Development and Implementation of Bayesian Computer Model Emulators

by

Danilo Lopes

Department of Statistical Science
Duke University

Date: _____

Approved:

_____
James O. Berger, Co-Supervisor

_____
Robert L. Wolpert, Co-Supervisor

_____
M. J. Bayarri

_____
Alan E. Gelfand

Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the Department of Statistical Science
in the Graduate School of Duke University
2011

Abstract
(Computer Models)

Development and Implementation of Bayesian Computer Model Emulators

by

Danilo Lopes

Department of Statistical Science
Duke University

Date: _____

Approved:

_____
James O. Berger, Co-Supervisor

_____
Robert L. Wolpert, Co-Supervisor

_____
M. J. Bayarri

_____
Alan E. Gelfand

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Statistical Science in the Graduate School of Duke University
2011

Abstract

Our interest is the risk assessment of rare natural hazards, such as large volcanic pyroclastic flows. Since catastrophic consequences of volcanic flows are rare events, our analysis benefits from the use of a computer model to provide information about these events under natural conditions that may not have been observed in reality.

A common problem in the analysis of computer experiments, however, is the high computational cost associated with each simulation of a complex physical process. We tackle this problem by using a statistical approximation (emulator) to predict the output of this computer model at untried values of inputs. Gaussian process response surface is a technique commonly used in these applications, because it is fast and easy to use in the analysis.

We explore several aspects of the implementation of Gaussian process emulators in a Bayesian context. First, we propose an improvement for the implementation of the plug-in approach to Gaussian processes. Next, we also evaluate the performance of a spatial model for large data sets in the context of computer experiments.

Computer model data can also be combined to field observations in order to calibrate the emulator and obtain statistical approximations to the computer model that are closer to reality. We present an application where we learn the joint distribution of inputs from field data and then bind this auxiliary information to the emulator in a calibration process.

One of the outputs of our computer model is a surface of maximum volcanic flow

height over some geographical area. We show how the topography of the volcano area plays an important role in determining the shape of this surface, and we propose methods to incorporate geophysical information in the multivariate analysis of computer model output.

Contents

List of Tables

List of Figures

Acknowledgements

First, I would like to thank my two advisors, Jim Berger and Robert Wolpert, for teaching me not only Statistical Science, but I also friendship, professionalism and some good jokes. They are truly my role models. I am also lucky for having Susie Bayarri as my non-official advisor, and I thank her for her kindness and valued inputs in my work.

I would like to thank all the people in the Department of Statistical Science, staff, faculty and students, for creating a great environment for academic research and for their contribution to my development not only as a professional, but also as human being.

I am grateful to SAMSI for giving me support as a graduate fellow during a year. I thank all the members of my research group, the Froggers, for giving me the opportunity to collaborate in such a wonderful project.

I thank my family for the constant support, even from such a long distance. Special thanks to Camilla Gomes for giving me strength even in the most difficult moments. The people you love are always with you, even when they are not physically close to you.

1

Introduction

Computer models are important when the researcher is interested in analyzing complex systems, where physical experimentation and data collection are too demanding and too expensive, or even impossible (Sacks *et al.* 1989). Some examples of applications are weather modeling, experiments in engineering and studies of catastrophic events, such as hurricanes and volcano eruptions. These computer models (also called *simulators*) often have high-dimensional inputs and a deterministic output (reruns of the code with the same inputs will give identical observations).

Due to the expense of running the code in some applications, validation analysis commonly uses approximations based on a limited number of code evaluations. One general approximation, motivated by kriging in the spatial statistics literature, is fitting a Gaussian process to the deterministic output as a model approximation, or *emulator*. The Gaussian Response Surface Approximation method (*GASP*) follows on work in Sacks *et al.* (1989), Currin *et al.* (1991), Welch *et al.* (1992), and Morris *et al.* (1993) and is the most common approach to the emulation of computer models. Gaussian processes give a wide range of flexible realizations that we can use to interpolate through our data points, allowing us to make simple probability

statements about our approximations at the same time.

Consider we have a computer model $y(\boldsymbol{z})$ as a function of the input vector $\boldsymbol{z}$. We assume a regression model for the expected value of our observations and model the residuals as a Gaussian process with constant variance and correlation that depends on the proximity of the input values. For a finite number of locations n in the input space, the Gaussian process reduces to a multivariate normal distribution. Let $\boldsymbol{y} = (y(\boldsymbol{z}_i), \cdots, y(\boldsymbol{z}_n))$ denote the vector of model evaluations. Then

$$\boldsymbol{y} \sim N_n(\Psi(\boldsymbol{z})\theta, \sigma^2 C)$$

where $\Psi(\cdot)$ is a vector of known functions, θ is a vector of unknown regression coefficients, σ^2 is the variance of the Gaussian process and C is the $n \times n$ correlation matrix whose entries (i, j) are given by the correlation function $c(\boldsymbol{z}_i, \boldsymbol{z}_j|\beta)$. In computer model applications, we typically assume a multidimensional power exponential correlation function:

$$c(\boldsymbol{z}_i, \boldsymbol{z}_k|\beta, \alpha) = \prod_{j=1}^{p} \exp\{-\beta_j|\boldsymbol{z}_{ij} - \boldsymbol{z}_{kj}|^{\alpha_j}\} \tag{1.1}$$

where p is the dimension of the input space, $\boldsymbol{\beta}_j$'s are positive range parameters that control the decay of the correlation function over each dimension and $\alpha_j \in [1, 2]$ are roughness parameters expressing the smoothness of the Gaussian process.

In the case where $\alpha_j = 2$ for all j we have that the Gaussian process realizations are infinitely differentiable, which seems plausible in many scientific applications. However, this extremely smooth correlation function can produce numerically singular correlation matrices. Ababou et al. (1994) show how the condition number of

Gaussian covariance matrices (equivalent to a power exponential correlation function with $\alpha = 2$) is very large compared to other covariance models, which means that numerical decompositions of the covariance matrix are very sensitive to small perturbations. Because of this numerical instability, we usually set α to some value close to 2, e.g. 1.9, when we want to reflect a prior belief that our computer model function is very smooth. As noted by Paulo (2005), inference is fairly insensitive to reasonable values of the roughness parameter in the context of computer experiments, since computer models are usually exercised at well-separated input points.

During, the emulation process, we assign a probability model to a deterministic function $y(\boldsymbol{z})$ that we have observed at some values of $\boldsymbol{z}$ with no measurement error. This probability model describes our uncertainty about the computer model at values of inputs that we have not run. We will also later assign prior beliefs to the emulator parameters and estimate them in a Bayesian fashion. However, since the emulator parameters do not have much interpretability in real life, the focus of our analysis is on the prediction of the computer model $y(\boldsymbol{z})$ at new values $\boldsymbol{z}$.

Emulation, however, is usually only a small part of computer model analysis. Some other aspects of the analysis of simulators include:

- **Uncertainty analysis and sensitivity analysis:** the objective of the uncertainty analysis is to verify how uncertainty and prior beliefs in the distribution of inputs propagate to the distribution of the output. We will discuss more about uncertainty analysis in Chapter 4 and Chapter 5. On the other hand, sensitivity analysis is related to the identification of relevant inputs to the simulator. See Saltelli *et al.* (2000) and Santner *et al.* (2003) for extended reviews.

- **Calibration:** Computer models may also include some inputs that do not exist in real life or that may have a physical meaning, but are not observed

in the field. Using field data of the actual process under consideration, we can find the values for these unknown inputs corresponding to model outputs that fit field observations best. Bayesian approaches to calibration have been addressed by Kennedy and O'Hagan (2001), Bayarri *et al.* (2007a), Craig *et al.* (1997) and Goldstein and Rougier (2003). Another type of calibration, discussed in Chapter 4, is when our beliefs are not represented by a fixed value (or a distribution of values) for the inputs, but rather by an association in the input space (motivated by field observations, for example).

- **Validation:** knowing that the computer model does represent reality in an adequate way, we can account for model bias and make probability statements about the accuracy of the simulator. Bayarri *et al.* (2007a) describe a mix of Bayesian and likelihood methods for quantifying uncertainty in computer models, accounting for tuning parameters and possible model bias. Other discussions on the validation process can be found in Roache (1998), Oberkampf and Trucano (2000), Cafeo and Cavendish (2001), Easterling (2001), Kennedy and O'Hagan (2001), Pilch *et al.* (2001), Trucano *et al.* (2002), and Santner *et al.* (2003).

The approximation to the computer model in the emulation step is usually intended to be used in some of the areas discussed above. Therefore, we cannot spend all of our computational effort in GaSP fitting. We need a fast surrogate that can be used repeatedly in several other parts of the analysis. This implies a computational constraint on the implementation of the emulator, that we roughly express as "being able to run the emulator on a regular laptop".

This dissertation analyzes several aspects of the development and implementation of computer model emulator in a Bayesian approach. In chapter 2, we consider methods for numerical estimation of Gaussian Process estimation parameters. We focus

our attention on the use of numerical estimates for prediction of computer model outputs in a plug-in approach and compare the performance of different alternatives to the numerical estimation. Chapter 3 analyzes alternatives to regular Gaussian process in the emulation of computer models with a large number of evaluations available. We are particularly interested in the use of the modified predictive process model (Banerjee *et al.*, 2008) as an approximation to the regular Gaussian process emulator, discussing the advantages and disadvantages of this approximation. We also analyze the use of modularization (Liu *et al.*, 2009) in conjunction with the modified predictive process in the emulation of computer experiments.

In chapter 4, we propose a hierarchical model for calibration of computer model emulators, using several sets of auxiliary data to learn about the joint distribution of inputs. We also propose a method for combining the output of the hierarchical analysis with the emulation of the computer model. Finally, in chapter 5, we discuss several problems associated with the multivariate analysis of a specific computer model output and propose the use of topographic information in the analysis. In particular, we analyze different aspects of the uncertainty in the topography and propose a probabilistic spatial model to describe these uncertainties.

2

Estimating Gaussian Process Correlation Parameters

2.1 Introduction

Computer models are increasingly being used to represent complex physical processes in several research applications. They are typically used to obtain information about events that are possible, but too expensive (or even impossible) to be observed in the field as a controlled experiment. For a set of initial conditions, expressed in the vector of inputs $\mathbf{x}$, a computer model (or simulator) generates a deterministic value y, the output. For simplicity, we assume that the output is a single scalar. Since the physical process in study is usually complex, the simulator requires several calculations and may be very time-consuming to run.

Our goal is to built a statistical approximation (also called emulator) to the output as a function of the input variables. The Gaussian Response Surface Approximation method ($GASP$) is the most common method for creating computationally cheap surrogates to computer codes (see Sacks *et al.* 1989). As Nagy *et al.* (2007) pointed out, trading off accuracy by computational speed of the statistical approximation is

acceptable when we can measure the deviations of the surrogate's predictions from the real computer model output.

A very common and simple approach that speeds up computation of emulators is the plug-in approach. It consists of using a optimization method to calculate numerical estimates of the GASP parameters (usually the maximum likelihood estimates, or MLEs) and then assuming that the true values of the parameters are equal to theirs estimates. This approach usually underestimates the variance of predictions because it does not take into account the uncertainties about the parameters, as shown by Abt (1999). The use of a Bayesian approach to estimate mean and variance parameters, however, usually helps to reduce the underestimation of the variance of predictions. Note that a full Bayes approach is the best solution to tackle this problem, but it is considered too computationally expensive in many applications.

The main problem of the plug-in approach, however, is that the likelihood of a Gaussian process is usually not a nice function of the correlation parameters. Depending on our observed values, the likelihood function can be maximized at zero or infinite values, rendering the plug-in approach based on MLEs useless.

In this chapter, we consider alternatives to the common use of MLEs in the plug-in method for fitting Gaussian processes. We divide the chapter as follows: in Section 2.2, we present the plug-in approach in details and some alternatives proposed in the literature; in Section 2.4, we show some problems related to the use of maximum likelihood estimates in GASP estimation; in Section 2.3, we determine common choices of objective priors for the GASP parameters; in Section 2.5, we propose different Laplace approximations to calculate posterior summaries of the correlation parameter. Section 2.7 offers some conclusions and directions for further work.

2.2 The MLE plug-in approach

Consider we have a computer model $y(z)$ as a function of the input vector z. Let $\mathbf{y}_1$ and $\mathbf{y}_2$ denote $n_1 \times 1$ and $n_2 \times 1$ vectors of output observations from our computer model. Assuming a Gaussian process model for $y(\boldsymbol{z})$, the joint distribution of $\mathbf{y}_1$ and $\mathbf{y}_2$ reduces to a multivariate normal distribution:

$$\boldsymbol{y} = \begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{pmatrix} \sim N_{n_1+n2} \left\{ \begin{pmatrix} \Psi_1\boldsymbol{\theta} \\ \Psi_2\boldsymbol{\theta} \end{pmatrix}, \sigma^2 \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^t & \Sigma_{22} \end{pmatrix} \right\} \tag{2.1}$$

where $\boldsymbol{\theta} \in \mathcal{R}_q$, $\sigma_2 > 0$, Ψ_i is a known full column rank matrix and the entries of the correlation matrix are given by the correlation function $c(\cdot, \cdot|\beta)$. In computer model applications, we typically assume a multidimensional power exponential correlation function:

$$c(\boldsymbol{z}_i, \boldsymbol{z}_k|\beta, \alpha) = \prod_{j=1}^{p} \exp\left\{\beta_j |z_{ij} - z_{kj}|^{\alpha_j}\right\} \tag{2.2}$$

where p is the dimension of the input space, $\boldsymbol{\beta}_j$'s are positive range parameters that control the decay of the correlation function over each dimension and $\alpha_j \in [1, 2]$ are roughness parameters expressing the smoothness of the Gaussian process. We consider the roughness parameters to be known.

There are two types of plug-in approach: *full plug-in* and *hyper-parameter plug-in*. In the full plug-in approach, all the parameters are assumed to be fixed and equal to their estimates, i.e. prior distributions are degenerated. In this case, the posterior predictive distribution of $\mathbf{y}_1$ given $\mathbf{y}_2$ is

$$
\begin{aligned}
\mathbf{y}_1|\mathbf{y}_2 &\sim N_{n_1}(\hat{m}_{1|2},\, \hat{\sigma}^2\hat{\Sigma}_{1|2}) \\
\hat{m}_{1|2} &= \Psi_1\hat{\boldsymbol{\theta}} + \hat{\Sigma}_{12}\hat{\Sigma}_{22}^{-1}(\mathbf{y}_2 - \Psi_2\hat{\boldsymbol{\theta}}) \\
\hat{\Sigma}_{1|2} &= \hat{\Sigma}_{11} - \hat{\Sigma}_{12}\hat{\Sigma}_{22}^{-1}\hat{\Sigma}_{12}^t
\end{aligned}
\tag{2.3}
$$

where $\hat{\Sigma}_{ij}$ is defined by plugging in the point estimate of $\boldsymbol{\beta}$.

On the other hand, hyper-parameter plug-in approach consists of plugging in the correlation parameters and assuming a noninformative prior distribution $\pi(\boldsymbol{\theta}, \sigma^2) \propto 1/\sigma^2$ for $\boldsymbol{\theta} \in \mathbb{R}^q$ and $\sigma^2 > 0$. In this case, the posterior predictive distribution becomes a n_1-variate noncentral Student-t distribution:

$$
\begin{aligned}
\mathbf{y}_1|\mathbf{y}_2 &\sim \mathcal{T}_{n_1}(\hat{m}_{1|2},\, \hat{\sigma}^2\hat{\Sigma}_{1|2},\, n_2 - q) \\
\hat{m}_{1|2} &= \Psi_1\hat{\boldsymbol{\theta}} + \hat{\Sigma}_{12}\hat{\Sigma}_{22}^{-1}(\mathbf{y}_2 - \Psi_2\hat{\boldsymbol{\theta}}) \\
\hat{\Sigma}_{1|2} &= \hat{\Sigma}_{11} - \hat{\Sigma}_{12}\hat{\Sigma}_{22}^{-1}\hat{\Sigma}_{12}^t + (\Psi_1 - \hat{\Sigma}_{12}\hat{\Sigma}_{22}^{-1}\Psi_2)(\Psi_2^t\hat{\Sigma}_{22}^{-1}\Psi_2)^{-1}(\Psi_1 - \hat{\Sigma}_{12}\hat{\Sigma}_{22}^{-1}\Psi_2)^t
\end{aligned}
\tag{2.4}
$$

where $\hat{\boldsymbol{\theta}} = (\Psi_2^t\hat{\Sigma}_{22}^{-1}\Psi_2)^{-1}(\Psi_2^t\hat{\Sigma}_{22}^{-1}\mathbf{y}_2)$, $\hat{\sigma}^2 = (\mathbf{y}_2^t\hat{\Sigma}_{22}^{-1}\mathbf{y}_2 - \hat{\boldsymbol{\theta}}^t(\Psi_2^t\hat{\Sigma}_{22}^{-1}\Psi_2)\hat{\boldsymbol{\theta}})/(n_2 - q)$ and $\hat{\Sigma}_{ij}$ is defined by plugging in the point estimate of $\boldsymbol{\beta}$.

The algorithm for the plug-in approach to Gaussian process emulators is extremely simple, consisting of two stages: estimation and prediction.

1. **Estimation:** Obtain the Maximum Likelihood Estimates (or posterior estimates) of the parameter $\hat{\boldsymbol{\beta}}$.

2. **Prediction:** Plug in the estimates into the predictive distribution of $y(\boldsymbol{z})$ in (2.3) or (2.4), depending on the model assumptions, and obtain point estimates and confidence (or credible) bounds.

The main reason for the use of MLE plug-in method is computational efficiency. Even if we choose a simpler correlation function than the one in (2.2), integrating the hyperparameters numerically with respect to the posterior distribution is still a computationally intensive task. This problem is aggravated when we increase the dimensionality of the input space or the number of observations. On the other side, the MLE plug-in approach requires the use of an optimization method, such as the Newton-Raphson method or the simulated annealing algorithm (Belisle 1992). Optimization methods have a large variety of algorithms and are usually computationally inexpensive compared to MCMC methods, presenting faster convergence to a solution. Another advantage of plug-in methods is that they can be used as black box methods, i.e., there is no need for tuning or problem-specific coding, so it can be used by non-specialists in statistical analysis.

The disadvantage of the plug-in approach is that the analysis does not account fully for computer code uncertainty, in the sense that we are assuming that we know the true value of the parameters. Not recognizing the uncertainty in the parameters affects the coverage of the credible bounds, leading to overly optimistic envelopes around our predictions. Abt (1999) evaluates the prediction error of plugging in estimates of the GASP parameters in (2.4), extends the approximation of the mean square prediction error and evaluates the proposed approximation under the power-exponential correlation function in (2.2). Another problem is the use of maximum likelihood estimates in the Gaussian process model: since the likelihood is a complex function of the correlation parameter, it is not hard to produce examples where the likelihood function is maximized at an unrealistic value of the parameter, e.g. zero or infinity. For more details, see Section 2.4.

A common alternative to the plug-in approach is the a fully Bayesian analysis, as detailed in Hancock and Stein (1993) (applied to geostatistics), and in Paulo (2005) and Bayarri *et al.* (2007b) (applied to computer experiments). It consists

of drawing posterior samples of the GASP parameters using MCMC techniques and then sampling from (2.3) given each value in the MCMC chain. Using posterior predictive samples of the vector $\mathbf{y}_1$, we can generate point estimates credible bounds for each component of $\mathbf{y}_1$. Alternatively, one can also calculate the Mean Squared Error of posterior predictions by computing the sample variance of the posterior expectation in (2.3) and adding it to the sample mean of the variance in (2.3).

Nagy *et al.* (2007) proposed an alternative method called *Fast Bayesian Inference*, or FBI. The FBI approximates the posterior distribution of the log-transformation of $\boldsymbol{\beta}$ by a multivariate normal distribution. This normal approximation is also used in the Laplace approximation methods discussed in Section 2.5. The FBI then generate samples from the approximate multivariate normal distribution of $\log \boldsymbol{\beta}$ and treat them as samples from the posterior distribution. Summaries of the posterior predictive distribution can then obtained in a similar fashion to the Full Bayes. By approximating the posterior distribution, the FBI avoids the computational cost of the MCMC procedure that is common to fully Bayesian approach. Nagy *et al.* (2007), however, recommends the use of uniform prior on the log scale for the correlation parameters, so the posterior modes are equivalent to maximum likelihood estimates. In the Section 2.4 we show that this choice of prior may be problemstic, mainly because the propriety of the posterior distribution is not guaranteed.

The plug-in method for Gaussian process emulators is certainly not a new idea. Pioneering works in the emulation of computer models, such as Currin *et al.* (1988) and Sacks *et al.* (1989), suggest the use of maximum likelihood estimation in the analysis. However, as Bayarri *et al.* (2007a) pointed out, a full justification of the use of MLE plug-in method is still an open theoretical issue. However, Abt (1999), Kennedy and O'Hagan (2001), and Bayarri *et al.* (2007b) have observed that the use of MLEs of the GASP correlation parameters yields approximately the same results as a full Bayesian analysis, and that acknowledging for uncertainty

in the correlation parameters produces a very small effect in the analysis, specially when calibration/tuning parameters are present in the computer model. Nonetheless, Bayarri *et al.* (2007b) recommend the use of MLE plug-in GASPs together with a Bayesian analysis of the validation/prediction process.

2.3 Objective priors and integrated likelihood

2.3.1 Overview

In a Bayesian approach to the emulation of computer models, prior distributions must be assigned to the parameters of the Gaussian process. However, these parameters do not have a straightforward interpretation and very little or no prior information about them is available. Analysts may have some prior knowledge about the real physical process in study, but it is difficult to extend these pieces of information to the computer model. As part of the validation process, we cannot simply assume that the computer model represents all the features of the reality in an adequate way. In addition, the implementation of the computer does not give much information about the output before we observe their runs.

Because of that, the analysis and validation of complex computer models usually requires the specification of automatic or default prior distributions. One needs to be specially careful about the propriety of the posterior distribution, since the likelihood is a complex function of the GASP correlation parameters that needs special attention at the tails. Berger *et al.* (2001) studied this problem in depth and developed default prior distributions of GASP parameters for the case where the range correlation parameter β in (2.2) is one-dimensional. Paulo (2005) extends these objective priors to the case where the input space (and, consequently, β) is multidimensional and investigates if the priors in study yield proper distributions.

Berger *et al.* (2001) and Paulo (2005) consider the roughness parameter α in (2.2) to be known when formally specifying prior distributions for the GASP parameters,

and so we do. Determining objective prior distributions when α is unknown and finding conditions that ensure posterior propriety is still an open problem. In this section, we present the objective prior distributions described in Paulo (2005) and show that they can be obtained in closed form (apart from a multiplicative constant) in the case where β is one-dimensional, $\alpha = 1$ and the data points are equally spaced.

2.3.2 General formulation

Assume the GASP model for any $n = n_1 + n_2$ runs of the simulator $\boldsymbol{y} = (\boldsymbol{y}(\boldsymbol{z}_1), \cdots, \boldsymbol{y}(\boldsymbol{z}_n))$ described in (2.1). Define $\Theta = (\sigma^2, \boldsymbol{\theta}^t, \boldsymbol{\beta}^t)^t$ and $\boldsymbol{X} = (\Psi_1^t \; \Psi_2^t)^t$. The resulting likelihood function is given by:

$$L(\Theta|\mathbf{y}) \;\propto\; (\sigma^2)^{-n/2}|\Sigma_{\boldsymbol{\beta}}|^{-1/2}\exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y}-\boldsymbol{X}\boldsymbol{\theta})^t\Sigma_{\boldsymbol{\beta}}^{-1}(\mathbf{y}-\boldsymbol{X}\boldsymbol{\theta})\right\} \tag{2.5}$$

We consider priors of the form:

$$\pi(\sigma^2, \boldsymbol{\theta}, \boldsymbol{\beta}) \;\propto\; \frac{\pi(\boldsymbol{\beta})}{(\sigma^2)^a} \tag{2.6}$$

for different $\pi(\boldsymbol{\beta})$ and $a \in \mathbb{R}$. $\pi(\boldsymbol{\beta})$ is here, in a abuse of terminology, called the marginal prior distribution of $\boldsymbol{\beta}$.

Using the prior in (2.6), we can integrate the product of the likelihood and the prior over $(\boldsymbol{\theta}, \sigma^2)$:

$$\int_{\mathbb{R}^q\times\mathbb{R}_+} L(\sigma^2, \boldsymbol{\beta}, \boldsymbol{\theta}|\mathbf{y})\pi(\boldsymbol{\theta}, \sigma^2, \boldsymbol{\beta})\, d\boldsymbol{\theta}\, d\sigma^2 \;=\; L^I(\boldsymbol{\beta}|\mathbf{y})\pi(\boldsymbol{\beta})$$

where

$$L^I(\boldsymbol{\beta}|\mathbf{y}) \;\propto\; |\boldsymbol{\Sigma}(\boldsymbol{\beta})|^{-1/2}|\boldsymbol{X}^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta})\boldsymbol{X}|^{-1/2}\left(S^2_{\boldsymbol{\beta}}\right)^{-((n-q)/2+a-1)} \tag{2.7}$$

here $S^2_{\boldsymbol{\beta}} = \mathbf{y}^t \boldsymbol{Q} \mathbf{y}$, $\boldsymbol{Q} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}) P$ corresponds to the generalized residual sum of squares, with $\boldsymbol{P} = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^t \boldsymbol{\Sigma}^{-1}_{\boldsymbol{\beta}} \boldsymbol{X})^{-1} \boldsymbol{X}^t \boldsymbol{\Sigma}^{-1}_{\boldsymbol{\beta}}$.

$L^I(\boldsymbol{\beta}|\mathbf{y})$ is called the *integrated likelihood* function of $\boldsymbol{\beta}$ (Berger *et al.* 1999). Having the integrated likelihood in closed form as in (2.7) is important in our analysis because we can combine this prior $\pi(\boldsymbol{\beta})$ to the integrated likelihood $L^I(\boldsymbol{\beta}|\boldsymbol{y})$ in order to obtain the marginal posterior distribution of $\boldsymbol{\beta}$:

$$\pi(\boldsymbol{\beta}|\mathbf{y}) \quad \propto \quad L^I(\boldsymbol{\beta}|\mathbf{y}) \pi(\boldsymbol{\beta})$$

We can see that the integrated likelihood describes how the observations affect the posterior distribution and can be used to evaluate posterior propriety, since $\pi(\boldsymbol{\beta}|\mathbf{y})$ is proper if, and only if, $0 < \int_\Omega L^I(\boldsymbol{\beta}|\mathbf{y}) \pi(\boldsymbol{\beta}) \, d\boldsymbol{\beta} < \infty$, where $\Omega \in \mathbb{R}^p$ is the parametric space of $\boldsymbol{\beta}$.

Define $\dot{\boldsymbol{\Sigma}}_k$ as the matrix resulting from $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ by differentiating each of its entries with respect to the k-th component of $\boldsymbol{\beta}$. Paulo (2005) derived different objective priors for the parameters of a Gaussian process. They are of the form (2.6) with the following choices for a and $\pi(\boldsymbol{\beta})$:

1. **Independence Jeffreys prior:** $a = 1$ and $\pi(\beta) \propto |\boldsymbol{I_J}(\boldsymbol{\beta})|^{1/2}$.

2. **Jeffreys-rule prior:** $a = 1 + \frac{q}{2}$ and $\pi(\beta) \propto |\boldsymbol{I_J}(\boldsymbol{\beta})|^{1/2} |\boldsymbol{X}^t \boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}) \boldsymbol{X}|^{1/2}$.

3. **Reference prior:** $a = 1$ and $\pi(\beta) \propto |\boldsymbol{I_R}(\boldsymbol{\beta})|^{1/2}$.

where

$$I_J(\boldsymbol{\beta}) \quad = \quad \begin{pmatrix} n & \mathrm{tr}\mathbf{U}_1 & \mathrm{tr}\mathbf{U}_1 & \cdots & \mathrm{tr}\mathbf{U}_p \\ & \mathrm{tr}\mathbf{U}_1^2 & \mathrm{tr}\mathbf{U}_1\mathbf{U}_2 & \cdots & \mathrm{tr}\mathbf{U}_1\mathbf{U}_p \\ & & \ddots & \cdots & \vdots \\ & & & & \mathrm{tr}\mathbf{U}_1^2 \end{pmatrix}$$

and

$$I_R(\boldsymbol{\beta}) = \begin{pmatrix} n-q & \mathrm{tr}\mathbf{W}_1 & \mathrm{tr}\mathbf{W}_1 & \cdots & \mathrm{tr}\mathbf{W}_p \\ & \mathrm{tr}\mathbf{W}_1^2 & \mathrm{tr}\mathbf{W}_1\mathbf{W}_2 & \cdots & \mathrm{tr}\mathbf{W}_1\mathbf{W}_p \\ & & \ddots & \cdots & \vdots \\ & & & & \mathrm{tr}\mathbf{W}_1^2 \end{pmatrix}$$

with $\mathbf{W}_k = \dot{\boldsymbol{\Sigma}}_k Q$ and $\mathbf{U}_k = \dot{\boldsymbol{\Sigma}}_k \boldsymbol{\Sigma}^{-1}$.

Paulo (2005) showed that, under reasonable assumptions such as separability of the correlation function and Cartesian product of the design set, the reference, the independence Jeffreys and the Jeffreys-rule priors yield proper posteriors. Although the reference prior is slightly more computationally expensive to calculate (we need to compute $\mathbf{W}_k$'s instead of $\mathbf{U}_k$'s), we recommend use of the reference prior as default prior for the analysis of Gaussian process models. Berger *et al.* (2001) also recommended the use of reference priors, observing that it always yields proper posteriors and credible intervals with satisfactory frequentist coverage when $\boldsymbol{\beta}$ is one-dimensional and the correlation function is isotropic.

An alternative to the above default priors is the proper flat prior based on an empirical Bayes approach (Paulo 2005). These proper priors are independent Exponential distributions centered at a multiple (e.g. 10) of the MLE's; the corresponding density function is very flat in the region of the parametric space where the likelihood has most of its mass, so we expect the prior to have very small effect on the posterior. Paulo (2005) showed that the choice of the multiplicative constant have very little impact in the analysis. However, we have observed some examples where the proper flat prior is not as vague as one would expect. Figure 2.1 provides a visual comparison between the impact of the empirical Bayes prior and the reference prior on their corresponding posterior distribution using the damped sine wave example with $n = 10$ (this example is going to be presented in details in Section 2.4). The prior density of β shows an extreme resemblance to the resulting posterior density in

the empirical Bayes approach (left panel of Figure 2.1), which is not observed when the reference prior is used (right panel). In the following chapters, we have several examples where proper flat priors are used, but we do recommend some caution when using them as vague priors.

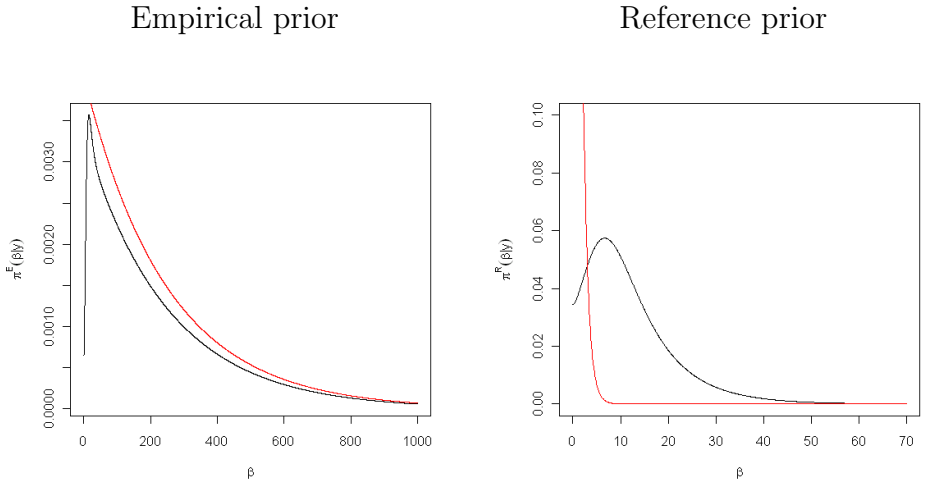


FIGURE 2.1: Comparison between prior (red curves) and posterior (black curves) distributions of the correlation parameter β using the damped sine wave example with $n = 10$. Left panel corresponds to empirical Bayes prior and posterior and right panel corresponds to regerence prior and posterior.

2.3.3 A special case

There is a special case when the objective priors can be obtained in closed form as a function of the GASP parameters. This special case is defined by the following assumptions: the input space is one-dimensional with data points $\boldsymbol{z}_1, \cdots, \boldsymbol{z}_n$ being equally spaced and ϵ units apart, the mean of the GASP is a constant ($\boldsymbol{X} = \mathbf{1}$ and $\boldsymbol{\theta} = \theta$), and the roughness correlation parameter α in (2.2) equals 1 (i.e., the power-exponential correlation function reduces to a simple exponential correlation function). Let $\rho = e^{-\epsilon\beta}$. Under these assumptions, the integrated likelihood in (2.7) becomes

$$L^I(\beta|\mathbf{y}) \propto [(1-\rho)(n-(n-2)\rho)]^{-1/2}\left\{\frac{1}{1-\rho^2}\right\}^{a-3/2}\left\{\sum_{i=1}^{n} y_i^2 - 2\rho\left(\sum_{i=1}^{n-1} y_i y_{i+1}\right)\right.$$
$$+\rho^2\left(\sum_{i=2}^{n-1} y_i^2\right) - \frac{1-\rho}{n-(n-2)\rho}\left[\sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j - 2\rho\left(\sum_{i=1}^{n}\sum_{j=1}^{n-1} y_i y_j\right)\right.$$
$$\left.\left.+\rho^2\left(\sum_{i=2}^{n-1}\sum_{j=2}^{n-1} y_i y_j\right)\right]\right\}^{-((n-q)/2+a-1)} \tag{2.8}$$

The marginal reference prior for the special case is

$$\pi^R(\beta) \propto \frac{\rho}{1-\rho^2}\left\{\frac{2(n-1)^2(1-\rho)^2}{(n-(n-2)\rho)^2} - 2(n-1)\left[\rho + \frac{1-\rho}{n-(n-2)\rho}\right]^2\right. \tag{2.9}$$
$$\left.+(n-1)(1+\rho^2) - \frac{2}{n-(n-2)\rho}\left[(2n-3)-(2n-1)\rho+\rho^{(n-1)}+\rho^n\right]\right\}^{1/2}$$

and the corresponding marginal reference posterior is

$$\pi^R(\beta|\mathbf{y}) \propto \rho\left[(1-\rho^2)(1-\rho)(n-(n-2)\rho)\right]^{-1/2} \times \Big\{(n-1)(1+\rho^2) \quad (2.10)$$

$$-\frac{2}{n-(n-2)\rho}\left[(2n-3)-(2n-1)\rho+\rho^{(n-1)}+\rho^n\right]$$

$$+\frac{2(n-1)^2(1-\rho)^2}{(n-(n-2)\rho)^2}-2(n-1)\left[\rho+\frac{1-\rho}{n-(n-2)\rho}\right]^2\Bigg\}^{1/2}$$

$$\times\left\{\sum_{i=1}^{n}y_i^2-2\rho\left(\sum_{i=1}^{n-1}y_iy_{i+1}\right)+\rho^2\left(\sum_{i=2}^{n-1}y_i^2\right)\right.$$

$$-\frac{1-\rho}{n-(n-2)\rho}\left[\sum_{i=1}^{n}\sum_{j=1}^{n}y_iy_j-2\rho\left(\sum_{i=1}^{n}\sum_{j=1}^{n-1}y_iy_j\right)\right.$$

$$\left.\left.+\rho^2\left(\sum_{i=2}^{n-1}\sum_{j=2}^{n-1}y_iy_j\right)\right]\right\}^{-(n-1)/2}$$

See Appendix A for proof. In Appendix A we also show the Jeffreys priors and posterior densities in closed form for the special case. We know that the assumptions of the special case are quite strong and not practical in many applications, but this special case is important for computational reasons. It does not require matrix algebra computations and is straightforward for both optimization and integration of density functions. We initially analyze the the behavior of the integrated likelihood and the reference posterior in the special case for simplicity and then verify if the same results hold for more realistic assumptions.

