{z}(\mathbf{s}) = \mathbb{W}(s),$$

$\Delta = \frac{\partial}{\partial s^2}$  is the Laplacian,  $\alpha$  controls the smoothness,  $\tau$  controls the variance.  $\mathbb{W}(s)$  assumes a Gaussian spatial white noise process. However,  $\mathbb{W}(s)$  can also assume a Laplace noise, especially, when the data exhibit spikes and heavy tail. The solution to the aforementioned differential equation is a stationary Gaussian Markov random field  $\mathbf{z}(\mathbf{s})$  with Matern

covariance structure  $\sigma p(s_i, s_j)$ , in which  $p(s_i, s_j)$  is a Matern correlation function between site  $i$  and  $j$  as defined previously,

$$\alpha = v + \frac{d}{2},$$

$$\sigma^2 = \frac{\Gamma(v)}{\Gamma(\alpha)(4\pi)^d/2k^{2v}\tau^2},$$

in which  $k = \frac{S_v}{r}$ ,  $S_v$ ,  $r$  and  $v$  are as previously defined in Equation (A11). The solution to the SPDE model is approximated by a basis function representation defined on a triangulation of the domain  $D$ . That is  $z(s) = \sum_{g=1}^G \varphi_g(s)\tilde{x}_g$ ,  $G$  is the total number of vertices of the triangulation,  $\{\varphi\}$  is the set of basis function, and  $\{\tilde{x}_g\}$  are zero-mean Gaussian distributed weights.

*Appendix A.4. Class of Non-Gaussian Random Fields Models*

A non-Gaussian process has increasingly been useful for modeling extreme ill-behaved random processes, such as predicting an earthquake and atmospheric temperature. This section discussed the asymmetric Laplace process, inverse Wishart distribution, Pareto process, and Dirichlet process.

**–Asymmetric Laplace Process**

Suppose  $Y(s)$  is an Asymmetric Laplace random vector with parameter  $p$  and  $\tau$ ,  $AL(p, 0, \tau)$ . According to Kuzobowski and Pogorski [85] and Fontanella et al. [91]  $Y(s)$  can be written as a sum of normal and exponential process, given by,

$$Y_p(\mathbf{s}) = \sqrt{\frac{2\zeta(\mathbf{s})}{\tau p(1-p)}}Z(\mathbf{s}) + \frac{1-2p}{p(1-p)}\zeta(\mathbf{s}),$$

$$Z(\mathbf{s}) \sim GP(\mathbf{0}, \rho_z(\mathbf{s}, \mathbf{s}^*; \boldsymbol{\theta})),$$

$$\zeta(\mathbf{s}) \sim \text{Gamma}(1, \tau),$$
(A13)

in which  $\rho_z(\mathbf{s}, \mathbf{s}^*; \boldsymbol{\theta})$  is a valid spatial covariance function.  $Z(\mathbf{s})$  is a Gaussian random field to account for spatial errors and it exists as a normal standard in its marginal form.  $\zeta(\mathbf{s})$  is marginally exponential distribution with rate  $\tau$ , and it is conditionally independent of  $Z(\mathbf{s})$ .

Thus, the conditional distribution of  $Y_p(s)$  given  $\zeta(\mathbf{s})$  is given by,

$$Y_p(\mathbf{s})|\zeta(\mathbf{s}) \sim N\left(\frac{1-2p}{p(1-p)}\zeta(\mathbf{s}), \frac{2}{\tau p(1-p)}\zeta(\mathbf{s})\right).$$
(A14)

Kristian and Alan [124] discussed approaches to model  $\zeta(\mathbf{s})$  in a generalized quantile regression. In the spatial case, they defined  $\zeta(\mathbf{s})$  through CDF or copula transformation by letting  $\zeta(\mathbf{s}) = -\frac{\log(\Phi(V_\zeta(\mathbf{s})))}{\tau} = F_\tau^{-1}(\Phi(V_\zeta(\mathbf{s})))$ , in which  $V_\zeta(\mathbf{s})$  is, once again, a Gaussian process with a valid spatial covariance,  $\Phi$  is a standard normal CDF, and  $F_\tau$  is an exponential distribution CDF.

**–Multivariate Log-Gamma process**

Let matrix  $V \in \mathbb{R}^n \times \mathbb{R}^n$ , and  $\boldsymbol{\mu} \in \mathbb{R}^n$ . Let  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_n)'$  be an  $n$  mutually independent log-Gamma random variables with corresponding shape and scale parameter  $\boldsymbol{\kappa} = (\kappa_1, \dots, \kappa_n)'$  and  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)'$  respectively. That is  $\gamma_i \sim LG(\kappa_i, \alpha_i)$ . According to Bradley et al. [87], a  $n$  dimensional vector  $\mathbf{q} = \boldsymbol{\mu} + V\boldsymbol{\gamma}$  is a multivariate log-Gamma random variable and its distribution, denoted by  $MLG(\boldsymbol{\mu}, V, \boldsymbol{\kappa}, \boldsymbol{\alpha})$ , mean and variance are given by,

$$\begin{aligned}
 f(q|\mu, V, \kappa, \alpha) &= \frac{1}{|\mathbf{V}\mathbf{V}'|^{1/2}} \left( \prod_{i=1}^n \frac{\alpha_i^{\kappa_i}}{\Gamma(\kappa_i)} \right) \exp \left[ \kappa' \mathbf{V}^{-1}(q - \mu) - \alpha' \exp\{\mathbf{V}^{-1}(q - \mu)\} \right]; \quad \mathbf{q} \in \mathbb{R}^n, \\
 \mathbb{E}(\mathbf{q}) &= \mu + \mathbf{V}(\omega_0(\kappa) - \log(\alpha)), \\
 Cov(\mathbf{q}) &= \mathbf{V} \text{diag}(\omega_1(\kappa)) \mathbf{V}',
 \end{aligned}
 \tag{A15}$$

in which  $|\mathbf{A}|$  is a determinant of a matrix  $\mathbf{A}$ .

Consider a spatial random process  $Z(s)$  distributed as a multivariate log-Gamma, Yang et al. [125], and Hu and Bradley [126] described  $Z(s)|\theta, \kappa, \alpha \sim MLG(\mu, \Sigma^{1/2}, \kappa\mathbb{I}, \alpha\mathbb{I})$ , in which  $\Sigma = \sigma^2\rho(\theta, d(s, s^*))$  is a valid spatial covariance matrix,  $\rho(\theta, d(s, s^*))$  is a correlation function,  $d(s, s^*)$  is a function of location  $s$  and  $s^*$ ,  $\theta$  is a vector of some parameters of interest, and  $\mathbb{I}$  is an identity matrix of appropriate dimension. An appropriate prior distribution is assigned to  $\alpha$  and  $\kappa$ .

**–Student-t Process**

Some random processes exhibit heavy tail property, which may be inappropriate for using a Gaussian distribution. The heavy tail property of Student-t distribution enables extreme observations to be captured. Hence, adopting Student-t distribution to model heavy tail spatial random process improves uncertainty quantification. Suppose  $Z(s)$ , as defined previously, is a random process that exhibits heavy tail property, observed at location  $s \in \mathcal{D}$ . This can be represented by setting  $Z(s)$  to be distributed as a centered multivariate Student-t distribution  $\mathbf{Z}(s) \sim t_n(0, \Sigma, v)$ , in which  $\Sigma$  is a covariance matrix which is determined by a valid covariance function. The  $i, j$ th element of  $\Sigma$  is given by  $\frac{v}{v-2}Cov(Z(s_i), Z(s_j))$ , and  $v$  is the degree of freedom. The  $\Sigma$  is used to account for spatial dependencies. Moreover, the covariance matrix can further account for spatially structured and unstructured (nugget) effects [127]. The main challenge of the Student-t process is the difficulty of assigning appropriate prior specification on  $v$  to make inferences in a Bayesian analysis. Fonseca [86] derived a Jeffreys-rule before  $v$ , which leads to a proper posterior distribution. Moreover, Ordoñez [77] extended it to a spatial domain and derived a reference prior to the joint spatial hyperparameter and the degrees of freedom  $v$ .

