Bayesian inference of earthquake parameters from buoy data using a polynomial chaos based surrogate

Loïc Giraldi[a,*], Olivier P. Le Maître[b], Kyle T. Mandli[c], Clint N. Dawson[d], Ibrahim Hoteit[e], Omar M. Knio[a]

[a] *Division of Computer, Electrical and Mathematical Sciences and Engineering, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia*
[b] *Univerité Paris-Saclay, LIMSI-CNRS, Orsay, France*
[c] *Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY*
[d] *Institute for Computational Engineering and Science, University of Texas at Austin, Austin, TX*
[e] *Division of Physical Sciences and Engineering, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia*

Abstract

This work addresses the estimation of the parameters of an earthquake model by the consequent tsunami, with an application to the Chile 2010 event. We are particularly interested in the Bayesian inference of the location, the orientation and the slip of an Okada based model of the earthquake ocean floor displacement. The tsunami numerical model is based on the GeoClaw software while the observational data is provided by a single DART® buoy. We propose in this paper a methodology based on polynomial chaos expansion to construct a surrogate model of the wave height at the buoy location. A correlated noise model is first proposed in order to represent the discrepancy between the computational model and the data. This step is necessary, as a classical independent Gaussian noise is shown to be unsuitable for modeling the error, and to prevent convergence of the Markov-Chain Monte-Carlo sampler. Second, the polynomial chaos model is subsequently improved to handle the variability of the arrival time of the wave, using a preconditioned non-intrusive spectral method. Finally, the construction of a reduced model dedicated to Bayesian inference is proposed. Numerical results are presented and discussed.

Keywords: uncertainty quantification, Bayesian inference, polynomial chaos expansion, noise model, low-rank representation, shallow water equation, tsunami, earthquake inversion

1. Introduction

Subduction zone tsunamis are one of the most devastating natural hazards on the planet, causing widespread destruction of life and property. Scientists have long attempted to mitigate the risk of another devastating tsunami, but unfortunately these efforts are often hampered by the large amount of uncertainty involved in predicting such events. These uncertainties include the lack of knowledge in bathymetry, parameterizations in the friction model, and perhaps most notably, the earthquake that triggers these large tsunamis. This paper is dedicated to addressing the last and most substantial source of input uncertainty, that of the earthquake parameterization itself.

Uncertainty quantification regarding tsunamis has not gone unnoticed in the literature. Studies include work done with existing earthquake parameterizations attempting to ascertain the best fit to observable data with only a temporal shift as a free parameter [20], while others looked at landslide generated tsunamis using a similar approach [28]. Bayesian inversion of the slip distribution was also studied on synthetic data [10]. In

*Corresponding author
Email addresses: loic.giraldi@kaust.edu.sa (Loïc Giraldi), olm@limsi.fr (Olivier P. Le Maître), kyle.mandli@columbia.edu (Kyle T. Mandli), clint@ices.utexas.edu (Clint N. Dawson), ibrahim.hoteit@kaust.edu.sa (Ibrahim Hoteit), omar.knio@kaust.edu.sa (Omar M. Knio)

the broader oceanic context there has been work to examine tidal components employing adjoint or Kalman filtering approaches [9, 19, 32, 15, 22].

The present work aims at extending previous effort in [29, 30] to assess uncertainty quantification and source inference in a more complex setting, and in a situation where measurements are scarce. The tsunami that occurred off the coast of Chile on February 27, 2010, was selected for this purpose. The complexity in this application arises from the fact that the epicenter of the earthquake that triggered the tsunami waves was located close to the coastline, and that wave height measurements were provided by a single DART® buoy only. With uncertainty in source location, this resulted in large variability in the wave arrival time, which was compounded by reflected waves from the boundaries.

We initially attempted to implement the methodology in [30] to construct a representation of the wave height response at the buoy location, and to exploit this representation to infer the earthquake source parameters. However, as further described later, a straightforward implementation of the approach in [30] was not satisfactory, for both the forward and inverse problems. Consequently, it became necessary to introduce alternative methodologies. These include the development of an elaborate likelihood model that incorporates a correlated noise term and the construction of preconditioned functional representations of the transient wave-height dependence on the source parameters. The development of the surrogate and noise models, and the assessment of their performance through applications to the 2010 Chile earthquake, are two of the main contributions of the present work. A third contribution concerns the evaluation of the potential of using a reduced surrogate model for the purpose of signal representation and parameter inference.

This paper is organized as follows. Section 2 describes the computational model, available measurements, and prior information. In Section 3, we provide a brief outline of the Bayesian framework used for the inverse problem, of the polynomial chaos (PC) expansion approach used for representing the model output, and of the non-intrusive projection approach adopted for the purpose of determining the expansion coefficients. Section 4 then outlines the design of the noise model used to describe the discrepancy between measured and predicted wave height. Section 5 discusses results of the simulations, analyzes the behavior of surrogate models in representing model outputs, and assesses the performance of the surrogate and likelihood models in the Bayesian inference methodology. Major conclusions are summarized in Section 6.

2. Test problem: Chile 2010 tsunami

As a case study we will consider the tsunami that occurred off the coast of Chile on February 27, 2010, modeled by the GeoClaw package [5]. A description of the numerical model is given in this section along with a description of the test problem, available data, and parameters of interest.

2.1. Numerical model

The GeoClaw package solves the two-dimensional shallow water equations

$$
\begin{aligned}
&h_t + (hu)_x + (hv)_y = 0,\\
&(hu)_t + \left(hu^2 + \frac{1}{2}gh^2\right)_x + (huv)_y = \\
&\qquad\qquad - ghb_x - gc^2h^{-5/3}\,\|\mathbf{u}\|\,u,\\
&(hv)_t + (huv)_x + \left(hv^2 + \frac{1}{2}gh^2\right)_y = \\
&\qquad\qquad - ghb_y - gc^2h^{-5/3}\,\|\mathbf{u}\|\,v,
\end{aligned}
$$

where h is the height of the water column, $\mathbf{u} = (u, v)$ is the depth-averaged velocity with components in the longitude and latitude directions, b is the bathymetry, represented with a 10-minute ETOPO2 topography, and $n = 0.025$ is the Manning's coefficient. The initial condition is taken to be at sea-level defined via the datum based on the zero-level of the bathymetric source (in this case ETOPO2). The boundaries of the flow are handled in two ways; (1) where the boundary is dry- land a numerical solver is employed that allows

for flooding of the dry-land as described in [12] and therefore allowed to move and (2) where a boundary is open ocean the boundary conditions allow for waves to exit the domain via a zero-extrapolation numerical boundary condition.

In the present study, we seek to infer the four parameters describing the geometry of the fault, namely the latitude and longitude of the epicenter, the strike angle, and the slip along the fault. The sea-floor deformation is prescribed and computed using the Okada model [25] which uses a Green's function approach to represent the static response of an elastic half- space to a single plane fault.

Concerning the prior knowledge on the four parameters of interest, we rely on the epicenter proposed in a preliminary U.S. Geological Survey (USGS) fault model of the earthquake [1]. Based on the information in [3], we define a large rectangular domain bordering the coast of Chile, from Concepción city (36.6 S, 73.0 W) to San Antonio city (33.25 S, 71.75 W), at a distance between 0 and 1.5 in latitude-longitude degrees. Figure 1 depicts the considered rectangular domain and shows the two nominal locations proposed by the USGS [3] and used in [5] (labeled GeoClaw). For convenience, the rectangular domain is linearly mapped to the reference square $(-1,1)^2$, and the location is expressed in term of two canonical coordinates ξ_1, ξ_2. A uniform distribution of the epicenter location (resp. canonical coordinates) is finally assumed over the rectangular (resp. reference square) domain. Similarly, the strike and the slip parameters are assumed to follow the uniform distributions $\mathcal{U}(10°, 30°)$ and $\mathcal{U}(5\text{ m}, 20\text{ m})$ respectively. Again, the strike and slip values are parameterized by canonical random variables ξ_3 and ξ_4, both uniformly distributed on $(-1,1)$. These four random variables are assumed to be independent, such that the vector of canonical parameter ξ has a uniform prior over a four-dimensional hypercube.