2.4 Problems with MLEs and posterior modes

2.4.1 The damped sine wave example

Consider the following synthetic computer model in one dimension:

$$y(z) = 10\sin(4\pi z^{0.9})\exp(-1.5z)$$

This function is an example of damped sine wave. We now observe the function at n equally-spaced points on $[0, 1]$ and we want to fit a GASP emulator to these observations based on the plug-in method. Figure 2.2 shows the predictions obtained by plugging maximum likelihood estimates into the predictuve distribution in (2.3) for $n = 10$ and $n = 9$.

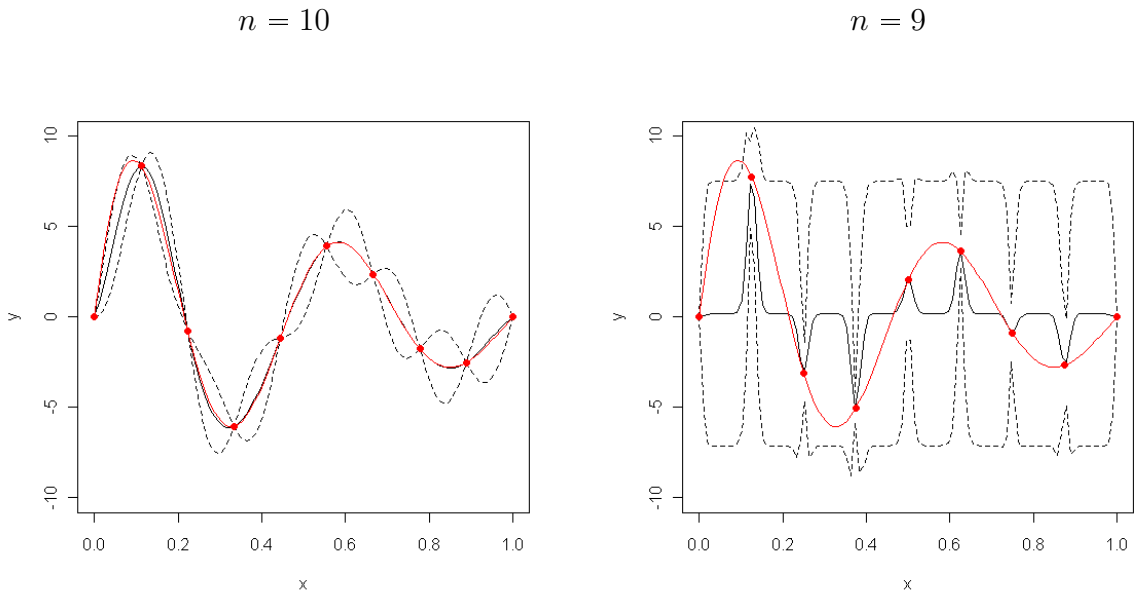


FIGURE 2.2: Plug-in GASP approximation to the damped sinewave using $n = 10$ (left panel) and $n = 9$ (right panel) equally-spaced data points and the MLE plug-in approach. Red solid line is the true value of the function, black and dashed solid lines are, respectively, the approximation and 95% confidence bounds from the predictive distribution.

We can clearly see drastic differences between the two predictions, although the computer model and the emulator are the same and the number of points are close to each other. Table 2.1 presents the values $\hat{\mu}, \hat{\sigma}^2$ and $\hat{\beta}$ that maximize the likelihood function in (2.5), for $n = 10$ and $n = 9$. α is fixed at 1.9. The main difference between the sets of MLE's for the two data sets is on the value of $\hat{\beta}$, the maximum likelihood estimate of the range correlation parameter. Because the estimate of β for $n = 9$ is extremely large, the plug-in approach gives unrealistic predictions as shown

in Figure 2.2: the approximation is constant with peaks at the data points and the confidence bounds are overconservative. Note that optimization algorithms usually return finite large values due to a tolerance stopping criterion; further analysis of the likelihood function indeed shows that the likelihood is maximized at infinity.

Table 2.1: MLE values for the damped sine wave example.

n	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\beta}$	α
10	0.107	15.59	74.95	1.9
9	0.197	13.34	4011.63	1.9

The above example shows how sensitive the MLE can be to slight changes to the data. We now focus on the problem of finding numerical estimates for the GASP correlation parameters $\boldsymbol{\beta}$. As shown in Equations (2.3) and (2.4), once we have the estimate of $\boldsymbol{\beta}$, we can easily obtain numerical estimates of σ^2 and $\boldsymbol{\theta}$, the variance and the regression parameters of the Gaussian process, respectively. For illustration, Figure 2.3 shows the MLEs of θ and σ^2, given in (2.3), as functions of β using the damped sine wave example. Note that both curves have asymptotes, because the larger the range correlation parameter is, the closer the Gaussian process gets to a process of independent variables with constant common mean and variance.

The estimation of β is particularly challenging because the likelihood is a complex function of the parameter at the same time that the value of β has a huge impact on the model. We deal with this problem by marginalizing likelihood and posterior distributions over $\boldsymbol{\theta}$ and σ^2. Berger *et al.* (1999) propose the use of the mode of the integrated likelihood as an estimate of the parameter in likelihood analysis and provide some examples where this estimate is consistent, but the MLE is not. Paulo (2005) states that there empirical evidence shows that integrated likelihood modes tend to be more stable and more meaningful than maximum likelihood estimates. Berger *et al.* (2001) show that, in the unidimensional case, $L^I(\beta|\mathbf{y}) \to c_0 > 0$ as

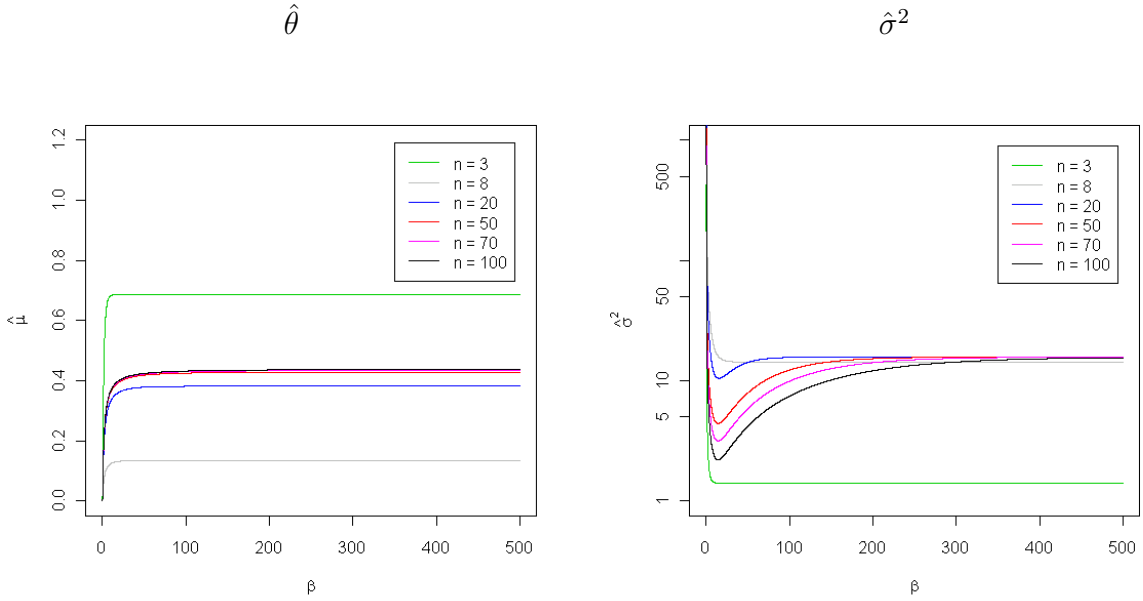


FIGURE 2.3: MLEs of θ and σ^2 as functions of β and n.

$\beta \to 0$. In the special case where the integrated likelihood can obtained in closed form, as shown in (2.8), the limits of $L^I(\beta|\mathbf{y})$ are:

$$\begin{array}{rcll} L^I(\beta|\mathbf{y}) & \to & \left\{\sum_{i=1}^{n-1}(y_{i+1}-y_i)^2\right\}^{-(n-1)/2} & \text{as} \quad \beta \to 0 \\ L^I(\beta|\mathbf{y}) & \to & \frac{1}{\sqrt{n}}\left\{\sum_{i=1}^{n}(y_i-\bar{y})^2\right\}^{-(n-1)/2} & \text{as} \quad \beta \to \infty \end{array} \tag{2.11}$$

In the simple case where data points are equally spaced and the roughness parameter is one, the integrated likelihood approaches nonzero constants when β goes to zero or infinity. Figure 2.4 shows integrated likelihood of β using the damped sine wave example in this special case for different values of n. The most intriguing conclusion from the several plots in the figure is that for some values of n the mode of the integrated likelihood is finite, and for others it is infinite. This is similar to what Table 2.1 tell us about the likelihood function for the 9 observations of the damped sine wave at equally-spaced points (once again, the optimization procedure returns finite large values when the integrated likelihood is maximized at infinity).

small n medium n several n's

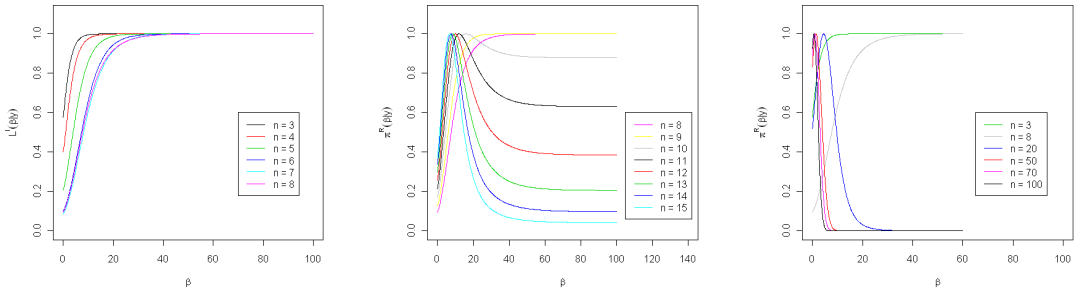


FIGURE 2.4: Integrated likelihood using the damped sine wave example at equally-spaced points for different ranges of n and $\alpha = 1$ (special case where the function can be obtained in closed form).

The limits in (2.11) show that the mode is infinite when $\sum_{i=1}^{n-1} (y_{i+1} - y_i)^2$ is very large. This statistic is inversely related to the first-order autocorrelation of $\mathbf{y}$, thus the integrated likelihood mode is infinite when our empirical function $y(z)$ is relatively wiggly (corresponding to a negative autocorrelation). One may say that this is a good feature once the Gaussian process model assumes positive correlations among observations at different points; however, when generating a small number of observations from a Gaussian process model, it is usually very common to obtain a vector $\mathbf{y}$ with negative autocorrelation, even though correlations are modeled as being positive.

We also analyze the reference posterior distribution given by (2.10). Figure 2.5 shows the reference posterior density of β using the damped sine wave example and different values of n. Note how the shape of the reference posterior changes drastically as we increase n. Because of the propriety of the reference posterior, its mode must be finite. However, for small n we observe that the reference posterior mode is zero. Unfortunately, zero estimates (or other values near zero) cannot be used in the plug-in method because the correlation matrix becomes computationally singular and the model is not well-defined.

small n medium n several n's

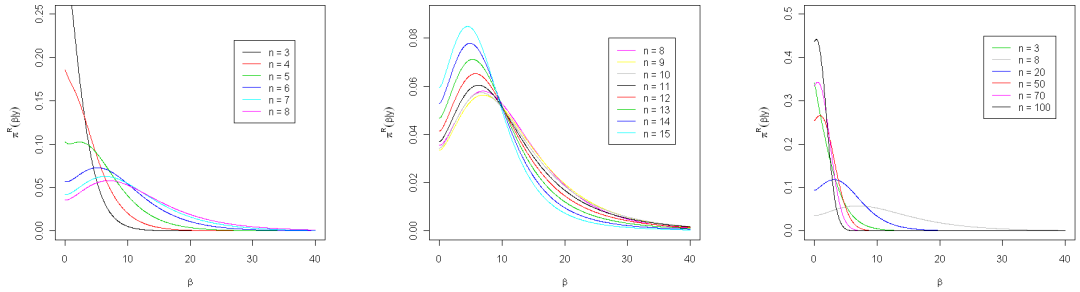


FIGURE 2.5: Reference posterior density using the damped sine wave example at equally-spaced points for different ranges of n and $\alpha = 1$ (special case where the function can be obtained in closed form).

2.4.2 Estimating β from simulated data

Next, we study the performance of MLEs and posterior reference modes when data come from the model in study. We generate one-dimensional equally-spaced data from a Gaussian process on $[0, 1]$ with $n = 10$, $\mu = 0$, $\sigma^2 = 1$, and $\beta = 1$. We are going to work with two different simulated data sets: one coming from a GASP model with $\alpha = 1$ and other from a GASP model with $\alpha = 1.9$. Either way, the roughness parameter is always assumed to be known during the estimation of other parameters. Figure 5.11 illustrates some realizations from this process.

Table 2.2 shows a summarized comparison between integrated likelihood mode and reference posterior mode for the 20,000 generated realizations. We maximize the unidimensional functions in (2.8) and (2.10). Note that even when data come from the GASP model, the integrated likelihood can be maximized at zero or at infinity. The reference posterior density cannot be maximized at infinity because the reference posterior is proper. However, the proportion of zeros is higher for the reference posterior mode than for the integrated likelihood mode. This is somehow expected because the reference prior density in (2.9) for the $n = 10$ is a monotonically decreasing function of β.

$\alpha = 1$ $\alpha = 1.9$

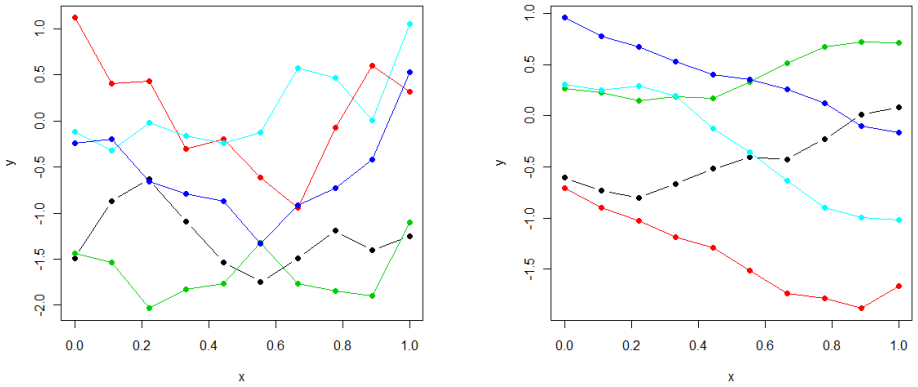


FIGURE 2.6: Realizations from Gaussian process model in one dimension with $\mu = 0$, $\sigma^2 = 1$, $\beta = 1$ and $\alpha = 1$ (left) or 1.9 (right).

Table 2.2: Integrated likelihood mode versus reference posterior mode for the simulated GASP realizations with $\alpha = 1$.

		Integrated Likelihood Mode			
		$\hat{\beta} = 0$	$0 < \hat{\beta} < \infty$	$\hat{\beta} = \infty$	Total
Posterior Mode	$\tilde{\beta} = 0$	24.5%	36.2%	0.2%	60.9%
	$0 < \tilde{\beta} < \infty$	0%	33.9%	5.2%	39.1%
	Total	24.5%	70.1%	5.4%	100%

The left panel of Figure 2.7 shows log-log plots of MLEs versus reference posterior modes of β estimated from different realizations of 10 observations from a GASP with $\mu = 0$, $\sigma^2 = 1$, $\beta = 1$ and $\alpha = 1$. Red circles correspond to realizations whose observed autocorrelation is negative. For 24.5% of the realizations, we obtained a reference posterior mode at zero, and the right panel of Figure 2.7 shows the corresponding MLE values. Maximum likelihood estimation does not return zero values, but, similarly to the maximization of integrated likelihood function, the likelihood function is also maximized at infinity when a negative autocorrelation is observed.

Figure 2.7 shows a cloud of red circles at a very large values of β because our optimization procedure stops due to a tolerance criterion, but further analysis of the likelihood function reveals that the function is approximately flat when β is big.

Non-zero modes vs
finite Non-zero MLEs

Finite non-zero MLEs corr.
to zero modes

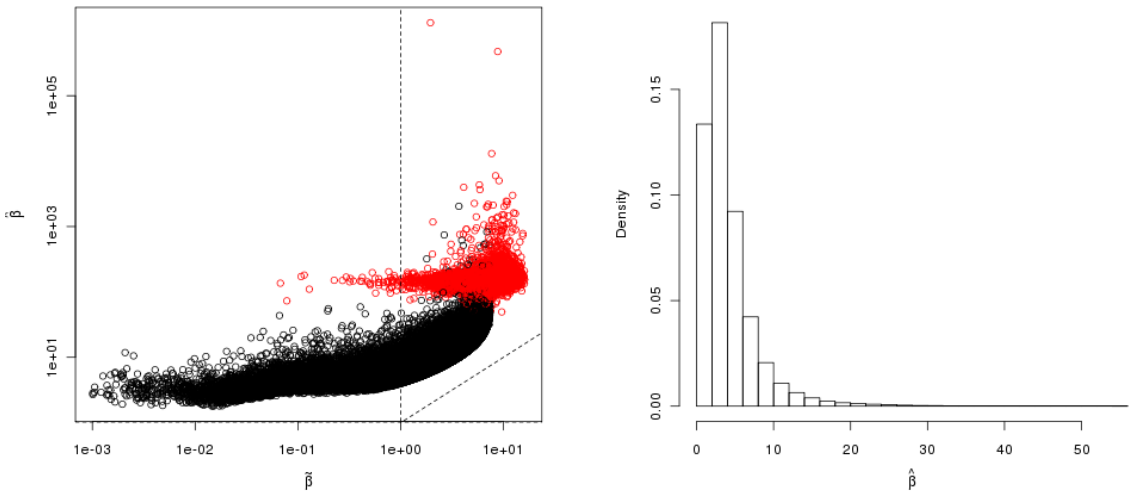


FIGURE 2.7: Left: Plot on log-log scale of reference posterior mode versus MLE of β calculated from realizations of a Gaussian process model with $n = 10$, $\mu = 0$, $\sigma^2 = 1$, $\beta = 1$, and $\alpha = 1$. Dashed lines correspond to the true value of the parameter and equality line. Red circles correspond to realizations with negative observed autocorrelation. Right: The reference posterior distribution is maximized at zero for some realizations; the histogram shows the corresponding MLE values for those realizations.

The problem of infinite MLEs and reference posterior modes is not restricted to the simple case where $\alpha = 1$. When we generate realizations from a Gaussian process model with larger α, we expect to see smoother empirical curves and a smaller number of realizations with negative autocorrelation. However, the extremely large values of MLE are still present even when the roughness parameter is close to 2. Figure 2.8 shows pairs of MLEs and reference posterior modes when we generate realizations and estimate parameters form the GASP model with $\alpha = 1.9$. The reference posterior modes are much closer to the true value of the parameter, but for many realizations,

the reference posterior density is maximized at zero with no secondary mode.

Non-zero modes vs
finite Non-zero MLEs

Finite non-zero MLEs corr.
to zero modes

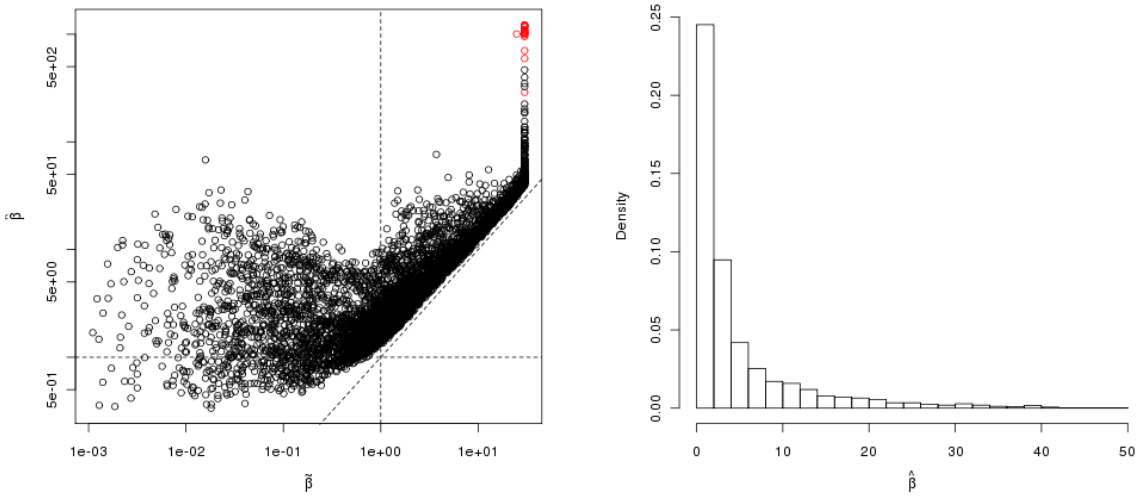


FIGURE 2.8: Left: Plot on log-log scale of reference posterior mode versus MLE of β calculated from realizations of a Gaussian process model with $n = 10$, $\mu = 0$, $\sigma^2 = 1$, $\beta = 1$, and $\alpha = 1.9$. Dashed lines correspond to the true value of the parameter and equality line. Red circles correspond to realizations with negative observed autocorrelation. Right: The reference posterior distribution is maximized at zero for some realizations; the histogram shows the corresponding MLE values for those realizations.

Recall that the plug-in approach is based on the use of numerical estimates of the correlation parameter in posterior predictive distributions, such as in (2.3) and (2.4). If the estimate is close to zero, we will not be able to numerically invert the correlation matrix Σ_β; if it is infinite, we will not be modeling the computer model output as an continuous function of the inputs. We need a point estimation method that does give us any of these undesirable values. Note that the plug-in approach is mainly used as a black box method: the analyst is not interested in studying why the likelihood or the reference posterior for his/her observations are not well-behaved, because the parameters of the emulator do not have a clear interpretation. The main goal here is to produce fast predictions from a flexible model that allows the analyst

to make probability statements from the model.

Up to this point, we have focused our attention on the case where the input space is unidimensional. Paulo (2005) discuss how many of the problems with integrated likelihood and default priors observed in one dimension are not common to the multidimensional case. We postpone the discussion of the multidimensional case to the next section, where we propose better point estimates based on the reference posterior.

2.5 Laplace approximation to posterior numerical summaries

2.5.1 Introduction

In the previous section, we analyzed the performance of posterior modes in the estimation of the GASP correlation parameter. Now, instead of maximizing the reference posterior density, we propose the use of estimates based on the integration of the function, such as means and medians. However, exact calculation of these estimates through numerical integration is usually unstable and computationally expensive compared to optimization methods, due to the matrix algebra involved in the general formulation. And, obviously, MCMC methods are better suited to the full Bayes tan to the plug-in approach.

We propose the use of the Laplace approximation method (de Brujin 1970, sec. 4.4; Tierney and Kadane 1986). This method consists of approximating posterior expectations using a transformation of the parameters and the Taylor expansion of the logarithm of the integrands in the expectation formula. The posterior density of the logarithm of the parameters is then approximated with a Gaussian distribution centered at its mode and with covariance matrix given by the curvature of the posterior density at the mode. Hence, this approximation works best when the posterior distribution of the transformed parameters is unimodal and highly concentrated around the posterior mode.

2.5.2 Formulation

Let f be a smooth positive function on the parametric space. Set $\gamma = \log \beta$. Then

$$\mathbb{E}f(\beta) \;=\; \mathbb{E}f(e^{\gamma}) = \frac{\int_{-\infty}^{\infty} f(e^{\gamma}) g(\gamma|y)\, d\gamma}{\int_{-\infty}^{\infty} g(\gamma|y)\, d\gamma} \tag{2.12}$$

where the posterior density of γ is given by $\pi(\gamma|y) \propto g(\gamma|y)$.

Set $l(\gamma|y) = \log g(\gamma|y)$ and let $\tilde{\gamma}_l = \arg\max l(\gamma|y)$ be the posterior mode of γ. Mosteller and Wallace (1964, sec. 4.6C) expanded both numerator and denominator integrands around the posterior mode $\tilde{\gamma}_l$:

$$\begin{aligned} \mathbb{E}f(\beta) \;&\approx\; \frac{\int_{-\infty}^{\infty} f(e^{\gamma}) \exp\left\{ l(\tilde{\gamma}_l|y) - \frac{1}{2\sigma^2}(\gamma - \tilde{\gamma}_l)^2 \right\} d\gamma}{\int_{-\infty}^{\infty} \exp\left\{ l(\tilde{\gamma}_l|y) - \frac{1}{2\sigma_l^2}(\gamma - \tilde{\gamma}_l)^2 \right\} d\gamma} \\ &=\; \int_{-\infty}^{\infty} f(e^{\gamma}) \frac{\exp\left\{ -\frac{1}{2\sigma_l^2}(\gamma - \tilde{\gamma}_l)^2 \right\}}{\int_{-\infty}^{\infty} \exp\left\{ -\frac{1}{2\sigma_l^2}(\gamma - \tilde{\gamma}_l)^2 \right\} d\gamma} d\gamma \end{aligned} \tag{2.13}$$

where $\sigma_l^2 = -\frac{1}{l''(\tilde{\gamma}_l|y)}$. If we use $f(\beta) = \beta$, then we can approximate the posterior mean of β by:

$$\mathbb{E}\beta \;\approx\; \exp\left\{ \tilde{\gamma}_l + \frac{1}{2}\sigma_l^2 \right\} \tag{2.14}$$

Similarly, in the multivariate case, for any vector $\boldsymbol{t}$:

$$\mathbb{E}e^{\boldsymbol{t}'\boldsymbol{\gamma}} \;\approx\; \exp\left\{ \boldsymbol{t}'\tilde{\boldsymbol{\gamma}}_l + \frac{1}{2}\boldsymbol{t}'\boldsymbol{\Sigma}_l\boldsymbol{t} \right\} \tag{2.15}$$

where $\boldsymbol{\Sigma}$ is minus the inverse Hessian of $l(\boldsymbol{\gamma}|y)$ at $\tilde{\boldsymbol{\gamma}}_l$.

Now let $h(\gamma|y) = f(\gamma) + l(\gamma|y)$ and $\tilde{\gamma}_h = \arg\max h(\gamma|y)$. Tierney and Kadane (1986) proposed the expansion of the numerator and denominator integrands at their respective modes:

$$\begin{aligned}
\mathbb{E}\beta &= \mathbb{E}e^{\gamma} = \frac{\int_{-\infty}^{\infty} e^{\gamma} g(\gamma|y)\, d\gamma}{\int_{-\infty}^{\infty} g(\gamma|y)\, d\gamma} = \frac{\int_{-\infty}^{\infty} e^{\gamma + l(\gamma|y)}\, d\gamma}{\int_{-\infty}^{\infty} e^{l(\gamma|y)}\, d\gamma} \\
&\approx \frac{\int_{-\infty}^{\infty} \exp\left\{h(\tilde{\gamma}_h|y) - \frac{1}{2\sigma_h^2}(\gamma - \tilde{\gamma}_h)^2\right\} d\gamma}{\int_{-\infty}^{\infty} \exp\left\{l(\tilde{\gamma}_l|y) - \frac{1}{2\sigma_l^2}(\gamma - \tilde{\gamma}_l)^2\right\} d\gamma} \\
&\approx \exp\{h(\tilde{\gamma}_h|y) - l(\tilde{\gamma}_l|y)\} \frac{\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2\sigma_h^2}(\gamma - \tilde{\gamma}_h)^2\right\} d\gamma}{\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2\sigma_l^2}(\gamma - \tilde{\gamma}_l)^2\right\} d\gamma} \\
&= \sqrt{\frac{\sigma_h^2}{\sigma_l^2}} \exp\{\tilde{\gamma}_h + l(\tilde{\gamma}_h|y) - l(\tilde{\gamma}_l|y)\} \qquad (2.16)
\end{aligned}$$

where $\tilde{\gamma}_l = \arg\max l(\gamma|y)$, $\sigma_h^2 = -\frac{1}{h''(\tilde{\gamma}_h|y)}$ and $\sigma_l^2 = -\frac{1}{l''(\tilde{\gamma}_l|y)}$. Also, we have in the multivariate case:

$$\mathbb{E}e^{\boldsymbol{t}'\boldsymbol{\gamma}} \approx \frac{\det(\boldsymbol{\Sigma}_h)}{\det(\boldsymbol{\Sigma}_l)} \exp\{\boldsymbol{t}'\tilde{\boldsymbol{\gamma}}_h + l(\tilde{\boldsymbol{\gamma}}_h|y) - l(\tilde{\boldsymbol{\gamma}}_l|y)\} \qquad (2.17)$$

where $\boldsymbol{\Sigma}_h$ and $\boldsymbol{\Sigma}_l$ are minus the inverse Hessians of h and l at $\tilde{\boldsymbol{\gamma}}_h$ and $\tilde{\boldsymbol{\gamma}}_l$, respectively.

We also propose the transformed posterior mode in $\log\beta$ space as an estimator of β:

$$\beta^* \approx \exp\{\tilde{\gamma}\} \qquad (2.18)$$

This is also the posterior mode in $\log\beta$ space. Heuristic arguments associate this estimator with the posterior median of β. Gu (1992) and Nagy *et al.* (2007),

for example, use the Laplace's method to approximate posterior distribution of non-Gaussian models and suggest the use of the approximate distribution to construct Bayesian credible intervals.

2.6 Results

2.6.1 The univariate case

We now return to the analysis of the simulated data presented in Subsection 2.4.2. Two data sets were generated (one with $\alpha = 1$ and other with $\alpha = 1.9$) from an one-dimensional Gaussian process at $n = 10$ equally-spaced points on $[0, 1]$. We are interested in evaluating the performance of the Laplace approximations compared to the maximum likelihood estimator. Recall that the true value of parameters used in the simulation are $\mu = 0$, $\sigma^2 = 1$, and $\beta = 1$.

Figure 2.9 shows plots of the Laplace approximation for the posterior mean of β in (2.14) compared to the value obtained from numerical integration (first column) and to the MLE (second column). When $\alpha = 1$ (first row) we can obtain the reference posterior in closed form for equally-spaced data points, as shown in (2.10). Unfortunately, applying the Laplace method to the case where $\alpha = 1.9$ sometimes produces unstable solutions, because we need to maximize the general formula for the posterior distribution. The first column of Figure 2.9 give us an idea of how good the approximation in (2.14). We observe a positive bias in the approximation for $\alpha = 1$, but not for $\alpha = 1.9$. As seen before, maximum likelihood estimation sometimes produces extremely large values, which are not convenient for the analysis.

Figure 2.10 shows similar plots, but now regarding the approximation for the posterior mean of β in (2.16). When we compare this approximation to the values obtained from numerical integration of the reference posterior, we can see that the Laplace approximation in (2.16) approximates the posterior expectation of β better than the approximation in (2.14) and it is also more stable. The Laplace approxi-

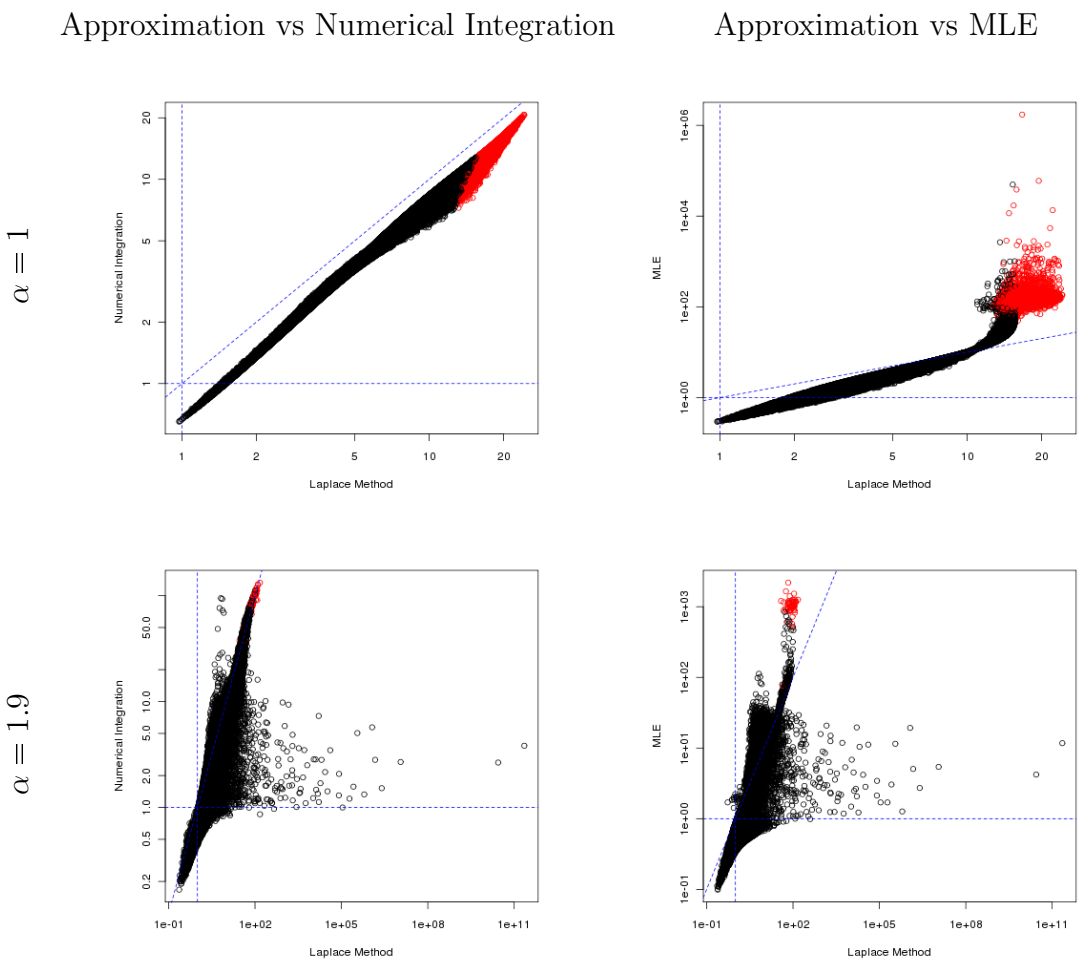


FIGURE 2.9: Comparison of the reference posterior mean approximation in (2.14) to the numerical integration value (first column) and to the MLE (second column) for data sets generated from unidimensional GASP model with $\alpha = 1$ (first row) and $\alpha = 1.9$ (second row) at equally-spaced points. Red circles correspond to realizations with negative observed autocorrelation. Blue dashed lines are represent the true value of the parameter and the equality line.

mation tends to give more reasonable estimates than the MLE.

