CIMAT Centro de Investigación en Matemáticas, A.C.

ADVANCES IN BAYESIAN UNCERTAINTY QUANTIFICATION

T E S I S

Que para obtener el grado de

Doctor en Ciencias

con Especialidad en

Probabilidad y Estadística

Presenta:

José Cricelio Montesinos López

Director de Tesis:

Dr. José Andrés Christen Gracia

Autorización de la versión final

Guanajuato, Gto., 11 de Enero de 2022

Acknowledgments

I want to thank my advisor, Dr. Andrés Christen, for believing in me and allowing me to be enriched with his valuable ideas, for his disposition, his patience, his motivation, and his support during the whole elaboration of this work.

I want to thank Dr. Maria L. Daza-Torres for their work in the paper; "*Error control in the numerical posterior distribution in the Bayesian UQ analysis of a semilinear evolution PDE. International Journal for Uncertainty Quantification*", which is the basis for Chapter 3. Also, for all the support given during the last three years of the doctorate.

I want to thank Drs. Antonio Capella and Josué Tago for their work in the paper; "*Uncertainty Quantification for Fault Slip Inversion (2021). Mathematical Geoscience*", which is the basis for Chapter 4.

I also thank my synods, Dr. Antonio Capella, Dr. Colin Fox, Dr. Emilien Joly, and Dr. Victor Cruz Atienza, for their time to review this work.

Thank you to my parents and my siblings for their unconditional support. I especially thank Osval and Abelardo for sharing their experiences with me and being aware of my progress in each doctorate stage.

Thanks to the Centro de Investigación en Matemáticas (CIMAT) for giving me a quality education and allow me to use their facilities and resources to complete this work. I am also grateful to the Consejo Nacional de Ciencia y Tecnología (CONACYT) for the four years of my doctoral studies' economic support through scholarship No. 333114.

Summary

Inverse problems (IPs) appear in many science and mathematics branches, mainly when quantities of interest are different from those we can measure; model parameter values must be estimated from the observed data. The Bayesian approach to IPs, recently known as Bayesian uncertainty quantification (UQ), is becoming ever more popular to approach these kinds of problems. The posterior distribution of the unknown parameters quantifies the uncertainty of these parameters' possible values consistent with the observed data. However, explicit analytic forms are usually not available for the posterior distributions, so sampling approaches such as the Markov chain Monte Carlo are required to characterize it. These methods involve repeated forward map (FM) solutions to evaluate the likelihood function. And, we usually do not have an analytical or computationally precise implementation of the FM and necessarily involves a numerical approximation, leading to a numerical/approximate posterior distribution. Thus, the numerical solution of the FM will introduce some numerical error in the posterior distribution.

In this thesis, a bound on the relative error (RE), in posterior expectations for some functional, is found. This bound is later used to justify a bound in the error of the numerical method used to approximate the FM for Bayesian UQ problems. Moreover, our contribution is a numerical method to compute the *after-the-fact* (i.e., *a posteriori*) error estimates of the numerical solution of a class of semilinear evolution partial differential equations and show the potential applicability of the result previously obtained for bounding the RE in posterior statistics for Bayesian UQ problems. Several numerical examples are presented to demonstrate the efficiency of the proposed algorithms.

We also present a Bayesian approach to infer a fault displacement from geodetic data in a slow slip event (SSE). Our physical model of the slip process reduces to a multiple linear regression subject to constraints. Assuming a Gaussian model for the geodetic data and considering a multivariate truncated normal (MTN) prior distribution for the unknown fault slip, the resulting posterior distribution is also MTN. Regarding the posterior, we propose an ad hoc algorithm based on Optimal Directional Gibbs that allows us to efficiently sample from the resulting high-dimensional posterior distribution without supercomputing resources. As a by-product of our approach, we are able to estimate moment magnitude for the 2006 Guerrero SSE with UQ.

List of Contributions

- A bound on the RE, in posterior expectations (PEs) for some functional, is found. This is later used to justify a bound in the error of the numerical method used to approximate the FM for Bayesian UQ problems (see Sec. 2.3).
- We implemented a method of lines for the solution of semilinear evolution PDEs with numerical after-the-fact (*a posteriori*) error estimates. This error estimation allows me to apply the results obtained in Sec. 2.3 for bounding the expected RE (ERE) in the respective numerical PE for Bayesian UQ problems (see Sec. 3). This represents a novel goal-oriented mesh design, considering the inference problem at hand (the posterior distribution).
- A Bayesian approach to infer a fault slip from geodetic data of a slow slip event is presented. I developed a framework in modeling the data as well as the prior, a numerical sampling method, and an improved analysis through posterior sample-based inference. I apply the methodology to a real data set, for the 2006 Guerrero Slow Slip Event (see Sec. 4.4).

List of Acronyms

ABIC Akaike's Bayesian Information Criterion

AGE absolute global error

AIC Akaike Information Criterion

BF Bayes factors

CFL Courant-Friedrichs-Lewy

CV coefficient of variation

DIC Deviance Information Criterion

EABF expected absolute Bayes factor

ERE expected relative error

FD finite difference

FM forward map

GGap Guerrero GAP

GP Gaussian process

IAT integrated autocorrelation time

IP inverse problem

MAP maximum a posteriori

MCMC Markov chain Monte Carlo

MLE Maximum Likelihood Estimation

MTN multivariate truncated normal

ODE ordinary differential equation

PDE partial differential equation

PDF probability density function

PE posterior expectation

QoI Quantity of Interest

RE relative error

RK Runge-Kutta

RKCK Runge-Kutta Cash–Karp

SSE slow slip event

TE truncation error

UQ uncertainty quantification

Contents

CHAPTER 1

Introduction

A wide range of applications are concerned with the solution of an IP (Kaipio and Fox 2011; Cai et al. 2011; Chadan et al. 1997; Holder 2004; Burggraf 1964; Snieder and Trampert 1999; Daza et al. 2017): given some observations of an output, $\mathbf{y} = (y_1, \ldots, y_n)$, determine the corresponding inputs θ such that $y_i = \mathcal{F}(\theta) +$ error.

We refer to the evaluation of $\mathcal{F}$ as solving the forward problem and, consequently, $\mathcal{F}$ is called the FM. In general, the FM is a complex non-linear map, with input parameters θ, defined by an initial/boundary value problem for a system of ordinary differential equations (ODEs) or partial differential equations (PDEs). Then, to evaluate $\mathcal{F}(\theta)$, we must solve an initial/boundary value problem for a system of (O, P)DEs.

IPs appear in many branches of science and mathematics, mainly in situations where quantities of interest are different from those we can measure. In IPs, model parameter values must be estimated from the observed data. Historically, IPs have been studied as deterministic problems in analysis, and only relatively recently have been correctly treated as problems in statistical inference. In deterministic inversion from noisy data, the IPs there may have no solution, or the solution may not be unique, or it may depend sensitively on measurements y_i (Kaipio and Somersalo 2006). A way to approach these difficulties is to formulate the IP in the Bayesian framework. Stuart (2010) studied conditions for the well-posedness of one particular Bayesian formulation of IPs. In this framework, a noise model is assumed for the observations, for example, an additive Gaussian noise model

$$y_i = \mathcal{F}(\theta) + \varepsilon_i,$$

where the errors, ε_i, follow a normal distribution with mean zero and variance σ^2. This observational model generates a probability density of $\mathbf{y}$ given the parameters $\Phi = (\theta, \sigma)$, namely

$P_{\mathbf{Y}|\Phi}(\mathbf{y}|\theta,\sigma)$. For fixed data $\mathbf{y}$, and as a function of Φ, we call this the likelihood function. Based on the available information, a prior model $P_{\Phi}(\cdot)$ is stated for Φ, and a posterior distribution $P_{\Phi|\mathbf{Y}}(\theta,\sigma|\mathbf{y})$ is obtained through Bayes theorem.

In a frequentist (classical) statistical paradigm, we often maximize the likelihood to obtain a single estimate for the parameter of interest. Uncertainty is defined by the sampling distribution based on the idea of infinite repeated sampling. In contrast, the goal of Bayesian inversion is not only to obtain a single estimate for the unknowns but to quantify their uncertainty consistently with the observed data. Therefore, we describe the unknowns by probability distributions. Before any observation is available, we have much uncertainty in the unknown. After making the measurements, the uncertainty is reduced, and the task is to quantify it and provide probabilistic answers to questions of interest (Kaipio and Fox 2011). Previous information regarding the physics of the problem, which is not specific enough to be incorporated into the direct problem, may be incorporated into the prior probability distribution. The posterior distribution of the unknown parameter quantifies the uncertainty of the possible values of these parameters consistent with the observed data.

However, explicit analytic forms are usually not available for the posterior distributions, so sampling approaches such as the Markov chain Monte Carlo (MCMC) are required to characterize it. These methods involve repeated FM solutions used to evaluate the likelihood function. Furthermore, we usually do not have an analytical or computationally precise and straightforward implementation of the FM. It necessarily requires a numerical approximation, $\mathcal{F}^{\alpha(n)}$, where $\alpha(n)$ represents a discretization used to approximate the FM, leading to a numerical/approximate posterior distribution. Thus, the numerical solution of the FM will introduce some numerical error in the posterior distribution, and in any PE. In general, numerical errors in the FM can be controlled and reduced to an arbitrarily low level through finer discretizations. What numerical error must be tolerated in the FM to obtain a correct and acceptable numerical posterior distribution or some desired numerical error in a posterior statistic?

Several approaches start by building computationally-cheap approximations of the FM and using these approximations as surrogates in the sampling procedure (Marzouk et al. 2007; Galbally et al. 2010; Lieberman et al. 2010; Rasmussen et al. 2003). Although such approaches can be quite effective at reducing computation cost, there has been little analysis of posterior inference approximation. Recently, adaptive multi-fidelity techniques have been developed to control the numerical posterior error for Bayesian UQ (Cui et al. 2015; Peherstorfer et al. 2018; Yan and Zhou 2019b;a; Li and Marzouk 2014). In Li and Marzouk (2014), the authors proposed an adaptive multi-fidelity polynomial chaos MCMC algorithm to find a distribution that is "close" to the posterior in the sense of Kullback–Leibler divergence. Similar approaches were proposed in Yan and Zhou (2019a) using an adaptive multi-fidelity polynomial chaos based ensemble Kalman inversion technique.

Close in spirit to the works mentioned, in Capistrán et al. (2016), the authors proposes the use of Bayes factors (BF; the odds in favor) between the true posterior distribution and the numerical approximated posterior distribution. In an ODE framework, they show that the BF converge to 1, in the same order as the numerical solver used.

Later, this idea was generalized in Capistrán et al. (2021) to consider also PDEs and the use of the expected absolute Bayes factor (EABF), before observing data. This results in more practical and workable guidelines in a more realistic multidimensional setting. The main result in Capistrán et al. (2021) is that convergence in a consistent discretization of the FM leads to convergence of the EABF and also shows that the EABF can be bounded by bounding the error in the numerical FM (from the true FM).

In Chapter 2, we present the main result of Capistrán et al. (2021) but justify the method not in term of BF but considering the PE in posterior statistics. The proposal is to bound the error, or the RE, in PEs for some functional. Indeed, as in Capistrán et al. (2021), a reliable estimate of the error for the numerical solver used is the central point in the calculation of this bound.

Current efforts to estimate the discretization error focus on *after-the-fact* methods (i.e., *a posteriori* methods, we prefer to call them *after-the-fact* to avoid the obvious confusion with the Bayesian jargon). These methods provide an error estimate only after the numerical solution has been computed. They use the computed solution of the discrete equations, possibly with additional information supplied by the equations, to estimate the RE to the exact solution of the mathematical model (Roy 2010). Most of the previous works are based on higher error bounds with asymptotic convergence when the mesh size tends to zero (Babuška and Rheinboldt 1978; Berzins 1988; De S. R. Gago et al. 1983; Ainsworth and Oden 2011; Rognes and Logg 2013). Unfortunately, these estimates imply "constants of stability" generally unknown and difficult to calculate. The resulting error estimation techniques, in practice, do not provide mathematically proven bounds that, in general, can be computed efficiently (Grätsch and Bathe 2005).

In Chapter 3, we present a numerical method for computing the *after-the-fact* error estimates for semilinear evolution PDEs, and show the potential applicability of a result in Chapter 2 for bounding the ERE in the respective numerical PE for Bayesian. We obtain the numerical solution with the method of lines; discretizing first in the space with the finite difference (FD) method and solving the resulting system in time with the Runge-Kutta Cash–Karp (RKCK) method (Cash and Karp 1990). Our error estimation uses the estimate available for the RKCK method, and the spatial discretization error is treated as the solution of an initial value problem, approximated by the RKCK method.

A common IP in Geophysics is to infer quantities about Earth's interior in terms of surface measurements. In specific, to estimate the magnitude and distribution of a slip along a fault plane from geodetic data. Fault slips may consist of complex and heterogeneous source processes, while limited geodetic data typically leads to an ill-posed IP. Bayesian methods provide a natural

method to explore the posterior distribution of IPs and provide formal UQ.

In Chapter 4, we present a novel Bayesian approach to infer a fault displacement from geodetic data in a SSE. Our physical model of the slip process reduces to a multiple linear regression subject to constraints. A Gaussian model is assumed for the geodetic data and we propose an MTN prior slip distribution, with a detailed correlation structure, to impose natural coherence in the fault movement. The resulting posterior distribution is also MTN. Regarding the posterior, we propose an ad hoc algorithm based on Optimal Directional Gibbs that allows us to efficiently sample from the resulting high-dimensional posterior distribution without resorting to supercomputing resources. We apply our method to quantify the uncertainty in the IP of seismic slip along the subduction interface in the 2006 Guerrero, Mexico, SSE. Moreover, with our method, we are able to provide the posterior distribution of the moment magnitude for this event in a straightforward manner.

The thesis is organized as follows. Chapter 2 describes the main result of Capistrán et al. (2021) and, using a similar mathematical approach, a bound on the RE, in PEs for some functional is proposed. It is also explained in detail how this bound can be used to find a discretization in the numerical method used to approximate the FM. We also present two examples considering Bayesian UQ problems for 1D and 2D heat equations, respectively. In Chapter 3, we propose a numerical method for computing the *after-the-fact* error estimates for semilinear evolution PDEs, and show the potential applicability of a result in Chapter 2. Chapter 4 presents a Bayesian approach to infer a fault displacement from geodetic data of a SSE. The methodology is applied to a real data set, for the 2006 Guerrero SSE. Finally, general conclusions and a discussion of future work are presented in Chapter 5.

CHAPTER 2

Bounds for numeric error in posterior statistics with two examples in an IP of the heat equation

This Chapter describes how to bound the numeric error in posterior statistics. Namely, a bound on the ERE, in PE for some functional, is found directly, in Theorem 2 (the proof is based on how convergence rates are exported using a similar mathematical approach to a combination of results from Christen and Pérez-Garmendia (2021) and Capistrán et al. (2021)). It is also detailed how this bound can be used to find a discretization level in the numerical method used to approximate the FM. Two examples are presented considering Bayesian UQ problems for 1D and 2D heat equations, respectively, to illustrate the proposed strategy performance.

2.1 Setting

Let $\Theta \subset \mathbb{R}^q$ be the parameter space (i.e., we only consider finite-dimensional space parameters). Let $\mathcal{F} : \Theta \to V$ be the FM (typically $\mathcal{F}(\theta)$, for all $\theta \in \Theta$, is the solution of a system of PDE's), and let $\mathcal{H} : V \to A \subseteq \mathbb{R}^m$ be the observation operator (e.g., $\mathcal{H}(\mathcal{F}(\theta))$ is one particular state variable, for which we have observations). The composition $\mathcal{H} \circ \mathcal{F}$ defines a mapping from the parameter space Θ to the data sample space in $\mathbb{R}^m$.

Assume that we observe a process $\mathbf{y} = (y_1, \ldots, y_m)$ at locations $z_1, \ldots, z_m \in D$. This is a general setting, to include PDEs and other IPs, in which the domain D may include, for example, space and time: $z_i = (x_i, t_i)$. That is, y_i is an observation at coordinates x_i and at time t_i. Also,

assume that $f(\mathbf{y}|\theta)$ is a density for data $\mathbf{y}$:

$$f(\mathbf{y}|\theta) := f_o(\mathbf{y}|\mathcal{H}(\mathcal{F}(\theta))); \quad \theta \in \Theta,$$

where $f_o(y|\eta(\theta))$ is a density function that interacts with θ only through $\eta(\theta) \in \mathbb{R}^m$.

Let $\mathcal{F}^{\alpha(n)}$ be a discretized version of the FM, $\mathcal{F}$. Here $\alpha(n)$ represents a discretization used to approximate the FM and as n increases the discretization becomes finer and the approximation error decreases. This is the actual numerical version of the FM defined on our computers. Let $f^n(\mathbf{y}|\theta) := f_o(\mathbf{y}|\mathcal{H}(\mathcal{F}^{\alpha(n)}(\theta)))$ be the resulting discretized numerical likelihood. In particular, note that θ is always finite dimensional and the same for the theoretical and numerical models, specifically, the prior π is the same in both cases (indeed, in some other contexts, θ may depend of the FM discretization, for example, but we do not consider this possibility).

In the rest of the Chapter, we take the following assumption.

Assumption 1. *Assume that, for all $\mathbf{y} \in \mathbb{R}^m$, the observation model $f_o(\mathbf{y}|\eta)$ is uniformly Lipschitz continuous on η, and for $\mathbf{y} \in \mathbb{R}^m$, $f_o(\mathbf{y}|\eta)$ is bounded. Moreover, the FMs $\mathcal{H} \circ \mathcal{F}$ and $\mathcal{H} \circ \mathcal{F}^{\alpha(n)}$ are continuous.*

Assumption 2. *Assume a global error control of the numeric FM as*

$$||\mathcal{H}(\mathcal{F}(\theta)) - \mathcal{H}(\mathcal{F}^{\alpha(n)}(\theta))||_\infty < K^{\alpha(n)}. \tag{2.1}$$

Note that this is a global bound, valid for all $\theta \in \Theta$, and includes already the observational operator. That is, it is a global bound, but is only a statement at the locations $\mathcal{H}_i$'s where each y_i is observed.

Finally, assuming a proper prior probability π for the parameter θ, we will use the notation $\mathbb{Q}_y[\pi]$ for the theoretical posterior to make the dependency on the prior explicit. Furthermore, when substituting the likelihood with the numerical approximation $f^n(\mathbf{y}|\theta)$, we denote the corresponding numeric posterior as $\mathbb{Q}_y^n[\pi]$.

2.2 Consistency and other results

If $\mathcal{H} \circ \mathcal{F}$ and $\mathcal{H} \circ \mathcal{F}^{\alpha(n)}$ are continuous then $\theta \mapsto f(\mathbf{y}|\theta)$ and $\theta \mapsto f^n(\mathbf{y}|\theta)$ are continuous. Therefore, the posterior measures $\mathbb{Q}_y[\pi]$ and $\mathbb{Q}_y^n[\pi]$ are well defined (see, for example, Christen and Pérez-Garmendia 2021, and references therein).

Since we have ensured the existence of the theoretical and numerical versions of the posterior measure, let's look at the following consistency result. If the FM discretization is consistent, that is $\mathcal{F}^{\alpha(n)} \to \mathcal{F}$, and under assumptions 1 and 2, the authors in Christen and Pérez-Garmendia

(2021) proved that the numeric posterior converge, i.e.,

$$\left\|\mathbb{Q}_y[\pi] - \mathbb{Q}_y^n[\pi]\right\|_{\text{TV}} \to 0$$

as n increases, in the Total Variation norm $\|\cdot\|_{\text{TV}}$, and that the normalization constants converge, $|Z(\mathbf{y}) - Z^n(\mathbf{y})| \to 0$, where

$$Z(\mathbf{y}) = \int f(\mathbf{y}|\theta)\pi(d\theta),$$

and $Z^n(\mathbf{y})$ is the corresponding numeric normalization constant when substituting the likelihood with the numerical approximation $f^n(\mathbf{y}|\theta)$.

To find reasonable guidelines, to choose a discretization level, Capistrán et al. (2016) compare the numeric posterior $\mathbb{Q}_y^n[\pi]$ with the theoretical posterior $\mathbb{Q}_y[\pi]$ using Bayesian model selection, namely BF. Assuming an equal prior probability π for both models, the BF is the ratio of the normalization constants $\frac{Z^n(\mathbf{y})}{Z(\mathbf{y})}$ for fixed $\mathbf{y}$.

Later, Capistrán et al. (2021) try to control the BF between the discretized model and the theoretical model, through the use of the Absolute BF. To do that, they bound the expected Absolute BF, namely EABF $= \mathbb{E}_{\mathbf{Y}}\left[\frac{1}{2}\left|\frac{Z^n(\mathbf{y})}{Z(\mathbf{y})} - 1\right|\right]$. In Theorem 1, we state the main result of Capistrán et al. (2021). Here we reproduce this theorem, and its proof in detail, since it will be instrumental to the prove of (our main) Theorem 2. In addition, two small errors in the proof of Capistrán et al. (2021) are corrected, mentioned below.

Theorem 1. *(Capistrán et al. (2021)) With assumptions 1–2, and assuming independent data,* $\mathbf{y}$*, arising from a location-scale family, with scale parameter* σ^2 *and location parameter* $\eta = \mathcal{H}(\mathcal{F}(\theta)) = (\mathcal{H}_1(\mathcal{F}(\theta)), \ldots, \mathcal{H}_m(\mathcal{F}(\theta)))^T$*, namely*

$$f_o(\mathbf{y}|\eta) = \prod_{i=1}^{m} \sigma^{-1}\rho\left(\frac{y_i - \eta_i}{\sigma}\right), \tag{2.2}$$

with ρ *a bounded* C^1 *symmetric Lebesgue density in* $\mathbb{R}$*,* $\rho'(x) < 0$ *for all* $x \in (0, \infty)$*, and* $\int_{-\infty}^{\infty} x^2\rho(x)dx = 1$*, then*

$$\textit{EABF} = \int \frac{1}{2}\left|\frac{Z^n(\mathbf{y})}{Z(\mathbf{y})} - 1\right| Z(\mathbf{y})\lambda(d\mathbf{y}) < \rho(0)\frac{K^{\alpha(n)}}{\sigma}m. \tag{2.3}$$

Note that the model in (2.2) can be written as

$$y_i = \mathcal{H}_i(\mathcal{F}(\theta)) + \sigma\varepsilon_i, \quad i = 1, \ldots, m,$$

where each ε_i has zero mean and unit variance, and its probability distribution function belongs to the location-scale family.

Proof. Notice that

$$\int \frac{1}{2}\left|\frac{Z^n(\mathbf{y})}{Z(\mathbf{y})}-1\right| Z(\mathbf{y})\lambda(d\mathbf{y}) \;=\; \int \frac{1}{2}\left|Z^n(\mathbf{y})-Z(\mathbf{y})\right|\lambda(d\mathbf{y}).$$

To bound the last integral, observe that

$$\begin{aligned}\left|Z^n(\mathbf{y})-Z(\mathbf{y})\right| &= \left|\int f^n(\mathbf{y}|\theta)\pi(d\theta)-\int f(\mathbf{y}|\theta)\pi(d\theta)\right| = \left|\int \left(f^n(\mathbf{y}|\theta)-f(\mathbf{y}|\theta)\right)\pi(d\theta)\right| \\ &= \left|\int f(\mathbf{y}|\theta)\left(R_n\left(\theta\right)-1\right)\pi(d\theta)\right| \leq \int f(\mathbf{y}|\theta)\left|R_n\left(\theta\right)-1\right|\pi(d\theta),\end{aligned}$$

with $R_n\left(\theta\right)=\frac{f^n(\mathbf{y}|\theta)}{f(\mathbf{y}|\theta)}$. For η close enough to η_1, for assumption 1, the likelihood ratio $\frac{f_o(\mathbf{y}|\eta)}{f_o(\mathbf{y}|\eta_1)}$ is near 1, and

$$\left|\frac{f_o\left(\mathbf{y}|\eta\right)}{f_o\left(\mathbf{y}|\eta_1\right)}-1\right| \cong \left|\log\left(\frac{f_o\left(\mathbf{y}|\eta\right)}{f_o\left(\mathbf{y}|\eta_1\right)}\right)\right| = \left|\phi_{\mathbf{y}}(\eta)-\phi_{\mathbf{y}}(\eta_1)\right| = \left|\phi_{\mathbf{y}}(\eta_1)-\phi_{\mathbf{y}}(\eta)\right|,$$

where $\phi_{\mathbf{y}}(\eta) = -\log f_o(\mathbf{y}|\eta)$. With the first-order Taylor approximation $\phi_{\mathbf{y}}(\eta) = \phi_{\mathbf{y}}(\eta_1) + \nabla\phi_{\mathbf{y}}(\eta_1)\cdot(\eta-\eta_1)+E$, we have[1]

$$\left|R_n\left(\theta\right)-1\right| = \left|\phi_{\mathbf{y}}(\eta_1)-\phi_{\mathbf{y}}(\eta)\right| = \left|\nabla\phi_y\left(\mathcal{H}(\mathcal{F}(\theta))\right)\cdot\left(\mathcal{H}(\mathcal{F}^{\alpha(n)}(\theta))-\mathcal{H}(\mathcal{F}(\theta))\right)+E\right|.$$

Ignoring the higher order terms in the residual E, we have

$$\left|R_n\left(\theta\right)-1\right| = \left|\sum_{j=1}^{m}\frac{\partial}{\partial\eta_j}\phi_{\mathbf{y}}\left(\mathcal{H}(\mathcal{F}(\theta))\right)\left(\mathcal{H}_j(\mathcal{F}^{\alpha(n)}(\theta))-\mathcal{H}_j(\mathcal{F}(\theta))\right)\right|$$

and using the global error bound in assumption 2, we have

$$f(\mathbf{y}|\theta)\left|R_n\left(\theta\right)-1\right| < f(\mathbf{y}|\theta)K^{\alpha(n)}\sum_{j=1}^{m}\left|\frac{\partial}{\partial\eta_j}\phi_{\mathbf{y}}\left(\mathcal{H}(\mathcal{F}(\theta))\right)\right| \tag{2.4}$$

since $\left|\sum a_j(\theta)b_j(\theta)\right| < c\sum\left|a_j(\theta)\right|$ with $\left|b_j(\theta)\right| < c$ for all θ. Then,

$$\begin{aligned}\int\frac{1}{2}\left|Z^n(\mathbf{y})-Z(\mathbf{y})\right|\lambda(d\mathbf{y}) &< \int\frac{1}{2}\int f(\mathbf{y}|\theta)\left|R_n\left(\theta\right)-1\right|\pi(d\theta)\lambda(d\mathbf{y}) \\ &< \frac{K^{\alpha(n)}}{2}\int\int f_o(\mathbf{y}|\eta)\sum_{j=1}^{m}\left|\frac{\partial}{\partial\eta_j}\phi_{\mathbf{y}}\left(\mathcal{H}(\mathcal{F}(\theta))\right)\right|\pi(d\theta)\lambda(d\mathbf{y})\end{aligned}$$

Note how the global error bound is needed for all j as in assumption 2, nonetheless, we

[1] In the proof given in Capistrán et al. (2021), the authors had $\nabla\phi_y\left(\mathcal{H}(\mathcal{F}^{\alpha(n)}(\theta))\right)$ instead of $\nabla\phi_y\left(\mathcal{H}(\mathcal{F}(\theta))\right)$.

could restrict this bound for $\theta \in supp\{\pi\} \subset \Theta$ since we are integrating w.r.t $\pi(\cdot)$.

