Additive Gaussian processes for interpretable ANOVA of [planar spatial structures]

Gabriel Riutort-Mayol*
Foundation for the Promotion of Health and Biomedical Research in the Valencian Community (FISABIO), Spain
Department of Mathematics, Autonomous University of Madrid, Spain
Yandira Cuvero
Department of Mathematics, University Jaume I, Spain;
Department of Mathematics, National Polytechnic School, Ecuador
Juan Ródenas-Gómez
Department of Mathematics, University of Valencia, Spain;
and
Jorge Mateu†
Department of Mathematics, University Jaume I, Spain

April 12, 2025

Abstract

Factorial experiments when observations are spatial point patterns are yet under developed, contrasting with the wealth of spatial analysis of continuous data. Motivated by a real problem of locations of bubbles in a mineral flotation experiment, where the interest is in analyzing if the spatial distribution might be affected by frother concentrations and volumetric airflow rates, we develop an approach for statistical testing of two-way factorial experiments for spatial point patterns. We consider additive Gaussian processes with categorical variables to implement the main and interaction Gaussian process-based effects of the factors. This strategy is used to model the underlying intensity of [~~zero-inflated~~] Poisson counts representing the [marked] spatial point pattern of bubbles. Inference is developed under the Bayesian framework, and we present the results testing for main and interaction effects of the factors over the spatial distribution of bubbles.

Keywords: Additive Gaussian processes, Bayesian inference, Bubbles, Categorical covariance functions, Mineral flotation experiments

*Gabriel Riutort-Mayol has been funded by grant CD21/00186—Sara Borrell Postdoctoral Fellowship (from Instituto de Salud Carlos III, Spain) and co-funded by the European Union.

†Jorge Mateu has been partially funded by funds CIAICO/2022/191 (from Generalitat Valenciana) and PID2022-141555OB-I00 (from Ministry of Science).

1 Introduction

Spatial statistics is a large field of statistics focused on the study of events that evolve in space (perhaps also in space-time) with a latent spatial correlation. From its formalization in the sixties, it has lived a significant expansion in the last decades embracing interdisciplinary collaboration from other fields. In this context, a spatial point process is a stochastic process that models and governs the random occurrence of points in space. In the eighties, point pattern analysis gained prominence, focusing on the spatial arrangement of points. One key element to describe a point pattern is its intensity function, which characterizes the average rate of events per unit area.

[In particular, Cox processes are widely used in statistics and stochastic processes because they can naturally model cluster events. They were first defined in the fifties (Cox, 1955), and soon gained importance being widely useful in time series, environmental science, ecology, and spatial epidemiology, to name just some fields of application. A log-Gaussian Cox process (LGCP) is a particular case of a Cox point process, where the logarithm of the intensity function is modeled through a Gaussian process (GP), or more generally, a Gaussian random field. The reader is referred to (Diggle et al., 2013) as a general modern reference, but many other references to particular scientific fields and problems are acknowledged, (see as examples Lindgren et al. (2011), Sørbye and Rue (2014), Datta et al. (2016)). As pointed out, to work with an LGCP, it is crucial to model the corresponding Gaussian random field, which is a random field where each point in the field is a random variable following a Gaussian distribution.

While Gaussian fields were defined as early as the mid-20th century, an important development came with the growth of spatial statistics in the eighties (Matheron, 1971), and the introduction of GPs in machine learning (Rasmussen and Williams, 2006). However, modeling a Gaussian random field can be cumbersome and overall challenging, while handling GPs can alleviate some of these problems (Cressie and Wikle, 2015).]

GPs are indeed powerful tools for modeling complex relationships, and can be completely determined by their underlying covariance functions (here we will be using the terms *covariance function* or *kernel function*, interchangeably). Since the process could be considered as centered, the determination of the covariance function is crucial, because it determines the smoothness and characteristics of the process (Wilson and Adams, 2013). Towards obtaining more flexibility, additive Gaussian processes (AGP) can be used, which are formed by the sum of multiple single GPs or composed GPs as the product of different GPs, and capture different aspects of the data (Duvenaud et al., 2011; Durrande et al., 2012). In this case, each kernel component has some associated hyper-parameters, which can be easily estimated using Bayesian inference. Bayesian methods are preferred to other techniques, because they give full probability distribution estimates and thus can quantify the uncertainty associated with all involved parameters (Gelman et al., 2013).

We are motivated by a real problem on mineral flotation (González et al., 2021), which provides locations of bubbles in a mineral flotation experiment, where the interest is in analyzing if the spatial distribution might be affected by frother concentrations and volumetric airflow rates. Indeed, the data set consists of 54 images containing a total of 8385 floating bubbles. The images of bubbles can be regarded as spatial point patterns where the centroids of the bubbles correspond to the points. In addition, we have three frother concentration levels (5 ppm, 10 ppm, 15 ppm) as well as three volumetric airflow rate levels (5 l/min, 8 l/min, 10 l/min), and we have six replicates of point patterns at each combination of levels of such factors.

We thus can approach this problem by adapting a two-way design ANOVA to the case when observations are point patterns to test for influence of the different factors and their interactions. González et al. (2021) looked at this problem using second-order K-functions, assuming homogeneity, a convenient simplifying assumption but quite far from data reality. We instead place ourselves in the framework of GPs, developing a two-factor ANOVA

GP model (indeed an AGP) to decompose the overall variability in the response variable into additive GP components for single and interaction effects between [factors (Shi and Choi, 2011; Kaufman and Sain, 2010). In addition to the functional ANOVA approach proposed by (Kaufman and Sain, 2010), the broader literature on functional data analysis provides extensive discussion on linear models and ANOVA for functional data. Notably, (Ramsay and Silverman, 2005) offer a comprehensive treatment of these methods, particularly emphasizing spline-based and parametric approaches. Furthermore, significant contributions have been made in the context of functional outputs from computer experiments, as highlighted by Higdon et al. (2008). These works provide valuable theoretical and methodological insights that complement our approach and contextualize our modeling framework within the broader landscape of functional data analysis.] In following this AGP strategy, while being close in nature to the framework of LGCP, we are avoiding working with complicated covariance structures necessary for the Gaussian random field that defines the intensity of the point process, while instead encoding the covariance structure into GP kernel functions of the Matérn class of covariance functions (Rasmussen and Williams, 2006) in a fully probabilistic (Bayesian) framework. In addition, using Bayesian inference we can handle more easily and naturally the uncertainty and credibility of our estimated parameters, as well as provide a more flexible statistical analysis (Gelman et al., 2013).

The remainder of the paper is structured as follows. In Section 2 we present necessary formalities for GPs, and covariance functions, both continuous and categorical covariance functions. Then Section 3 developes a two-way ANOVA GP model. Section 4 presents a complete analysis of the mineral flotation problem. The paper ends with some final discussion.

2 Gaussian processes

A GP is a stochastic process which defines the distribution over a collection of random variables $\{f(\boldsymbol{x}) : \boldsymbol{x} \in \mathcal{X}\}$ with $f : \mathbb{R}^D \to \mathbb{R}$ a function in some input domain $\mathcal{X} \subset \mathbb{R}^D$. For generality, we will be using a general D-dimensional space, noting that when using spatial point patterns $D = 2$. [One of the principal advantages of this model are the properties that the GP can directly provide.]. GPs have the defining property that the marginal distribution of any finite subset of random variables, $\{f(\boldsymbol{x}_1), f(\boldsymbol{x}_2), \ldots, f(\boldsymbol{x}_n)\}$, follows a multivariate Gaussian distribution (Neal, 1997; Rasmussen and Williams, 2006). In this paper, GPs will take the role of a prior distribution over function spaces for non-parametric latent functions in a Bayesian setting.

Let us assume a GP prior for $f \sim \mathsf{GP}(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}'))$, where $m : \mathbb{R}^D \to \mathbb{R}$ is the mean and $k : \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}$ is a positive [~~semi~~] definite covariance function that defines the covariance between any two realizations of f, $f(\boldsymbol{x})$ and $f(\boldsymbol{x}')$; thus

$$
\begin{aligned}
m(\boldsymbol{x}) &= \mathbb{E}\left(f(\boldsymbol{x})\right), \\
k(\boldsymbol{x}, \boldsymbol{x}') &= \operatorname{cov}\left(f(\boldsymbol{x}), f(\boldsymbol{x}')\right) = \mathbb{E}\left(\left(f(\boldsymbol{x}) - m(\boldsymbol{x})\right)\left(f(\boldsymbol{x}') - m(\boldsymbol{x}')\right)\right).
\end{aligned}
$$

The mean and covariance functions completely characterize the GP prior, and control the *a priori* behavior of the function f. Consider $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ data inputs, with $\boldsymbol{x}_i \in \mathbb{R}^D$, and let $\boldsymbol{f} = (f(\boldsymbol{x}_1), f(\boldsymbol{x}_2), \ldots, f(\boldsymbol{x}_n))^\intercal$, then the resulting prior distribution for $\boldsymbol{f}$ is a multivariate Gaussian distribution $\boldsymbol{f} \sim \mathcal{N}(\boldsymbol{m}, \boldsymbol{K})$, where $\boldsymbol{m} = (m(\boldsymbol{x}_1), m(\boldsymbol{x}_2), \ldots, m(\boldsymbol{x}_n))^\intercal$ is the mean and $\boldsymbol{K}$ the covariance matrix, where $K_{i,j} = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$, $i, j \in \{1, \ldots, n\}$. [We define a log-Gaussian Cox processes (LGCP) which is a point process with intensity function $\lambda(\cdot)$ and a counting function N that verifies:

$$
\begin{aligned}
N(B)|\Lambda &\sim \mathsf{Poisson}\left(\int_B \Lambda(\boldsymbol{x}) dx\right), && \forall B \subseteq \mathcal{X} \\
\Lambda(\boldsymbol{x}) &= \exp\left(f(\boldsymbol{x})\right), && \forall \boldsymbol{x} \in \mathcal{X}.
\end{aligned}
$$

For more detail on marked point process please refer to (Gavrikov and Stoyan, 1995), and for in general for point process consider (Baddeley et al., 2006; Van Lieshout, 2019)]

The mean of the GP can be any arbitrary function of $\boldsymbol{x}$, but it is often assumed to be centered, i.e., $m(\boldsymbol{x}) = 0$, and the covariance function might depend on a set of hyper-parameters θ, i.e. $k(\boldsymbol{x}, \boldsymbol{x}'|\theta)$. The joint distribution of $\boldsymbol{f}$ and a new $\hat{f} = f(\hat{\boldsymbol{x}})$ evaluated at a new input $\hat{\boldsymbol{x}}$ is also a multivariate Gaussian, as we have

$$p(\boldsymbol{f}, \hat{f}) = \mathcal{N}\left(\begin{bmatrix} \boldsymbol{f} \\ \hat{f} \end{bmatrix} \middle| 0, \begin{bmatrix} K_{\boldsymbol{f},\boldsymbol{f}} & \boldsymbol{k}_{\boldsymbol{f},\hat{f}} \\ \boldsymbol{k}_{\hat{f},\boldsymbol{f}} & k_{\hat{f},\hat{f}} \end{bmatrix} \right),$$

where $\boldsymbol{k}_{\boldsymbol{f},\hat{f}}$ is the covariance between $\boldsymbol{f}$ and $\hat{f}$, and $k_{\hat{f},\hat{f}}$ is the prior variance of $\hat{f}$.

The covariance function is a crucial ingredient in a GP as it encodes our prior assumptions about the variation of the function, and defines a correlation structure which characterizes the correlations between function values at different inputs. A covariance function can be expressed as a positive[~~semi-~~]definite kernel function, such that the Gram matrix corresponding to the covariance function is positive [~~semi-~~]definite (Rasmussen and Williams, 2006). Here, the terms *covariance function* and *kernel* might be used interchangeably. Readers are referred to Duvenaud (2014) and Rasmussen and Williams (2006) for a deeper description of covariance functions.