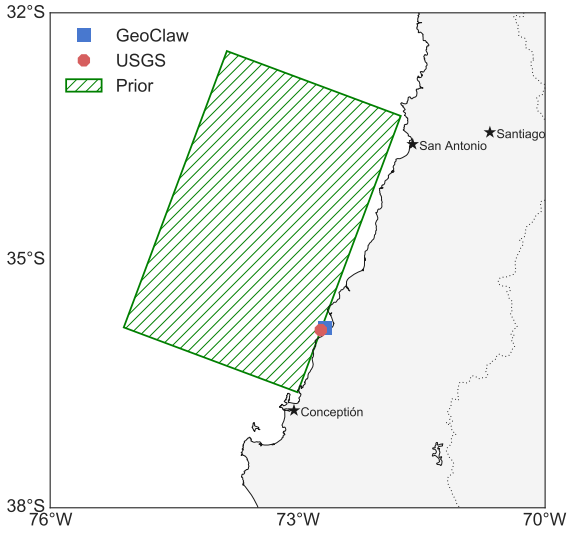

Figure 1: Support of the prior distribution of the location of the epicenter, USGS and GeoClaw nominal value.

2.2. Available observations

For the purpose of inferring the model parameters we rely on the comparison of the sea surface height anomaly predicted by the model and measured at a buoy. The buoy is one of the 39 DART® stations of the national oceanic and atmospheric administration (NOAA) network for early warning detections. It is composed of a bottom pressure recorder at the seafloor coupled with a surface buoy which transmits the recorded data to a satellite [23]. Specifically, the data used in the present work are detided measured sea surface anomalies. The buoy in question is station 32412 located southwest of Lima, at 17.980 S, 86.330 W, and shown in Figure 2. The data recorded by the DART® station 32412 are plotted in Figure 3. It is

[1] See https://earthquake.usgs.gov/earthquakes/eventpage/official20100227063411530_30

seen that the buoy location is a few thousand kilometers away from the epicenter prior domain. We can notice some oscillations at the beginning of the record that are due to the earthquake waves. Since the GeoClaw model does not take into account seismic waves propagation, the data corresponding to the first two recorded hours will be discarded from the analysis. The measured anomaly after 5.5 hours will be also ignored due to the loss the veracity of model equations. The observations used for the inference therefore consist of a set of n sea surface anomalies $y_{i=1,n}$ at times $2 \leq t_{i=1,n} \leq 5.5$ hours after the event.

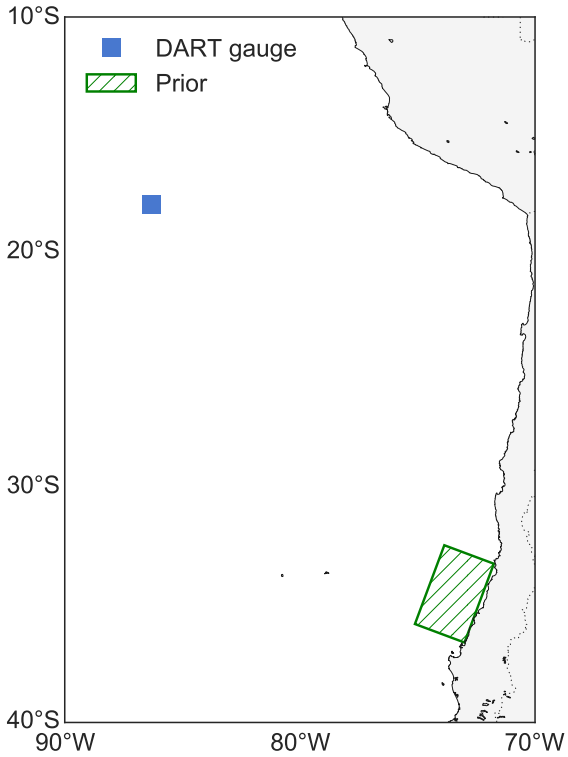

Figure 2: DART® gauge 32412 and support of the prior distribution of the location of the epicenter and comparison with the default GeoClaw model.

We notice that the evaluation of the model predictions of the sea surface anomaly at the buoy location, given the value of the canonical parameter ξ, requires the solution of an elaborate model with significant computational cost. The wall clock time associated with the computation of the solution GeoClaw model is 26.7 s on a single thread. In the present study, all computations are performed on a processor with 24 cores, each core having a 2.50 GHz base frequency. In the following section we discuss the non-intrusive construction of an efficient surrogate model that can be used to model predictions, and subsequently used to perform the inference task.

3. Background

In this section we introduce the notation and briefly cast the Bayesian inference problem and the Polynomial Chaos expansion method to efficiently solve it.

3.1. Bayesian inference

Let $(\Omega, \mathscr{B}, \mathbb{P})$ be an abstract probability space with Ω a sample space, $\mathscr{B}$ a σ-algebra, and $\mathbb{P}$ a probability measure. We are interested in the inference of a random vector $\xi : \Omega \to \Xi \subset \mathbb{R}^p$, with $p < \infty$, using a forward model $M(\xi) = \ell(h(\xi))$, composed of the solution of (an eventually discretized) partial differential

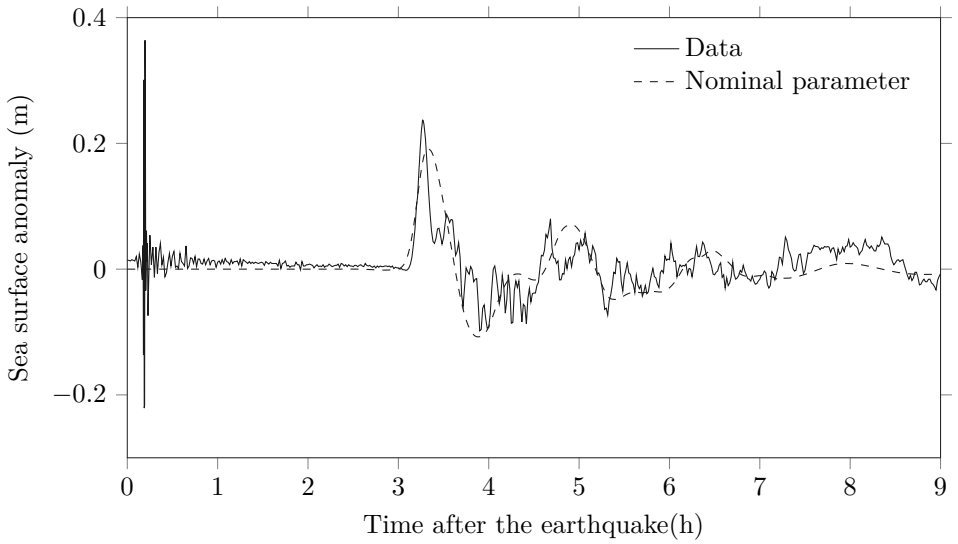

Figure 3: Sea surface anomaly recorded by the DART® station 32412. Also shown is the GeoClaw output using the nominal parameter.

equation $h(\xi)$, and a quantity of interest ℓ, which predicts an observation. In order to perform the inference, an additive noise ε is added to the model prediction, yielding finally

$$Y = M(\xi) + \varepsilon, \tag{1}$$

where Y is a $\mathbb{R}^n$-valued random vector representing the predicted observations.