We calculate the remaining double integral by changing the order of integration to calculate $\int M(\mathcal{H}(\mathcal{F}(\theta)))\pi(d\theta)$, where

$$M(\eta) = \sum_{i=1}^{m} \int \left| \frac{\partial}{\partial \eta_i} \phi_{\mathbf{y}}(\eta) \right| f_o(\mathbf{y}|\eta)\lambda(d\mathbf{y}). \tag{2.5}$$

From the above equation, assuming data arising from model (2.2), note that

$$\begin{aligned}
\int \left| \frac{\partial}{\partial \eta_i} \phi_{\mathbf{y}}(\eta) \right| f_o(\mathbf{y}|\eta)\lambda(d\mathbf{y}) &= \int_{-\infty}^{\infty} \left| \frac{\partial}{\partial \eta_i} \phi_{\mathbf{y}}(\eta) \right| f_o(y_i|\eta)\lambda(dy_i) \underbrace{\int_{-\infty}^{\infty} f_o(\mathbf{y}_{-i}|\eta)\lambda(d\mathbf{y}_{-i})}_{1} \\
&= \int_{-\infty}^{\infty} \left| \sigma^{-1} \frac{\rho'\left(\frac{y_i-\eta_i}{\sigma}\right)}{\rho\left(\frac{y_i-\eta_i}{\sigma}\right)} \right| \sigma^{-1}\rho\left(\frac{y_i-\eta_i}{\sigma}\right) dy_i \\
&= \sigma^{-1} \int_{-\infty}^{\infty} \left|\rho'(x)\right| dx
\end{aligned}$$

Now[2], if $\rho'(x) < 0$ for all $x \in (0, \infty)$, then the integral on the right side is in fact equal to $\int_{-\infty}^{\infty} \left|\rho'(x)\right| dx = -2\int_{0}^{\infty} \rho'(x)\, dx = 2\rho(0)$. Thus,

$$\int M(\mathcal{H}(\mathcal{F}(\theta)))\pi(d\theta) = \int \sum_{i=1}^{m} \sigma^{-1} 2\rho(0)\, \pi(d\theta) = 2\frac{\rho(0)}{\sigma} m. \tag{2.6}$$

Therefore,

$$\int \frac{1}{2} \left| Z^n(\mathbf{y}) - Z(\mathbf{y}) \right| \lambda(d\mathbf{y}) < \frac{K^{\alpha(n)}}{2} 2\frac{\rho(0)}{\sigma} m = \rho(0)\, \frac{K^{\alpha(n)}}{\sigma} m.$$

□

2.3 Setting bounds on posterior statistics

To fix a discretization level to compute the posterior distribution, we do not follow the above approach and find an alternative, using a similar mathematical procedure but with a different purpose and justification. Namely, let $h : \Theta \to \mathbb{R}^+$ be a positive functional, and we are interested in the PE of this functional (e.g., a Quantity of Interest, or QoI), $\hat{h}_y = \int h(\theta)\mathbb{Q}_y[\pi](d\theta)$. Moreover, we have the numeric version of such posterior statistic $\hat{h}_y^n = \int h(\theta)\mathbb{Q}_y^n[\pi](d\theta)$. The

[2] In the proof given in Capistrán et al. (2021), the authors do not explicitly assume that the derivative is negative and forgot the minus sign after removing the absolute value.

proposal is to bound the error, or the RE, in $\hat{h}_y^n$,

$$\left|\hat{h}_y^n - \hat{h}_y\right| \quad \text{or} \quad \frac{\left|\hat{h}_y^n - \hat{h}_y\right|}{\hat{h}_y}.$$

Theorem 2. *With the setting of Theorem 1, let* $h : \Theta \to \mathbb{R}^+$ *be measurable and* $\hat{h}_y^n = \int h(\theta)\mathbb{Q}_y^n[\pi](d\theta)$ *and* $\hat{h}_y = \int h(\theta)\mathbb{Q}_y[\pi](d\theta)$ *exists. Then*

$$\textit{ERE} = \int \frac{\left|\hat{h}_{\mathbf{y}}^n - \hat{h}_{\mathbf{y}}\right|}{\hat{h}_{\mathbf{y}}} Z(\mathbf{y})\lambda(d\mathbf{y}) < 4\rho(0)\frac{K^{\alpha(n)}}{\sigma}m. \tag{2.7}$$

Moreover,

$$\int \left\|\mathbb{Q}_y^n[\pi] - \mathbb{Q}_y[\pi]\right\|_{TV} Z(\mathbf{y})\lambda(\mathbf{y}) < 2\rho(0)\frac{K^{\alpha(n)}}{\sigma}m.$$

Proof. From inequality (2.4) it follows that

$$\left|f^n\left(\mathbf{y}|\theta\right) - f\left(\mathbf{y}|\theta\right)\right| < f\left(\mathbf{y}|\theta\right) K^{\alpha(n)} k_1\left(\mathbf{y},\theta\right).$$

where $k_1\left(\mathbf{y},\theta\right) = \sum_{j=1}^m \left|\frac{\partial}{\partial\eta_j}\phi_{\mathbf{y}}\left(\mathcal{H}(\mathcal{F}^{\alpha(n)}(\theta))\right)\right|$. Therefore,

$$\begin{aligned}
\left|Z_n\left(\mathbf{y}\right) - Z\left(\mathbf{y}\right)\right| &= \left|\int f^n\left(\mathbf{y}|\theta\right)\pi(d\theta) - \int f\left(\mathbf{y}|\theta\right)\pi(d\theta)\right| \\
&\leq \int \left|f^n\left(\mathbf{y}|\theta\right) - f\left(\mathbf{y}|\theta\right)\right|\pi(d\theta) \\
&< K^{\alpha(n)} \underbrace{\int f\left(\mathbf{y}|\theta\right) k_1\left(\mathbf{y},\theta\right)\pi(d\theta)}_{k_2(\mathbf{y})}.
\end{aligned}$$

Then, working with a modified version of the proof of Lemma 3.1 in Christen and Pérez-Garmendia (2021), we have

$$\left|\frac{f^n(\mathbf{y}|\theta)}{Z_n(\mathbf{y})} - \frac{f(\mathbf{y}|\theta)}{Z(\mathbf{y})}\right| < \frac{f(\mathbf{y}|\theta)}{Z(\mathbf{y})}\frac{K^{\alpha(n)}k_2(\mathbf{y})}{Z(\mathbf{y})} + \frac{f(\mathbf{y}|\theta)}{Z(\mathbf{y})}K^{\alpha(n)}k_1(\mathbf{y},\theta).$$

Thus, $\forall \mathbf{y}$ we have

$$\begin{aligned}
\left|\hat{h}^n_{\mathbf{y}} - \hat{h}_{\mathbf{y}}\right| &= \left|\int h(\theta)\frac{f^n(\mathbf{y}|\theta)}{Z^n(\mathbf{y})}\pi(d\theta) - \int h(\theta)\frac{f(\mathbf{y}|\theta)}{Z(\mathbf{y})}\pi(d\theta)\right| \\
&\leq \int h(\theta)\left|\frac{f^n(\mathbf{y}|\theta)}{Z^n(\mathbf{y})} - \frac{f(\mathbf{y}|\theta)}{Z(\mathbf{y})}\right|\pi(d\theta) \\
&< \int h(\theta)\frac{f(\mathbf{y}|\theta)}{Z(\mathbf{y})}\frac{K^{\alpha(n)}k_2(\mathbf{y})}{Z(\mathbf{y})}\pi(d\theta) + \int h(\theta)\frac{f(\mathbf{y}|\theta)}{Z(\mathbf{y})}K^{\alpha(n)}k_1(\mathbf{y},\theta)\pi(d\theta) \\
&< \hat{h}_y\frac{K^{\alpha(n)}k_2(\mathbf{y})}{Z(\mathbf{y})} + \hat{h}_y K^{\alpha(n)}k_1(\mathbf{y},\theta)
\end{aligned}$$

or

$$\frac{\left|\hat{h}^n_{\mathbf{y}} - \hat{h}_{\mathbf{y}}\right|}{\hat{h}_{\mathbf{y}}} < \frac{K^{\alpha(n)}k_2(\mathbf{y})}{Z(\mathbf{y})} + K^{\alpha(n)}k_1(\mathbf{y},\theta).$$

The bound is not directly computable, however, for the expected value of this RE, $\text{ERE} = \mathbb{E}_{\mathbf{Y}}\left[\frac{|\hat{h}^n_{\mathbf{y}} - \hat{h}_{\mathbf{y}}|}{\hat{h}_{\mathbf{y}}}\right]$, we have

$$\text{ERE} = \int \frac{\left|\hat{h}^n_{\mathbf{y}} - \hat{h}_{\mathbf{y}}\right|}{\hat{h}_{\mathbf{y}}} Z(\mathbf{y})\lambda(d\mathbf{y}) < K^{\alpha(n)}\int k_2(\mathbf{y})\lambda(d\mathbf{y}) + \int K^{\alpha(n)}k_1(\mathbf{y},\theta)Z(\mathbf{y})\lambda(d\mathbf{y}).$$

Note that, for location-scale families and from Eq. (2.6), it follows that

$$\int k_2(\mathbf{y})\lambda(d\mathbf{y}) = \int M(\mathcal{H}(\mathcal{F}(\theta)))\pi(d\theta) < 2\frac{\rho(0)}{\sigma}m,$$

where $M(\mathcal{H}(\mathcal{F}(\theta)))$ is given in Eq. (2.5). Now, for the second term,

$$\begin{aligned}
\int K^{\alpha(n)}k_1(\mathbf{y},\theta)Z(y)\lambda(d\mathbf{y}) &= K^{\alpha(n)}\int k_1(\mathbf{y},\theta)\int f(\mathbf{y}|\theta)\pi(d\theta)\lambda(d\mathbf{y}) \\
&= K^{\alpha(n)}\int\int f(\mathbf{y}|\theta)k_1(\mathbf{y},\theta)\pi(d\theta)\lambda(d\mathbf{y}) \\
&= K^{\alpha(n)}\int k_2(\mathbf{y})\lambda(d\mathbf{y}) < 2\rho(0)\frac{K^{\alpha(n)}}{\sigma}m.
\end{aligned}$$

Therefore

$$\text{ERE} < 4\rho(0)\frac{K^{\alpha(n)}}{\sigma}m.$$

Moreover, if $|h(\theta)| < 1$ then

$$\begin{aligned}\mathbb{E}_{\mathbf{Y}}\left[\left\|\mathbb{Q}_y^n[\pi] - \mathbb{Q}_y[\pi]\right\|_{\text{TV}}\right] &= \int \frac{1}{2}\max_{|h|\leq 1}\left|\int h(\theta)\,\mathbb{Q}_y^n[\pi](d\theta) - \int h(\theta)\,\mathbb{Q}_y[\pi](d\theta)\right| Z(\mathbf{y})\lambda(\mathbf{y}) \\ &< 2\rho(0)\frac{K^{\alpha(n)}}{\sigma}m,\end{aligned}$$

and we obtain the second result. □

2.3.1 Choosing a solver discretization

The bound given in Eq. (2.7) allows us to define precision to run the solver. The idea is to keep the ERE below a threshold b (e.g., $b = \frac{1}{10}$) so that the ERE between the numeric and theoretical posterior statistic is of, at most, $b \times 100\%$. We can guarantee ERE $< 4\rho(0)\frac{K^{\alpha(n)}}{\sigma}m < b$ by choosing the numerical error in the FM in (2.1) to satisfy

$$K^{\alpha(n)} < \frac{\sigma}{m}\frac{b}{4\rho(0)}. \tag{2.8}$$

In Capistrán et al. (2021), the authors bound the EABF in (2.3) with $\rho(0)\frac{K^{\alpha(n)}}{\sigma}m < b^*$ and suggest $b^* = \frac{1}{20}$, which corresponds to ERE $< 4b^* = \frac{1}{5}$. That is, in this new perspective they allow a maximum RE of 20%. Instead, here, b should be set in the perspective of the precision needed to estimate the QoI, for example, and is decided by the user in a case by case basis.

Note that, in practice, it is not necessary to establish the global bound (2.1) a priori, but rather by a careful strategy using an after-the-fact error estimation. That is, in most cases, the posterior distribution is sampled using MCMC, which requires evaluation of the approximated likelihood at each of many iterations; an automatic process of global error estimation and control will be necessary to comply with (2.8). This we explain in detail in the next Section.

2.3.2 Choosing the MCMC sample size

MCMC methods have an approximation error due to the use of a finite sample, and certainly, as the MCMC sample size increases, this error decreases. Thus, necessarily, the bound on the ERE, in any statistic, must be larger than that of the MCMC to see the desired effect. Indeed, the MCMC estimate of $\hat{h}_{\mathbf{y}}^n$ should have a lower RE than the desired bound on the ERE for $\hat{h}_{\mathbf{y}}^n$.

To calculate the MCMC error, we will use the result of (Molina and Christen 2021). In Molina and Christen (2021), the authors present a criterion to determine the sample size of a

stochastic simulation with MCMC, which guarantees a certain precision in the estimation of the parameters. This criterion is based on simple classical usage of the Centra Limit Theorem to build a confidence interval on MC estimates. This simple procedure is put in terms of the desired significant digits for the estimate. It includes a heuristic procedure to produce a rough estimate of the coefficient of variation (CV) of the functional from an initial small sample size.

Consequently, as in Molina and Christen (2021), to select the MCMC sample size to obtain the desired precision, we proceed as follows. Let $a \in \mathbb{R}^+$ be a QoI, and expressing this number in scientific notation, we have:

$$a = m_a 10^q = a_0.a_1a_2...10^q \quad \text{with} \quad a_0 \neq 0 \quad \text{and} \quad q \in \mathbb{Z}, \tag{2.9}$$

m_a is referred to as the mantissa of a, $1 \leq m_a \leq 10$, and $a_0.a_1a_2...$ is the decimal expansion of m_a. The idea is to establish how many digits p are required to be correct with probability close to 1 in an MC estimator of a. Thus, they are correct from a_0 to a_p and the exponent q.

For our case, we consider a functional $h : \Theta \to \mathbb{R}^+$, as in Sec. 2.3, and the QoI is the PE of this functional, $\hat{h}_{\mathbf{y}} = \int h(\theta)\mathbb{Q}_{\mathbf{y}}[\pi](d\theta)$. Let $X_1, X_2, \ldots, X_T$ be an independent sample of the posterior distribution $\mathbb{Q}_{\mathbf{y}}[\pi]$, obtained from MC simulation, and let

$$\hat{h}_T = \frac{1}{T}\sum_{i=1}^{T} h\left(X_i\right)$$

be an estimator of $\hat{h}_{\mathbf{y}}$. Expressing $\hat{h}_{\mathbf{y}}$ in scientific notation gives $\hat{h}_{\mathbf{y}} = m_h 10^q$. Now, taking the exponent of $\hat{h}_{\mathbf{y}}$ we will express $\hat{h}_T$ in the following form, $h_T = m_T 10^q$. The integrated autocorrelation time (IAT) of the MCMC sample is estimated. Then we thin our chains, i.e., we discard all but every k-th (k is the calculated IAT) observation, to obtain a pseudo-independent sample.

In Molina and Christen (2021), the authors ensure, with high probability ($> .9999$), that with a sample size of $T = 16\,(\text{CV})^2\,10^{2p}$ in the MCMC, an accuracy of precision of p significant figures is achieved in the estimation of the mantissa of $\hat{h}_{\mathbf{y}}$, i.e.,

$$|m_h - m_T| < \frac{1}{2}10^{-(p-1)}.$$

Note that, in general, one cannot directly estimate the CV $= \frac{\sigma}{\mu}$, since doing so incurs a circular argument concerning the primary goal of guaranteeing accuracy in the estimation of m_h. First, a preliminary estimate of the CV is made with a small sample size to see roughly the sample size needed. Then a more accurate conventional estimate of the CV is made.

Now, expressing the numeric version of the QoI, $\hat{h}^n_{\mathbf{y}} = \int h(\theta)\mathbb{Q}^n_{\mathbf{y}}[\pi](d\theta)$, in scientific nota-

tion, taking the exponent of $\hat{h}_y$, we have $\hat{h}_{\mathbf{y}}^n = m_h^n 10^q$. Then, the RE in $\hat{h}_{\mathbf{y}}^n$ is given by

$$\left|\frac{\hat{h}_{\mathbf{y}} - h_{\mathbf{y}}^n}{\hat{h}_{\mathbf{y}}}\right| = \left|\frac{m_h 10^q - m_h^n 10^q}{m_h 10^q}\right| = \left|\frac{m_h - m_h^n}{m_h}\right| < |m_h - m_h^n|\,,$$

since $1 \leq m_h < 10$.

So, if we want an RE of l significant figures in the mantissa (i.e., $|m_h - m_h^n| < \frac{1}{2}10^{-(l-1)}$), the error in the MCMC must have $p > l$ significant figures (i.e., $|m_h - m_T| < \frac{1}{2}10^{-(p-1)}$), to be able to see the effect correctly.

The way to proceed in the examples that we present in this thesis will be as follows. First, define the desired ERE ℓ (number of significant figures), and choose a sample size T in the MCMC that guarantees a precision $p = \ell + 1$. We will then round the estimate obtained with the MCMC to p decimal places, both for the theoretical and the numerical version, and then we will calculate the RE with these estimates.

In the next Sections, we present two examples considering Bayesian UQ problems for 1D and 2D heat equations, respectively, to illustrate the proposed strategy and its performance for bounding ERE.

2.4 A 1D heat equation inferring the thermal conductivity

The determination of the thermal conductivity is essential in many thermal design problems since the calculated thermal results depend on the estimated thermal conductivity. In the following example, we present one application of inverse heat conduction problem.

Let us consider the thermal conductivity problem for the stationary heat equation in 1D. Namely

$$-\frac{d}{dx}\left(a\,(x)\,\frac{du\,(x)}{dx}\right) = f(x), \qquad x \in (0,1)\,, \tag{2.10}$$

subject to Dirichlet boundary conditions $u\,(0) = u\,(1) = 0$, with forcing term $f\,(x) = \sin\,(\pi x)$ and thermal conductivity $a(x) > 0$ that varies with the space parameter x. Here, the functions a and f are assumed to be continuous on $[0,1]$, and $0 < \alpha_0 \leq a(x) \leq \alpha_1 < \infty$, $\forall x \in (0,1)$. The FM in this case is: given $a(x)$, solve the above differential equation to find u, namely $\mathbf{F}[a] = u$.

In this example, the FM is not available analytically, and a numeric FM is used. We use an error estimation in the FM to bound the ERE of the PE of the function $a(x)$.

The Finite Element Method (FEM) is a numerical method for solving PDEs (see Brenner and Scott 2008; Reddy 2006, for example). The numerical solution of (2.10) is computed using the FEM, which allows us to calculate a local error estimation in the L_2 norm (see Babuška and

Rheinboldt 1978, for more details), given by

$$\|u_h - u\|_{L_2(I_i)} = \left(\int_{x_{i-1}}^{x_i} (u_h - u)^2 \, dx \right)^{1/2} \leq \frac{h^2}{\pi^2 a^i_{\text{min}}} \|r\|_{L_2(I_i)}, \quad i = 1, \ldots, n,$$

where n is the number of elements, u_h the numerical solution with step size h, $I_i = [x_{i-1}, x_i]$, $r(x) = f(x) + \frac{d}{dx}\left(a(x)\frac{du_h(x)}{dx}\right)$ is the residual, and

$$a^i_{\text{min}} = \min_{x \in I_i} a(x) = a\left(\frac{x_i + x_{i-1}}{2}\right)(1 + O(h)) \text{ as } h \to 0.$$

Then, the error estimation $\hat{K}_0$ is computed by

$$\hat{K}_0 = \max_{I_i} \frac{h^2}{\pi^2 a^i_{\text{min}}} \|r\|_{L_2(I_i)}, \quad i = 1, \ldots, n. \tag{2.11}$$

The inference problem is the estimation of the function $a(x)$ (in particular, we are going to infer the logarithmic transformation $b(x) = \log(a(x))$) given the observation functional $\mathcal{H}(u) = (u(x_1), u(x_2), \ldots, u(x_m))'$ at a fixed locations x_j, $j = 1, \ldots, m$.

We simulate a synthetic data set with the true thermal conductivity given by $a(x) = k_0 - r\frac{k_0}{1+\exp\left(-xa+\frac{a}{s}\right)}$, and error model $y_j = u(x_j) + \sigma\varepsilon_j$, where $\varepsilon_j \sim N(0,1)$, with the following parameters $k_0 = 5$, $r = 0.9$, $a = 20$, $s = 2$ and $\sigma = 0.0005$ (to maintain a 0.01 signal-to-noise ratio). The data are plotted in Figure 2.1(b). We consider $m = 30$ observations at locations x_j regularly spaced between 0 and 1.

In order to define the parametric space, the function $a(x)$ is represented as a third-order b-spline that passes through the set of points $\{a_i\}_{i=0}^k$, where $a_i = \exp(b(x_i))$. Therefore, the parameter space is defined by $\theta = \{b_i\}_{i=0}^k$, where $b_i = b(x_i)$. In this case, since the FEM used is numerically demanding we keep the prior truncation fixed to $k+1$ terms; namely the control points $\{a_i\}_{i=0}^k$, at locations $\{x_i\}_{i=0}^k$, for the third-order b-spline. In this example we take $k = 10$.

Regarding the prior distribution for the parameters $\{b_i\}_{i=0}^k$, we define their prior using a Gaussian Markov random field with zero mean and a sparse precision matrix (inverse-covariance), encoding statistical assumptions regarding the value of each element b_i based on the values of its neighbors (see details in Bardsley and Kaipio 2013). We restrict the support of $-\infty < \beta_1 \leq b(x) \leq \beta_2 < \infty$, this implies that $a_i = \exp(b_i) \in [\exp(\beta_1), \exp(\beta_2)]$. Then the parameter space is compact and there exists a global bound for (2.11), complying with (2.1). To sample from the posterior distribution, we use the t-walk (Christen and Fox 2010).

Note that, considering independent data with a Gaussian model ($\rho(0) = \frac{1}{\sqrt{2\pi}}$), the first part of assumption 1 is right, and furthermore, as the observation operator is the identity, $\mathcal{H} \circ \mathcal{F}$ and $\mathcal{H} \circ \mathcal{F}^{\alpha(n)}$ are continuous. Thus, all the necessary assumptions for Theorem 1 are satisfied.

	True	PE-n=500	PE-n=150	RE	b
a_0	5.00×10^0	5.07×10^0	5.06×10^0	$1.8{\times}10^{-3}$	$1.5{\times}10^{-1}$
a_1	5.00×10^0	5.08×10^0	5.12×10^0	$8.5{\times}10^{-3}$	$1.5{\times}10^{-1}$
a_2	4.99×10^0	4.97×10^0	4.93×10^0	$8.7{\times}10^{-3}$	$1.5{\times}10^{-1}$
a_3	4.92×10^0	4.95×10^0	4.96×10^0	$2.0{\times}10^{-3}$	$1.5{\times}10^{-1}$
a_4	4.46×10^0	4.39×10^0	4.37×10^0	$4.6{\times}10^{-3}$	$1.5{\times}10^{-1}$
a_5	2.75×10^0	2.86×10^0	2.85×10^0	$1.0{\times}10^{-3}$	$1.5{\times}10^{-1}$
a_6	1.04×10^0	1.29×10^0	1.26×10^0	$1.6{\times}10^{-2}$	$1.5{\times}10^{-1}$
a_7	$5.81{\times}10^{-1}$	$6.08{\times}10^{-1}$	$6.44{\times}10^{-1}$	$5.9{\times}10^{-2}$	$1.5{\times}10^{-1}$
a_8	$5.11{\times}10^{-1}$	$4.92{\times}10^{-1}$	$4.64{\times}10^{-1}$	$5.7{\times}10^{-2}$	$1.5{\times}10^{-1}$
a_9	$5.02{\times}10^{-1}$	$4.60{\times}10^{-1}$	$4.76{\times}10^{-1}$	$3.5{\times}10^{-2}$	$1.5{\times}10^{-1}$
a_{10}	$5.00{\times}10^{-1}$	$5.14{\times}10^{-1}$	$4.59{\times}10^{-1}$	$1.1{\times}10^{-1}$	$1.5{\times}10^{-1}$

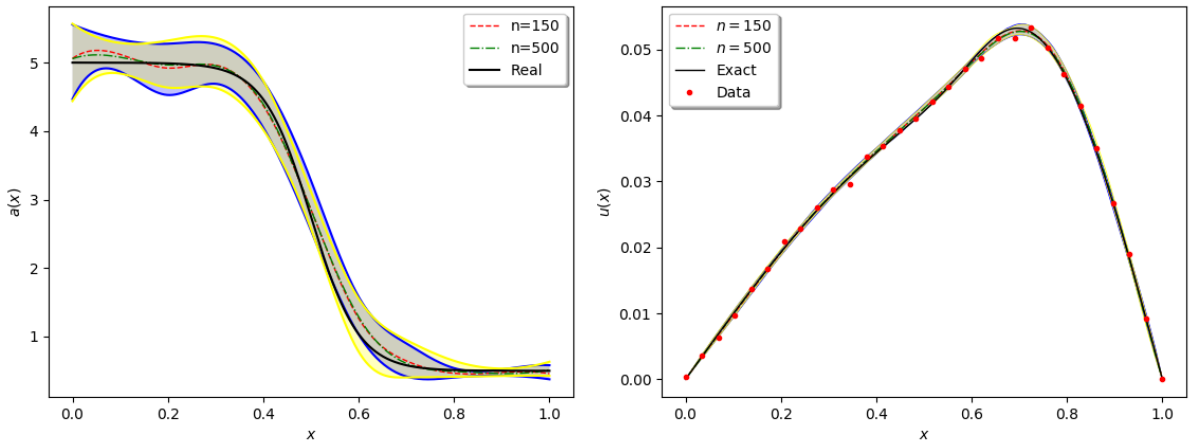
Table 2.1: Comparison of the PE for parameters $\{a_i\}_{i=0}^k$ using $n = 150$ and $n = 500$ elements in the FEM. The estimated RE for each parameter and the tolerance b for the ERE is also presented. It is observed that the RE is less than b in all cases.

Therefore, if we want to guarantee a ERE of 15% ($b = 1.5{\times}10^{-1}$) in the numerical PE of parameters $\{a_i\}_{i=0}^k$, then the threshold in (2.8), for the numerical error in the FM, is $B := \frac{\sigma}{m}\frac{b}{4\rho(0)} = 1.68 \times 10^{-6}$.

Regarding the numerical solver, we begin with a relatively large step size $h = 0.02$ (considering $n = 50$ elements in the FEM) and start the MCMC. At each iteration the FM is first computed along with its error estimation $\hat{K}_0$. If the solution u_h does not satisfy the estimated global bound, i.e., $\hat{K}_0 > B$, we increase the number of elements by 50 ($h = 1/(n + 50)$), until the bound is met. For $h = 0.0066$, $n = 150$ elements in the FEM, the bound is achieved for all iterations.

For comparison, a finer grid is considered with $h = 0.002$, $n = 500$ elements. It is important to point out that we will use this as a reference since we do not have the exact solution of $u(\cdot)$. The results are shown in Figure 2.1. We took 100,000 iterations of the t-walk, with the resulting IAT this guarantees that the error in the MCMC ($0.5 \times 10^{-1.1}$) is less than $b = 0.5 \times 10^{-.52}$ (15%) for all parameters $\{a_i\}_{i=0}^k$. Note that with an ERE of 15% 1 significant figures are achieved, while with the MCMC 2 significant figures are achieved, as explained in Sec. 2.3.2. With $n = 150$ the sampling took 3 min and with $n = 500$, 16 min; in a standard 2.6Ghz processor computer.

We can see that the RE is less than the tolerance b in all parameters. We observe that for several a_i's the RE was much lower than b, but for a_{10}, the RE was very close to tolerance. This may speak about the tightness of the bound in (2.8).



(a) Thermal conductivity $a(x)$ (b) Solution $u(x)$

Figure 2.1: (a) The true conductivity $a(x)$ (black), the PE with $n = 150$ elements (red) and $n = 500$ elements (green) in the FEM. (b) The exact solution $u(x)$ (black), the PE with $n = 150$ elements (red) and $n = 500$ elements (green). Shaded areas represent the uncertainty in the model fit, as draws from the posterior distribution, using 150 elements (blue) and 500 elements (yellow).

2.5 A 2D heat equation inferring the initial condition

In this example, we consider a 2D PDE inverse problem. The FM is available analytically, and a numeric FM is also used. In this case, the numeric error is directly calculated, and we infer only two parameters.

We present a 2D heat equation problem to determine the initial conditions from observations of transient temperature measurements taken within the domain at a time $t = t_1$. The heat transfer PDE is given by

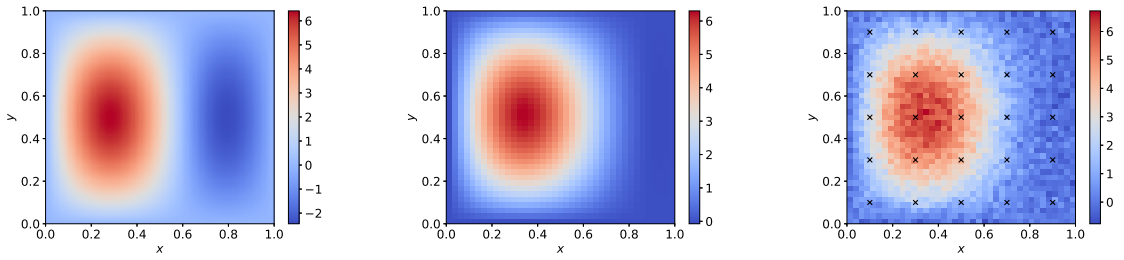
$$\begin{aligned} \frac{\partial u}{\partial t} &= \alpha \Delta u, \text{ in } \quad D = (0,1) \times (0,1)\,, \\ u(x,y,t) &= 0 \quad \text{on} \quad \partial D \\ u(x,y,0) &= f\,(x,y)\,. \end{aligned} \tag{2.12}$$

Taking the forcing term $f\,(x,y) = b \sin(\pi x) \sin(\pi y) + c \sin(2\pi x) \sin(\pi y)$ as initial condition, the PDE has an analytical solution

$$u(x,y,t) = b \exp\left(-2\alpha\pi^2 t\right) \sin(\pi x) \sin(\pi y) + c \exp\left(-5\alpha\pi^2 t\right) \sin(2\pi x) \sin(\pi y)\,.$$

A numerical solution of the boundary value represented in (2.12) is also computed using the

FEM within FEniCS (Martin et al. 2015), which allows us to calculate the error in the numerical solver using the exact solution.