**–Spatial Mixture model**

Green and Richardson [84] broaden the application of hidden Markov models for the random component  $z_i$  by extending it to a spatial domain. It uses the model benefits of the Hidden Potts model. The allocation of the mixture components to clusters uses a spatial dependence structure, and the numbers of clusters are considered to be random. Suppose  $\mathbf{z}$  is a spatial random field, according to Best and Thomson [128], the model specification is given in (A16). Areas estimated to belong to the same clusters do not need to be contiguous,

$$\begin{aligned}
 z_i &= \log(\eta_{w_i}), i = 1, 2, \dots, n, \\
 w_i &\in \{1, 2, \dots, k\} \\
 p(\mathbf{w}|\boldsymbol{\psi}, k) &= \exp(\boldsymbol{\psi}U(\mathbf{w}) - \delta_{\mathbf{k}}(\boldsymbol{\psi})) \\
 \eta_j &\sim \text{Gamma}(\alpha, \beta), j = 1, 2, \dots, k, \\
 k &\sim \text{Uniform}(1, c_{max}),
 \end{aligned}
 \tag{A16}$$

in which  $\boldsymbol{\psi} > 0$ ,  $c_{max}$  is the upper bound of the number of clusters,  $U(\mathbf{w}) = \sum_{i \sim i'} \mathbb{I}[z_i = z_{i'}]$  is the number of the same labels pairs to  $i$  in a neighboring area  $i' \neq i$ , parameter  $\eta_j$  is associated with each component, and  $\delta_{\mathbf{k}}(\boldsymbol{\psi}) = \log(\sum_{w \in \{1, 2, \dots, k\}^n} \exp^{\boldsymbol{\psi}U(w)})$  is the normalizing constant, in which the sum is the total amount of possible ways for the allocation for the  $n$  areas. The estimation of the numbers of clusters or components is obtained through reversible jump Markov chain Monte Carlo. Notice that  $p(\mathbf{w}|\boldsymbol{\psi}, k)$  represents the probability function of the Potts model [129].

**–Spatial Partition Model**

Leonhard and Raßer [18] proposed an approach for cluster detection, implemented using the reversible jump Markov chain Monte Carlo. The location of clusters, the number of clusters, and the random process are unknown. Best and Thomson [128] summarized the model as follows. A random number  $k$  areas are selected as clusters,  $g_j, j = 1, 2, \dots, k$ . Conditioning on these  $k$  areas, the remaining areas are assigned to their closest clusters  $j \in \{1, 2, \dots, k\}$ . For a given cluster  $k$ , the positioning of clusters is assumed to be equally probable. All areas are assumed to be contiguous, unlike the mixture model. The model specification is given by,

$$\begin{aligned} z_i &= \eta_{\gamma_i}, i = 1, 2, \dots, n, \\ \gamma_i &\in \{1, 2, \dots, k\} \\ \log(\eta_j) &\sim N(\alpha, \sigma^2), j = 1, 2, \dots, k, \\ k &\sim \text{Uniform}(1, c_{max}) \text{ or Geometry,} \end{aligned} \tag{A17}$$

in which  $\mathbf{z}$  is a random field, and parameter  $\eta_j$  is associated with each distinct geographic cluster.

**–Global Spline Mode**

Similarly to spatial models for area data, the spatial spline model assumes that each random field is centered on the centroid of a specific area in spatial domain  $D$ . Aiming to separate a global geographical trend and local spatial trends, Lee and Durbán [82] proposed a two-dimensional P-spline at the centroids of each area in  $D$  and this was further extended to incorporate a random effect which is modeled using a CAR model. It was termed smooth-CAR models. According to Lee and Durbán, the spatial P-spline model is described as follows. Suppose that  $(s_{1i}, s_{2j}, z_{ij})$  are normally distributed spatial data, in which  $s_{1i}, s_{2j}$  are respectively the longitude and latitude of the  $i$ th centroid, and  $z_{ij}$  is the response variable. The spline model is defined as:

$$\mathbf{z} = f(s_1, s_2) + \epsilon = \mathbf{B}\theta + \epsilon,$$

in which  $\theta$  is a vector of coefficients,  $\mathbf{B}$  is a regression basis constructed from the coordinates  $s_1$  and  $s_2$ , and  $\epsilon \sim N(0, \sigma^2\mathbb{I})$ . Then, the P-spline approach penalizes the squared error loss with a penalty matrix  $P$  depending on  $\lambda$ . That is,

$$S(\theta; z, \lambda) = (z - \mathbf{B}\theta)'(z - \mathbf{B}\theta) + \theta'P\theta.$$

When observed data violates the normality assumption and generally belongs to an exponential family with link function  $g$ , and  $\eta = g(\mu) = \mathbf{B}\theta$  with penalized sum of squares  $l_p(\theta) = l(\theta) + \theta'P\theta$ , in which  $l(\theta)$  is the data likelihood. The penalized score is given as:

$$(\mathbf{B}'\tilde{\mathbf{W}}_\delta\mathbf{B} + \mathbf{P})\hat{\theta} = \mathbf{B}'\tilde{\mathbf{W}}_\delta\mathbf{B}\tilde{\theta} + \mathbf{B}'(y - \tilde{\mu}),$$

in which  $\mathbf{W}_\delta$  is a diagonal matrix with element  $w_{ii} = \left(\frac{\partial \eta_i}{\partial \mu_i}\right)^2 \text{var}(y_i)$ , and terms with tilde represent an approximate solution and terms with hat are the improved approximation.

**–Dirichlet Process (DP)**

Dirichlet process is a random process with sample functions almost certainly a probability measure proposed by Ferguson [130]. In the Bayesian framework, DP is a technique of analyzing nonparametric problems [131]. Let  $\{Z(\mathbf{s}) : \mathbf{s} \in \mathbf{D}\}, \mathbf{D} \subset \mathbb{R}^d$  be the realizations with replicates of a spatial random field at distinct locations  $\mathbf{s}$ . Gelfand et al. [88] described the spatial Dirichlet process as follows. Let  $Z_t = (Z_t(s_1), \dots, Z_t(s_n))'$ , in which  $t$  represents replicates at site  $s_j$ . A Dirichlet process is a random probability measure defined on a measurable space  $(\Omega, \mathbb{B})$ , denoted by  $DP(vG_0)$ , in which  $v > 0$  is the precision parameter and  $G_0$  is a specific base probability distribution over  $(\Omega, \mathbb{B})$ . Let  $k_i, i = 1, 2, \dots$  be independent and identically distributed  $Beta(1, v)$ . The resulting random process for  $Z(\mathbf{s})$  from Dirichlet

process  $DP(vG_0)$  defined on  $(\Omega, \mathbb{B})$  can be written as  $\sum_{i=1}^{\infty} \lambda_i \delta_{\theta_i, D}$ , in which  $\delta_k$  represents a point mass at  $k$  and  $\lambda_1 = k_1$ ,  $\lambda_i = k_i \prod_{j=1}^{i-1} (1 - k_j)$ ,  $i = 2, 3, \dots$ , and  $\{\theta_i(s) : s \in D\}$  are realizations from base probability  $G_0$  which can possibly be a stationary Gaussian process. Notice that the DP process is independent, however, MacEachern [132] relaxed this condition by deriving the dependent DP. Moreover, Gelfand et al. [88] extended the dependent DP to account for spatial dependence. The property that the DP process is almost certainly discrete restricts its applications to a wide class of continuous problems. Hence, Antoniak [131] derived a mixtures of DP to circumvent the problems, and thus extended it to handle continuous cases.

Appendix A.5. Statistical Models

**–Generalized Linear Mixed Model**

A classical linear model formulation is given by:

$$\begin{aligned}
 y_i &\sim N(\mu_i, \sigma^2) \\
 y_i &= \beta_0 + \sum_{j=1}^J \beta_j x_{1i} + \epsilon_i \\
 \epsilon_i &\sim N(0, \sigma^2), \quad i = 1, 2, \dots, n \\
 \mathbb{E}(y_i | \mu_i) &= \mu_i \\
 &= \beta_0 + \sum_{j=1}^J \beta_j x_{1i},
 \end{aligned}
 \tag{A18}$$

and the matrix form is given by,

$$Y = X\beta + \epsilon,
 \tag{A19}$$

in which the latent field  $\beta = (\beta_0, \beta_1, \dots, \beta_p)'$  defines the relationship between response variable  $Y$  and covariate  $X$  [133]. Each outcome  $y_i$  is assumed to be generated according to a Gaussian distribution. The mean depends on related covariates through  $\mathbb{E}(y|\mu)$ . A generalized form of (A19) relaxes the assumption that the errors are Gaussian distributed and each outcome is generated from a non-Gaussian distribution such as Binomial, Poisson, Beta, among others [134]. It enables these random processes to be modeled through a link function and enables the magnitude of the measurement error to be a function of the predicted estimates. That is,

$$\begin{aligned}
 y_i &\sim \mathbb{P}(\cdot | \theta), \\
 g(\theta) &= \beta_0 + \sum_{j=1}^J \beta_j x_{ji} + \epsilon_i,
 \end{aligned}
 \tag{A20}$$