Plots for the transformed posterior mode in $\log \beta$ space in (2.18) are shown in Figure 2.11. This estimator is extremely simple, corresponding to the exponentiation of the posterior mean of γ, the logarithm of β, it is also very efficient. It may not be as good as an approximation to the posterior median as the approximation in (2.16) is to the the posterior mean, but it provides quite good results without the

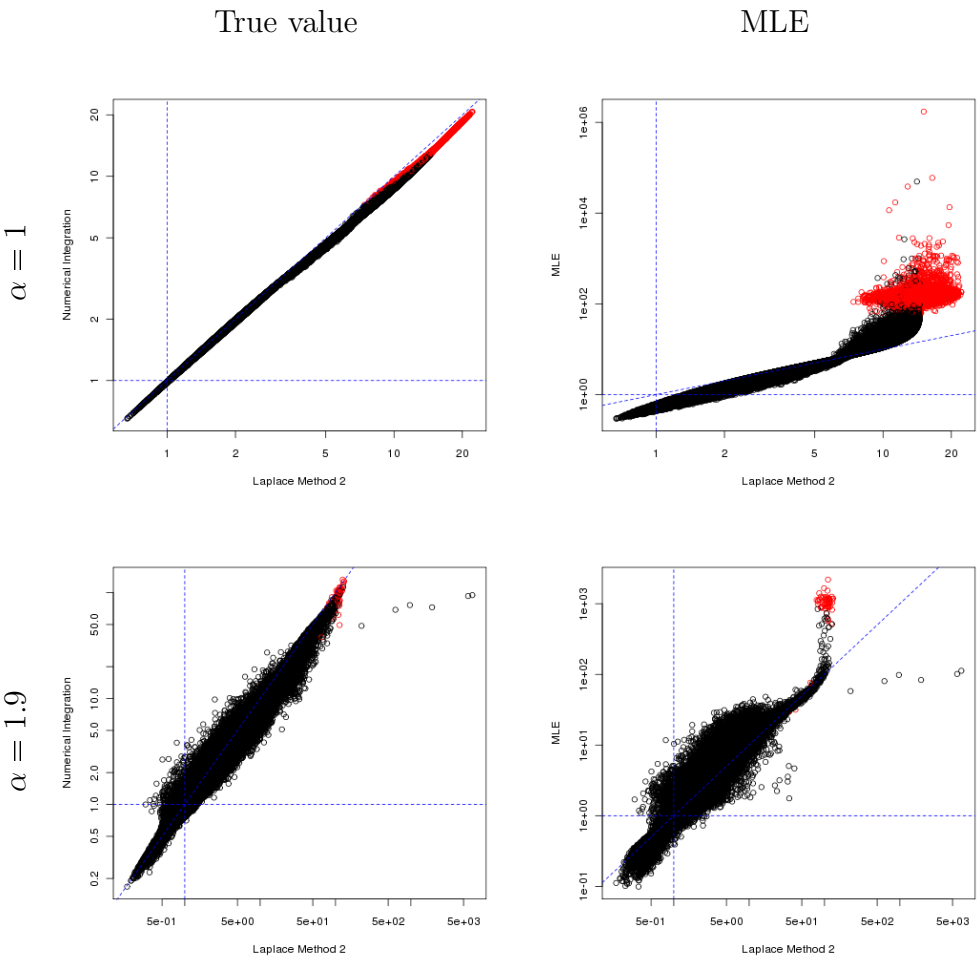


FIGURE 2.10: Comparison of the reference posterior mean approximation in (2.16) to the numerical integration value (first column) and to the MLE (second column) for data sets generated from unidimensional GASP model with $\alpha = 1$ (first row) and $\alpha = 1.9$ (second row) at equally-spaced points. Red circles correspond to realizations with negative observed autocorrelation. Blue dashed lines are represent the true value of the parameter and the equality line.

need of second derivatives. Also, there are few cases where the estimator in (2.16) is extremely unstable, whereas the transformed posterior mode never is (presumably because of the second derivative issue). For both cases $\alpha = 1$ and $\alpha = 1.9$ the approximation is very close to the numerical integration value. The transformed posterior mode also seems to be an excellent alternative to MLEs.

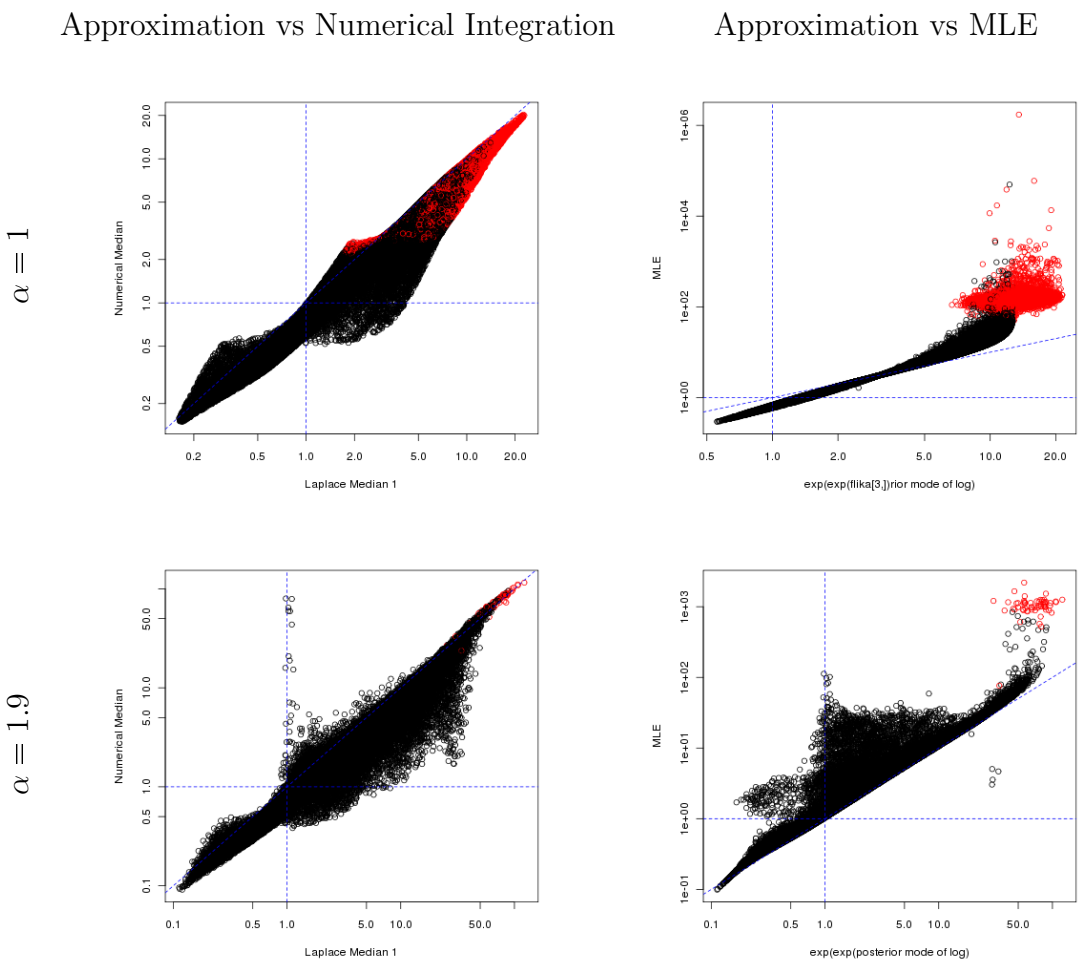


FIGURE 2.11: Comparison of the reference transformed posterior mode in (2.18) to the numerical integration value (first column) and to the MLE (second column) for data sets generated from unidimensional GASP model with $\alpha = 1$ (first row) and $\alpha = 1.9$ (second row) at equally-spaced points. Red circles correspond to realizations with negative observed autocorrelation. Blue dashed lines are represent the true value of the parameter and the equality line.

2.6.2 The multivariate case

We evaluate the performance of Laplace approximations to posterior means using data simulated from a bidimensional GASP model. We generate several realizations of a Gaussian process model on the 20-point maxmin Latin Hypercube Design (LHD) on $[0, 1] \times [0, 1]$ illustrated by Figure 2.12. LHDs are constructed by selecting points in the input space according to a space-filling strategy. The true values of the GASP

parameters are $\theta = 0$, $\sigma^2 = 1$, $\beta_1 = 1$, $\beta_2 = 1$ and $\alpha = 1.9$. We are interested in calculating the reference posterior expectation of β_1 and β_2 jointly, using the Laplace approximations presented in Section 2.5. We will also compare these approximate values to the values obtained from numerical integration and to the MLEs. We use adaptive integration over hypercubes to calculate the two integrands in (2.12); since the posterior distribution was truncated inside a large hypercube, our numerical integration values are also approximations to the exact expectation.

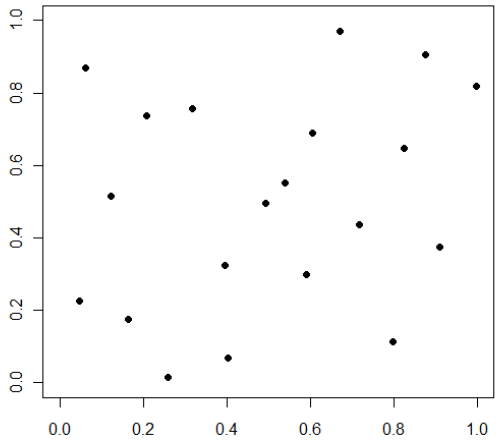


FIGURE 2.12: 20-point Latin Hypercube Design used to generate observations from GASP model in the multivariate case.

Figure 2.13 compares the values of the Laplace approximation in (2.15) to the values from numerical integration and maximum likelihood estimation. For both range correlation parameters β_1 and β_2, we observe that the value of the Laplace approximation is usually very close to the one from numerical integration. The values of the Laplace approximation are usually much closer to the true values of the parameters than the MLEs, as we observed in the univariate case. However, instead of getting extremely large MLEs, in the multivariate example we observe MLE values very close to zero.

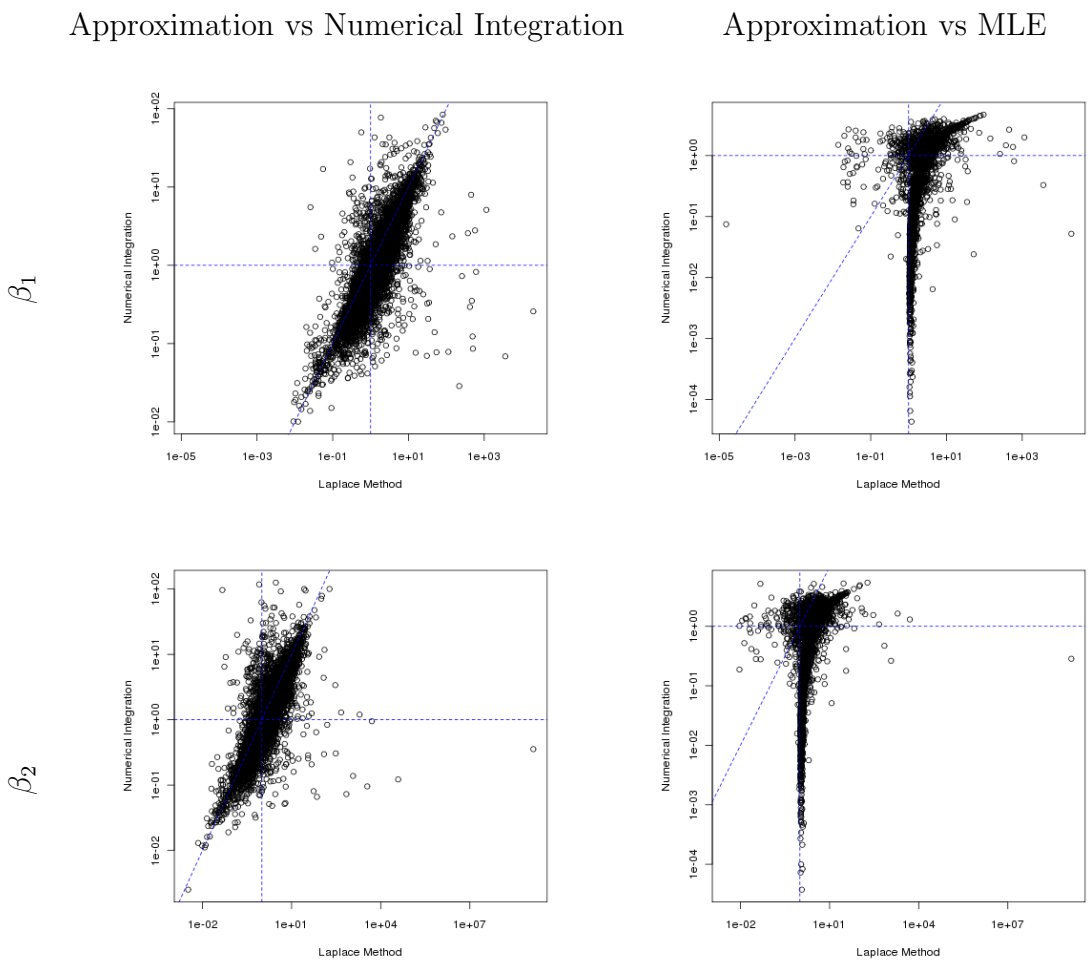


FIGURE 2.13: Comparison of the reference posterior mean approximation in (2.15) to the numerical integration value (first column) and to the MLE (second column) for data sets generated from bidimensional GASP model with $\alpha = 1.9$ at equally-spaced points. Each row represent a different range correlation parameter. Red circles correspond to realizations with negative observed autocorrelation. Blue dashed lines are represent the true value of the parameter and the equality line.

Figure 2.14 shows similar comparisons for posterior mean approximation in (2.16). As we have noted in the univariate case, this approximation seems better than the approximation in (2.15) with respect to proximity to the numerical integration value. The use of different expansions on the numerator and denominator integrands of (2.12) represents a significant improvement in the approximation. From our examples, we have empirical evidence that the reference posterior mean gives very

reasonable estimated values for the parameters, avoiding situation where predictions based on the plug-in approach are unrealistic, as illustrated in Figure 2.2.

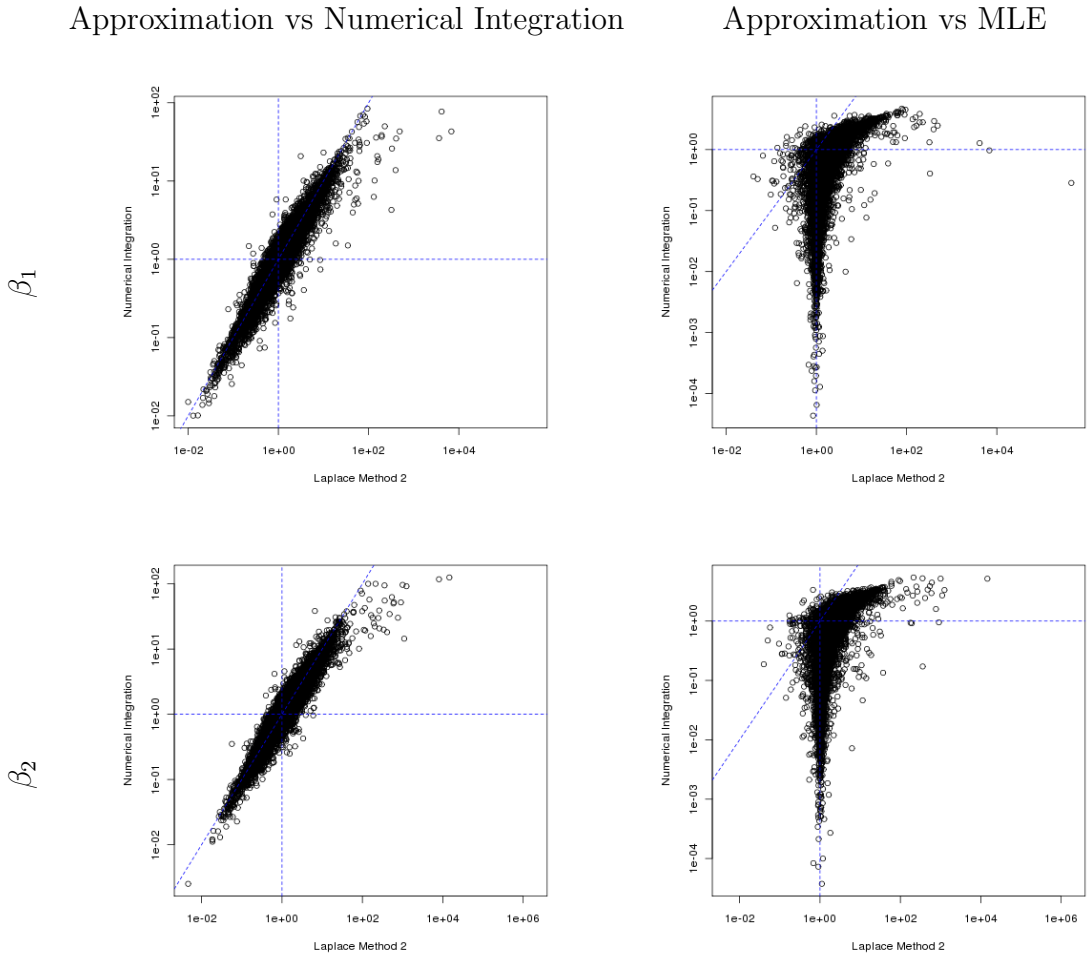


FIGURE 2.14: Comparison of the reference posterior mean approximation in (2.17) to the numerical integration value (first column) and to the MLE (second column) for data sets generated from bidimensional GASP model with $\alpha = 1.9$ at equally-spaced points. Each row represent a different range correlation parameter. Red circles correspond to realizations with negative observed autocorrelation. Blue dashed lines are represent the true value of the parameter and the equality line.

2.7 Discussion

In this chapter, we presented some problems associated with common estimators of GASP correlation parameters, such as MLEs and posterior modes. It is not hard

to find examples where these are either zero (or near zero) or infinite (or extremely large). These values are not useful when applying the widely used plug-in method for emulators.

We have recommended the use of reference prior combined with Laplace approximation to the posterior mean or the transformed posterior mode in $\log \beta$ space. The reference prior yields proper posterior, so we can expect a finite nonzero posterior mean for the correlation parameter, and the Laplace approximation has shown very small error in our applications when compared to the numerical integration of the reference posterior density.

Our major suggestions for point estimators are the posterior mean approximation in (2.16) if the analyst is also interested in accuracy of the estimation, or the transformed posterior mode in (2.18) if the analyst is mostly interested in computational speed. One can also use the Gaussian approximation to the reference posterior to obtain approximate posterior samples, using the FBI method proposed by Nagy *et al.* (2007).

3

Analysis of emulators under computational constraints

3.1 Introduction

Gaussian processes are commonly used in the approximation of deterministic functions, such as computer models. If we imagine the simulator output as a smooth function of its inputs, it is reasonable to assume that output values from computer model runs with similar inputs values are more alike than outputs from runs that are far apart in input space. Therefore, models with a correlation structure that depends on the input values seem very natural.

Historically, there has been some difficulties in the implementation of spatial modeling techniques for large data sets. If we look at a Gaussian process at a finite number of observations, it corresponds to a multivariate normal distribution. If there is a large amount of data, fitting Gaussian process with Markov chain Monte Carlo methods is practically unfeasible, because it involves matrix decompositions (e.g. matrix inversions) whose complexity increases as $O(n^3)$ in the number of locations, n, at every iteration of the MCMC algorithm. See Quiñonero-Candela *et al.* (2007)

for a discussion of this issue.

Computer experiment data sets are usually not as large as data sets from other spatial applications, because of the computational cost associated with running complex simulators. However, with advances in cluster-based computing and parallel computing, we expect that more and more simulator runs become to available in a short period of time. Also, as discussed in Section 3.1, there is a need for faster implementations of emulators, so they can be used in other steps of computer model analysis.

Several approaches have been proposed to address the problem of spatial modeling with large data sets. Many approaches focus on approximating the spatial process realizations with basis functions (Wikle and Cressie, 1999), splines (Lin *et al.*, 2000), moving averages (Higdon, 2001; Ver Hoef *et al.*, 2004) and kernel convolutions (Xia and Gelfand, 2006). Other approaches consist of approximations to the spatial likelihood using spectral methods (Fuentes, 2007; Paciorek, 2007) or conditional distributions (e.g. Vecchia (1988); Stein *et al.* (2004)). There are also methods that approximate the process by a Markov random field when points are observed in a regular grid (Rue and Tjelmeland, 2002; Rue and Held, 2006).

Of note, approaches based on Fourier transforms cannot be applied to our choice of correlation function in (1.1), because the power exponential correlation function with $\alpha < 2$ is not mean square differentiable. Also, we are interested in a closed and directly tractable probabilistic model rather than approximations to the final model.

We want to explore and compare different statistical approaches to the approximation of computer model evaluations, given some computational limits. Modified Predictive Process (Banerjee et al, 2008) is a class of model developed for similar problems in spatial statistics; this class of models does not involve complex matrix decompositions, and has been showing some promising results in the solution of our problem.

In this chapter, we analyze the performance of different methods that improve the speed of the emulation process. These methods are variations of the Gaussian response surface technique and we usually treat them as approximations to the conventional GaSP. Gaussian processes are widely used in the emulation of computer models, but they are computationally expensive in some applications; our alternatives are less computationally expensive, but they are not able to capture all the complexity involved in a GaSP model. Our objective is to verify how good these alternatives are in approximating the conventional Gaussian response surface technique.

This chapter is divided as follows: in Section 3.2, we present the predictive process (Banerjee *et al.*, 2008) and discuss the impact of the choice of the number of knots in the analysis; in Section 3.3, we show different modularization approaches (Liu *et al.*, 2009) and analyze their performances in emulation of computer models; in Section 3.4, we apply these models to a real computer model and compare their results to the ones from a regular GASP.

3.2 Emulators for large data sets

3.2.1 Modified Predictive Process

The best model we have found that give us flexibility in the choice of the correlation function and is directly tractable (in the sense that it is not defined as an approximation to another process) is the Predictive Process, initially proposed by Banerjee *et al.* (2008) and later improved by Finley *et al.* (2009). The basic idea is to define a *parent process* $w(\boldsymbol{z})$ as a Gaussian process specified at a small set of locations, called the the set of *knots* $\mathcal{S}^* = \{\boldsymbol{z}_1^*, \cdots, \boldsymbol{z}_m^*\}$, that may or may not be a subset of our observations. The predictive process $\tilde{w}(\boldsymbol{z})$ is then defined as a projection of $w(\boldsymbol{z})$ using the "kriging" interpolator:

$$\begin{aligned} \boldsymbol{w}^* &= \left[w(\boldsymbol{z}_{\boldsymbol{i}}^*)\right]_{i=1}^m \sim N_m(\mathbf{0}, \sigma^2 \mathcal{C}^{*-1}) \\ \tilde{\boldsymbol{w}} &= \left[\tilde{w}(\boldsymbol{z}_{\boldsymbol{i}})\right]_{i=1}^n = \mathcal{C}^t \mathcal{C}^{*-1} \boldsymbol{w}^* \end{aligned} \tag{3.1}$$

where $\mathcal{C}^* = \left[c(z_i^*, z_j^* | \boldsymbol{\beta})\right]_{i,j=1}^{m,m}$ is $m \times m$ and $\mathcal{C}^t = \left[c(z_i, z_j^* | \boldsymbol{\beta})\right]_{i,j=1}^{n,m}$ is $n \times m$.

Note that the predictive process $\tilde{w}(\boldsymbol{z})$ deterministically interpolates the parent process $w(\boldsymbol{z})$ at the knots. Also, the correlation matrix of $\tilde{\boldsymbol{w}}$ is given by $\mathcal{C}^t \mathcal{C}^{*-1} \mathcal{C}$. Finley *et al.* (2009) define the modified predictive process model as:

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{\Psi\theta} + \mathcal{C}^t \mathcal{C}^{*-1} \boldsymbol{w}^* + \tilde{\boldsymbol{\epsilon}} + \boldsymbol{\epsilon} \\ \tilde{\boldsymbol{\epsilon}} &\sim N_n(\mathbf{0}, \sigma^2 \Sigma_{\tilde{\epsilon}}) \\ \boldsymbol{\epsilon} &\sim N_n(\mathbf{0}, \tau^2 I_n) \end{aligned} \tag{3.2}$$

where $\boldsymbol{y} = (y(\boldsymbol{s}_1), \cdots, y(\boldsymbol{s}_n))^t$ is the $n \times 1$ response vector, $\boldsymbol{\Psi}$ is the $n \times q$ matrix of regressors, $\boldsymbol{\theta}$ is the $q \times 1$ vector of regression coefficients, and $\Sigma_{\tilde{\epsilon}} = I_n - \text{Diag}([\mathcal{C}^t \mathcal{C}^{*-1} \mathcal{C}]_{ii})$.

Banerjee *et al.* (2008) suggest either a regular grid or a subset of observed location for selection of knots. We prefer the latter, because this allows our emulator to interpolate the computer model at the knots (unfortunately, the modified predictive process model cannot interpolate the simulator at the other locations). Finley *et al.* (2008) presents a strategy for optimal knot design, that we will discuss in subsection 3.2.2.

In addition to the regression mean and the predictive process, the modified predictive process in (3.3) is composed by two other processes: $\boldsymbol{\epsilon}$, an independent white noise process, called *nugget*, and $\tilde{\boldsymbol{\epsilon}}$, a process of independent variables with spatially adaptive variances. $\tilde{\boldsymbol{\epsilon}}$ is added to model in order to correct for the fact that $\text{Var}[\tilde{w}(\boldsymbol{z}_i)] = \sigma^2 \mathcal{C}_i^t \mathcal{C}^{*-1} \mathcal{C}_i \neq \sigma^2$.

In computer model analysis, the nugget $\boldsymbol{\epsilon}$ is usually dropped, because of the deterministic nature of the simulator and the absence of measurement error while observing the output value. In this case, the process $\tilde{\boldsymbol{\epsilon}}$ plays an important role in the decomposition of the correlation matrix in the modified predictive process model. From here on, we are going to consider the modified predictive process without nugget, a set of knots that is a subset of our observations and constant mean ($\boldsymbol{\Psi} = \mathbf{1}$ and $\boldsymbol{\theta} = \theta$), we can represent the likelihood function associated with (3.3) as:

$$\boldsymbol{y} = \begin{pmatrix} \boldsymbol{y}^* \\ \tilde{\boldsymbol{y}} \end{pmatrix} \sim N_n \left\{ \begin{pmatrix} \theta\mathbf{1} \\ \theta\mathbf{1} \end{pmatrix}, \sigma^2 \begin{pmatrix} \mathcal{C}^* & \mathcal{C} \\ \mathcal{C}^t & \mathcal{C}^t\mathcal{C}^{*-1}\mathcal{C} + (\boldsymbol{I} - \mathcal{D}) \end{pmatrix} \right\} \tag{3.3}$$

where $\mathcal{D} = \text{Diag}([\mathcal{C}^t\mathcal{C}^{*-1}\mathcal{C}]_{ii})$.

Note that the covariance function in (3.3) is identical to the covariance matrix corresponding to a standard Gaussian process, apart from the lower right block, corresponding to the covariance of the regular points. Using the Sherman-Woodbury-Morrison formula (Hager, 1989), we can show that the inverse of the covariance matrix of $\boldsymbol{y}$ is:

$$\frac{1}{\sigma^2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}) = \frac{1}{\sigma^2} \begin{pmatrix} \mathcal{C}^{*-1} + \mathcal{C}^{*-1}\mathcal{C}^t(\boldsymbol{I} - \mathcal{D})^{-1}\mathcal{C}^t\mathcal{C}^{*-1} & -\mathcal{C}^{*-1}\mathcal{C}^t(\boldsymbol{I} - \mathcal{D})^{-1} \\ -(\boldsymbol{I} - \mathcal{D})^{-1}\mathcal{C}^t\mathcal{C}^{*-1} & (\boldsymbol{I} - \mathcal{D})^{-1} \end{pmatrix} \tag{3.4}$$

and the determinant of the covariance matrix is given by:

$$|\sigma^2\boldsymbol{\Sigma}(\boldsymbol{\beta})| = (\sigma^2)^n|\mathcal{C}^*(\boldsymbol{\beta})||\boldsymbol{I} - \mathcal{D}(\boldsymbol{\beta})|$$

The formulas above show that evaluations of the likelihood of the modified predictive process require calculation of the determinants and inverse matrix for the diagonal matrix $\Sigma_{\tilde{\epsilon}} = \boldsymbol{I} - \mathcal{D}$ and the $m \times m$ matrix $\mathcal{C}^{*-1}$. We can choose a small size for the matrix $\mathcal{C}^{*-1}$, depending on our computational limits; however, the smaller

the size m is, the further apart the modified predictive process will be from a conventional Gaussian process model. For a detailed comparison between these two models, see Section 3.4.

3.2.2 Knot selection

One arbitrary choice we need to make when implementing modified predictive process models is the selection of knots. Since different choices of knot sets yield different models, the results may differ drastically. However, knot design is a challenge, as in any other knot-based method, and the difficulty of its implementation increase with the dimensionality of the input space. In this chapter, we focus on methods that select the set of knots as subset of our observed locations.

Because computer model runs are controlled experiments, we typically use space-filling strategies to select points in the input space where we observe the simulator. The most common approach to the design of computer experiments is the use of Latin hypercube sampling, introduced by McKay *et al.* (1979) in computer model analysis. Latin Hypercube Designs (LHD) select points in a hypercube such that the minimum distance between points is maximized, i.e we choose $\boldsymbol{z}_1, \cdots, \boldsymbol{z}_n$ subject to

$$\max_{LHD} \min_{i,j} \delta(\boldsymbol{z}_i, \boldsymbol{z}_j)$$

where δ is the Euclidean distance. We usually rescale the input space to a unit hypercube, so the the Euclidean distance will not be affected by differences in units. It is also common to perform some transformations on the variables to make the isotropy assumptions more plausible.

For knot design, we may choose similar space-filling strategies, that do not take the spatial model into account. A naive approach is to generate an m-point maximin LHD and then select the m corresponding closest points in the set of observed locations. This approach clearly does not work very well when the design space does

not fill a hypercube (for example, we may thin set set of design points according to some criterion, such as computer model run failures).

A more sophisticated alternative is the *point swapping algorithm* proposed by Royle and Nychka (1998), which selects a subset of points from a larger candidate set based on the maximization of a "coverage" criterion. In terms of our problem, this space-filling algorithm can be represented as:

1. Specify a set of m points as the initial configuration for the set of knots $\mathcal{S}^*$; we typically select a random subset of the set of observed locations $\mathcal{S}$. Let $\tilde{\mathcal{S}} = \mathcal{S} \backslash \mathcal{S}^*$

2. Compute the coverage criterion

$$C_{p,q}(\mathcal{S}^*) = \left(\sum_{\boldsymbol{z} \in \tilde{\mathcal{S}}} \left(\sum_{\boldsymbol{z}^* \in \mathcal{S}^*} \delta(\boldsymbol{z}, \boldsymbol{z}^*)^p \right)^{q/p} \right)^{1/q}$$

3. For $i = 1, \cdots, m$

   - For $j = 1, \cdots, n-m$ replace $\boldsymbol{z}_i^* \in \mathcal{S}^*$ by $\boldsymbol{z}_j \in \tilde{\mathcal{S}}$ and recompute $C_{p,q}(\mathcal{S}^*)$
   - Swap $\boldsymbol{z}_i^*$ with $\boldsymbol{z}_j$ that produces the largest decrease over the initial criterion.

4. Repeat step 3 until no further swap can be made.

This coverage criterion is similar to a criterion associated with minimax space filling set of points when $p \to -\infty$ and $q \to \infty$. However, Royle and Nychka (1998) suggest the use of arbitrary finite values ($p = -5$ and $q = 1$) for stability in the computations.

In contrast to the above, many authors have tackled the problem of geostatistical design by minimizing a model-based design criterion, such as the mean or the

maximum of prediction variance. See, for example, McBratney and Webster (1981), Ritter (1996), and Zhu (2002). While initial works considered the covariance parameters to be known, there are some by Zhu and Stein (2005), Diggle and Lophaven (2006), and Zimmerman (2006) that take estimation uncertainty into account whilst constructing spatial designs that are efficient in model prediction.

Finley *et al.* (2009) proposed a knot selection approach based on the modified predictive process model. The idea is to find a set of knots for which the predictive process $\tilde{w}(\boldsymbol{z})$ in (3.2) best approximates the parent process $w^*(\boldsymbol{z})$. In other words, under the conventional Gaussian process model, we want to evaluate the predictive variance conditional on the observations at the knots:

$$V_\beta(\boldsymbol{z}, \mathcal{S}^*) = \mathrm{Var}[w^*(\boldsymbol{z})|w^*(\cdot), \mathcal{S}^*, \beta] = \sigma^2 - \sigma^2 \boldsymbol{c}(\boldsymbol{z}, \beta)^t \mathcal{C}^{*-1} \boldsymbol{c}(\boldsymbol{z}, \beta)$$

Note that $V_\beta(\boldsymbol{z}, \mathcal{S}^*)$ is also the difference in variance between the parent process and the predictive process at location $\boldsymbol{z}$. The design criterion is similar to the spatially averaged prediction variance used by Diggle and Lophaven (2006):

$$V_\beta(\mathcal{S}^*) = \int_A V_\beta(\boldsymbol{z}, \mathcal{S}^*) d\,\boldsymbol{z}$$

In practice, we compute $V_\beta(\boldsymbol{z}, \mathcal{S}^*)$ as an average over all the observed locations:

$$V_\beta(\mathcal{S}^*) = \frac{\sum_{i=1}^n \mathrm{Var}[w^*(\boldsymbol{z}_i)|w^*(\cdot), \mathcal{S}^*, \beta]}{n} \tag{3.5}$$

The covariance parameter β has to be estimated from the assumed model. Finley *et al.* (2009) suggest the use of a subset of the original data or fitting the predictive process on a regular lattice of knots. Following Diggle and Lophaven (2006), we may also assign a prior to β and then compute $E_\beta(V_\beta(\mathcal{S}^*))$ in a Bayesian fashion.

Once we are able to compute the design criterion $V_\beta(\mathcal{S}^*)$, we can implement

the sequential search algorithm suggested by Finley *et al.* (2009) in the following manner:

1. Specify a set of $n_0 < m$ points as the initial configuration for the set of knots $\mathcal{S}^*$; we typically select a random subset of the set of observed locations $\mathcal{S}$. Let $\tilde{\mathcal{S}} = \mathcal{S} \backslash \mathcal{S}^*$

2. At each step $t+1$

   - For each point $\boldsymbol{z}_i \in \tilde{\mathcal{S}}$, compute $V_\beta(\{\mathcal{S}^{*(t)}, \boldsymbol{z}_i\})$
   - Remove the point in $\tilde{\mathcal{S}}$ with largest decrease in V and add it to the knot set $\mathcal{S}^*$.