(a) (b) (c)

Figure 2.2: Heat equation in 2D, (a) exact solution at $t = t_1$, (b) numerical solution using FEM with FEniCS with mesh 40×40 with $\Delta t = 0.067$ and (c) numerical solution with an additive Gaussian noise with variance $\sigma = 0.3$ and data point locations.

The inferential problem is to estimate $\theta = (b, c)$ given measurements of u at time $t_1 = 0.3$. A priori we took independent truncated Gamma distributions for b and c with parameters $(2, 0.7)$ and $(2, 0.4)$ respectively, both restricted to $[0, 8]$. Certainly, the theoretical and the numeric FMs are continuous, and since the support is compact we may conclude that the error bound in (2.1) exists for all θ.

We simulate a synthetic data set with the error model

$$Y_i = u(x_i, y_i, t_1) + \sigma\varepsilon_i,$$

where $\varepsilon_i \sim N(0, 1) \quad i = 1, \ldots, m$, $\sigma = 0.3$ (using a the signal to noise ratio of 5%), with $b = 3$ and $c = 5$. The data are plotted in Figure 2.2(b). We consider $m = 25$ observations, $(x_i, y_i), \quad i = 1, 2, \ldots, m$ regularly spaced on D. Since we have an analytic solution, if we run the PDE solver we may calculate the maximum absolute error, K_0, exactly.

Here, as in Example 2.4, all the requirements of the Theorem 1 are fulfilled. Therefore, if we want to guarantee an ERE of 20% ($b = 2.0e - 1$) in the numerical PE of parameters, then the error bound for the FM as stated in (2.8) is $B \simeq 0.0015$. To sample from the posterior distribution, we also use the t-walk.

Regarding the numerical solver we start with a large step size of $\Delta x = \Delta y = 0.1$ and $\Delta t = 0.268$, and calculate K_0. If the solution does not comply with the bound, that is if $K_0 > B$, a new solution is attempted by reducing the step-size in $\Delta x, \Delta y$ and Δt by half, until the global absolute errors are within the bound, i.e., $K_0 \leqslant B$. The resulting mesh is $\Delta x = \Delta y = 0.025$ and $\Delta t = 0.067$.

We compare the above FEM numerical FM with the exact FM, with 200,000 iterations of our MCMC, to guarantee the error in the MCMC ($0.5 \times 10^{-1.45}$) is less than $b = 0.5 \times 10^{-0.39}$

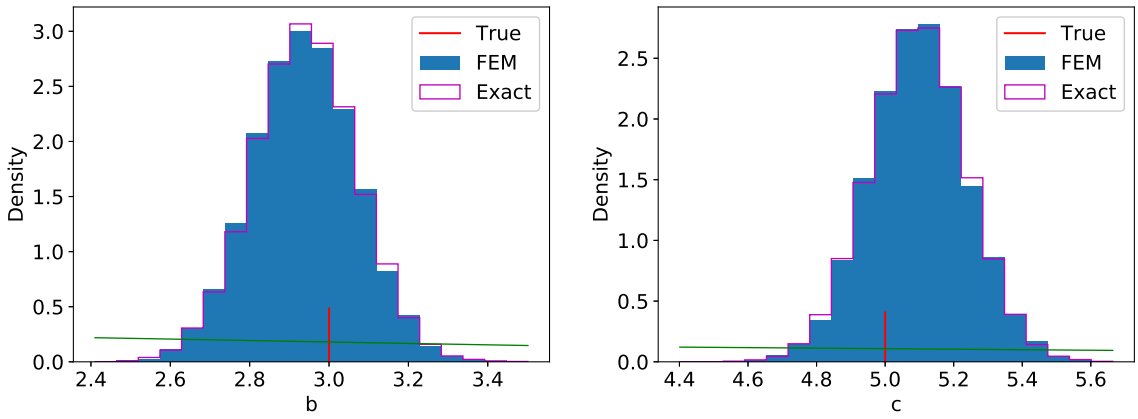


Figure 2.3: Comparison between numerical (blue) a theoretical (magenta) posteriors for parameters b (left) and c (right), of the initial conditions of the 2D heat equation. The prior, in this scale, appears in green.

for both parameters. The results are shown in Figure 2.3 and in Table 2.2. We see that the RE of the PEs of the two parameters are less than the tolerance b, in this case, by several orders of magnitude. Note that with an ERE of 20% 1 significant figure is required, while with the MCMC 2 significant figures are achieved, as explained in Sec. 2.3.2. In table 2.2, the PE was rounded to 3 figures because if it were rounded to 2 figures, the RE would give 0 in both parameters.

	True	PE-Exact	PE-FEM	RE	b
b	3.00×10^0	2.936×10^0	2.938×10^0	$6.8{\times}10^{-4}$	$2.0{\times}10^{-1}$
c	5.00×10^0	5.101×10^0	5.098×10^0	$7.8{\times}10^{-4}$	$2.0{\times}10^{-1}$

Table 2.2: Comparison of the PE for parameters b and c using the exact FM and the FEM approximate FM. The estimated RE for each parameter and the tolerance b for the ERE is also presented. It is observed that the RE is less than b in both cases. Note that the PE was rounded to 3 figures because if it was rounded to 2 the RE would give 0 in both parameters.

2.6 Discussion

Here, we found a bound on the RE in PEs for some functional. Also we detailed how this bound can be used to find a discretization in the numerical method used to approximate the FM. We also presented two examples (we have experimented with several others). In all cases, using the bound (2.8), the RE in the PEs of parameters is less than the established tolerance b, i.e., the numerical error in the posterior was successfully bounded. With these examples, we showed

the importance of having an estimator for the absolute global error (AGE) of the numeric FM, which is feasible, general, and applicable.

Our results lead to the error bound for the FM's error in (2.8) that is quite simple to calculate, provided, as we said, that after-the-fact error estimates of the FM numeric approximation are available. This bound depends on the specific maximum RE (b) selected and on the noise for the data (σ, and on $\rho(0)$). The posterior variance, or specifically, the CV of the statistic, enters through the MCMC estimation, and statistics with larger CV will require larger effective sample sizes for their estimation.

We have not discussed the scenario when error parameters σ are unknown and/or there is a correlation in the data. In this case, we may consider that a priori θ and σ are independent, a correlation structure for y, and equivalent results should follow. This was discussed in an unpublished manuscript (Christen et al. 2016) with a previous version of Capistrán et al. (2021). We only need to prove that the new likelihood follows assumption 1, in particular, that it is bounded.

In general, one needs to decide, sooner or later, a discretization level n. Although there are multilevel methods that work with several discretizations simultaneously (Cliffe et al. 2011; Katsiolides et al. 2018), even, in that case, one nevertheless needs, a priori, a sense of what is a large or a small FM error. Our bound in (2.8) is an attempt to provide precisely that, a sense of what is a large or a small FM error in the perspective of the Bayesian UQ problem at hand and the selected QoI's, taking into account the noise level in the data (σ) and the sample size.

Nonetheless, our approach only makes sense in the case when working with different discretizations for the FM is computationally feasible.

CHAPTER 3

Bounding numerical posterior error in the Bayesian UQ analysis of a semilinear evolution PDE

In this Chapter, we elaborate on results obtained in Chapter 2 for bounding the numerical posterior error for Bayesian UQ problems, now considering FMs arising from the solution of a semilinear evolution PDE. Results in Chapter 2 demand an estimate for the AGE of the numeric FM. Our contribution is a numerical method for computing the AGE for semilinear evolution PDEs and shows the potential applicability of our results in Chapter 2 in this important wide range family of PDEs. Numerical examples are given to illustrate the efficiency of the proposed method by keeping an ERE in the PEs of parameters less than an established tolerance b. We will experiment with $b = 0.05$ and $b = 0.1$.

3.1 Introduction

Physical models involving semilinear evolution PDE's arise in several fields of science and engineering (Tzafestas 2013; Diagana 2018). These PDEs are used to describe many complex nonlinear settings in applications such as vibration and wave propagation, fluid mechanics, plasma physics, quantum mechanics, nonlinear optics, solid-state physics, chemical kinematics, physical chemistry, population dynamics, and many other areas of mathematical modeling.

In this Chapter, we derive an after-the-fact error estimate for a numerical approximation of

semilinear evolution PDEs of the form

$$\frac{\partial u}{\partial t} = D\frac{\partial^2 u}{\partial x^2} + F\left(u, \frac{\partial u}{\partial x}, \theta\right), \tag{3.1}$$

defined on the region $t \in [0, \tau], \quad x \in [a, b]$, with left and right boundary conditions

$$u(a,t) = g(t) \text{ and } u(b,t) = h(t), \quad 0 \leq t \leq \tau, \tag{3.2}$$

and initial condition

$$u(x,0) = f(x), \quad a \leq x \leq b. \tag{3.3}$$

In Eq. (3.1), D is the diffusion coefficient, θ is a parameter (possibly a vector) of interest, and F is a non-linear operator.

We obtain the numerical solution for Eq. (3.1) with the method of lines; discretizing first in the space with the FD method and solving the resulting system in time with the RKCK method. This scheme is widely used to solve numerically evolution PDEs (Bellman and Kalaba 1965; Bastani and Salkuyeh 2012; Sari and Güraslan 2011). However, computable after-the-fact error estimates for these methods have not yet been derived.

The idea behind our construction of the error estimates for the PDE in Eq. (3.1) is the error estimate available for the RKCK method. Our numerical method uses these error estimates, in time, for the resulting ODE system. The spatial discretization error is treated as the solution of an initial value problem, approximated by the RKCK method. The truncation error (TE) introduced for the approximation with FDs is computed using the solutions in two different mesh sizes. In modern computers, the added computational effort can be reduced to an equivalent result to that solving the PDE conventionally (on a single mesh) since evaluating the solution in two different meshes may be easily parallelized.

We will then use this after-the-fact error estimate to apply the results of Chapter 2, for the solution of the Bayesian UQ problem associated with the PDE given in Eq. (3.1), to bound the ERE in a QoI. Numerical examples are given to illustrate the efficiency of the proposed method. We obtain RE in the PEs of unknown parameters less than the established tolerance b.

The Chapter is organized as follows. In Section 3.2, we present a numerical method used for solving evolution PDEs numerically. In Section 3.3, we derive our after-the-fact error estimate for semi-linear evolution differential equations. The accuracy of our error estimate is evaluated for some classic examples. In Section 3.4, we propose an algorithm that incorporates the after-the-fact error estimate to bound the posterior numerical error in a QoI. Numerical examples are given in Section 3.5 to illustrate the efficiency of the proposed Algorithm. Finally, a conclusion is given in Section 3.6.

3.2 Numerical solution

This Section introduces a particular version of a standard numerical procedure, namely, the method of lines, to solve semilinear evolution PDEs. This procedure has been widely used for solving evolution PDEs (Bellman and Kalaba 1965; Bastani and Salkuyeh 2012; Sari and Güraslan 2011). The basic idea of the method is to replace the spatial derivatives on the PDE with an algebraic approximation in order to obtain an ODE system. The resulting system is then solved with a standard ODE solver. We discretize in the space with the FD method and solve the ODE system with the RKCK method. We called this method FD-RKCK. In the next Section, we will then use this method to build a computable after-the-fact error estimate in order to be able to use the posterior error bound results of Chapter 2.

For simplicity, we denote $\dot{u} := \frac{\partial u}{\partial t}$, $u' := \frac{\partial u}{\partial x}$, and $F\left(u, u'\right)$ instead of $F\left(u, u', \theta\right)$. Moreover, without losing generality, we can set $D = 1$ in Eq. (3.1). We consider a one-dimensional uniform mesh, Ω_h, on the region $[a, b]$, with nodes x_i, for $i = 0, 1, \ldots, N$, where

$$\Omega_h : a = x_0 < x_1 < \cdots < x_N = b, \tag{3.4}$$

and a constant step size h between any two successive nodes (i.e., $h = x_i - x_{i-1}$).

To solve the PDE in Eq. (3.1), we start by linearizing it by using the quasilinearization method which was proposed by Bellman and Kalaba (1965). We split the function F into its linear (L) and nonlinear (N) components, and rewriting Eq. (3.1) in the form

$$\dot{u} = u'' + \mathsf{L}[u, u'] + \mathsf{N}[u, u']. \tag{3.5}$$

For example, in Section 3.3 we use the Fisher equation where $F = ru(1-u)$, thus $\mathsf{L} = ru$ and $\mathsf{N} = -ru^2$. Afterwards, the nonlinear operator N is approximated with a Taylor series, assuming that the difference $u_{i+1,\cdot} - u_{i,\cdot}$ and all its spatial derivatives are small. Hence

$$\mathsf{N}[u_{i+1,\cdot}, u'_{i+1,\cdot}] \approx \mathsf{N}[u_{i,\cdot}, u'_{i,\cdot}] + \phi_{0,i}[u_{i,\cdot}, u'_{i,\cdot}] \cdot (u_{i+1,\cdot} - u_{i,\cdot}) + \phi_{1,i}[u_{i,\cdot}, u'_{i,\cdot}] \cdot (u'_{i+1,\cdot} - u'_{i,\cdot}), \tag{3.6}$$

where $u_{i,\cdot} := u(x_i, t)$ is the solution of Eq. (3.1) evaluated in (x_i, t), and

$$\phi_{k,i}[u_{i,\cdot}, u'_{i,\cdot}] := \frac{\partial \mathsf{N}[u_{i,\cdot}, u'_{i,\cdot}]}{\partial u^{(k)}}, \quad k = 0, 1.$$

For simplicity, $u^{(i)}$ denotes the i-th derivative. Substituting Eq. (3.6) into Eq. (3.5), we get

$$\dot{u}_{i+1,\cdot} \approx u''_{i+1,\cdot} + \mathsf{L}[u_{i+1,\cdot}, u'_{i+1,\cdot}] + \mathsf{N}[u_{i,\cdot}, u'_{i,\cdot}] + \phi_{0,i}[u_{i,\cdot}, u'_{i,\cdot}] \cdot (u_{i+1,\cdot} - u_{i,\cdot}) + \phi_{1,i}[u_{i,\cdot}, u'_{i,\cdot}] \cdot (u'_{i+1,\cdot} - u'_{i,\cdot}), \tag{3.7}$$

for $i = 1, ..., N-2$. Now, the spatial partial derivatives are approximated using the central difference formula. For simplicity, we use the simplest spatial derivative approximations here, while the analysis can be extended for other (e.g., five-point stencil) approximations as well,

$$u'_{i,\cdot} \approx \frac{u_{i+1,\cdot} - u_{i-1,\cdot}}{2h}, \quad u''_i \approx \frac{u_{i+1,\cdot} - 2u_{i,\cdot} + u_{i-1,\cdot}}{h^2}, \tag{3.8}$$

for $i = 1, \ldots, N-1$, and

$$u'_{0,\cdot} \approx \frac{u_{1,\cdot} - u_{0,\cdot}}{h}. \tag{3.9}$$

Substituting Eqs. (3.8)–(3.9) in to Eq. (3.7), joint with the boundary condions (3.2) and initial condition (3.3), we get the following semi-discrete differential equation:

$$\dot{\mathbf{V}}_h(t) = \frac{1}{h^2}\mathbf{A}_{xx}\mathbf{V}_h(t) + \mathbf{F}(t, \mathbf{V}_h(t)) \tag{3.10}$$

$$\mathbf{V}_h(0) = \mathbf{U}(0) \tag{3.11}$$

where $\mathbf{V}_h(t) = (v_{1,\cdot}, v_{2,\cdot}, \ldots, v_{N-1,\cdot})^T$ approximates

$$\mathbf{U}(t) = (u_{1,\cdot}, u_{2,\cdot}, \ldots, u_{N-1,\cdot})^T, \tag{3.12}$$

which $\mathbf{U}$ is the exact solution of the PDE (3.1) on the mesh Ω_h, and $\mathbf{F}$ is the approximate operator F in matrix form, see A.1 for details.

Remark. *The semi-discrete differential equation (3.10) have a TE $O(h^p)$:*

(i) If F does not have a nonlinear component, the quasi-linear approximation (3.6) is not necessary. Thus, the TE for the central difference formula is not affected ($p = 2$).

(ii) If N is non-linear in u', the quasi-linear approximation (3.6) introduces a TE of first-order, which is propagated when u' is approximated using the central difference formula. Thus, the order of the TE for (3.10) is less than 2 ($p < 2$).

(iii) If N is non-linear in u and linear in u', the TE introduced for the quasi-linear approximation (3.6) is not propagated as in case (ii). Then, the TE in (3.10) is slightly affected ($p \approx 2$).

In order to solve the resulting ODE system (3.10)-(3.11), with $N-2$ equations, we use the RKCK method. This method uses six function evaluations to calculate fourth and fifth-order accurate solutions. The difference between these solutions is then taken to be the error (fourth-order) of the solution; see Burden and Faires (2011) for details. The available error estimate is

the reason to solve the resulting ODE's system with this Runge-Kutta (RK) method, and it will be used in turn, in Section 3.3, for computing the after-the-fact error of the numerical solution of Eq. (3.1).

Setting $\mathbf{G}(t, \mathbf{V}_h(t)) = \frac{1}{h^2}\mathbf{A}_{xx}\mathbf{V}_h(t) + \mathbf{F}(t, \mathbf{V}_h(t))$, a RK scheme applied to the ODE system (3.10), at a uniform time grid

$$0 = t_0 < t_1 < \cdots < t_n < \cdots < t_{M-1} < t_M = \tau; \quad t_{n+1} = t_n + k, \tag{3.13}$$

is given by

$$\mathbf{K}_{1,n} = \mathbf{G}\left(t_n, \mathbf{W}_{\cdot,n}\right), \tag{3.14}$$

$$\mathbf{K}_{l,n} = \mathbf{G}\left(t_n + c_l k, \mathbf{W}_{\cdot,n} + k\sum_{j=1}^{l-1} a_{lj}\mathbf{K}_{j,n}\right), \quad l = 2, 3, \ldots, 6, \tag{3.15}$$

$$\mathbf{W}_{\cdot,n+1} = \mathbf{W}_{\cdot,n} + k\sum_{l=1}^{6} b_l \mathbf{K}_{l,n}, \quad n = 1, \ldots, M-1,$$

where $\mathbf{W}_{\cdot,n+1}$ is the approximation for $\mathbf{V}_h(t_{n+1})$, (a_{lj}) are the Runge-Kutta coefficients, $\mathbf{b} = (b_1, b_2, \ldots, b_6)$ are the quadrature nodes, and $\mathbf{c} = (c_1, c_2, \ldots, c_6)$ are the quadrature weights of the RK scheme. $k = \Delta t > 0$ is the step size in time and define a uniform grid.

In order to have stable solutions in explicit schemes, the step size in time is related to the discretization through the Courant-Friedrichs-Lewy (CFL) condition (Courant et al. 1928), which restricts the step size in time based on the eigenspectrum of the discretized spatial operator. The CFL condition for the FD-RKCK scheme considering only the pure diffusion is

$$\frac{\Delta t}{\Delta x^2} \leq \frac{1}{4} B_m,$$

where $B_m = \max_i b_i$, and $b_1, b_2, \ldots, b_6$ are the quadrature nodes for the RK method used, see A.2 for details.

3.3 After-the-fact error estimates

In this Section, we propose a numerical procedure to obtain an after-the-fact error estimate of the AGE for the numerical solution of Eq. (3.1) presented above. We use the error estimation in the time-stepping given for the RKCK method and estimate the leading term of the TE in space stepping. This scheme can be extended for differential equations of non-linear evolution, but some additional considerations regarding the stability of the solution must be taken into account.

In Section 3.2, we obtained the semi-discrete differential equation (3.10)–(3.11), with a unique solution vector, $\mathbf{V}_h(t)$, being a grid function on Ω_h. This initial value problem solved with the RKCK method yields approximations $\mathbf{W}_{.,n}$ to $\mathbf{V}_h(t_n)$. The global error at the spatial mesh points at knot t_n is defined by

$$\mathbf{E}_h(t_n) := \mathbf{W}_{.,n} - \mathbf{U}(t_n), \tag{3.16}$$

where $\mathbf{U}$ is the exact solution of the PDE (3.1) on the mesh Ω_h defined in (3.12). The vector $\mathbf{E}_h$ may also be written as a combination of the ODE global error, this is defined as the error made by the solver, i.e.,

$$e_h(t_n) = \mathbf{W}_{.,n} - \mathbf{V}_h(t_n), \tag{3.17}$$

and the spatial discretization error defined by

$$\eta_h(t_n) = \mathbf{V}_h(t_n) - \mathbf{U}(t_n). \tag{3.18}$$

The function $\eta(t)$ represents the accumulation of the spatial TE when we solve (3.10)-(3.11),

$$\mathrm{TE}_h(t) = \mathbf{G}(t, \mathbf{U}) - \dot{\mathbf{U}}(t). \tag{3.19}$$

From Eqs. (3.17)–(3.18), the global error $\mathbf{E}_h(t_n)$ may be written as the sum of the global time and spatial error, i.e.,

$$\mathbf{E}_h(t_n) = e_h(t_n) + \eta_h(t_n). \tag{3.20}$$

We assume that $u(t, x)$ is p-times differentiable with respect to x and fourth-times continuously differentiable with respect to t. Then, it holds for the global space and time error that $||\eta_h|| = O(h^p)$ and $||e_h(t_n)|| = O(k^4)$, $n = 1, \ldots, M$, respectively.

Although there are already articles that propose an error estimation of this type of PDE's (Bellman and Kalaba 1965; Bastani and Salkuyeh 2012; Sari and Güraslan 2011), they do not show a direct way for computing such error. These estimations leave constants that cannot be easily computed. More importantly, these methods do not guarantee that the error estimate is an upper bound for the actual error.

Note that, $||\mathbf{E}_h(t_n)||_\infty \leq ||e_h(t_n)||_\infty + ||\eta_h(t_n)||_\infty$, i.e., what we are proposing, is a bound for the global error, which is required in Assumption 2. The ODE global error (3.17) is calculated using the error estimation of RKCK (Bellman and Kalaba 1965). The spatial discretization error (3.18) is treated as the solution of an initial value problem (3.21)-(3.22), approximated by the RKCK method, which involves an estimation for the TE. The Richardson extrapolation provides a suitable estimate of the TE (Roy 2010). The idea is to calculate the solution using a one-step size h and then compute them again with half the space step ($h/2$). The result obtained using

two steps size is more accurate than using the single-step size h. Their difference can be used as an estimate of the TE, which is proportional to h^p. In the following Sections, we find an entirely computable upper bound (our error estimate) of the numerical error, with the same asymptotic behavior as the true error.

3.3.1 Spatial discretization error

We may obtain an equation for the evolution of $\eta(t)$ by adding terms to both sides of (3.10):

$$\dot{\mathbf{V}}_h(t) - \dot{\mathbf{U}}(t_n) = \mathbf{G}(t, \mathbf{V}_h) - \mathbf{G}(t, \mathbf{U}) + \mathbf{G}(t, \mathbf{U}) - \dot{\mathbf{U}}(t_n).$$

From the initial condition (3.11) and using the definition $\eta(t)$ in the above equation, the accumulation of the spatial discretization error is the solution to the initial value problem:

$$\dot{\eta}(t) = \mathbf{G}(t, \mathbf{V}_h) - \mathbf{G}(t, \mathbf{U}) + \mathrm{TE}_h(t), \quad t \in (0, \tau] \tag{3.21}$$

$$\eta(0) = 0. \tag{3.22}$$

Assuming $\mathbf{G}$ to be twice continuously differentiable, we use the approximation:

$$\frac{\partial \mathbf{G}}{\partial \mathbf{V}_h} \approx \frac{\mathbf{G}(t, \mathbf{V}_h) - \mathbf{G}(t, \mathbf{U})}{\mathbf{V}_h - \mathbf{U}}. \tag{3.23}$$

Finally, we rewrite (3.21)-(3.22) to get

$$\dot{\eta}(t) = \frac{\partial \mathbf{G}}{\partial \mathbf{V}_h}\eta(t) + \mathrm{TE}_h(t), \quad t \in (0, \tau] \tag{3.24}$$

$$\eta(0) = 0. \tag{3.25}$$

The integration of (3.24)–(3.25) is performed using M steps of size k of the RKCK method, as in the solution of the semi-discrete differential equation (3.10)–(3.11). In each RKCK step, $\frac{\partial \mathbf{G}}{\partial \mathbf{V}_h}$ is approximated using the approximations $\mathbf{W}_{.,n}$ and $\mathbf{W}_{.,n+1}$ to $\mathbf{V}_h$ at time t_{n+1}, i.e.,

$$\frac{\partial \mathbf{G}}{\partial \mathbf{V}_h} \approx \frac{\mathbf{G}(t_{n+1}, \mathbf{W}_{.,n+1}) - \mathbf{G}(t_{n+1}, \mathbf{W}_{.,n})}{\mathbf{W}_{.,n+1} - \mathbf{W}_{.,n}}. \tag{3.26}$$

3.3.2 Spatial and time error

The ODE global error (3.17) is computed by the error estimation given by the RKCK method. This scheme uses an RK method with a fifth-order local TE to estimate the local error in an RK method of fourth-order. Both with the same number of stages $s = 6$, Runge Kutta matrix $\mathbf{A}$,

and weights $\mathbf{c}$, while their nodes $\hat{\mathbf{b}}$ and $\mathbf{b}$, respectively, are different; see Bellman and Kalaba (1965) for details.

Let $\mathbf{W}_{\cdot,n+1}$ the $n+1$ approximation of $\mathbf{V}_h(t_{n+1})$ of fourth-order, and let $\mathbf{Y}_{\cdot,n+1}$ be obtained by the fifth-order method starting at $\mathbf{W}_{\cdot,n}$, namely

$$\mathbf{W}_{\cdot,n+1} = \mathbf{W}_{\cdot,n} + k\sum_{i=1}^{s} b_i \mathbf{K}_{i,n} \quad \text{and} \quad \mathbf{Y}_{\cdot,n+1} = \mathbf{W}_{\cdot,n} + k\sum_{i=1}^{s} \hat{b}_i \mathbf{K}_{i,n}, \tag{3.27}$$

The local TE $\hat{\tau}_{\cdot,n+1}$ at node t_{n+1} of the RK method is defined as the error made in step $n+1$ of the solver if starting at the exact value $\mathbf{W}_{\cdot,n}$. The estimation of $\hat{\tau}_{\cdot,n+1}$ for the RKCK method is given by

$$\hat{\tau}_{\cdot,n+1} = \mathbf{Y}_{\cdot,n+1} - \mathbf{W}_{\cdot,n+1} = k\sum_{i=1}^{s} (\hat{b}_i - b_i)\mathbf{K}_{i,n},$$

and the global error at knot t_{n+1} is

$$\hat{e}_{\cdot,n+1} = \sum_{j=1}^{n+1} \tilde{\tau}_{\cdot,j}. \tag{3.28}$$

In each RK iteration, we solve the equation for the spatial discretization error (3.24)–(3.25),

$$\hat{\eta}_{n+1} = \hat{\eta}_n + k\sum_{i=1}^{s} b_i \hat{K}_{i,n}, \tag{3.29}$$

where

$$\begin{aligned} \hat{\mathbf{K}}_{1,n} &= \mathbf{H}(t_n, \hat{\eta}_n) \\ \hat{\mathbf{K}}_{l,n} &= \mathbf{H}\left(t_n + c_l k, \eta_n + k\sum_{j=1}^{l-1} a_{lj}\hat{K}_{j,n}\right), \quad l = 2, 3, \ldots, 6. \end{aligned}$$

$\mathbf{H}$ is the right side of (3.24). From (3.28)–(3.29), an estimation for the global error $\mathbf{E}_h$ (3.20) at knot t_n is given by,

$$\hat{\mathbf{E}}_{\cdot,n+1} \approx \hat{e}_{\cdot,n+1} + \hat{\eta}_{n+1}. \tag{3.30}$$

Note that to solve fully (3.24)–(3.25), we need an estimate for the TE. This estimation is done in parallel to be used in (3.29). Below we give details for computing the TE.