The parameter ξ and the additive noise ε are random vectors following prior distributions with probability density functions f_ξ and f_ε respectively. These distributions represent the prior belief regarding ξ and ε. When an observation y of Y is available, the posterior probability density $f_\xi(\xi|Y=y)$ of ξ is given by Bayes' theorem:

$$f_\xi(\xi|Y=y) = \frac{f_Y(y|\xi)}{f_Y(y)} f_\xi(\xi).$$

We shall assume that the noise follows a multivariate normal distribution, *i.e.* $\varepsilon \sim \mathcal{N}(\bar{\varepsilon}(\theta), C_\varepsilon(\theta))$. Here, we denote by θ the vector of hyperparameters defining the noise bias $\bar{\varepsilon}(\theta)$ and the covariance matrix $C_\varepsilon(\theta)$. In fact, we will also infer the value of θ with Bayes' theorem, leading to the posterior probability density function $f_{\xi,\theta}(\xi,\theta|Y=y)$ defined by

$$f_{\xi,\theta}(\xi,\theta|Y=y) = \frac{f_Y(y|\xi,\theta)}{f_Y(y)} f_\xi(\xi) f_\theta(\theta). \tag{2}$$

It is supposed that the parameter and hyperparameter vectors are *a priori* independent. The likelihood of the observation, $f_Y(y|\xi,\theta)$, is then directly deduced from the distribution of the noise

$$f_Y(y|\xi,\theta) = (2\pi)^{-\frac{n}{2}} |C_\varepsilon(\theta)|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left\|y - M(\xi) - \bar{\varepsilon}(\theta)\right\|^2_{C_\varepsilon(\theta)^{-1}}\right), \tag{3}$$

hence the posterior distribution $f_{\xi,\theta}(\xi,\theta|Y=y)$ is known up to a normalizing constant, the marginal likelihood $f_Y(y)$.

In order to sample the ξ and θ according to the posterior distribution, we shall use the adaptive Metropolis algorithm [13]. It is a Markov Chain Monte Carlo (MCMC) method which enables sampling from any

arbitrary distribution provided that its density is known up to a constant. The main drawback of this sampling strategy is that it requires a large number of evaluations of the forward model $M(\xi)$. As a consequence, if one evaluation of the forward model requires the solution of a transient nonlinear PDE in multiple dimensions, the computational cost of the sampling strategy becomes prohibitive. This problem is circumvented by computing an approximation of the forward model with a truncated polynomial chaos expansion as proposed initially in [21].

3.2. Polynomial chaos acceleration of the inference

Let $(\Xi, \mathscr{B}_\Xi, \mu_\xi)$ be a probability space where $\mathscr{B}_\Xi$ is the Borel σ-algebra, and μ_ξ is the push-forward measure of $\mathbb{P}$ by ξ. We first assume that the quantity of interest mapping M has a finite second moment, meaning that

$$M \in L^2_{\mu_\xi}(\Xi, \mathbb{R}^n),$$

where $L^2_{\mu_\xi}(\Xi, \mathbb{R}^n) = \left\{ v : \Xi \to \mathbb{R}^n;\ \mathbb{E}\left(\|v\|^2 \right) < \infty \right\}$, $\mathbb{E}$ is the mathematical expectation, and $\|\cdot\|$ is the Euclidean norm. The space $L^2_{\mu_\xi}(\Xi, \mathbb{R}^n)$ is a Hilbert space when equipped with the inner product defined by $\langle u, v \rangle_{L^2} = \mathbb{E}(u^T v)$ and the associated norm. Let $(\psi_\alpha)_{\alpha \in \mathscr{I}}$ be a Hilbert basis of $L^2_{\mu_\xi}(\Xi, \mathbb{R})$ where $\mathscr{I}$ is a countably infinite set, an expansion of M is given by

$$M(\xi) = \sum_{\alpha \in \mathscr{I}} M_\alpha \psi_\alpha(\xi), \qquad M_\alpha \in \mathbb{R}^n,$$

where the sum converges in the L^2 sense [33, 8].

In the univariate case, a pertinent choice for the Hilbertian basis $(\psi_\alpha)_{\alpha \in \mathscr{I}}$ is the family of orthonormal polynomials with respect to the probability measure μ_ξ. This strategy corresponds to the generalized polynomial chaos expansion (PCE) of M which exhibits an exponential convergence rate when M is sufficiently smooth [34]. For the multivariate case, the definition of the basis depends on the independence of the variables $(\xi_i)_{i=1}^p$. In this work, these are assumed to be independent. The basis is thus the tensorization of univariate polynomials, exploiting the tensor product structure of the space $L^2_{\mu_\xi}(\Xi, \mathbb{R})$ [14].

An approximation of M is then obtained by truncating the series; consider a finite subset of indices $\mathscr{I}_m \subset \mathscr{I}$, the approximation is defined by

$$M(\xi) \approx M^m(\xi) = \sum_{\alpha \in \mathscr{I}_m} M_\alpha \psi_\alpha(\xi).$$

Once the vectors $(M_\alpha)_{\alpha \in \mathscr{I}_m}$ are computed, the forward model $M(\xi)$ is approximated by $M^m(\xi)$, called the surrogate model in the following, whose evaluation is inexpensive owing to the product form of the polynomials ψ_α. It enables efficient sampling of the posterior using MCMC methods.

3.3. Non-intrusive spectral projection

The construction of the surrogate model of M requires the calculation of its PC modes M_α. To this purpose, we will rely on a non-intrusive method, meaning that the M_α are determined for evaluations of the forward model $M(\xi)$ at selected values of ξ. Such an approach uses a numerical solver for $h(\xi)$ as a black box and no modifications of the original code are required. Several non-intrusive methods are available, including stochastic collocation methods [4] and regression techniques [6]. We choose here to use the non-intrusive spectral projection (NISP) method [27] which consists in writing the L^2 optimality condition for the approximation of M by the surrogate model. Formally, thanks to the L^2 orthogonality of the basis, the solution of the minimization problem

$$\min_{(M_\alpha)_{\alpha \in \mathscr{I}_m}} \mathbb{E}\left(\|M - \sum_{\alpha \in \mathscr{I}_m} M_\alpha \psi_\alpha\|^2 \right)$$

satisfies

$$M_\alpha = \frac{\mathbb{E}(M\psi_\alpha)}{\mathbb{E}(\psi_\alpha^2)}, \quad \forall \alpha \in \mathscr{I}_m.$$

The term $\mathbb{E}(\psi_\alpha^2)$ has an analytical expression when using classical orthogonal polynomials like Hermite or Legendre ones [1]. In the NISP method, a quadrature rule is introduced to estimate the correlations $\mathbb{E}(M\psi_\alpha)$. Denoting $(w_z)_{z\in\mathscr{Z}}$ and $(\xi_z)_{z\in\mathscr{Z}}$ the weights and points of the quadrature respectively, the vectors $(M_\alpha)_{\alpha\in\mathscr{I}_m}$ are estimated by

$$M_\alpha \approx \frac{1}{\mathbb{E}(\psi_\alpha^2)} \sum_{z\in\mathscr{Z}} w_z M(\xi_z)\psi_\alpha(\xi_z), \quad \forall \alpha \in \mathscr{I}_m.$$

It is seen that the method requires only the evaluation of the forward model at the quadrature points $(M(\xi_z))_{z\in\mathscr{Z}}$.

In the following, we will exploit the assumed independence of the variable ξ_i and the tensorized form of the polynomial $(\psi_\alpha)_{\alpha\in\mathscr{I}}$, using fully tensorized Gauss quadrature rules. In particular, the order of the Gauss quadrature rules are selected such that it prevents internal aliasing, by insuring

$$\sum_{z\in\mathscr{Z}} w_z \psi_\alpha(\xi_z)\psi_\beta(\xi_z) = \delta_{\alpha\beta}\mathbb{E}(\psi_\alpha^2), \quad \forall \alpha,\beta \in \mathscr{I}_m.$$

4. Noise modeling

In this section we illustrate the importance of selecting an appropriate noise model to achieve a correct estimation of the model parameter vector ξ. Equations (2), for the posterior distribution, and (3), for the likelihood, shows how the noise model intervenes in the definition of the posterior, in particular through the selected noise covariance $C_\varepsilon(\theta)$ and bias $\bar{\varepsilon}(\theta)$. We shall contrast two noise models: a white noise model and a correlated noise model, in Sections 4.1 and 4.2 respectively. In both cases, we rely on the same surrogate model of $M(\xi)$. The parameter ξ having a uniform prior, the surrogate model consists in a PC expansion of fully tensorized Legendre polynomials of a maximum partial degree 9. The expansion coefficients are computed by the NISP method discussed above, using the full tensorization of the 10 points Gauss-Legendre quadrature rules (integrating exactly polynomials with degree up to 19).