Continuous covariance functions

Covariance functions for continuous input variables can be stationary or non-stationary, depending on whether the covariance is invariant or not to translations. A stationary covariance function is a function of $\boldsymbol{\tau} = \boldsymbol{x} - \boldsymbol{x}' \in \mathbb{R}^D$, such that we can write $k(\boldsymbol{x}, \boldsymbol{x}') = k(\boldsymbol{\tau})$. Isotropic covariance functions are those that are a function only of the distance between observations, $k(\boldsymbol{x}, \boldsymbol{x}') = k(|\boldsymbol{x} - \boldsymbol{x}'|) = k(r)$, $r \in \mathbb{R}^+$, which means that the covariance is both translation and rotation invariant. Dot-product (e.g., linear covariance function) and neural network covariance functions are examples of non-stationary covariance functions [,

for more detail, please refer to Paciorek and Schervish (2003); Schabenberger and Gotway (2017)].

The most commonly used class of isotropic covariance functions in practical applications is the Matérn class of covariance functions, which depend on the order ν of the kernel (Rasmussen and Williams, 2006). The particular case where $\nu = \infty$ and $\nu = 3/2$ are probably the most commonly used kernels, which are given by

$$k_\infty(\boldsymbol{\tau}) = \sigma \exp\left(-\frac{1}{2}\frac{\boldsymbol{\tau}^2}{\ell^2}\right),$$
$$k_{\frac{3}{2}}(\boldsymbol{\tau}) = \sigma\left(1 + \frac{\sqrt{3}\boldsymbol{\tau}}{\ell}\right)\exp\left(-\frac{\sqrt{3}\boldsymbol{\tau}}{\ell}\right),$$

where $\ell > 0$ is known as the length-scale of the kernel and controls the degree of decay of correlation between inputs, and that ultimately determines the smoothness or non-linear effects of the function f. The marginal variance of the kernel is $\sigma > 0$, that determines the magnitude of the variance of the function f. The length-scale ℓ and magnitude σ are the hyper-parameters θ of the kernel. The kernel k_∞ is commonly known as squared exponential (exponentiated quadratic) covariance function, and the kernel $k_{\frac{3}{2}}$ as Matérn covariance function. Assuming the Euclidean distance between observations, $r = |\boldsymbol{x} - \boldsymbol{x}'|_{L_2} = \sqrt{\sum_{i=1}^{D}(x_i - x_i')^2}$, the kernels written above take the form

$$k_\infty(|\boldsymbol{x} - \boldsymbol{x}'|_{L_2}) = \sigma \exp\left(-\frac{1}{2}\sum_{i=1}^{D}\frac{(x_i - x_i')^2}{\ell_i^2}\right),$$
$$k_{\frac{3}{2}}(|\boldsymbol{x} - \boldsymbol{x}'|_{L_2}) = \sigma\left(1 + \sqrt{\sum_{i=1}^{D}\frac{3(x_i - x_i')^2}{\ell_i^2}}\right)\exp\left(-\sqrt{\sum_{i=1}^{D}\frac{3(x_i - x_i')^2}{\ell_i^2}}\right).$$

Notice that the previous expressions have been easily generalized to using a multidimensional length-scale $\boldsymbol{\ell} \in \mathbb{R}^D$. The use of a multidimensional length-scale basically turns the isotropic covariance function into non-isotropic.

Categorical covariance functions

Categorical covariance functions are commonly constructed to encode the effects of categorical variables on a response variable as static category-specific offset effects (Timonen et al., 2021). Given $c, c' \in \mathcal{C}$, where $\mathcal{C} = \{1, 2, \ldots, M_{\mathcal{C}}\}$ denotes a finite set of categorical levels to which the data are related, a categorical kernel takes the form

$$k_{cat}(c,\, c') = \begin{cases} 1, & \text{if } c = c', \\ 0, & \text{if } c \neq c'. \end{cases} \tag{1}$$

Similarly, given $d, d' \in \mathcal{D}$, where $\mathcal{D} = \{1, 2, \ldots, M_{\mathcal{D}}\}$ denotes a finite set of categorical levels of another categorical variable, a kernel can be constructed to model the interaction between two categorical variables by encoding the effects of combined categories in the co-interaction variables

$$k_{cat}((c, d), (c', d')) = \begin{cases} 1, & \text{if } c = c' \;\&\; d = d', \\ 0, & \text{if } c \neq c' \text{ or } d \neq d'. \end{cases} \tag{2}$$

A particular case of a categorical kernel is the binary (mask) kernel which only considers the effects of one single category of a variable

$$k_{bin}(c,\, c' \,|\, \varsigma) = \begin{cases} 1, & \text{if } c = c' = \varsigma, \\ 0, & \text{if } c = c' \neq \varsigma \text{ or } c \neq c', \end{cases}$$

with $\varsigma \in \mathcal{C}$ one of the M categories of the variable.

To facilitate interpretation by separating the effect of categorical variables from the overall mean, the zero sum of the effects of different categories of a variable can be implemented in a sum-to-zero categorical kernel

$$k_{cat}(c,\, c') = \begin{cases} 1, & \text{if } c = c', \\ \dfrac{1}{1 - M}, & \text{if } c \neq c'. \end{cases} \tag{3}$$

A sum-to-zero kernel that models the interaction between two categorical variables by encoding the effects of combined categories in the co-interaction variables can be specified

as follows

$$k_{cat}((c,d),(c',d')) = \begin{cases} 1, & \text{if } c = c' \;\&\; d = d', \\ \dfrac{1}{1 - M_{\mathcal{C}} M_{\mathcal{D}}}, & \text{if } c \neq c' \text{ or } d \neq d'. \end{cases} \tag{4}$$

Alternative sum-to-zero categorical kernel constructions can be found in Kaufman and Sain (2010).

When dealing with complex data sets, more than two of these categorical factors may arise. Nonetheless, for the remaining of the paper, the number of categorical factors will be taken as two. This means that a group structure of two crossed factors, $\mathcal{C}$ and $\mathcal{D}$, with their respective $M_{\mathcal{C}}$ and $M_{\mathcal{D}}$ categorical levels, is taken into account, which is a frequent case in practice.

Those categorical kernels introduced above do not model covariance among categories of the categorical variable, that is, category-specific effects are independent of each other. Nevertheless, they can be easily generalized to take into account correlations between categories by defining the categorical kernel as follows

$$k_{cat}(c,\, c') = \begin{cases} 1, & \text{if } c = c', \\ \rho(c,c'), & \text{if } c \neq c', \end{cases} \tag{5}$$

where $\rho(\cdot)$ is a function $\rho : \mathcal{C} \times \mathcal{C} \to \mathbb{R}$ such that $k_{cat}(c,c')$ is a positive definite function in the domain $\mathcal{C} \times \mathcal{C}$ that generates a positive definite correlation matrix $\Theta \in \mathbb{R}^{M_{\mathcal{C}} \times M_{\mathcal{C}}}$ between categories $\mathcal{C}$. As long as matrix Θ is positive definite, k_{cat} is a valid categorical kernel that models prior correlations between categories. This can be easily generalized to the case of two crossed factors, $\mathcal{C}$ and $\mathcal{D}$, as follows

$$k_{cat}((c,d),(c',d')) = \begin{cases} 1, & \text{if } c = c' \;\&\; d = d', \\ \rho((c,d),(c',d')), & \text{if } c \neq c' \text{ or } d \neq d', \end{cases} \tag{6}$$

where $c, c' \in \mathcal{C}$ and $d, d' \in \mathcal{D}$. They can also be defined to fulfill the zero-sum of the effects of different categories of a variable or of the co-interaction of two variables by imposing the

following constraints

$$1 + \sum_{c' \neq c}^{M_{\mathcal{C}}} \rho(c, c') = 0, \quad \forall c \in \mathcal{C} \tag{7}$$

and

$$1 + \sum_{(c',d') \neq (c,d)}^{M_{\mathcal{C}} \cdot M_{\mathcal{D}}} \rho((c,d),(c',d')) = 0, \quad \forall (c,d) \in \mathcal{C} \times \mathcal{D}, \tag{8}$$

to the categorical kernels in (5) and (6), respectively.

Considering correlations between categories can be useful to improve modeling performance in the cases of unbalanced groups/categories and/or with few observations in some of the groups. Another novel and powerful approach to deal with covariances between groups has been recently proposed by Li et al. (2021), who theoretically and mathematically extended the Matérn class of covariance functions to be defined in the mixed domain of continuous $\mathbb{R}^p$ and categorical $\mathcal{C}$ variables ($\mathbb{R}^p \times \mathcal{C}$). These new mixed-domain kernels proposed by Li et al. (2021) can also be useful for discovering similarities/correlations between groups; however further work is needed to analyze its performance in practice.

Just to mention a traditional way to considering covariances among category-specific effects is to convert the variable with M categorical levels into a M-dimensional set of dummy variables by means of *one-hot encoding*, and place these multivariate set of dummy variables in a continuous kernel (Duvenaud, 2014). Short length-scales on any particular dimension of the continuous kernel indicate that the effects corresponding to that category are uncorrelated with the others, and vice versa. More flexible parameterizations are also possible (Pinheiro and Bates, 1996).

Product kernel of a continuous and categorical kernel

Effects of categorical variables can be modeled as category-specific deviations from the continuous kernel effects by the interaction of a continuous kernel with a categorical kernel (i.e., product of both kernels) (Rasmussen and Williams, 2006). For example, the product

kernel between a squared exponential kernel of $\boldsymbol{x}$ and a categorical kernel of $c \in \mathcal{C}$,

$$k_{\infty \times cat}((\boldsymbol{x}, c), (\boldsymbol{x}', c')) = k_{\infty}(\boldsymbol{x}, \boldsymbol{x}') \cdot k_{cat}(c, c'), \tag{9}$$

will produce deviations in the effects of the continuous kernel of $\boldsymbol{x}$ as a function of the categories of $\mathcal{C}$. Product kernels can be built between any set of valid kernels, either continuous or categorical.

2.1 Additive Gaussian processes

GPs can be made more flexible and interpretable by making them additive with q additive kernels, such as

$$f(\boldsymbol{x}) = f^{(1)}(\boldsymbol{x}^1) + f^{(2)}(\boldsymbol{x}^2) + \ldots + f^{(q)}(\boldsymbol{x}^q),$$

where $\boldsymbol{x} = (\boldsymbol{x}^1 \cup \boldsymbol{x}^2 \cup \ldots \cup \boldsymbol{x}^q) \in \mathbb{R}^D$, and each $f^{(j)}(\boldsymbol{x}^j) \sim \mathsf{GP}(0, k^{(j)}(\boldsymbol{x}^j, \boldsymbol{x}'^j|\theta^{(j)}))$, $j \in \{1, \ldots, q\}$, is a separate GP with input values $\boldsymbol{x}^j \in \mathbb{R}^{D^j}$, with D^j the dimension of vector of input values $\boldsymbol{x}^j$, and specific hyper-parameters $\theta^{(j)}$ for the kernel $k^{(j)}(\boldsymbol{x}^j, \boldsymbol{x}'^j|\theta^{(j)})$ (Duvenaud et al., 2011; Durrande et al., 2012; Timonen et al., 2021). Assuming the components $f^{(j)}$ are *a priori* independent, by the properties of the multivariate normal distribution, the sum f is also a zero-mean Gaussian process with kernel

$$k(\boldsymbol{x}, \boldsymbol{x}'|\theta) = \sum_{j=1}^{q} k^{(j)}(\boldsymbol{x}^j, \boldsymbol{x}'^j|\theta^{(j)})),$$

where $\theta = \left(\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(q)}\right)^{\intercal}$ is the vector with the hyper-parameters of every single kernel. In order to cover most common modeling needs and properly set up an additive Gaussian process (AGP) model, it is essential to choose appropriate kernels or combination of kernels (e.g., product kernel of categorical and continuous variables) for every single GP function.