3.3.3 Spatial truncation error

The TE is the difference between the discretized equations and the original partial differential equations. It contains the errors due to the discretization of the PDE and the errors due to the grid. For the finite difference scheme used to approximate the spatial operator, we have that the

TE at time t has rate order $O(h^p)$,

$$\text{TE}_h(t) \approx O(h^p).$$

An efficient strategy to estimate the spatial TE by Richardson extrapolation is proposed in Roy (2010). We will adopt this approach to our setting. The actual mesh used to compute the numerical solution to the PDE is used as the fine mesh in the Richardson extrapolation process. Suppose we are given a second semi-discretization of the PDE system (3.1), now with doubled local mesh sizes defined as follows,

$$\Omega_{2h} : a = z_0 < z_1 < z_2 < \ldots < z_{N/2} = b, \quad z_i = x_{2i}, \quad i = 0, \ldots, N.$$

This mesh is called the coarse mesh. We assume that the solution $V_{2h}(t)$ to the discretized PDE, on the coarse mesh $2h$, exists and is unique. The Richardson extrapolation gives an estimation of the TE for the fine mesh at time t,

$$\widehat{\text{TE}}(t) \approx \frac{\mathbf{V}_{2h} - R_{2h}(\mathbf{V}_h)}{2^p - 1}, \tag{3.31}$$

where $R_{2h}(\mathbf{V}_h)$ is the operator that returns the solution $\mathbf{V}_h$ in the grid Ω_{2h}.

Remark. *The after-the-fact error estimate of the AGE for the FD-RKCK solution of* (3.1)*, on* $\Omega_h \times (0, \tau]$*, is given by*

$$\hat{K} = ||\hat{\mathbf{E}}_{\cdot,M}||_\infty,$$

where Ω_h *is defined in* (3.4)*, the discretization grid for* $(0, \tau]$ *is defined in* (3.13)*, and* $\hat{\mathbf{E}}_{\cdot,M}$ *is an estimate for the global error* (3.20)*.*

In Algorithm 1, we describe the steps necessary to compute the numerical solution of Eq. (3.1), with the after-the-fact error estimation. We call this algorithm DF-RKCK.

To test our algorithm, we consider three classical semi-linear PDEs, of the form (3.1): Example 1 (Fisher equation), Example 2 (Fitzhugh-Nagumo equation), and Example 3 (Burgers-Fisher equation). The three examples used also have analytic solutions, allowing us to compute the actual numerical error and compare it with our estimates. In Figure 3.1, a graphical comparison is shown between our numerical implementation approximations and the exact solution for the three examples. Table 3.1 shows the convergences order of the solution obtained with the DF-RKCK Algorithm. It can be seen that the method achieves full convergence for the error (order 2) for Examples 1 and 2, but the order of convergence for Example 3 is 1, and this is due to the non-linearity of F in u', as was mentioned before.

Remark. *To compute the numerical convergence rate, we use*

$$p = \log_2 \left(\frac{||u_{4h} - u_{2h}||_\infty}{||u_{2h} - u_h||_\infty} \right).$$

Algorithm 1: DF-RKCK

Step 1. **Initialization:**

- Spatial step size h. The step size in time is given for keeping the stability condition $k = \alpha h^p$
- Initial conditions W_0^h and W_0^{2h}; initial time t_0; parameter θ; $\hat{\mathbf{e}}_{\cdot,0} = \mathbf{0}$ and $\widehat{\mathrm{TE}} = 0$
- The RK matrix $\mathbf{A} = (a_{ij})$, the nodes $\mathbf{b}$ and $\hat{\mathbf{b}}$, and the weights $\mathbf{c}$

Step 2. Discretizing (3.1) with the FD method for h and $2h$, as is described in Section 3.2;

Step 3. Solve (3.10) with the Cash–Karp method for the step size h and $2h$;

For $n = 1, 2, \ldots, M$:

$$\mathbf{K}_{1,n}^h = \mathbf{G}\left(t_n, \mathbf{W}_{\cdot,n}\right); \qquad \hat{\mathbf{K}}_{1,n} = \mathbf{H}\left(t_n, \hat{\eta}_n\right)$$
$$\mathbf{K}_{1,n}^{2h} = \mathbf{G}\left(t_n, \mathbf{W}_{\cdot,n}\right)$$

Step 4. For $i = 2, 3, \ldots, 6$:

$$\mathbf{K}_{i,n}^h = \mathbf{G}\left(t_n + c_i k, \mathbf{W}_{\cdot,n} + k\sum_{j=1}^{i-1} a_{ij}\mathbf{K}_{j,n}^h\right)$$
$$\hat{\mathbf{K}}_{i,n} = \mathbf{H}\left(t_n + c_i k, \hat{\eta}_n + k\sum_{j=1}^{i-1} a_{lj}\hat{\mathbf{K}}_{j,n}\right)$$
$$\mathbf{K}_{i,n}^{2h} = \mathbf{G}\left(t_n + c_i k, \mathbf{W}_{\cdot,n} + k\sum_{j=1}^{i-1} a_{ij}\mathbf{K}_{j,n}^{2h}\right)$$

Step 5. Compute

$$\mathbf{W}_{\cdot,n+1}^h = \mathbf{W}_{\cdot,n}^h + k\sum_{i=1}^{6} b_i\mathbf{K}_i^h; \quad \hat{\eta}_{n+1} = \hat{\eta}_n + k\sum_{i=1}^{6} b_i\hat{\mathbf{K}}_{i,n}$$
$$\mathbf{W}_{\cdot,n+1}^{2h} = \mathbf{W}_{\cdot,n}^{2h} + k\sum_{i=1}^{6} b_i\mathbf{K}_i^{2h}; \quad \widehat{\mathrm{TE}} = \frac{\left\|R_{2h}(\mathbf{W}_{\cdot,n+1}^h) - \mathbf{W}_{\cdot,n+1}^{2h}\right\|_\infty}{h^p\left(2^p - 1\right)}$$
$$\hat{\mathbf{e}}_{\cdot,n+1} = \hat{\mathbf{e}}_{\cdot,n} + k\sum_{i=1}^{6}(b_i - \hat{b}_i)\mathbf{K}_i^h$$
$$\widehat{\mathbf{E}}_{\cdot,n+1} = \hat{\mathbf{e}}_{\cdot,n+1} + \hat{\eta}_{n+1}$$

Step 6. Compute the maximum AGE in the solution approximated $\mathbf{W}^h$,

$$\widehat{K} = \left\|\widehat{\mathbf{E}}\right\|_\infty$$

Step 7: Output: $\mathbf{W}^h$, $\widehat{K}$

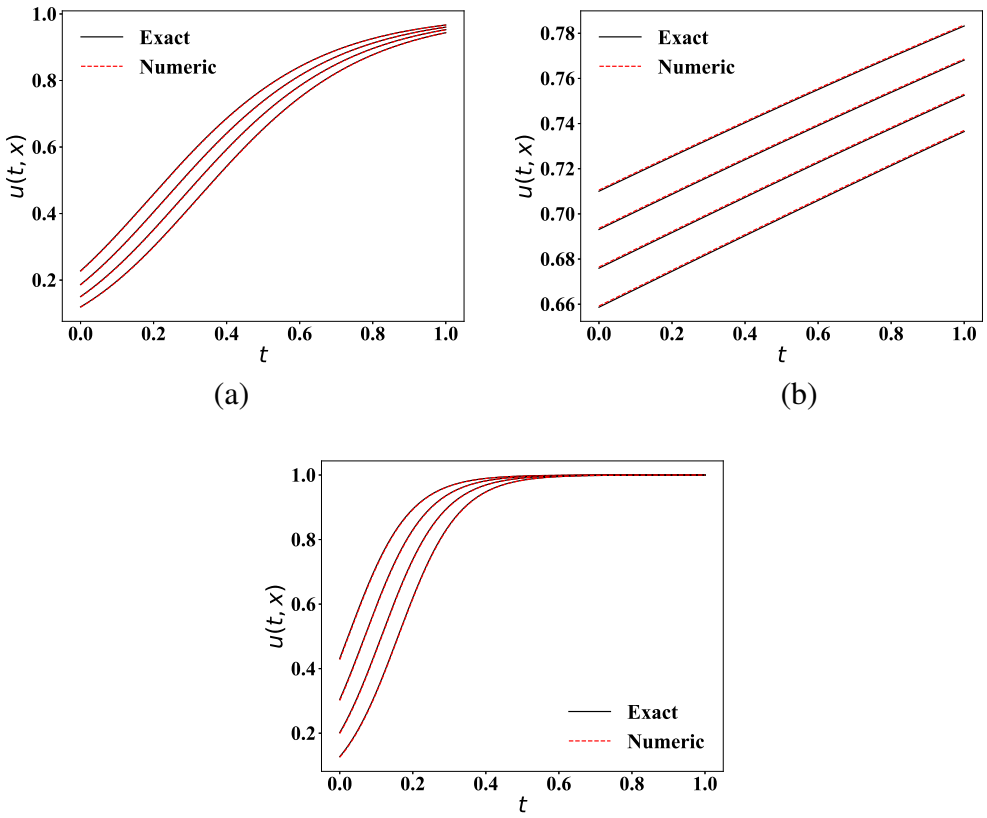


Figure 3.1: The analytical and approximate solution in $x = 0.1, 0.3, 0.5, 0.7$, with step sizes in space $h = 0.0125$ and in time $k = 0.0001$, for (a) the Fisher's equation, (b) the Fitzhugh-Nagumo equation, and (c) the Burgers-Fisher equation. Both solutions overlap in all cases.

In Figure 3.2, we show the maximum error between the exact solution and the numerical solution for the three examples considered, comparing it to our error estimates. We can see that the estimation proposed for the AGE is an upper bound for the exact error. The numerical implementation has been performed in Python, using the scipy, numpy, and matplotlib packages. For the sake of reproducibility, all code is available in a Github repository (see Daza-Torres and Montesinos-López (2020)).

Example 1 (Fisher equation)**.**

Fisher's equation belongs to the class of reaction-diffusion equation and is encountered in chemical kinetics and population dynamics applications. The equation is given by

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial^2 x} + ru\left(1 - u\right), \tag{3.32}$$

h	Example 1	Example 2	Example 3
0.0125	2.098279	1.992578	1.075049
0.0083	2.068255	1.995410	1.052625
0.00625	2.052187	1.996692	1.040450
0.005	2.042208	1.997418	1.032826

Table 3.1: Convergence Order.

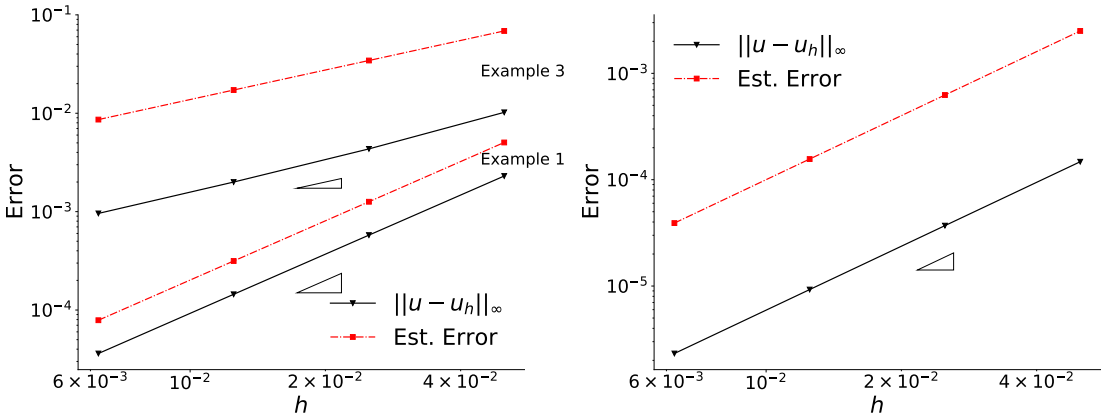


Figure 3.2: The maximum error between the exact solution and our DF-RKCK method against the error estimation: (a) the Fisher and the Burgers-Fisher equation, (b) the Fitzhugh-Nagumo equation. Different step sizes (h) in space are taken and for time we let $k = \alpha h^2$, with $\alpha = 3/4$. The triangles indicate the slope of the exact error.

with boundary and initial conditions

$$\begin{aligned} u(0,t) &= \frac{1}{(1+e^{-5t})^2}, \quad 0 \le t \le \tau, \\ u(1,t) &= \frac{1}{(1+e^{1-5t})^2}, \quad 0 \le t \le \tau, \\ u(x,0) &= \frac{1}{(1+e^{x})^2}, \quad 0 \le x \le 1. \end{aligned}$$

This PDE has the following analytic close form solution

$$u\left(x,t\right) = \frac{1}{\left[1+\exp\left(\sqrt{\frac{r}{6}}x - \frac{5r}{6}t\right)\right]^2}, \quad x \in \left[0,1\right], \quad \text{and} \quad t \in \left[0,\tau\right],$$

where r is a parameter. The non-linear operator is $F(u,u',r) = ru(1-u)$; hence the appropriate linear component is $\mathsf{L} = ru$, and the non-linear component is $\mathsf{N} = -ru^2$; we see that the operator F does not depend on u' so the method achieves order 2, as can be seen in Table 3.1. For the

examples in Figs. (3.1)–(3.2) and Table 3.1, we use $r = 4$, and this parameter will be tried to be identified using synthetic data in Section 3.5.

Example 2 (Fitzhugh-Nagumo equation)**.**

The Fitzhugh-Nagumo equation is given by

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial^2 x} + u\left(1-u\right)\left(u-a\right), \quad 0 < a < 1, \tag{3.33}$$

with boundary and initial conditions

$$\begin{aligned} u(0,t) &= \frac{1}{2}\left(1+a\right) + \frac{1}{2}\left(1-a\right)\tanh\left(\frac{(1-a^2)}{4}t\right), \quad 0 \le t \le \tau, \\ u(1,t) &= \frac{1}{2}\left(1+a\right) + \frac{1}{2}\left(1-a\right)\tanh\left(\sqrt{2}\left(1-a\right)\frac{1}{4} + \frac{(1-a^2)}{4}t\right), \quad 0 \le t \le \tau, \\ u(x,0) &= \frac{1}{2}\left(1+a\right) + \frac{1}{2}\left(1-a\right)\tanh\left(\sqrt{2}\left(1-a\right)\frac{x}{4}\right), \quad 0 \le x \le 1. \end{aligned}$$

The analytic solution for this PDE is given by

$$u\left(x,t\right) = \frac{1}{2}\left(1+a\right) + \frac{1}{2}\left(1-a\right)\tanh\left(\sqrt{2}\left(1-a\right)\frac{x}{4} + \frac{(1-a^2)}{4}t\right),\ x \in [0,1]\,,\ \text{and}\ t \in [0,\tau]\,,$$

where a is a parameter. The non-linear operator is $F(u,u',a) = u\left(1-u\right)\left(u-a\right)$; hence the appropriate linear component is $\mathsf{L} = -au$, and the non-linear component is $\mathsf{N} = u^2\left(1-u+a\right)$; we see that the operator F does not depend on u', so the method achieves order 2, as can be seen in Table 3.1. For the examples in Figs. (3.1)–(3.2) and Table 3.1, we use $a = 0.3$ and this parameter will be tried to identify using synthetic data in Section 3.5.

Example 3 (Burgers-Fisher equation)**.**

The Burgers-Fisher equation is given by

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial^2 x} - ruu' + su\left(1-u\right), \tag{3.34}$$

with the initial condition

$$u\left(x,0\right) = \frac{1}{2} + \frac{1}{2}\tanh\left(-\frac{r}{4}x\right), \quad 0 \le x \le 1,$$

and with boundary and initial conditions

$$\begin{aligned} u(0,t) &= \frac{1}{2}+\frac{1}{2}\tanh\left(\left(\frac{r^2}{8}+\frac{s}{2}\right)t\right), \quad 0\le t\le\tau, \\ u(1,t) &= \frac{1}{2}+\frac{1}{2}\tanh\left(-\frac{r}{4}\left[1-\left(\frac{r}{2}+\frac{2s}{r}\right)t\right]\right), \quad 0\le t\le\tau, \\ u(x,0) &= \frac{1}{2}+\frac{1}{2}\tanh\left(-\frac{r}{4}x\right), \quad 0\le x\le 1. \end{aligned}$$

This problem also has an analytic solution given by

$$u\left(x,t\right)=\frac{1}{2}+\frac{1}{2}\tanh\left(-\frac{r}{4}\left[x-\left(\frac{r}{2}+\frac{2s}{r}\right)t\right]\right), \quad x\in\left[0,1\right], \quad \text{and} \quad t\in\left[0,\tau\right],$$

where r and s are parameters. The non-linear operator is $F(u,u',a)=-ruu'+su\left(1-u\right)$; hence the appropriate linear component is $\mathsf{L}=su$, and the non-linear component is $\mathsf{N}=-ruu'-su^2$. Note that the order of convergence for the error is 1 because F is nonlinear in u'. For the examples in Figs. (3.1)–(3.2) and Table 3.1, we use $r=4.5$ and $s=5.5$, and these parameters will be tried to identify using synthetic data in the next Section.

3.4 Numeric posterior error bounds

In this Section, we discuss how to incorporate the after-the-fact error estimate, proposed in Section 3.3, in the results of Chapter 2, to bound the error on posterior statistics. To do that, we propose a MCMC algorithm with refinement to fulfill the global bound (2.8) for all θ in the parametric space of interest.

Algorithm 2 describes a strategy for incorporating the bound (2.8) and the after-the-fact error estimate, proposed in Section 3.3, to bound the error on posterior statistics.

The basic idea of Algorithm 2 is to start with a relatively large step size (e.g., $h=0.1$), and the step size in time is established to keep the stability condition $k=\frac{3}{4}h^2$. At each iteration, θ_i, of the MCMC, the FM, $\mathcal{F}^h(\theta_i)$, is computed, including the after-the-fact error estimate $\hat{K}^h_{\theta^i}$, using Algorithm 1. If the error in the FM does not comply with the bound in (2.8), then run the solver again reducing the spatial step size by half. In this process, we therefore assure (2.1) for all $\theta\in\Theta$.

3.5 Numerical examples

In this Section, we use the three previous examples to show the performance of Algorithm 2, in the solution of the corresponding Bayesian UQ problem, using simulated data sets.

Algorithm 2: Numerical refinement for the FM in the MCMC algorithm

Step 1: **Initialization:**

- Spatial step size h (large)
- Standard error (σ), sample size (m), and $\rho(0)$
- Calculate the error bound $B = \frac{\sigma}{m}\frac{b}{4\rho(0)}$, with a tolerance b
- Initial value for the parameter, θ^0
- MCMC length M (number of simulations)

for $i \leftarrow 1$ **to** M **do**

Step 2. Compute the FM, $\mathcal{F}^h(\theta^{i-1})$, and the error estimation $\hat{K}^h_{\theta^{i-1}}$, using Algorithm 1;

Step 3.;

if $\hat{K}^h_{\theta^{i-1}} > B$ **then**

- Set $h = h/2$;
- Return to *Step 2*;

else

- Simulate θ^i with some MCMC algorithm;

Step 4: Output: $(\theta^0, \theta^1, \ldots, \theta^M)$

We simulate data as follows. The (synthetic) observations, $\mathbf{y} = (y_1, \ldots, y_m)$, are generated under an independent Gaussian model

$$f_o(\mathbf{y}|\eta) = \prod_{i=1}^{m} \sigma^{-1} \rho\left(\frac{y_i - \eta_i}{\sigma}\right)$$

with $\rho(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, i.e.,

$$y_i = \mathcal{H}_i(\mathcal{F}(\theta)) + \sigma\varepsilon_i, \tag{3.35}$$

where the ε_i's are independent and identically distributed as $\mathcal{N}(0,1)$, $\theta \in \Theta \subset \mathbb{R}^d$ is a vector of unknown parameters, and $\mathcal{F}(\theta)$ represents the FM. In all our examples, we consider the variance, σ^2, to be known.

The solution of (3.1) with its initial and boundary conditions defines our FM, and we take $\mathcal{H}(x) = x$ as the observation operator. We consider the Bayesian UQ problem to estimate the parameter θ given observations $\mathcal{H}_i(\mathcal{F}(\theta)) = u(x_i, t_1, \theta)$ at some points in space x_i, for $i = 1, 2, \ldots, m$ and at a fixed time $0 < t_1 \leq \tau$. We let the system evolve until time t_1 and then observe it at the spacial locations x_i's. The resulting observations are $y_i = u(x_i, t_1, \theta) + \sigma\varepsilon_i; \varepsilon_i \sim \mathcal{N}(0,1)$. In particular, the QoI is the PE of the parameter Θ. The implementation was done using

MCMC, through a generic MCMC algorithm, called the t-walk (Christen and Fox 2010).

Note that considering independent data with a Gaussian model, the first part of assumption 1 is right, and we only require to verify that $\mathcal{H} \circ \mathcal{F}$ and $\mathcal{H} \circ \mathcal{F}^{\alpha(n)}$ are continuous. Indeed, the latter is true if the observation operator is the identity. With this scheme, all the necessary assumptions for Theorem 1 are satisfied. And, in this case, $\rho(0) = \frac{1}{\sqrt{2\pi}}$ and the threshold $B := \frac{\sigma}{m}\frac{b}{4\rho(0)}$ in (2.8), for the numerical error in the FM, is

$$B = \frac{b}{4}\frac{\sigma}{m}\sqrt{2\pi}. \tag{3.36}$$

Example 4 (IP, Fisher's equation)**.**

We consider the IP to estimate $\theta = r$ in Fisher's equation of Example 1, given measurements of $\mathcal{H}_i(\mathcal{F}(\theta))$ at time $t_1 = 0.4$. The synthetic data are simulated with the error model (3.35), using the analytical solution for the FM, and the following parameters: $\theta = 4$ and $\sigma = 0.008$. The solution of (3.32) with its initial and boundary conditions defines our FM. We consider $n = 8$ observations at locations x_i regularly spaced between 0 and 1. The data are plotted in Fig. 3.3 (a).

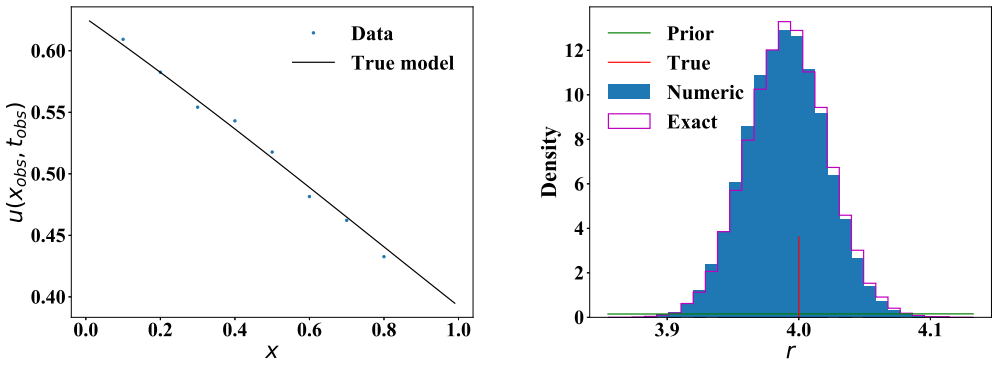
Considering a tolerance in the ERE of 5% ($b = 0.5\times10^{-1}$), and with the standard error and sample size used, the error bound for the FM is $B = 3.52\times10^{-5}$. We require a prior distribution, $\pi(\cdot)$, for the parameter θ; it is assumed $\theta \sim$ Gamma(α_1, β_1) with all known hyperparameters. Regarding the numerical solver, we begin with a (relatively) large step size, $h = 0.05$, and the step size in time is established to keep the stability condition $k = \frac{3}{4}h^2$. Then, we start the Algorithm 2. For $h = 0.0067$, the bound is achieved for all iterations.

We compare the posterior distributions using the numerical FM vs. the exact FM, with 350,000 iterations of the t-walk, to guarantee the precision of 3 decimal places in the PE (i.e., the error in the MCMC, $0.5 \times 10^{-2.4}$, is less than $b = 0.5\times10^{-1}$). Results are reported with 345,000 samples since the first (burn-in) 5,000 are discarded. The results are shown in Fig. 3.3 (b) and Table 3.2. The PE was rounded to 3 figures; it is observed that the RE of the posterior means is less than the tolerance b.

Example 5 (IP, Fitzhugh-Nagumo equation)**.**

For this example, the IP is to estimate $\theta = a$ in the Fitzhugh-Nagumo equation of Example 2, given measurements $\mathcal{H}_i(\mathcal{F}(\theta))$ at time $t_1 = 0.3$. The synthetic data are simulated with the error model (3.35), using the analytical solution for the FM, and the following parameters: $\theta = 0.3$ and $\sigma = 0.009$. The solution of (3.33) with its initial and boundary conditions defines our FM. We consider $n = 8$ observations at locations x_i regularly spaced between 0 and 1. The data are plotted in Fig. 3.4 (a).

To be able to get the posterior distributions, we assume that $\theta \sim$ Gamma(α_2, β_2) with all known hyperparameters. With the standard error and sample size used, and considering



(a) (b)

Figure 3.3: (a) Fisher equation example data (blue points) and true model (black line), considering $r = 4$. (b) Comparison between numerical (blue) and theoretical (magenta) posterior for parameter r. The green line represents the prior distribution.

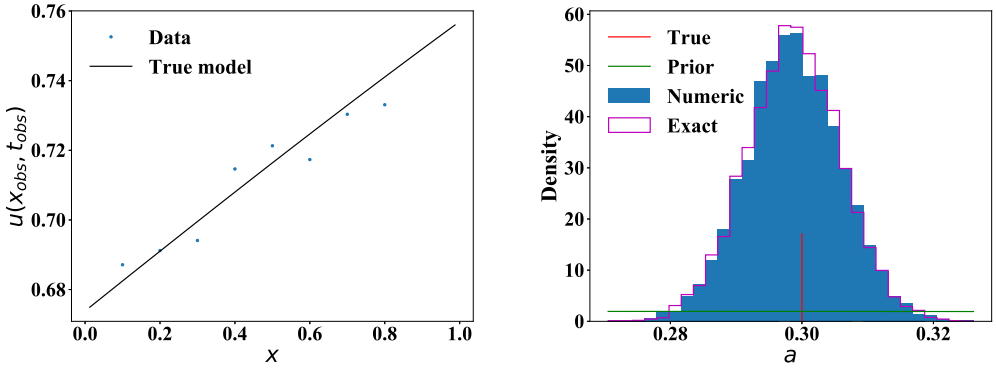
a tolerance in the ERE of 5% ($b = 0.5{\times}10^{-1}$), we have that the error bound for the FM is $B = 3.52 \times 10^{-5}$. Regarding the numerical solver, we begin with a step size, $h = 0.1$, and the step size in time $k = \frac{3}{4}h^2$. Then, we start the Algorithm 2. For $h = 0.0077$, the bound is achieved for all iterations.

We compare the posterior distributions using the numerical FM vs. the exact FM, with 200,000 iterations of the t-walk, to guarantee that the error in the MCMC, $0.5 \times 10^{-2.4}$, is less than the tolerance $b = 0.5{\times}10^{-1}$. Results are reported with 180,000 samples since the first (burn-in) 20,000 are discarded. The results are shown in Fig. 3.4 (b) and in Table 3.2. Note that with an ERE of 5%, 2 significant figures are achieved, while with the MCMC 2.6 significant figures are achieved, as explained in Sec. 2.3.2. In Table 3.2, the PE was rounded to 4 figures because if it were rounded to 2 or 3 figures, the RE would give 0; it is observed that the RE of the PE is less than the tolerance b.

Example 6 (IP, Burgers-Fisher equation)**.**

For this example, the IP is to estimate $\theta = (r, s)$ in the Burgers-Fisher equation of Example 4, given measurements $\mathcal{H}_i(\mathcal{F}(\theta))$ at time $t_1 = 0.2$. The synthetic data are simulated with the error model (3.35), using the analytical solution for the FM, and the following parameters: $\theta = (4.5, 5.5)$ and $\sigma = 0.05$. The solution of (3.34) with its initial and boundary conditions defines our FM. We consider $n = 10$ observations at locations x_i regularly spaced between 0 and 1. The data are plotted in Fig. 3.5 (a).

To get the posterior distributions, we assume independent priors between the parameters of the model. We assume $r \sim$ Gamma (α_r, β_r) and $s \sim$ Gamma (α_s, β_s) with all known hyperparameters. With the standard error and sample size used, and considering a tolerance in the ERE of 10% ($b = 0.5 \times 10^{-0.7}$), we have that the error bound for the FM is $B = 6 \times 10^{-4}$. Regarding



(a) (b)

Figure 3.4: (a) Fitzhugh-Nagumo equation example data (blue points) and true model (black line), considering $a = 0.3$. (b) Comparison between numerical (blue) and theoretical (magenta) posterior for parameter a. The green line represents the prior distribution.