Regarding the sampling of the posterior distribution, the convergence of the Markov chains needs to be assessed. Besides the visual inspection of the simulated chains and the computation of their autocorrelation functions, we use in this work the Gelman-Rubin (GR) diagnostic [11]. This diagnostic will be also used to analyze the quality of the inference, depending on the noise model. The GR diagnostic requires the computation of several chains with different starting points. Let $(\xi^{q,k})_{q=1,k=1}^{q=Q,k=K}$ be Q chains of K samples. In the scalar case, $\xi^{q,k} \in \mathbb{R}$, the GR statistic compares the between-chain variance B and the within-chain variance W to assess the convergence of the chains. These variances are respectively given by

$$B = \frac{K}{Q-1}\sum_{q=1}^{Q}(\overline{\xi^q} - \overline{\xi})^2 \quad \text{and} \quad W = \frac{1}{Q}\sum_{q=1}^{Q} s_q^2,$$

where

$$\overline{\xi^q} = \frac{1}{K}\sum_{k=1}^{K}\xi^{q,k}, \quad \overline{\xi} = \frac{1}{Q}\sum_{q=1}^{Q}\overline{\xi^q},$$

$$\text{and} \quad s_q^2 = \frac{1}{K-1}\sum_{k=1}^{K}(\xi^{q,k} - \overline{\xi^q})^2.$$

An unbiased estimator of the posterior variance is

$$V = \frac{K-1}{K}W + \frac{1}{K}B.$$

The convergence of the chain is finally assessed by the potential scale reduction R defined by

$$R = \sqrt{\frac{V}{W}}.$$

It can be shown that for over dispersed starting points (with respect to the posterior), we have $R \geq 1$ and it converges to 1 with the number of samples K, provided that the sampler is robust enough. As a consequence, the case of $R \gg 1$ denotes a situation where the chains have not converged. If the symptom remains for $K \to \infty$, it means that the sampler is severely challenged and not able to properly visit the different posterior modes. In the case of random vectors, the GR statistic is computed for each component ξ_i.

4.1. White noise

We first analyze an unbiased white noise model in the inference of the parameter. This is the simplest and certainly mostly widely used noise model. The quantities $(\varepsilon_i)_{i=1}^n$ are independent and identically drawn from a normal distribution with mean 0 and variance σ^2, or more concisely $\varepsilon \sim \mathcal{N}(0, \sigma^2 \mathrm{I}_n)$, where I_n is the $n \times n$ identity matrix.

The standard deviation σ is defined through a single hyperparameter θ by

$$\sigma = 10^\theta, \quad \text{where} \quad \theta \sim \mathcal{U}(-5, 0).$$

This $\log_{10}$-uniform distribution for σ gives more importance to small values of σ with the property that $\mathbb{P}(10^a \leq \sigma \leq 10^b) = (b-a)/5$ for $-5 \leq a \leq b \leq 0$.

For this white noise model, $Q = 5$ chains are run with the adaptive Metropolis algorithm. The initial 10^5 samples are discarded and $K = 4 \times 10^5$ subsequent samples are recorded and used to compute the GR statistics for the 4 model parameters ξ_1 to ξ_4. The computed potential scale reductions are found

$$R = (19.09, 193.4, 55.27, 10.72).$$

Each component is far greater than 1, by one or two orders of magnitudes, and it can be concluded that the Q sampled chains are far off the posterior distribution. In fact, the chains are concentrating in different regions of the parameter domain, indicating that the posterior distribution is multi-modal. This is illustrated in Figure 4 where the joint density of the parameter pair (ξ_1, ξ_2) obtained for two chains with different starting points is demonstrated.

We observe that the length of the chains are large and that considering longer chains (larger K) would not help reducing the reported R values. Similarly, the MCMC sampler can hardly be blamed for being responsible for the convergence failure. Instead we claim that it is the poor quality of the noise model that causes an inference problem with a badly conditioned posterior. This claim is supported by the plots in Figure 5, where is plotted the discrepancy between the data and the GeoClaw prediction (top plot). This error is compared to a random sample of the Gaussian white noise as inferred by one of the chains (bottom plot). It is seen that if the standard deviation of the noise is correctly estimated, globally over the whole time domain, the noise model is still missing the temporal structures that are present in the actual discrepancy between the observation and the prediction. In particular we can distinguish two periods in time, before the arrival of the tsunami ($t \leq 3$ h) and after ($t > 3$ h). Hence the stationary assumption of the Gaussian white noise does not hold, and the inferred noise does not fit properly any of these two periods of time. A proper noise model is proposed below to remedy these problems and to ensure proper convergence of the MCMC sampler.

4.2. Noise design

In order to represent the measurement noise as well as the model error, we design in this section a suitable Gaussian process capable of accounting for the discrepancies between typical model predictions and the measurements, as illustrated in Figure 5 for the case of the prediction with nominal parameter values. As discussed previously, the time domain is separated into two phases, with separation time $t \approx 3$ h

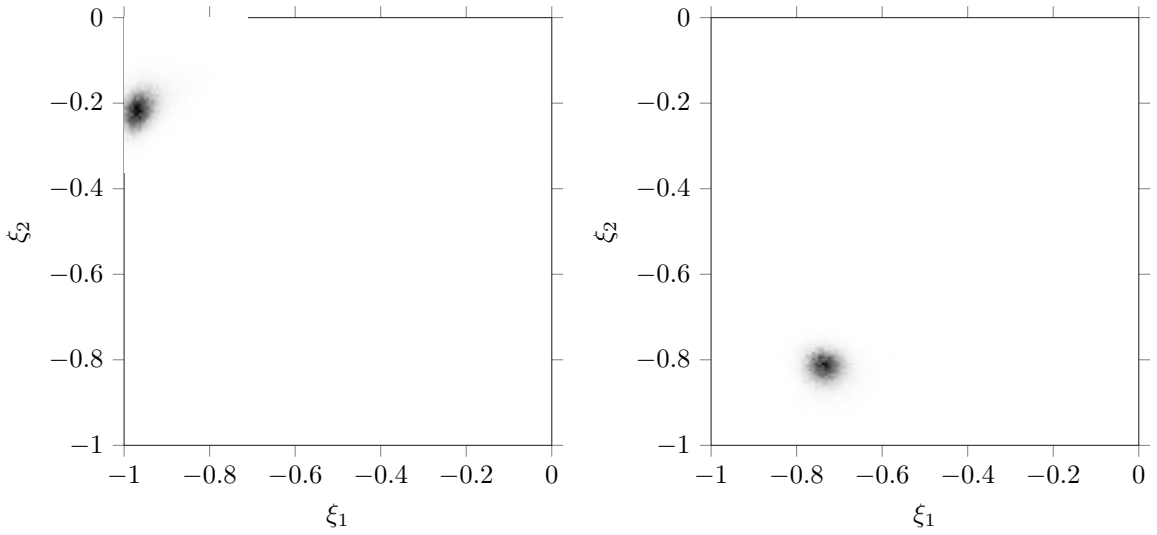

Figure 4: Joint density of the pair (ξ_1, ξ_2) for two different chains; shown is one fourth of the domain (lower left quadrant).

corresponding to the arrival time of the wave at the measurement buoy. Prior to the arrival of the waves, the discrepancies are low and can be properly modeled with a Gaussian noise. After the arrival of the waves, the discrepancy is on the contrary correlated with a significantly higher level. These features suggest a composite noise model.