3 Functional ANOVA model with Gaussian process functions

Consider two crossed categorical factors, $\mathcal{C} = \{1, 2, \ldots, M_{\mathcal{C}}\}$ and $\mathcal{D} = \{1, 2, \ldots, M_{\mathcal{D}}\}$, with levels denoted by c and d, respectively. Consider a given data set $\{\boldsymbol{x}_i, c_i, d_i, r_i, y_i\}_{i=1}^n$, with level $c_i \in \mathcal{C}$, level $d_i \in \mathcal{D}$, and r_i, for observation i, and y_i is modeled conditionally as $p(y_i|f(\boldsymbol{x}_i, c_i, d_i), \phi)$. Here, p is some parametric distribution with parameters f and ϕ, and f is an unknown function with an AGP prior, which depends on a continuous input vector $\boldsymbol{x}_i \in \mathbb{R}^D$, and levels (c_i, d_i), at the i-th observation.

Applying functional two-way ANOVA modeling (Kaufman and Sain, 2010), the latent function f of functional responses y_i is splitted into main effects and interactions of the two crossed factors, as follows

$$f(\boldsymbol{x}_i, c_i, d_i) = \mu(\boldsymbol{x}_i) + \alpha(\boldsymbol{x}_i, c_i) + \beta(\boldsymbol{x}_i, d_i) + (\alpha\beta)(\boldsymbol{x}_i, c_i, d_i), \tag{10}$$

where $\mu(\boldsymbol{x}_i)$ is an overall mean function, $\alpha(\boldsymbol{x}_i, c_i)$ is the main effect function of level c_i, $\beta(\boldsymbol{x}_i, d_i)$ is the main effect function of level d_i, and $(\alpha\beta)(\boldsymbol{x}_i, c_i, d_i)$ is the interaction effect function of combined levels c_i and d_i, at observation i. As it is well-known in ANOVA models, we need to impose sum-to-zero constraints

$$\sum_{c_i \in \mathcal{C}} \alpha(\boldsymbol{x}_i, c_i) = 0,\ \sum_{d_i \in \mathcal{D}} \beta(\boldsymbol{x}_i, d_i) = 0,\ \sum_{c_i \in \mathcal{C}} \alpha\beta(\boldsymbol{x}_i, c_i, d_i) = 0,\ \sum_{d_i \in \mathcal{D}} \alpha\beta(\boldsymbol{x}_i, c_i, d_i) = 0,$$

or level-specific constraints,

$$\alpha(\boldsymbol{x}_i, \varsigma) = 0,\ \beta(\boldsymbol{x}_i, \delta) = 0,\ \alpha\beta(\boldsymbol{x}_i, \varsigma, d_i) = 0,\ \alpha\beta(\boldsymbol{x}_i, c_i, \delta) = 0, \tag{11}$$

where $\varsigma \in \mathcal{C}$ and $\delta \in \mathcal{D}$ are specific levels of factors $\mathcal{C}$ and $\mathcal{D}$, on functions α, β, and $(\alpha\beta)$ for model identifiability.

In GP ANOVA, function μ is considered as a GP with a continuous kernel, and factor functions α, β, and $(\alpha\beta)$ are considered as GPs with kernels that are interaction kernels of

continuous and categorical covariates (i.e., product kernel between a continuous kernel and a categorical kernel) as seen in eq. (9). Note that a GP ANOVA model fits the framework of the AGP model.

Thus, $\mu(\boldsymbol{x})$ is a GP with a continuous kernel function of $\boldsymbol{x}$ and hyper-parameters $\theta^{(\mu)}$

$$\mu(\boldsymbol{x}) \sim \mathsf{GP}(0, k^{(\mu)}(\boldsymbol{x}, \boldsymbol{x}'|\theta^{(\mu)})).$$

The functional factor $\alpha(\boldsymbol{x}, c)$ is a GP with a kernel function which is the product of a continuous kernel of $\boldsymbol{x}$ with parameters $\theta^{(\alpha)}$ and a categorical kernel of $c \in \mathcal{C}$

$$\alpha(\boldsymbol{x}, c) \sim \mathsf{GP}(0, k^{(\alpha)}(\boldsymbol{x}, \boldsymbol{x}'|\theta^{(\alpha)}) \cdot k_{cat}(c, c')).$$

Also, factor $\beta(\boldsymbol{x}, d)$ is a GP with a kernel function which is the product of a continuous kernel of $\boldsymbol{x}$ with parameters $\theta^{(\beta)}$ and a categorical kernel of $d \in \mathcal{D}$

$$\beta(\boldsymbol{x}, c) \sim \mathsf{GP}(0, k^{(\beta)}(\boldsymbol{x}, \boldsymbol{x}'|\theta^{(\beta)}) \cdot k_{cat}(d, d')),$$

Finally, the interaction factor $(\alpha\beta)(\boldsymbol{x}, c, d)$ is a GP with a kernel function which is the [Hadamard] product of a continuous kernel of $\boldsymbol{x}$ with parameters $\theta^{(\alpha\beta)}$ and a categorical kernel of the level combination $(c, d) \in \mathcal{C} \times \mathcal{D}$

$$(\alpha\beta)(\boldsymbol{x}, c, d) \sim \mathsf{GP}(0, k^{(\alpha\beta)}(\boldsymbol{x}, \boldsymbol{x}'|\theta^{(\alpha\beta)}) \cdot k_{cat}((c, d), (c', d')).$$

Using sum-to-zero categorical kernels as defined in equations (3) and (4), for uncorrelated category effects, or equations (5) and (6) with constraints in equations (7) and (8), for correlated category effects, constraints for GP ANOVA model identifiability are implicitly implemented. However, if using categorical kernels as defined in equations (1) and (2) or equations (5) and (6), instead, those have to be re-defined to encode a level-specific constraint, such as

$$k_{cat}(c, c') = \begin{cases} 1, & \text{if } c = c' \neq \varsigma, \\ 0, & \text{if } c = \varsigma \text{ or } c' = \varsigma, \\ \rho(c, c'), & \text{if } c \neq c' \ \& \ (c \neq \varsigma \text{ or } c' \neq \varsigma), \end{cases} \tag{12}$$

$$k_{cat}((c,d),(c',d')) = \begin{cases} 1, & \text{if } (c,d) = (c',d') \neq (\varsigma,\delta), \\ 0, & \text{if } (c = \varsigma \ \& \ d = \delta) \text{ or } (c' = \varsigma \ \& \ d' = \delta), \\ \rho((c,d),(c',d')), & c \neq c' \neq \varsigma \text{ or } d \neq d' \neq \delta, \end{cases} \tag{13}$$

with $\varsigma \in \mathcal{C}$ and $\delta \in \mathcal{D}$ one of the $M_{\mathcal{C}}$ and $M_{\mathcal{D}}$ levels of the categorical covariates $\mathcal{C}$ and $\mathcal{D}$, respectively, and where $\rho(c,c') = 0$ and $\rho((c,d),(c',d')) = 0$ if independence between group effects is considered, or $\rho(c,c') : \mathcal{C} \times \mathcal{C} \to \mathbb{R}$ and $\rho((c,d),(c',d')) : (\mathcal{C} \times \mathcal{C}) \times (\mathcal{D} \times \mathcal{D}) \to \mathbb{R}$ are positive definite covariance functions to handle dependence between groups, as commented in previous Section 2. It is straightforward to prove kernels in (12) and (13) generate positive definite correlation matrices as they are the result of multiplying kernels in (1) and (2) or in (5) and (6), which are indeed positive definite, by matrices

$$\begin{cases} 0, & \text{if } c = \varsigma \text{ or } c' = \varsigma, \\ 1, & \text{otherwise,} \end{cases} \tag{14}$$

and

$$\begin{cases} 0, & \text{if } (c = \varsigma \ \& \ d = \delta) \text{ or } (c' = \varsigma \ \& \ d' = \delta), \\ 1, & \text{otherwise,} \end{cases} \tag{15}$$

respectively, which are also positive definite matrices.

The AGP model structure of function f allows us to formulate the GP ANOVA model in a more compact way, as seen in Section 2.1,

$$\begin{aligned} f(\boldsymbol{x}) &\sim \mathsf{GP}(0, k^f(\boldsymbol{x}, \boldsymbol{x}'|\theta)), \\ k(\boldsymbol{x}, \boldsymbol{x}'|\theta) &= k^{(\mu)}(\boldsymbol{x}^{(\mu)}, \boldsymbol{x}'^{(\mu)}|\theta^{(\mu)}) \\ &\quad + k^{(\alpha)}(\boldsymbol{x}^{(\alpha)}, \boldsymbol{x}'^{(\alpha)}|\theta^{(\alpha)}) \cdot k_{cat}(c,c') \\ &\quad + k^{(\beta)}(\boldsymbol{x}^{(\beta)}, \boldsymbol{x}'^{(\beta)}|\theta^{(\beta)}) \cdot k_{cat}(d,d') \\ &\quad + k^{(\alpha\beta)}(\boldsymbol{x}^{(\alpha\beta)}, \boldsymbol{x}'^{(\alpha\beta)}|\theta^{(\alpha\beta)}) \cdot k_{cat}((c,d),(c',d')), \end{aligned} \tag{16}$$

where $\boldsymbol{x} = (\boldsymbol{x}^{(\mu)} \cup \boldsymbol{x}^{(\alpha)} \cup \boldsymbol{x}^{(\beta)} \cup \boldsymbol{x}^{(\alpha\beta)})$ and $\theta = \left(\theta^{(\mu)}, \theta^{(\alpha)}, \theta^{(\beta)}, \theta^{(\alpha\beta)}\right)^{\intercal}$.

3.1 Model inference

The likelihood function of the observations $\boldsymbol{y} = \{y_i\}_{i=1}^n$ given the latent function $f(\boldsymbol{x}_i, c_i, d_i) = \mu(\boldsymbol{x}_i) + \alpha(\boldsymbol{x}_i, c_i) + \beta(\boldsymbol{x}_i, d_i) + (\alpha\beta)(\boldsymbol{x}_i, c_i, d_i)$ and parameter ϕ is of the form

$$\begin{aligned} p(\boldsymbol{y}|\boldsymbol{f}, \phi) &= \prod_i p(y_i| \exp(f(\boldsymbol{x}_i, c_i, d_i)), \phi) \\ &= \prod_i p(y_i|\mu(\boldsymbol{x}_i), \alpha(\boldsymbol{x}_i, c_i), \beta(\boldsymbol{x}_i, d_i), (\alpha\beta)(\boldsymbol{x}_i, c_i, d_i), \phi), \end{aligned}$$

where p is some parametric distribution with parameters $f(\boldsymbol{x}_i, c_i, d_i)$ and ϕ. Note that for point patterns the likelihood function will be a Poisson distribution.