	Example 1	**Example 2**	**Example 3**	
Parameter	r	a	r	s
True	4×10^{0}	$3{\times}10^{-1}$	4.5×10^{0}	5.5×10^{0}
PE-Exact	3.992×10^{0}	$2.9853{\times}10^{-1}$	4.18×10^{0}	5.54×10^{0}
PE-Numeric	3.990×10^{0}	$2.9859{\times}10^{-1}$	4.19×10^{0}	5.53×10^{0}
RE	$5.00{\times}10^{-4}$	$3.40{\times}10^{-4}$	$2.4{\times}10^{-3}$	$3.6{\times}10^{-3}$
b	$5.00{\times}10^{-2}$	$5.00{\times}10^{-2}$	$1.0{\times}10^{-1}$	$1.0{\times}10^{-1}$

Table 3.2: Comparison of the PE of each parameter using the exact and the numeric FM. The PE for each example was rounded to the first decimal place where the RE was not 0, because if we had rounded to the significant figures guaranteed by the MCMC (3,2, and 1, respectively), the RE would have been 0 for all parameters. In all cases, the RE of the PE is less than the tolerance b.

.

the numerical solver, we begin with a step size $h = 0.1$, and the step size in time $k = \frac{3}{4}h^2$. Then, we start the Algorithm 2. For $h = 0.0017$, the bound is achieved for all iterations.

We compare the posterior distributions using the numerical FM vs. the exact FM, with 100,000 iterations of the t-walk, to guarantee that the error in the MCMC, $0.5 \times 10^{-1.5}$, is less than the tolerance $b = 0.5 \times 10^{-0.7}$. Results are reported with 99,500 samples since the first (burn-in) 500 are discarded. The results are shown in Fig. 3.5 (b)–(c) and in Table 3.2. Note that for this case, with the MCMC, only 1 decimal figure is guaranteed in the PE, and with an ERE of 10% less than 1 decimal figure is required. Still, in Table 3.2, the PE was rounded to 2 figures because if it were rounded to 1 figure, the RE would give 0. It is observed that the RE of the PE is less than the tolerance b.

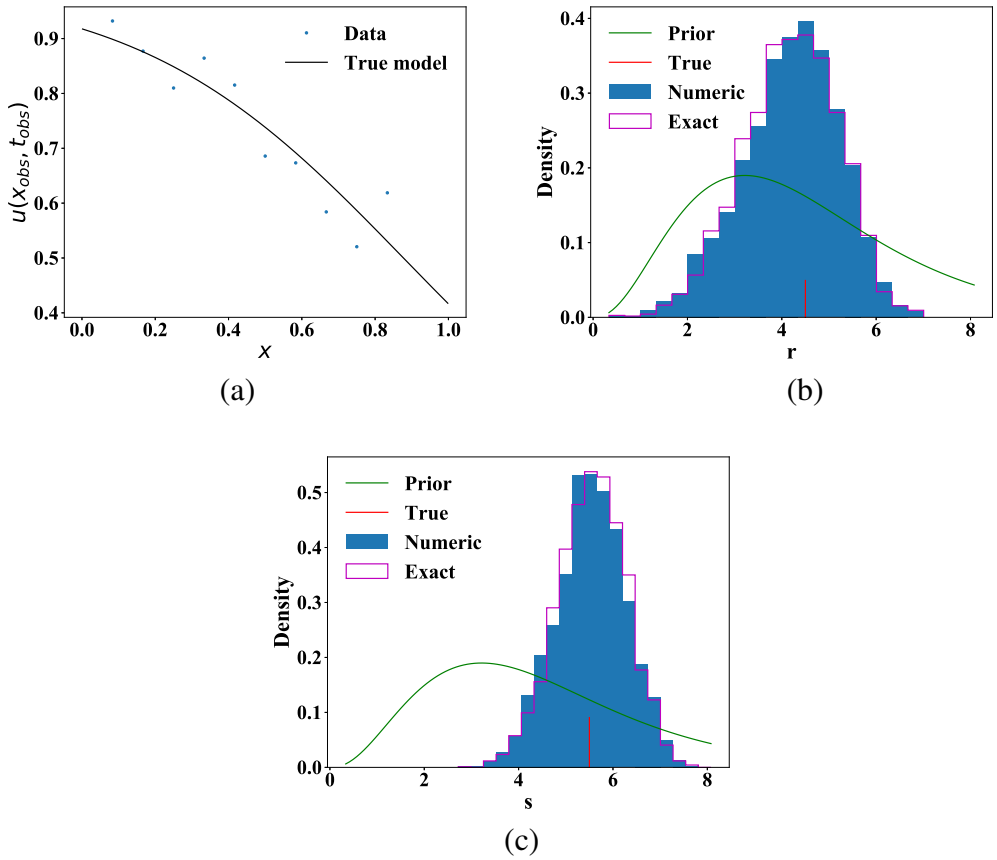


Figure 3.5: (a) Burgers-Fisher equation example data (blue points) and true model (black line), considering $\theta = (4.5, 5.5)$. Histogram from the numerical (blue) and theoretical (magenta) posterior distribution for: (b) parameter r and (c) parameter s. The green line represents the prior distribution.

3.6 Conclusion

This Chapter proposed an error estimation for a class of PDEs motived by its application in the UQ area. Our error estimation allows us to apply the results obtained in Chapter 2 for bounding the ERE in the respective numerical posterior statistics for IPs when the forward mapping involves a semilinear evolution PDE.

We presented three worked examples considering different levels of accuracy for both MCMC and ERE; in all cases, the RE in the PE of parameters is less than the established tolerance b, i.e., the RE in the numerical posterior statistics was successfully bounded.

Although two numerical solutions are required for the error estimation, the added computational effort can be reduced to an equivalent result to that solving the PDE conventionally (on a single mesh) since evaluating the solution in two different meshes may be easily parallelized.

CHAPTER 4

Uncertainty quantification for fault slip inversion

We propose an efficient Bayesian approach to infer a fault displacement from geodetic data in a SSE. Our physical model of the slip process reduces to a multiple linear regression with constraints. Assuming a Gaussian model for the geodetic data and considering a MTN prior distribution for the unknown fault slip, the resulting posterior distribution is also MTN. We propose a prior slip distribution with a detailed correlation structure, to impose natural coherence in the fault slip. Regarding the posterior, we propose an ad hoc algorithm based on Hybrid Optimal Directional Gibbs sampler that allows us to efficiently sample from the resulting high-dimensional posterior slip distribution without supercomputing resources. A synthetic fault slip example illustrates the flexibility and accuracy of the proposed approach. The methodology is also applied to a real data set, for the 2006 Guerrero, Mexico, SSE, where the objective is to recover the fault slip on a known interface that produces displacements observed at ground geodetic stations. As a by-product of our approach, we are able to estimate moment magnitude for the 2006 Guerrero SSE with UQ.

4.1 Introduction

A major task of geophysics is to make quantitative statements about Earth's interior in terms of surface measurements. One fundamental element of earthquake investigations is to estimate the magnitude and distribution of slip along a fault plane. Fault slips may consist of complex and heterogeneous source processes, while limited geodetic data typically leads to an ill-posed IP. Conventionally, regularization is used to transform such IPs into a well-posed optimization

problem for a single-source model. The most common approach is to add Tikhonov regularization terms to smooth the solution (Calvetti et al. 2000; McCaffrey et al. 2007; Wallace and Beavan 2010; Radiguet et al. 2011), as well as including positivity constraints and reducing the solution space (Tago et al. 2021). All these regularization terms try to be justified by the physical processes being modeled. These strategies lead to the solution of a well-posed approximate and computationally feasible problem. However, the uncertainty involved in the solution is not methodologically sound solely using regularization. These schemes produce only limited point-wise solution estimates. Moreover, the lack of sufficient physical interpretation of some critical regularization terms may introduce bias in the solutions without proper justification. For more robust and informative IP solutions, and to formally quantify their uncertainty, we require a different approach.

The Bayesian statistical approach provides a rigorous framework to handle constrains and UQ of IPs. A model for observations is assumed (leading to a likelihood) and prior information is incorporated using probability density functions (PDFs) to determine the posterior distribution through Bayes theorem, which quantifies our inference's uncertainty. A prior PDF is established through its interpretation as a modeling device of the probabilistic prior knowledge available on source movements and imposing model restrictions using truncated PDFs (Fukuda and Johnson 2008; Minson et al. 2013; Amey et al. 2018; Nocquet 2018).

Bayesian techniques have been of limited use for slip inversions, mainly because the simulation from the posterior distribution is not straightforward. This typically occurs when there are many parameters, such as in SSEs (see below for more details).

In modern Bayesian analyses, MCMC algorithms (Robert and Casella 2013) are standard tools to sample from the posterior distribution. Many versions of the MCMC method have been proposed in the literature, but the Metropolis-Hastings and the Gibbs sampler algorithms are the most common (Robert and Casella 2013).

In Minson et al. (2013), the authors developed a framework for Bayesian inversion of finite fault earthquake models. They combined a Metropolis algorithm with simulated annealing and genetic algorithms to sample high-dimensional problems in a parallel computing framework. The method remains computationally expensive despite parallelization. A finite-dimensional Gaussian processes approximation to allow for inequality constraints in the entire domain is proposed in Maatouk and Bay (2017). Their problem is equivalent to simulating a Gaussian vector restricted to convex sets, and they use an improved rejection sampling (see Maatouk and Bay (2016)) in which only the random coefficients in the convex set are selected. They mention that the MTN simulation can be accelerated by MCMC or Gibbs sampling methods (Geweke 1991).

There are many methods available for simulating the MTN distribution (Breslaw 1994; Kotecha and Djuric 1999; Robert 1995; Yu and Tian 2011). Most of these are based on the

Gibbs sampler, which is simple to use and has the advantage of accepting all proposals generated and, therefore, is not affected by poor acceptance rates, such as rejection sampling. These methods work well in many situations but may become very slow in the presence of high correlation and/or high dimensionality.

The authors in Nocquet (2018) show that a MTN prior can be applied to achieve positivity or bound constraints. He employs recent findings in MTN probability calculations (Genz and Bretz 2009) to derive relevant posterior statistics (e.g., posterior marginal PDF, mean and covariances) without performing MCMC sampling. However, the evaluation of these quantities require complex numerical integration over an hyper-rectangle while quantifying the uncertainty of a function f of the parameters (e.g., moment magnitude M_w) is not straightforward. In contrast, when we have available Monte Carlo samples from the posterior, the posterior uncertainty of the quantities of interest may be directly obtained. In Michalak (2008), the author provides a statistically rigorous methodology for geostatistical interpolation and inverse modeling, subject to multiple and spatially-variable inequality constraints. The approach uses a Gibbs sampler to characterize the marginal probability distribution at each estimation point, using a MTN prior probability distribution. This kind of algorithms are systematic Gibbs samplers which makes CPU time increase linearly with dimension (Christen et al. 2017). The authors in Christen et al. (2017) explore an optimality criterion for the MCMC Direction Gibbs algorithm to simulate from a MTN distribution. This criterion consists of minimizing the Mutual Information between two consecutive steps of the Markov chain. The algorithm proposed in Christen et al. (2017) is especially suited for high correlation and high dimensionality; one of the main advantages is that the number of iterations to obtain an independent sample only increases linearly with dimensionality and is commonly close to theoretical limits.

In this work, we propose a Bayesian approach for estimating the parameters in a constrained multiple linear regression model. We propose a prior slip distribution with a detailed correlation structure, to impose natural coherence in the fault movement. Combining the algorithm proposed in Christen et al. (2017) and in Montesinos-López and Christen (2016), we propose an Hybrid Optimal Directional Gibbs algorithm that allows us to sample from high-dimensional problems efficiently when the posterior distribution is a MTN distribution. With the methodology from Christen et al. (2017) the burn-in period is reduced, while with the Montesinos-López and Christen (2016) approach the chain's correlations is reduced. Besides presenting a synthetic example, we apply our method to quantify the uncertainty in the IP of seismic slip along the subduction interface in the 2006 Guerrero, Mexico, SSE. Moreover, with our method, we are able to provide the posterior distribution of the moment magnitude for this event in a straightforward manner.

A SSE is a slip produced at a fault that does not generate seismic waves. However, the induced deformation may be registered at the surface from SSEs lasting several weeks to a

couple of months. SSEs have been observed in different fault configurations around the world (Gao et al. 2012), and the role they play in the seismic cycle is an active research topic (Kato et al. 2012; Ruiz et al. 2014; Radiguet et al. 2016; Cruz-Atienza et al. 2021).

In Mexico, SSEs have been identified in different segments along the subduction region, on the Pacific coast, where the Cocos Plate and the North American Plate collide. In the so-called Guerrero GAP (GGap), before the great Mw8.2 Tehuantepec event on 8 September 2017, SSEs showed a periodicity of approximately four years and a duration from six to twelve months (Cruz-Atienza et al. 2021). The SSE, occurred within Guerrero state in 2006, one of the most studied globally, was recorded at 15 continuous GPS stations (Radiguet et al. 2011; 2012; Cavalié et al. 2013; Bekaert et al. 2015; Tago et al. 2021) as millimetrical displacements on the surface. This event offers the opportunity to analyze the slip's spatial evolution and delimit the characteristics of a typical SSE in the GGap. In this case, the IP consists of recovering the slip along a known interface that produces displacements at the surface observed at those 15 continuous GPS geodetic stations.

4.2 Methodology

4.2.1 Forward map

For the direct problem in our particular setting, we begin with the elastostatic equations representation theorem that models the displacement $\mathbf{u}(\mathbf{x})$, at the coordinates $\mathbf{x}$ of the GPS station, due to a slip $\mathbf{d}(\boldsymbol{\xi})$, produced at a fault Γ, as

$$u_j(\mathbf{x}) = \int_\Gamma T_k\left(S_{ij}(\boldsymbol{\xi};\mathbf{x}), \hat{\mathbf{n}}(\boldsymbol{\xi})\right) d_k(\boldsymbol{\xi})\, dS(\boldsymbol{\xi}), \qquad j \in \{x, y, z\} \tag{4.1}$$

where $T_k(\cdot,\cdot)$ is the k-component of the traction on the fault computed through the Somigliana tensor, $S_{ij}(\boldsymbol{\xi};\mathbf{x})$, and the fault normal vector $\hat{\mathbf{n}}(\boldsymbol{\xi})$ (Udías et al. 2014). If the traction and the slip are projected along the dip component, d-direction, and along the strike direction, s-direction, Eq. (4.1) can be written in matrix form as

$$\begin{bmatrix} u_x(\mathbf{x}) \\ u_y(\mathbf{x}) \\ u_z(\mathbf{x}) \end{bmatrix} = \int_\Gamma \begin{bmatrix} T_s\left(S_{ix}(\boldsymbol{\xi};\mathbf{x}), \hat{\mathbf{n}}(\boldsymbol{\xi})\right) & T_d\left(S_{ix}(\boldsymbol{\xi};\mathbf{x}), \hat{\mathbf{n}}(\boldsymbol{\xi})\right) \\ T_s\left(S_{iy}(\boldsymbol{\xi};\mathbf{x}), \hat{\mathbf{n}}(\boldsymbol{\xi})\right) & T_d\left(S_{iy}(\boldsymbol{\xi};\mathbf{x}), \hat{\mathbf{n}}(\boldsymbol{\xi})\right) \\ T_s\left(S_{iz}(\boldsymbol{\xi};\mathbf{x}), \hat{\mathbf{n}}(\boldsymbol{\xi})\right) & T_d\left(S_{iz}(\boldsymbol{\xi};\mathbf{x}), \hat{\mathbf{n}}(\boldsymbol{\xi})\right) \end{bmatrix} \begin{bmatrix} d_s(\boldsymbol{\xi}) \\ d_d(\boldsymbol{\xi}) \end{bmatrix} dS(\boldsymbol{\xi}),$$

or in a more compact vector notation as

$$\mathbf{u}(\mathbf{x}) = \int_\Gamma \mathbf{T}(\boldsymbol{\xi};\mathbf{x})\, \mathbf{d}(\boldsymbol{\xi})\, dS(\boldsymbol{\xi}).$$

We assume to know the fault's geometry, which is discretized in M subfaults, with coordinates $\{\boldsymbol{\xi}^1, \boldsymbol{\xi}^2, \ldots, \boldsymbol{\xi}^M\}$, such that the integral can be approximated as

$$\mathbf{u}(\mathbf{x}) \simeq \sum_{i=1}^{M} A^i \mathbf{T}\left(\boldsymbol{\xi}^i; \mathbf{x}\right) \mathbf{d}\left(\boldsymbol{\xi}^i\right),$$

where A^i is the i-subfault area. Finally, if we want to compute the displacement for N receivers, we can order the displacements in a single vector such that the entire computation is reduced to a simple matrix-vector product as

$$\begin{bmatrix} \mathbf{u}(\mathbf{x}^1) \\ \mathbf{u}(\mathbf{x}^2) \\ \vdots \\ \mathbf{u}(\mathbf{x}^N) \end{bmatrix} = \begin{bmatrix} A^1\mathbf{T}\left(\boldsymbol{\xi}^1; \mathbf{x}^1\right) & A^2\mathbf{T}\left(\boldsymbol{\xi}^2; \mathbf{x}^1\right) & \cdots & A^M\mathbf{T}\left(\boldsymbol{\xi}^M; \mathbf{x}^1\right) \\ A^1\mathbf{T}\left(\boldsymbol{\xi}^1; \mathbf{x}^2\right) & A^2\mathbf{T}\left(\boldsymbol{\xi}^2; \mathbf{x}^2\right) & \cdots & A^M\mathbf{T}\left(\boldsymbol{\xi}^M; \mathbf{x}^2\right) \\ \vdots & \vdots & \ddots & \vdots \\ A^1\mathbf{T}\left(\boldsymbol{\xi}^1; \mathbf{x}^N\right) & A^2\mathbf{T}\left(\boldsymbol{\xi}^2; \mathbf{x}^N\right) & \cdots & A^M\mathbf{T}\left(\boldsymbol{\xi}^M; \mathbf{x}^N\right) \end{bmatrix} \begin{bmatrix} \mathbf{d}\left(\boldsymbol{\xi}^1\right) \\ \mathbf{d}\left(\boldsymbol{\xi}^2\right) \\ \vdots \\ \mathbf{d}\left(\boldsymbol{\xi}^M\right) \end{bmatrix},$$

or more compactly as

$$\mathbf{U} = \mathbf{X}\mathbf{D}, \tag{4.2}$$

where $\mathbf{U} \in \mathbb{R}^{3N}$ is the vector of surface displacements, North, East and Vertical for the N stations stored in a single ordered vector, $\mathbf{X} \in \mathbb{R}^{3N \times 2M}$ is the discretized FM operator built with the subfault tractions and areas, and $\mathbf{D} \in \mathbb{R}^{2M}$ is the unknown vector of sub-fault slips, along dip and along strike for each of the M sub-faults stored in a single ordered vector.

4.2.2 Data likelihood

The IP consists of recovering the slip at each subfault, of a known interface, that produces displacements observed at geodetic stations. Due to the linearity of the FM in Eq. (4.2), we solve the Bayesian inversion as a multiple linear regression model with constraints on the coefficients. We use a simple representation of observation and modeling errors by assuming a Gaussian multiple linear model of the form

$$\mathbf{Y} = \mathbf{X}\mathbf{D} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{\varepsilon}$ follows a Gaussian distribution, $\boldsymbol{\varepsilon} \sim N_{3N}(\mathbf{0}, \boldsymbol{\Sigma})$, and $\boldsymbol{\Sigma} = \mathbf{I} \otimes \gamma$ is a known covariance matrix associated with the data errors, with $\gamma = \text{diag}\left(\left[\sigma_x^2, \sigma_y^2, \sigma_z^2\right]\right)$ are the North, East, and Vertical standard deviations of the measurements process, $\mathbf{I}$ is an identity matrix of order N, and $\otimes$ denote the Kronecker product. That is,

$$\mathbf{Y}|\mathbf{D} \sim N_{3N}(\mathbf{X}\mathbf{D}, \boldsymbol{\Sigma}), \tag{4.3}$$

where $\mathbf{Y} \in \mathbb{R}^{3N}$ are the observed displacements (3 components for each station) at the N geodetic stations stored in a single ordered vector, as in Eq. (4.2). Therefore, the likelihood is given by

$$\pi\left(\mathbf{Y}|\mathbf{D}\right) = \left(2\pi\right)^{-3N/2} \left|\mathbf{A}\right|^{1/2} \exp\left\{-\frac{1}{2}\left(\mathbf{Y}-\mathbf{XD}\right)^{T}\mathbf{A}\left(\mathbf{Y}-\mathbf{XD}\right)\right\}, \tag{4.4}$$

where $\mathbf{A} = \boldsymbol{\Sigma}^{-1}$ is the precision matrix.

4.2.3 Prior elicitation

Bayesian formulation of IPs requires that we specify a prior distribution for each model parameter. Proposal of the prior density is an essential step of Bayesian analyses and is often the most challenging and critical part of the approach. Usually, the major problem while proposing an adequate prior density lies in the nature of the prior information. The prior specification is less critical for large sample sizes since the likelihood typically dominates the posterior distribution. The prior distribution plays a much more crucial role in small sample sizes because the posterior distribution represents a compromise of the prior knowledge and the observed evidence. On the one hand, for the 2006 Guerrero SSE, there are only 45 observations ($N = 15$ GPS stations) and 1178 parameters ($M = 589$ subfaults) to estimate (for the proposed model). On the other hand, the problem is so ill-posed that the inversion becomes useless unless prior knowledge is put into the problem in terms of at least simple restrictions on the possible solutions for $\mathbf{D}$. Thus, defining an adequate prior distribution is crucial for the inversion within the framework of a formal Bayesian approach. We use a Gaussian process prior for the joint PDF of the fault slip displacements $\mathbf{D}$, to be explained next.

Gaussian process priors

In statistics, a Gaussian process (GP) is a stochastic process (a collection of random variables indexed by time or space), such that every finite collection of those random variables has a multivariate normal distribution. The most commonly used probability densities in statistical IPs are undoubtedly Gaussian since they are easy to construct. However, they form a much more versatile class of densities than is usually believed (Kaipio and Somersalo 2006).

For the slip vector $\mathbf{D} = (d_s^1, d_d^1, d_s^2, d_d^2, \ldots, d_s^M, d_d^M)^T$, we consider a GP prior distribution, that is, $\mathbf{D} \sim N\left(\mathbf{0}, \boldsymbol{\Sigma}_0\right)$, but with truncated support, along strike $d_s^i \in (a_s, b_s)$ and along dip $d_d^i \in (a_d, b_d)$, $i = 1, \ldots, M$, where $\boldsymbol{\Sigma}_0 = \sigma_\beta^2 \mathbf{A}_0^{-1}$ denote the covariance matrix, with $\mathbf{A}_0 = \boldsymbol{\beta}\mathbf{W}\mathbf{C}^{-1}\mathbf{W}\boldsymbol{\beta}$ and σ_β^2 is an unknown scale factor that characterizes the magnitude of $\mathbf{D}$ (see Appendix B.1 with further considerations on how to establish the hyperparameter σ_β), and $\boldsymbol{\beta} = \mathbf{I} \otimes \underline{\beta}$, with $\underline{\beta} = \text{diag}\left([\beta_s, \beta_d]\right)$, where different precisions, β_s and β_d, are considered for the

along strike and along dip components, respectively. We consider an along dip variance five times greater than the along strike variance(i.e., $\beta_s = 1$ and $\beta_d = 1/5$); since we expect most of the slip along the opposite of the subduction direction. The matrix $\mathbf{W}$ of weights were included in the inversion scheme to penalize movements at depths greater than $z_{lim} = 50$ km

$$\mathbf{W}(i,j) = \begin{cases} 1 + 0.5\,(\text{depth}\,(i,j) - z_{lim})\,/1e3 & \text{depth}\,(i,j) > z_{lim} \\ 1 & \text{depth}\,(i,j) \leq z_{lim}. \end{cases} \tag{4.5}$$

The correlation matrix $\mathbf{C}$ is used to introduce correlation between parameters of nearby subfaults, and it is constructed using the Matérn covariance function, explained in Sect. 4.2.3. Thus, the prior density is

$$\pi(\mathbf{D}) = \frac{1}{Z_{prior}} \exp\left\{-\frac{1}{2\sigma_\beta^2}\mathbf{D}^T\mathbf{A}_0\mathbf{D}\right\} \mathbb{1}_{(\mathbf{a},\mathbf{b})}(\mathbf{D}), \tag{4.6}$$

where $\mathbb{1}_{(\mathbf{a},\mathbf{b})}$ is the indicator function, $\mathbf{a} = \mathbf{1}_M \otimes [a_s, a_d]$, $\mathbf{b} = \mathbf{1}_M \otimes [b_s, b_d]$, with $\mathbf{1}_M$ the all-ones vector of length M, and Z_{prior} is an unknown normalization constant of this MTN distribution. The constraints $a_s \leq d_s^i \leq b_s$ and $a_d \leq d_d^i \leq b_d$, $i = 1, \ldots, M$, imposed on $\mathbf{D}$, are based on available information on the physical processes being modeled. If we assume that the coupling has been removed from the GPS data, then we should only consider the displacement due to an SSE. However, we allow some negative slip ($a_d < 0$) for the dip component since, on one hand, we expect that there will be subfaults where there is no slip[1], on the other hand, the coupling removal is not precise. Together with the constraints on the support, the density function Eq. (4.6) is indeed a MTN distribution.

Matérn covariance

The Matérn covariance (Minasny and McBratney 2005) is a covariance function widely used in spatial statistics to define the covariance between measurements made at two points separated by d distance units.

In a GP, the essential ingredient is the covariance function, and this is used to introduce a correlation between nearby points (i.e., spatial smoothing). To construct the correlation matrix $\mathbf{C}$, we use the most simplified form of the Matérn covariance function corresponding to C^1 function (i.e., the space of functions that admit derivatives of first order). In the discretized

[1] If a random variable is positive, then its expected value is positive. In this way, if we consider $a_d > 0$, we would be forcing small slips in all subfaults.

subfaults model, the element (i, j) of $\mathbf{C}$ is given by the relation

$$\mathbf{C}\,(i,j) = \gamma^2 \left(1 + \sqrt{3}\frac{d\,(i,j)}{\lambda}\right) \exp\left\{-\sqrt{3}\frac{d\,(i,j)}{\lambda}\right\}, \tag{4.7}$$

where $d(i, j)$ is the distance between the subfault i and the subfault j, γ^2 represent the variance, and λ is the correlation length. Note that as λ increases, more coefficients of the matrix $\mathbf{C}$ become relevant (i.e., more subfaults are correlated).

Different correlation lengths, λ_s and λ_d, are considered for the along -strike and along -dip components, respectively. Also, we consider that the strike component of subfault i has zero correlation with the dip component of the subfault j. It is important to mention that fault continuity signifies that a slip in the dip direction of a subfault induces a slip in the strike direction of neighboring subfaults, if they are strongly coupled. However, we do not know how to determine this dependence so that we will assume independence between the strike and the dip components. The correlation lengths, λ_s and λ_d, are chosen to minimize the Deviance Information Criterion (DIC); see Sect 4.2.6 for more details.

We consider $\gamma^2 = 1$ in Eq. (4.7) to control the variance of the slips with the scale factor σ_β^2, the weight matrix $\mathbf{W}$, and the vector $\underline{\beta} = \text{diag}\,([\beta_s, \beta_d])$, throughout the precision matrix, $\mathbf{A}_0 = \frac{1}{\sigma_\beta^2}\boldsymbol{\beta}\mathbf{W}\mathbf{C}^{-1}\mathbf{W}\boldsymbol{\beta}$, of the density function given in Eq. (4.6).