To this end, we first arrange the measurement times in increasing order, that is $t_1 \leq t_2 \leq \ldots \leq t_n$, and denote by i_0 the largest integer such that $t_{i_0} \leq 3$ h. A white Gaussian noise with a non-zero mean is first considered over the whole time domain. This first contribution, denoted ε_W, involves two hyper parameters: a mean value θ_1 representing a measurement bias and a standard deviation σ_1. The bias θ_1 is given a uniform prior distribution, while the standard deviation σ_1 is given a $\log_{10}$-uniform distribution as in Section 4.1. The noise contribution at time t_i is then given by

$$\varepsilon_W(t_i) \sim \mathcal{N}(\theta_1, \sigma_1^2),$$

where $\theta_1 \sim \mathcal{U}(0, 0.025)$ and $\sigma_1 = 10^{\theta_2}$, $\theta_2 \sim \mathcal{U}(-5, 0)$.

For the correlated part, say $\epsilon_C(t)$, we assume that the model error is a centered Gaussian process. The process is then fully specified by its covariance. For simplicity, we shall assume the process to have zero variance for $t < 3$ h, and to be stationary for $t \geq 3$ h with a Matérn 3/2 covariance [26]. Denoting r_{ij} the time lag between two measurement times, $r_{ij} = |t_i - t_j|$, the process covariance can be thus written as

$$C_M(t_i, t_j; \sigma_2, T) = \sigma_2^2 \left(1 + \frac{\sqrt{3} r_{ij}}{T}\right) \exp\left(-\frac{\sqrt{3} r_{ij}}{T}\right),$$

when $\min(t_i, t_j) \geq 3$ and $C_M(t_i, t_j; \sigma_2, T) = 0$ otherwise. The covariance has two additional hyperparameters, the standard deviation σ_2, for $t \geq 3$ h, and a time scale T. The standard deviation is once more given a $\log_{10}$-uniform prior distribution, while the time scale is given a uniform prior distribution, as follows

$$\sigma_2 = 10^{\theta_3},\ \theta_3 \sim \mathcal{U}(-5, 0) \text{ and } T = \theta_4 \sim \mathcal{U}(5 \text{ s}, 500 \text{ s}).$$

Finally, the noise model is written as

$$\varepsilon(t) = \varepsilon_W(t) + \varepsilon_C(t).$$

The noise model now has a total of 4 hyperparameters $\theta_{i=1,\ldots,4}$.

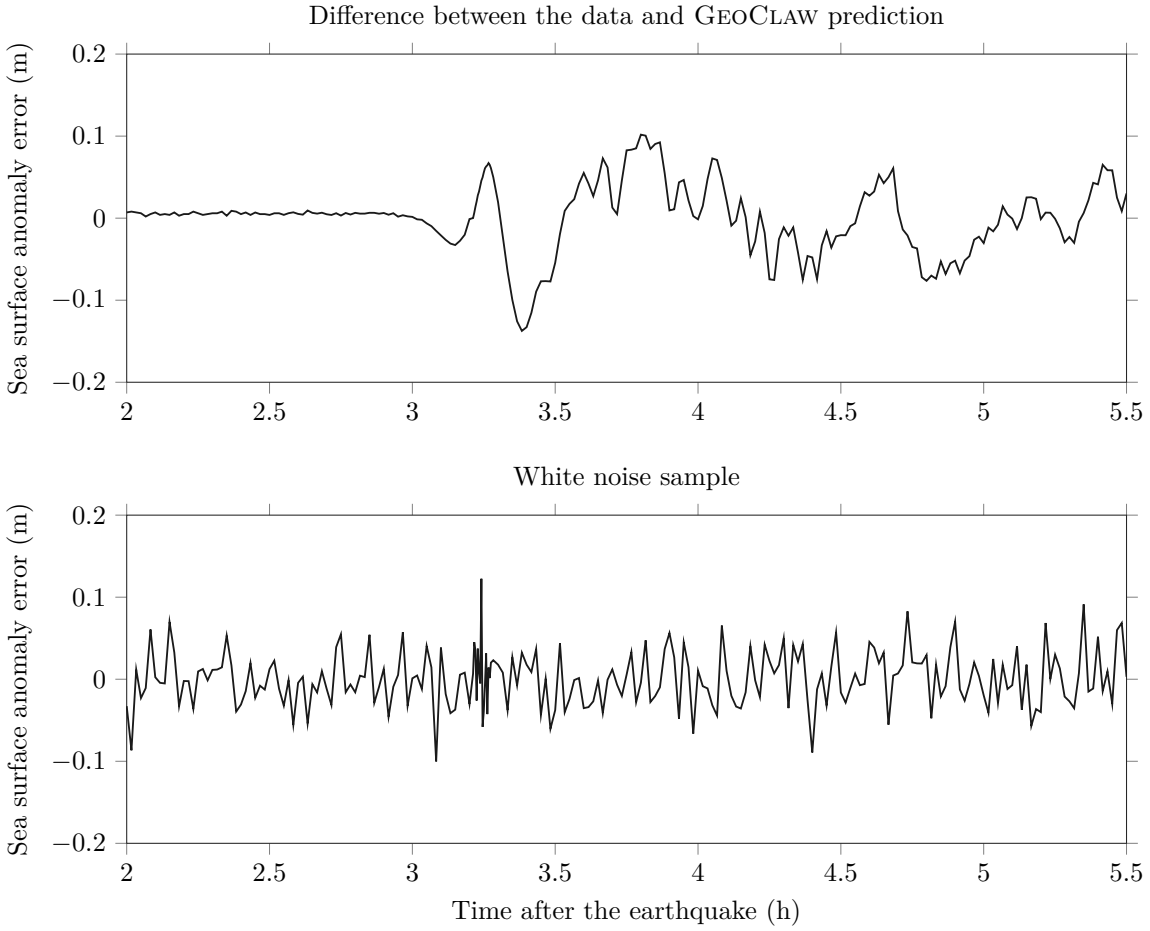

Figure 5: Error between the data and the GeoClaw prediction using nominal parameters $\xi = \xi^{\mathrm{GC}}$ (top plot), and a random sample of the Gaussian white noise model with standard deviation equal to 3.35×10^{-2}m (bottom plot).

Using the improved noise model, we compute the GR statistics using $Q = 5$ chains each with $K = 4 \times 10^5$ samples (after a burning period of 10^5 samples); the potential scale reductions are now equal to

$$R = (1.00200, 1.00188, 1.00127, 1.00087),$$

indicating that the MCMC samples are now well converged, in contrast with the previous experiment.

Being confident in the solution of the inference problem, we characterize its solution using the maximum a posteriori (MAP) values of the parameter ξ^{MAP} and the hyperparameter θ^{MAP}. These quantities maximize the posterior probability density function, thus satisfy

$$\xi^{\mathrm{MAP}}, \theta^{\mathrm{MAP}} = \arg\max_{\xi,\theta} f_{\xi,\theta}(\xi, \theta|Y).$$

Note that since the prior distributions are uniform, the MAP estimate is also the maximum likelihood estimate. The MAP values are computed with the differential evolution algorithm [31], a genetic type method. We used here the function implemented in the Python SciPy library [16]. The best solution computed among several runs is

$$\xi^{\mathrm{MAP}} \approx (-7.67 \times 10^{-1}, -9.41 \times 10^{-1}, 1.01 \times 10^{1}, 9.43)$$

and

$$\theta^{\mathrm{MAP}} \approx (5.57 \times 10^{-3}, -2.85, -1.43, 4.35 \times 10^{-2}). \tag{4}$$

The quality of the result can be qualitatively assessed with the marginals shown below in Figures 8, 9 and 10. Moreover, this MAP estimate is compared to the posterior distribution computed for each element of the Markov chains used at the beginning of Section 4.2. We deduce in particular that $\sigma_1^{\text{MAP}} \approx 1.40 \times 10^{-3}$ m and $\sigma_2^{\text{MAP}} \approx 3.71 \times 10^{-2}$ m. These values reflects the large differences in the discrepancy level before and after the arrival of the tsunami at the buoy. The MAP correlation time is $T^{\text{MAP}} \approx 4.35 \times 10^{-2}$ h $\approx 1.57 \times 10^{2}$ s.