We follow Bayesian inference over the joint posterior distribution of parameters and hyper-parameters given the data, which is proportional to the likelihood and priors, assuming independent priors among the effects

$$\begin{aligned} p(\boldsymbol{f}, \phi|\boldsymbol{y}) &= p(\boldsymbol{y}|\boldsymbol{f}, \phi)p(\boldsymbol{f})p(\phi) = p(\boldsymbol{y}|\boldsymbol{f}, \phi)p(\boldsymbol{\mu})p(\boldsymbol{\alpha})p(\boldsymbol{\beta})p(\boldsymbol{\alpha\beta})p(\phi) \\ &= \prod_i p\big(y_i|\mu(\boldsymbol{x}_i), \alpha(\boldsymbol{x}_i, c_i), \beta(\boldsymbol{x}_i, d_i), (\alpha\beta)(\boldsymbol{x}_i, c_i, d_i), \phi\big) \times \prod_i p\big(\mu(\boldsymbol{x}_i, c_i)|\theta^{(\mu)}\big) \\ &\quad \times \prod_i p\big(\alpha(\boldsymbol{x}_i, c_i)|\theta^{(\alpha)}\big) \times \prod_i p\big(\beta(\boldsymbol{x}_i, d_i)|\theta^{(\beta)}\big) \times \prod_i p\big((\alpha\beta)(\boldsymbol{x}_i, c_i, d_i)|\theta^{(\alpha\beta)}\big) \\ &\quad \times p\big(\theta^{(\mu)}\big) \times p\big(\theta^{(\alpha)}\big) \times p\big(\theta^{(\beta)}\big) \times p\big(\theta^{(\alpha\beta)}\big) \times p\big(\phi\big) \end{aligned}$$

where $p(\boldsymbol{y}|\boldsymbol{f}, \phi) = \prod_i p\big(y_i|\mu(\boldsymbol{x}_i), \alpha(\boldsymbol{x}_i, c_i), \beta(\boldsymbol{x}_i, d_i), (\alpha\beta)(\boldsymbol{x}_i, c_i, d_i), \phi\big)$ is the likelihood function and $p(\boldsymbol{\mu}) = \prod_i p\big(\mu(\boldsymbol{x}_i)|\theta^{(\mu)}\big)$, $p(\boldsymbol{\alpha}) = \prod_i p\big(\alpha(\boldsymbol{x}_i, c_i)|\theta^{(\alpha)}\big)$, $p(\boldsymbol{\beta}) = \prod_i p\big(\beta(\boldsymbol{x}_i, d_i)|\theta^{(\beta)}\big)$, and $p\big((\boldsymbol{\alpha\beta})\big) = \prod_i p\big((\alpha\beta)(\boldsymbol{x}_i, c_i, d_i)|\theta^{(\alpha\beta)}\big)$ are the GP prior distributions for functional parameters $\mu(\boldsymbol{x}_i)$, $\alpha(\boldsymbol{x}_i, c_i)$, $\beta(\boldsymbol{x}_i, d_i)$ $(\alpha\beta)(\boldsymbol{x}_i, c_i, d_i)$, respectively, with hyper-prior distributions $p(\theta^{(\mu)})$, $p(\theta^{(\alpha)})$, $p(\theta^{(\beta)})$ and $p(\theta^{(\alpha\beta)})$ for hyper-parameters $\theta^{(\mu)}$, $\theta^{(\alpha)}$, $\theta^{(\beta)}$ and $\theta^{(\alpha\beta)}$, respectively. $p(\phi)$ is the hyper-prior distribution for hyper-parameter ϕ.

Marginal posterior probability distributions of parameters of interest, $p(\mu(\boldsymbol{x}_i, c_i))$, $p(\alpha(\boldsymbol{x}_i, c_i))$, $p(\beta(\boldsymbol{x}_i, d_i))$, and $p((\alpha\beta)(\boldsymbol{x}_i, c_i, d_i))$ are in general not available in closed form. Hence, to estimate them, simulation methods and distributional (Gelman et al., 2013) or numeri-

cal (Rue et al., 2009; Gómez-Rubio, 2020) approximation methods must be used. Simulation methods based on the Markov chain Monte Carlo (MCMC) (Brooks et al., 2011), and more recently, the Hamiltonian Monte Carlo (HMC) (Neal et al., 2011), are general sampling methods to obtain samples from the joint and marginal posterior distributions. Here, HMC methods are used to make inferences about the posterior distributions.

3.2 Between-group differences evaluation

Magnitude and uncertainty of the main effects functions $\alpha(\boldsymbol{x}_i, c_i)$ and $\beta(\boldsymbol{x}_i, d_i)$ and of the interaction effect function $(\alpha\beta)(\boldsymbol{x}_i, c_i, d_i)$ for all categories $c_i \in \mathcal{C}$, $d_i \in \mathcal{D}$, and crossed categories $(c_i, d_i) \in \mathcal{C} \times \mathcal{D}$, are given by their marginal posterior distributions $p(\alpha(\boldsymbol{x}_i, c_i))$, $p(\beta(\boldsymbol{x}_i, d_i))$, and $p((\alpha\beta)(\boldsymbol{x}_i, c_i, d_i))$, respectively. Credibility of $\alpha(\boldsymbol{x}_i, c_i)$, $\beta(\boldsymbol{x}_i, d_i)$, and $(\alpha\beta)(\boldsymbol{x}_i, c_i, d_i)$ can be determined by evaluating whether their accumulated probabilities $P(\cdot)$ of being less than or greater than zero are lower than a probability threshold η

$$\begin{cases} \eta > P\big(\alpha(\boldsymbol{x}_i, c_i) \leq 0\big) > 1 - \eta, & \forall c_i \in \mathcal{C}, \\ \eta > P\big(\beta(\boldsymbol{x}_i, c_i) \leq 0\big) > 1 - \eta, & \forall d_i \in \mathcal{D}, \\ \eta > P\big((\alpha\beta)(\boldsymbol{x}_i, c_i, d_i) \leq 0\big) > 1 - \eta, & \forall (c_i, d_i) \in \mathcal{C} \times \mathcal{D}, \end{cases} \tag{17}$$

where common values for η are used to assess credibility of parameters. To clarify the notation used, note that $\eta > P(\alpha \leq 0) > 1 - \eta$ is equivalent to saying $P(\alpha \leq 0) < \eta$ or $P(\alpha \leq 0) > 1 - \eta$, so that if $P(\alpha \leq 0) < \eta$ is satisfied, then α is credibly positive and when $P(\alpha \leq 0) > 1 - \eta$ then α es credibly negative.

Similarly, differences of model components $\alpha(\boldsymbol{x}_i, c_i)$, $\beta(\boldsymbol{x}_i, d_i)$, and $(\alpha\beta)(\boldsymbol{x}_i, c_i, d_i)$ among categories $c_i \in \mathcal{C}$, $d_i \in \mathcal{D}$, and $(c_i, d_i) \in \mathcal{C} \times \mathcal{D}$, respectively, can be determined and their credibility assessed through evaluating the difference of their marginal posterior distribu-

tions

$$\begin{cases} \eta > P\big(\alpha(\boldsymbol{x}_i, \varsigma) - \alpha(\boldsymbol{x}_i, \varsigma') \leq 0\big) > 1 - \eta, & \forall \varsigma, \varsigma' \in \mathcal{C}, \\ \eta > P\big(\beta(\boldsymbol{x}_i, \delta) - \beta(\boldsymbol{x}_i, \delta') \leq 0\big) > 1 - \eta, & \forall \delta, \delta' \in \mathcal{D}, \\ \eta > P\big((\alpha\beta)(\boldsymbol{x}_i, \varsigma, \delta) - (\alpha\beta)(\boldsymbol{x}_i, \varsigma', \delta') \leq 0\big) > 1 - \eta, & \forall (\varsigma, \delta), (\varsigma', \delta') \in \mathcal{C} \times \mathcal{D}. \end{cases} \tag{18}$$

4 Analysis of bubbles from a mineral flotation process

The copper production process consists of a large number of steps that allow obtaining cathodes from a mineral deposit. For sulphide ores, the processing requires steps of concentration by flotation and pyrometallurgy. A mineral flotation process allows separating particles from non-metallic gangue particles. The comminuted material is deposited in reactors (flotation cells) forming a pulp with water and some reagents. The recovered mineral particles attach to the bubbles and rise to the surface where they form a blanket of froth containing the mineral. See (González et al., 2021) for a more detailed explanation on bubbles formation.

Understanding the bubbles spatial distribution can be essential for a proper operation of flotation machines, and the recovery of metal particles. This distribution may depend on the variables that are controlled in the experiment. It is known that the flotation characteristics are strongly dependent on a variety of important operating and design factors. Indeed, the volumetric airflow rate (in L min^{-1}) and the specific frother concentration (in ppm) are two factors that particularly influence the physical properties. Therefore, we are interested in these two factors as they could be potentially influential in the spatial distribution of the bubbles.

High-resolution images (170 pixels/mm) were recorded with dimensions (our region of interest) $W = 29.0mm \times 19.2mm$. The volumetric airflow (VA) rate was measured at three levels, 5L min^{-1}, 8 L min^{-1} and 10 L min^{-1}, while frother concentration (FC) was analyzed at 5ppm, 10ppm and 15ppm. Additionally, we have six replicates at each combination of

levels of such factors/treatments. The treatment combinations of the experiment, as well as the observed bubble patterns, are depicted in Figure 1. It is evident that the bubbles can be considered realizations of spatial point patterns in W measured at the cross-combinations of levels of VA and FC. In this context, we are interested in analyzing if the spatial distribution might be affected by frother concentrations and volumetric airflow rates, and we use an AGP-based ANOVA to provide evidences on factor dependencies.

We first consider partitioning the observation window $W \subset \mathbb{R}^2$ into a regular set of cells $\{W_i\}_{i=1}^{100}$ of center $\boldsymbol{x}_i \in \mathbb{R}^2$ and assign the mean of the marks $z_{ij}^{cd} \in \mathbb{R}$ to each grid cell W_i and replicate j for the different point patterns measured at crossed categories of factors $c \in VA$ and $d \in FC$. In this context, the original point patterns transforms into a new random variable y_{ij}^{cd} that counts the number of bubbles on each cell W_i, replicate j and point pattern at levels c and d (see Figure 2), i.e., $y_{ij}^{cd} = \mathrm{N}(Y_j^{cd} \cap W_i)$ where Y_j^{cd} is the point pattern at cross-category (c, d) and replicate j, with $i = 1, \ldots, s$, where s denotes the number of grid cells, and $j = 1, \ldots, r$, where r denotes the number of replicates.

The stochastic model underlying the number of events y_{ij}^{cd} is considered to be a [Poisson model, as follows

$$p(y_{ij}^{cd} \mid \boldsymbol{x}_i, z_{ij}^{cd}, c, d, \tau_j^{cd}, \theta^{cd}) = \mathsf{Poisson}\big(y_{ij}^{cd} \mid \mathsf{exp}(f(\boldsymbol{x}_i, z_{ij}^{cd}, c, d) + \tau_j^{cd})\big)$$

with $i = 1, \ldots, s$ and $j = 1, \ldots, r$. Parameter τ_j^{cd} is a random effect associated to replicate j of the point pattern at levels c and d. The Poisson distribution with mean parameter $\mathsf{exp}(f(\boldsymbol{x}_i, z_{ij}^{cd}, c, d) + \tau_j^{cd})$ that represents the mean number of events in cell W_i and replicate j for the point pattern at levels c and d. The function $f(\boldsymbol{x}_i, z_{ij}^{cd}, c, d)$ is a real function defined in the mixed domain of the continuous variables $\boldsymbol{x}_i \in \mathbb{R}^2$ and $z_{ij}^{cd} \in \mathbb{R}$ and the categorical variables $c \in \mathrm{VA}$ and $d \in \mathrm{FC}$, i.e. $f : \mathbb{R}^3 \times (\mathrm{VA} \times \mathrm{FC}) \to \mathbb{R}$. Note that within groups (crossed categories of factors VA and FC), the function f is a 3-dimensional real function of $\boldsymbol{x}_i \in \mathbb{R}^2$ and $z_{ij}^{cd} \in \mathbb{R}$, i.e. $f(\boldsymbol{x}_i, z_{ij}^{cd}) : \mathbb{R}^3 \to \mathbb{R}$.]