4.2.4 Posterior distribution

The likelihood function given in Eq. (4.4) and the prior distribution given in Eq. (4.6) are combined to form the posterior distribution, namely

$$\begin{aligned}
\pi\,(\mathbf{D}|\mathbf{Y}) &\propto \pi\,(\mathbf{Y}|\mathbf{D})\,\pi\,(\mathbf{D}) \\
&\propto \exp\left\{-\frac{1}{2}\,(\mathbf{Y}-\mathbf{X}\mathbf{D})^T\,\mathbf{A}\,(\mathbf{Y}-\mathbf{X}\mathbf{D})\right\} \exp\left\{-\frac{1}{2\sigma_\beta^2}\mathbf{D}^T\mathbf{A}_0\mathbf{D}\right\} \mathbb{1}_{(\mathbf{a},\mathbf{b})}(\mathbf{D}) \\
&\propto \exp\left\{-\frac{1}{2}\left[\mathbf{D}^T\mathbf{X}^T\mathbf{A}\mathbf{X}\mathbf{D} - 2\mathbf{Y}^T\mathbf{A}\mathbf{X}\mathbf{D} + \frac{1}{\sigma_\beta^2}\mathbf{D}^T\mathbf{A}_0\mathbf{D}\right]\right\} \mathbb{1}_{(\mathbf{a},\mathbf{b})}(\mathbf{D}) \\
&= \exp\left\{-\frac{1}{2}\left[\mathbf{D}^T\left(\mathbf{X}^T\mathbf{A}\mathbf{X} + \frac{1}{\sigma_\beta^2}\mathbf{A}_0\right)\mathbf{D} - 2\mathbf{D}^T\mathbf{X}^T\mathbf{A}\mathbf{Y}\right]\right\} \mathbb{1}_{(\mathbf{a},\mathbf{b})}(\mathbf{D}) \\
&= \exp\left\{-\frac{1}{2}\left[\mathbf{D}^T\mathbf{A}_p\mathbf{D} - 2\mathbf{D}^T\mathbf{A}_p\mathbf{A}_p^{-1}\mathbf{X}^T\mathbf{A}\mathbf{Y}\right]\right\} \mathbb{1}_{(\mathbf{a},\mathbf{b})}(\mathbf{D}) \\
&= \exp\left\{-\frac{1}{2}\left[\left(\mathbf{D}-\boldsymbol{\mu}_p\right)^T\,\mathbf{A}_p\,\left(\mathbf{D}-\boldsymbol{\mu}_p\right)\right]\right\} \mathbb{1}_{(\mathbf{a},\mathbf{b})}(\mathbf{D}),
\end{aligned}$$

where $\mathbf{A}_p = \mathbf{X}^T\mathbf{A}\mathbf{X} + \frac{1}{\sigma_\beta^2}\mathbf{A}_0$, $\mathbf{A}_0 = \boldsymbol{\beta}\mathbf{W}\mathbf{C}^{-1}\mathbf{W}\boldsymbol{\beta}$ and $\boldsymbol{\mu}_p = \mathbf{A}_p^{-1}\mathbf{X}^T\mathbf{A}\mathbf{Y}$. Thus,

$$\pi\left(\mathbf{D}|\mathbf{Y}\right) = \frac{1}{Z_{post}} \exp\left\{-\frac{1}{2}\left[\left(\mathbf{D}-\boldsymbol{\mu}_p\right)^T \mathbf{A}_p \left(\mathbf{D}-\boldsymbol{\mu}_p\right)\right]\right\} \mathbb{1}_{(\mathbf{a},\mathbf{b})}(\mathbf{D}),$$

where Z_{post} is an unknown normalization constant. Therefore, $\mathbf{D}|\mathbf{Y}$ has a MTN distribution.

To obtain information from $\mathbf{D}|\mathbf{Y}$ one needs to calculate relevant posterior statistics, e.g., marginals of the subfaults movements, expected values, quantifying the uncertainty of a function of $\mathbf{D}$, etc. This is not trivial and accordingly we propose, as already mentioned, to use a MCMC sampler. In the next section, we propose an Algorithm to simulate from the MTN distribution.

4.2.5 Posterior exploration and MCMC

The Gibbs sampler (Gelfand and Smith 1990) is an MCMC algorithm that systematically or randomly simulates conditional distributions on a set of directions. A general case of the Gibbs sampler is the Optimal Direction Gibbs sampling, which chooses an arbitrary direction $\boldsymbol{e} \in \mathbb{R}^n$ such that $\|\boldsymbol{e}\| = 1$, and samples from the conditional distribution along such direction[2]. This can be written as,

$$\mathbf{X}^{(t+1)} = \mathbf{X}^{(t)} + r\mathbf{e},$$

where the length $r \in \mathbb{R}$ has distribution proportional to $\pi(\mathbf{x}^{(t)} + r\mathbf{e})$ (Liu 2008).

The authors in Christen et al. (2017) propose the use of the mutual information as a measure of dependence between two consecutive iterations of the Gibbs sampler. The mutual information of two random variables $\mathbf{X}$ and $\mathbf{Y}$, $I(\mathbf{X},\mathbf{Y})$, is the Kullback-Leibler divergence between the joint model $f_{\mathbf{Y},\mathbf{X}}$ and the independent alternative $f_{\mathbf{Y}}\left(\mathbf{y}\right) f_{\mathbf{X}}\left(\mathbf{x}\right)$, namely,

$$I\left(\mathbf{Y},\mathbf{X}\right) = \int\int f_{\mathbf{Y},\mathbf{X}}\left(\mathbf{y},\mathbf{x}\right) \log \frac{f_{\mathbf{Y},\mathbf{X}}\left(\mathbf{y},\mathbf{x}\right)}{f_{\mathbf{Y}}\left(\mathbf{y}\right) f_{\mathbf{X}}\left(\mathbf{x}\right)}\, d\mathbf{x}\, d\mathbf{y}.$$

From the properties inherited from the Kullback-Leibler divergence $I \geq 0$ and, from the Jensen inquality it is easy to prove that $I = 0$ if and only if $f_{\mathbf{Y},\mathbf{X}} = f_{\mathbf{Y}}\left(\mathbf{y}\right) f_{\mathbf{X}}\left(\mathbf{x}\right)$, i e., if and only if $\mathbf{X}$ and $\mathbf{Y}$ are independent.

From mutual information, the authors in Christen et al. (2017) explore a criterion of optimality for the Direction Gibbs algorithm. This criterion consists in minimizing the mutual information between two consecutive steps, $\mathbf{X}^{(t)}$ and $\mathbf{X}^{(t+1)}$, of the Markov chain generated by the algorithm. They also propose, in a heuristic way, a direction distribution for the case where the target distribution is the MTN distribution. They take the directions, $\mathbf{e}$, as the eigenvectors of the precision matrix $\mathbf{A}$, so $\mathbf{e} = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$. The i-th direction will be selected with

[2] If we choose the directions systematically or randomly from the canonical base, we obtain the standard Gibbs sampler or the Random Scan Gibbs, respectively.

probability proportional to λ_i^{-b}, where λ_i is the eigenvalue corresponding to the i-th eigenvector, $i = 1, 2, \ldots, n$, and b is a random variable with a $Beta$ distribution. Then, the probability of selecting the i-th direction is given by

$$h_1(\mathbf{e}_i) = \lambda_i^{-b}/k_1,$$

where $k_1 = \sum_{i=1}^{n} \lambda_i^{-b}$. See Christen et al. (2017) for more details.

Now, let $\mathbf{X} = \mathbf{X}^{(t)}$ and $\mathbf{Y} = \mathbf{X}^{(t)} + r\mathbf{e}$ be two consecutive steps, and denote by X_i, Y_i, $i = 1, \ldots, n$, the elements of $\mathbf{X}$ and $\mathbf{Y}$, respectively. In Montesinos-López and Christen (2016), the authors propose the Mutual Information as dependence measure, but now no longer on the complete vectors $\mathbf{X}$ and $\mathbf{Y}$, instead, they obtain it with Y_i and the full vector $\mathbf{X}$. They call this the *marginal mutual information*

$$I_{\mathbf{e}}(Y_i, \mathbf{X}) := \int \int f_{Y_i,\mathbf{X}}(y, \mathbf{x}) \log \frac{f_{Y_i,\mathbf{X}}(y, \mathbf{x})}{f_{Y_i}(y) f_{\mathbf{X}}(\mathbf{x})} \, d\mathbf{x} \, dy.$$

The idea is to choose directions for which $I_{\mathbf{e}}(Y_i, \mathbf{X})$, $\forall i = 1, \ldots, n$, is minimized. In this way, the dependency of each entry of the new generated vector $\mathbf{Y}$ with the current state $\mathbf{X}$ is reduced.

Suppose we have a multivariate normal distribution, with precision matrix $\mathbf{A}_{n\times n}$ and mean vector $\boldsymbol{\mu}_{n\times 1}$, but with truncated support, $x_i \in (a_i, b_i)$, $-\infty \leq a_i < b_i \leq \infty$, $i = 1, \ldots, n$. The probability density function of this MTN can be written as

$$\pi(\mathbf{x}; \boldsymbol{\mu}, \mathbf{A}, \mathbf{a}, \mathbf{b}) = \frac{\exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{A} (\mathbf{x} - \boldsymbol{\mu})\right\}}{\int_{\mathbf{a}}^{\mathbf{b}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{A} (\mathbf{x} - \boldsymbol{\mu})\right\} d\mathbf{x}} \mathbb{1}_{(\mathbf{a},\mathbf{b})}(\mathbf{x}).$$

To generate samples from the MTN distribution, in Montesinos-López and Christen (2016) the authors take the directions $\mathbf{e}$ as the standardized columns of the covariance matrix $\mathbf{A}^{-1}$, so $\mathbf{e} = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$. The $i-th$ direction will be selected with probability $(h_2(\mathbf{e}_i))$ proportional to I_i^{-1}, with

$$I_i := \sum_{j=1}^{n} I_{\mathbf{e}_i}(Y_j, \mathbf{X}) = -\frac{1}{2} \sum_{j=1}^{n} \log\left(\rho_{ij}^2\right),$$

where ρ_{ij} is the correlation between the variables Z_i and Z_j, with $\mathbf{Z} \sim \pi$. Then, the probability of selecting the i-th direction is given by

$$h_2(\mathbf{e}_i) = I_i^{-1}/k_2,$$

where $k_2 = \sum_{i=1}^{n} I_i^{-1}$. See Montesinos-López and Christen (2016) for more details of its derivation. Thus, they give more weights to the directions that make the I_i's small.

In this article, we slightly modify the probability of selecting the i-th direction as follows,

$$h_3\left(\mathbf{e}_i\right) = I_i^{-b}/k_3,$$

where $k_3 = \sum_{i=1}^{n} I_i^{-b}$ and b is a random variable with $Beta$ distribution. This modification slightly reduces the time to obtain a pseudo-independent sample (by 3%) and also reduces the initial burn-in length.

We will refer to these algorithms as ODG1 or ODG3 when the direction distribution used is h_1 or h_3, respectively. This paper combines the ODG1 and ODG3 algorithms in a hybrid Gibbs. The ODG1 algorithm provides a faster convergence to the target distribution reducing the burn-in, while with the ODG3 algorithm the chain's correlations is reduced, saving 60% of the time to obtain a pseudo-independent sample compared to the ODG1 algorithm. By combining both algorithms, we reduce the chain's correlations, and the support is better explored. The resulting algorithm, which we call the HODG algorithm, is described in Algorithm 3.

4.2.6 Pragmatic approach to correlation lengths selection using the DIC

Note that if the full conditional distribution for the slip vector $\mathbf{D}$ had a known tractable form, then full MCMC sampling could be performed on the low-dimensional marginal posterior distribution over hyperparameters (Norton et al. 2018; Fox and Norton 2016), namely λ_s and λ_d. In Fox and Norton (2016), the authors propose sampling the low-dimensional marginal posterior distribution over hyperparameters in a linear-Gaussian inverse problem. Then, the full conditional for the latent variable is sampled using only the approximately independent samples of the hyperparameters obtained from a MCMC run over its marginal distribution. However, in our case, sampling from the marginal distribution of λ_s and λ_d can be challenging since the way these parameters interact in the correlation matrix C makes it impossible to factorize them. In addition, marginalizing these hyperparameters would require evaluating the normalization constant of the MTN posterior, which requires a very high computational cost (Genz and Bretz 2009; Nocquet 2018).

Setting a prior distribution for λ_s and λ_d and including these parameters in the posterior would be the theoretical correct way to proceed, but it is computationally unfeasible, as just noted. Instead, we follow a more pragmatic approach. Using prior information on the values of these hyperparameters, we proposed a grid search applying the DIC to select appropriate correlation lengths.

The correlation length selection for the MTN prior model must be done carefully since each correlation length defines a different matrix $\mathbf{C}$, hence different results. Several criteria have been proposed to select between competing models. In the maximum-likelihood framework, the most well-known criterion for model comparison is the Akaike Information Criterion (AIC), which

Algorithm 3: HODG: Multivariate Truncated Normal

input : The means vector $\boldsymbol{\mu}_{n\times 1}$, the precision matrix $\mathbf{A}_{n\times n}$, the support $(\mathbf{a}, \mathbf{b})$, a initial value $\mathbf{X}^{(0)}$, and the number of simulations M.

output: A sample of size M from $\mathbf{X} \sim MTN(\boldsymbol{\mu}_{n\times 1}, \mathbf{A}_{n\times n}, \mathbf{a}, \mathbf{b})$, with the truncated support $x_i \in (a_i, b_i)$, $i = 1, \ldots, n$.

Step 1. Compute the eigenvectors and eigenvalue of the precision matrix $\mathbf{A}$, $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ and $\lambda_1, \lambda_2, \ldots, \lambda_n$, respectively;

Step 2. Normalize the columns of the covariance matrix $\Sigma = \mathbf{A}^{-1}$, these will be $\{\mathbf{e}_1^c, \mathbf{e}_2^c, \ldots, \mathbf{e}_n^c\}$;

Step 3. Compute the correlation matrix (ρ) corresponding to Σ;

Step 4. Compute the weights I_i,

$$I_i = -\frac{1}{2}\sum_{j=1}^{n} \log\left(\rho_{ij}^2\right),$$

where ρ_{ij} is the correlation between the variables X_i and X_j;

for $t \leftarrow 1$ **to** M **do**

Set $\mathbf{x} = \mathbf{X}^{(t-1)}$;

Simulate $b \sim B(2., 9)$, where B is the Beta distribution;

Simulate $p \sim U(0, 1)$, where U is the uniform distribution;

if $p < 0.5$ **then**

$h(\mathbf{e}_i) = \lambda_i^{-b}/k_1$, where $k_1 = \sum_{i=1}^{n} \lambda_i^{-b}$;

else

$h(\mathbf{e}_i^c) = I_i^{-b}/k_2$, where $k_2 = \sum_{i=1}^{n} I_i^{-b}$;

Step 5. Propose a direction $\mathbf{e}$ from the direction distribution $h(\cdot)$;

Step 6. Simulate $r \sim TN(\mu_r, \tau_r, c, d)$, where TN is the univariate truncated normal distribution, $\mu_r = -\frac{\mathbf{e}^T\mathbf{A}(\mathbf{x}-\boldsymbol{\mu})}{\mathbf{e}^T\mathbf{A}\mathbf{e}}$ is the mean, $\tau_r = \mathbf{e}^T\mathbf{A}\mathbf{e}$ is the precision, and

$$c = \max_{i\in\{1,\ldots,n\}} \left(\left\{ \frac{a_i - x_i}{e_i} : e_i > 0 \right\} \cup \left\{ \frac{b_i - x_i}{e_i} : e_i < 0 \right\} \right),$$

$$d = \min_{i\in\{1,\ldots,n\}} \left(\left\{ \frac{a_i - x_i}{e_i} : e_i < 0 \right\} \cup \left\{ \frac{b_i - x_i}{e_i} : e_i > 0 \right\} \right)$$

Step 7. Set $\mathbf{X}^{(t)} = \mathbf{x} + r\mathbf{e}$;

involves the marginal likelihood (Akaike 1974). The Deviance Information Criterion (DIC) has been proposed as Bayesian alternative to the AIC (Spiegelhalter et al. 2002) to select the model that better fits the data between a pool of competing models. The DIC is particularly useful in Bayesian model selection problems where the model's posterior distributions have been obtained by MCMC simulation. The DIC's advantages is that it reduces each model to a single number summary and that the models to be compared do not need to be nested.

For a model with parameters θ and data y, the DIC is calculated as

$$DIC = \overline{D(\theta)} + p_D = D(\bar{\theta}) + 2p_D.$$

where $D(\theta) := -2\log(likelihood) = -2\log(p(y|\theta))$ is called the model deviance, $\overline{D(\theta)} = E[D(\theta)|y]$ is the posterior expected deviance, $\bar{\theta} = E[\theta|y]$ is the posterior mean, and $p_D = \overline{D(\theta)} - D(\bar{\theta})$ is called the effective number of parameters. In our case, the expectations may be easily calculated using the MCMC sample. We then proceed to calculate the DIC for all models of interest and choose the one with the smallest DIC value. For further discussion of the DIC see Spiegelhalter et al. (2002). Note that –up to our knowledge– DIC has not been applied for model comparison in geophysical inverse modeling.

4.3 Results

4.3.1 Synthetic example

To verify how much fault movements and correlation lengths are recoverable, using the actual data collecting design we have for the real data presented in Sec.4.4, we proceed as follows. We generate a synthetic data set based on the same fault geometry and geodetic stations configuration as for the 2006 Guerrero SSE. For this, we assume a priori that the slips ($\mathbf{D}$) have a MTN with zero mean vector (i.e., in average we consider that there is no slipping), restricting the support according to the information we have on the GGap. That is, $\mathbf{D} \sim N\left(\mathbf{0}, \frac{1}{\sigma_\beta^2}\mathbf{A}_0\right)$, subject to $-.0804 \leq d_d^i \leq 0.4$, $-0.1 \leq d_s^i \leq 0.1$, $i = 1, 2, \ldots, M$, where $\mathbf{A}_0 = \boldsymbol{\beta}\mathbf{W}\mathbf{C}^{-1}\mathbf{W}\boldsymbol{\beta}$, $\mathbf{C}$ is the correlation matrix given by Eq. (4.7), and $\mathbf{W}$ is the matrix of weights computed with Eq. (4.5). For the remaining parameters we consider $\lambda_s = 40$, $\lambda_d = 45$, $\sigma_\beta^2 = 0.0002$, $\beta_s = 1$ and $\beta_d = 0.2$. To generate the synthetic example, we fix a slip in a subfault $\mathbf{D}_i$ (approximately in the same place where the maximum slip is suspected in the real 2006 GGap SSE) and the rest of the subfault slips are simulated from the conditional distribution $\mathbf{D}_{-i}|\mathbf{D}_i = d_i$ (also a MTN).

Once the slip vector $\mathbf{D}$ has been generated, we solve the forward problem ($\mathbf{U} = \mathbf{X}\mathbf{D}$) adding Gaussian noise to obtain our synthetic observations. Figure 4.2 (a) shows the true slip movement $\mathbf{D}$ and the synthetic measurements. Following this strategy of simulation of the synthetic data and considering the statistical model given in Sec. 4.2.1, we obtain simulations of the MTN distribution using the HODG algorithm. We set the hyperparameter σ_β^2 as explained in appendix B.1.

We consider different correlation lengths for λ_s and for λ_d in the a priori distributions, see Fig. 4.1 and 4.2. To choose the correlation lengths, we computed the DIC in a grid search along the hyperparameter space $\lambda_s : [35, 45]$ km $\times$ $\lambda_d : [45, 60]$ km, see Fig. 4.1. The correlation lengths of $\lambda_s = 40$ km and $\lambda_d = 55$ km were finally selected.

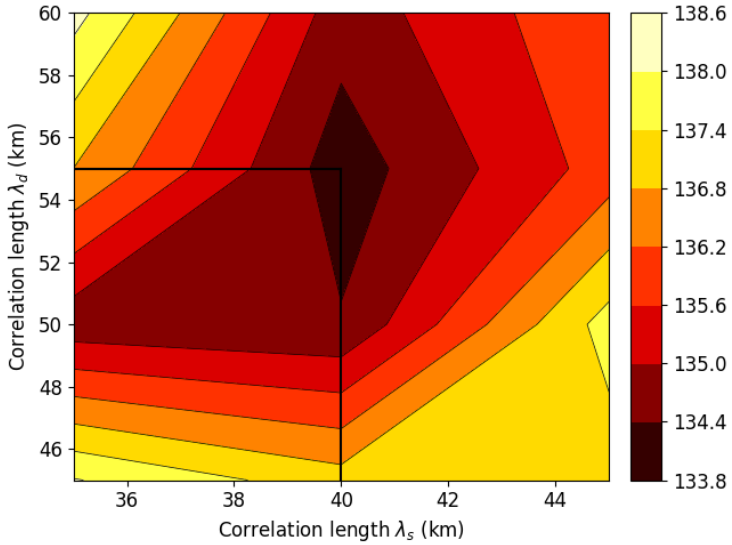


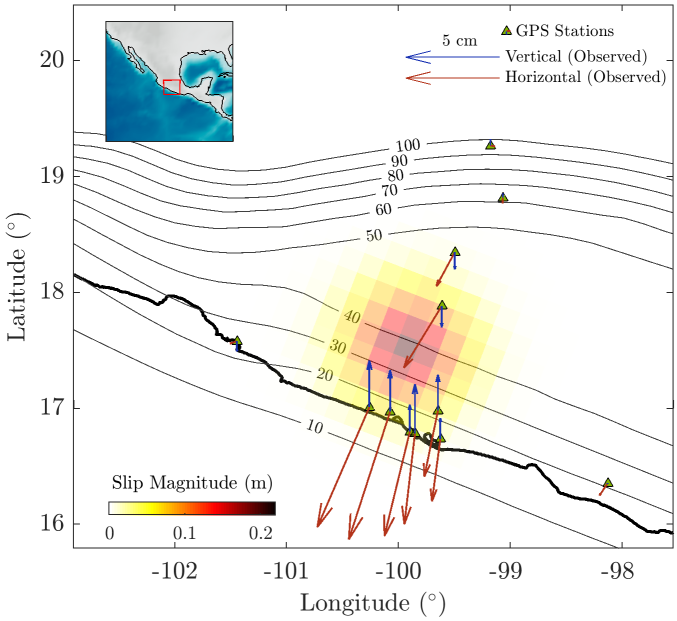
Figure 4.1: The DIC's obtained with the MCMC output considering synthetic data, for different combinations of correlation lengths. The correlation lengths of $\lambda_s = 40$ km and $\lambda_d = 55$ km were finally selected, although the other DIC's are similar. The actual correlation lengths used for simulation are $\lambda_s = 40$ and $\lambda_d = 45$. The Bayesian inversion is quite similar in all cases (see Figure 4.2), as predicted by the similarities in the DIC's. The selection of the correlation length in this scale is robust and, on the other hand, in general estimating correlation is commonly a difficult statistical problem.

4.3.2 Posterior distribution and uncertainty representations

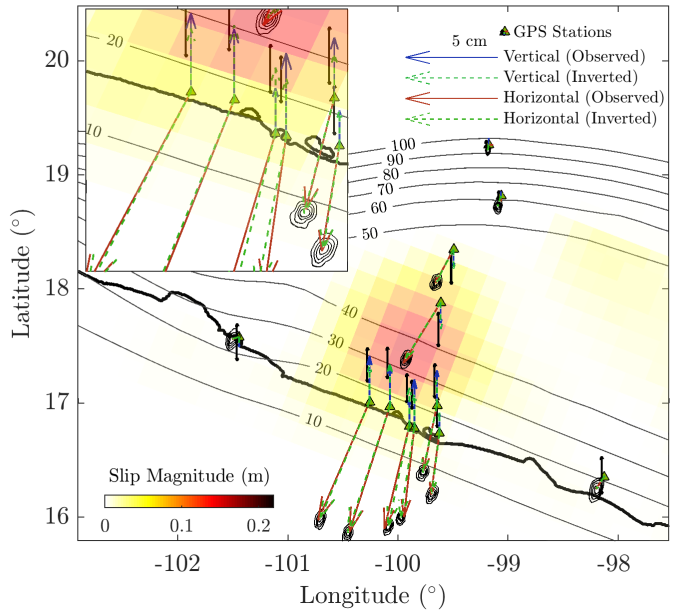
The median of the posterior samples of each subfault was plotted for the chosen correlation lengths in Fig. 4.2 (b) and Fig. 4.3 (a), with a heat map. We also plot the GPS stations locations with triangles and the their corresponding data using arrows. The black contours at the arrowheads represent the posterior error calculated for the inversion in the data. We can see that the data fit is excellent in all cases and the slip solution is almost perfect, in relation to the true slip seen in Fig. 4.2 (a).

Since we have access to the conditional posterior distribution of $\mathbf{D}$, conditioned on estimates of the covariance parameters, we can look at point estimators such as the posterior mean, the posterior median, the maximum a posteriori (MAP). However, these point estimators may be unrepresentative of the actual posterior. The mean and median may be misleading for long-tailed asymmetric PDFs, and the MAP may be unrepresentative in the presence of skewness. In this synthetic case, we obtain far better results with the median of the posterior samples, shown in Figs. 4.2 (b) and 4.3 (a). For comparisons, and to observe the crucial importance of the inclusion of correctly modeled prior information, we include a Maximum Likelihood Estimation (MLE) estimation in Fig. 4.3 (b).

An advantage of the Bayesian approach is that it does not only produce one optimal model,

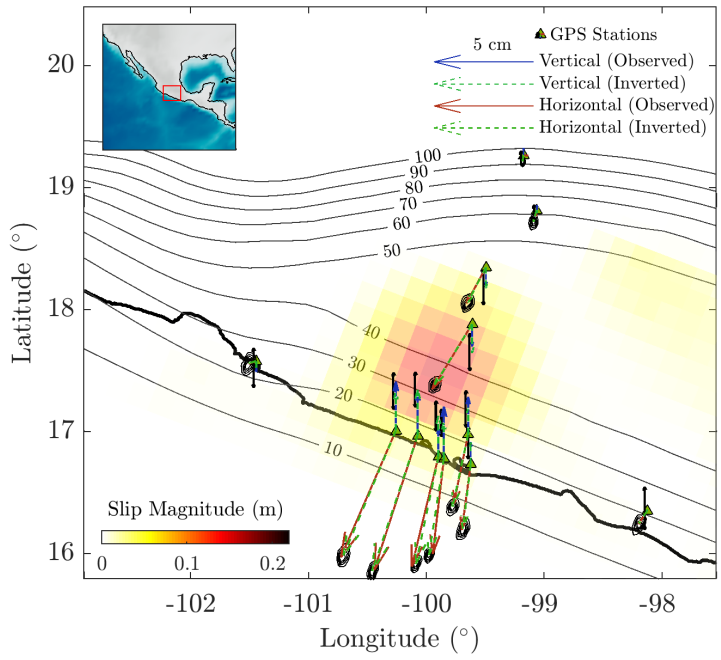


(a) True synthetic displacements

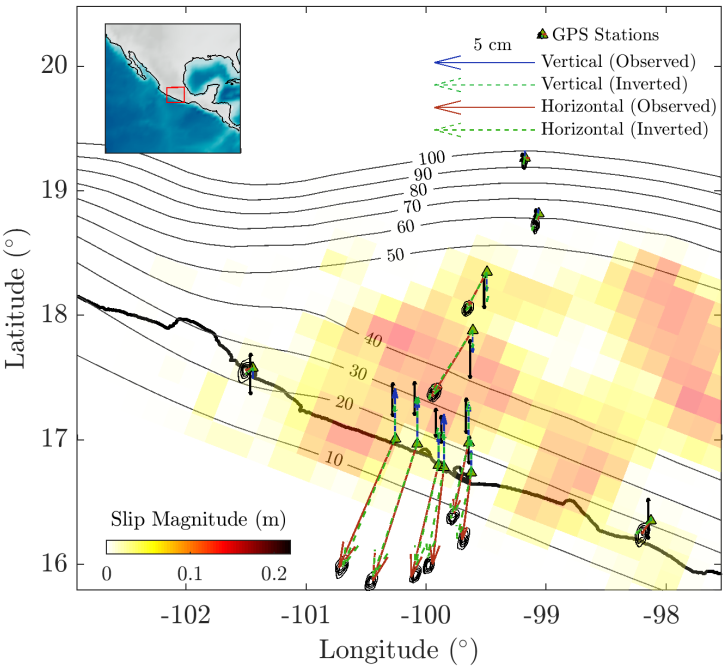


(b) Bayesian inversion: $\lambda_s = 40, \lambda_d = 45$

Figure 4.2: Slip models on the plate interface (heat colors) and the associated model surface displacement predictions (arrows): (a) True synthetic displecements and (b) the median of the posterior samples for the slip inversion using correlation lengths $\lambda_s = 40, \lambda_d = 45$. Blue-solid and red-solid arrows show the observed (synthetic) surface displacements while dashed arrows show the predictions. The black contours at the arrowheads represent the data uncertainty; see the zoomed insert in (b). Green triangles show the GPS station locations. Black lines represent the isodepth contours (in km) of the subducted oceanic slab.



(a) Bayesian inversion: $\lambda_s = 35, \lambda_d = 50$



(b) MLE

Figure 4.3: Slip models on the plate interface (heat colors) and the associated model surface displacement predictions (arrows): (a) The median of the posterior samples for the slip inversion using correlation lengths $\lambda_s = 35, \lambda_d = 50$ and (b) a MLE estimation.

but the sampling yields a large ensemble of probable models, sampled from the posterior distribution. In Figs. 4.4 (a) and (b), we represent the median and the uncertainty of the slips considering $\lambda_s = 40$ and $\lambda_d = 55$, respectively. These correlation lengths provide the lowest DIC (see Fig. 4.1).