In Figure 6, the model prediction for the MAP estimate ($\xi = \xi^{\text{MAP}}$) and the GeoClaw nominal value ($\xi = \xi^{\text{GC}}$) are compared to the data, taking the inferred bias θ_1^{MAP} into account for the MAP output. We remark that the MAP differs significantly from the nominal GeoClaw prediction. The latter appears to better predict the first extrema of the data signals, while The MAP prediction fits better the signal in average over the whole range of data (as measured by the likelihood). Moreover, the data are contained within the interval $M(\xi^{\text{MAP}}) \pm 3\ \sigma^{\text{MAP}}$ where $\sigma^{\text{MAP}}(t)$ is the inferred total standard deviation of the noise $\varepsilon(t)$ (summing σ_1^{MAP} and σ_2^{MAP}). In fact, it is likely that the GC nominal values are overfitting the noisy data.

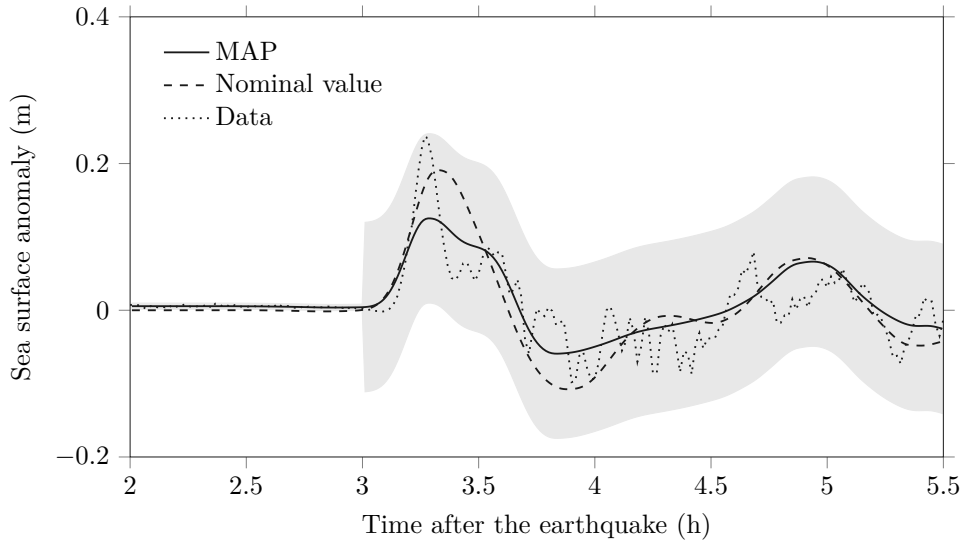

Figure 6: Comparison between the data and GeoClaw predictions with the MAP estimate and the nominal value. The bias is taken into account for the MAP output. The shaded area corresponds to $\theta_1^{\text{MAP}} + M(\xi^{\text{MAP}}) \pm 3\sigma^{\text{MAP}}$, where $\sigma^{\text{MAP}} = \sigma_1^{\text{MAP}} + \sigma_2^{\text{MAP}}$.

In Figure 7, a comparison of the actual error between the observations and numerical model predictions at $\xi = \xi^{\text{MAP}}$ is provided. A random sample of the noise $\varepsilon(t)$ with the MAP hyperparameter (θ^{MAP}) is also plotted for illustration. We observe a similarity between the actual prediction error and the sample of the noise model. In particular, the amplitude of the error and the correlation time are respected.

A more quantitative comparison of the two noise models can be obtained by computing their Akaike information criterion (AIC) and the Bayesian information criterion (BIC) [7]. These criteria are defined respectively by

$$AIC = -2\log L + 2(p+q),$$
$$\text{and} \quad BIC = -2\log L + (p+q)\log n,$$

where L is the maximum of the likelihood function, $p+q$ is the number of free parameters associated to the noise model and n is the number of observations. They both measure a trade-off between the goodness of fit ($-2\log L$) and the complexity of the model $p+q$ or $(p+q)\log n$. Note that $n = 223$, so $\log n = 5.41$ and therefore the BIC is more stringent than the AIC. The smaller the value of the criteria, the better the noise model to represent the data given the prediction model.

We denote by AIC_W (resp. AIC_C) the AIC associated to the white Gaussian noise model (resp. proposed non-stationary correlated noise model), and BIC_W and BIC_C their corresponding BIC. Computing the maximum likelihood L with the differential evolution algorithm, their values are

$$\begin{aligned} AIC_W &= -877.01, \quad AIC_C = -1508.46, \\ \text{and} \quad BIC_W &= -854.53, \quad BIC_C = -1479.08. \end{aligned}$$

These results clearly show that the proposed non-stationary noise is the preferred model. According to Kass and Raftery [17], the fact that $BIC_W - BIC_C > 10$ indicates that the evidence against the white Gaussian noise model is strong.

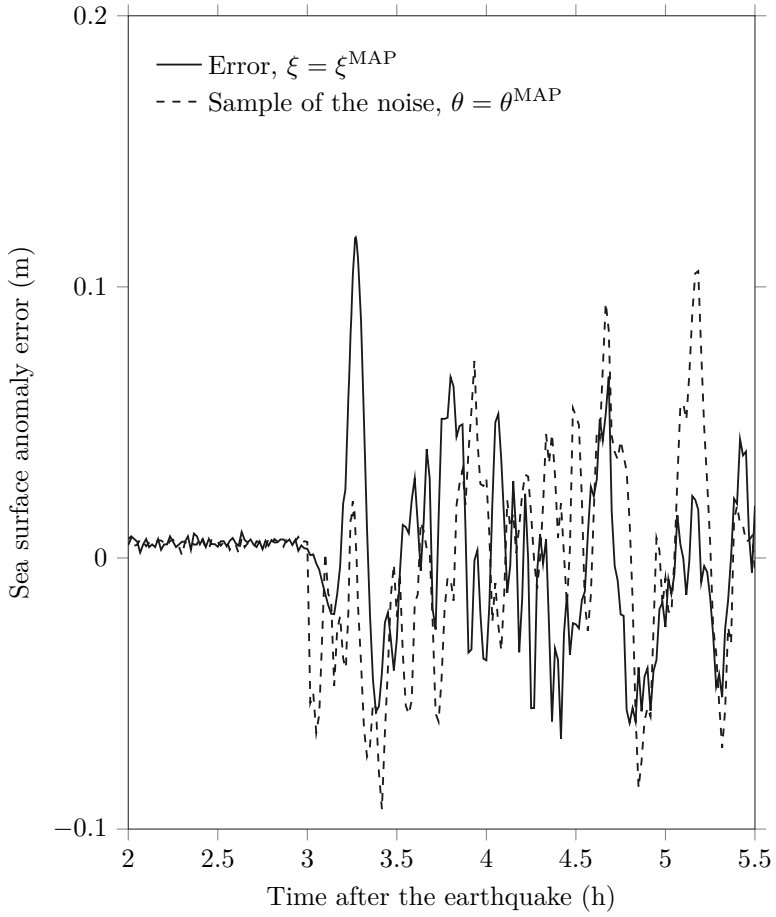

Figure 7: Comparison of the error between the data and the model prediction for $\xi = \xi^{\text{MAP}}$ (solid line) and a random sample of the inferred noise with hyperparameters $\theta = \theta^{\text{MAP}}$ (dashed line).

To complete this section, we provide the posterior marginal distributions of the inferred model parameters ξ and hyperparameters θ. These densities were estimated from the MCMC samples, using a standard kernel density estimation (KDE) method, in the case of the correlated noise model. Figure 8 shows the marginal posterior distribution of the earthquake location, in terms of canonical coordinates (ξ_1, ξ_2) (left plot) and corresponding latitude and longitude coordinates (right plot). The bimodal character of these inferred parameters is clearly visible, with two pronounced maximums in the marginal posterior. The MAP point roughly coincides with the maximum of the most northern pick, which is however located at the south of the GeoClaw nominal point. The unidimensional marginal posterior distributions of the model parameters, latitude, longitude, strike and slip, and hyperparameters, θ_i, are shown in Figures 9 and 10

respectively. Again, the multimodality of the posterior marginals of the parameters is clearly visible. In contrast, the posterior marginal distributions of the hyperparameters are highly concentrated around their MAP point, θ^{MAP}, confirming that the proposed noise model is accurately inferred and suitable to represent the discrepancy between the model prediction and the observation data.