3. Repeat step 2 until we obtain m points in the knot set.

We may use the methods discussed in Chapter 2: evaluate the marginal objective posterior of β under the parent process model (conventional Gaussian process) and then calculate $E_\beta(V_\beta(\mathcal{S}^*))$ using Laplace approximation:

$$V_\beta(\mathcal{S}^*) = 1 - \frac{1}{n} \frac{\int_{-\infty}^{\infty} f(\mathcal{S}^*, e^\gamma) g(\gamma|\boldsymbol{y})\, d\gamma}{\int_{-\infty}^{\infty} g(\gamma|\boldsymbol{y})\, d\gamma} \approx \sqrt{\frac{\sigma_h^2}{\sigma_l^2}} \exp\{\tilde{\gamma}_h + h(\tilde{\gamma}_h|\mathbf{y}) - l(\tilde{\gamma}_h|\mathbf{y})\}$$

where $f(\mathcal{S}^*, \beta) = \sum_{i=1}^n V_\beta(\boldsymbol{z}_i, \mathcal{S}^*)$, $\gamma = \log \beta$, $\pi(\gamma|\mathbf{y}) \propto g(\gamma|\mathbf{y})$, $l(\gamma|\mathbf{y}) = \log g(\gamma|\mathbf{y})$, $h(\gamma|\mathbf{y}) = \log f(\mathcal{S}^*, e^\gamma) + l(\gamma|\mathbf{y})$, $\tilde{\gamma}_h = \arg\max h(\gamma|\mathbf{y})$, $\tilde{\gamma}_l = \arg\max l(\gamma|\mathbf{y})$, $\sigma_h^2 = -\frac{1}{h''(\tilde{\gamma}_h|\mathbf{y})}$, and $\sigma_l^2 = -\frac{1}{l''(\tilde{\gamma}_h|\mathbf{y})}$. For more details on how to implement the Laplace approximation, see Chapter 2.

This approach, however, makes the minimization of the model-based criterion extremely difficult to compute, because we would need to evaluate the posterior distribution of β at each step of the iterative algorithm. This is an example where

the plug-in approach becomes very handy. We can use the Laplace approximation methods discussed in Chapter 2 to obtain point estimates for the objective posterior median or mean of β under the parent process model. We then plug in the estimate in (3.5) and minimize the averaged prediction variance according to the sequential search algorithm.

3.2.3 Simulation examples

In order to illustrate the methods presented in the previous sections, and get some insight on how the modified predictive process behaves with varying number of data points, we have selected some synthetic example, where truth is known.

In the Bayesian analysis of our examples, we opted for vague prior distributions of the GASP parameters:

$$\begin{aligned} \beta_j &\sim \text{Exp}(1/(10\hat{\beta}_j)) \\ \pi(\theta, \sigma^2) &\propto 1/\sigma^2 \end{aligned}$$

The choice of prior distribution for the vector of correlation parameters $\boldsymbol{\beta}$ is particularly challenging. Our typical choice here corresponds to independent Exponential priors centered at a multiple (e.g. 10) of the MLE's; these proper prior distributions are very flat in the region of the parametric space where the likelihood has most of its mass, so we expect the prior to have very small effect on the posterior. Paulo (2005) showed that the choice of the multiplicative constant have very little impact in the analysis. The finite tail of this prior is important to control the likelihood in many cases.

The full conditionals for θ and τ (the inverse of σ^2) can be easily obtained in closed form. In the MCMC algorithm, we update these two parameters using Gibbs

sampling, while the correlation parameter $\boldsymbol{\beta}$ is updated by a Metropolis-Hastings step; our choice for proposal distribution is a normal random walk on the logarithm of $\boldsymbol{\beta}$.

Impact of the number of knots in the analysis

For comparison between different methods for knot design, we have reanalyze the pedagogic example proposed by Bayarri *et al.* (2007b). Inspired by mathematical models for chemical reactions, the computer model produces simulated concentrations of a compound according to the following rule:

$$y(t, u) = y_0 \exp(-ut)$$

were $y_0 = 5$ is the initial concentration, and the inputs t and u are time and rate of reaction, respectively.

The computer model is then exercised on a 75-point maximin LHD on the 2-dimensional rectangle $[0.0, 3.0] \times [0.5, 2.0]$ in (t, u) space, as shown on the left side of Figure (3.1). In the middle and on the right side of the Figure (3.1), solid circles represent $m = 30$ locations selected as knots using the space-filling algorithm and the sequential search algorithm, respectively.

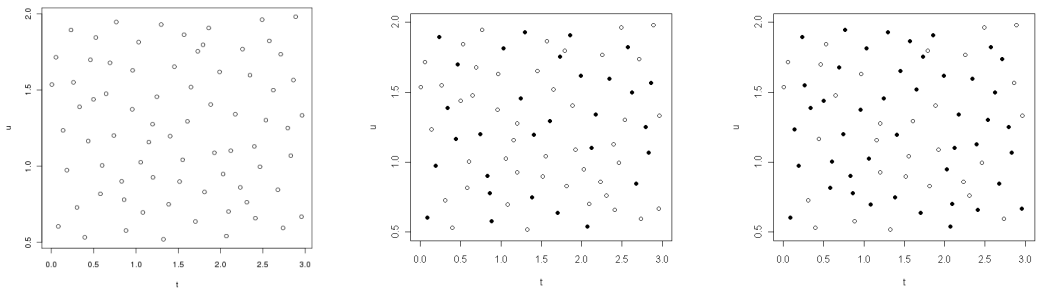


FIGURE 3.1: Left: 75-point maximin Latin Hypercube design selected for pedagogic example; Middle: on the same design, solid circles denote 30 locations selected as knots by minimizing a coverage criterion; Right: 30 locations selected as knots by minimizing the spatially averaged prediction variance.

Royle and Nychka (1998) discussed how designs based on a coverage criterion are nearly optimal for a spatial prediction. In practice, we have observed that knot designs generated by the space-filling algorithm and the sequential search algorithm are very similar. However, since our model is nicely defined, we choose to use strategies based on the minimization of the spatially averaged prediction variance.

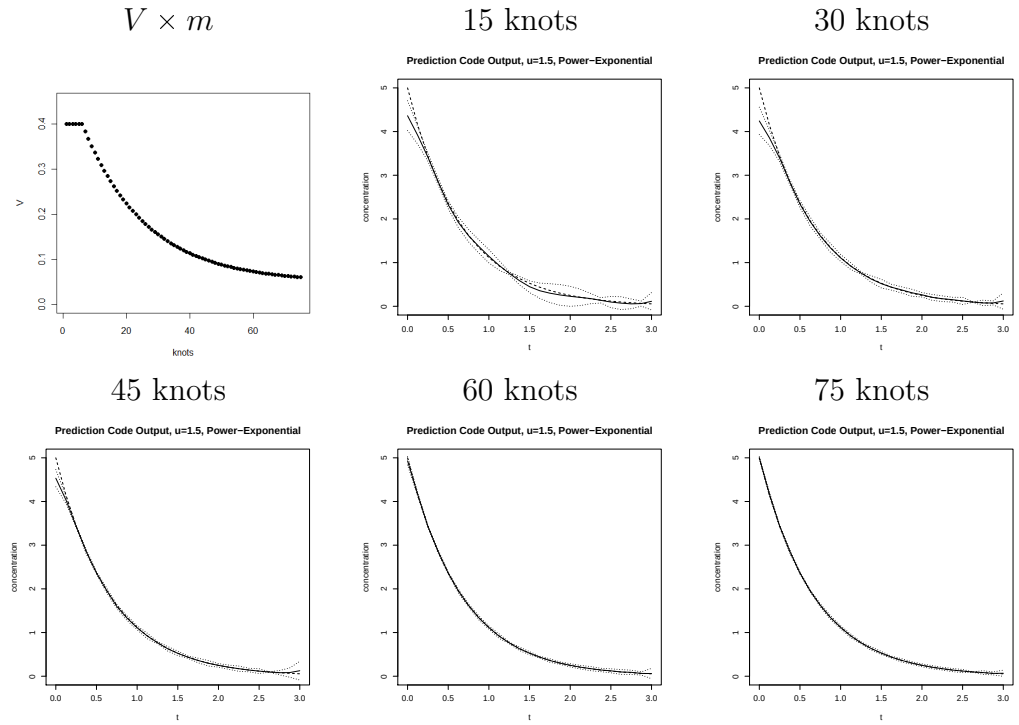


FIGURE 3.2: Top-left panel: Average prediction variance (V) versus number of knots (m). Other panels: posterior median and 95% confidence bounds for predictions of y as a function of t at $u = 1.5$, using the full Bayesian analysis on the modified predictive process model, for different number of knots. The bottom-right panel ($m = 75$) corresponds to the conventional Gaussian process.

Another important aspect of knot design is the choice of the number of knots m. The predictive gets closer to the parent process when m increases, so a large m is very desired. However, because the computation complexity increases as $O(m^3)$, the choice of m should also take into account the computational constraints of the problem. We recommend to evaluate the sensitivity of the analysis to the choice

of m. The top-left panel on Figure (3.2) is a plot of average prediction variance versus number of knots for the sequential search algorithm; the other panels show approximations to the output of the model (and corresponding 95% credible bounds), as a function of t, at the input value $u = 1.5$. The other panels show predictions and corresponding 95% confidence bounds for optimal knot designs (with respect to minimum average prediction variance) of varying sizes. As initial configuration for the sequential search algorithm we selected 5 locations from the design points that minimized the coverage criterion.

The plot of average prediction variance versus number of knots shows that the model gets less sensitive to the number of knots as m increases. This is confirmed by our plots of predicted output, where predictions and confidence bounds regarding the modified predictive process model with m larger than 30 do not present very noticeable differences from the conventional Gaussian process (the limit as m increases).

In Section 3.3 we will apply knot selection techniques to a real computer model example and discuss some other problems that may arise during the analysis.

3.2.4 Consistency under the modified predictive process

In the previous example, we analyzed how sensitive predictions are to an increasing number m of knots when the total number of data points is fixed. We are also interested in seeing the behavior of the modified predictive process model when the number of knots is fixed and the number of regular locations increases. Imagine that we have obtained new runs of the computer model; however, we cannot increase the number of knots, because our computational limits remain the same. Our only choice is to use the new observations as regular observations inside the vector $\tilde{\boldsymbol{y}}$. What will be the gain in both estimation and prediction under the modified predictive process?

We cannot expect much prediction improvement by increasing the number of

regular points, because the predictive distribution does not depend directly on the regular observations:

$$\begin{aligned}
\pi(y_{pred}|\boldsymbol{y}) &= \int \pi(y_{pred}|\boldsymbol{y},\Theta)\pi(\Theta|\boldsymbol{y})\,d\Theta \\
&= \int \pi(y_{pred}|\boldsymbol{y}^*,\Theta)\pi(\Theta|\boldsymbol{y}^*,\tilde{\boldsymbol{y}})\,d\Theta \\
&\propto \int \exp\left(\frac{(y_{pred}-m(\boldsymbol{y}^*,\Theta))^2}{v(\Theta)}\right)\pi(\Theta|\boldsymbol{y})\,d\Theta \\
\boldsymbol{m}(\boldsymbol{y}^*,\Theta) &= \theta\mathbf{1}+\boldsymbol{c}_{pred}^t\mathcal{C}^{*-1}\left[\boldsymbol{y}^*-\theta\mathbf{1}\right] \\
\boldsymbol{v}(\Theta) &= \sigma^2(1-\boldsymbol{c}_{pred}^t\mathcal{C}^{*-1}\boldsymbol{c}_{pred})
\end{aligned}$$

where $\Theta=(\theta,\sigma^2,\boldsymbol{\beta})$ is the vector of parameters and $\boldsymbol{c}_{pred}$ is the vector of correlations between the predicted value y_{pred} and the observations at the knots $\boldsymbol{y}^*$.

Note that the predictive distribution of y_{pred} does not depend directly on the values of the regular observations $\tilde{\boldsymbol{y}}$. The regular observations affect the posterior predictive distribution through the posterior distribution of Θ. Because of that, the modified predictive process model cannot interpolate through the regular observations.

In this pedagogic example, we generate realizations from a conventional Gaussian Process Model on different 2-dimensional designs. We first generate our set of knots using nested maximin LHD's of different sizes $m=15,60$ and 135. We augmented each one of these sets of knots with nested sets of regular locations of sizes $k=n-m=15,60$ and 135, totaling 9 different designs. Then, for each one of these 9 designs, we generate 2,000 realizations of a GASP with power-exponential correlation function, $\theta=0$, $\sigma^2=1$, and $\boldsymbol{\beta}=\mathbf{1}$. Since our typical choices of prior distributions are non-informative, we focus on the likelihood of the modified predictive process and calculate the MLE for each simulated GASP realization. MLE's were computed

by the Fisher scoring method described in Paulo (2005). See details in the Appendix.

Figure 3.3 shows the distribution of observed MLE for each pair of number of knots and number of regular locations (m, k). On the rows, we have configurations with the same number m of knots and, on the columns, the configurations have the same number k of regular locations. The empirical distribution of the MLE is centered at the true value of the parameter, even though we are using different models to generate samples and estimate parameters.

When we compare different rows of Figure 3.3 we note that the addition of knots to our data points improve the estimation of our parameters, since the empirical distributions of MLEs get more and more concentrated around the true value of the parameter. However, when we compare different columns of the figure, we can see that additional regular points do not change the observed spread of the distribution of MLES. The MLEs of other parameters show similar patterns. This suggests that the modified predictive process model is consistent with respect to knots but not with respect to regular points, i.e., adding regular points will not make the likelihood concentrate its mass around the true value of the parameter. This unusual feature of the predictive process model make us question the role of regular points in the estimation of parameters. In Section 3.3, we present a method for analysis of the model that takes into account our suspicions about this portion of the data set.

Different estimators for a fixed number of knots

The previous example suggested that the addition of regular points does not represent a great improvement on the estimation of parameters in the modified predictive process model. A possible solution to this problem is the partition of our data points into batches: for each batch, we calculate the MLE under the modified predictive process model and then we average the MLE values over different batches. If these MLE values were independent, then we could expect to obtain consistent estimates

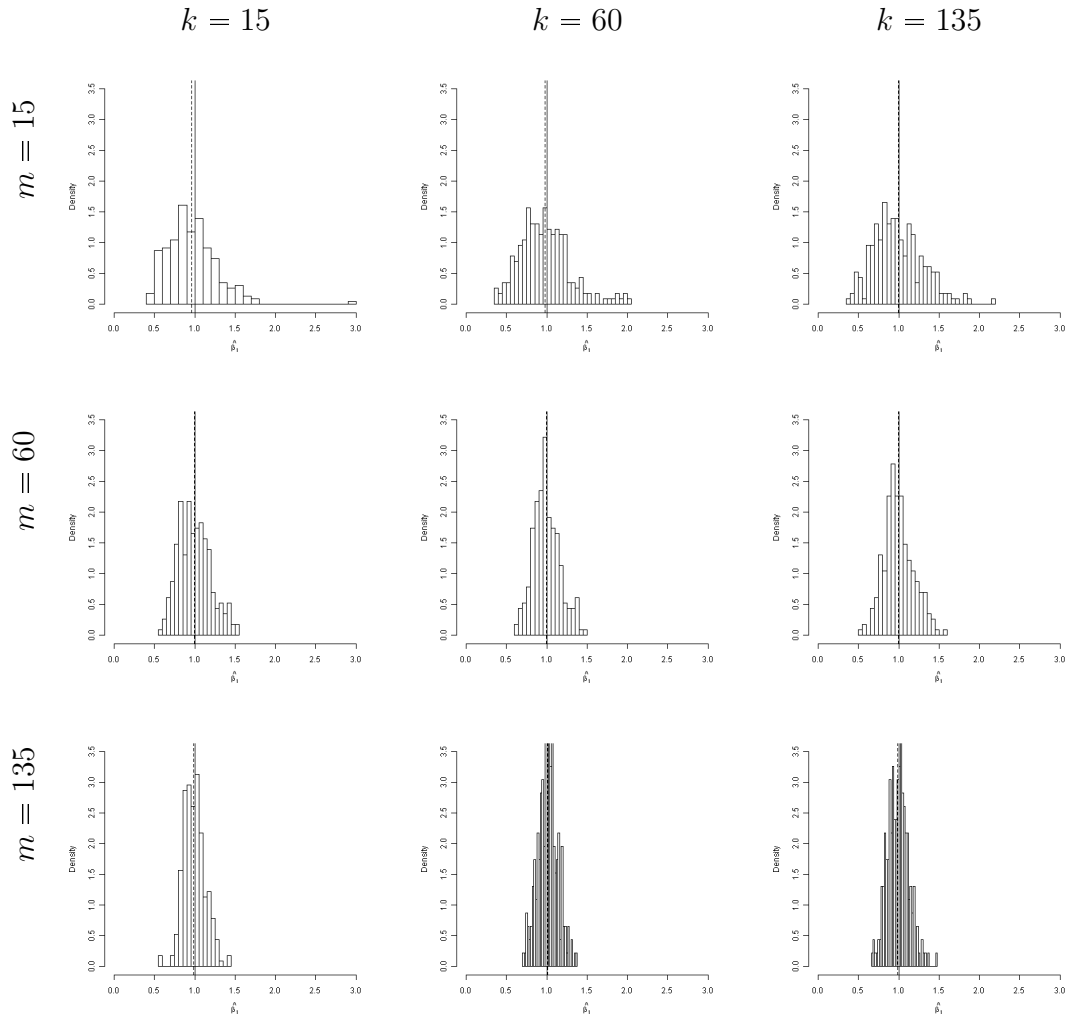


FIGURE 3.3: Empirical distribution of MLEs of β_1 calculated from GASP realizations on nested design points under the modified predictive process model. On the rows, from left to right, we augment each design with regular points, while on the columns, from top to bottom, the designs are augmented with knots. The solid vertical lines are averages of the MLEs, and the dashed vertical lines are the true value of the parameter $\beta_1 = 1$.

from the averages by the Central Limit Theorem.

Similarly to the previous example, we generate several realizations of a Gaussian process model on a 1,015-point maximin LHD. We fix the $m = 15$ knot locations in the design space and use different number of regular locations $k = 10, 20, 50, 100, 500$ and $1,000$. We then separate at random the regular points into $1000/k$ batches of k

points. The MLE is calculated for each batch and the average of the batches' MLEs is computed. We also have tried an approach where each batch has a different set of 15 knots, but similar results were obtained.

Figure 3.4 presents the empirical distribution of the MLE averages of β. The case where $k = 1000$ corresponds to the same analysis as the previous example, where only one MLE was calculated per realization and no average was taken. We can see that the different estimators have similar distributions, and we conclude that the strategy of partitioning the design points into batches and then averaging estimates over batches does not improve estimation of the parameters in the modified predictive process model. If we use the predictive process model as an approximation to the conventional GASP model, it seems that the choice of the number of knots will restrict how good our estimates and predictions can be, no matter how many regular points we have.

3.3 Modularization

Recall that we denote the vector of observations at the knots by $\boldsymbol{y}^*$ and the vector of other, regular, observations by $\tilde{\boldsymbol{y}}$. If we look at the predictive process model in (3.3), we can easily see that the conditional distribution of the emulator at regular points given the observations at the knots is given by:

$$\begin{aligned} (\tilde{\boldsymbol{y}}|\boldsymbol{y}^*) &\sim N_{n-m}(\boldsymbol{m}(\boldsymbol{y}^*), \boldsymbol{V}(\sigma^2, \boldsymbol{\beta})) \\ \boldsymbol{m}(\boldsymbol{y}^*) &= \theta\mathbf{1} + \mathcal{C}^t\mathcal{C}^{*-1}\left[\boldsymbol{y}^* - \theta\mathbf{1}\right] \\ \boldsymbol{V}(\sigma^2, \boldsymbol{\beta}) &= \sigma^2(\boldsymbol{I} - \boldsymbol{D}) \end{aligned} \tag{3.6}$$

That is, the regular points are conditionally independent given the knots with

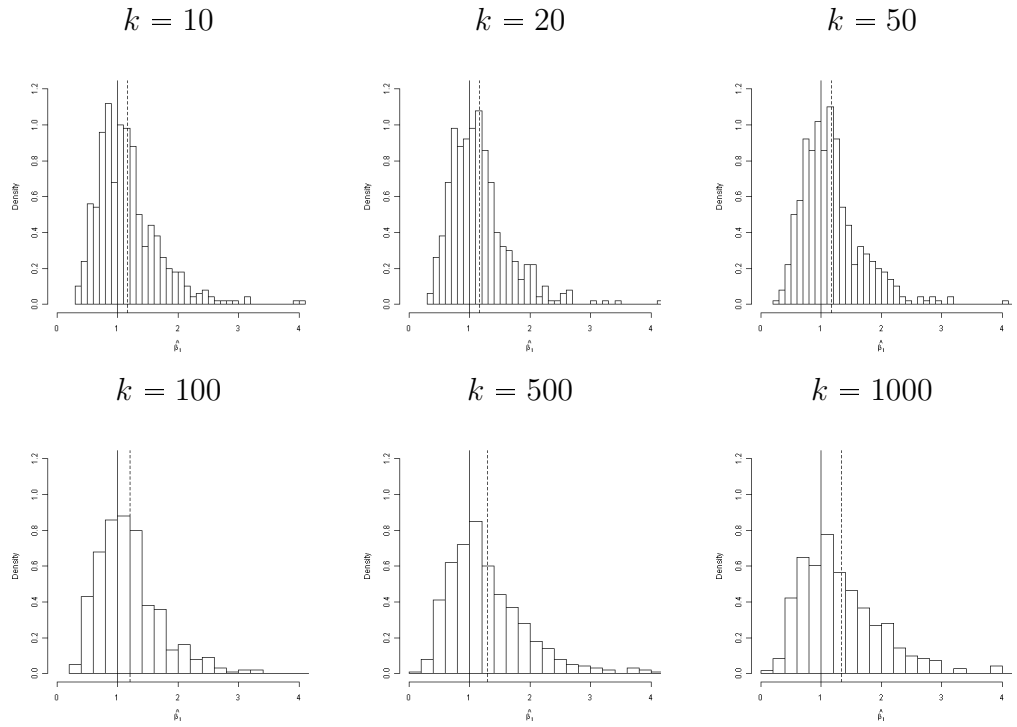


FIGURE 3.4: 500 simulated averages of the MLE of β_1 under the predictive process model with 15 knots and k regular points. The total number of regular points is 1000, and we partitioned them into $1000/k$ batches. The MLE is calculated for each batch and the average of the batches' MLEs is computed.

$$\begin{aligned}
(\tilde{y}_i|\boldsymbol{y}^*) &\overset{\text{ind}}{\sim} N(m_i(\boldsymbol{y}^*), v_i(\sigma^2, \boldsymbol{\beta})) \\
m_i(\boldsymbol{y}^*) &= \theta + \boldsymbol{c}_i^t \mathcal{C}^{*-1} [\boldsymbol{y}^* - \theta \mathbf{1}] \\
v_i(\sigma^2, \boldsymbol{\beta}) &= \sigma^2 (1 - \boldsymbol{c}_i^t \mathcal{C}^{*-1} \boldsymbol{c}_i)
\end{aligned}$$

where $\boldsymbol{c}_i$ is the i-th column of $\mathcal{C}(\boldsymbol{\beta})$.

Hence, the likelihood in (3.3) can be rewritten as

$$\begin{aligned}
L(\boldsymbol{y}; \theta, \sigma^2, \boldsymbol{\beta}) \propto\ & (\sigma^2)^{-1/2} |\mathcal{C}^*|^{-1/2} \exp\left(-\frac{1}{2\sigma^2} (\boldsymbol{y}^* - \theta \mathbf{1})^t \mathcal{C}^{*-1} (\boldsymbol{y}^* - \theta \mathbf{1}) \right) \\
& \times \prod_{i=1}^{n-m} v_i(\sigma^2, \boldsymbol{\beta})^{-1/2} \exp\left(-\frac{1}{2 v_i(\sigma^2, \boldsymbol{\beta})} (\tilde{y}_i - m_i(\boldsymbol{y}^*))^2 \right)
\end{aligned}$$

When analyzing the likelihood function, we note that the correlation parameter $\boldsymbol{\beta}$ in the modified predictive process model plays two different roles at the same time:

- Describing the correlation among observations at the knots; this is the usual role of the correlation parameter in a Gaussian process model.
- Interpolating conditional means and controlling conditional variances of the regular observations given observations at the knots; while in a conventional Gaussian process this would be a simple consequence of the correlation structure, in the predictive process this is the only effect of the correlation parameter in the conditional distribution.

The reduction in complexity of the modified predictive process is achieved by eliminating the correlation structure from the conditional distribution of the regular observations given observations at the knots. Although conditional independence is not an assumption we would like to make, because it may not capture all the possible interactions among our observations, this is a price we may need to pay due to our computational limits.

However, our major concern is how different portions of the data give information about the correlation parameter. In terms of likelihood, observations at the knots tell which values of parameters are most likely to give the observed correlation structure; on the other side, regular observations tell which values of covariance parameters fit best a series of marginal conditional distributions. Since we usually want the number of regular locations to be much larger than the number of knots (so we can justify the use of the modified predictive process instead of a conventional GASP), it is important to know how the unconventional pieces of information given by regular observations will affect both estimation and prediction.

Liu *et al.* (2009) proposed a Bayesian technique for settings where one or more sources of information are suspect, called *modularization*. The authors give several

reasons for the use of modularization, but we are particularly interested in the case of contamination: a 'good' component of the model (or module) might need to be kept separate from a 'suspect' module to prevent the analysis to be unreasonably influenced. The basic idea of modularization is to ignore the contribution of some modules in estimating parameters that we believe may be unreasonably influenced by those modules during the first stage. At a second stage, the other, more stable, parameters can be estimated using all the modules and the posterior samples generated at the first stage.

One may argue that the existence of a suspect module in the model should be attacked before embarking in the analysis, but there are many applications where improving the modeling is not feasible. In the modified predictive process model, for example, the existence of a suspect model is motivated by computational costs. Note that scientific reasons should be provided for the division of the model into modules and that the modular approach must be carried out with care since there is no guarantee of coherence in the Bayesian analysis.

A useful application of modularization in the analysis of computer models to estimate model parameters separately from its bias. As any other model, computer models are simple versions of reality, and because of that they may differ systematically from the real process. However, the bias cannot be observed directly. Using field data, computer modelers are interested in assessing this discrepancy for future improvement of the computer model. Even though there are similarities between computer models and the real processes, if we use field observations when estimating model parameters in a full Bayesian analysis, the discrepancy is usually underestimated because the emulator will choose to approximate computer model estimates to the reality. The modular approach tends to correct that by fitting the emulator to computer model runs only and then estimating the bias from the difference between field observations and the emulator. For details on how to estimate discrepancy

of computer models using modularization see Bayarri *et al.* (2007a) and Liu *et al.* (2009).

Similarly, in our study, we know that observations at the knots give correct information about correlation parameters, while we suspected that regular observations may affect the analysis in a undesired way. We then separate the knots and regular locations into two modules and carry out the following modular approach:

- **Stage 1:** Infer $\boldsymbol{\beta}$ from the knots only using a conventional Gaussian process model.
- **Stage 2:** Do inferences about $\boldsymbol{\theta}$, σ^2 and the "kriging prediction" from the modified predictive process model given β.

In the modularization of the modified predictive process model, we ignore the contribution of regular observations to the estimation of the correlation parameter $\boldsymbol{\beta}$, using only observations at the knots. In terms of the MCMC method for posterior inference, we first sample from the marginal posterior distribution of $\boldsymbol{\beta}$ using the GASP model at the knots only and then sample from the full conditionals of the other parameters under the modified predictive process model given the sampled values of $\boldsymbol{\beta}$. MCMC algorithms are presented in details in the next subsection with all conditional distributions in closed form.

3.3.1 MCMC Algorithm

Our main interest is to compare the original modified predictive process model to the 'modularized model' to check which model has better performance in both estimation and prediction. Modularization may also give us some insight on how information from regular observations affect the analysis.

We specify prior distributions for the parameters as:

$$\begin{aligned} \beta_j &\sim \text{Exp}(1/(10\hat{\beta}_j)) \\ \pi(\theta, \sigma^2) &\propto 1/\sigma^2 \end{aligned}$$

Next, we describe how to sample the next state state of the MCMC chain, given the current state of the chain $\boldsymbol{\beta}^{(t-1)}$, $\theta^{(t-1)}$, $\lambda^{(t-1)}$ for different approaches.

Non-modular approach:

- Generate θ from its full conditional given the predictive process model and all the observations $\boldsymbol{y} = (\boldsymbol{y}^{*t}, \tilde{\boldsymbol{y}}^t)^t$:

$$(\theta|\boldsymbol{\beta}, \lambda, \mathbf{y}) \sim N\left((\mathbf{1}^t\Sigma^{-1}(\boldsymbol{\beta})\mathbf{1})^{-1}(\mathbf{1}^t\Sigma^{-1}(\boldsymbol{\beta})\mathbf{y}),\ \lambda^{-1}(\mathbf{1}^t\Sigma^{-1}(\boldsymbol{\beta})\mathbf{1})^{-1}\right)$$

- Generate θ from its full conditional given the predictive process and all the observations:

$$(\lambda|\boldsymbol{\beta}, \theta, \mathbf{y}) \sim Ga\left(n/2,\ (\mathbf{y} - \theta\mathbf{1})^t\Sigma^{-1}(\boldsymbol{\beta})(\mathbf{y} - \theta\mathbf{1})\right)$$

- Generate $\boldsymbol{\beta}$ using a Metropolis-Hastings step and a proposal distribution equivalent to a normal random walk on $\log(\boldsymbol{\beta})$.

$$\begin{aligned} \pi(\boldsymbol{\beta}|\theta, \lambda, \mathbf{y}) \propto\ & \left[\det\left(\Sigma(\boldsymbol{\beta})\right)\right]^{-1/2} \exp\left\{-\frac{\lambda}{2}(\mathbf{y} - \theta\mathbf{1})^t\Sigma^{-1}(\boldsymbol{\beta})(\mathbf{y} - \theta\mathbf{1})\right\} \\ & \times \exp\left\{-\sum_j \beta_j/(10\hat{\beta}_j)\right\} \end{aligned}$$

Modular approach:

Stage 1:

- Generate the auxiliary variable θ^* from the full conditional of θ given the knots $\boldsymbol{y}^*$ only:

$$(\theta^*|\boldsymbol{\beta},\theta,\lambda,\lambda^*,\boldsymbol{y}^*) \sim N\left((\mathbf{1}^t\mathcal{C}^{*-1}(\boldsymbol{\beta})\mathbf{1})^{-1}(\mathbf{1}^t\mathcal{C}^{*-1}(\boldsymbol{\beta})\boldsymbol{y}^*),\ \lambda^{*-1}(\mathbf{1}^t\mathcal{C}^{*-1}(\boldsymbol{\beta})\mathbf{1})^{-1}\right)$$

- Generate auxiliary variable from the full conditional of λ given the knots only:

$$(\lambda^*|\boldsymbol{\beta},\theta,\theta^*,\lambda,\mathbf{y}) \sim Ga\left(m/2,\ (\boldsymbol{y}^*-\theta^*\mathbf{1})^t\mathcal{C}^{*-1}(\boldsymbol{\beta})(\boldsymbol{y}^*-\theta^*\mathbf{1})\right)$$

- Generate $\boldsymbol{\beta}$ using a Metropolis-Hastings step given θ^*,λ^* and the knots only. The proposal distribution equivalent to a normal random walk on $\log(\boldsymbol{\beta})$.

$$\begin{aligned}\pi(\boldsymbol{\beta}|\theta,\theta^*,\lambda,\lambda^*,\mathbf{y}) \quad &\propto \quad [\det(\mathcal{C}^*(\boldsymbol{\beta}))]^{-1/2}\exp\left\{-\sum_j \beta_j/(10\hat{\beta}_j)\right\} \\ &\quad \times\exp\left\{-\frac{\lambda^*}{2}(\boldsymbol{y}^*-\theta^*\mathbf{1})^t\mathcal{C}^{*-1}(\boldsymbol{\beta})(\boldsymbol{y}^*-\theta^*\mathbf{1})\right\}\end{aligned}$$

- Discard θ^* and λ^*. After convergence of the MCMC, $\boldsymbol{\beta}$ is an observation from the posterior corresponding to the knots only.

Stage 2:

- Generate θ^* from its full conditional given the predictive process model, $\boldsymbol{\beta}$ from the Stage 1, and all the observations $\boldsymbol{y} = (\boldsymbol{y}^{*t},\tilde{\boldsymbol{y}}^t)^t$:

$$(\theta|\boldsymbol{\beta},\theta^*,\lambda,\lambda^*,\mathbf{y}) \sim N\left((\mathbf{1}^t\Sigma^{-1}(\boldsymbol{\beta})\mathbf{1})^{-1}(\mathbf{1}^t\Sigma^{-1}(\boldsymbol{\beta})Y\mathbf{y}),\ \lambda^{-1}(\mathbf{1}^t\Sigma^{-1}(\boldsymbol{\beta})\mathbf{1})^{-1}\right)$$

- Generate λ^* from its full conditional given the predictive process model, $\boldsymbol{\beta}$ from the Stage 1, and all the observations

$$(\lambda|\boldsymbol{\beta},\theta,\theta^*,\lambda^*,\mathbf{y}) \sim Ga\left(n/2,\ (\mathbf{y}-\theta\mathbf{1})^t\Sigma^{-1}(\boldsymbol{\beta})(\mathbf{y}-\theta\mathbf{1})\right)$$

Note that the two stages can be implemented in a unique loop. The advantage of doing so is the reduction in computational effort by using $\mathcal{C}^{*-1}(\boldsymbol{\beta})$ form Stage 1 to calculate $\Sigma^{-1}(\boldsymbol{\beta})$ as indicated in (3.4).

3.3.2 Simulation examples

In this section we assume that the true distribution of our data is a regular Gaussian process. Because a Gaussian process model is considered too computationally expensive to be used in our applications, we evaluate the performance of the modified predictive process model as an approximation to the conventional GASP. Note that our conclusions may be different if we assume that the true model for our data is a modified predictive process.

Influence of regular observations

In our first example, we want to check what is the impact of the information provided by regular observations in the analysis of the modified predictive process model. Imagine we have two computer models $y_1(t,u)$ and $y_2(t,u)$, of the form:

$$y_1(t,u) = 5\exp(-ut) \quad , \quad y_2(t,u) = 5\exp(-ut) + \sin(5ut)/10$$

We then exercise each computer model on the on the 55-point maximin LHD showed on the left side of Figure 3.5, where the solid circles denote our selected knot locations. We forced these knots to lie on the the isolines (dashed lines), where the two models are identical, i.e., $y_1(t,u) = y_2(t,u)$. We are interested in analyzing the impact of different values for the regular points on estimation and prediction when the information provided by the set of knots is exactly the same, for both modular and non-modular approaches.