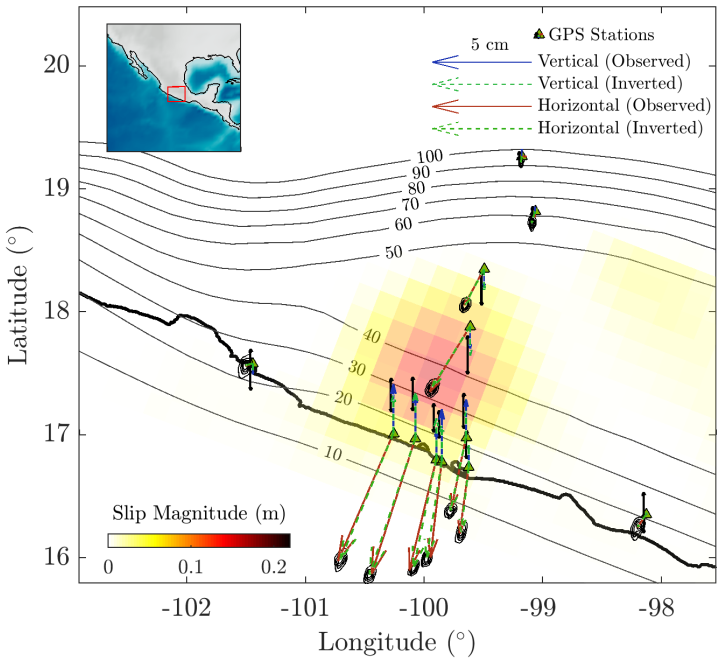
Now we turn to the question of uncertainty representation. This is not a straight forward task for a posterior distribution on a high dimensional vector field. Here we resort to the posterior coefficient of variation (CV) at each sub-fault. The CV is a statistical measure of the dispersion of a probability distribution around its mean. The CV represents the ratio of the standard deviation σ to the mean μ (CV = $\frac{\sigma}{\mu}$), showing a relative quantity of the degree of variation, independent of the scale of the variable. This metric provides a tool to compare the data dispersion between different data series. We use the MCMC simulations to estimate the CV at each subfault. In Fig. 4.4 (b) the CV is plotted to compare the posterior uncertainty in the inferred (inverted) movements in each subfault. We can see that in the areas where the largest slip was found, we have the least relative uncertainty, that is, the median is more representative. Note also that where the GPS stations are located, we have a clear decrease in uncertainty. These regions of low uncertainty are consistent with the regions with maximum restitution index computed by Tago et al. (2021) through a mobile checkerboard strategy. The map with the displacement medians along with the map with their corresponding CVs, as in Fig. 4.4, is our representation of the posterior distribution, and the UQ representation of this inversion. In the next section we use the same strategy to study the 2006 GGap SSE inversion.

4.4 Real case: 2006 Guerrero Slow Slip Event

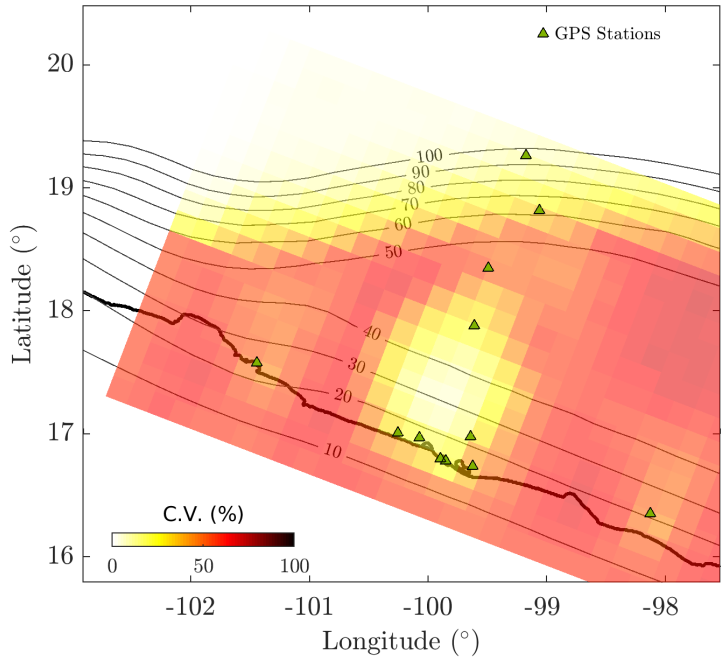
In this section we present a real data application to illustrate the performance of our approach. We study the 2006 Guerrero SSE with data collected by the Instituto de Geofísica (IGF), Universidad Nacional Autónoma de México (UNAM), and the Servicio Sismológico Nacional (SSN). In 2006, a SSE in Guerrero was recorded by $N = 15$ GPS stations. The stations are located mainly along the coast and on a transect perpendicular to the trench, between Acapulco and the north of Mexico city (Radiguet et al. 2011). We used these same locations in the synthetic analysis presented in the previous Section.

4.4.1 Observations and data preprocessing

The GPS data must be preprocessed taking into account the time-varying climate phenomena. Besides, the inter-SSE steady-state motion is subtracted to isolate the GPS data related with an SSE event, that is, the tectonic coupling is removed. For the actual GPS data in the GGap 2006 event, we used the data processed by Radiguet et al. (2011) with the their proposed standard



(a) Bayesian inversion: $\lambda_1 = 40, \lambda_2 = 55$



(b) Coefficient of variation

Figure 4.4: Slip posterior median (a) and its CV (b), resulting from the Bayesian inversion of synthetic GPS cumulative displacements for the correlation lengths $\lambda_s = 40$ and $\lambda_d = 55$, see Fig. 4.2 for more details. The CV is expressed as a percentage and quantifies the uncertainty in the inverted slip shown in (a). Higher (darker) CV values implies more uncertainty in the inferred slip.

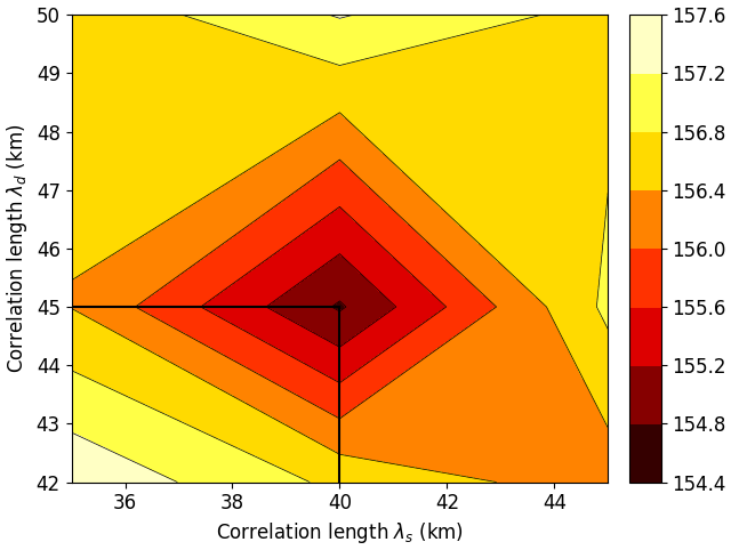


Figure 4.5: Results for the 2006 Guerrero SSE Bayesian inversions: The DICs obtained with the MCMC output, for different combinations of correlation lengths. The correlation lengths of $\lambda_s = 40$ km and $\lambda_d = 45$ km were finally selected.

deviations, $\sigma_x = 0.0021$, $\sigma_y = 0.0025$, and $\sigma_z = 0.0051$ in the north, east and vertical directions, respectively. All these quantities are measured in meters. The time window that was considered to compute the displacements was from January 2, 2006 to May 15, 2007.

We solve the Bayesian inversion as a multiple linear regression model with constraints on the coefficients, considering the statistical model (4.3) as explained in Sect. 4.2.1. For the GPS data, we took $\sigma_x^2 = 0.0021$, $\sigma_y^2 = 0.0025$, and $\sigma_z^2 = 0.0051$ as the standard deviations in the north, east and vertical directions, respectively (Radiguet et al. 2011). As in the synthetic case, the regularization parameter, σ_β^2, is obtained by minimizing Eq. (B.4) given in Appendix B.1.

To sample from the resulting MTN posterior distribution we use the HODG sampler explained in Algorithm 3. To choose the correlation lengths, we computed the DIC in a grid search along the hyperparameter space $\lambda_s : [30, 45]$ km $\times$ $\lambda_d : [42, 50]$ km, see Fig. 4.5. The correlation lengths of $\lambda_s = 40$ and $\lambda_d = 45$ were finally selected.

The median of the posterior samples for the static inversion was plotted (heat colors) in Fig. 4.6. The black contours at the arrowheads represent the data uncertainty, and the horizontal lines in the vertical component represent the quantiles $0.025, 0.5, 0.975$, respectively. The CV is plotted in Fig. 4.7.

All GPS data is well recovered by the method within the estimated uncertainty bounds. The median of the posterior show a compact region where most of the displacement took place. It is consistent with the most recent inversions, where the region of maximum slip is located from 30 to 40 km depth and with a slight updip penetration in the north-west section (Bekaert et al. 2015; Tago et al. 2021). Recent offshore observations showed that the mechanical properties

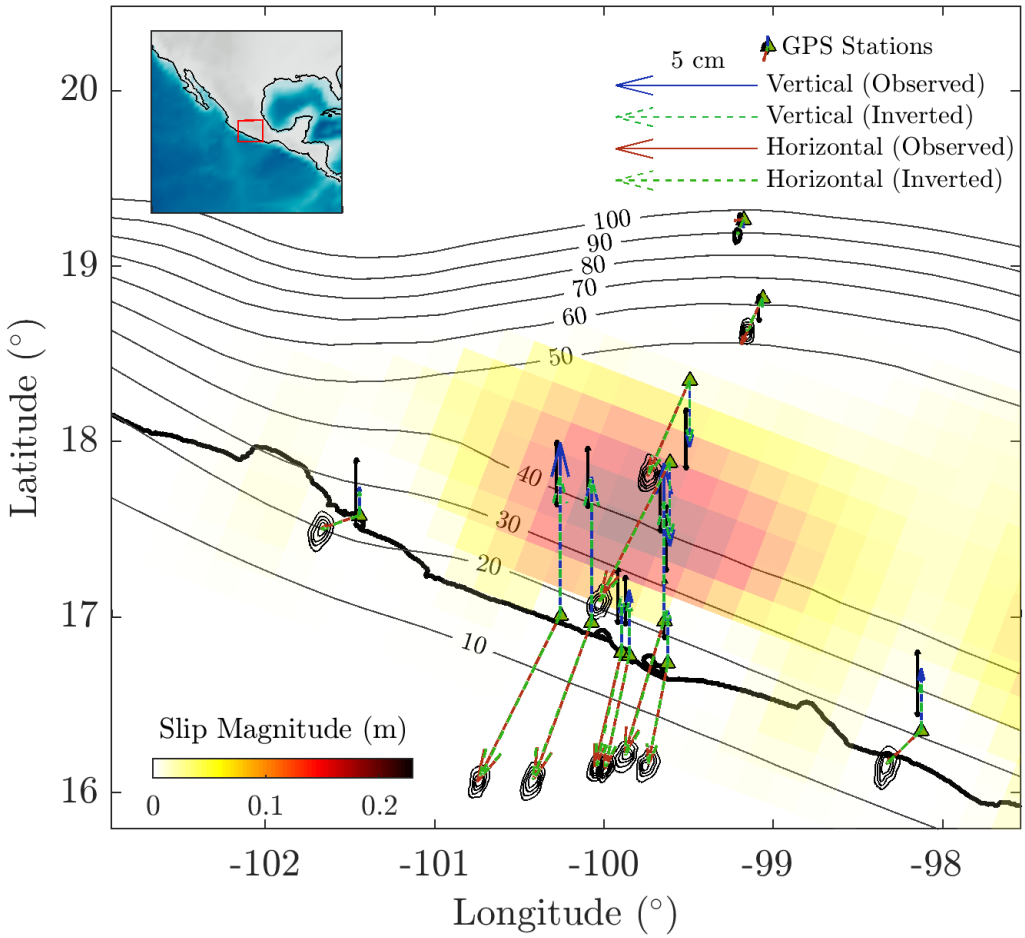


Figure 4.6: Bayesian inversion (Posterior median) of fault slips of the 2006 Guerrero SSE for correlation lengths $\lambda_s = 40, \lambda_d = 45$, following the same conventions as in Fig. 4.2.

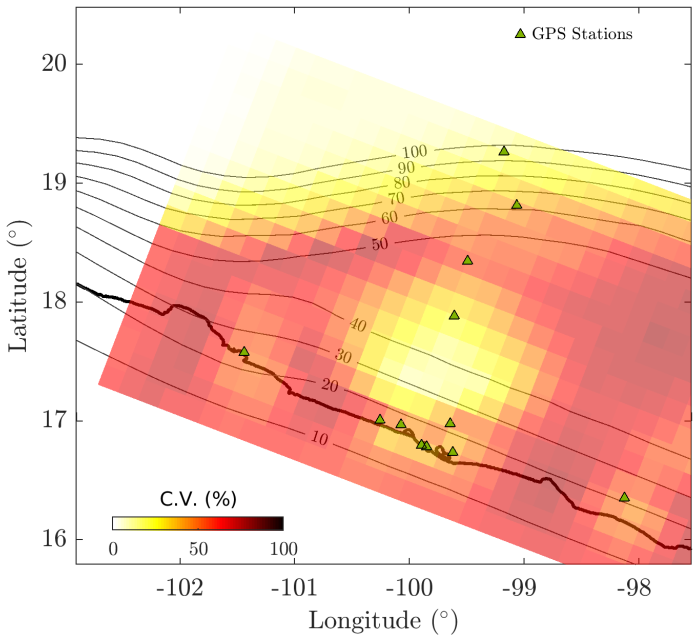
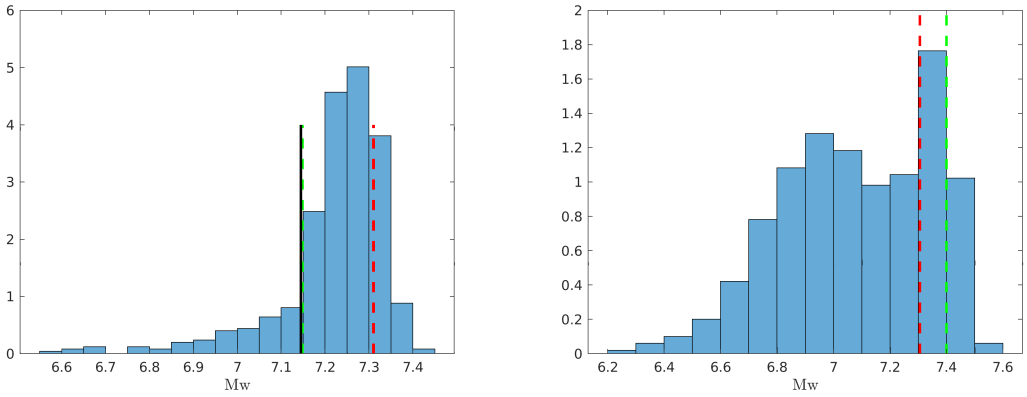


Figure 4.7: Bayesian inversion (CV, expressed as a percentage) of fault slips of the 2006 Guerrero SSE for correlation lengths $\lambda_s = 40, \lambda_d = 45$, following the same conventions as in Fig. 4.2.

in that segment of the subduction slab are different and it may explain the inferred updip slip (Plata-Martínez et al. 2021). However, since in that region the CV is high, $> 50\%$, it should be taken with caution. A better instrumentation deployment should be considered to avoid any misinterpretation. Despite the similarities with previous studies, it is important to mention that most of previous works are supported on constrained optimization framework, a different point-wise estimate than the one presented here (e.g., Radiguet et al. (2011); Tago et al. (2021)). As explained in the introduction, these estimates may be biased and the comparison with our results should be made carefully.

A novelty of our procedure is depicted in Fig. 4.6 (b) where we are able to estimate the posterior CV, as measure of the uncertainty in our solution. Uncertainty is low in the region where the fault's displacement is concentrated, and nearby coastline GPS stations. The former is a consequence of solution to the IP, the latter is expected since the GPS station illuminates the nearby faults. On the upper part of the color map, where the Cocos plate dives into the mantle, the uncertainty is also low. This is a consequence of the prior information built into our prior distribution. Specifically, by using the weight matrix $\mathbf{W}$, and represents our knowledge that the Cocos and North American plates are not coupled at such depths. Therefore, very low or no uncertainty in the displacement is to be observed in this region.



(a) Synthetic (b) 2006 Guerrero SSE

Figure 4.8: Posterior distributions of the moment magnitude M_w, along with the MAP (red) and median (green) point estimates: (a) Synthetic case fault displacement inversion presented in Fig. 4.4, the black marker represents the true M_w. (b) 2006 Guerrero SSE fault displacement inversion presented in Fig. 4.6.

4.4.2 Uncertainty Quantification of the moment magnitude

Given a particular displacements vector $\mathbf{D}$, the moment magnitude $M_w(\mathbf{D})$ is computed as

$$M_w(\mathbf{D}) = \frac{2}{3}\left(\log_{10}\frac{M_0(\mathbf{D})}{N\cdot m} - 9.1\right),$$

where $M_0(\mathbf{D}) = \mu A\mathbf{D}$ is the seismic moment in N·m, μ is the crustal rigidity in Pa and A is the surface that slipped in m^2 (Stein and Wysession 2009). We take the 1 cm slip contour as the effective SSE area, and we consider a typical crustal rigidity $\mu = 32 \cdot 10^9$ Pa.

Using the Bayesian approach has a further advantage; we can consistently produce estimates and UQ of inferred parameters. That is, the posterior distribution of $M_w(\mathbf{D})$ is well defined, as the transformation of the random vector $\mathbf{D}|\mathbf{Y}$. Moreover, since we already have a Monte Carlo sample, $\mathbf{D}^{(1)}, \mathbf{D}^{(2)}, \ldots, \mathbf{D}^{(T)}$, of the posterior $\mathbf{D}|\mathbf{Y}$, $M_w^{(i)} = M_w(\mathbf{D}^{(i)})$, $i = 1, 2, \ldots, T$, is a Monte Carlo sample from the posterior distribution of the moment magnitude. The posterior distributions for the M_w of the synthetic and 2006 Guerrero SSE examples are presented in Fig. 4.8. For comparisons, also the moment magnitude of the MAP and the median displacements are plotted in Fig. 4.8. In Fig. 4.8 (a), the synthetic case, the true M_w is plotted; indeed calculated from the true displacements used to simulate the exercise (see Fig. 4.2 (a)). Note how the posterior of M_w is skewed and would suggest larger values that what is estimated using the MAP or the median. The M_w of the median coincides nicely with the true value and both are contained in the posterior (note that for this non-linear functional, the M_w of the median displacements need not to coincide with the median of the posterior for M_w).

For the 2006 Guerrero SSE, we can extract from the posterior distribution of M_w the point estimates $M_{w_{\text{median}}} = 7.4$ and $M_{w_{\text{MAP}}} = 7.3$. The former is consistent with the value computed by Tago et al. (2021), the later with study by Bekaert et al. (2015) and both are below the $M_w = 7.5$ estimated by Radiguet et al. (2011). Clearly our point estimates are consistent with previous studies, however the skewed posterior distribution has not been showed before. From the synthetic data, it seems that the median is a better point estimate but further research must be done.

4.5 Discussion

Solutions to IPs that include UQ in geophysics are challenging problems. Computationally feasible methods such as Tikhonov regularization may introduce biases due to non-physically justified regularization terms and solutions may misrepresent part of the phenomena. Moreover, different considerations in the regularization terms may produce dissimilar solutions. Bayesian methods provide a natural alternative to explore the posterior distribution of IPs and provide formal UQ.

In this work, we developed an efficient Bayesian approach to estimate fault slip induced by a SSE. We calculated the fault slip posterior through a constrained multiple linear regression model and geodetic observations. We postulate a Gaussian model for geodetic data, and a model of the prior information as MTN prior distribution for the unknown slip. The resulting fault slip posterior PDF is also a MTN. Regarding the posterior, we proposed an efficient Optimal Direction Gibbs sampler algorithm to sample from this high-dimensional MTN. The computations can be carried out in a personal computer as opposed to other MCMC samplers that require high-performance computing. An advantage of our MCMC algorithm is that no parameter has to be adapted or tuned.

Prior elicitation is a fundamental part of the modeling process regarding the particular physical problem under study. We use the Matérn covariance function to control the subfault autocovariance and to impose physically-consistent slip restrictions (on the prior distribution). Different correlation lengths are considered in the prior distribution. As a model selection criteria, we propose the DIC to choose the correlation lengths.

In many applications the MAP estimator is chosen as a representative solution to the IP due to its computational feasibility by usual regularization schemes. Our results in the synthetic data case shows that the MAP is biased. Instead, we propose the posterior median as an alternative to fault slips's point-wise estimate. Moreover, UQ is represented by the CV. We compare variability between subfaults and show areas where we have the most certainty. Since we have Monte Carlo samples of the full conditional posterior distribution, conditioned on estimates of the covari-

ance parameters, both these quantities are readily available, in sharp contrast to regularization methods, where these quantities can not be recovered. For the 2006 Guerrero SSE, the median of the posterior distribution of the slip shows a compact slip patch where most of the slip is located from 30 to 40 km depth with a slight updip penetration in the north-west section. Both of these main characteristics are consistent with the most recent studies of Bekaert et al. (2015) and Tago et al. (2021). Besides those coincidences, through the CV we can assess the uncertainty which is lower where the most of the slip is located. With the posterior distribution of the slip, we could easily compute the posterior distribution of M_w which showed to be skewed with the median and MAP estimates also consistent with the above mentioned studies.

One limitation of our current approach is that the FM needs to be linearized in order to obtain a MTN posterior. A more general FM could also be analyzed, although adding severe computational burden and difficulties in a highly multidimensional MCMC.

The Bayesian framework allows us to also consider different representations of uncertainty. It is clear that more SSE's should be analyzed for the GGap, and elsewhere. A further improvement would be to learn parameters of the particular fault slip from multiple SSE's analyses such as the correlation length. Moreover, border effects should be formally included in the covariance matrix, by improving the covariance operator (Daon and Stadler 2018). For the moment, border effects do not seem apparent in the maps produced by the median, as seen in the examples presented here. However, we leave these ideas for future research.

Computationally efficient Bayesian methods are being developed for many IPs in geophysics. In many cases, they provide access to full posterior distributions, which provide better and more informative estimates for the solutions as well as UQ. Our proposed methodology is applied to a real data set, for the 2006 Guerrero SSE, where the objective was to recover the slip on a known interface from observations at few geodetic stations. Lastly, once the slip inversion is available, calculating the seismic moment, with formal UQ, is a simple subproduct of our methodology, a result that –up to our knowledge– is new.

CHAPTER 5

Discussion

The Bayesian UQ analysis of IPs continues to be a very challenging problem. In this thesis, we addressed two fairly broad research questions. First, what numerical error may be tolerated in the FM to obtain a correct and acceptable numerical posterior distribution? Second, how may we estimate the magnitude and distribution of slip along a fault plane with limited geodetic data? The provided answers and some of the following issues are summarized and discussed in these concluding remarks. In each Chapter, particular conclusions have already been drawn.

For the first question, a bound on the ERE, in posterior expectations for some functional, is found. It is also detailed how this bound can be used to find a discretization in the numerical method used to approximate the FM. Moreover, we proposed a numerical method of computing the *after-the-fact* error estimates of the numerical solver of a class of semilinear evolution PDEs. Our error estimation allows us to apply the result obtained previously for bounding the ERE in the respective numerical posterior expectation for Bayesian UQ problems. In all the presented examples, RE in the numerical posterior statistics was successfully bounded, i.e., the RE in the posterior expectation of parameters was smaller than the established tolerance b, this considering different levels of accuracy for both MCMC and ERE. Although the bound appears to not be quite tight, i.e., the margin between b and the RE is considerable, the point here is that there are no similar results for comparison in the UQ literature or elsewhere, to our knowledge.

For future work, we plan to extend the method used for computing the error estimates of nonlinear evolution PDEs, but some considerations regarding the solution's stability and the convergence orders most likely will need to be added.

For the second question, we developed an Bayesian approach to estimate fault slips in an SSE. We calculated the fault slip posterior through a constrained multiple linear regression model and geodetic observations. We postulate a Gaussian model for geodetic data and a MTN

prior distribution for the unknown slip. The resulting fault slip posterior PDF is also a MTN. Regarding the posterior, we proposed an efficient Optimal Direction Gibbs sampler algorithm to sample from this high-dimensional MTN.

Computationally efficient Bayesian methods are being developed for many IPs in geophysics. In many cases, they provide access to full posterior distributions, which provide better and more informative estimates for the solutions as well as UQ. Our proposed methodology is applied to a real data set for the 2006 Guerrero SSE, where the objective was to recover the slip on a known interface from observations at few geodetic stations. However, our method may in principle be used for any earthquake slip inversion as long as the FM can be linearized. Lastly, once the slip inversion is available, calculating the seismic moment with formal UQ is a simple subproduct of our methodology. A new and very interesting result, we believe.

One limitation of our approach is that the FM needs to be linearized to obtain a MTN posterior. A more general FM could also be analyzed, although adding severe computational burden and difficulties in a highly multidimensional MCMC.

Bibliography

Ainsworth, M. and Oden, J. T. (2011). *A posteriori error estimation in finite element analysis*, volume 37. John Wiley & Sons.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723.

Akaike, H. (1980). Likelihood and the Bayes procedure, in "Bayesian Statistics", ed. by JM Bernardo, MH DeGroot, DV Lindley, and AFM Smith.

Amey, R., Hooper, A., and Walters, R. (2018). A Bayesian method for incorporating self-similarity into earthquake slip inversions. *Journal of Geophysical Research: Solid Earth*, 123(7):6052–6071.

Babuška, I. and Rheinboldt, W. C. (1978). A-posteriori error estimates for the finite element method. *International Journal for Numerical Methods in Engineering*, 12(10):1597–1615.

Bardsley, J. M. and Kaipio, J. (2013). Gaussian Markov random field priors for inverse problems. *Inverse Problems & Imaging*, 7(2):397–416.

Bastani, M. and Salkuyeh, D. K. (2012). A highly accurate method to solve Fisher's equation. *Pramana*, 78(3):335–346.

Bekaert, D. P. S., Hooper, A., and Wright, T. J. (2015). Reassessing the 2006 Guerrero slow-slip event, Mexico: Implications for large earthquakes in the Guerrero gap. *Journal of Geophysical Research: Solid Earth*, 120(2):1357–1375.

Bellman, R. and Kalaba, R. E. (1965). *Quasilinearization and nonlinear boundary-value problems*, volume 3. New York: American Elsevier.

Berzins, M. (1988). Global error estimation in the method of lines for parabolic equations. *SIAM Journal on Scientific and Statistical Computing*, 9(4):687–703.

Brenner, S. and Scott, R. (2008). *The Mathematical Theory of Finite Element Methods*, volume 15 of *Texts in Applied Mathematics*. Springer-Verlag, New York.

Breslaw, J. (1994). Random sampling from a truncated multivariate normal distribution. *Applied Mathematics Letters*, 7(1):1–6.

Burden, R. L. and Faires, J. D. (2011). *Numerical Analysis*. Brooks/Cole, Boston, MA, ninth edition.

Burggraf, O. (1964). An exact solution of the inverse problem in heat conduction theory and applications. *Journal of Heat transfer*, 86(3):373–380.

Cai, C., Mohammad-Djafari, A., Legoupil, S., and Rodet, T. (2011). Bayesian data fusion and inversion in X-ray multi-energy computed tomography. In *Proc. 18th IEEE International Conference on Image Processing*, pages 1377–1380, Brussels, Belgium.

Calvetti, D., Morigi, S., Reichel, L., and Sgallari, F. (2000). Tikhonov regularization and the L-curve for large discrete ill-posed problems. *Journal of computational and applied mathematics*, 123(1-2):423–446.

Capistrán, M., Christen, J., and Donnet, S. (2016). Bayesian analysis of ODE's: solver optimal accuracy and Bayes factors. *Journal of Uncertainty Quantification*, 4(1):829–849.

Capistrán, M. A., Christen, J. A., Daza-Torres, M. L., Flores-Arguedas, H., and Montesinos-López, J. C. (2021). Error control of the numerical posterior with Bayes factors in Bayesian uncertainty quantification. *Bayesian Analysis*, 1:1 – 23.

Cash, J. R. and Karp, A. H. (1990). A variable order Runge-Kutta method for initial value problems with rapidly varying right-hand sides. *ACM Transactions on Mathematical Software (TOMS)*, 16(3):201–222.

Cavalié, O., Pathier, E., Radiguet, M., Vergnolle, M., Cotte, N., Walpersdorf, A., Kostoglodov, V., and Cotton, F. (2013). Slow slip event in the Mexican subduction zone: Evidence of shallower slip in the Guerrero seismic gap for the 2006 event revealed by the joint inversion of InSAR and GPS data. *Earth and Planetary Science Letters*, 367:52–60.

Chadan, K., Colton, D., Päivärinta, L., and Rundell, W. (1997). *An Introduction to Inverse Scattering and Inverse Spectral Problems*. Society for Industrial and Applied Mathematics.

Christen, J. A., Capistrán, M. A., and Moreles, M. Á. (2016). Numerical posterior distribution error control and expected Bayes factors in the Bayesian uncertainty quantification of inverse problems. *arXiv preprint arXiv:1607.02194*.

Christen, J. A. and Fox, C. (2010). A general purpose sampling algorithm for continuous distributions (the t-walk). *Bayesian Analysis*, 5(2):263–281.