The random effects $\{\tau_j^{cd}\}_{j=1}^{r}$ for the point pattern at levels c and d model differences

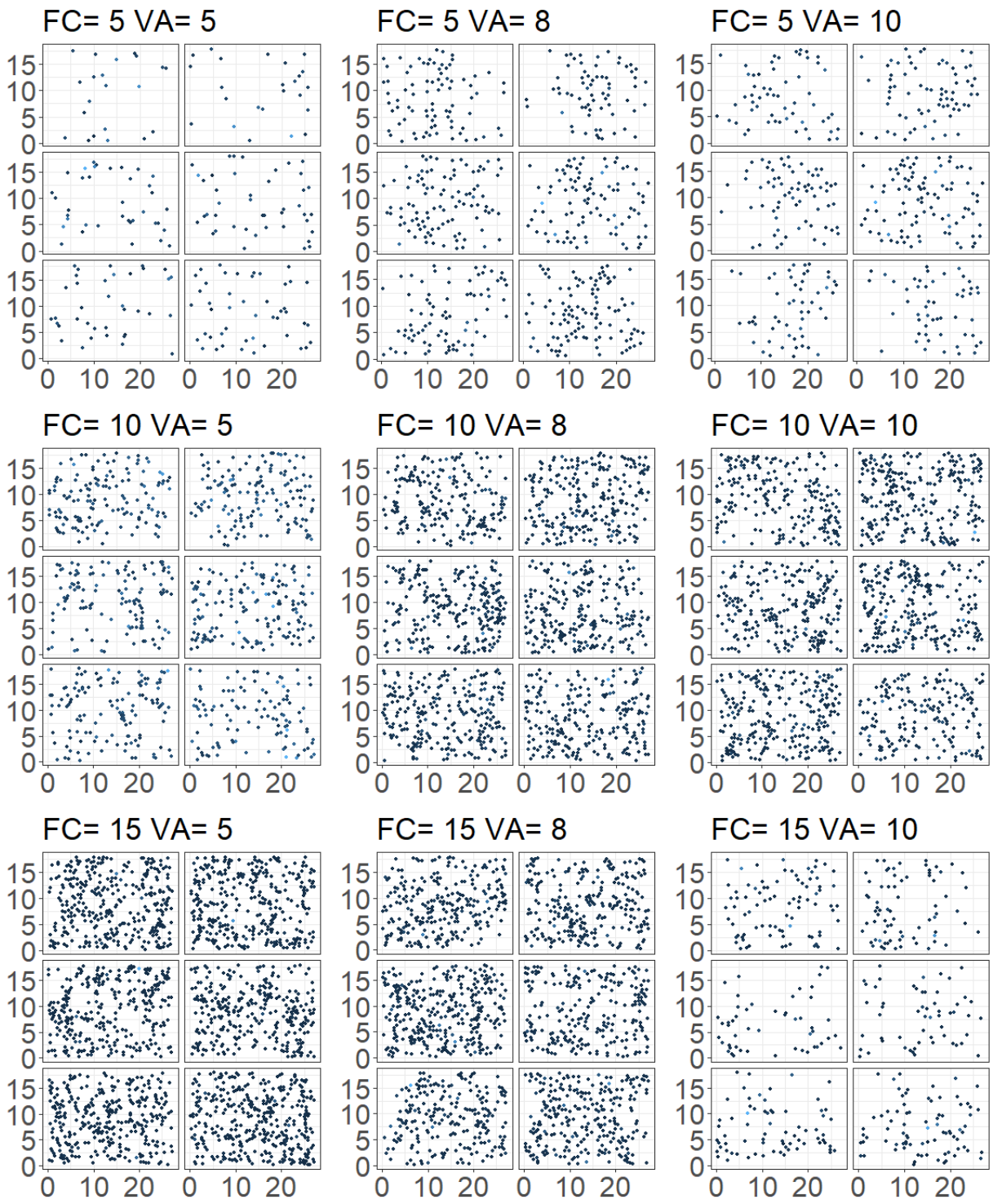

Figure 1: Point patterns of floating bubbles data. Rows represent the three frother concentration levels (FC), while columns the three volumetric airflow rate levels (VA). Each cell contains six spatial point patterns. Each bubble is represented by its center location

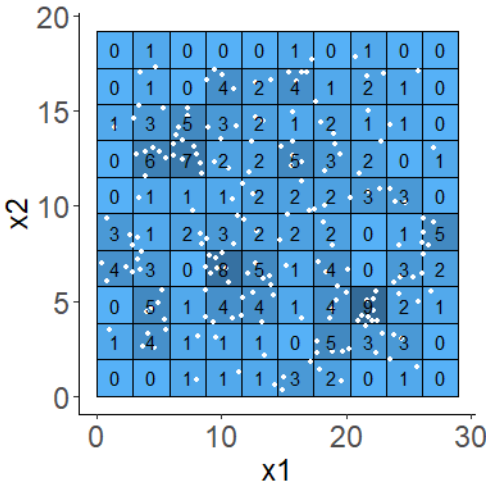

Figure 2: Discretization by a 10×10 ($h = 100$) regular grid of a realization of the bubbles point pattern at the crossed categories VA = 8 and FC = 10 (represented in white dots), and the count y_{ij}^{cd} of events in every grid cell W_i, $c = 8$ and $d = 10$, and replicate j

between replicates as structured random effects from a multivariate normal:

$$\{\tau_1^{cd}, \ldots, \tau_r^{cd}\} \sim \text{Normal}(0, \Sigma^{cd}), \tag{19}$$

where Σ^{cd} is the variance-covariance matrix between replicate effects for the point pattern at levels c and d. Using a multivariate normal for modeling replicate effects may consider possible relationship structures between replicates as well as check for possible lack of complete randomness (i.e., i.i.d.) of replicates.

Following the modeling framework in Section 3, eq. (10), function f is modeled as AGP with a two-way ANOVA structure of crossed factors VA and FC as follows,

$$f(\boldsymbol{x}_i, z_{ij}^{cd}, c, d) = \mu(\boldsymbol{x}_i, z_{ij}^{cd}) + \alpha(\boldsymbol{x}_i, z_{ij}^{cd}, c) + \beta(\boldsymbol{x}_i, z_{ij}^{cd}, d) + (\alpha\beta)(\boldsymbol{x}_i, z_{ij}^{cd}, c, d), \tag{20}$$

with level-specific constraints at the first level of each categorical variables, $c = 5$ and $d = 5$, as seen in eq. (11). To simplify the interpretation of the function f in terms of spatial coordinates $\boldsymbol{x}$ and marks z, we can separate the contribution of $\boldsymbol{x}$ and z in the amount and distribution of bubbles such that they are additive contributions:

$$f(\boldsymbol{x}_i, z_{ij}^{cd}, c, d) = g(\boldsymbol{x}_i, c, d) + h(z_{ij}^{cd}, c, d), \tag{21}$$

where $g : \mathbb{R}^2 \times (\text{VA} \times \text{FC}) \to \mathbb{R}$ and $h : \mathbb{R} \times (\text{VA} \times \text{FC}) \to \mathbb{R}$ are additive functions of $\boldsymbol{x}_i$ and z_{ij}^{cd}, respectively, with a two-way ANOVA functional structure of crossed factors VA and FC based on GPs as follows,

$$g(\boldsymbol{x}_i, c, d) = \mu^g(\boldsymbol{x}_i) + \alpha^g(\boldsymbol{x}_i, c) + \beta^g(\boldsymbol{x}_i, d) + (\alpha\beta)^g(\boldsymbol{x}_i, c, d), \tag{22}$$

$$h(z_{ij}^{cd}, c, d) = \mu^h(z_{ij}^{cd}) + \alpha^h(z_{ij}^{cd}, c) + \beta^h(z_{ij}^{cd}, d) + (\alpha\beta)^h(z_{ij}^{cd}, c, d), \tag{23}$$

with level-specific constraints at the first level of each categorical variables, $c = 5$ and $d = 5$, which implies (as seen in eq. (11))

$$\alpha^g(\boldsymbol{x}_i, 5) = 0,\ \beta^g(\boldsymbol{x}_i, 5) = 0,\ (\alpha\beta)^g(\boldsymbol{x}_i, 5, d) = 0,\ (\alpha\beta)^g(\boldsymbol{x}_i, c, 5) = 0,$$

$$\alpha^h(z_{ij}^{cd}, 5) = 0,\ \beta^h(z_{ij}^{cd}, 5) = 0,\ (\alpha\beta)^h(z_{ij}^{cd}, 5, d) = 0,\ (\alpha\beta)^h(z_{ij}^{cd}, c, 5) = 0.$$

Thus, ANOVA components of function f have been decomposed into additive components of $\boldsymbol{x}$ and z such that,

$$\mu(\boldsymbol{x}_i, z_{ij}^{cd}) = \mu^g(\boldsymbol{x}_i) + \mu^h(z_{ij}^{cd}),$$

$$\alpha(\boldsymbol{x}_i, z_{ij}^{cd}, c) = \alpha^g(\boldsymbol{x}_i, c) + \alpha^h(z_{ij}^{cd}, c),$$

$$\beta(\boldsymbol{x}_i, z_{ij}^{cd}, d) = \beta^g(\boldsymbol{x}_i, d) + \beta^h(z_{ij}^{cd}, d),$$

$$(\alpha\beta)(\boldsymbol{x}_i, z_{ij}^{cd}, c, d) = (\alpha\beta)^g(\boldsymbol{x}_i, c, d) + (\alpha\beta)^h(z_{ij}^{cd}, c, d).$$

We fitted our ANOVA model with $s = 100$ grid cells and $r = 6$ available point pattern replications per combination of levels. We experimentally used different values for s and the fitted results were basically the same, so we decided to present the results based on a number of grid cells of $s = 100$.

[We formulate the Bayesian model with Stan. For the parameters we considered a HMC sampling method using the modelization and probabilistic inference of Stan.]

4.1 Posterior latent function f: Log of the intensity function

The description of the components of function f can be done in a similar way to the classic linear ANOVA models with level-specific constraints, but where now each component is a

GP function of $\boldsymbol{x}$ and z. Thus, the function f at the reference cross-category $(c = 5, d = 5)$ is given by the overall mean function $\mu(\boldsymbol{x}_i, z_{ij}^{cd})$ for all cross-categories of factors $c \in VA$ and $d \in FC$. This overall mean function also participates in all other components of the other cross-categories as their mean, so it can be understood as a baseline. The main effect function of the factor VA is given by the function $\alpha(\boldsymbol{x}_i, z_{ij}^{cd}, c)$ that represents the variation of f as a function of the factor VA given that it is null at $c = 5$. Similarly, the main effect function of factor FC is given by the function $\beta(\boldsymbol{x}_i, z_{ij}^{cd}, d)$ that represents the variation of f as a function of factor FC given that it is null at $d = 5$. The interaction effect function between factors VA and FC is given by the function $(\alpha\beta)(\boldsymbol{x}_i, z_{ij}^{cd}, c, d)$ that represents the variation of f as a function of the two crossed factors VA and FC varying simultaneously from the reference cross-categories, given that its effects for either $c = 5$ and $d = 5$ are null, for which variation of f is given by the main effect functions $\alpha(\boldsymbol{x}_i, z_{ij}^{cd}, c)$ and $\beta(\boldsymbol{x}_i, z_{ij}^{cd}, d)$.

Thus, the value of function f at each individual cross-category can be extracted from function f as follows:

$$\begin{cases} \mu(\boldsymbol{x}_i, z_{ij}^{cd}) & \text{for } c = 5 \;\&\; d = 5, \\ \mu(\boldsymbol{x}_i, z_{ij}^{cd}) + \alpha(\boldsymbol{x}_i, z_{ij}^{cd}, c) & \text{for } c \in \{8, 10\} \;\&\; d = 5, \\ \mu(\boldsymbol{x}_i, z_{ij}^{cd}) + \beta(\boldsymbol{x}_i, z_{ij}^{cd}, d) & \text{for } c = 5 \;\&\; d \in \{10, 15\}, \\ \mu(\boldsymbol{x}_i, z_{ij}^{cd}) + \alpha(\boldsymbol{x}_i, z_{ij}^{cd}, c) + \beta(\boldsymbol{x}_i, z_{ij}^{cd}, d) + (\alpha\beta)(\boldsymbol{x}_i, z_{ij}^{cd}, c, d) & \text{for } c \in \{8, 10\} \;\&\; d \in \{10, 15\}, \end{cases} \tag{24}$$

with $i = 1, \ldots, 100$ and $j = 1, \ldots, 6$.

The function f is a stochastic function and, using Bayesian inference, its full probability distribution can be estimated together with its components μ, α, β, $(\alpha\beta)$, similarly as seen in Section 3.1. Figure 3 shows, as an example, the posterior probability distribution of the log of the intensity function, $f(\boldsymbol{x}_i, z_{ij}^{cd}, c, d)$ (log of the mean number of events), at cell $W_{i=25}$ and replicate $j = 1$. [As a difference of the classic methods, for each cell W_i, we have the distribution of the function. In this Figure, we notice that in the same cell, for

different factors, the distribution is not the same, meaning that factors modify the amount of bubbles of the process.]