in which  $g$  is an appropriate link function such as the *logit* for a Binomial model and  $\log_e$  for the Poisson model. The mixed form of Equation (A20) incorporates a function  $f(\cdot)$  to relax the linearity assumption on covariates or to introduce a random effect usually modeled as a random walk, auto-regressive, or penalized spline models [134]. The mixed form is given in (A21):

$$\begin{aligned}
 y_i &\sim \mathbb{P}(\cdot | \theta) \\
 g(\theta_i) &= \beta_0 + \sum_{j=1}^J \beta_j x_{ji} + \sum_{k=1}^K f_k(z_{ki}) + f^*(v_i) + \epsilon_i,
 \end{aligned}
 \tag{A21}$$

in which  $z_k$  is the  $k$ th random effect variable and  $v_i$  is a spatial effect variable assigned spatial prior distribution. The latent field of interest is given as  $\Theta = \{\beta, f, v\}$ . Equation (A21)

is termed the Generalized Linear Mixed Model (GLMM). In a Bayesian setting, all the parameters in the model are considered to be random, and appropriate prior distributions are assigned. In the absence of prior information, a noninformative prior is assigned to  $\beta \sim Normal(0, 10^6)$  and  $\log(1/\sigma^2) \sim loggamma(1, 10^{-5})$  [69]. A Bayesian hierarchical model contains the data, prior, and hyperprior levels,

$$\begin{aligned} y_i | \Theta, \phi &\sim \mathbb{P}(y | \Theta, \phi), \\ \Theta | \phi &\sim \mathbb{P}(\Theta | \phi), \\ \phi &\sim \mathbb{P}(\phi), \end{aligned} \tag{A22}$$

in which  $\theta_i \in \Theta | y$  is of main interest. These models are usually referred to as latent Gaussian models, which are flexible and can accommodate a wide range of models.

**–Survival Model**

A survival data analysis models the time of the event occurrence. This can be time to death of subject under study, process failure time, or time to radioactive emission. Let  $T_i$  be a random variable of survival times, then  $S(t|\tau) = \mathbb{P}(T > t|\tau)$  is the survival function that, for example, determines the probability of a patient surviving (default/failure occurs) over time  $t$ . It is assumed that all subjects are alive,  $S(0|\tau) = 1$ , and all subjects will eventually die  $\lim_{t \rightarrow +\infty} S(t|\tau) = 0$ . The survival function is expressed as a distribution function  $F(t) = \mathbb{P}(T < t|\tau) = 1 - S(t|\tau)$  with a probability distribution  $f(t|\tau)$ . The hazard function  $h(t)$  measures the probability that an event will occur at a small instance  $\Delta t$  after the subject has survived through time  $t$ . That is,

$$\begin{aligned} h(t) &= \lim_{\Delta t \rightarrow 0} \frac{\mathbb{P}(T < t + \Delta t | T > t)}{\Delta t}, \\ H(t) &= \int_0^t h(u) du, \end{aligned}$$

in which  $H(t)$  is the cumulative hazard function.

Let  $(t_i, \delta_i, x_i)$  represent survival data, in which  $T_i$  represents survival times of the  $i$ th subject of interest,  $x_i$  is a predictor variable, and  $\delta_i = \mathbb{I}(T_i \leq C)$ , in which  $C$  is a censoring threshold set a priori. Considering the covariates, the likelihood of the model is expressed by,

$$\begin{aligned} g(\tau_i) &= \beta_0 + \sum_{j=1}^J \beta_j x_{ji} + \sum_{k=1}^K f(z_k) + v_i + \epsilon_i, \\ L(\theta|t) &= \prod_{i=1}^n f(t_i|\tau)^{\delta_i} S(t_i|\tau)^{1-\delta_i}, \end{aligned} \tag{A23}$$

in which  $\theta = \{\beta_1, \dots, \beta_j, \tau, \Phi\}$ ,  $\beta_i$  is a fixed effect,  $z_k$  is a random effect,  $v_i$  with parameter  $\Phi$ , is a spatial random effect assigned a spatial prior in a Bayesian framework [135,136].  $f(t_i|\tau)$  can assume Exponential, Weibull, Logistics, or lognormal distribution to number a few. The censoring status  $\delta$  controls the contribution to the likelihood of subjects that experienced the event and those that survived through the entire study period. Kaplan-Meier provides nonparametric estimates of the survival curves. The proportional hazards model and accelerated failure time model are the most frequently used frequentist parametric methods to analyze survival data [99].

**–Bayesian Spatial Econometrics**

Spatial analytical tools are widely used in Economics to quantify the spatial dependencies and heterogeneity of economic variables. Referred to as spatial econometrics, they extend the traditional econometrics to a spatial domain. The most frequently used models are the Spatial Autoregressive (SAR), Spatial Error Model (SEM), and Spatial Durbin Model (SDM). The models are briefly described as follows.

Let  $Y$  be a response variable assuming a linear relationship with explanatory variables  $X$ , the SAR model is represented by,

$$\begin{aligned} Y &= \rho WY + X\beta + \epsilon, \\ \epsilon &\sim MVN(0, \sigma^2 I). \end{aligned} \tag{A24}$$

$W$  is a spatial weight matrix,  $\rho$  is a spatial autocorrelation parameter, and  $\beta$  is a vector of the regression parameters. The SAR model assumes that the dependent variable is spatially autocorrelated [137]. In contrast, the SEM model assumes the error is a spatial correlation. The SEM formulation is given by,

$$\begin{aligned} Y &= X\beta + u, \\ u &= \rho Wu + \epsilon, \\ \epsilon &\sim MVN(0, \sigma^2 I). \end{aligned} \tag{A25}$$

The SDM is an extension of the SAR model which assumes that the dependent variable and covariates are spatially correlated. The formulation is given by,

$$\begin{aligned} Y &= \rho WY + WX\beta + \epsilon, \\ \epsilon &\sim MVN(0, \sigma^2 I). \end{aligned} \tag{A26}$$

According to LeSage and Pace [138], SDM is appropriate when the included covariates are correlated with spatially correlated variables not included in the model. In the next sections, we present each class main idea of the spatial literature computation techniques.

Appendix A.6. Computation Technique

–Maximum Likelihood Estimation

The maximum likelihood estimation approach involves an optimization problem to determine the best sets of distribution parameters representing data. In spatial statistics, the MLE method is usually used to estimate global spatial dependencies in a spatial econometric model. Authors often compare its estimates with one obtained in a Bayesian inferential framework and highlight the MLE approach’s inadequacies to spatial statistical inferences [139–142].

Let  $y_i \sim f(\cdot | \theta_i)$ , in which  $\theta_i \in \Theta$  is a parameter of interest and  $\Theta$  is the parameter space.  $y_i$  do not need to be independent and identical. The estimation procedure involves computing the likelihood (log-likelihood) of the data distribution and determining the parameter sets  $\theta^* = \{\theta_i^* : \theta_i^* \in \Theta, i = 1, 2, \dots, n\}$  in which the likelihood is maximum. For independent  $y_i, i = 1, 2, \dots, n$ , the likelihood is given by,

$$\begin{aligned} L(\theta | \mathbf{y}) &= \prod_{i=1}^n f(y_i | \theta_i), \\ \theta^* &= \arg \max_{\theta} \log(L(\theta | \mathbf{y})). \end{aligned} \tag{A27}$$

An equivalent approach to the classical MLE when the dimension of the model parameters is large or complex for estimation is the Quasi-Maximum Likelihood Estimation (QMLE). Unlike the MLE, the QMLE maximizes a function  $\log L^*(\theta | \mathbf{y})$  that is related to the logarithm of the likelihood function  $L(\theta | \mathbf{y})$ . Su and Yang [143] proposed a QMLE of dynamic panel models with spatial errors and this was further broadened by Yu, De Jong, and Lee et al. [142]. An equivalent approach to MLE in the Bayesian setting is the Maximum A Posteriori (MAP). It involves maximizing the posterior conditional probability,

$$\theta^* = \arg \max_{\theta} p(\theta) p(\mathbf{y} | \theta), \tag{A28}$$

in which  $p(\theta)$  is the prior distribution and  $p(\mathbf{y}|\theta)$  is the data likelihood defined as  $L(\mathbf{y}|\theta)$  in (A27).

**–Expectation–Maximization (EM Algorithm)**

The maximum likelihood estimation limitation is the assumption that the variables that generate the process are all observable. In practice, this assumption rarely stands. One possible choice to overcome the limitation is the estimation through the EM algorithm.

The EM algorithm, proposed by Dempster et al. [144], fits a model to a latent representation of data rather than merely fitting distribution models. It can work well in data that contains unobserved (latent) variables. The algorithm, an iterative method, has two major stages: estimating the latent variables (E-step) and maximizing the model parameters given the data and the estimated variables (M-step).