For posterior inference, we use the prior distributions and MCMC procedures described in the previous section. We fit emulators for both computer models using the non-modular and the modular approach. Note that, since the Stage 1 of modular approach depends only on the knots and the information provided by observations at the knots is the same for both computer models in the example, the posterior

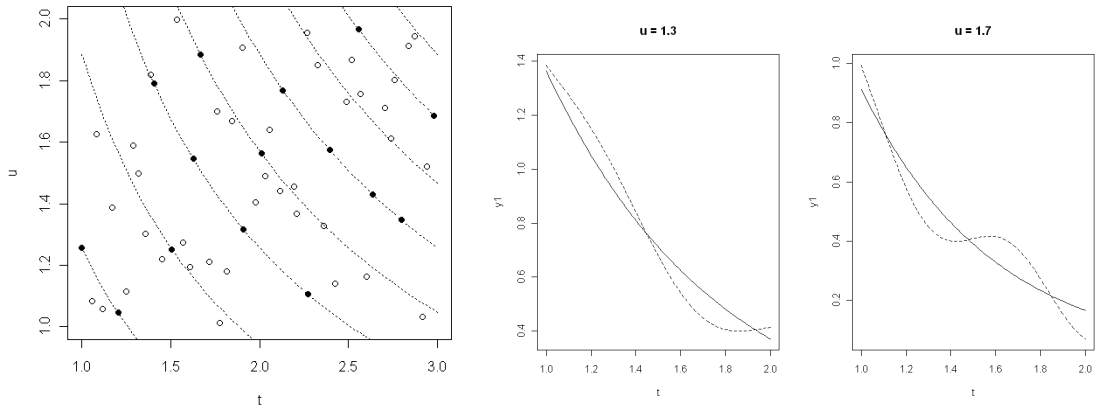


FIGURE 3.5: Left Panel: Design points where we exercise our computer models. Solid circles are knots and dashed lines are isolines where the two models are identical. Middle and Right Panels: computer models y_1 (solid line) and y_2 (dashed line) as a function of t at $u = 1.5$ and $u = 1.7$, respectively.

distribution of $\boldsymbol{\beta}$ for the modular emulator should be the same, regardless of the computer model in study.

Figure 3.6 presents posterior samples of our two correlation parameters for each model. Comparing the first two columns, corresponding to samples from the original model for computer models y_1 and y_2, respectively, we conclude that regular observations do affect the estimation of the correlation parameters, even in the case where the information coming from observations at the knots is the same. In particular, the emulator for the first computer model tends to generate smaller values of correlation parameters than the emulator for the second computer model. The results from the modular approach on the third column show that ignoring the information from the regular observations does have an impact on the posterior samples of β.

On the other side, when we look at predictions, the differences between modular and non-modular approaches are almost non-noticeable, as shown in Figure 3.7. As we pointed out before, the modified predictive process interpolates through the knots only and, since the information provided by observations at the knots is the same for the two computer models, we cannot distinguish point predictions (solid lines) over

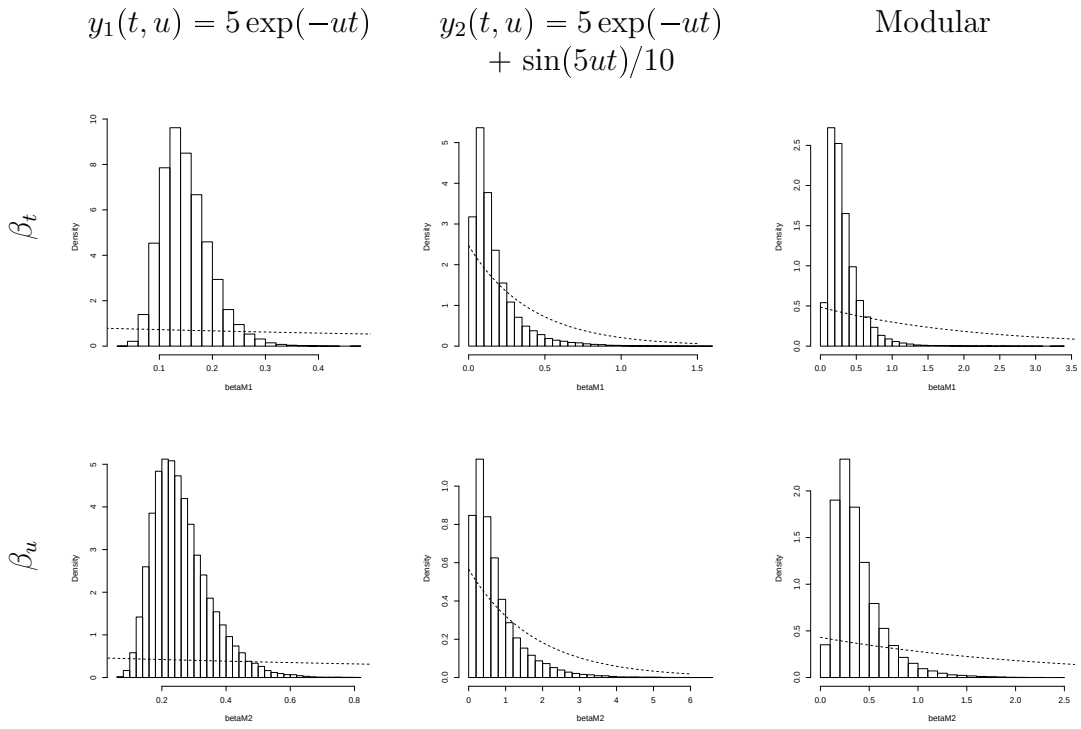


FIGURE 3.6: Histograms of posterior draws of correlation parameters β_t (first row) and β_u (second row) for computer models y_1 (first column) and y_2 (second column) using the non-modular and modular (third column) approaches. Dashed lines represent the corresponding prior distribution.

the columns of Figure 3.7. However, the modified predictive process tries to capture the wiggly pattern of the computer model y_2 by providing wider credible bounds compared to the bounds of predictions for computer model y_1. This example shows that we need to be careful when presenting predictions from the modified predictive process model: one should not present only point estimates, but also some measure of uncertainty, e.g., credible bounds.

Performance of the modular approach

Our objective in this analysis is to use the modified predictive process as an approximation to the conventional Gaussian process for emulation of computer models. In Subsection 3.2.3 we have showed that the modified predictive process model produces estimates of the parameters of a conventional GASP that are centered at their true

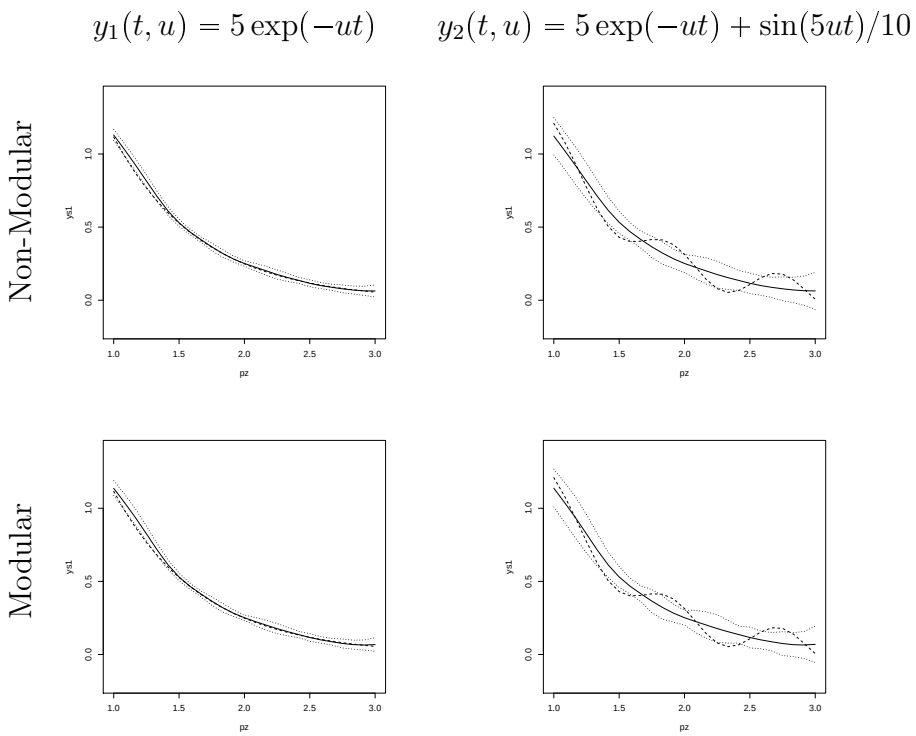


FIGURE 3.7: Posterior predictions (solid lines) and 95% credible bounds (dotted lines) as a function of t at $u = 1.5$ for computer models y_1 (first column) and y_2 (second column) using the non-modular (first row) and modular (second row) approaches. Dashed lines represent the true value of the computer model.

values, even though they are completely different models.

We now return to the use of the predictive process as an approximation to the regular GASP, but focusing on the comparison between non-modular and modular approaches. Does modularization improve the performance of the predictive process model as an approximation to a GASP?

As we showed before, regular observations affect posterior predictions of the modified predictive process only through the posterior distribution of parameters. Since, the difference between the non-modular and the modular approaches is how we treat the information provided by regular observations, we may focus our attention on the estimation of parameters. Assuming that the real overall model is a Gaussian Process, and we want to do inferences with the predictive process model using the

two different approaches.

We generate realizations from a conventional Gaussian Process Model on the 2-dimensional design pictured in Figure 3.1. Using these design points, we select optimal sets of knots (according to the minimization of the average prediction variance) of sizes $m = 15, 30$ and 45. We set the GASP mean $\theta = 0$ and draw β_j from $Exponential(1)$ and λ from $Ga(2, 1)$ distributions for each realization (we use a Gamma distribution with the rate parametrization). For each design, we generate 500 realization and, for each realization, we run the MCMC algorithm for posterior inference.

Figure 3.8 shows plots of posterior median versus the true value generated fot β_t, one of the correlation parameters, for each realization, with different number of knots, using modularization or not. Both modular and non-modular approaches present a cloud of points that converges to the equality line as the number of knots increases, but we cannot visually determine which approach has better performance for each number of knots.

Figure 3.9 shows plots of posterior Mean Absolute Errors (MAE) using the non-modular approach versus posterior MAEs using the modular approach. We calculate the Mean Absolute Error as the average absolute value of the difference between each value in our posterior sample and the true value of the parameter. Now we can clearly see that the non-modular approach yields posterior samples with smaller errors around the true value when compared to the modular approach, regardless of the number of knots. The observed percentages of realizations that that had smaller MAE under the modular approach for $m = 14, 30$ and 45 are, respectively, 10%, 18.4% and 27.4%.

In our pedagogic examples, modularization has not proven to improve estimation or predictions when we use the modified predictive process model as approximation to the regular Gaussian process model.

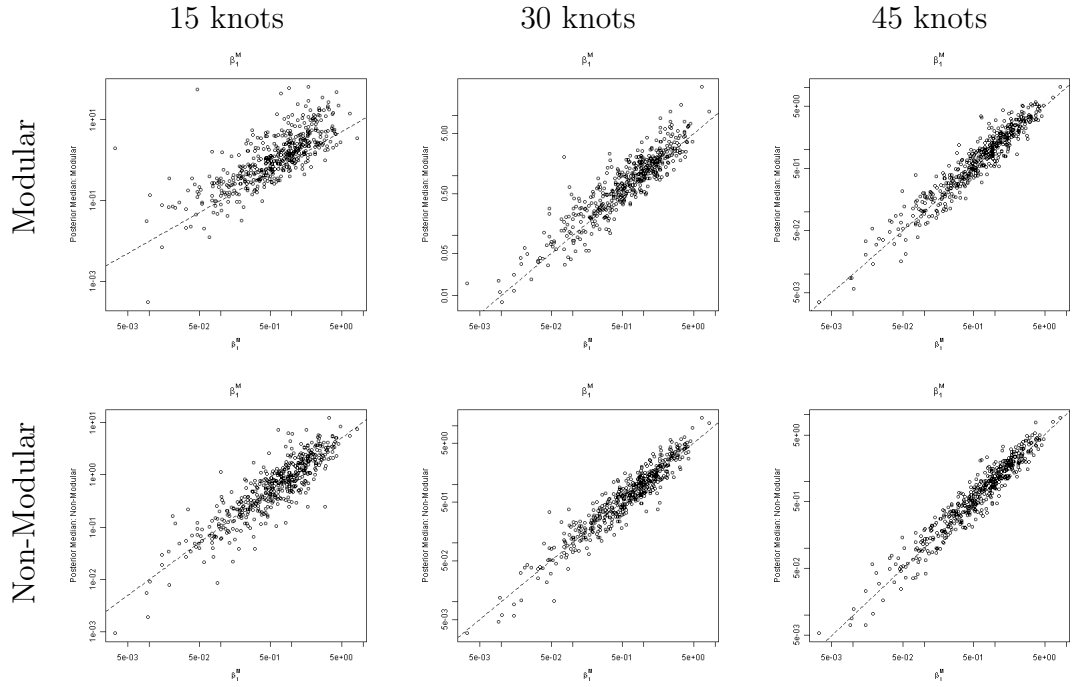


FIGURE 3.8: Plots of posterior median of β_t versus its true value for different simulated GASP realizations on a 75-point maximin LHD. We use the modified predictive process model with different number of knots in the analysis for comparison. Dashed lines represent the equality line.

3.4 Comparison between models for emulation

In this section, we illustrate the analysis by a real computer model example. TITAN2D is a simulation tool developed at University of Buffalo for modeling pyroclastic flows (Patra *et al.*, 2005; Sheridan *et al.*, 2005). TITAN2D uses a digital elevation model of geographic region and a system of partial differential equations to describe the behavior of a pyroclastic flow on that region over time given a set of initial conditions. These conditions, the inputs of the computer model, are the initial volume of the erupted material, the initial direction of the flow and two friction angles that describe the thickness of the flow material. The output is the maximum flow-height over time on a grid of points.

The test site used in the analysis is the Soufrière Hills Volcano on the Caribbean

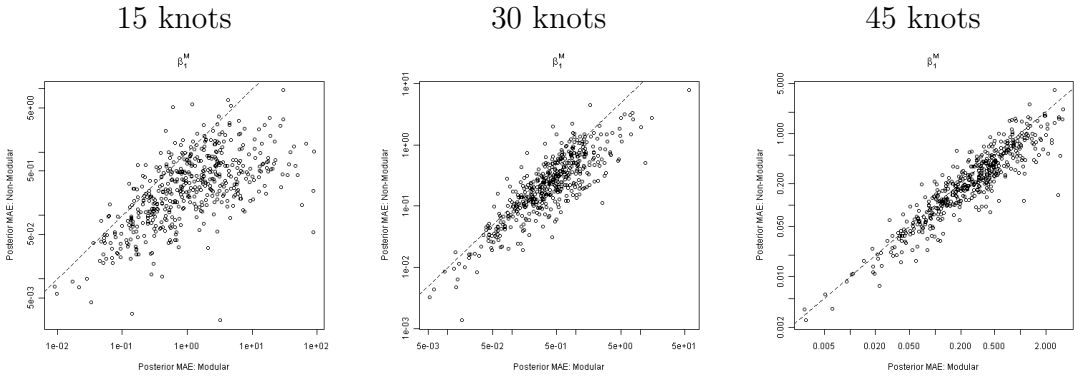


FIGURE 3.9: Plots of posterior MAE for the non-modular approach versus posterior MAE for the modular approach, for designs with different number of knots. Dashed lines represent the equality line.

island of Montserrat. We select two different locations on the island for separate univariate analysis of the output: the former capital city of Plymouth, and the former Bramble Airport. Both locations were destroyed by pyroclastic flows in real life in 1997.

Our data set consists of 2,048 flow simulations at design points corresponding to a maximin LHD in 4-dimensional hypercube of the input space. For each one of these simulations we recorded the maximum flow-height at our two study locations. Simulated pyroclastic flows often do not hit one of our locations on the island, which is equivalent to a zero flow-height. Unfortunately, those zero observations become a problem when we emulate TITAN2D, because Gaussian processes are not good to describe functions with regions of constant value. Determining which regions in the input space give a zero output should be a separate problem, and we analyze the computer model conditioned on the fact that a pyroclastic flow has reached the location in study. For more details on the analysis of the zero problem, see Bayarri *et al.* (2009).

In Figure 3.10, design points corresponding to zero flow-heights are represented by small blue dots. After discarding these observations, we have 323 points for Bram-

ble Airport and 431 points for Plymouth. We use the sequential search algorithm described in Subsection 3.2.2 to select 50 knots for Bramble Airport and 100 knots for Plymouth (represented by solid circles in Figure 3.10). We will study the two locations in separate univariate models; in Section 5 we will discuss some methods that consider several spatial locations on the island at the same time.

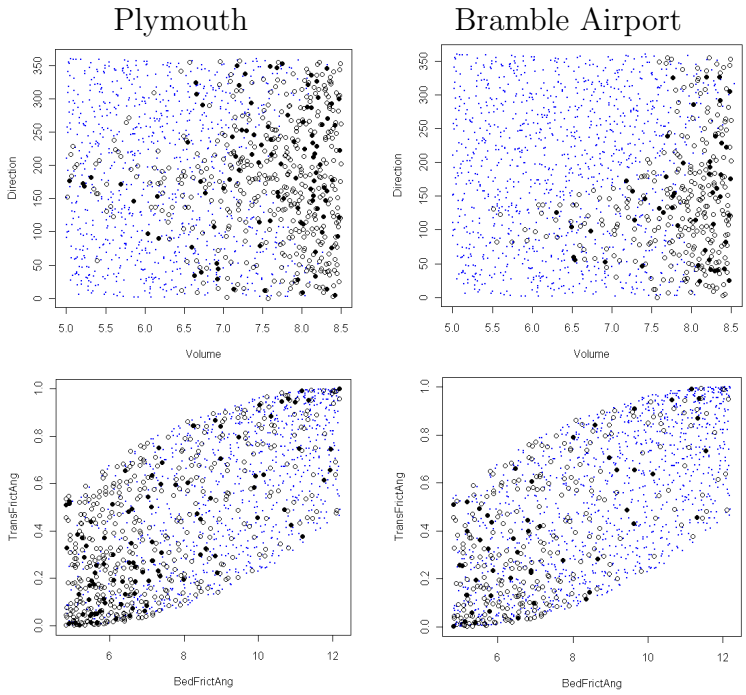


FIGURE 3.10: Representation of the design inputs covering the input space for our two locations in study: Plymouth (first column) and Bramble Airport (second column). Blue dots represent locations were the flow-height is zero and will be discarded from the analysis. Open circles identify regular locations and filled circles are knots used in the modified predictive process model.

The main goal of our analysis is to evaluate the performance of the modified predictive process model as an approximation to the conventional Gaussian process. The regular GASP model can be seen as limiting case to the predictive process model as the number of knots increases, so we expect the two models to yield similar results for both estimation and prediction. For comparison to the modified predictive process model, we are going to fit a model of similar computational cost: the conventional

GASP observed at the knots only. This is equivalent to discarding all the regular observations and considering the set of knots as our new data set.

For posterior inference, we use the prior distributions and MCMC algorithms described in Subsection 3.3.1. However, we make a slight change to the correlation function in order to ensure that our emulated height surface is a periodic function of the initiation angle:

$$\mathbf{\Sigma}_{ij} = c(\mathbf{x}_i, \mathbf{x}_j) = \prod_{k=1}^{4} c_k(\mathbf{x}_{ik}, \mathbf{x}_{jk})$$

where $c_k(\mathbf{x}_{ik}, \mathbf{x}_{jk}) = \exp\{-\beta_k|\mathbf{x}_{ik} - \mathbf{x}_{jk}|^{1.9}\}$ for the initial volume and the two friction angles and

$$c_4(\mathbf{x}_4, \mathbf{y}_4) = \frac{1}{K(\beta_4)} \sum_{k=-\infty}^{\infty} \exp(-\beta_4|\mathbf{x}_{i4} - \mathbf{x}_{j4} + 2\pi k|^{1.9})$$

for the initial direction of the flow, where $K(\beta) = 1 + 2\sum_{k=1}^{\infty} e^{-\beta|2\pi k|^{1.9}}$.

Figure 3.11 shows posterior samples of the correlation parameter β_V of the initial volume for each spatial location (Plymouth and Bramble Airport) and each univariate model for which we are are doing the analysis. The conventional Gaussian process at all the points is the most complete model, but it is also the most expensive one. Hence, rather than considering this model as an alternative, we see it as a target for the other models. If we the results from the predictive process model or from the GASP at the knots only are similar to results from the GASP at all points, then we have a evidence that the simpler models are going in the right direction.

Comparing the columns of Figure 3.11, we note that the GASP at the knots only has posterior distributions that are closer to the GASP at all points than the predictive process model for Bramble Airport. For Plymouth, both models seem to be far off our target. The posterior distribution of β_V under the modified predicative

process model is usually concentrated on small values when compared to the two regular Gaussian process models, regardless of the spatial location in study. When we look at the posterior distribution of other parameters, we note that discarding the regular observations usually gives better posterior estimates than using the entire data set with the predictive process model.

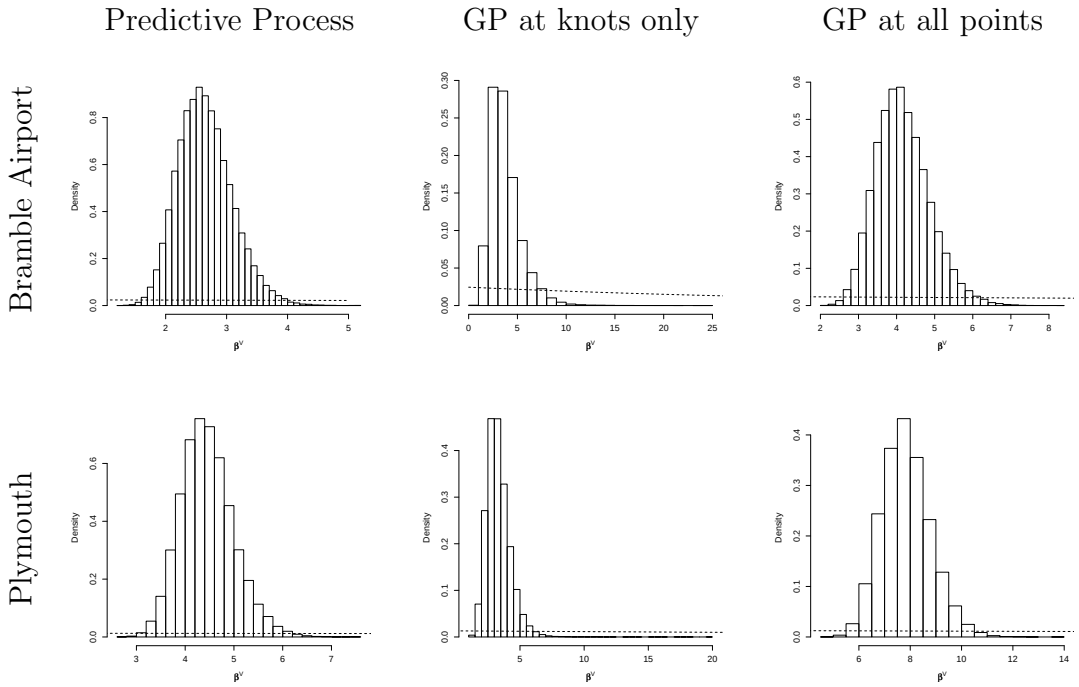


FIGURE 3.11: Histograms of posterior samples of the correlation parameter corresponding to initial volume for Bramble Aiport (first row) and Plymouth (second row). On the columns, we have different models, which are from left to right: modified predictive process model, regular Gaussian process model at the knots only, and regular Gaussian process model at all points. Dashed lines represent the corresponding prior distributions.

Figure 3.12 presents predictions and 95% credible bounds as functions of initial volume, with all the other inputs fixed (Basal friction angle = 8.5°, internal friction angle = 29.5°, and Direction = 75° for Bramble Airport and 200° for Plymouth), for different spatial locations and models. Here we observe something very unusual: the predictions for Bramble Airport for the GASP at all points have very narrow confidence bounds, while the corresponding bounds the predictive process model are

very wide. On the other side, predictions for Plymouth show the opposite behavior: when the bounds for the GASP at all points are wide, the bounds for the predictive process model are relatively narrow. The Gaussian process model at the knots only tends to follow our target model more closely with respect to predictions.

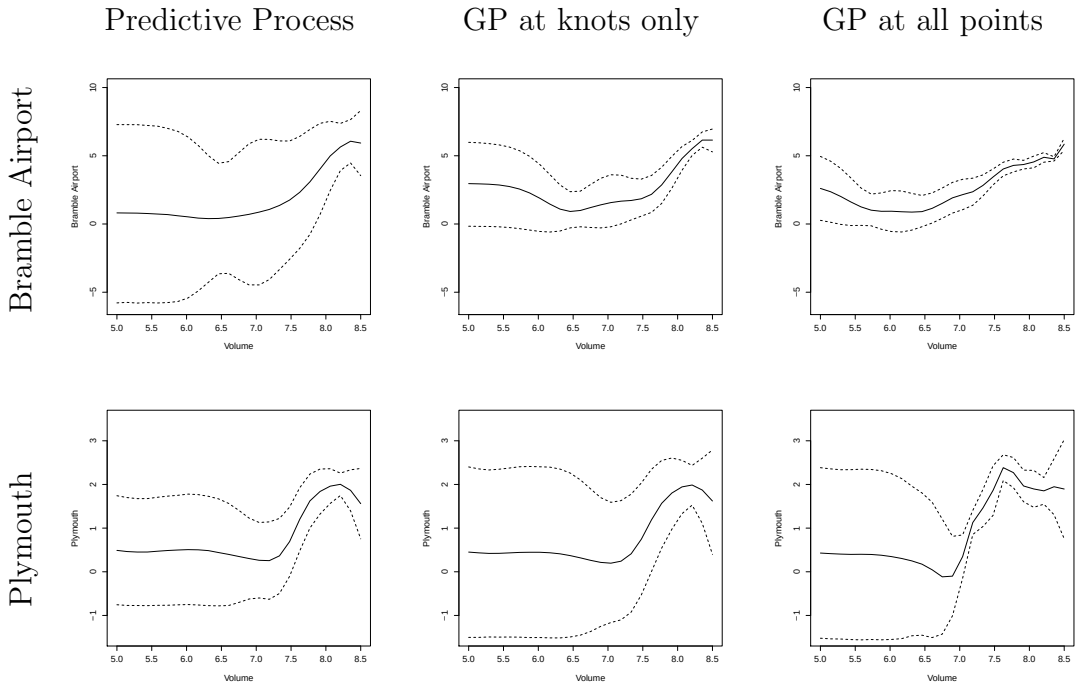


FIGURE 3.12: Posterior predictions (solid lines) and 95% credible bounds (dashed lines) as function of initial volume, with all the other inputs fixed (Basal friction angle = 8.5°, internal friction angle = 29.5°, and Direction = 75° for Bramble Airport and 200° for Plymouthh) for Bramble Aiport (first row) and Plymouth (second row). On the columns, we have different models, which are from left to right: modified predictive process model, regular Gaussian process model at the knots only, and regular Gaussian process model at all points.

From this example, we may come to the conclusion that discarding observations is better for estimation and prediction than using the modified predictive process model. Obviously, this may be a feature of our particular example. Nonetheless, this example shows that we should exercise caution when using this model as an approximation to the regular GASP.

3.5 Discussion

We have observed some downsides of the modified predictive process model as an approximation to conventional Gaussian process in the emulation of computer models. First, it does not encourage the collection of new observations: if we increase the size of the set of knots, we will also be increasing the complexity of the model; if we increase the size of the set of regular observations, after some point, we will not expect to observe improvements in estimation or prediction.

Second, we have found some examples of unusual features of the modified predictive process model: there are some situations where discarding some data may give results closer to our target than using the predictive process model with the entire data set. Also, the use of modularization did not improve the inference from our model. Hence, our recommendation is to use a Gaussian process at as many points as are computationally feasible.

We are not questioning the usefulness of the modified predictive process as a model here; in fact, it is a very attractive model and there are several applications that could benefit from its properties. However, as an emulator, the modified predictive process model does not seem very appropriate, because most of its easiness in computation is achieved at the cost of accuracy in predictions.

We have focused our attention on the modified predictive process because it is a well-defined model. However, there exist several methods for the large data set problem based on approximations to Gaussian process that might be worth considering in the future. For an extensive review of some of these methods, please see Quiñonero-Candela *et al.* (2007).

4

Hierarchical Bayes calibration of inputs

4.1 Introduction

Computer codes use mathematical models to represent physical systems that are difficult to be observed in reality as controlled experiments. They take *inputs*, which are variables describing initial conditions or characteristics of the system, and return *outputs*, variables representing the state of the physical process resulting from the given values of inputs.

In chapters 2 and 3, we have studied the emulation of computer models, where a statistical approximation of the simulator is used to predict the outputs at values of inputs that we have not run. In real life, however, not all the components of the input vector may be observed. Some of them may be variables exclusive to the computer code (also called *tuning parameters*), others may have a physical meaning to the problem but could not be observed in our field data (the *calibration parameters*).

An important aspect of the analysis of computer models is *calibration*: it consists of combining simulator and field observations to find the values of unknown inputs that fit best the observed data. This is typically an *ad hoc* inverse problem, where we

search the input space for values that correspond to simulator (or emulator) outputs that are close, in some sense, to the field observations. In a Bayesian approach, we can find a posterior distribution for the tuning/calibration parameters. For details on the calibration of computer models, see Craig *et al.* (2001), Kennedy and O'Hagan (2001), Goldstein and Rougier (2006) and Bayarri *et al.* (2007a).

In this chapter we describe a different approach to calibration. We combine several sets of auxiliary data to model hierarchically the joint distribution of inputs. Similarly to inverse problem approach, we use observed data to estimate unknown inputs and calibrate the model. We also propose a method for combining the output of the hierarchical analysis with emulation of the computer model in a calibration process. We illustrate the method with the example of a pyroclastic flow simulator. In section 4.2 we give a brief introduction to the test site and the simulator in the example.

4.2 Example: a volcanic hazard model

4.2.1 Test site

We focus our attention on the study of volcanic pyroclastic flows, avalanche-like streams of debris, ash and superheated gas that travel at extremely high speeds (Calder *et al.* 1999). These events may have devastating consequences on both life and property of nearby cities.

The test site used in the analysis is the Soufrière Hills Volcano on the Caribbean island of Montserrat. A dense collection of data about pyroclastic flows and other volcanic activities at Soufrière Hills Volcano is available due to the constant monitoring of the Montserrat Volcano Observatory scientists.

The eruption of the Soufrière Hills Volcano usually involves the growth of a dome by extrusion of lava. The lava dome may grow stably up to the point where it becomes propense to suddenly collapse and break into coarse-grained flows of blocks

and ash. The pyroclastic flow resulting from these gravitational dome collapses may occur with little or no apparent warning and pose a major hazard to even distal areas from the volcano. For more details on the lava dome growth process at Soufrière Hills Volcano, see Hale *et al.* (2009).

4.2.2 Computer model

TITAN2D is a simulator at University of Buffalo (Patra *et al.* 2005) that uses topographical features of the study region to describe the movement of pyroclastic flows. The basic idea behind the computer code is to drop an ellipsoidal pile of granular material from the top of the volcano edifice and then, under a set of initial conditions, describe the gravitational movement of the grains over time using a set of partial differential equations. These differential equations take into account different physical forces, such as gravity, friction between grains and between the erupted material and the basal surface (Pitman *et al.* 2003).

The output is a surface that gives the maximum height of the flow over time at different points on a grid over the study region. In chapter 5, we will discuss some methods to analyze the flow height at several grid points simultaneously, but for now we fix a location on the study region and treat the output as being a scalar.

TITAN2D takes several inputs describing the initial conditions of the pyroclastic flow, but we focus our attention on four variables of special interest for the study of the geological process:

- **Volume:** logarithm to base 10 of the initial flow volume in m^3; this variable corresponds to the volume of the ellipsoidal pile used in the simulation.

- **Direction:** angle in the horizontal plane at which the discharged material begins to move. The initiation angle is measured in degrees from east as reference direction (0°) and runs counterclockwise.

- **Basal Friction Angle:** angle (in degrees) measuring the resistance of the basal surface to the flow material. TITAN2D allows the use of spatially varying basal friction angles, but, in our analysis, we consider this variable to be constant for each simulation.

- **Internal Friction Angle:** angle (in degrees) measuring the rigidity of the flow material.

As noted by Dalbey *et al.* (2008), TITAN2D simulations are relatively insensitive to the value of the internal friction angle, but the other three inputs are very important to determine aspects of the simulated flow. Our approximations to TITAN2D have shown very similar results.

In our analysis, we treat the computer model as a 'black box', making no use of the mathematical model implemented in the code to improve our predictions. The computer model structure is usually too complex to be regarded in the emulator and we approximate simulated flow heights as completed unknown functions of the inputs.

4.3 Relation between inputs in real life

The risk assessment of natural hazards, such as pyroclastic flows, cannot rely solely on field observations, because catastrophic consequences of such events on areas of interest are extremely rare. A computer model is of great importance in this analysis because it can provide information about events resulting from conditions that are possible but have not been observed in the field yet. However, the selection of design points at which we exercise the computer code usually does not take into account which combinations of input values are more realistic, but rather follows a space-filling strategy. Evaluating the computer model at different regions of the input

space allows the emulator to better approximate the shape of the output surface as a function of the inputs.

Hence, the most recommended procedure is to emulate the computer model without imposing any prior knowledge about the real physical process in study. The emulator is just an approximation of the output as a function of the inputs determined by the underlying mathematical model. Information about input variables in real life can be later introduced in the statistical approximation through a joint distribution during the calibration step of the analysis. This type of calibration is very context-specific. In this section, we present a motivation for the model that is going to be introduced in Section 4.4.

4.3.1 Variation of friction coefficient with the size of the pyroclastic flow

The volume V of a pyroclastic flow can be estimated by different methods, including a photographic method and kinematic surveys (Sparks *et al.*, 1998). The photographic method consists of comparing photographs of the same region before and after the volcanic event using a topographic feature of known dimensions as reference, while kinematic surveys uses triangulation and the reflection of a laser shot from a helicopter to measure positions on the study region. The volume of a flow can be estimated by either measuring the dome scar or by quantifying the flow deposit. Another common method of volume estimation is the multiplication of an estimate of the area covered by the flow and the estimated thickness (Legros, 2002).

Heim (1932) was the first to propose a model for gravitational motion of granular material based on dissipation of potential energy by friction forces along the topography. This simple model proposes the use of the basal friction angle ϕ as a measure of flow mobility and defines it as the angle in the vertical direction connecting the flow source to its most distal location downhill:

$$\tan\phi = \frac{H_{\max}}{L_{\max}}$$

where $H_{\max}$ is the maximum vertical drop distance and $L_{\max}$ is the maximum horizontal runout distance.

Some authors criticize the physical meaning of the friction angle as a measure of flow mobility, because it may lead to the false conclusion that the friction is constant along the topography (Legros, 2002). However, we assume the hypothesis of constant friction as a simplification of the model, in connection with our runs of the computer model TITAN2D.

Figure 4.1 shows observed pairs of volume and basal friction angles of pyroclastic flows, with colors representing four different dome-collapsing volcanoes (Montserrat observations are green). These data are described more fully in Section 4.4. The first thing we note is that, instead of being scattered all over the plot, the data points seem to be confined in a very restrict area. Volume of the flow appears to have a negative linear relationship with friction angle, where the correlation between the variables seems to depend on the volcano where the flow was originated.

The relation between the two variables has already been observed in the previous literature. Calder *et al.* (1999) studied the pyroclastic flows produced by the Soufrière Hills Volcano during its 1995-1998 eruption and noted that the largest flows did not have a large mobility, being confined to a few valleys. The effect of these flows on distal areas was associated with the production of dilute surges of ash that traveled much larger distances than their parent flows. Calder *et al.* (1999) have also observed that pyroclastic flows exhibit mobilities very similar to cold rock landslides from other locations on Earth, Moon and Mars.