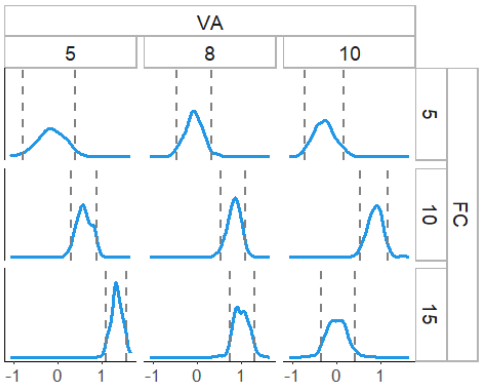

Figure 3: Posterior probability distribution of $f(\boldsymbol{x}_i, z_{ij}^{cd}, c, d)$ (log of the mean number of events) at grid cell W_{25} and replicate $j = 1$, depicted per crossed categories of factors VA and FC. The gray dashed vertical lines in the graphs represent the 2.5 and 97.5 percentiles of the distributions.

For a better understanding of the obtained distribution, on the left Figure 4 depicts the mean of the posterior probability distribution of function $f(\boldsymbol{x}_i, z_{ij}^{cd}, c, d)$ over the whole spatial region and per crossed categories of factors $c \in$ VA and $d \in$ FC. Notice that although function f does not contain replicate's random effects τ_j^{cd}, does depend on marks z_j^{cd} that vary across replicates j. So, in Figure 4, function f is depicted with marks z_j^{cd} measured at replicate $j = 1$. [As for each cell we have a posterior distribution, a surface is constructed. On the left of Figure 4, the mean values of the density for each cell is shown. We observe that the obtained surface are rough, having deep valleys. On the right, the horizontal cut of the surface is shown. The corresponding valleys in each section are represented. This shows that the distribution has a negative or positive mean impact to the intensity function. We can see how the mean value of the distributions is significant.

In this study, we have constructed credible intervals using a pointwise methodology, which provides an intuitive and localized measure of uncertainty for each spatial location.

This approach allows for a clear interpretation of the variability in the estimated intensity function at individual points. However, an alternative and potentially more comprehensive approach is to consider jointly credible spatial regions, which account for the spatial dependence structure and provide a global assessment of uncertainty across the entire domain. While such jointly credible intervals are beyond the scope of this work, they represent an important direction for future research. Readers interested in methodologies for constructing jointly credible spatial regions may refer to (Flury and Furrer, 2022) and (Flury et al., 2021), where these concepts are explored in greater depth.]

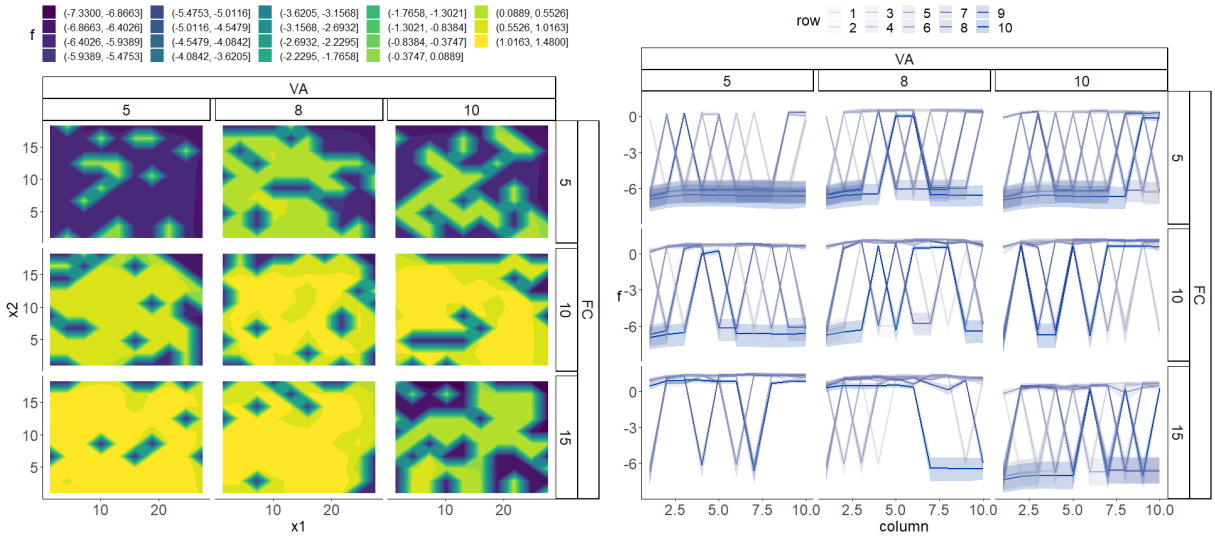

Figure 4: Mean of the posterior probability distribution of log of the intensity function, $f(\boldsymbol{x}_i, z_{ij}^{cd}, c, d)$, as a function of spatial coordinates x^1 and x^2, marks z_j^{cd} belonging to replicate $j = 1$, and cross-categories of factors $c \in$ VA and $d \in$ FC (Left). Similarly, the mean and 2.5 and 97.5 percentiles of horizontal profiles per row of function f, i.e., horizontal cuts per row on surface/function f, (Right).

4.1.1 Spatial function component g

Log of the intensity function, f, was decomposed into two additive components in equation (21): function $g(\boldsymbol{x}_i, c, d)$ of the spatial coordinates $\boldsymbol{x}$ and function $h(z_{ij}^{cd}, c, d)$ of the

bubbles marks z. Figure 5 shows the mean of the posterior distribution of the spatial function component g over the whole spatial region and per crossed categories of factors $c \in$ VA and $d \in$ FC. [The obtained surface is smoother than the spatial component, and its magnitude is also higher. Then the randomness of the process is significantly impacted by the component h.]

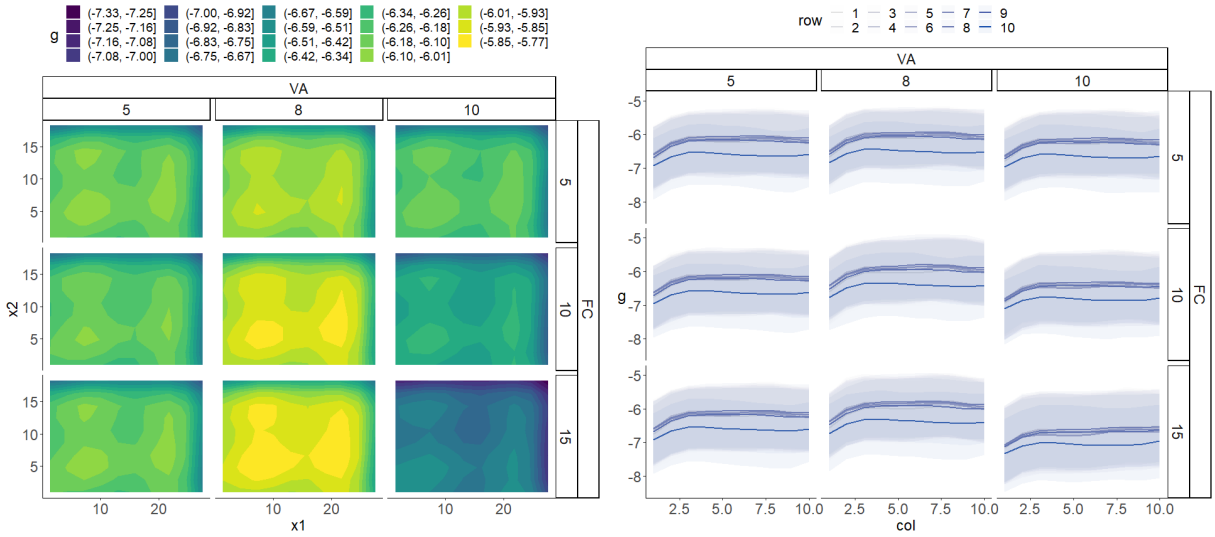

Figure 5: Mean of the posterior probability distribution of the additive spatial function component g as a function of spatial coordinates x^1 and x^2 and per crossed categories of factors VA and FC (Right). Similarly, the mean and 2.5 and 97.5 percentiles of horizontal profiles per row of function g, i.e., horizontal cuts per row on surface/function g, (Left).

4.1.2 Mark function component h

Figure 6 shows the mean and 2.5 and 97.5 percentiles of the posterior distribution of the additive mark function component $h(z_{ij}^{cd}, c, d)$ as a function of bubble marks z_{ij}^{cd} at replicate $j = 1$ and per cross-categories of factors $c \in$ VA and $d \in$ FC. [As we see in Figure 6, the functions are significant different based on the values of the factor, having a wider impact for higher values of VA and more skewed values as FC increases. Another relevant aspect is that despite the change of replicate, the corresponding h function is practically the same.

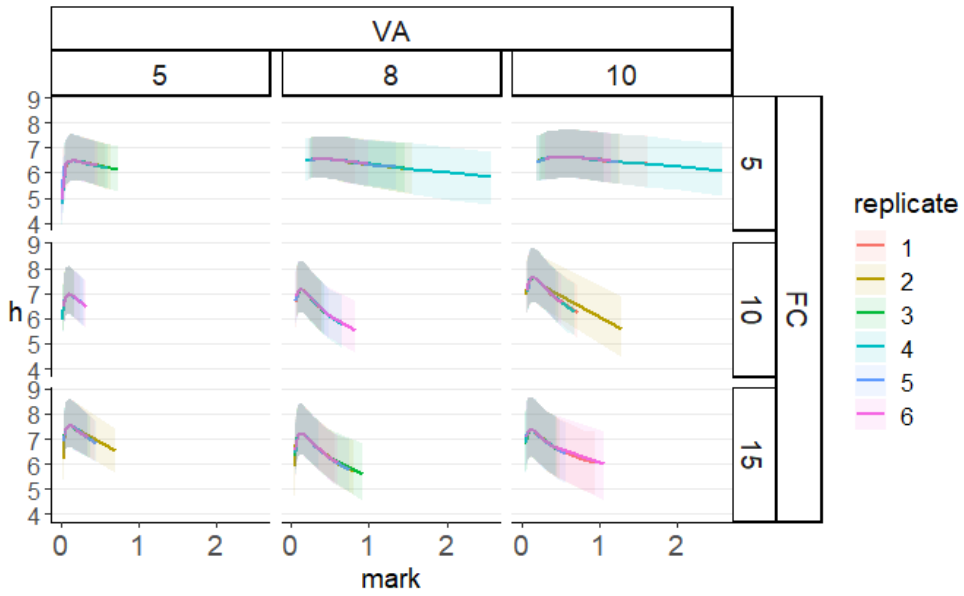

Figure 6: Mean and 2.5 and 97.5 percentiles of the posterior probability distribution of the mark function component h as a function of bubble marks at replicate $j = 1$ and per crossed categories of factors VA and FC.

Then, we can conclude that the diameter of the bubble is more determinant that the space structure for the amount of bubbles, but also that the factors are also determinant.]

4.2 Replicate effects

Replicate effects were included as additive multivariate normal random effects to the log of the intensity function, f, as in equation (19). Using covariances between random effects in a multivariate normal allows for certain non-random structure between replicate effects. Having similar variances between replicate's random effects and covariances close to zero points to complete randomness (i.i.d) in the replicated effects. Figure 7 shows the mean and 2.5 and 97.5 percentiles of the posterior distribution of the replicate's random effects τ_j^{cd} and their variances from the multivariate normal, for replicates j and per cross-categories of factors $c \in$ VA and $d \in$ FC. [As we can see, the mean interval includes the zero which implies that the replicates have not significant relevance on the model.] Figure 8 shows the

mean and 5, 25, 75, and 95 percentiles of the posterior distribution of the correlations from the multivariate normal between replicate's random effects per cross-categories of factors $c \in$ VA and $d \in$ FC. [The random effects are similar to each other. This implies that the random effect is not affected by the category to which it belongs.]