Christen, J. A., Fox, C., and Santana-Cibrian, M. (2017). Optimal direction Gibbs sampler for truncated multivariate normal distributions. *Communications in Statistics-Simulation and Computation*, 46(4):2587–2600.

Christen, J. A. and Pérez-Garmendia, J. L. (2021). Weak and TV consistency in Bayesian uncertainty quantification using disintegration. *Boletín de la Sociedad Matemática Mexicana*, 27(2):1–23.

Cliffe, K., Giles, M., Scheichl, R., and Teckentrup, A. L. (2011). Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients. *Computing and Visualization in Science*, 14(3):https://doi.org/10.1007/s00791–011–0160–x.

Courant, R., Friedrichs, K., and Lewy, H. (1928). On the partial difference equations op mathematical physics. *Mathematische Annalen*, 100(1):32–74.

Cruz-Atienza, V., Tago, J., Villafuerte, C., Wei, R., Garza-Girón, R., Domínguez, L., Kostoglodov, V., Nishimura, T., Franco, S., Real, J., Santoyo, M., Ito, Y., and Kazachkina, E. (2021). Short-term interaction between silent and devastating earthquakes in Mexico. *Nature Communications*, 12(2171).

Cui, T., Marzouk, Y. M., and Willcox, K. E. (2015). Data-driven model reduction for the Bayesian solution of inverse problems. *International Journal for Numerical Methods in Engineering*, 102(5):966–990.

Daon, Y. and Stadler, G. (2018). Mitigating the influence of the boundary on PDE-based covariance operators. *Inverse Problems & Imaging*, 12:1083.

Daza, M. L., Capistrán, M. A., Christen, J. A., and Guadarrama, L. (2017). Solution of the inverse scattering problem from inhomogeneous media using affine invariant sampling. *Mathematical Methods in the Applied Sciences*, 40(9):3311–3319.

Daza-Torres, M. L. and Montesinos-López, J. C. (2020). Codes. https://github.com/mdazatorres/Error_control.

De S. R. Gago, J. P., Kelly, D. W., Zienkiewicz, O. C., and Babuska, I. (1983). A posteriori error analysis and adaptive processes in the finite element method: Part II—adaptive mesh refinement. *International Journal for Numerical Methods in Engineering*, 19(11):1621–1656.

Diagana, T. (2018). *Semilinear Evolution Equations and Their Applications*. Springer.

Fox, C. and Norton, R. A. (2016). Fast sampling in a linear-Gaussian inverse problem. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):1191–1218.

Fukuda, J. and Johnson, K. M. (2008). A fully Bayesian inversion for spatial distribution of fault slip with objective smoothing. *Bulletin of the Seismological Society of America*, 98(3):1128–1146.

Galbally, D., Fidkowski, K., Willcox, K., and Ghattas, O. (2010). Non-linear model reduction for uncertainty quantification in large-scale inverse problems. *International journal for numerical methods in engineering*, 81(12):1581–1608.

Gao, H., Schmidt, D. A., and Weldon, Ray J., I. (2012). Scaling relationships of source parameters for slow slip events. *Bulletin of the Seismological Society of America*, 102(1):352–360.

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

Genz, A. and Bretz, F. (2009). *Computation of multivariate normal and t probabilities*, volume 195. Springer Science & Business Media.

Geweke, J. (1991). Efficient simulation from the multivariate normal and Student-t distributions subject to linear constraints and the evaluation of constraint probabilities. In *Computing science and statistics: Proceedings of the 23rd symposium on the interface*, volume 571, page 578. Citeseer.

Grätsch, T. and Bathe, K.-J. (2005). A posteriori error estimation techniques in practical finite element analysis. *Computers & structures*, 83(4-5):235–265.

Holder, D. S. (2004). *Electrical impedance tomography: methods, history and applications*. CRC Press.

Kaipio, J. and Somersalo, E. (2006). *Statistical and computational inverse problems*, volume 160. Springer Science & Business Media.

Kaipio, J. P. and Fox, C. (2011). The Bayesian framework for inverse problems in heat transfer. *Heat Transfer Engineering*, 32(9):718–753.

Kato, A., Obara, K., Igarashi, T., Tsuruoka, H., Nakagawa, S., and Hirata, N. (2012). Propagation of slow slip leading up to the 2011 Mw 9.0 Tohoku-Oki earthquake. *Science*, 335(6069):705–708.

Katsiolides, G., Müller, E. H., Scheichl, R., Shardlow, T., Giles, M. B., and Thomson, D. J. (2018). Multilevel Monte Carlo and improved timestepping methods in atmospheric dispersion modelling. *Journal of Computational Physics*, 354:320 – 343.

Kotecha, J. and Djuric, P. (1999). Gibbs sampling approach for generation of truncated multivariate Gaussian random variables. In *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No.99CH36258)*, volume 3, pages 1757–1760 vol.3.

Li, J. and Marzouk, Y. M. (2014). Adaptive construction of surrogates for the Bayesian solution of inverse problems. *SIAM Journal on Scientific Computing*, 36(3):A1163–A1186.

Lieberman, C., Willcox, K., and Ghattas, O. (2010). Parameter and state model reduction for large-scale statistical inverse problems. *SIAM Journal on Scientific Computing*, 32(5):2523–2542.

Liu, J. S. (2008). *Monte Carlo strategies in scientific computing*. Springer Science & Business Media.

Maatouk, H. and Bay, X. (2016). A new rejection sampling method for truncated multivariate Gaussian random variables restricted to convex sets. In *Monte Carlo and quasi-monte carlo methods*, pages 521–530. Springer.

Maatouk, H. and Bay, X. (2017). Gaussian process emulators for computer experiments with inequality constraints. *Mathematical Geosciences*, 49(5):557–582.

Martin, S. A., Blechta, J., Hake, J., Johansson, A., Kehlet, B., Logg, A., Richardson, C., Ring, J., Rognes, M. E., and Wells, G. N. (2015). The FEniCS Project Version 1.5. *Archive of Numerical Software*, 3(100):9–23.

Marzouk, Y. M., Najm, H. N., and Rahn, L. A. (2007). Stochastic spectral methods for efficient Bayesian solution of inverse problems. *Journal of Computational Physics*, 224(2):560–586.

McCaffrey, R., Qamar, A. I., King, R. W., Wells, R., Khazaradze, G., Williams, C. A., Stevens, C. W., Vollick, J. J., and Zwick, P. C. (2007). Fault locking, block rotation and crustal deformation in the Pacific Northwest. *Geophysical Journal International*, 169(3):1315–1340.

Michalak, A. M. (2008). A Gibbs sampler for inequality-constrained geostatistical interpolation and inverse modeling. *Water Resources Research*, 44(9).

Minasny, B. and McBratney, A. B. (2005). The Matèrn function as a general model for soil variograms. *Geoderma*, 128:192–207.

Minson, S., Simons, M., and Beck, J. (2013). Bayesian inversion for finite fault earthquake source models I—Theory and algorithm. *Geophysical Journal International*, 194(3):1701–1726.

Molina, J. D. and Christen, J. A. (2021). Criterio para determinar el tamaño de muestra en procesos de simulación estocástica. Master's thesis, Centro de Investigación en Matemáticas, CIMAT.

Montesinos-López, J. C. and Christen, J. A. (2016). Distribución de direcciones en el Gibbs sampler generalizado. Master's thesis, Centro de Investigación en Matemáticas, CIMAT.

Nocquet, J.-M. (2018). Stochastic static fault slip inversion from geodetic data with non-negativity and bound constraints. *Geophysical Journal International*, 214(1):366–385.

Norton, R. A., Christen, J. A., and Fox, C. (2018). Sampling hyperparameters in hierarchical models: Improving on Gibbs for high-dimensional latent fields and large datasets. *Communications in Statistics-Simulation and Computation*, 47(9):2639–2655.

Peherstorfer, B., Willcox, K., and Gunzburger, M. (2018). Survey of multifidelity methods in uncertainty propagation, inference, and optimization. *Siam Review*, 60(3):550–591.

Plata-Martínez, R., Ide, S., Shinohara, M., García, E., Mizuno, N., Domínguez, L., Taira, T., Yamashita, Y., Toh, A., Yamada, T., Real, J., Husker, A., Cruz-Atienza, V., and Ito, Y. (2021). Shallow slow earthquakes to decipher future catastrophic earthquakes in the Guerrero seismic gap. *Nature Communications*, 12(1):1–8.

Radiguet, M., Cotton, F., Vergnolle, M., Campillo, M., Valette, B., Kostoglodov, V., and Cotte, N. (2011). Spatial and temporal evolution of a long term slow slip event: the 2006 Guerrero slow slip event. *Geophysical Journal International*, 184(2):816–828.

Radiguet, M., Cotton, F., Vergnolle, M., Campillo, M., Walpersdorf, A., Cotte, N., and Kostoglodov, V. (2012). Slow slip events and strain accumulation in the Guerrero gap, Mexico. *Journal of Geophysical Research: Solid Earth*, 117(B4).

Radiguet, M., Perfettini, H., Cotte, N., Gualandi, A., Valette, B., Kostoglodov, V., Lhomme, T., Walpersdorf, A., Cabral-Cano, E., and Campillo, M. (2016). Triggering of the 2014 Mw7.3 Papanoa earthquake by a slow slip event in Guerrero, Mexico. *Nature Geoscience*, 9:829–833.

Rasmussen, C. E., Bernardo, J., Bayarri, J., Berger, J., Dawid, A., Heckerman, D., and Smith, A, e. a. E. (2003). Gaussian processes to speed up hybrid Monte Carlo for expensive Bayesian integrals. In *Proc. Bayesian Statistics 7*, pages 651–659, Oxford, UK: Oxford University Press.

Reddy, J. N. (2006). *Introduction to the Finite Element Method, Third Edition*. McGraw-Hill Education, New York.

Robert, C. and Casella, G. (2013). *Monte Carlo statistical methods*. Springer Science & Business Media.

Robert, C. P. (1995). Simulation of truncated normal variables. *Statistics and computing*, 5(2):121–125.

Rognes, M. E. and Logg, A. (2013). Automated goal-oriented error control I: Stationary variational problems. *SIAM Journal on Scientific Computing*, 35(3):C173–C193.

Roy, C. (2010). Review of discretization error estimators in scientific computing. In *Proc. 48th AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition*, page 126, Orlando, Florida.

Ruiz, S., Metois, M., Fuenzalida, A., Ruiz, J., Leyton, F., Grandin, R., Vigny, C., Madariaga, R., and Campos, J. (2014). Intense foreshocks and a slow slip event preceded the 2014 Iquique Mw 8.1 earthquake. *Science*, 345(6201):1165–1169.

Sari, M. and Gürarslan, G. (2011). A sixth-order compact finite difference method for the one-dimensional sine-Gordon equation. *International Journal for Numerical Methods in Biomedical Engineering*, 27(7):1126–1138.

Snieder, R. and Trampert, J. (1999). Inverse problems in geophysics. In *Wavefield inversion*, pages 119–190. Springer.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the royal statistical society: Series b (statistical methodology)*, 64(4):583–639.

Stein, S. and Wysession, M. (2009). *An introduction to seismology, earthquakes, and earth structure*. John Wiley & Sons.

Stuart, A. M. (2010). Inverse problems: a Bayesian perspective. *Acta numerica*, 19:451–559.

Tago, J., Cruz-Atienza, V. M., Villafuerte, C., Nishimura, T., Kostoglodov, V., Real, J., and Ito, Y. (2021). Adjoint slip inversion under a constrained optimization framework: revisiting the 2006 Guerrero slow slip event. *Geophysical Journal International*, 226(2):1187–1205.

Tzafestas, S. G. (2013). *Distributed parameter control systems: Theory and application*. Elsevier.

Udías, A., Vallina, A. U., Madariaga, R., and Buforn, E. (2014). *Source mechanisms of earthquakes: theory and practice*. Cambridge University Press.

Wallace, L. M. and Beavan, J. (2010). Diverse slow slip behavior at the Hikurangi subduction margin, New Zealand. *Journal of Geophysical Research: Solid Earth*, 115(B12).

Xu, P. (2019). Akaike's Bayesian information criterion (ABIC) or not ABIC for geophysical inversion. *arXiv preprint arXiv:1911.06564*.

Yan, L. and Zhou, T. (2019a). An adaptive multifidelity pc-based ensemble Kalman inversion for inverse problems. *International Journal for Uncertainty Quantification*, 9(3):205–220.

Yan, L. and Zhou, T. (2019b). Adaptive multi-fidelity polynomial chaos approach to Bayesian inference in inverse problems. *Journal of Computational Physics*, 381:110–128.

Yu, J.-w. and Tian, G.-l. (2011). Efficient algorithms for generating truncated multivariate normal distributions. *Acta Mathematicae Applicatae Sinica, English Series*, 27(4):601.

APPENDIX A

A.1 Details of the numerical solution

For the reader's convenience, here we describe in detail the numerical method introduced in Section 3.2.

To solve the PDE in Eq. (3.1), we start by separating the function F into a linear (L) and a nonlinear (N) component and rewriting the Eq. (3.1) in the form

$$\dot{u} = u'' + \mathsf{L}[u, u'] + \mathsf{N}[u, u']. \tag{A.1}$$

The nonlinear operator N is approximated with a Taylor series, assuming that the difference $u_{i+1,\cdot} - u_{i,\cdot}$ and all its spatial derivatives are small, hence

$$\mathsf{N}[u_{i+1,\cdot}, u'_{i+1,\cdot}] \approx \mathsf{N}[u_{i,\cdot}, u'_{i,\cdot}] + \phi_{0,i}[u_{i,\cdot}, u'_{i,\cdot}] \cdot (u_{i+1,\cdot} - u_{i,\cdot}) + \phi_{1,i}[u_{i,\cdot}, u'_{i,\cdot}] \cdot (u'_{i+1,\cdot} - u'_{i,\cdot}), \tag{A.2}$$

where $u_{i,\cdot} := u(x_i, t)$ is the solution of Eq. (3.1) evaluated in (x_i, t) and $\phi_{k,i}[u_{i,\cdot}, u'_{i,\cdot}] := \frac{\partial^k \mathsf{N}}{\partial u^{(k)}}[u_{i,\cdot}, u^{(k)}_{i,\cdot}]$, $k = 0, 1$.

Substituting Eq. (A.2) into Eq. (A.1), we get

$$\dot{u}_{i+1,\cdot} = u''_{i+1,\cdot} + \mathsf{L}[u_{i+1,\cdot}, u'_{i+1,\cdot}] + \mathsf{N}[u_{i,\cdot}, u'_{i,\cdot}] + \phi_{0,i}[u_{i,\cdot}, u'_{i,\cdot}] \cdot (u_{i+1,\cdot} - u_{i,\cdot}) + \phi_{1,i}[u_{i,\cdot}, u'_{i,\cdot}] \cdot (u'_{i+1,\cdot} - u'_{i,\cdot}), \tag{A.3}$$

for $i = 1, ..., N - 2$.

Now, the spatial partial derivatives are approximated using the central difference formula (3.8)-(3.9). We write in matrix form the central differences approximations

$$\mathbf{V}'_h = \frac{1}{2h}\mathbf{A}_x \mathbf{V}_h + \mathbf{C}_x, \tag{A.4}$$

and

$$\mathbf{V}_h'' = \frac{1}{h^2}\mathbf{A}_{xx}\mathbf{V}_h + \mathbf{C}_{xx}, \tag{A.5}$$

where $V_h = (v_{1,\cdot}, v_{2,\cdot}, \ldots, v_{N-1,\cdot})^T$ approximates $\mathbf{U} = (u_{1,\cdot}, u_{2,\cdot}, \ldots, u_{N-1,\cdot})^T$,

$$\mathbf{A}_x = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ -1 & 0 & 1 & \ldots & 0 \\ 0 & -1 & 0 & \ldots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 1 \\ 0 & \ldots & 0 & -1 & 0 \end{pmatrix}_{(N-2)\times(N-2)}, \quad \mathbf{C}_x = \frac{1}{2h}\begin{pmatrix} -u_{0,\cdot} \\ 0 \\ \vdots \\ 0 \\ u_{N,\cdot} \end{pmatrix}_{(N-2)\times 1},$$

$$\mathbf{A}_{xx} = \begin{pmatrix} -2 & 1 & 0 & \ldots & 0 \\ 1 & -2 & 1 & \ldots & \vdots \\ 0 & 1 & -2 & \ldots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 1 \\ 0 & \ldots & 0 & 1 & -2 \end{pmatrix}_{(N-2)\times(N-2)}, \text{ and } \quad \mathbf{C}_{xx} = \frac{1}{h^2}\begin{pmatrix} u_{0,\cdot} \\ 0 \\ \vdots \\ 0 \\ u_{N,\cdot} \end{pmatrix}_{(N-2)\times 1}.$$

Now, $u'_{0,\cdot}$ is approximated with the forward difference scheme (3.9), which leaves us

$$\widehat{\mathbf{V}}_h' = \frac{1}{2h}\widehat{\mathbf{A}}_x\mathbf{V}_h + \widehat{\mathbf{C}}_x, \tag{A.6}$$

$$\widehat{\mathbf{A}}_x = \begin{pmatrix} 2 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & \vdots \\ -1 & 0 & 1 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 1 \\ 0 & \ldots & 0 & 1 & 0 \end{pmatrix}_{(N-2)\times(N-2)}, \text{ and } \quad \widehat{\mathbf{C}}_x = \frac{1}{2h}\begin{pmatrix} -2u_{0,\cdot} \\ -u_{0,\cdot} \\ \vdots \\ 0 \\ u_{N,\cdot} \end{pmatrix}_{(N-2)\times 1}.$$

where $\widehat{\mathbf{V}}_h = (v_{0,\cdot}, v_{1,\cdot}, \ldots, v_{N-2,\cdot})^T$.

Finally, substituting the approximate derivatives in Eqs. (A.4)–(A.6) into Eq. (3.7), joint with the boundary conditions given in Eq. (3.2), we get the following semi-discrete differential

equation:

$$\begin{aligned}\dot{\mathbf{V}}_h &= \frac{1}{h^2}\mathbf{A}_{xx}\mathbf{V}_h + \mathbf{C}_{xx} + \mathsf{L}\left(\mathbf{V}_h, \frac{1}{2h}\mathbf{A}_x\mathbf{V}_h + \mathbf{C}_x\right) + \mathsf{N}\left(\widehat{\mathbf{V}}_h, \frac{1}{2h}\widehat{\mathbf{A}}_x\mathbf{V}_h + \widehat{\mathbf{C}}\right) \\ &\quad + \mathbf{\Phi_0}\left[\widehat{\mathbf{V}}, \frac{1}{2h}\widehat{\mathbf{A}}_x\mathbf{V}_h + \widehat{\mathbf{C}}\right] \cdot \left(\mathbf{V}_h - \widehat{\mathbf{V}}_h\right) \\ &\quad + \mathbf{\Phi_1}\left[\widehat{\mathbf{V}}_h, \frac{1}{2h}\widehat{\mathbf{A}}_x\mathbf{V} + \widehat{\mathbf{C}}\right] \cdot \left(\frac{1}{2h}(\mathbf{A}_x - \widehat{\mathbf{A}}_x)\mathbf{V}_h + (\mathbf{C}_x - \widehat{\mathbf{C}}_x)\right), \end{aligned} \tag{A.7}$$

where $\mathbf{\Phi_k}[\widehat{\mathbf{V}}, \widehat{\mathbf{V}}'] = (\phi_{k,0}[v_{0,\cdot}, v'_{0,\cdot}], \phi_{k,1}[v_{1,\cdot}, v'_{1,\cdot}], \ldots, \phi_{k,N-3}[v_{N-3,\cdot}, v'_{N-3,\cdot}])^T$.

Note that the right-hand side of equation (A.7) only depends on $\mathbf{V}$ and t, due $v_{0,\cdot}$ and $v_{N,\cdot}$ are known (3.2). Thus, we can write (A.7) in a compact form,

$$\dot{\mathbf{V}} = \frac{1}{h^2}\mathbf{A}_{xx}\mathbf{V} + \mathbf{F}(t, \mathbf{V}) \tag{A.8}$$

with

$$\begin{aligned}\mathbf{F}(t, \mathbf{V}) &= \mathbf{C}_{xx} + \mathsf{L}\left(\mathbf{V}_h, \frac{1}{2h}\mathbf{A}_x\mathbf{V}_h + \mathbf{C}_x\right) + \mathsf{N}\left(\widehat{\mathbf{V}}_h, \frac{1}{2h}\widehat{\mathbf{A}}_x\mathbf{V}_h + \widehat{\mathbf{C}}\right) \\ &\quad + \mathbf{\Phi_0}\left[\widehat{\mathbf{V}}, \frac{1}{2h}\widehat{\mathbf{A}}_x\mathbf{V}_h + \widehat{\mathbf{C}}\right] \cdot \left(\mathbf{V}_h - \widehat{\mathbf{V}}_h\right) \\ &\quad + \mathbf{\Phi_1}\left[\widehat{\mathbf{V}}_h, \frac{1}{2h}\widehat{\mathbf{A}}_x\mathbf{V} + \widehat{\mathbf{C}}\right] \cdot \left(\frac{1}{2h}(\mathbf{A}_x - \widehat{\mathbf{A}}_x)\mathbf{V}_h + (\mathbf{C}_x - \widehat{\mathbf{C}}_x)\right). \end{aligned}$$

A.2 Stability considerations

We briefly describe stability considerations for the DFRK method introduced in Section 3.2.

Let

$$\begin{aligned}\mathbf{W}_{\cdot,n+1} = \mathbf{W}_{\cdot,n} &+ \left(\frac{b_1 k}{h^2}\mathbf{A}_{xx}\mathbf{W}_{\cdot,n} + k b_1 \mathbf{F}\left(t_n, \mathbf{W}_{\cdot,n}\right)\right) + \sum_{i=2}^{6}\frac{b_i k}{h^2}\mathbf{A}_{xx}\left(\mathbf{W}_{\cdot,n} + k\sum_{j=1}^{i-1} a_{ij}\mathbf{K}_j\right) \\ &+ k\sum_{i=2}^{6} b_i \mathbf{F}\left(t_n + c_i k, \mathbf{W}_{\cdot,n} + k\sum_{j=1}^{i-1} a_{ij}\mathbf{K}_j\right)\end{aligned}$$

the solution of (3.1) using the FD-RKCK method (see Section 3.2). To determine the CFL condition, we consider only the pure diffusion. Thus, the scheme is stable only if $\rho(\mathbf{A}) \leq 1$ (Burden and Faires 2011), where

$$\mathbf{A} = \frac{B_m k}{h^2}\mathbf{A}_{xx} \quad B_m = \max_i b_i.$$

The eigenvalues of $\mathbf{A}$ can be shown to be

$$\mu_i = -4\lambda \left(\sin\left(\frac{i\pi}{2N} \right) \right)^2, \quad \text{for} \quad i = 1, 2, \ldots, N-2,$$

where $\lambda = B_m \frac{k}{h^2}$. So, the condition for stability consequently reduces to determining if

$$\rho\left(\mathbf{A}\right) = \max_{1 \le i \le N-2} \left| -4\lambda \left(\sin\left(\frac{i\pi}{2N} \right) \right)^2 \right| \le 1,$$

and this simplifies to

$$0 \le \lambda \left(\sin\left(\frac{i\pi}{2N} \right) \right)^2 \le \frac{1}{4}, \quad \forall i = 1, 2, \ldots, N-2.$$

Stability requires that this inequality condition hold as $h \to 0$, or, equivalently, as $N \to \infty$,

$$\lim_{N \to \infty} \left[\sin\left(\frac{(N-1)\,\pi}{2N} \right) \right]^2 = 1.$$

Thus, stability occurs if only if $0 \le \lambda \le \frac{1}{4}$. By definition $\lambda = B_m \frac{k}{h^2}$, so this inequality requires that h and k be chosen such that

$$B_m \frac{k}{h^2} \le \frac{1}{4}.$$

The method converges to the solution with a rate of convergence $O\left(h^p + k^4\right)$, provided $B_m \frac{k}{h^2} \le \frac{1}{4}$. For the numerical implementation, we take $k = \alpha h^2$, with $\alpha = \frac{1}{4B_m}$.

APPENDIX B

B.1 Determining the variance σ_β^2

Note that the variance σ_β^2 could be considered a random variable. The full conditional σ_β^2 has also a known form, namely $\sigma_\beta^2|\cdots \sim$ Inv-Gamma $\left(\frac{d}{2}+a_\beta, \frac{1}{2}\mathbf{D}^T\mathbf{A}_0\mathbf{D}+b_\beta\right)$, an could in principle be included in the Gibbs sampling. However, this requires a computationally expensive reevaluation and inversion of the precision matrix $\mathbf{A}_p$ at each MCMC step. For this reason, we decided to simulate only from the conditional distribution of $\mathbf{D}$, conditioned on the value of σ_β^2 that maximizes the marginal posterior distribution of this parameter, as follows.

Akaike's Bayesian Information Criterion (ABIC), proposed by Akaike (1980), has been widely applied in geophysical inversion to determine the regularization parameters. Following ABIC, we propose a criterion by maximizing the marginal posterior distribution of this parameter, that is,

$$\max : f\left(\sigma_\beta^2|\mathbf{Y}\right) = \int f\left(\mathbf{D}, \sigma_\beta^2|\mathbf{Y}\right) d\mathbf{D} = \frac{1}{f\left(\mathbf{Y}\right)} \int f\left(\mathbf{Y}, \mathbf{D}, \sigma_\beta^2\right) d\mathbf{D}. \tag{B.1}$$

For this, we consider the follow hierarchical model

$$\begin{aligned} \mathbf{Y}|\mathbf{D} &\sim N\left(\mathbf{X}\mathbf{D}, \mathbf{A}\right), \\ \mathbf{D}|\sigma_\beta^2 &\sim N_d\left(\mathbf{0}, \frac{1}{\sigma_\beta^2}\mathbf{A}_0\right); \quad \mathbf{A}_0 = \boldsymbol{\beta}\mathbf{W}\mathbf{C}^{-1}\mathbf{W}\boldsymbol{\beta}, \\ \sigma_\beta^2 &\sim \text{Inv-Gamma}\left(a_\beta, b_\beta\right), \end{aligned} \tag{B.2}$$

where $\mathbf{A} = \boldsymbol{\Sigma}^{-1}$ is the precision matrix, and Inv-Gamma (α, β) denote a inverse Gamma distribution with shape parameter α and scale parameter β.

Note that,

$$\begin{aligned}\int f\left(\mathbf{Y},\mathbf{D},\sigma_\beta^2\right)d\mathbf{D} &= \int f\left(\mathbf{Y},\mathbf{D}|\sigma_\beta^2\right)\pi\left(\sigma_\beta^2\right)d\mathbf{D}\\ &= \pi\left(\sigma_\beta^2\right)\int f\left(\mathbf{Y},\mathbf{D}|\sigma_\beta^2\right)d\mathbf{D}\\ &= \pi\left(\sigma_\beta^2\right)m\left(\mathbf{Y}|\sigma_\beta^2\right),\end{aligned}$$

where $\pi\left(\sigma_\beta^2\right)$ is the prior distribution for σ_β^2, and

$$m\left(\mathbf{Y}|\sigma_\beta^2\right) := \int f\left(\mathbf{Y},\mathbf{D}|\sigma_\beta^2\right)d\mathbf{D}. \tag{B.3}$$

So, maximizing (B.1) is equivalent to minimizing

$$\min : \ell\left(\sigma_\beta^2\right) = -\log\left(m\left(\mathbf{Y}|\sigma_\beta^2\right)\right) - \log\left(\pi\left(\sigma_\beta^2\right)\right).$$

Now, with the hierarchical model (B.2) and using the derivation of (B.3) given in Xu (2019), that is,

$$m\left(\mathbf{Y}|\sigma_\beta^2\right) = \frac{1}{(2\pi)^{n/2}\sqrt{|\Sigma_{py}|}}\exp\left\{-\frac{1}{2}\left(\mathbf{Y}-\mathbf{X}\boldsymbol{\mu}_0\right)^T\Sigma_{py}^{-1}\left(\mathbf{Y}-\mathbf{X}\boldsymbol{\mu}_0\right)\right\},$$

where $\boldsymbol{\mu}_0 = \mathbf{0}$ is the prior mean of $\mathbf{D}$, and $\Sigma_{py} = \mathbf{A}^{-1}+\mathbf{X}\mathbf{A}_0^{-1}\mathbf{X}^T\sigma_\beta^2$, it follows that, the optimal prior variance is obtained by minimizing

$$\min : \ell\left(\sigma_\beta^2\right) = \ln\left\{\det\left(\Sigma_{py}\right)\right\} + \mathbf{Y}^T\Sigma_{py}^{-1}\mathbf{Y} + \left(a_\beta+1\right)\log\left(\sigma_\beta^2\right) + \frac{b_\beta}{\sigma_\beta^2}. \tag{B.4}$$

Minimizing the above expressions is straightforward since is a function defined in $\mathbb{R}^1$.