Legros (2002) compiled a collection of volcanic and non-volcanic flows and showed that the coefficient of solid friction (the arctangent of the friction angle) decreases

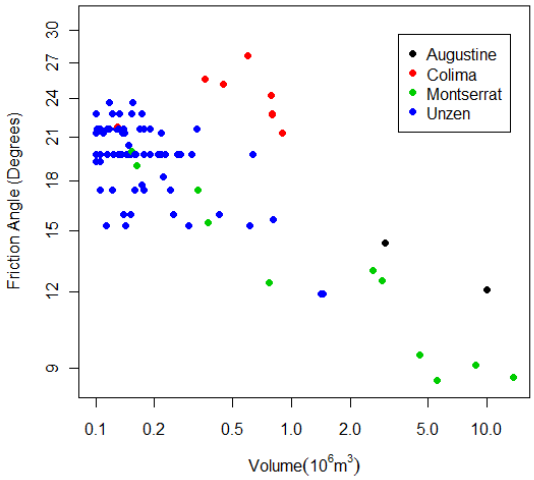


FIGURE 4.1: Friction angle-versus-volume plot for pyroclastic flows of different dome-collapsing volcanoes.

when the volume of the flow increases. Staron and Lajeunesse (2009) presented some plots in log-log scale of inverse coefficient of friction versus volume for several flows, suggesting a very strong linear correlation between the transformed variables. Although some fitted lines were drawn on these plots (separately for terrestrial and Martian flows), Staron and Lajeunesse (2009) did not present a model for the relation between the variables.

There are many hypothesis proposed to explain this inverse relationship between volume and friction angle. One may assume that large flows have a different structure from small flows (e.g., larger flows may have higher gas content, or they might travel in a way that later flows cover recent layers of material instead of dry rock), but such assumptions still need further study. Davies (1982) and Staron and Lajeunesse (2009) suggest that the empirical relation between friction angle and volume is an effect of the topography geometry: small flows have a runout dominated by sliding through the topography, while large flows have a bigger contribution of spreading on their runouts. Legros (2002) argues that the effect of volume on the coefficient

of friction is due to the fact that the slope of the topography generally decreases as the flow moves away from its sources. Thus, larger flows tend to rest on gentler slopes. Sheridan *et al.* (2005) and Charbonnier and Gertisser (2009) propose the use of spatially-varying basal friction angles in TITAN2D in order to account for possible topography effects. This procedure, however, must be implemented in the computer model only and requires a profound study of friction on the topography.

4.4 Hierarchical Bayes model

The data presented in Figure 4.1 represents measurements of volume and basal friction angle of pyroclastic flows generated at different dome-collapsing volcanoes: Augustine Volcano, Alaska (Vallance *et al.*, 2010), Colima Volcano, Mexico (Saucedo *et al.*, 2002; Saucedo *et al.*, 2004; Macías *et al.*, 2006), Soufrière Hills Volcano, Montserrat (E. Calder, personal communication) and Unzen Volcano, Japan (S. Takarada, personal communication). The numbers of observations for each volcano are, respectively, 2, 8, 11 and 72.

Although some volcanoes have very few observations available, we expect to borrow some information from similar volcanoes to better understand the process behind pyroclastic flows. Recall from from Section 4.3 that pyroclastic flows share similarities in the relation between volume and friction angle even with non-volcanic or extraterrestrial flows. The approximately linear relationship between the logarithm of the volume and the logarithm of the tangent of the friction angle for each volcano suggests the use of a conditionally independent hierarchical linear model (Kass and Steffey, 1989). Next, we are going to present two variations of this model and show some results.

4.4.1 A hierarchical model for both intercepts and slope coefficients

Let n_j denote the number of observations available for each volcano $i = 1, \cdots, J$. For each volcano, we observe the $n_j \times 1$ vector $\mathbf{y}_j = (\log\tan\phi_{j,1}, \cdots \log\tan\phi_{j,n_j})$ of logarithms of coefficient of friction for different pyroclastic flows. Let $\boldsymbol{X}_j = (\mathbf{1} \quad \mathbf{x}_j)$, where $\mathbf{x}_j = (\log V_{j,1}, \cdots \log V_{j,n_j})$ is the corresponding vector of log-volumes.

We consider the following hierarchical model:

$$\begin{aligned} \mathbf{y}_j &\sim N_{n_j}(\boldsymbol{X}_j\boldsymbol{\theta}_j, \sigma_y^2\boldsymbol{I}) \\ \boldsymbol{\theta}_j &= \begin{pmatrix} \theta_{j1} \\ \theta_{j2} \end{pmatrix} \sim N_2(\boldsymbol{\mu}, \Sigma) \end{aligned} \tag{4.1}$$

This model fits a separate line for each volcano, where the corresponding regression coefficients are tied together hierarchically by a common mean $\boldsymbol{\mu}$ and a common variance structure at the second stage. The variance σ_y^2 represents the "within-volcano variation", while Σ is the covariance matrix of the regression parameters between volcanoes.

Let p be the number of regression coefficients for each volcano (in our case, $p = 2$), and set $\Lambda = \Sigma/\sigma^2$. We perform a Bayesian analysis of the model, assigning to the hyperparameters the reference prior with respect to the Group Ordering $\{\beta, \sigma^2, \Lambda\}$, $\{\sigma^2, \beta, \Lambda\}$ or $\{\sigma^2, \Lambda, \beta\}$ (Yang and Chen, 1995):

$$p(\boldsymbol{\mu}, \Lambda, \sigma_y^2) \propto |G\sum_{i=1}^{n}(B_i \otimes B_i)G^t|^{1/2}/\sigma_y^2 \tag{4.2}$$

where $B_i = (I_p + X_i^t X_i \Lambda)^{-1} X_i^t X_i$ and G is a $(p(p+1)/2) \times p^2$ matrix of zeros and ones $\partial \text{vec}V/\partial \text{vecp}V$, where V is a $p \times p$ symmetric matrix.

One needs to be very careful when specifying prior distributions for the variance components of a hierarchical model, because standard choices such as constant prior

density or scale-invariant priors may lead to improper posteriors (Hobert and Casella, 1996). The prior distribution in (4.2) yields a proper posterior when we have at least $2p + 1$ full rank design matrices.

MCMC algorithm

We choose to update the precision $\tau_y = 1/\sigma_y^2$ instead of the corresponding variance for simplicity. Posterior samples for the regression coefficients and τ_y can be obtained through Gibbs sampling, while the Λ must be updated by a Metropolis-Hastings step. The full conditionals are:

$$
\begin{aligned}
\boldsymbol{\theta}_j|\boldsymbol{\mu}, \tau_y, \Sigma, \mathbf{y}, X &\sim N(\left[X_j^t X_j + \Lambda^{-1}\right]^{-1} \left[X_j^t \mathbf{y}_j + \Lambda^{-1}\boldsymbol{\mu}\right], \\
&\qquad \tau_y^{-1} \left[X_j^t X_j + \Lambda^{-1}\right]^{-1}) \\
\boldsymbol{\mu}|\tau_y, \Sigma, \mathbf{y}, X &\sim N\left(\boldsymbol{m},\, \tau_y^{-1}\boldsymbol{V}\right) \\
\tau_y|\Sigma, \mathbf{y}, X &\sim Ga\left(\frac{n-p}{2},\, \frac{S_y^2}{2}\right) \\
p(\Lambda|\tau_y, \mathbf{y}, X) &\propto |\boldsymbol{V}|^{1/2} \prod_{i=1}^{J} \left[|I_p + X_i^t X_i \Lambda|^{-1/2}\right] \exp\left\{-\frac{\tau_y}{2}(S_y^2 - \boldsymbol{m}^t \boldsymbol{V}^{-1} \boldsymbol{m})\right\} \\
&\qquad \times |G \sum_{i=1}^{n} (B_i \otimes B_i) G^t|^{1/2}
\end{aligned}
$$

where $\boldsymbol{V} = [\sum_{i=1}^{J} B_i]^{-1}$, $\boldsymbol{m} = \boldsymbol{V} \sum_{i=1}^{J} A_i$, $S_y^2 = \sum_i (\mathbf{y}_i^t \mathbf{y}_i - \mathbf{y}_i^t X_i \Lambda A_i)$, $A_i = (I_p + X_i^t X_i \Lambda)^{-1} X_i^t \mathbf{y}_i$ and $B_i = (I_p + X_i^t X_i \Lambda)^{-1} X_i^t X_i$.

We update Λ using the one-step of a metropolisized hit-and-run sampler described by Yang and Chen (1995).

Results

The posterior distributions for first-stage regression coefficients corresponding to pyroclastic flows at the Soufrière Hills Volcano are in Figure 4.2. The reasonable amount of data observed in Montserrat results in very small uncertainty in the regression coefficients, and the trace plots show that the MCMC is mixing well. However, as seen in Figure 4.3, the small number of volcanoes and the existence of volcanoes with very few observations seem to hurt the estimation of second-stage parameters. Note that the posterior distributions of the precisions of the intercept (τ_1) and the slope (τ_2) are very similar in shape to the distribution of the first-stage precision τ_y due to the poor mixing of Λ.

4.4.2 A hierarchical model for the slope coefficients only

Figure 4.1 suggests that the slopes of the regression lines of log-friction coefficient on log-volume for different volcanoes are similar, but the intercepts seem to be very different. This suggests that connecting the intercepts through a second-stage common distribution may not be adequate. As shown before, our data give very poor estimation of second-stage parameters under the full hierarchical model in (4.1).

We now consider a model where only the slopes of the regression lines are tied together. For each volcano j:

$$\begin{aligned} \mathbf{y}_j &\sim N_{n_j}(\boldsymbol{X}_j\boldsymbol{\theta}_j, \sigma_y^2\boldsymbol{I}) \\ \boldsymbol{\theta}_j &= \begin{pmatrix} \theta_{j1} \\ \theta_{j2} \end{pmatrix}, \text{ where } \quad \theta_{j2} \overset{iid}{\sim} N(\mu, \sigma_\theta^2) \end{aligned} \tag{4.3}$$

For this new model, we set $\lambda = \sigma_\theta^2/\sigma_y^2$ and specify a prior distribution similar to (4.2):

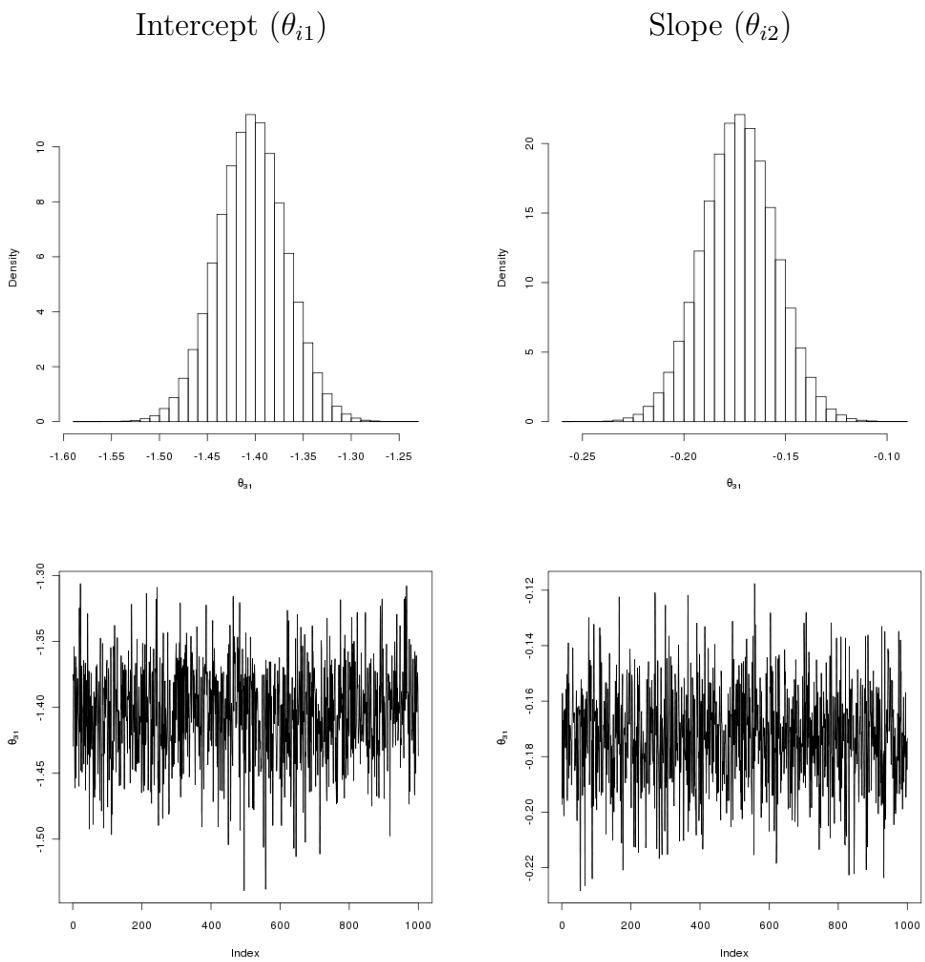


FIGURE 4.2: Histograms (first row) and trace plots (second row) of posterior samples of the hierarchical model intercept (first column) and slope (second column) coefficients for the Soufrière Hills Volcano. These parameters are tied together with intercepts and slopes for other volcanoes through a second-stage common distribution.

$$p(\boldsymbol{\mu}, \Lambda, \sigma_y^2) \quad \propto \quad (\sum_{i=1}^{n} b_i^2)^{1/2} / \sigma_y^2$$

where $b_i = (1 + \lambda \mathbf{x}_i^t \mathbf{x}_i)^{-1} \mathbf{x}_i^t \mathbf{x}_i$.

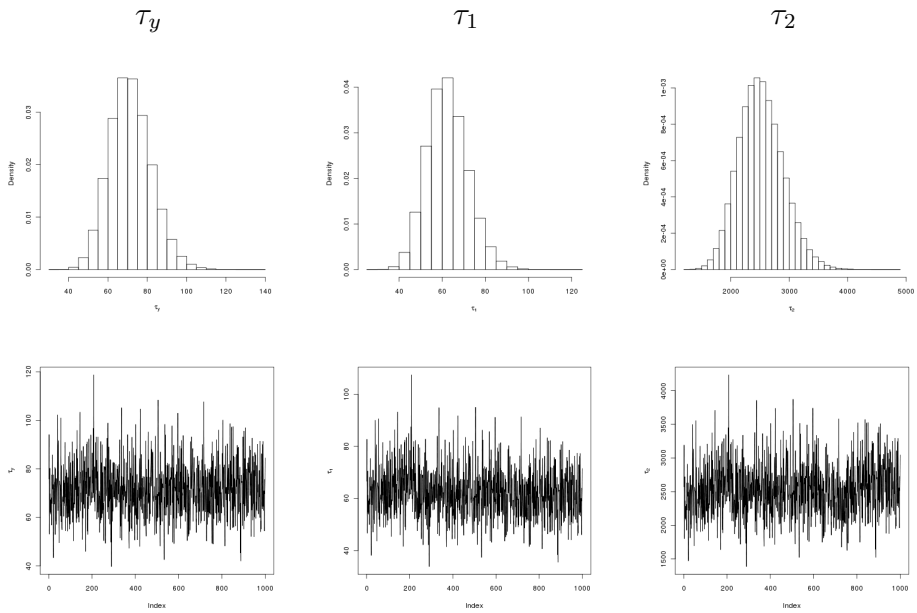


FIGURE 4.3: Histograms (first row) and trace plots (second row) of posterior samples of the "within-volcano" (first column), intercept (second column) and slope (third column) precisions.

MCMC Algorithm

There are some slight differences between the MCMC algorithms for the models in (4.2) and in (4.3). Slopes must now be updated in a different block from the intercepts. We use a Metropolis-Hastings step to draw samples of λ, with the a proposal distributions corresponding to a random walk on $\log \lambda$, while all the other parameters are sampled using Gibbs sampling. The full conditional distributions for the second hierarchical model are:

$$
\begin{aligned}
\theta_{j1}|\boldsymbol{\theta}_{-j,-1}, \boldsymbol{\mu}, \tau_y, \tau_\theta, \mathbf{y}, X &\sim N\left(\frac{\sum_k y_{jk} - \theta_{j2}\sum_k x_{jk}}{n_j}, \frac{1}{\tau_y n_j}\right) \\
\theta_{j2}|\boldsymbol{\theta}_{-j,-2}, \boldsymbol{\mu}, \tau_y, \tau_\theta, \mathbf{y}, X &\sim N\left(\frac{\tau_y\left(\sum_k x_{jk}y_{jk} - \theta_{j1}\sum_k x_{jk}\right) + \tau_\theta\mu}{\tau_y\sum_k x_{jk}^2 + \tau_\theta},\right. \\
&\qquad \left.\frac{1}{\tau_y\sum_k x_{jk}^2 + \tau_\theta}\right)
\end{aligned}
$$

$$
\begin{aligned}
\mu|\tau_y, \theta_{\bullet 1}, \Sigma, \mathbf{y}, X &\sim N\left(m,\, \tau_y^{-1}v\right) \\
\tau_y|\Sigma, \theta_{\bullet 1}, \mathbf{y}, X &\sim Ga\left(\frac{n-p}{2},\, \frac{S_y^2}{2}\right) \\
p(\lambda|\tau_y, \theta_{\bullet 1}, \mathbf{y}, X) &\propto v^{1/2}\prod_{i=1}^{J}\left[(1+\lambda\mathbf{x}_i^t\mathbf{x}_i)^{-1/2}\right]\exp\left\{-\frac{\tau_y}{2}(S_y^2 - m^2/v)\right\} \\
&\quad \times(\sum_{i=1}^{n} b_i^2)^{1/2}
\end{aligned}
$$

where $v = \left[\sum_{i=1}^{J} b_i\right]^{-1}$, $m = v\sum_{i=1}^{J} a_i$, $S_y^2 = \sum_i((\mathbf{y}_i - \theta_{i1}\mathbf{1})^t(\mathbf{y}_i - \theta_{i1}\mathbf{1}) - \lambda a_i\mathbf{x}_i^t(\mathbf{y}_i - \theta_{i1}\mathbf{1})$, $A_i = (1+\lambda\mathbf{x}_i^t\mathbf{x}_i)^{-1}\mathbf{x}_i^t(\mathbf{y}_i - \theta_{i1}\mathbf{1})$ and $b_i = (1+\lambda\mathbf{x}_i^t\mathbf{x}_i)^{-1}\mathbf{x}_i^t\mathbf{x}_i$.

Results

Figure 4.4 shows the posterior distributions of the second hierarchical model, where only slopes of regression lines for different volcanoes are tied together hierarchically. We note that the posterior distribution of the slope corresponding to the Soufrière Hills Volcano has not changed very much with the modification in the model, but the intercept shows some reduction in the posterior variability.

Figure 4.5 shows that the estimation of second-stage parameters has not improved very much with the changes in the model. The precision of the slope τ_θ is highly

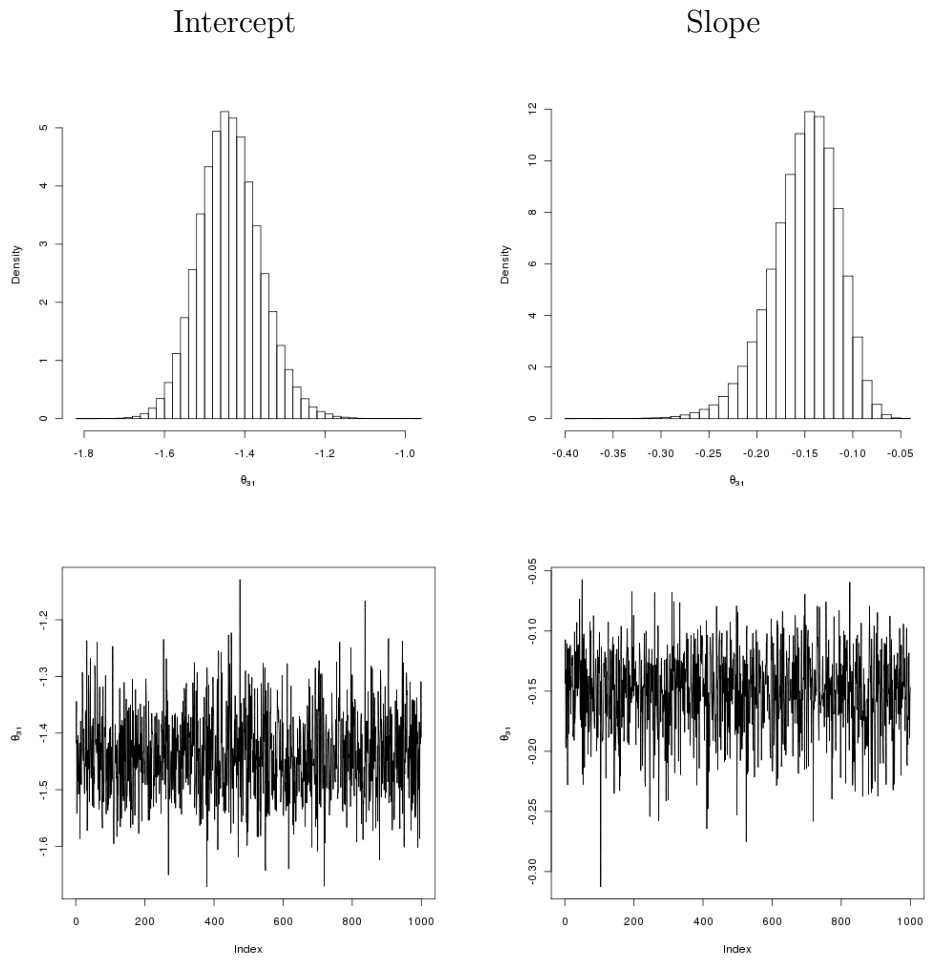


FIGURE 4.4: Histograms (first row) and trace plots (second row) of posterior samples of the hierarchical model intercept (first column) and slope (second column) coefficients for the Soufrière Hills Volcano. Only the slopes are tied together with intercepts and slopes for other volcanoes through a second-stage common distribution.

correlated with the first-stage precision due to poor mixing on the MCMC chain of λ.

The posterior regression lines of log-coefficient of friction on log-volume for each volcano given by the second hierarchical model are in Figure 4.6. The data points shown in Figure 4.1 are repeated in each panel of Figure 4.6 and the colors of the regression lines also follow the legend of that figure. The panels of Figure 4.6 differ in which volcano has a 95% confidence bounds for the regression line (dashed

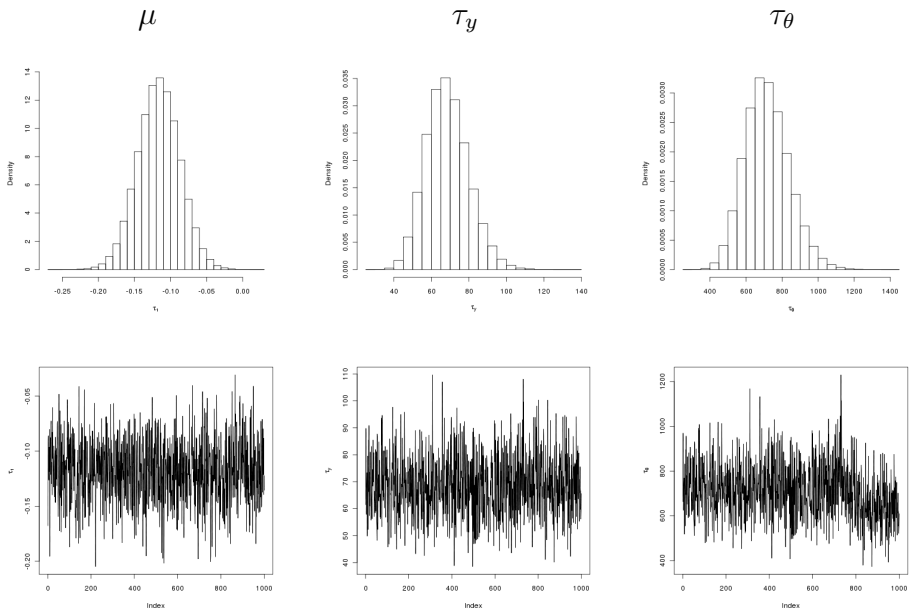


FIGURE 4.5: Histograms (first row) and trace plots (second row) of posterior samples of the "within-volcano" (first column), intercept (second column) and slope (third column) precisions.

lines) and for the data points (dotted lines) being represented. We note that these bounds change from volcano to volcano according to the number of points and the concentration of data points, as we would expect.

Although the second-stage parameters of the hierarchical model cannot be estimated very well, we saw that the posterior samples of first-stage parameters (regression coefficients) are very reasonable. We choose the results of the second hierarchical model, where only the slopes are tied together by common mean and variance, for further analysis because this hypothesis seems to agree with our data set and historical data (see for example Staron and Lajeunesse, 2009).

4.5 Binding emulators with auxiliary information

As discussed in chapters 2 and 3, Gaussian process emulators are fast surrogates for the computer models. Statistical approximations to the computer model are

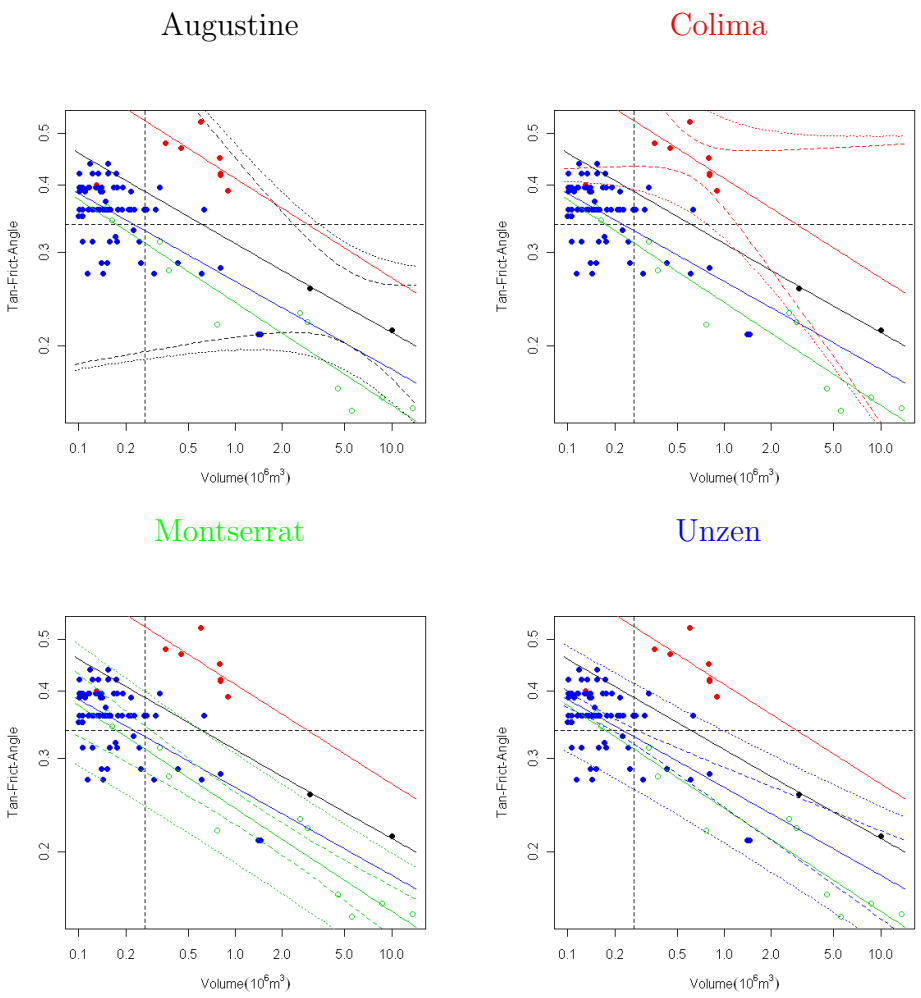


FIGURE 4.6: Posterior hierarchical model regression lines for each volcano with corresponding 95% confidence bounds for the regression line (dashed lines) and for the data points (dotted lines).

important for the analysis of the effect of inputs on the output, because it is unfeasible to use the computer model in the analysis directly due to the long running time usually needed for each simulation.

We are now interested in combining emulators with the model for the relation between inputs presented in Section 4.4. Using computer model data and the methods for emulation of computer models described in Chapters 2 and 3, assume that we have posterior samples of GASP parameters available. These samples can be used

to create a posterior predictive distribution for the output at values of inputs that we have not used to evaluate the computer model yet.

Before the emulation process, it is important to select values of inputs that will provide good statistical approximations to the computer model. Once the emulator is ready to be used in the analysis, we can focus our attention on combinations of values of inputs that are more probable to be observed in reality.

Figure 4.7 shows some realizations from random curves obtained by transforming the posterior distribution of the regression lines of log-coefficient of friction on log-volume at the Soufrière Hills Volcano, summarized in Figure 4.6, back to the original units. Assuming a linear model for the transformed variables, the hierarchical model gives us a distribution of the basal friction angle $\phi(V)$ as a function of the volume V for a pyroclastic flow generated in Montserrat.

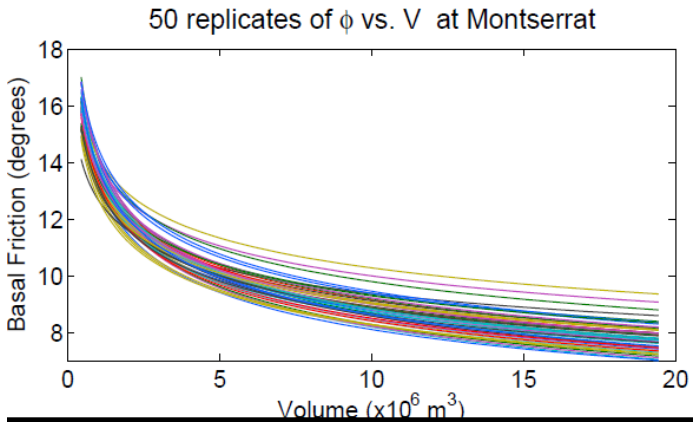


FIGURE 4.7: 50 replicates of ϕ vs V at Monsterrat using the hierarchical model in (4.3).

Let $y(V, \phi, \nu)$ be a scalar output of our computer model (e.g. maximum flow height at a specific location in the study region) as a function of the inputs, where V is the volume of the flow, ϕ is the basal friction angle and ν is a vector of other inputs. Assume that the emulator $\hat{y}(V, \phi, \nu)$ is available, using the posterior predictive distribution of a Gaussian process. The calibration process consists of plugging

posterior realizations of the function $\phi(V)$ into the emulator predictions. Hence, the calibrated prediction of the computer model prediction is $\hat{y}(V, \phi(V), \nu)$.

It is a scientific knowledge, usually confirmed by empirical observations, that the basal friction angle falls asymptotically to an approximate value of 8° for flows with volume much larger than $10^7 m^3$. We include this information in our analysis by truncating each random curve of $\phi(V)$ at 8°, i.e. we consider the new function $\tilde{\phi}(V) = \max(\phi(V), 8)$.

In terms of the MCMC algorithm, for each sampled values of GASP parameter generated during the emulation process, we draw a point from the posterior cloud of slopes and intercepts from one of the hierarchical models presented in Section 4.4. We then construct the regression line and transform it back to the same units of the computer model, obtaining a curve similar to the ones in Figure 4.7. The calibrated emulator should be constrained to values of inputs lying on the sampled curve $\tilde{\phi}(V)$. At each new step of the chain, we repeat this process.

Figure 4.8 shows a comparison between the emulator predictions of flow height at two different locations in Montserrat as a function of the flow volume before (first column) and after (second column) the calibration procedure. For the non-calibrated emulator predictions, we fixed the friction angle at 8.5°. We note that point predictions seem to have about the same shape regardless of the calibration, but the variability of the predictions expressed in the confidence bounds has increased with the calibration of inputs. The confidence bounds are approximately parallel for small values of volume because the hierarchical model gives friction angles much higher than those in our computer model data. Since we are evaluating the emulator in a region with no observations, the GASP gives predictions with constant mean and constant variance. The calibration also introduce more variability in the emulator and wider confidence bounds are expected.

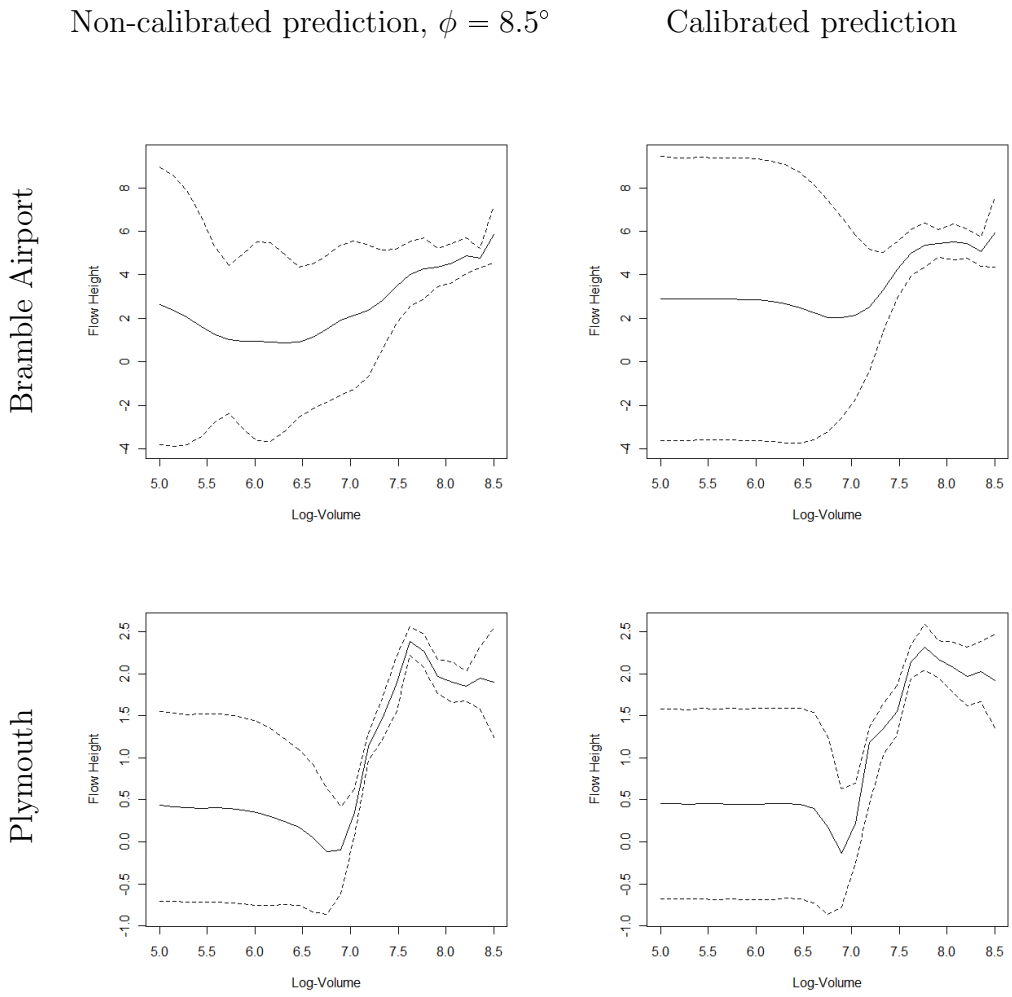


FIGURE 4.8: Posterior predictions (solid lines) and 95% credible bounds (dashed lines) as function of initial volume, with some of the other inputs fixed (Internal friction angle = 29.5°, Direction = 75 for Bramble Airport and 200 for Plymouth) for Bramble Aiport (first row) and Plymouth (second row). On the first column, we have the non-calibrated emulator with basal friction angle fixed at 8.5° and on the second column we have the calibrated emulator with friction angle as random function of volume.