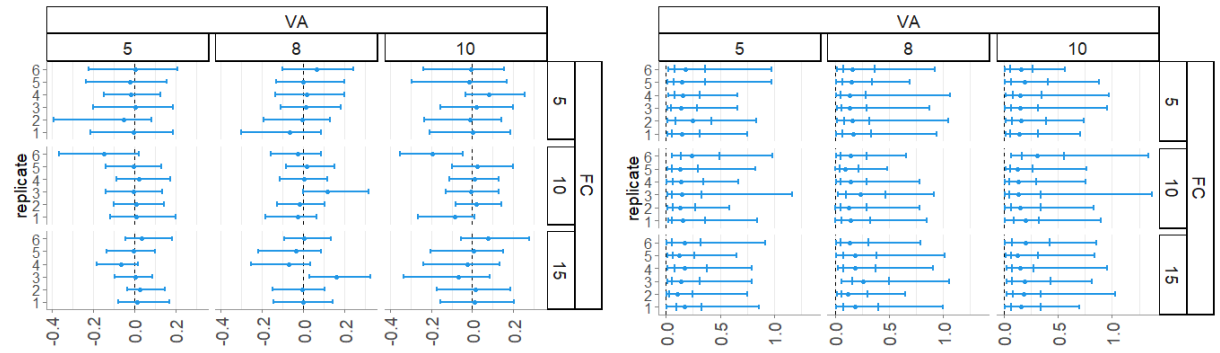

Figure 7: Mean and 2.5 and 97.5 percentiles of the posterior probability distribution of the random effects τ_j^{cd} due to replicates j per crossed categories of factors VA and FC. [Figura de la derecha ???]

4.3 Differences in intensity function between categories

The function at the reference cross-category ($c = 5$, $d = 5$) is given by, as seen above, the overall mean function $\mu(\boldsymbol{x}_i, z_{ij}^{cd})$ across all categories of crossed factors $c \in$ VA and $d \in$ FC. The main effect function $\alpha(\boldsymbol{x}_i, z_{ij}^{cd}, c)$ represents the effects of factor VA given that it is null at VA $= 5\,(c = 5)$ independently of factor FC and relative to the overall mean function $\mu(\boldsymbol{x}_i, z_{ij}^{cd})$. Similarly, the main effect function $\beta(\boldsymbol{x}_i, z_{ij}^{cd}, d)$ represents the effects of factor FC given that it is null at FC $= 5\,(d = 5)$ independently of factor VA and relative to the overall mean function $\mu(\boldsymbol{x}_i, z_{ij}^{cd})$. And the interaction effect function $(\alpha\beta)(\boldsymbol{x}_i, z_{ij}^{cd}, c, d)$ represents the additive effects to the main effect functions $\alpha(\boldsymbol{x}_i, z_{ij}^{cd}, c)$ and $\beta(\boldsymbol{x}_i, z_{ij}^{cd}, d)$ due to the fact that the two factors are varying simultaneously from the reference cross-category $(c = 5, d = 5)$. In Figure 9 the ANOVA components α, β, and $(\alpha\beta)$ are depicted and placed

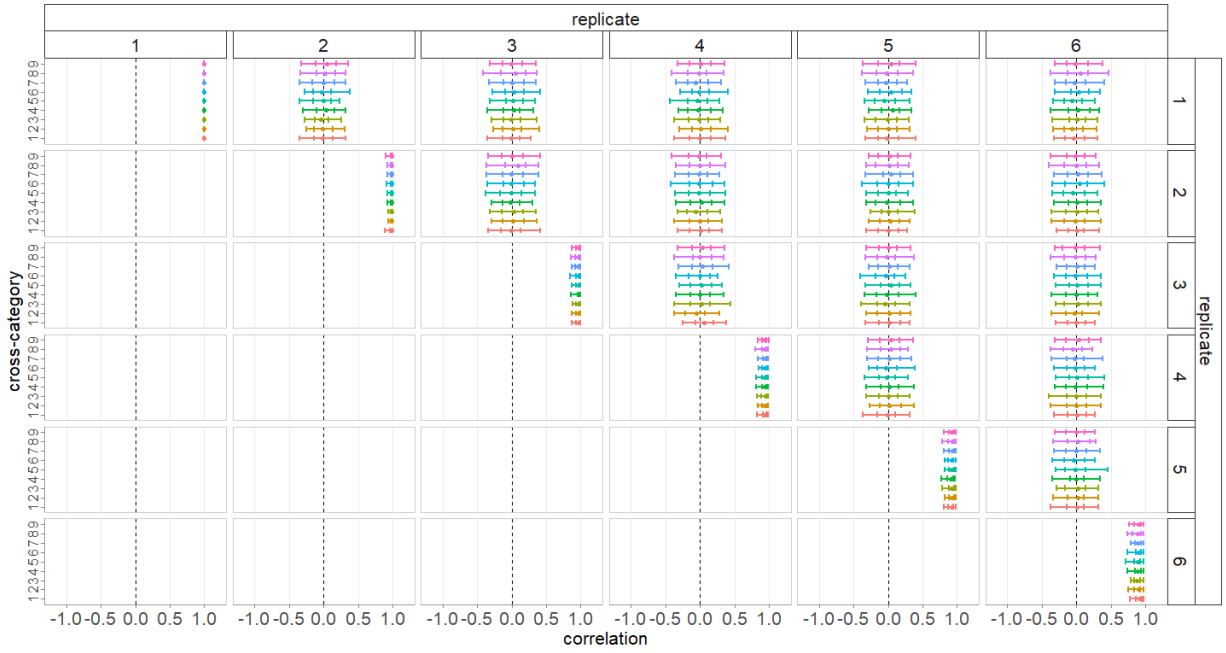

Figure 8: Mean and 5, 25, 75, and 95 percentiles of the posterior distribution of the correlations between replicated random effects per cross-categories of factors $c \in$ VA and $d \in$ FC.

across cross-categories as follows:

$$\begin{cases} \alpha(\boldsymbol{x}_i, z_{ij}^{cd}, c) & \text{at } c \in \{8, 10\} \ \& \ d = 5, \\ \beta(\boldsymbol{x}_i, z_{ij}^{cd}, d) & \text{at } c = 5 \ \& \ d \in \{10, 15\}, \\ (\alpha\beta)(\boldsymbol{x}_i, z_{ij}^{cd}, c, d) & \text{at } c \in \{8, 10\} \ \& \ d \in \{10, 15\}, \end{cases} \tag{25}$$

with $i = 1, \ldots, 100$ and $j = 1, \ldots, 6$. Notice that to form the total function f, the ANOVA components have to be composed following equation (24). [The components are significant, as shown in the figure, given that each combination determines a positive or significant value for all cells. β shows a positive impact in general. For VA $= 8$, α is positive, and for VA $= 10$ is rather negative. The $\alpha\beta$ components are on their part negative, except for VA $= 10$ and FC $= 10$. This results proof that different factors have a significant impact on the intensity of the process.]

The assessment of the differences in the log of the intensity function, f, between cate-

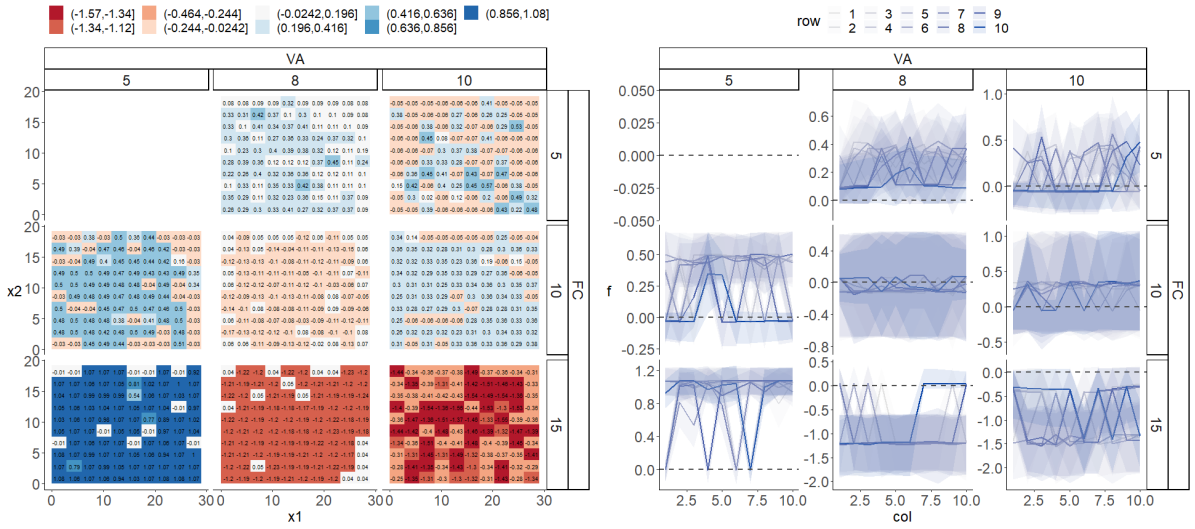

Figure 9: Mean of the posterior probability distributions of the ANOVA components of function f: the main effect function of factor $c \in VA$, $\alpha(\boldsymbol{x}_i, z_{ij}^{cd}, c)$, the main effect function of factor $d \in FC$, $\beta(\boldsymbol{x}_i, z_{ij}^{cd}, d)$, and the interaction effect function of the two factors, $(\alpha\beta)(\boldsymbol{x}_i, z_{ij}^{cd}, c, d)$, over the whole spatial region and per crossed categories of factors VA and FC (Right). Similarly, the mean and 2.5 and 97.5 percentiles of horizontal profiles per row of the ANOVA components, i.e., horizontal cuts per row on the surfaces/functions of the ANOVA components, (Left).

gories is given by the relevance of its ANOVA components α, β, and $(\alpha\beta)$. Their relevance can be determined through their probability distributions $p\big(\alpha(\boldsymbol{x}_i, z_{ij}^{cd}, c)\big)$, $p\big(\beta(\boldsymbol{x}_i, z_{ij}^{cd}, d)\big)$ and $p\big((\alpha\beta)(\boldsymbol{x}_i, z_{ij}^{cd}, c, d)\big)$, evaluating their accumulated probabilities of being less than or greater than zero, as presented in equations (17). Figure 10 shows the accumulated probability of $\alpha^g(\boldsymbol{x}_i, c) + \alpha^h(z_{ij}^{cd}, c)$, $\beta^g(\boldsymbol{x}_i, d) + \beta^h(z_{ij}^{cd}, d)$ and $(\alpha\beta)^g(\boldsymbol{x}_i, c, d) + (\alpha\beta)^h(z_{ij}^{cd}, c, d)$ of being less than or equal to zero in every grid cell W_i. Probabilities close to zero mean that most of the probability mass of the parameter is placed above the zero value of the parameter, and probabilities close to one mean that most of the probability mass of the parameter is placed below the zero value of the parameter. Commonly, probabilities less than the threshold $\eta = 0.05$ or $\eta = 0.10$ indicate that the parameter is credibly greater than zero, and probabilities greater than the threshold $1-\eta = 0.90$ or $1-\eta = 0.95$ indicate that the parameter is credibly less than zero.

The main effects $\alpha(\boldsymbol{x}_i, z_{ij}^{cd}, c)$, which represent the change of the function f as a function of the factor VA given FC = 5 with respect to the general mean effects $\mu(\boldsymbol{x}_i, z_{ij}^{cd})$, were credibly positive since their probabilities of being less than zero are less than 0.05 or 0.10 over most of grid cells. However, the main effects of factor FC, that is, $\beta(\boldsymbol{x}_i, z_{ij}^{cd}, d)$, were not credibly different from the general mean effects $\mu(\boldsymbol{x}_i, z_{ij}^{cd})$ since their probabilities of being less than zero are lower than 0.90 and greater than 0.10 over most grid cells. These results led to the conclusion that by increasing the factor VA, the intensity function of the spatial point pattern increases, however the intensity function does not change significantly as a function of the factor FC, considering the independence between both factors.