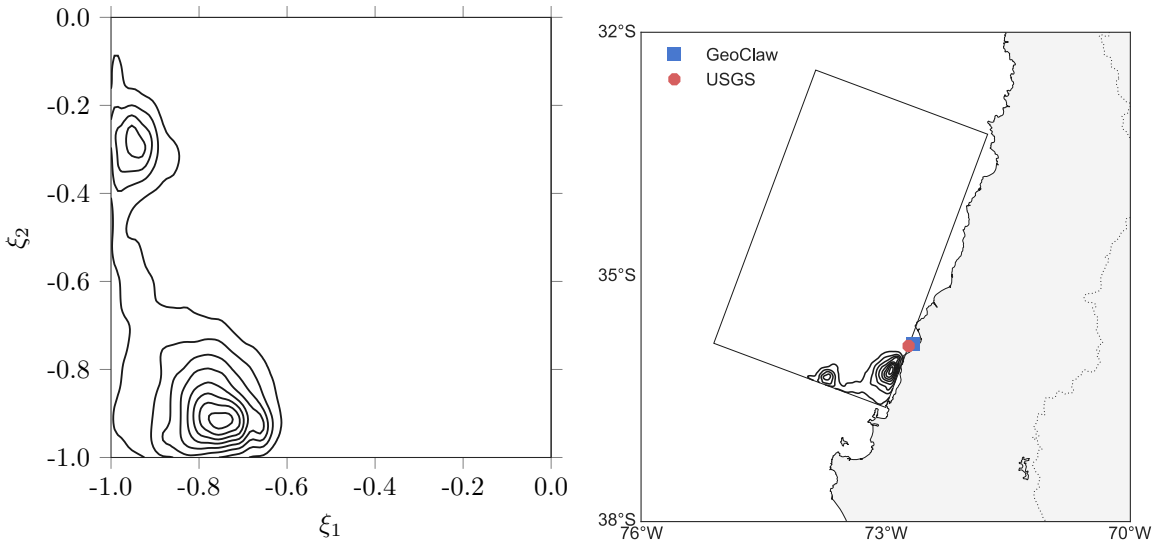

Figure 8: Kernel density estimation of the marginal posterior distribution f_{ξ_1,ξ_2} (top plot), and corresponding latitude and longitude coordinates.

5. Surrogate modeling

The objective of this section is to compute an accurate surrogate model of the signal at the buoy with respect to the posterior distribution. First, a novel non-intrusive spectral projection method is introduced in Section 5.1 that yields accurate metamodels for low-order polynomial chaos expansions by smoothing the discontinuity related to the arrival of the wave. The model is further reduced in Section 5.2 by selecting the features of the signal that are important for the inference. This is achieved by using a singular value decomposition with a suitable norm. The accuracy of the models with respect to the posterior distribution are assessed by using the Kullback-Leibler divergence, a standard tool in information theory.

5.1. Preconditioning of the polynomial chaos approximation

As mentioned above, one of the key features of the prediction of the sea-surface anomaly is the presence of a pre- and post wave-arrival periods. This feature reflects the wave propagation dynamics and the hyperbolic nature of the prediction model. For the present setting, the PC representation of the dependence with the parameter ξ of the model prediction is challenging. Indeed, it is well known that purely hyperbolic models can introduce discontinuities and shocks that compromise the spectral convergence of PC expansion on smooth bases, requiring some ad-hoc treatment such as multi-resolution schemes at the parametric level.

Such effects are illustrated in Figure 11, which compares the model prediction at a particular value of the parameters ξ and its truncated PC approximation (for a polynomial degree of 2). The model prediction consists in a constant elevation, until $t \approx 3$h, followed by a sharp increase of the sea surface anomaly and subsequent oscillations. The PC approximation exhibits spurious oscillations prior to the actual wave arrival time, that is for $t < 3$h. This is caused by a slow convergence of the PC approximation and Gibbs oscillations, as the sea surface anomaly is finite or exactly zero prior to the arrival time of the waves which varies with ξ. In addition, it is seen that at later time the predicted surface anomaly is not perfectly represented, because of complex dependences with ξ arising from multiple reflection waves.

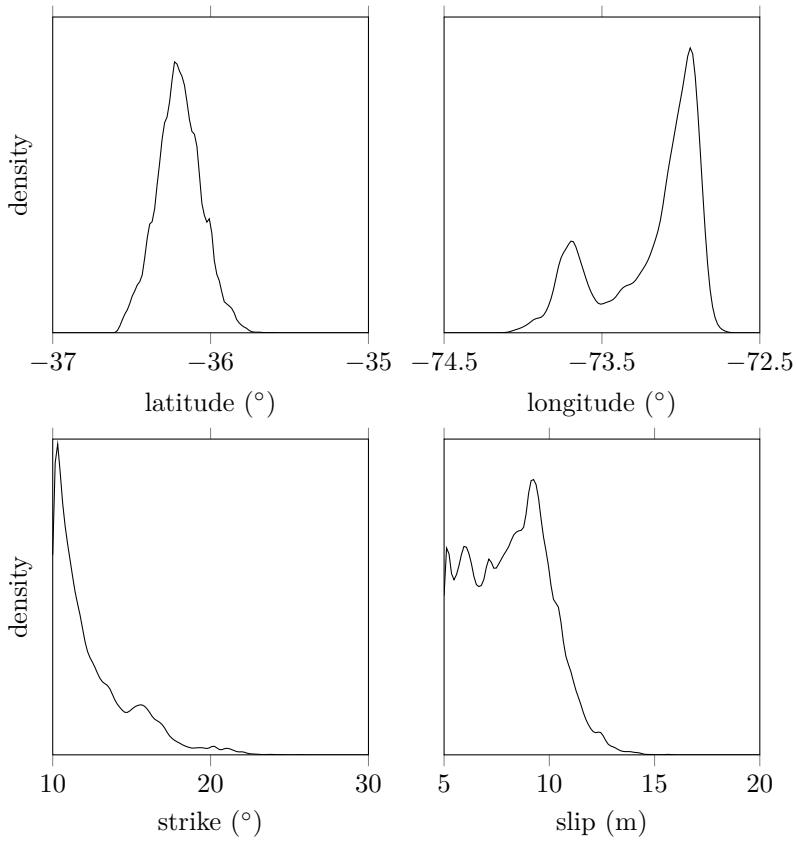

Figure 9: Marginal posterior distribution of model parameters as indicated.

We propose in this section a preconditioned NISP (PNISP) [2] strategy to improve the PC approximation of the model prediction in the neighborhood of the arrival time of the wave. The technique is based on two successive NISP approximations associated with an appropriate stochastic time shift which aims at smoothing the dependence with respect to ξ of the sea surface anomaly for $t \approx 3$h.

5.1.1. Preconditioned non-intrusive spectral projection

As before, we denote by $M(t;\xi)$ the predicted sea surface anomaly at parameter ξ and time t. We first introduce the arrival time random variable $a : \Xi \to \mathbb{R}$ defined by

$$a(\xi) = \min \left\{ t \in (2\text{ h}, 5.5\text{ h});\ M(t;\xi) \geq m_{\text{tol}} \right\}. \tag{5}$$

It is in fact the first time at which the sea surface anomaly predicted by the numerical model exceeds a prescribed tolerance $m_{\text{tol}} = 10^{-2}$ m. We have chosen the time interval such that the set

$$\{ t \in (2\text{ h}, 5.5\text{ h});\ M(t;\xi) \geq \text{tol} \}$$

is never empty and the arrival time is always defined.