4.6 Discussion

In this chapter, we presented some methods for calibration of computer model inputs applied to an example of pyroclastic flows. The calibration process is extremely important for the improvement of computer models (and their statistical approximations) as representations of the reality.

As we showed in our applications, the calibration process is usually very context-specific. Calibration usually requires some extensive knowledge of the physical process being studied, in order to transmit many physical system aspects to the statistical model.

The proposed model for the relation between computer model inputs should also vary between different applications. As part of a sensitivity analysis, one can also fit different models to the field observations in order to evaluate the effect of different calibrations on the uncertainties of the Gaussian process emulator.

5

Uncertainty in the analysis of geophysical features

5.1 Introduction

TITAN2D is a simulator of pyroclastic flows that, among other things, returns as an output a surface of maximum flow height over time for a given study region. For simplicity, we have considered only scalar functions of this output in our analysis, e.g. the value of the surface at a fixed location on the study region. However, the risk assessment of a pyroclastic flow event usually requires the analysis of several locations on the study region simultaneously, as we are interested in the construction of hazard maps.

We want to emulate a vector of outputs, representing the simulated flow height at different locations on the study region, as a function of the computer model inputs. Typically, we are interested in selecting locations from a small portion of the region, where it is more populated and human facilities are present. Although there are many methods available for multivariate Gaussian process modeling, we show in Section 5.2 that, without additional information about the behavior of pyroclastic flows, these methods cannot provide satisfactory results in the emulation of our computer model.

We believe that the topography of the study region is crucial to determine dependence between the values of flow height at two different geographical locations. For example, if the two locations are along the same valley, their corresponding flow heights for the same values of inputs should a much higher correlation than if the two locations were at the top of different hills. Section 5.3 presents some topographical features that may be of interest in the analysis of pyroclastic flow height.

The topographic surface of a region is usually represented by a regular array of elevation values, called a Digital Elevation Model (DEM). There are several methods for construction of DEMs, such as the use of GPS units, aerial photogrammetry and airborne laser scanning. Even though these methods are expensive to implement, they produce imperfect DEMs. Very specific measurement errors and interpolation can produce distortions from reality that cannot be well described by additive random fields. Also, the DEM may differ from the real topography because the latter is constantly changing as an effect of erosion and deposit of debris by pyroclastic flows.

We present different sources of uncertainty in the DEM in Section 5.3, while in Section 5.4 we propose a probabilistic model model for distortions in the Digital Elevation Model. In Section 5.4.4 we discuss how our proposed model can be used to propagate the uncertainty in the DEM to the statistical approximations of the computer model. Section 5.5 presents our concluding remarks.

5.2 Multivariate emulators and pyroclastic flows

In August 2008, the government of Montserrat divided the southern two-thirds of the island into zones of controlled access according to the hazard level of volcanic activity at the Soufrière Hills Volcano. Since then, locations that could be easily reached by simulated pyroclastic flows at reasonable values of inputs, such as the ones studied in the previous chapters, Plymouth and Bramble Airport, have lost most of their practical importance in municipal guidance and risk assessment. These

locations are now in evacuated areas and have only theoretical importance in the study of pyroclastic flows, for better understanding of the physical process.

A region of growing practical interest for risk assessment in Montserrat is the Belham River Valley, in the western part of the island. The government of Montserrat restricts the access to this area only for very high hazard level, i.e. when there is a large and unstable dome on the volcano. The Belham River Valley has started to be occupied recently, which brings the need for constant monitoring of pyroclastic flow hazard in that area. Figure 5.1 shows several points of interest along the Belham valley (E. Calder, personal communication).

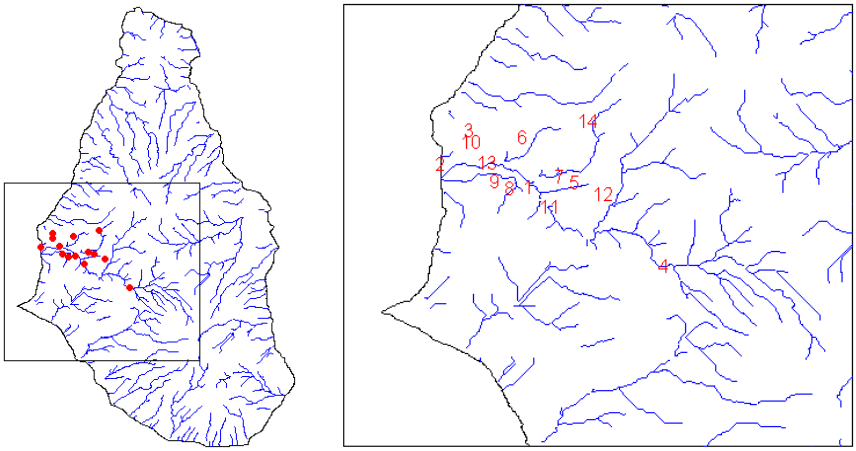


FIGURE 5.1: Locations of interest for the risk assessment of pyroclastic flows in Montserrat. The right panel correspond to a zoom of the square area on the left panel.

For reference, the location numbered by 11 in the right panel of Figure 5.1 represents the historical limit of pyroclastic flows in that direction until January 2010. Many of the locations represented in the figure were never hit by a pyroclastic flow. This rarity of events is also observed in our computer model. Out of 1,430 TITAN2D runs, only 174 had non-zero flow-height at the four closest locations to the volcano. Figure 5.2 shows a scatter plot matrix of these flow-heights and TITAN2D inputs. Note that the simulated flow-height at location 1 is highly correlated with the heights

at locations 11 and 12.

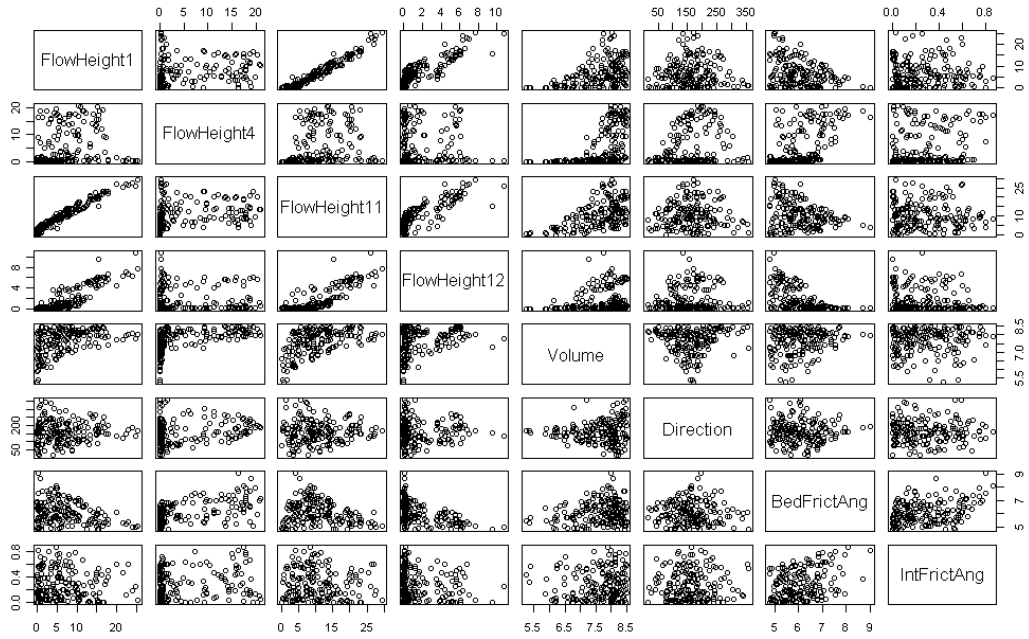


FIGURE 5.2: Scatter plot matrix of the inputs and the flow-height at the 4 closest locations to the volcano.

We are interested in constructing an emulator for the vector of outputs $\boldsymbol{y}(\boldsymbol{z}) = (y_1(\boldsymbol{z}) \cdots y_k(\boldsymbol{z}))^t$ (whose components are flow-heights at different locations) as a function of the vector of inputs $\boldsymbol{z}$. Gelfand *et al.* (2004) give an extensive review of multivariate spatial processes. We adopt their method of choice, the Linear Model of Coregionalization (also discussed in Wackernagel, 2003). This model is particularly interesting for emulation because it is based on dimension reduction and it is computationally inexpensive compared to other multivariate methods. The linear model of coregionalization assumes that, apart from the regression mean, each component of the output vector $y_i(\boldsymbol{z})$ is a linear combination of k independent Gaussian processes with unit variance and correlation function $c(\boldsymbol{z}, \boldsymbol{z}'; \boldsymbol{\beta}_i)$. The multivariate spatial model is of the form

$$\boldsymbol{y}(\boldsymbol{z}) = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{A}(\boldsymbol{z})\boldsymbol{v}(\boldsymbol{z}) \tag{5.1}$$

where $\boldsymbol{v}(\boldsymbol{z}) = [v_i(\boldsymbol{z})]_{i=1}^k$ and

$$v_i(\boldsymbol{z}) \overset{ind}{\sim} GP(\mathbf{0}, c(\boldsymbol{z}, \boldsymbol{z}'; \boldsymbol{\beta}_i))$$

where $c(\boldsymbol{z}, \boldsymbol{z}'; \boldsymbol{\beta}_i)$ is a correlation function, e.g. the power-exponential correlation function.

Let $\Gamma = \boldsymbol{A}(\boldsymbol{z})\boldsymbol{A}(\boldsymbol{z})^t$ be the covariance matrix for the k geographical locations at input point $\boldsymbol{z}$. If we assume stationarity, $\boldsymbol{A}(\boldsymbol{z}) = \boldsymbol{A}$.

Unfortunately, our experience with this method shows that it does not work very well when the components of the vector of outputs $\boldsymbol{y}(\boldsymbol{z})$ are highly correlated. In this case, the number of independent Gaussian processes could be reduced to accommodate the high linear correlations, but there is no way of telling how many v_i's we should use beforehand. If we insist in fitting the full model to the data, there will be very poor mixing in the MCMC chain and some correlation parameters might converge to zero in order to reflect the nonexistence of one or more of the independent Gaussian processes in (5.1).

Basically, there are two problems associated with the use of multivariate spatial processes to model the height values of a pyroclastic flows at different geographic locations. The first one is the *zero problem*: for many runs, the pyroclastic flow does not get to our points of interest on the study region and we register a zero flow-height. There are regions in the input space that produce a value of output (for one or more components of $\boldsymbol{y}(\boldsymbol{z})$) constant at zero, which is a problem for methods based on Gaussian processes. The second problem is that we are interested in analyzing the height of a smooth surface at different points that are close in space, so high correlations between observations should be expected and so coregionalization should work poorly here.

We believe that the topographic information should play a big role in the multivariate analysis of flow heights. For example, if two points are along the same valley,

we should expect their flow-heights to have similar values for each simulation. In the next, we present some topographic feature that may be important for the analysis of pyroclastic flows.

5.3 DEM and important geophysical features

5.3.1 Watersheds and Valleys

Watersheds are sets of directed segments representing possible streams in different valleys. The endpoints of these segments lie on the bottom of a valley and go downhill through the topography. We use the hydrological modelling package of the free open source GRASS GIS software (GRASS Development Team, 2008) to produce watershed lines using a Digital Elevation Model. The blue lines in Figure 5.3 represent the watersheds in Montserrat and the red lines represent 1-m contours of flow-height for different TITAN2D runs.

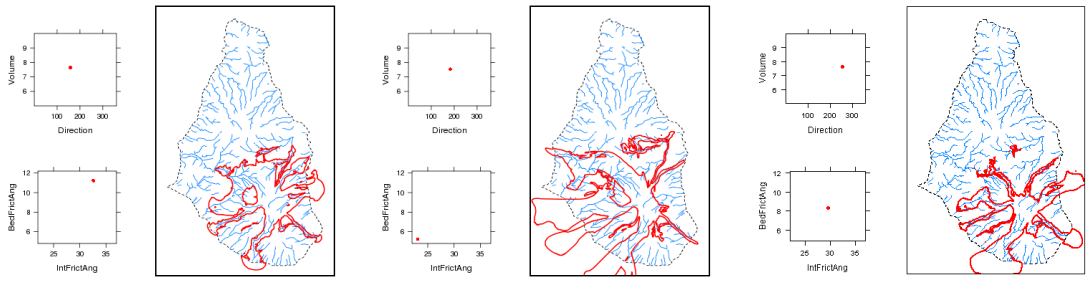


FIGURE 5.3: Plots of 1-m contour flow-height (red lines) and watersheds (blue lines) using simulations of TITAN2D at Montserrat. The small plots on the left of each panel represent the input values of $\log_{10}$ volume, initiation angle and (basal and internal) friction angles.

Figure 5.3 clearly shows that simulated pyroclastic flow contours tend to follow the direction of different watersheds. How closely each branch of the flow surface will follow a watershed, however, depends on the values of inputs in the simulation. We also expect the the shape of flow to be similar in shape to its valley.

We expect to use watersheds to describe where each branch of the flow surface will be centered at and the valleys on the island to model the spreading of the simulated flow. In order to obtain a representation of a given valley shape, we developed and implemented an algorithm that uses the Digital Elevation Model to produce valley contours. The basic idea of the algorithm is use the watershed as a reference to obtain lines on the valley walls that are at a specified height h relative to the watershed line. Theoretically, these valley contours represent the projection on the horizontal plane of the level of a pyroclastic flow filling the valley with a constant height h. This model, however, is an oversimplification of the flow system, since is does not consider the momentum of the flow material downhill.

We now describe our algorithm for obtaining valley contours. Each step of the procedure is represented by a panel in Figure 5.4 For each point $\mathbf{x}$ in the watershed line:

- **Step 1:** Find the elevation $z(\mathbf{x})$ in the DEM;
- **Step 2:** Find the isohypse (contour line of the DEM) corresponding to the elevation $z(\mathbf{x}) + h$;
- **Step 3:** In a perpendicular direction to the watershed at $\mathbf{x}$, find the two closest points in the isohypse to $\mathbf{x}$: one to the left and one to the right of the watershed;

For each relative height h the valley contour consists of connecting the points on the left, and those on the right of the watershed line. Figure 5.5 shows valley contours obtained from the algorithm with different h's and compares these contours to the contour of a simulated pyroclastic flow.

Future work includes modeling the pyroclastic flow contours as function of the valley contour. We plan to use a new coordinate system in terms of distances from the watershed and distances down the watershed. Each curves in this new coordinate

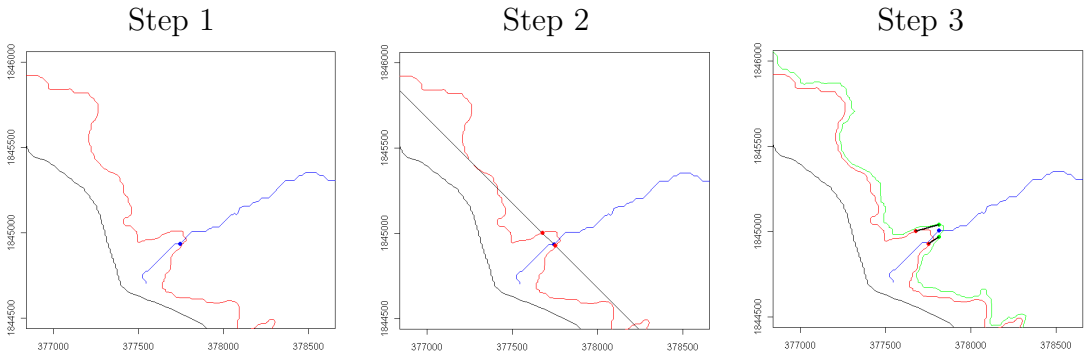


FIGURE 5.4: Graphical representation of the algorithm for obtaining valley contours.

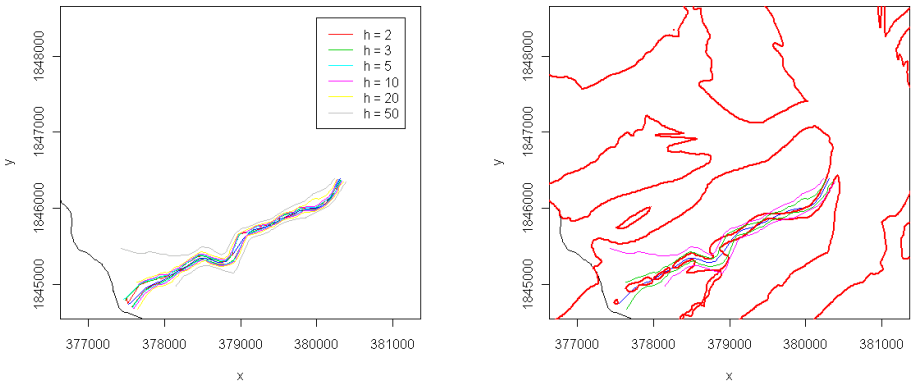


FIGURE 5.5: Left panel: valley contours for different values of height h relative to the watershed. Right panel: comparison between the valley contours and a 1-m contour of pyroclastic flow simulated at Volume $= 10^{7.4} m^3$, Direction $= 179.6°$, Basal friction angle $= 7.1°$, and Internal friction angle $= 30.5°$

system would then be decomposed using a wavelets analysis and the wavelet coefficients of the pyroclastic flow contour would be modeled as functions of the computer model inputs and the valley contour wavelet coefficients.

5.3.2 Uncertainty in the DEM

A Digital Elevation Model (DEM) is a representation of the elevation in a topographic surface that has an application in different areas of research, such as hydrologic

modeling and analysis of surface erosion. It represents a continuous variable (height) over a two-dimensional surface by a regular array of elevation values. DEM can be generated from a variety of methods, including ground measurements with a GPS unit, aerial photogrammetry, airborne laser scanning and the use of satellites. The elevation measurements are usually collected at a irregularly distributed points, thus an interpolation method is commonly used to produce a regular array of values.

Both data measurement and interpolation introduce errors in the generation of DEMs, and these errors can often be systematic (Hebeler and Purves, 2009). For example, a very common and relatively precise method for measuring elevations is the LIght Detection And Ranging (lidar) mapping (Csatho *et al.*, 2008), which is based on the calculation of the reflection of a laser ranging from a plane. The calculation of an airborne laser mapping can be affected by several error sources, such as surface orientation, laser calibration and recovery (or not) of the laser point. Errors in calculation of laser reflection not only lead to uncertainty in the inferred elevation value, but also in the *location* at which the elevation value was recorded. Is possible that the elevation of a topographic feature, such as a cliff, is measured correctly, but the location of the feature in the study region is not, resulting in a distortion of the Digital Elevation Model. This distortion in the DEM cannot be captured by models of additive errors, typically used to measure uncertainties in the DEM. In Section we propose a new probability model for elastic distortions in the DEM.

The DEM can also be seen as an input of the computer model TITAN2D, since it needs to be provided to the computer code in order to produce simulated pyroclastic flows. The recent work of Stefanescu *et al.* (2010) showed that uncertainty in DEM does affect the outcome of TITAN2D. They present a procedure to analyze the sensitivity of the output to the DEM as input.

5.4 Probability model for elastic distortions

5.4.1 Overview

There are many applications where the interest is to estimate a surface $\mu(x)$ as a function of the location x. Typically, we collect observations Y and treat them as random variables using an error model to estimate the value of $\mu(x)$. We consider two types of error model:

- *Additive error*: For some zero-mean random variables ϵ_i:

$$Y(x_i) = \mu(x_i) + \epsilon_i$$

- *Elastic distortion*: For some near-by points $t_i \approx x_i$, we have

$$Y(x_i) = \mu(t_i)$$

  i.e., Y is approximating μ at the wrong place.

Note that these two models can be combined in a single approach. As discussed in subsection 5.3.2, approaches that consider only additive errors may not be appropriate for data such as DEMs, because not only residuals ϵ_i may be uncertain, but also the location x_i.

For the problem of modeling spatial distortions, we propose an approach motivated by elasticity in mechanical physics. Consider a spring of length ϵ_{ij}. This spring will be compressed or stretched, and let $\mathbf{x}_i$ and $\mathbf{x}_j$ be the new positions of the spring edges. The elastic force acting on the spring is given by $-k|d(\mathbf{x}_i, \mathbf{x}_j) - \epsilon_{ij}|$, where d is the Euclidean distance and k is the elastic modulus (Hooke's constant).

Assume that we have little masses connected by springs, as represented in Figure 5.6. If we consider a bidimensional problem, we can think of a grid of points connected by springs.

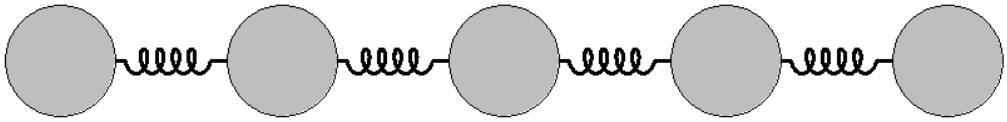


FIGURE 5.6: Representation of blocks connected by springs for illustration of the elastic model.

We want to move the blocks randomly. How easy it is to get a new configuration is may depend on the total potential energy of the system:

$$E(\mathbf{x}) \quad = \quad \frac{k}{2}\sum_{i<j} w_{ij}(d(\mathbf{x}_i,\mathbf{x}_j)-\epsilon_{ij})^2$$

where w_{ij} is an indicator that locations x_i and x_j are connected by a spring in the model.

We then can assign to $\{x_i\}$ a joint probability density function of the same form as the Boltzmann Distribution:

$$f(\{\mathbf{x}_1,\cdots,\mathbf{x}_n\}) \quad \propto \quad \exp(-E(\mathbf{x})) = \exp\left(-\frac{k}{2}\sum_{i<j} w_{ij}(d(\mathbf{x}_i,\mathbf{x}_j)-\epsilon_{ij})^2\right) \quad (5.2)$$

The parameter k, called elasticity constant in physical applications, controls the rigidity of the imaginary springs connecting points. Large values of k generates configurations where the locations $\mathbf{x}$ stay very close to their original positions with relaxed springs. Generalizations of this model may include different elasticity constants k_{ij} for each spring in the grid. The parameters ϵ_{ij} is the length of the spring connecting locations $\mathbf{x}_i$ and $\mathbf{x}_j$ at equilibrium. For simplicity, we consider regular grids, where $\epsilon_{ij} = \epsilon$ for any i and j.

In order to make this distribution proper, we need to fix the location of some points, usually points at the boundary of the study region. This restriction requires observations at some boundary points. For some examples, such as DEMs in

Montserrat, this is not a problem, because we know that the elevation should be zero at the sea.

It is also important to apply a geometric constraint to the model in 5.2, so the locations x_i's do not cross one another. We want to produce distortions that do not produce drastic changes in the orientation of points (we do not want to generate configurations where points exchange places inelastically). In one dimension, this constraint can be simply implemented as an order constraint, but the implementation of such a constraint becomes a challenging problem in two or more dimensions.

5.4.2 One-Dimensional Elastic Distortion

We have a collection of pairs $\{(x_i, y_i)\}_{i=1}^n$ of heights y_i at locations x_i and we are interested in assigning a probability distribution to the x_i locations and observing the effect of this distribution on our inferred heights.

For one dimension, our elastic model correspond to a multivariate normal distribution with order restriction:

$$f(\{x_i\}_{i=2}^{n-1}|x_0 = 0, x_n = T) \propto \text{MVN}(\mathbf{x}; \boldsymbol{m}, \boldsymbol{\Lambda}^{-1}) I(0 = x_1 < x_2 < \cdots < x_{n-1} < x_n = T) \tag{5.3}$$

with mean vector $m_i = iT/n$ and with precision matrix $\boldsymbol{\Lambda}$ given by:

$$\Lambda_{ij} \quad = \quad \begin{cases} 2k & i = j \\ -k & i = j \pm 1 \\ 0 & |i - j| > 1 \end{cases}$$

When k is large, the variability of the x-coordinates is very small and we can approximate the elastic model in one dimensional by an unrestricted multivariate normal distribution. This result is particularly useful to evaluate the effect of the number of points n, which corresponds to the resolution at which we observe the

function $y(x)$. In Appendix B we show that every time we divide the springs into two pieces by including new points between two other points, the elasticity constant of the marginal distribution of the original points increases by a factor of 2. Interestingly, this result does not depend on the location where we place the new point in between the the two original points.

Simulation Example

We now return to the damped sine wave example, discussed in Chapter 2. Recall that the computer model for this one-dimensional example is of the form:

$$y(z) = 10\sin(4\pi z^{0.9})\exp(-1.5z)$$

Assume that observations of the function are available at 9 equally-spaced points on $[0, 1]$. We fix the endpoints ($x_1 = 0$ and $x_9 = 1$) and wiggle the other x_i's randomly using the model in (5.3). The values of y are fixed, but their locations corresponding locations in x vary from one realization to the other. Figure 5.7 gives an idea of how the x-coordinate variability for different values of the elasticity constant k.

$k = 100$ $k = 1000$ $k = 9000$

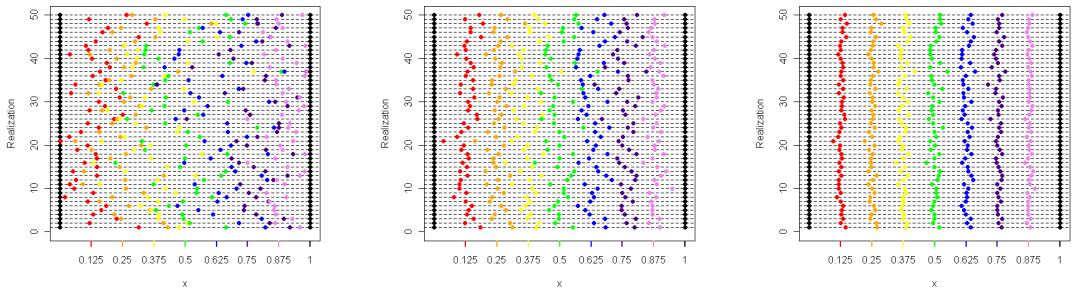


FIGURE 5.7: 50 realizations of the elastic model in one dimension. Colors represent points of the same position in the line.

For each generated sequence $\{x_i\}$, we interpolate through the pair of points $\{(x_i, y_i)\}$ using a posterior mean plug-in GASP. The result is a curve on the in-

terval $[0, 1]$ and this curve is treated as a realization of a random function. Some of these realizations are represented in Figure 5.8.

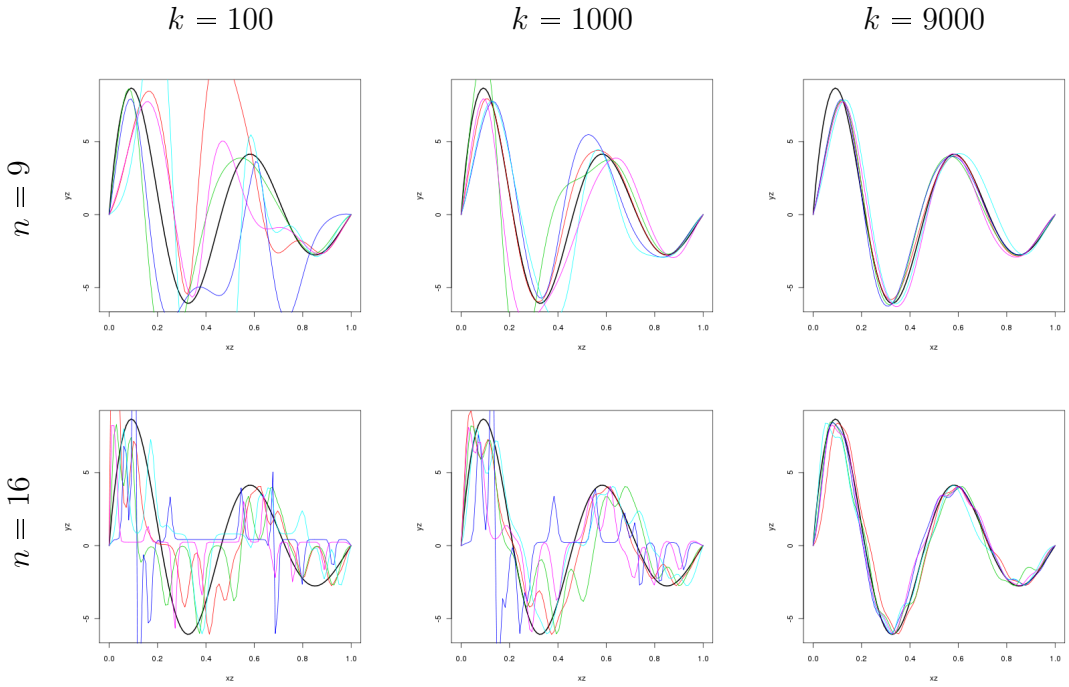


FIGURE 5.8: 5 random curves generated from the unidimensional elastic model. The curve in black is the original damped sine wave curve.

Figure 5.9 shows pointwise medians and 95% confidence bounds corresponding to 200,000 random curves generated by interpolating the pair of points $\{(x_i, y_i)\}$, where the x_i's are sampled from the elastic model. Note that the confidence bounds are wider when k is small. It is easier to obtain x_i's that are close to each other when k is small and, in that case, the interpolated function $y(x)$ has a big derivative. Since moderate changes in the x's become big changes in the y's, the MLE plug-in approach assigns very large variances to their predictions.

5.4.3 Two-dimensional Elastic Model

In order to evaluate the elastic model in two dimensions, we consider three regular grids on the region $[0, 8] \times [\sqrt{3}/2, 5\sqrt{3}]$ with different shapes: triangular, diamond

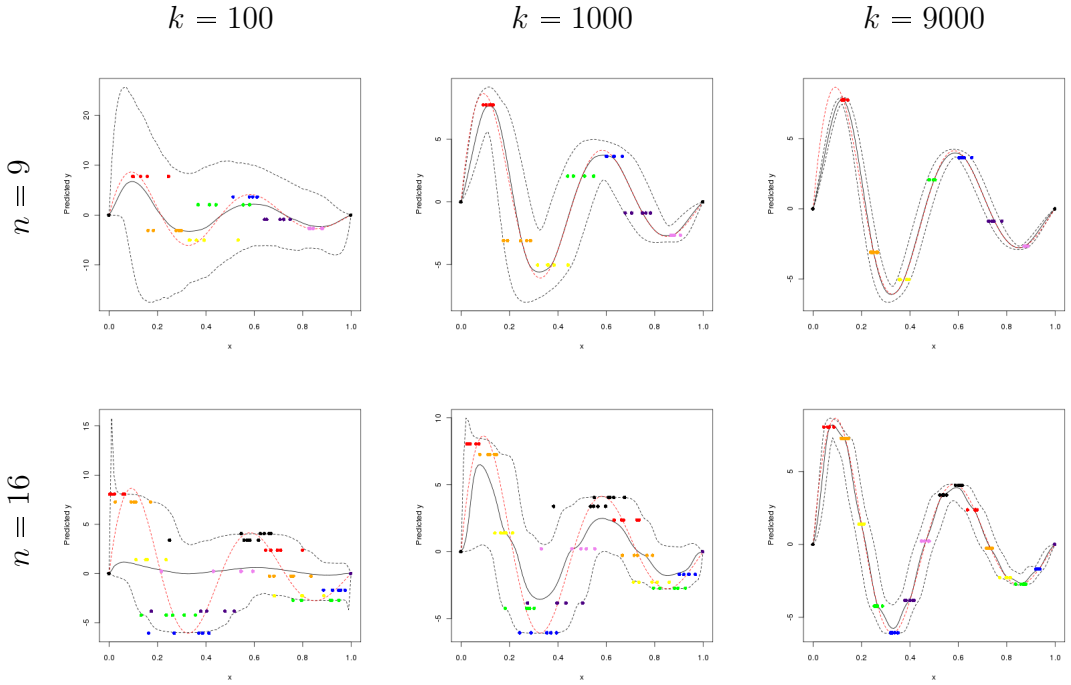
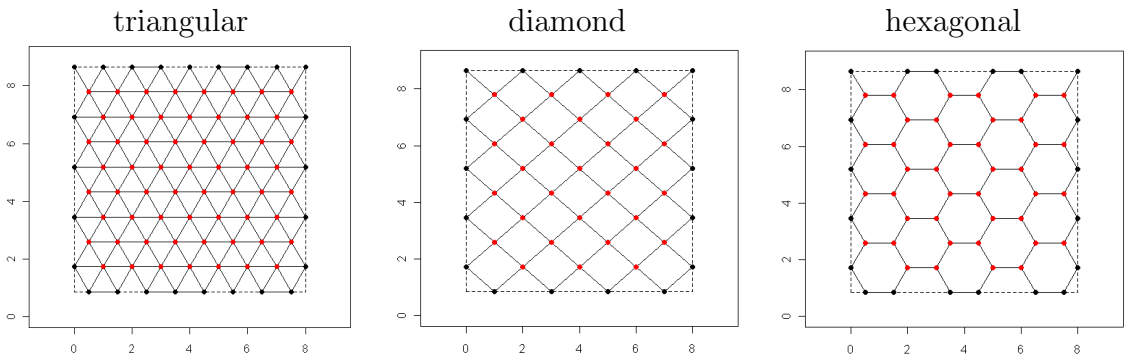


FIGURE 5.9: Pointwise median (black solid line) and 95% confidence bounds (black dashed lines) for random curves generated by assigning an elastic model to the x-coordinates while the y-coordinates are fixed. The red dashed line is the original damped sine wave curve. The circles represent some realizations from the elastic model.

and hexagonal. These three grids are shown in Figure 5.4.3.



For each one of the grids in Figure 5.4.3, we fix the black points on the boundary of the study region and wiggle the red points according to the elastic model in (5.2). The number of boundary black points is $m = 25$, 17 and 20, for the triangular, diamond and hexagonal grids, respectively. The respective numbers of red points

are $n = 60$, 28 and 40 and the distance between points in the original configuration is $\epsilon_{ij} = \sqrt{7}/2$ in the triangular and hexagonal grids and $\epsilon_{ij} = 1$ in the diamond grid, for all i and j.

We generate realizations from the bidimensional elastic model using an MCMC algorithm with Metropolis-Hastings step for each red location in the grid. The proposal distribution for each grid point is uniform on a small rectangle in the study region, attaining reasonable acceptance rates.

Figure 5.10 shows realizations of the two-dimensional elastic model for different grids and different values of the elastic constant k. We note that the hexagonal grid is the most susceptible to distortions from the elastic model because of the small number of spring connections for each grid point.

5.4.4 Propagation of uncertainty to geophysical features

The elastic model in two dimensions can be used to produce elastic distortions and assess uncertainty in the DEM. Elastic distortions usually produce irregular grids, and thus an interpolation method must be used to produce an array of values that can be compared to the original Digital Elevation Model, similarly to what we have done for the one-dimensional elastic model.

Unfortunately, there is an error associated with the interpolation process and the final result may depend on the technique being used. One possible way to reduce the errors associated with interpolation is to produce elastic distortions on the original points used to generate the DEM instead of the DEM itself. These points may be irregularly distributed in the study region, but the can be connected by springs according to some proximity criterion.