When dependence between both factors is considered and they vary simultaneously, credible interaction effects appear when the factor VA is at the level 10 and the factor FC varies, effects that are given by the parameters $(\alpha\beta)(\boldsymbol{x}_i, z_{ij}^{cd}, c = 10, d = 10)$ and $(\alpha\beta)(\boldsymbol{x}_i, z_{ij}^{cd}, c = 10, d = 15)$, causing the function f decreases as FC increases, since its probabilities of being less than zero are greater than 0.95 in most grid cells. However, when

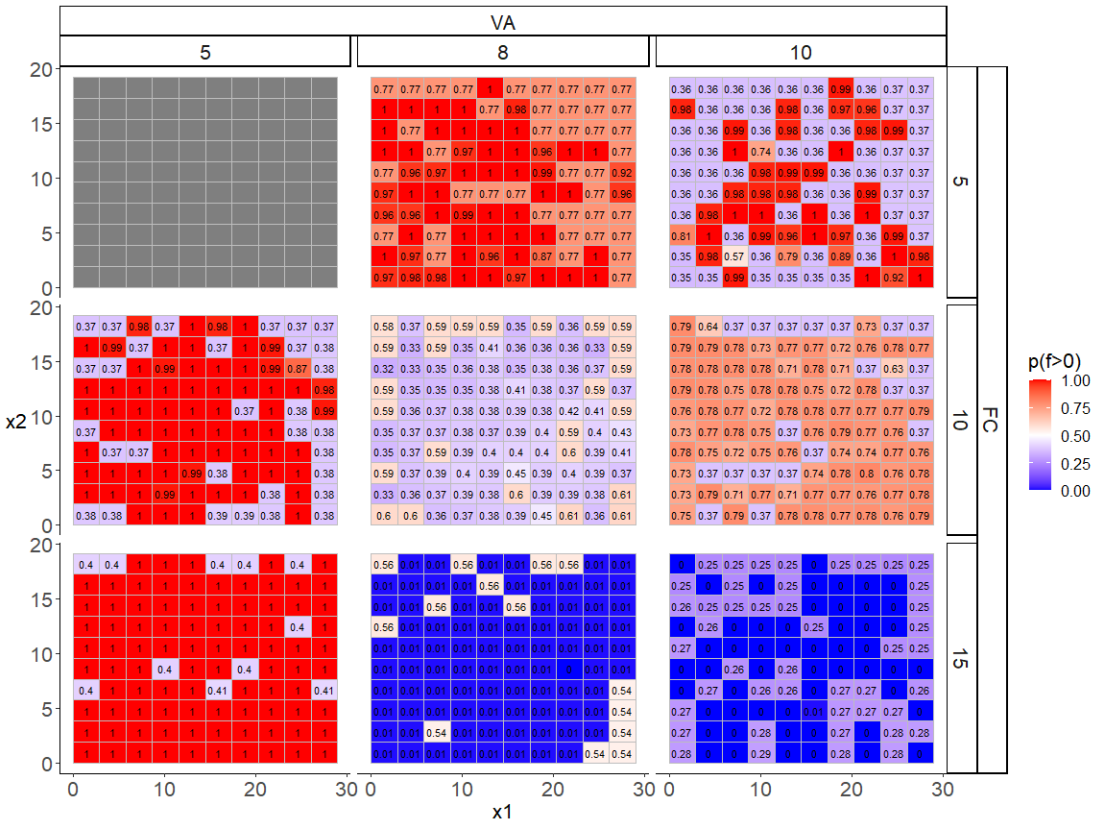

Figure 10: Probability of ANOVA components of function f: the main effect function of factor $c \in VA$, $\alpha(\boldsymbol{x}_i, z_{ij}^{cd}, c)$, the main effect function of factor $d \in FC$, $\beta(\boldsymbol{x}_i, z_{ij}^{cd}, d)$, and the interaction effect function of the two factors, $(\alpha\beta)(\boldsymbol{x}_i, z_{ij}^{cd}, c, d)$, of being less than or equal to zero in every grid cell W_i

the factor VA is at the level 8, the function f does not change significantly when the factor FC varies, it only does so in a small area to the right of the surface of the point pattern. These results led to the conclusion that the factors VA and FC interact with each other when the factor VA is at the level 10, causing a decrease in the intensity function of the point pattern when the factor FC increases.

5 Discussion

The bubbles dataset motivates the use of factorial experiment analysis when observations are spatial point patterns. This change of support, from random variables to spatial patterns, calls for adapting ANOVA methodology to a more complex scenario. The literature in this context is scarse and we thus provide a challenging methodology able to handle these new complex scenarios.

Although LGCPs play a benchmark framework in our problem, we remark that they depend on Gaussian random fields based on perhaps complicated forms of covariance functions that also pose computational issues. We thus present a method based on AGP that alleviate some of these problems while are able to handle complex relationships in spatial point processes. We use Bayesian inference to provide a complete inferential framework able to report full probability distribution estimates and thus quantify the uncertainty associated with all involved parameters.

We provide a complete analysis of a two-way ANOVA experiment for the bubbles data highlighting which factors are significant as being informative on the changes of the intensities of the point patterns, and if there is an inherent interaction between such factors. In our analysis we selected some particular forms of the involved kernels. We leave open the possible use of other forms perhaps more suitable to the problem at hand. We have also restricted to regular cells when calculating the number of points per cells. Additional grid structures, such as Voronoi cells could be implemented.

[The incorporation of marks in the model proved to be valuable, as it directly influenced the spatial structure. Specifically, larger bubbles led to a decrease in the overall number of bubbles, thereby amplifying the effect of the marks relative to the spatial distribution. In this study, the mean and kernel of the marked point pattern were considered collectively; however, future research could investigate the modeling of the mean and kernel specifically for the marks.

Stan was selected as the modeling framework due to its flexibility and transparency. Its capability to construct customized models facilitates modifications with ease. While the computational time is nontrivial, all results in this study were obtained without reliance on a server, indicating that the required computational resources remain manageable.

(González et al., 2021) considers the same data set as we use develops a methodology for statistical testing of two-way factorial experiments for spatial point patterns. It demonstrates the performance of new Fisher-based statistics through simulation experiments. The study focuses on balanced experiments with two factors and equal replicates, motivated by the dataset used. The methodology is validated through simulation experiments, demonstrating its effectiveness in analyzing the impact of frother concentrations and volumetric airflow rates on bubble distributions. Since our methodology differs from that of (González et al., 2021), a direct comparison is not feasible. However, regarding the application, as we both utilize the same dataset, our conclusions align in that the impact of the factor under study is indeed significant for the process.]

References

Baddeley, A., Gregori, P., Mateu, J., Stoica, R., and Stoyan, D. (2006), *Case studies in spatial point process modeling*, volume 185, Springer.

Brooks, S., Gelman, A., Jones, G., and Meng, X.-L. (2011), *Handbook of markov chain monte carlo*, CRC press.

Cox, D. R. (1955), "Some statistical methods connected with series of events," *Journal of the Royal Statistical Society: Series B (Methodological)*, 17, 129–157.

Cressie, N. and Wikle, C. K. (2015), *Statistics for spatio-temporal data*, John Wiley & Sons.

Datta, A., Banerjee, S., Finley, A. O., and Gelfand, A. E. (2016), "Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets," *Journal of the American Statistical Association*, 111, 800–812.

Diggle, P., Moraga, P., Rowlingson, B., and Taylor, B. (2013), "Spatial and spatio-temporal log-Gaussian Cox processes: extending the geostatistical paradigm," *Statistical Science*, 28, 542–563.

Durrande, N., Ginsbourger, D., and Roustant, O. (2012), "Additive covariance kernels for high-dimensional Gaussian process modeling," in *Annales de la Faculté des sciences de Toulouse: Mathématiques*, number 3 in 21.

Duvenaud, D. (2014), *Automatic model construction with Gaussian processes*, Ph.D. thesis, University of Cambridge.

Duvenaud, D. K., Nickisch, H., and Rasmussen, C. (2011), "Additive gaussian processes," *Advances in neural information processing systems*, 24.

Flury, R. and Furrer, R. (2022), "Pipeline to Identify Dominant Features in Spatial Data," *Journal of Computational Mathematics and Data Science*, 5, 100063.

Flury, R., Gerber, F., Schmid, B., and Furrer, R. (2021), "Identification of Dominant Features in Spatial Data," *Spatial Statistics*, 41, 100483.

Gavrikov, V. and Stoyan, D. (1995), "The use of marked point processes in ecological and environmental forest studies," *Environmental and ecological statistics*, 2, 331–344.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013), *Bayesian data analysis*, CRC press.

Gómez-Rubio, V. (2020), *Bayesian inference with INLA*, CRC Press.

González, J. A., Lagos-Álvarez, B. M., and Mateu, J. (2021), "Two-way layout factorial experiments of spatial point pattern responses in mineral flotation," *TEST*, 30, 1046–1075.

Higdon, D., Gattiker, J., Williams, B., and Rightley, M. (2008), "Computer model calibration using high-dimensional output," *Journal of the American Statistical Association*, 103, 570–583.

Kaufman, C. G. and Sain, S. R. (2010), "Bayesian functional {ANOVA} modeling using Gaussian process prior distributions," *Bayesian Anal. 5 (1)*, 123 – 149.

Li, D., Jones, A., Banerjee, S., and Engelhardt, B. E. (2021), "Multi-group Gaussian processes," *arXiv preprint arXiv:2110.08411*.

Lindgren, F., Rue, H., and Lindström, J. (2011), "An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 73, 423–498.

Matheron, G. (1971), "The theory of regionalised variables and its applications," *Les Cahiers du Centre de Morphologie Mathématique*, 5, 212.

Neal, R. M. (1997), "Monte Carlo implementation of Gaussian process models for Bayesian regression and classification," *arXiv preprint physics/9701026*.

Neal, R. M. et al. (2011), "MCMC using Hamiltonian dynamics," *Handbook of markov chain monte carlo*, 2, 2.

Paciorek, C. and Schervish, M. (2003), "Nonstationary covariance functions for Gaussian process regression," *Advances in neural information processing systems*, 16.

Pinheiro, J. C. and Bates, D. M. (1996), "Unconstrained parametrizations for variance-covariance matrices," *Statistics and computing*, 6, 289–296.

Ramsay, J. and Silverman, B. (2005), *Functional Data Analysis*, Springer Series in Statistics, Springer, URL `https://books.google.es/books?id=mU3dop5wY_4C`.

Rasmussen, C. E. and Williams, C. K. (2006), *Gaussian processes for machine learning*, volume 1, Cambridge, MA: MIT press.

Rue, H., Martino, S., and Chopin, N. (2009), "Approximate Bayesian Inference for Latent Gaussian models by using Integrated Nested Laplace Approximations," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 71, 319–392, URL `https://doi.org/10.1111/j.1467-9868.2008.00700.x`.

Schabenberger, O. and Gotway, C. A. (2017), *Statistical methods for spatial data analysis*, Chapman and Hall/CRC.

Shi, J. Q. and Choi, T. (2011), *Gaussian process regression analysis for functional data*, CRC press.

Sørbye, S. H. and Rue, H. (2014), "Scaling intrinsic Gaussian Markov random field priors in spatial modelling," *Spatial Statistics*, 8, 39–51.

Timonen, J., Mannerström, H., Vehtari, A., and Lähdesmäki, H. (2021), "lgpr: an interpretable non-parametric method for inferring covariate effects from longitudinal data," *Bioinformatics*, 37, 1860–1867.

Van Lieshout, M. N. M. (2019), *Theory of spatial statistics: a concise introduction*, CRC Press.

Wilson, A. and Adams, R. (2013), "Gaussian process kernels for pattern discovery and extrapolation," in *International conference on machine learning*, PMLR.