Let  $\theta$  be an initialized model parameter, the E-step is used to update the latent space variables  $z$ , usually discrete or cluster in particular, through  $p(\mathbf{z}|\mathbf{y}, \theta)$ , in which  $\mathbf{y}$  is the observed data. To update  $\theta$ , the expectation  $\mathbb{E}_{\mathbf{z}|\mathbf{y}, \theta^*} \log(p(\mathbf{y}|\mathbf{z}, \theta))$  is computed.  $\theta^*$  represents the previous parameter and  $\theta$  is the potential new parameter of the model:

$$\mathbb{E}_{\mathbf{z}|\mathbf{y}, \theta^*} \log(p(\mathbf{y}|\mathbf{z}, \theta)) = \sum_{i=1}^n \sum_{j=1}^k p(z_i = j|\mathbf{y}, \theta^*) \log[p(z_i = j|\theta)p(y_i|z_i = j, \theta)]. \tag{A29}$$

In the M-step, the EM algorithm maximizes the model parameter in the equation:

$$\theta^* = \arg \max_{\theta} \mathbb{E}_{\mathbf{z}|\mathbf{y}, \theta^*} \log(p(\mathbf{y}|\mathbf{z}, \theta)). \tag{A30}$$

The iteration continues until the difference between the current and the previous expectation is less than  $\epsilon > 0$  set at the initial stage.

**–Markov Chain Monte Carlo (MCMC)**

Given a posterior distribution:

$$p(\theta, \psi|\mathbf{y}) \propto p(\mathbf{y}|\theta, \psi) \times p(\theta, \psi), \tag{A31}$$

and assuming that the posterior  $p(\theta|\mathbf{y}, \Psi)$  is of known form, such as a standard probability distribution, we can resort to the Monte Carlo approach to approximate posterior quantities  $h(\theta)$ ,

$$\mathbb{E}(h(\theta)|\mathbf{y}) = \int_{\theta \in \Theta} p(\theta|\mathbf{y}) d\theta = \int_{\theta \in \Theta} \int_{\psi \in \Psi} h(\theta) p(\theta|\psi, \mathbf{y}) p(\psi|\mathbf{y}) d\psi d\theta,$$

which could be the mean, median or higher moments. The procedure consists of simulating  $m$  random samples from  $p(\theta|\mathbf{y})$ , say  $\{\theta^1, \theta^2, \dots, \theta^m\}$  and evaluating the unknown quantity  $h(\theta)$  using the empirical average:

$$\mathbb{E}(h(\hat{\theta})|\mathbf{y}) = \frac{\sum_{i=1}^m h(\theta^i)}{m}. \tag{A32}$$

Under the Law of Large Numbers, the empirical distribution will converge to the true distribution. In the case of a joint posterior distribution, an approximation of the posterior marginals is achieved by first sampling from a marginal distribution of a subset of parameters given its complements and then using it to evaluate the full joint distribution.

In practice, the posterior distribution’s functional form is unknown or complex, and independent samples are not feasible. An alternative approach comprises generating independent samples from an importance distribution  $\pi(\theta|\mathbf{y})$  which is a close distribution to  $p(\theta|\mathbf{y})$ . Empirically, the quantity  $h(\theta)$  is obtained as:

$$\mathbb{E}(h(\hat{\theta})) = \frac{\sum_{i=1}^m h(\theta_i) \pi(\theta_i)}{\sum_{i=1}^m \pi(\theta_i)}. \tag{A33}$$

This approach is not trivial for a large number of dimensions of  $\theta$  [145]. A more widely used alternative approach comprises generating correlated samples by running a Markov chain whose stationary distribution converges to the posterior density. The posterior summaries are computed from these samples using the empirical method, as described above. Suppose  $\chi$  is the state space of the posterior distribution. As stated by Blangiardo [69], the convergence of the Markov chain stationary distribution to the posterior distribution requires the Markov chains to be irreducible (the chain has a positive probability of reaching all region of  $\chi$  regardless of the starting point), recurrence (the limit of the probability of the chain visiting a subset  $\chi$  infinitely many times is 1), and aperiodic (the chain does not circle when exploring  $\chi$ ). The highlighted procedure is referred to as MCMC. The Gibbs sampler and Metropolis-Hastings algorithms are the most frequently used standard MCMC algorithm in Bayesian inference literature. For a description of these algorithms, see [69] (pp. 91–103).

**–Integrated Nested Laplace Approximation (INLA)**

INLA, proposed by Rue et al. [79], is an alternative approach to the estimation of posterior marginals. It has gained considerable attention and has been proven to outperform the MCMC approach in computational speed [79]. The availability of the  $R - INLA$  simplifies the implementation of the approach and authors from various fields have found it useful and easy.

Again, consider the posterior distribution:

$$p(\theta_i | \mathbf{y}) \int p(\theta_i, \boldsymbol{\psi} | \mathbf{y}) d\boldsymbol{\psi} = \int p(\theta_i | \boldsymbol{\psi}, \mathbf{y}) p(\boldsymbol{\psi} | \mathbf{y}) d\boldsymbol{\psi}. \tag{A34}$$

The objective is to obtain the posterior marginals  $p(\theta_i | \mathbf{y})$  for each parameter in the vector and the estimates of the hyperparameters given by,

$$p(\boldsymbol{\psi}_k | \mathbf{y}) \int p(\boldsymbol{\psi}_k | \mathbf{y}) d\boldsymbol{\psi}_{-k}. \tag{A35}$$

The INLA approach uses the model assumptions to approximate the marginal posterior distribution and its moments based on Laplace approximation [146]. According to [69,147], INLA approximation follows the following steps. Firstly, the posterior marginals of the hyperparameters are approximated, that is,

$$\begin{aligned} p(\boldsymbol{\psi} | \mathbf{y}) &= \frac{p(\boldsymbol{\theta}, \boldsymbol{\psi} | \mathbf{y})}{p(\boldsymbol{\theta} | \boldsymbol{\psi}, \mathbf{y})} \propto \frac{p(\boldsymbol{\psi}) p(\boldsymbol{\theta} | \boldsymbol{\psi}) p(\mathbf{y} | \boldsymbol{\theta})}{p(\boldsymbol{\theta} | \boldsymbol{\psi}, \mathbf{y})}, \\ &\approx \frac{p(\boldsymbol{\psi}) p(\boldsymbol{\theta} | \boldsymbol{\psi}) p(\mathbf{y} | \boldsymbol{\theta})}{\tilde{p}(\boldsymbol{\theta} | \boldsymbol{\psi}, \mathbf{y})} \Big|_{\boldsymbol{\theta} = \boldsymbol{\theta}^*(\boldsymbol{\psi})} \tilde{p}(\boldsymbol{\psi} | \mathbf{y}), \end{aligned}$$

in which  $\tilde{p}(\boldsymbol{\theta} | \mathbf{y})$  is a Gaussian approximation for  $p(\boldsymbol{\theta} | \mathbf{y})$  and  $\boldsymbol{\theta}^*$  is the mode. Secondly, the parameter vector is partitioned in a way that  $\boldsymbol{\theta} = (\theta_i, \boldsymbol{\theta}_{-i})$  and again approximated using the Laplace procedure to obtain:

$$p(\theta_i | \boldsymbol{\psi}, \mathbf{y}) = \frac{p(\theta_i, \boldsymbol{\theta}_{-i} | \boldsymbol{\psi}, \mathbf{y})}{p(\boldsymbol{\theta}_{-i} | \theta_i, \boldsymbol{\psi}, \mathbf{y})} \approx \frac{p(\boldsymbol{\theta}, \boldsymbol{\psi} | \mathbf{y})}{\tilde{p}(\boldsymbol{\theta}_{-i} | \theta_i, \boldsymbol{\psi}, \mathbf{y})} \Big|_{\boldsymbol{\theta}_{-i} = \boldsymbol{\theta}_{-i}^*(\theta_i, \boldsymbol{\psi})} \tilde{p}(\theta_i | \boldsymbol{\psi}, \mathbf{y}). \tag{A36}$$

To bypass the computational complexity of computing  $\tilde{p}(\theta_i | \boldsymbol{\psi}, \mathbf{y})$ , INLA explores the marginal joint posterior for the hyperparameters  $p(\boldsymbol{\psi} | \mathbf{y})$  in a grid search to select the important points  $\{\boldsymbol{\psi}_k\}$  jointly with a corresponding set of weights  $\{\Delta_k\}$  to give approximates to the posterior to the hyperparameters. Each marginal  $\tilde{p}(\boldsymbol{\psi}_k | \mathbf{y}) \forall k$  can be obtained using log-spline interpolation bases on selected  $\boldsymbol{\psi}_k$  and  $\Delta_k$ . For each  $k$ , the conditional posterior  $\tilde{p}(\theta_i | \theta_i | \boldsymbol{\psi}, \mathbf{y})$  is computed and a numerical integration:

$$\tilde{p}(\theta_i | \mathbf{y}) \approx \sum_{k=1}^K \tilde{p}(\theta_i | \boldsymbol{\psi}_k, \mathbf{y}) \tilde{p}(\boldsymbol{\psi}_k | \mathbf{y}). \tag{A37}$$

is, then, used to obtain  $\tilde{p}(\theta_i|\psi, y)$ .