Once we have obtained a random array of elevation values, we may compute some other topographical features such as watershed lines and valley contours. For illustrative purpose, we show in Figure 5.11 the watersheds corresponding to the original

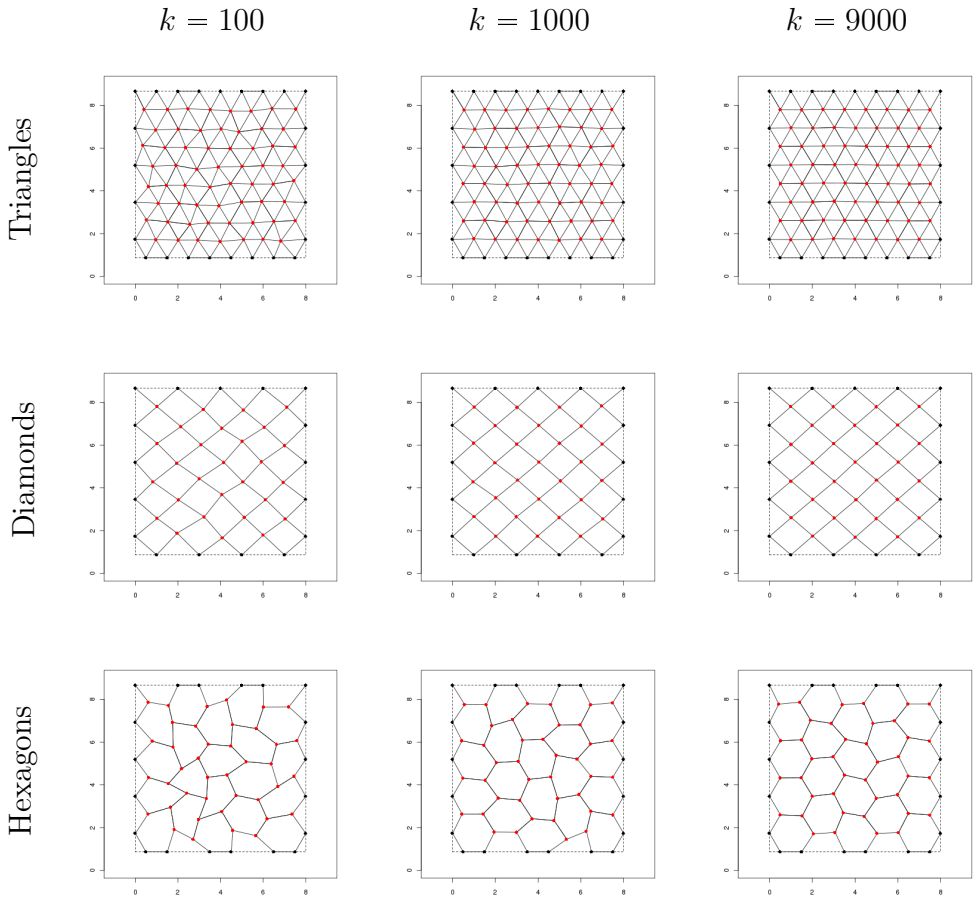


FIGURE 5.10: Realizations from Elastic Model on regular grids

DEM and some superimposed random watersheds generated from an additive-error model. This figure gives an notion of how errors in the DEM can affect the location of the watershed and possibly affect the path of simulated flows.

5.5 Discussion

The multivariate analysis of the TITAN2D output is certainly of great importance to the risk assessment of pyroclastic flows. However, the usual methods for multivariate spatial processes are not appropriate for our data.

We have proposed some initial ideas for modeling the flow-height surface as a

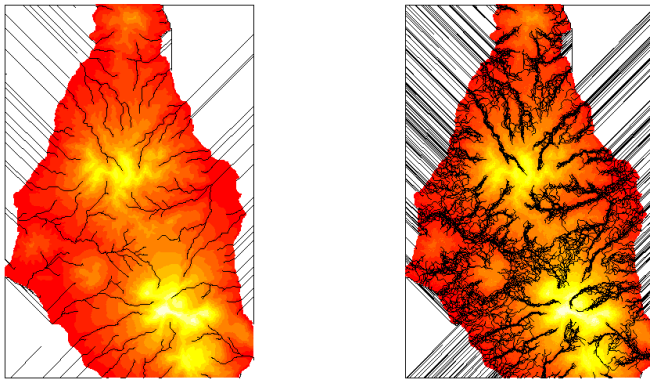


FIGURE 5.11: Comparison between watersheds corresponding to the original DEM and some superimposed random watersheds generated from an additive-error model.

function of the computer model inputs by using topographical information of the study region in the analysis. These ideas certainly need further developments, but they have been showing promising results in the analysis of pyroclastic flows.

We also emphasize the importance of the proposed method for modeling uncertainty in the DEM. The elastic model can be used in conjunction with additive-error models for better description of the errors associated with the construction of a Digital Elevation Model. Future steps of our analysis include the development of a method for estimation of the elastic constant k from replications of the DEM and a full analysis of the uncertainty propagation to other topographical features.

Appendix A

Objective priors in a simple case

Consider a simple one-dimensional experiment consisting of n runs, equally spaced at intervals of ϵ units. We want to fit a Gaussian process emulator to the computer model evaluations $\mathbf{y} = (y_1, \cdots, y_n)$, using a constant mean θ and exponential correlation function, which is equivalent to the correlation function in (2.2) with $\alpha = 1$.

The correlation matrix is a Toeplitz matrix of the form:

$$\Sigma_\beta = \begin{pmatrix} 1 & e^{-\epsilon\beta} & e^{-2\epsilon\beta} & \cdots & \cdots & e^{-(n-1)\epsilon\beta} \\ e^{-\epsilon\beta} & 1 & e^{-\epsilon\beta} & e^{-2\epsilon\beta} & & \vdots \\ e^{-2\epsilon\beta} & e^{-\epsilon\beta} & 1 & e^{-\epsilon\beta} & \ddots & \vdots \\ \vdots & e^{-2\epsilon\beta} & e^{-\epsilon\beta} & \ddots & \ddots & e^{-2\epsilon\beta} \\ \vdots & & \ddots & \ddots & 1 & e^{-\epsilon\beta} \\ e^{-(n-1)\epsilon\beta} & \cdots & \cdots & e^{-2\epsilon\beta} & e^{-\epsilon\beta} & 1 \end{pmatrix}$$

Let $\rho = e^{-\epsilon\beta}$. We can rewrite Σ_β as

$$\Sigma_\beta = \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \cdots & \rho^{(n-1)} \\ \rho & 1 & \rho & \rho^2 & & \vdots \\ \rho^2 & \rho & 1 & \rho & \ddots & \vdots \\ \vdots & \rho^2 & \rho & \ddots & \ddots & \rho^2 \\ \vdots & & \ddots & \ddots & 1 & \rho \\ \rho^{(n-1)} & \cdots & \cdots & \rho^2 & \rho & 1 \end{pmatrix}$$

Following Rybicki and Press (1995), we can show that the inverse of the correlation matrix is tridiagonal:

$$\Sigma_\beta^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho & 0 & 0 & \cdots & 0 \\ -\rho & 1+\rho^2 & -\rho & 0 & & \vdots \\ 0 & -\rho & 1+\rho^2 & -\rho & \ddots & \vdots \\ 0 & 0 & -\rho & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & 1+\rho^2 & -\rho \\ 0 & \cdots & \cdots & 0 & -\rho & 1 \end{pmatrix}$$

Since each element of Σ_β^{-1} can be obtained explicitly, we are able to perform matrix algebra calculations and obtain each piece of the integrated likelihood and objective priors presented in Section 2.3. We skip the matrix multiplications and present some fundamental results for the special case in study:

$$\begin{aligned} \mathrm{tr}\mathbf{U} &= \mathrm{tr}\left(\dot{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}\right) = \frac{2\epsilon(n-1)\rho^2}{1-\rho^2} \\ \mathrm{tr}\mathbf{U}^2 &= \mathrm{tr}\left[\left(\dot{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}\right)^2\right] = \frac{2\epsilon^2(n-1)\rho^2(1+\rho^2)}{(1-\rho^2)^2} \\ \mathrm{tr}\mathbf{W} &= \mathrm{tr}\mathbf{U} - \frac{\mathbf{1}^t\mathbf{\Sigma}^{-1}\dot{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^t\mathbf{\Sigma}^{-1}\mathbf{1}} \\ \mathrm{tr}\mathbf{W}^2 &= \mathrm{tr}\mathbf{U}^2 - 2\frac{\mathbf{1}^t\mathbf{\Sigma}^{-1}\dot{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}\dot{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^t\mathbf{\Sigma}^{-1}\mathbf{1}} + \left(\frac{\mathbf{1}^t\mathbf{\Sigma}^{-1}\dot{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^t\mathbf{\Sigma}^{-1}\mathbf{1}}\right)^2 \end{aligned}$$

$$
\begin{aligned}
\mathbf{1}^t\boldsymbol{\Sigma}^{-1}\mathbf{1} &= \frac{(1-\rho)(n-(n-2)\rho)}{1-\rho^2} \\
\mathbf{1}^t\boldsymbol{\Sigma}^{-1}\dot{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}\mathbf{1} &= \frac{-2\epsilon(m-1)\rho}{(1+\rho)^2} \\
\mathbf{1}^t\boldsymbol{\Sigma}^{-1}\dot{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}\dot{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}\mathbf{1} &= \frac{2\epsilon^2\rho^2(1-\rho)\left[(2n-3)-(2n-1)\rho+\rho^{(n-1)}+\rho^n\right]}{(1-\rho^2)^3} \\
|\boldsymbol{\Sigma}| &= \left(1-\rho^2\right)^{n-1}
\end{aligned}
$$

Using these fundamental results, we can construct the integrated likelihood function:

$$
\begin{aligned}
L^I(\beta|\mathbf{y}) \propto\ & \left[(1-\rho)(n-(n-2)\rho)\right]^{-1/2}\left\{\frac{1}{1-\rho^2}\right\}^{a-3/2}\left\{\sum_{i=1}^{n}y_i^2-2\rho\left(\sum_{i=1}^{n-1}y_iy_{i+1}\right)\right. \\
& +\rho^2\left(\sum_{i=2}^{n-1}y_i^2\right)-\frac{1-\rho}{n-(n-2)\rho}\left[\sum_{i=1}^{n}\sum_{j=1}^{n}y_iy_j-2\rho\left(\sum_{i=1}^{n}\sum_{j=1}^{n-1}y_iy_j\right)\right. \\
& \left.\left.+\rho^2\left(\sum_{i=2}^{n-1}\sum_{j=2}^{n-1}y_iy_j\right)\right]\right\}^{-((n-q)/2+a-1)}
\end{aligned}
$$

Similarly, the marginal objective priors of β can be expressed as

1. Independence Jeffreys prior:

$$
\pi^{J1}(\beta) \propto \frac{\rho}{1-\rho^2}\sqrt{n-(n-2)\rho^2}
$$

2. Jeffreys-rule prior:

$$
\pi^{J2}(\beta) \propto \frac{\rho}{(1-\rho^2)}\left[(n-(n-2)\rho^2)(n-(n-2)\rho)\right]^{1/2}\left[1+\rho\right]^{-1/2}
$$

3. Reference prior:

$$\begin{aligned}\pi^R(\beta) \;\propto\; & \frac{\rho}{1-\rho^2}\left\{\frac{2(n-1)^2(1-\rho)^2}{(n-(n-2)\rho)^2} - 2(n-1)\left[\rho + \frac{1-\rho}{n-(n-2)\rho}\right]^2\right. \\ & \left. +(n-1)(1+\rho^2) - \frac{2}{n-(n-2)\rho}\left[(2n-3)-(2n-1)\rho+\rho^{(n-1)}+\rho^n\right]\right\}^{1/2}\end{aligned}$$

As shown in Section 2.3, the marginal posterior density of β is equivalent to the product of the integrated likelihood and the corresponding marginal prior density:

1. Independence Jeffreys posterior:

$$\begin{aligned}\pi^{J1}(\beta|\mathbf{y}) \;\propto\; & \rho\left[(1-\rho^2)(1-\rho)(n-(n-2)\rho)\right]^{-1/2}\sqrt{n-(n-2)\rho^2} \\ & \times\left\{\sum_{i=1}^{n} y_i^2 - 2\rho\left(\sum_{i=1}^{n-1} y_i y_{i+1}\right) + \rho^2\left(\sum_{i=2}^{n-1} y_i^2\right)\right. \\ & -\frac{1-\rho}{n-(n-2)\rho}\left[\sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j - 2\rho\left(\sum_{i=1}^{n}\sum_{j=1}^{n-1} y_i y_j\right)\right. \\ & \left.\left.+\rho^2\left(\sum_{i=2}^{n-1}\sum_{j=2}^{n-1} y_i y_j\right)\right]\right\}^{-(n-1)/2}\end{aligned}$$

2. Jeffreys-rule posterior:

$$\begin{aligned}\pi^{J2} \;\propto\; & \rho\left[1-\rho^2\right]^{-3/2}\left[(n-(n-2)\rho^2)\right]^{1/2} \\ & \times\left\{\sum_{i=1}^{n} y_i^2 - 2\rho\left(\sum_{i=1}^{n-1} y_i y_{i+1}\right) + \rho^2\left(\sum_{i=2}^{n-1} y_i^2\right)\right. \\ & \left.-\frac{1-\rho}{n-(n-2)\rho}\left[\sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j - 2\rho\left(\sum_{i=1}^{n}\sum_{j=1}^{n-1} y_i y_j\right) + \rho^2\left(\sum_{i=2}^{n-1}\sum_{j=2}^{n-1} y_i y_j\right)\right]\right\}^{-n/2}\end{aligned}$$

3. Reference posterior:

$$\begin{aligned}
\pi^R(\beta|\mathbf{y}) \;\propto\; & \frac{\rho}{1-\rho^2}\left\{\frac{2(n-1)^2(1-\rho)^2}{(n-(n-2)\rho)^2} - 2(n-1)\left[\rho + \frac{1-\rho}{n-(n-2)\rho}\right]^2\right. \\
& \left. +(n-1)(1+\rho^2) - \frac{2}{n-(n-2)\rho}\left[(2n-3) - (2n-1)\rho + \rho^{(n-1)} + \rho^n\right]\right\}^{1/2} \\
& \times \left\{\sum_{i=1}^{n} y_i^2 - 2\rho\left(\sum_{i=1}^{n-1} y_i y_{i+1}\right) + \rho^2\left(\sum_{i=2}^{n-1} y_i^2\right)\right. \\
& -\frac{1-\rho}{n-(n-2)\rho}\left[\sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j - 2\rho\left(\sum_{i=1}^{n}\sum_{j=1}^{n-1} y_i y_j\right)\right. \\
& \left.\left. +\rho^2\left(\sum_{i=2}^{n-1}\sum_{j=2}^{n-1} y_i y_j\right)\right]\right\}^{-(n-1)/2}
\end{aligned}$$

It is presumably possible to obtain closed forms of the integrated likelihood and the objective priors in higher dimensions when data points are equally-spaced, $\alpha = 1$, the correlation function is separable and the design set forms a Cartesian product. These new assumptions allow us to represent the correlation matrix as a convenient Kronecker product:

$$\Sigma = \Sigma_1 \otimes \Sigma_2 \otimes \cdots \otimes \Sigma_p = \bigotimes_{k=1}^{p} \Sigma_k$$

We then can use the properties of the Kronecker product to perform the matrix operations, as detailed by Paulo (2005). However, the simple case in one dimension is only an illustrative example and we are not interested here in extending it further. In higher dimensions, we want to study more complex examples.

Appendix B

Relation Between Elasticity Constant and the Number of Points in the Unidimensional Elastic Model

We adopt the one-dimensional elastic model in 5.3 to represent the distribution of the three black points in Figure B.1.

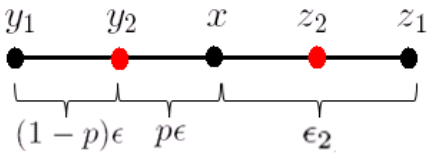


FIGURE B.1: Each spring is broken into two pieces at arbitrary points y_2 and z_2.

We later add two new points y_2 and z_2 ,breaking each spring in the figure into two pieces. Our objective is to integrate the model over the new red points for comparison with the elastic model using the old configuration.

The elastic model for the 5 points is given by (assume that k is large enough, so we do not need to worry about constraints):

$$p(\mathbf{x}) \quad \propto \quad \exp\left(-\frac{k}{2}\sum_{i=1}^{n}(x_i - x_{i-1} - \epsilon)^2\right)$$

where the two endpoints are fixed.

The distribution of x given y_1 and z_1 in the model with 3 points is:

$$p(x|y_1, z_1) \quad \propto \quad \exp\left(-\frac{k'(x - y_1 - \epsilon)^2}{2} - \frac{k'(z_1 - x - \epsilon_2)^2}{2}\right)$$

Note that the left and right sides of the segment around x are separable ($p(x|y_1, z_1) = g(x|y_1)g_2(x|z_1)$ - g and g_2 are functions, not densities). From now on, we are going to work with the left side of x only ($g(x|y_1)$), but the derivation of the right side follows from the same reasoning.

If we consider the distribution of x given y_1 and z_1 in the model with 5 points, we get:

$$g(x|y_1) \quad \propto \quad \int_{y_1}^{x} \exp\left(-\frac{k(y_2 - y_1 - (1-p)\epsilon)^2}{2} - \frac{k(x - y_2 - p\epsilon)^2}{2}\right) dy_2$$

We want to find the relation between the elastic constants k' (3 points) and k (5 points).

$$
\begin{aligned}
g(x|y_1) &\propto \int_{y_1}^{x} \exp\left(-\frac{k(y_2-y_1-(1-p)\epsilon)^2}{2} - \frac{k(x-y_2-p\epsilon)^2}{2}\right) dy_2 \\
&= \int_{y_1}^{x} \exp\left(-k\frac{y_2^2 - 2y_2(y_1+(1-p)\epsilon) + [y_1+(1-p)\epsilon]^2}{2}\right) \\
&\quad \times \exp\left(-k\frac{y_2^2 - 2y_2(x-p\epsilon) + [x-p\epsilon]^2}{2}\right) dy_2 \\
&= \int_{y_1}^{x} \exp\left(\frac{-2ky_2^2 + 2ky_2[y_1+x+(1-2p)\epsilon]}{2}\right) dy_2 \\
&\quad \times \exp\left(-\frac{k(y_1+(1-p)\epsilon)^2 + k(x-p\epsilon)^2}{2}\right) \\
f(x|y_1,z_1) &\propto \exp\left(k\frac{[y_1+x+(1-2p)\epsilon]^2}{4} - k\frac{(y_1+(1-p)\epsilon)^2 + (x-p\epsilon)^2}{2}\right)
\end{aligned}
$$

$$
\begin{aligned}
g(x|y_1) &\propto \exp\left(k\frac{[y_1+x+(1-2p)\epsilon]^2}{4} - k\frac{(y_1+(1-p)\epsilon)^2 + (x-p\epsilon)^2}{2}\right) \\
&\propto \exp\left(k\frac{x^2 + 2x[y_1+(1-2p)\epsilon]}{4} - k\frac{(x^2 - 2xp\epsilon)}{2}\right) \\
&= \exp\left(-k\frac{x^2}{4} + k\frac{2x[p\epsilon + (y_1+(1-2p)\epsilon)/2]}{2}\right) \\
&= \exp\left(-k\frac{x^2}{4} + k\frac{2x(y_1+\epsilon)/2}{2}\right) \\
&\propto \exp\left(-\frac{k/2[x-(y_1+\epsilon)]^2}{2}\right)
\end{aligned}
$$

Similarly, $g_2(x|z_1) \propto \exp\left(-\frac{k/2[x-(z_1-\epsilon)]^2}{2}\right)$.

So, if we want the conditional distributions under the two models to be the same, we should set $k = 2k'$.

Bibliography

[1] Ababou, R., Bagtzoglou, A. C., Wood, E.F. (1994) "On the Condition Number of Covariance Matrices in Kriging, Estimation, and Simulation of Random Fields". *Mathematical Geology*, 26, 99-133.

[2] Abt, M. (1999). "Estimating the prediction mean squared error in Gaussian stochastic processes with exponential correlation structure." *Scandinavian Journal of Statistics*, 26, 563-578.

[3] Assunção, R. and Lopes, D. (2007). "Testing Association Between Origin-Destination Spatial Locations". In *Advances in Geoinformatics*, eds. Junior, C. and Monteiro, A., 293-304. Berlin: Springer.

[4] Bayarri, M. J.; Berger, J. O.; Cafeo, J. A.; Garcia-Donato, G.; Liu, F.; Paulo, R.; Sacks, J.; and Walsh, D. (2007a). "Computer Model Validation With Functional Output," *The Annals of Statistics*, 35, 1874-1906.

[5] Bayarri, M. J.; Berger, J. O.; Paulo, R.; Sacks, J.; Cafeo, J. A.; Cavendish, J.; Lin, C. H. and Tu, J. (2007b). "A framework for validation of computer models," *Technometrics*, 49, 138-154.

[6] Bayarri, M. J. , Berger, J. O.,Calder, E. S., Dalbey, K., Lunagomez, S., Patra, A. K., Pitman, E. B., Spiller, E., Wolpert., R. L. (2009). "Using Statistical and Computer Models to Quantify Volcano Hazards," *Technometrics*, 51, 402-413.

[7] Banerjee, S.; Gelfand A. E.; Finley, A. O. and Sang, H. (2008). "Gaussian predictive process models for large spatial data sets," *Journal Of The Royal Statistical Society Series B*, 70, 825-848.

[8] Belisle, C. J. P. (1992), "Convergence theorems for a class of simulated annealing algorithms on Rd," *Journal of Applied Probability*, 29, 885-895.

[9] Berger, J. O. and Bernardo, J. M. (1992). "Reference priors in variance components problems." In *Proc. Indo-USA Workshop on Bayesian Analysis Statist.*

Econometrics (Edited by P. K. Goel and N. S. Iyenger), 177-194. Springer-Verlag, New York.

[10] Berger, J. O., De Oliveira, V. and Sansó, B. (2001). "Objective Bayesian analysis of spatially correlated data". *J. Amer. Statist. Assoc.*, 96, 1361-1374.

[10] Berger, J. O., Liseo, B., and Wolpert, R. (1999), "Integrated Likelihood Methods for Eliminating Nuisance Parameters (with discussion)," *Statistical Science*, 14, 1-28.

[11] Cafeo, J. and Cavendish, J. (2001), "A Framework For Verification And Validation Of Computer Models and Simulations," Internal general motors document, to be published, GM Research & Development Center.

[12] Calder, E. S., Cole, P. D., Dade, W. B., Druitt, T. H., Hoblitt, R. P., Huppert, H. E., Ritchie, L. J., Sparks, R. S. J., and Young, S. R. (1999), "Mobility of pyroclastic flows and surges at the Soufriere Hills Volcano, Montserrat," *Geophysical Research Letters*, 26, 537-540.

[13] Charbonnier, S.J. and Gertisser, R. (2009), "Numerical simulations of block and ash flows using the Titan2D flow model: examples from the 2006 eruption of Merapi Volcano, Java, Indonesia", *Bulletin of Volcanology*, 71, 953-959.

[14] Craig, P. S., Goldstein, M., Seheult, A. H., and Smith, J. A. (1997). "Pressure matching for hydrocarbon reservoirs: a case study in the use of bayes linear strategies for large computer ex- periments." In *Case Studies in Bayesian Statistics: Volume III.* (C. Gatsonis, J. S. Hodges, R. E. Kass, R. McCulloch, P. Rossi and N. D. Singpurwalla, eds.) 36-93. Springer, New York.

[15] Craig, P. S., Goldstein, M., Rougier, J. C., Seheult, A. H. (2001) "Bayesian forecasting for complex systems using computer simulators," *Journal of the American Statistical Association*, 96, 717-729.

[16] Csatho, B., Schenk, T., Kyle, P., Wilson, T., Krabill, W. B. (2008) "Airborne laser swath mapping of the summit of Erebus volcano, Antarctica: Applications to geological mapping of a volcano," *Journal of Volcanology and Geothermal Research*, 177, 531-548.

[17] Currin, C., Mitchell, T., Morris, M., and Ylvisaker, D. (1988). " A Bayesian approach to the design and analysis of computer experiments". ORNL Technical Report 6498, available from the National Technical Information Service, Springfield, VA, 22161.

[18] Currin, C., Mitchell, T., Morris, M., and Ylvisaker, D. (1991). "Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments". *Journal of the American Statistical Association*, 86, 953-963.

[19] Dalbey, K., A. K. Patra, E. B. Pitman, M. I. Bursik, and M. F. Sheridan (2008), "Input uncertainty propagation methods and hazard mapping of geophysical mass flows, *Journal of Geophysical Research*, 113, B05203, doi:10.1029/2006JB004471.

[20] Davies, T. R. H. (1982) "Spreading of rock avalanche debris by mechanical fluidization," *Rock Mechanics and Rock Engineering*, 15, 9-24.

[21] De Brujin, N. G. (1970), *Asymptotic Methods in Analysis*, Amsterdan: North-Holland.

[22] Diggle, P., Lophaven, S., (2006), "Bayesian geostatistical design," *Scandinavian Journal of Statistics*, 33, 5364.

[23] Easterling, R. G. (2001), "Measuring the Predictive Capability of Computational Models: Principles and Methods, Issues and Illustrations," Tech. Rep. SAND2001-0243, Sandia National Laboratories.

[24] Finley A. O., Sang H., Banerjee S., Gelfand A. E. (2009), "Improving the Performance of Predictive Process Modeling for Large Datasets." *Computational Statistics and Data Analysis*, 53, 2873-2884.

[25] Gelfand A. E., Schmidt, A., Banerjee S. and Sirmans C. F. (2004) "Nonstationary multivariate process modelling through spatially varying coregionalization (with discussion)." *Test*, 13, 1-50.

[26] Goldstein, M. and Rougier, J. C. (2003). "Calibrated bayesian forecasting using large computer simulators." Tech. rep., Statistics and Probability Group, University of Durham, http://www.maths.dur.ac.uk/stats/physpred/papers/CalibratedBayesian.ps.

[27] Goldstein, M. and Rougier, J. C. (2006). "Bayes linear calibrated prediction for complex systems," *Journal of the American Statistical Association*, 101, 1132-1143.

[28] GRASS Development Team, (2008). *Geographic Resources Analysis Support System (GRASS) Software.* Open Source Geospatial Foundation Project. http://grass.osgeo.org

[29] Gu, C. (1992). "Penalized likelihood regression: A Bayesian Analysis." *Statistica Sinica*, 2, 255-264.

[30] Hager, W. W. (1989). "Updating the inverse of a matrix". *SIAM Review*, 31, 221-239.

[31] Hale, A. J., Calder E. S., Loughlin, S., and Wadge, G. (2009) "Modelling the Lava Dome Extruded at Soufriere Hills Volcano, Montserrat, August 2005 - May 2006. Part I: Dome shape and internal structure," *Journal of Volcanology and Geothermal Research*, 187, 53-68.

[32] Handcock, M. S. and Stein, M. L. (1993). "A Bayesian analysis of kriging." *Technometrics*, 35, 403-410.

[33] Hebeler, F., Purves, R. S., (2009) "The influence of elevation uncertainty on derivation of topographic indices," *Geomorphology*, 111, 4-16.

[34] Heim, A.(1932). *Bergsturz und Menschenleben (Landslides and Human Lives).* Translation by N. Skermer. BiTech Publishers Ltd, Vancouver, Canada.

[35] Kass, R. E. and Steffey, D. (1989), "Approximate Bayesian Inference in Conditionally Independent Hierarchical Models (Parametric Empirical Bayes Models), " *Journal of the American Statistical Association*, 84, 717-726.

[36] Kennedy, M. C. and O'Hagan, A. (2001). "Bayesian calibration of computer models (with discussion)," *Journal of the Royal Statistical Society B*, 63, 425-464.

[37] Legros, F. (2002) "The mobility of long-runout landslides," *Engineering Geology*, 63, 301-331.

[38] Liu, F., Berger, J., Bayarri, M.J. (2009). "Modularization in Bayesian Analysis, with Emphasis on Analysis of Computer Models". *Bayesian Analysis*, 4, 119-150.

[39] Lopes, D. and Assunção, R. (*in press.*), "Visualizing Marked Spatial and Origin-destination Point Patterns With Dynamically Linked Windows". Journal of Computational and Graphical Statistics.

[40] Macías, J.L., Saucedo, R., Gavilanes, J.C., Varley, N., Velasco Garcia, S., Bursik, M.I., Vargas Gutierres, V. and Cortes, A., (2006). "Flujos piroclsticos asociados a la actividad explosiva del volcn de Colima y perspectivas futuras," *GEOS*, 25, 340-351.

[41] Morris, M. D., Mitchell, T. J., and Ylvisaker, D. (1993). "Bayesian design and analysis of computer experiments: Use of derivatives in surface prediction". *Technometrics*, 35, 243-255.

[42] Mosteller, F. and Wallace, D. L. (1964). *Inference and Disputed Authorship: The Federalists*, Reading, MA,: Addison-Wesley.

[43] Nagy, B., Loeppky, J. and Welch, W. J. (2007). "Fast Bayesian inference for Gaussian process models." Technical Report 230, Dept. Statistics, Univ. British Columbia.

[44] Oberkampf, W. and Trucano, T. (2000). "Validation Methodology in Computational Fluid Dynamics," Tech. Rep. 2000-2549, American Institute of Aeronautics and Astronautics.

[45] Paulo, R. (2005). "Default priors for Gaussian processes." *The Annals of Statistics*, 33, 556-582.

[46] Patra A. K., Bauer A. C., Nichita C. C., Pitman E. B., Sheridan M. F., Bursik M. I., Rupp B., Webber A., Stinton A. J., Namikawa L. M., Renschler C.S. (2005). "Parallel adaptive simulation of dry avalanches over natural terrain." *Journal of Volcanology and Geothermal Research*, 139, 122.

[47] Pilch, M., Trucano, T., Moya, J. L., Froehlich, G. Hodges, A., and Peercy, D. (2001), "Guidelines for Sandia ASCI Verification and Validation Plans - Content and Format: Version 2.0," Tech. Rep. SAND 2001-3101, Sandia National Laboratories.

[48] Pitman E. B., Patra A. K., Bauer A., Sheridan M. F., Bursik M. I. (2003), "Computing debris flows and landslides." *Physics of Fluids*, 15, 3638-3646.

[49] Quiñonero-Candela, J., Rasmussen, C. E., and Williams, C. K. I.(2007) "Approximation Methods for Gaussian Process Regression." In *Large Scale Kernel Machines, Neural Information Processing*, pages 203223. MIT Press, Cambridge, MA, USA.

[50] Roache, P. (1998), *Verification and Validation in Computational Science and Engineering*, Albuquerque: Hermosa Publishers.

[51] Royle, J. A.; Nychka, D. (1998) "An algorithm for the construction of spatial coverage designs with implementation in SPLUS". *Computers & Geosciences*, 24, 479-488.

[52] Rybicki, G. B. and Press, W. H. (1995) "Class of Fast Methods for Processing Irregularly Sampled or Otherwise Inhomogeneous One-Dimensional Data". *Physical Review Letters*, 74, 1060-1063.

[53] Sacks, J.; Welch, W. J.; Mitchell, T. J.; Wynn, H. P. (1989). "Design and Analysis of Computer Experiments," *Statistical Science*, 4, 409-423.

[54] Saltelli, A., Chan, K., and Scott, M., eds. (2000). *Sensitivity Analysis.* Wiley, Chichester.

[55] Santner, T., Williams, B., and Notz, W. (2003). *The Design and Analysis of Computer Experiments.* Springer-Verlag.

[56] Saucedo, R., Macías, J.L., Bursik, M.I., Gavilanes, J.C., Cortés, A., (2002). "Emplacement of pyroclastic flows during the 19981999 eruption of volcán de Colima, México," *Journal of Volcanology and Geothermal Research*, 117, 129-153.

[57] Saucedo, R., Macías, J.L., Bursik, M.I., (2004). "Pyroclastic flow deposits of the 1991 eruption of Volcán de Colima, Mexico." *Bulletin of Volcanology*, 66, 291-306.

[58] Sheridan, M. F., Stinton, A. J., Patra, A. K., Pitman, E. B., Bauer, A., Nichita, C. C. (2005), "Evaluating Titan2D mass-flow model using the 1963 Little Tahoma Peak avalanches, Mount Rainier, Washington." *Journal of Volcanology and Geothermal Research*, 139, 89102.

[59] Sparks, R.S., Young, S. R., Barclay, J., Calder, E. S., Cole, P. D., Darroux, B., Davies, M.A., Druitt, T. H., Harford, C., Herd, R. A., James, M. R., LeJeune, A. M., Loughlin, S., Norton, G., Skerrit, G., Stasiuk, M. V., Stevens, N.F., Toothill, J., Wadge, G. and Watts, R., (1998) "Magma production and growth of the lava dome of the Soufriere Hills Volcano, Montserrat, West Indies: November 1995 to November 1997," *Geophysical Research Letters*, 25, 3421-3684.

[60] Staron, L., and Lajeunesse, E. (2009), "Understanding how volume affects the mobility of dry debris flows," *Geophysical Research Letters*, 36, L12402, doi:10.1029/2009GL038229.

[61] Stefanescu, E. R., M. Bursik, Dalbey, K., Jones, M., Patra, A. and Pitman E.B. (2010). "DEM uncertainty and hazard analysis using a geophysical flow model," International Congress on Environmental Modelling and Software, Ottawa, Canada, July 2010.

[62] Tierney, L. and Kadane, J. (1986). "Accurate approximations for posterior moments and marginal densities." *Journal of the American Statistical Association*, 81, 82-86.

[63] Trucano, T., Pilch, M., and Oberkampf, W. O. (2002), "General Concepts for Experimental Validation of ASCII Code Applications," Tech. Rep. SAND 2002-0341, Sandia National Laboratories.

[64] Vallance, J.W., Bull, K.F., and Coombs, M.L., (2010), "Pyroclastic flows, lahars, and mixed avalanches generated during the 2006 eruption of Augustine Volcano," *The 2006 eruption of Augustine Volcano, Alaska: U.S. Geological Survey Professional Paper 1769*, (Power, J.A., Coombs, M.L., and Freymueller, J.T., eds.), 219267.

[65] Wackernagel, H., (2003). *Multivariate Geostatistics: An Introduction with Applications.* Springer.

[66] Welch, W. J., Buck, R. J., Sacks, J., Wynn, H. P., Mitchell, T. J., and Morris, M. D. (1992). "Screening, predicting, and computer experiments". *Technometrics*, 34, 15-25.

[67] Yang, R. and Chen, M. H. (1995), "Bayesian analysis for random coefficient regression models using noninformative priors", *Journal of Multivariate Analysis*, 55, 283-311.

Biography

Danilo Lourenço Lopes was born in Belo Horizonte, MG, Brazil on April 25, 1985. He received his BS in Actuarial Science from Federal University of Minas Gerais (UFMG, Brazil) in 2006. In 2007, Danilo started his graduate studies at Duke University and he received his MS in Statistics from Federal University of Minas Gerais (UFMG, Brazil) in 2008. Later, in 2009, we received his MS from Duke University. Danilo has a published book chapter (Assunção and Lopes, 2007) and a recently accepted paper in the JCGS (Lopes and Assunção, *in press*).