## References

1. Tao, W. Interdisciplinary urban GIS for smart cities: Advancements and opportunities. *Geo-Spat. Inf. Sci.* **2013**, *16*, 25–34. [[CrossRef](#)]
2. Roche, S. Geographic information science II: Less space, more places in smart cities. *Prog. Hum. Geogr.* **2016**, *40*, 565–573. [[CrossRef](#)]
3. Marjani, M.; Nasaruddin, F.; Gani, A.; Karim, A.; Hashem, I.A.T.; Siddiqa, A.; Yaqoob, I. Big IoT data analytics: Architecture, opportunities, and open research challenges. *IEEE Access* **2017**, *5*, 5247–5261.
4. Razafimandimby, C.; Loscri, V.; Vegni, A.M.; Neri, A. A Bayesian and smart gateway based communication for noisy IoT scenario. In Proceedings of the 2017 International Conference on Computing, Networking and Communications (ICNC), Silicon Valley, CA, USA, 26–29 January 2017; pp. 481–485.
5. Harquel, S.; Diard, J.; Raffin, E.; Passera, B.; Dall’Igna, G.; Marendaz, C.; David, O.; Chauvin, A. Automated set-up procedure for transcranial magnetic stimulation protocols. *Neuroimage* **2017**, *153*, 307–318. [[CrossRef](#)]
6. Meincke, J.; Hewitt, M.; Batsikadze, G.; Liebetanz, D. Automated TMS hotspot-hunting using a closed loop threshold-based algorithm. *NeuroImage* **2016**, *124*, 509–517. [[CrossRef](#)]
7. Derado, G.; Bowman, F.D.; Zhang, L.; Initiative, A.D.N. Predicting brain activity using a Bayesian spatial model. *Stat. Methods Med. Res.* **2013**, *22*, 382–397. [[CrossRef](#)]
8. Kang, J.; Johnson, T.D.; Nichols, T.E.; Wager, T.D. Meta analysis of functional neuroimaging data via Bayesian spatial point processes. *J. Am. Stat. Assoc.* **2011**, *106*, 124–134. [[CrossRef](#)] [[PubMed](#)]
9. Kang, S.Y.; Cramb, S.; White, N.; Ball, S.J.; Mengersen, K. Making the most of spatial information in health: A tutorial in Bayesian disease mapping for areal data. *Geospat. Health* **2016**, *11*, 190–198. [[CrossRef](#)]
10. Fonseca, V.P.; Pennino, M.G.; de Nóbrega, M.F.; Oliveira, J.E.L.; de Figueiredo Mendes, L. Identifying fish diversity hot-spots in data-poor situations. *Mar. Environ. Res.* **2017**, *129*, 365–373. [[CrossRef](#)]
11. Kennedy, M.C.; O’Hagan, A. Bayesian calibration of computer models. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **2001**, *63*, 425–464. [[CrossRef](#)]
12. Higdon, D.; Nakhleh, C.; Gattiker, J.; Williams, B. A Bayesian calibration approach to the thermal problem. *Comput. Methods Appl. Mech. Eng.* **2008**, *197*, 2431–2441. [[CrossRef](#)]
13. Romanowski, A.; Grudzien, K.; Williams, R.A. Analysis and interpretation of hopper flow behaviour using electrical capacitance tomography. *Part. Part. Syst. Character.* **2006**, *23*, 297–305. [[CrossRef](#)]
14. Conti, S.; O’Hagan, A. Bayesian emulation of complex multi-output and dynamic computer models. *J. Stat. Plan. Inference* **2010**, *140*, 640–651. [[CrossRef](#)]
15. Rappel, H.; Beex, L.A.; Noels, L.; Bordas, S. Identifying elastoplastic parameters with Bayes’ theorem considering output error, input error and model uncertainty. *Probabilistic Eng. Mech.* **2019**, *55*, 28–41. [[CrossRef](#)]
16. Ozturk, D.; Kilic, F. Geostatistical approach for spatial interpolation of meteorological data. *An. Acad. Bras. Ciências* **2016**, *88*, 2121–2136. [[CrossRef](#)]
17. Lindgren, F.; Rue, H.V.; Lindström, J. An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2011**, *73*, 423–498. [[CrossRef](#)]
18. Knorr-Held, L.; Raßer, G. Bayesian detection of clusters and discontinuities in disease maps. *Biometrics* **2000**, *56*, 13–21. [[CrossRef](#)] [[PubMed](#)]
19. Corpas-Burgos, F.; Martinez-Beneito, M.A. An Autoregressive Disease Mapping Model for Spatio-Temporal Forecasting. *Mathematics* **2021**, *9*, 384. [[CrossRef](#)]
20. Griffith, D.A.; Chun, Y. *Gis and Spatial Statistics/Econometrics: An Overview*; University of Texas at Dallas: Richardson, TX, USA, 2018.
21. Hernandez-Lemus, E. Random fields in physics, biology and data science. *Front. Phys.* **2021**. [[CrossRef](#)]
22. Aswi, A.; Cramb, S.; Moraga, P.; Mengersen, K. Bayesian spatial and spatiotemporal approaches to modeling dengue fever: A systematic review. *Epidemiol. Infect.* **2019**, *147*, e33. [[CrossRef](#)]
23. Duncan, E.W.; White, N.M.; Mengersen, K. Spatial smoothing in Bayesian models: A comparison of weights matrix specifications and their impact on inference. *Int. J. Health Geogr.* **2017**, *16*, 1–16. [[CrossRef](#)] [[PubMed](#)]
24. Wah, W.; Ahern, S.; Earnest, A. A systematic review of Bayesian spatial-temporal models on cancer incidence and mortality. *Int. J. Public Health* **2020**, *65*, 673–682. [[CrossRef](#)]
25. Fisher, R.A. *The Design of Experiments*; Hafner Publishing Company: New York, NY, USA, 1935.
26. Ibrahim, J.G.; Chen, M.H.; Gwon, Y.; Chen, F. The power prior: Theory and applications. *Stat. Med.* **2015**, *34*, 3724–3749. [[CrossRef](#)]
27. Bachoc, F. Parametric Estimation of Covariance Function in Gaussian-Process Based Kriging Models. Application to Uncertainty Quantification for Computer Experiments. Ph.D. Thesis, Université Paris-Diderot-Paris VII, Paris, France, 2013.
28. Hutton, B.; Salanti, G.; Caldwell, D.M.; Chaimani, A.; Schmid, C.H.; Cameron, C.; Ioannidis, J.P.; Straus, S.; Thorlund, K.; Jansen, J.P.; et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: Checklist and explanations. *Ann. Intern. Med.* **2015**, *162*, 777–784. [[CrossRef](#)]

29. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G.; Altman, D.; Antes, G.; Atkins, D.; Barbour, V.; Barrowman, N.; Berlin, J.A.; et al. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement (Chinese edition). *J. Chin. Integr. Med.* **2009**, *7*, 889–896. [[CrossRef](#)]
30. Hsieh, H.F.; Shannon, S.E. Three approaches to qualitative content analysis. *Qual. Health Res.* **2005**, *15*, 1277–1288. [[CrossRef](#)]
31. Lee, D.; Mitchell, R. Boundary detection in disease mapping studies. *Biostatistics* **2012**, *13*, 415–426. [[CrossRef](#)]
32. Riley, S.; Eames, K.; Isham, V.; Mollison, D.; Trapman, P. Five challenges for spatial epidemic models. *Epidemics* **2015**, *10*, 68–71. [[CrossRef](#)] [[PubMed](#)]
33. Karimi, O.; Mohammadzadeh, M. Bayesian spatial regression models with closed skew normal correlated errors and missing observations. *Stat. Pap.* **2012**, *53*, 205–218. [[CrossRef](#)]
34. Berliner, L.M. Spatial Statistical Methods. *Int. Encycl. Soc. Behav. Sci.* **2015**, *23*, 191–197.
35. Isard, W. The general theory of location and space-economy. *Q. J. Econ.* **1949**, *63*, 476–506. [[CrossRef](#)]
36. Sparks, P.J.; Sparks, C.S.; Campbell, J.J. An application of Bayesian spatial statistical methods to the study of racial and poverty segregation and infant mortality rates in the US. *GeoJournal* **2013**, *78*, 389–405. [[CrossRef](#)]
37. Krige, D. *Lognormal-de Wijsian Geostatistics for Ore Evaluation*; South African Institute of Mining and Metallurgy: Johannesburg, South Africa, 1978.
38. Gracia, E.; López-Quílez, A.; Marco, M.; Lladosa, S.; Lila, M. The Spatial Epidemiology of Intimate Partner Violence: Do Neighborhoods Matter? *Am. J. Epidemiol.* **2015**, *182*, 58–66. [[CrossRef](#)] [[PubMed](#)]
39. Luan, H.; Law, J.; Lysy, M. Diving into the consumer nutrition environment: A Bayesian spatial factor analysis of neighborhood restaurant environment. *Spat. Spatio-Temporal Epidemiol.* **2018**, *24*, 39–51. [[CrossRef](#)] [[PubMed](#)]
40. Morris, R.S. Diseases, dilemmas, decisions—Converting epidemiological dilemmas into successful disease control decisions. *Prev. Vet. Med.* **2015**, *122*, 242–252. [[CrossRef](#)] [[PubMed](#)]
41. Müller, I.; Betuela, I.; Hide, R. Regional patterns of birthweights in Papua New Guinea in relation to diet, environment and socio-economic factors. *Ann. Hum. Biol.* **2002**, *29*, 74–88. [[CrossRef](#)]
42. Short, M.; Carlin, B.P.; Bushhouse, S. Using hierarchical spatial models for cancer control planning in Minnesota (United States). *Cancer Causes Control CCC* **2002**, *13*, 903–916.:1021980820140. [[CrossRef](#)]
43. Ward, M. Spatial Epidemiology: Where Have We come in 150 Years? In *Geospatial Technologies and Homeland Security*; The GeoJournal Library: Dordrecht, The Netherlands, 2008; Volume 94, pp. 257–282.
44. Da Silva-Buttkus, P.; Marcelli, G.; Franks, S.; Stark, J.; Hardy, K. Inferring biological mechanisms from spatial analysis: Prediction of a local inhibitor in the ovary. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 456–461. [[CrossRef](#)]
45. Arpòn, J.; Sakai, K.; Gaudin, V.; Andrey, P. Spatial modeling of biological patterns shows multiscale organization of Arabidopsis thaliana heterochromatin. *Sci. Rep.* **2021**, *11*, 1–17. [[CrossRef](#)]
46. Li, Q.; Zhang, M.; Xie, Y.; Xiao, G. Bayesian Modeling of Spatial Molecular Profiling Data via Gaussian Process. *arXiv* **2020**, arXiv:2012.03326.
47. Adegboye, O.A.; Adekunle, A.I.; Pak, A.; Gayawan, E.; Leung, D.H.; Rojas, D.P.; Elfaki, F.; McBryde, E.S.; Eisen, D.P. Change in outbreak epicentre and its impact on the importation risks of COVID-19 progression: A modeling study. *Travel Med. Infect. Dis.* **2021**, *40*, 101988. [[CrossRef](#)] [[PubMed](#)]
48. Saavedra, P.; Santana, A.; Bello, L.; Pacheco, J.M.; Sanjuán, E. A Bayesian spatiotemporal analysis of mortality rates in Spain: Application to the COVID-19 2020 outbreak. *Popul. Health Metrics* **2021**, *19*, 1–10. [[CrossRef](#)]
49. Clements, A.C.; Lwambo, N.J.; Blair, L.; Nyandindi, U.; Kaatano, G.; Kinung'hi, S.; Webster, J.P.; Fenwick, A.; Brooker, S. Bayesian spatial analysis and disease mapping: Tools to enhance planning and implementation of a schistosomiasis control programme in Tanzania. *Trop. Med. Int. Health* **2006**, *11*, 490–503. [[CrossRef](#)] [[PubMed](#)]
50. Huertas, I.; Oldehinkel, M.; van Oort, E.S.; Garcia-Solis, D.; Mir, P.; Beckmann, C.F.; Marquand, A.F. A Bayesian spatial model for neuroimaging data based on biologically informed basis functions. *NeuroImage* **2017**, *161*, 134–148. [[CrossRef](#)]
51. Benassi, F.; Naccarato, A. Households in potential economic distress. A geographically weighted regression model for Italy, 2001–2011. *Spat. Stat.* **2017**, *21*, 362–376. [[CrossRef](#)]
52. Simões, P.; Carvalho, M.; Aleixo, S.; Gomes, S.; Natário, I. A spatial econometric analysis of the calls to the portuguese national health line. *Econometrics* **2017**, *5*, 24. [[CrossRef](#)]
53. Osama, A.; Sayed, T. A novel approach for identifying, diagnosing, and treating active transportation safety issues. *Transp. Res. Rec.* **2019**, *2673*, 813–823. [[CrossRef](#)]
54. Law, J.; Quick, M.; Jadavji, A. A Bayesian spatial shared component model for identifying crime-general and crime-specific hotspots. *Ann. GIS* **2020**, *26*, 65–79. [[CrossRef](#)]
55. Potter, J.E.; Schmertmann, C.P.; Assunção, R.M.; Cavenaghi, S.M. Mapping the timing, pace, and scale of the fertility transition in Brazil. *Popul. Dev. Rev.* **2010**, *36*, 283–307. [[CrossRef](#)]
56. Gregory, A.; Lau, F.D.H.; Girolami, M.; Butler, L.J.; Elshafie, M.Z. The synthesis of data from instrumented structures and physics-based models via Gaussian processes. *J. Comput. Phys.* **2019**, *392*, 248–265. [[CrossRef](#)]
57. Rappel, H.; Wu, L.; Noels, L.; Beex, L.A. A Bayesian framework to identify random parameter fields based on the copula theorem and Gaussian fields: Application to polycrystalline materials. *J. Appl. Mech.* **2019**, *86*, 121009. [[CrossRef](#)]
58. Koutsourelakis, P.S. A multi-resolution, nonparametric, Bayesian framework for identification of spatially-varying model parameters. *J. Comput. Phys.* **2009**, *228*, 6184–6211. [[CrossRef](#)]

59. Koutsourelakis, P.S. A novel Bayesian strategy for the identification of spatially varying material properties and model validation: An application to static elastography. *Int. J. Numer. Methods Eng.* **2012**, *91*, 249–268. [[CrossRef](#)]
60. Sharkey, P.; Winter, H.C. A Bayesian spatial hierarchical model for extreme precipitation in Great Britain. *Environmetrics* **2019**, *30*, e2529. [[CrossRef](#)]
61. Robinson, N.; Rampant, P.; Callinan, A.; Rab, M.; Fisher, P. Advances in precision agriculture in south-eastern Australia. II. Spatio-temporal prediction of crop yield using terrain derivatives and proximally sensed data. *Crop Pasture Sci.* **2009**, *60*, 859–869. [[CrossRef](#)]
62. Yousefi, K.; Swartz, T.B. Advanced putting metrics in golf. *J. Quant. Anal. Sport.* **2013**, *9*, 239–248. [[CrossRef](#)]
63. Woods, J. Two-dimensional discrete Markovian fields. *IEEE Trans. Inf. Theory* **1972**, *18*, 232–240. [[CrossRef](#)]
64. Besag, J. Spatial interaction and the statistical analysis of lattice systems. *J. R. Stat. Soc. Ser. B (Methodol.)* **1974**, *36*, 192–225. [[CrossRef](#)]
65. El-Baz, A.; Farag, A.A. Image segmentation using GMRF models: Parameters estimation and applications. In Proceedings of the Proceedings 2003 International Conference on Image Processing (Cat. No. 03CH37429), Barcelona, Spain, 14–17 September 2003; Volume 2, pp. II–177. [[CrossRef](#)]
66. Sidén, P.; Lindsten, F. Deep gaussian markov random fields. In Proceedings of the International Conference on Machine Learning PMLR, Virtual Event, 12–18 July 2020; pp. 8916–8926.
67. Vemulapalli, R.; Tuzel, O.; Liu, M.Y. Deep gaussian conditional random field network: A model-based deep network for discriminative denoising. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 4801–4809.
68. Lee, J.; Bahri, Y.; Novak, R.; Schoenholz, S.S.; Pennington, J.; Sohl-Dickstein, J. Deep neural networks as gaussian processes. *arXiv* **2017**, arXiv:1711.00165.
69. Blangiardo, M.; Cameletti, M. *Spatial and Spatiotemporal Bayesian Models with R-INLA*; John Wiley and Sons, Ltd.: Hoboken, NJ, USA, 2015.
70. Banerjee, S.; Carlin, B.; Gelfand, A. *Hierarchical Modelling and Analysis of Spatial Data*; Monographs on Statistics and Applied Probability; Chapman and Hall: New York, NY, USA, 2004.
71. Cressie, N. *Statistics for Spatial Data*; Wiley: Hoboken, NJ, USA, 1993.
72. Muñoz, F.; Pennino, M.G.; Conesa, D.; López-Quílez, A.; Bellido, J.M. Estimation and prediction of the spatial occurrence of fish species using Bayesian latent Gaussian models. *Stoch. Environ. Res. Risk Assess.* **2013**, *27*, 1171–1180. [[CrossRef](#)]
73. Leininger, T.J.; Gelfand, A.E. Bayesian inference and model assessment for spatial point patterns using posterior predictive samples. *Bayesian Anal.* **2017**, *12*, 1–30. [[CrossRef](#)]
74. Rue, H.v.; Tjelmeland, H.k. Fitting Gaussian Markov random fields to Gaussian fields. *Scand. J. Statistics. Theory Appl.* **2002**, *29*, 31–49. [[CrossRef](#)]
75. Walker, M.; Curtis, A. Eliciting spatial statistics from geological experts using genetic algorithms. *Geophys. J. Int.* **2014**, *198*, 342–356. [[CrossRef](#)]
76. Moala, F.A.; O’Hagan, A. Elicitation of multivariate prior distributions: A nonparametric Bayesian approach. *J. Stat. Plan. Inference* **2010**, *140*, 1635–1655. [[CrossRef](#)]
77. Ordoñez, J.A.; Prates, M.O.; Matos, L.A.; Lachos, V.H. Objective Bayesian analysis for spatial Student-t regression models. *arXiv* **2020**, arXiv:math.ST/2004.04341.
78. Simpson, D.; Rue, H.v.; Riebler, A.; Martins, T.G.; Sø rbye, S.H. Penalising model component complexity: A principled, practical approach to constructing priors. *Stat. Sci. A Rev. J. Inst. Math. Stat.* **2017**, *32*, 1–28. [[CrossRef](#)]
79. Rue, H.; Martino, S.; Chopin, N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **2009**, *71*, 319–392. [[CrossRef](#)]
80. Besag, J.; York, J.; Mollié, A. Bayesian image restoration, with two applications in spatial statistics. *J. R. Stat. Soc. Ser. B (Methodol.)* **1991**, *43*, 1–59. [[CrossRef](#)]
81. Leroux, B.G.; Lei, X.; Breslow, N. Estimation of disease rates in small areas: A new mixed model for spatial dependence. In *Statistical Models in Epidemiology, the Environment, and Clinical Trials*; Springer: Berlin/Heidelberg, Germany, 2000; pp. 179–191.
82. Lee, D.J.; Durbán, M. Smooth-CAR mixed models for spatial count data. *Comput. Stat. Data Anal.* **2009**, *53*, 2968–2979. [[CrossRef](#)]
83. Dean, C.; Ugarte, M.; Militino, A. Detecting interaction between random region and fixed age effects in disease mapping. *Biometrics* **2001**, *57*, 197–202. [[CrossRef](#)] [[PubMed](#)]
84. Green, P.J.; Richardson, S. Hidden Markov models and disease mapping. *J. Am. Stat. Assoc.* **2002**, *97*, 1055–1070. [[CrossRef](#)]
85. Kozubowski, T.; Podgorski, K. A multivariate and asymmetric generalization of Laplace distribution. *Comput. Stat.* **2000**, *15*, 531–540. [[CrossRef](#)]
86. Fonseca, T.C.O.; Ferreira, M.A.R.; Migon, H.S. Objective Bayesian analysis for the Student-*t* regression model [erratum to MR2521587]. *Biometrika* **2014**, *101*, 252. [[CrossRef](#)]
87. Bradley, J.R.; Holan, S.H.; Wikle, C.K. Computationally efficient multivariate spatiotemporal models for high-dimensional count-valued data (with discussion). *Bayesian Anal.* **2018**, *13*, 253–310. [[CrossRef](#)]
88. Gelfand, A.E.; Kottas, A.; MacEachern, S.N. Bayesian nonparametric spatial modeling with Dirichlet process mixing. *J. Am. Stat. Assoc.* **2005**, *100*, 1021–1035. [[CrossRef](#)]

89. Obaromi, D. Spatial modeling of some conditional autoregressive priors in a disease mapping model: The Bayesian approach. *Biomed. J. Sci. Tech. Res.* **2019**, *14*.
90. Egbon, O.A.; Somo-Aina, O.; Gayawan, E. Spatial Weighted Analysis of Malnutrition Among Children in Nigeria: A Bayesian Approach. *Stat. Biosci.* **2021**, *13*, 495–523. [[CrossRef](#)]
91. Fontanella, L.; Ippoliti, L.; Sarra, A.; Valentini, P.; Palermi, S. Hierarchical generalised latent spatial quantile regression models with applications to indoor radon concentration. *Stoch. Environ. Res. Risk Assess.* **2015**, *29*, 357–367. [[CrossRef](#)]
92. Nott, D.J.; Seah, M.; Al-Labadi, L.; Evans, M.; Ng, H.K.; Englert, B.G. Using prior expansions for prior-data conflict checking. *Bayesian Anal.* **2021**, *16*, 203–231. [[CrossRef](#)]
93. Egidi, L.; Pauli, F.; Torelli, N. Avoiding prior–data conflict in regression models via mixture priors. *Can. J. Stat.* **2021**. [[CrossRef](#)]
94. Staubach, C.; Schmid, V.; Knorr-Held, L.; Ziller, M. A Bayesian model for spatial wildlife disease prevalence data. *Prev. Vet. Med.* **2002**, *56*, 75–87. [[CrossRef](#)]
95. Teerapabolarn, K.; Jaioun, K. Approximation of binomial distribution by an improved poisson distribution. *Int. J. Pure Appl. Math.* **2014**, *97*, 491–495. [[CrossRef](#)]
96. Burr, I.W. Some approximate relations between terms of the hypergeometric, binomial and Poisson distributions. *Commun. Stat.-Theory Methods* **1973**, *1*, 297–301.
97. Alexander, N.; Moyeed, R.; Stander, J. Spatial modeling of individual-level parasite counts using the negative binomial distribution. *Biostatistics* **2000**, *1*, 453–463. [[CrossRef](#)] [[PubMed](#)]
98. Krisztin, T.; Piribauer, P.; Wögerer, M. A spatial multinomial logit model for analysing urban expansion. *Spat. Econ. Anal.* **2021**, *1–22*. [[CrossRef](#)]
99. Kwak, S.G.; Kim, J.H. Central limit theorem: The cornerstone of modern statistics. *Korean J. Anesthesiol.* **2017**, *70*, 144. [[CrossRef](#)]
100. Aswi, A.; Cramb, S.; Duncan, E.; Hu, W.; White, G.; Mengersen, K. Bayesian spatial survival models for hospitalisation of Dengue: A case study of Wahidin hospital in Makassar, Indonesia. *Int. J. Environ. Res. Public Health* **2020**, *17*, 878. [[CrossRef](#)]
101. Palacios, M.B.; Steel, M.F.J. Non-gaussian bayesian geostatistical modeling. *J. Am. Stat. Assoc.* **2006**, *101*, 604–618. [[CrossRef](#)]
102. Kibria, B.G.; Sun, L.; Zidek, J.V.; Le, N.D. Bayesian spatial prediction of random space-time fields with application to mapping PM<sub>2.5</sub> exposure. *J. Am. Stat. Assoc.* **2002**, *97*, 112–124. [[CrossRef](#)]
103. Fragoso, T.M.; Bertoli, W.; Louzada, F. Bayesian model averaging: A systematic review and conceptual classification. *Int. Stat. Review. Rev. Int. De Stat.* **2018**, *86*, 1–28. [[CrossRef](#)]
104. Sun, W.; Le, N.D.; Zidek, J.V.; Burnett, R. Assessment of a Bayesian multivariate interpolation approach for health impact studies. *Environmetr. Off. J. Int. Environmetr. Soc.* **1998**, *9*, 565–586. [[CrossRef](#)]
105. Neill, D.; Moore, A.; Cooper, G. A Bayesian spatial scan statistic. *Adv. Neural Inf. Process. Syst.* **2005**, *18*, 1003–1010.
106. Morris, T.P.; White, I.R.; Crowther, M.J. Using simulation studies to evaluate statistical methods. *Stat. Med.* **2019**, *38*, 2074–2102. [[CrossRef](#)] [[PubMed](#)]
107. Gelman, A.; Meng, X.L.; Stern, H. Posterior predictive assessment of model fitness via realized discrepancies. *Stat. Sin.* **1996**, *733–760*.
108. Akseer, N.; Bhatti, Z.; Mashal, T.; Soofi, S.; Moineddin, R.; Black, R.E.; Bhutta, Z.A. Geospatial inequalities and determinants of nutritional status among women and children in Afghanistan: An observational study. *Lancet Glob. Health* **2018**, *6*, e447–e459. [[CrossRef](#)]
109. Gelfand, A.E.; Diggle, P.; Guttorp, P.; Fuentes, M. *Handbook of Spatial Statistics*; CRC Press: Boca Raton, FL, USA, 2010.
110. Taniguchi, A.; Hagiwara, Y.; Taniguchi, T.; Inamura, T. Online spatial concept and lexical acquisition with simultaneous localization and mapping. In Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vancouver, BC, Canada, 24–28 September 2017; pp. 811–818.
111. Song, Y.; Nathoo, F.; Babul, A. A Potts-mixture spatiotemporal joint model for combined magnetoencephalography and electroencephalography data. *Can. J. Stat.* **2019**, *47*, 688–711. [[CrossRef](#)]
112. Taschler, B.; Ge, T.; Bendfeldt, K.; Müller-Lenke, N.; Johnson, T.D.; Nichols, T.E. Spatial modeling of multiple sclerosis for disease subtype prediction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 797–804.
113. Hachicha, W.; Ghorbel, A. A survey of control-chart pattern-recognition literature (1991–2010) based on a new conceptual classification scheme. *Comput. Ind. Eng.* **2012**, *63*, 204–222. [[CrossRef](#)]
114. Rue, H.v.; Held, L. *Gaussian Markov Random Fields: Theory and Applications*; Chapman and Hall/CRC: London, UK, 2005.
115. Freni-Sterrantino, A.; Ventrucci, M.; Rue, H. A note on intrinsic conditional autoregressive models for disconnected graphs. *Spat. Spatio-Temporal Epidemiol.* **2018**, *26*, 25–34. [[CrossRef](#)]
116. Lee, D. A comparison of conditional autoregressive models used in Bayesian disease mapping. *Spat. Spatio-Temporal Epidemiol.* **2011**, *2*, 79–89. [[CrossRef](#)]
117. Cramb, S.; Duncan, E.; Baade, P.; Mengersen, K. *Investigation of Bayesian Spatial Models*; Cancer Council Queensland and Queensland University of Technology (QUT): Brisbane City, Australia, 2017.
118. Lee, D.; Mitchell, R. Locally adaptive spatial smoothing using conditional auto-regressive models. *J. R. Stat. Soc. Ser. C Appl. Stat.* **2013**, *62*, 593–608. [[CrossRef](#)]
119. Lee, D.; Rushworth, A.; Sahu, S.K. A Bayesian localized conditional autoregressive model for estimating the health effects of air pollution. *Biometrics* **2014**, *70*, 419–429. [[CrossRef](#)]

120. Griffith, D.A. Selected challenges from spatial statistics for spatial econometricians. *Comp. Econ. Res.* **2012**, *15*, 71–85. [[CrossRef](#)]
121. Sun, X.; He, Z.; Kabrick, J. Bayesian spatial prediction of the site index in the study of the Missouri Ozark Forest Ecosystem Project. *Comput. Stat. Data Anal.* **2008**, *52*, 3749–3764. [[CrossRef](#)]
122. Karimi, O.; Mohammadzadeh, M. Bayesian spatial prediction for discrete closed skew Gaussian random field. *Math. Geosci.* **2011**, *43*, 565–582. [[CrossRef](#)]
123. Rivaz, F. Optimal network design for Bayesian spatial prediction of multivariate non-Gaussian environmental data. *J. Appl. Stat.* **2016**, *43*, 1335–1348. [[CrossRef](#)]
124. Lum, K.; Gelfand, A.E. Spatial quantile multiple regression using the asymmetric Laplace process. *Bayesian Anal.* **2012**, *7*, 235–258. [[CrossRef](#)]
125. Hu, G.; Bradley, J. A Bayesian spatial-temporal model with latent multivariate log-gamma random effects with application to earthquake magnitudes. *Stat* **2018**, *7*, e179. [[CrossRef](#)]
126. Yang, H.C.; Hu, G.; Chen, M.H. Bayesian Variable Selection for Pareto Regression Models with Latent Multivariate Log Gamma Process with Applications to Earthquake Magnitudes. *Geosciences* **2019**, *9*, 169. [[CrossRef](#)] [[PubMed](#)]
127. Schemmer, R.C.; Uribe-Opazo, M.A.; Galea, M.; Assumpção, R.A.B. Spatial variability of soybean yield through a reparameterized t-student model. *Eng. Agrícola* **2007**, *37*, 760–770. [[CrossRef](#)]
128. Best, N.; Richardson, S.; Thomson, A. A comparison of Bayesian spatial models for disease mapping. *Stat. Methods Med Res.* **2005**, *14*, 35–59. [[CrossRef](#)] [[PubMed](#)]
129. Li, Q.; Wang, X.; Liang, F.; Yi, F.; Xie, Y.; Gazdar, A.; Xiao, G. A Bayesian hidden Potts mixture model for analyzing lung cancer pathology images. *Biostatistics* **2019**, *20*, 565–581. [[CrossRef](#)]
130. Ferguson, T.S. A Bayesian analysis of some nonparametric problems. *Ann. Stat.* **1973**, *1*, 209–230. [[CrossRef](#)]
131. Antoniak, C.E. Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Ann. Stat.* **1974**, *2*, 1152–1174. [[CrossRef](#)]
132. MacEachern, S.N. *Dependent Dirichlet Process*; Technical Report; Ohio State University, Dept of Statistics: Columbus, OH, USA, 2000.
133. Dey, D.K.; Ghosh, S.K.; Mallick, B.K. *Generalized Linear Models: A Bayesian Perspective*; CRC Press: Boca Raton, FL, USA, 2000.
134. Fahrmeir, L.; Tutz, G. *Multivariate Statistical Modelling Based on Generalized Linear Models*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013.
135. Dasgupta, P.; Cramb, S.M.; Aitken, J.F.; Turrell, G.; Baade, P.D. Comparing multilevel and Bayesian spatial random effects survival models to assess geographical inequalities in colorectal cancer survival: A case study. *Int. J. Health Geogr.* **2014**, *13*, 36. [[CrossRef](#)] [[PubMed](#)]
136. Onicescu, G.; Lawson, A.B. Bayesian cure-rate survival model with spatially structured censoring. *Spat. Stat.* **2018**, *28*, 352–364. [[CrossRef](#)] [[PubMed](#)]
137. Jensen, C.D.; Lacombe, D.J.; McIntyre, S.G. A Bayesian spatial econometric analysis of the 2010 UK General Election. *Pap. Reg. Sci.* **2013**, *92*, 651–666. [[CrossRef](#)]
138. LeSage, J.P. *An Introduction to Spatial Econometrics*; Chapman and Hall/CRC: London, UK, 2008.
139. Neill, D.B.; Moore, A.W.; Cooper, G.F. A Bayesian spatial scan statistic. In *Advances in Neural Information Processing Systems*; MIT Press: Vancouver, BC, Canada, 2006; pp. 1003–1010.
140. Pilz, J.; Kazianka, H.; Spöck, G. Some advances in Bayesian spatial prediction and sampling design. *Spat. Stat.* **2012**, *1*, 65–81. [[CrossRef](#)]
141. Porto, E.D.; Revelli, F. Tax-limited reaction functions. *J. Appl. Econom.* **2013**, *28*, 823–839. [[CrossRef](#)]
142. Yu, J.; De Jong, R.; Lee, L.f. Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. *J. Econom.* **2008**, *146*, 118–134. [[CrossRef](#)]
143. Su, L.; Yang, Z. QML estimation of dynamic panel data models with spatial errors. *J. Econom.* **2015**, *185*, 230–258. [[CrossRef](#)]
144. Dempster, A.P.; Laird, N.M.; Rubin, D.B. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B (Methodol.)* **1977**, *39*, 1–22.
145. Hastings, W.K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **1970**. [[CrossRef](#)]
146. Tierney, L.; Kadane, J.B. Accurate approximations for posterior moments and marginal densities. *J. Am. Stat. Assoc.* **1986**, *81*, 82–86. [[CrossRef](#)]
147. Blangiardo, M.; Cameletti, M.; Baio, G.; Rue, H. Spatial and spatiotemporal models with R-INLA. *Spat. Spatio-Temporal Epidemiol.* **2013**, *4*, 33–49. [[CrossRef](#)] [[PubMed](#)]