The shifted (or reference) time τ is defined by $\tau = t - a(\xi)$, and the transformed signal $M_\tau(\tau;\xi)$ by

$$\begin{aligned} & M_\tau(\tau;\xi) = M(\tau + a(\xi);\xi), \\ \text{or} \quad & M(t;\xi) = M_\tau(t - a(\xi);\xi). \end{aligned}$$

The transformed signals are therefore synchronized, in the sense that $M_\tau(0;\xi) = M(a(\xi);\xi)$; hence, in the reference time τ, all the waves are arriving at $\tau = 0$ for every $\xi \in \Xi$. This implies that M_τ satisfies the

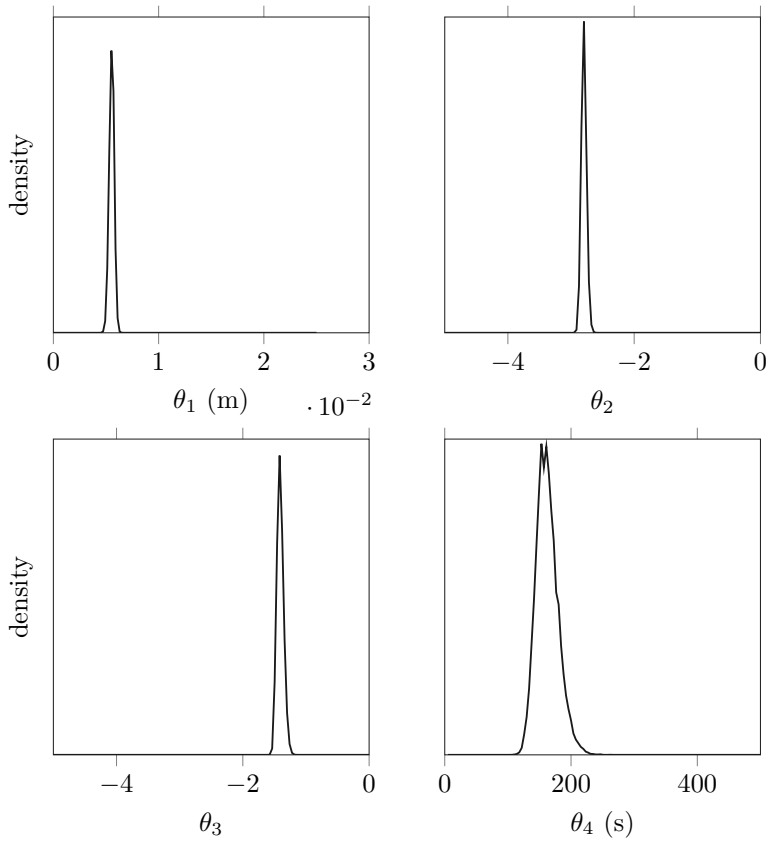

Figure 10: Marginal posterior distribution of the proposed correlated noise model hyperparameters θ_i as indicated.

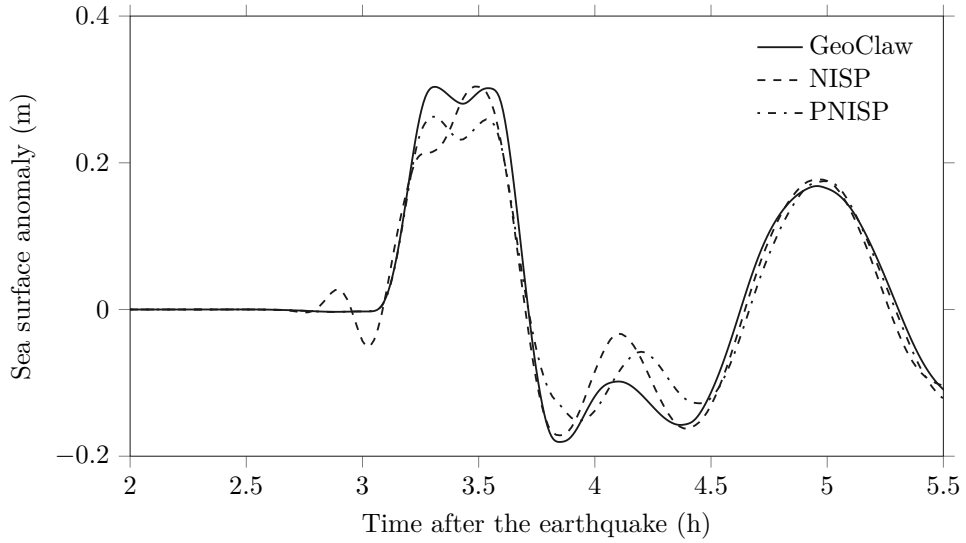

Figure 11: Comparison between the output of GeoClaw for a particular value of the parameter ξ and the evaluation of the corresponding truncated PCE of the model with degree 2 in all dimensions.

condition

$$M_\tau(0;\xi) = m_{\text{tol}}, \quad \forall \xi \in \Xi.$$

The synchronization procedure is illustrated in Figure 12. The left plot shows the model predictions M in the original time t for two sampled values of ξ. The right plot shows the corresponding transformed predictions M_τ, highlighting the synchronization of the two wave arrival times.

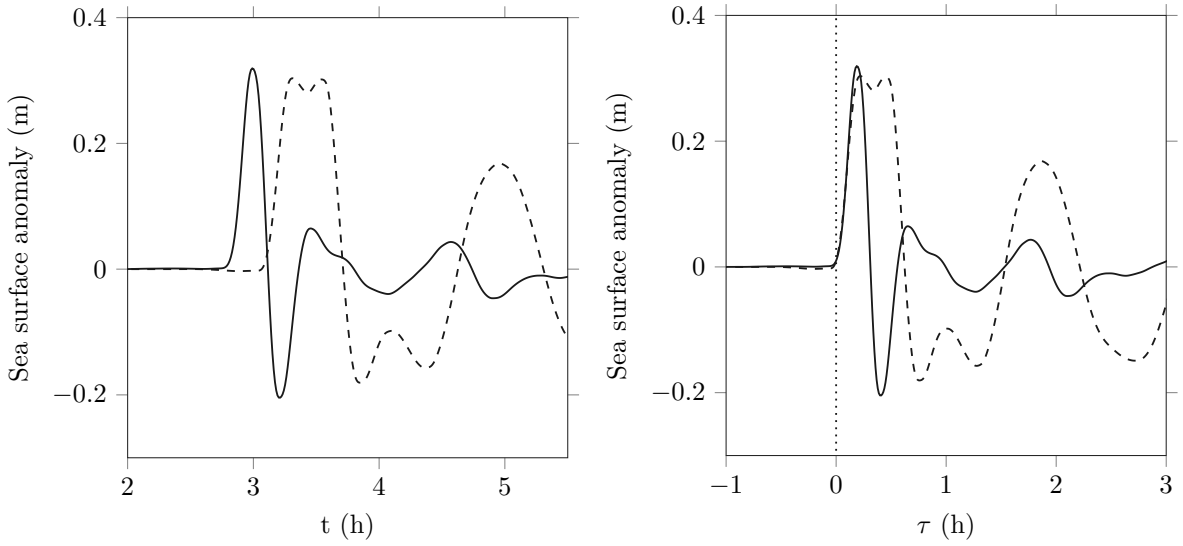

Figure 12: Evaluation of predictions M (left panel) and shifted predictions M_τ (right panel) for two sampled values of ξ.

In practice, the PNISP method uses the set of model predictions $M(t, \xi_z)$ at the quadrature points ξ_z, to first determine $a(\xi_z)$ and subsequently the PCE of the arrival time. With the notations introduced in Section 3.2, the PCE coefficients of $a(\xi)$ are given by

$$a_\alpha \approx \frac{1}{\mathbb{E}(\psi_\alpha^2)} \sum_{z \in \mathscr{Z}} w_z a(\xi_z) \psi_\alpha(\xi_z), \quad \forall \alpha \in \mathscr{I}_m.$$

The model prediction is finally approximated by the transformation

$$\begin{aligned} M(t;\xi) &\approx \widetilde{M}(t,\xi) \\ &= \sum_{\alpha \in \mathscr{I}_m} M_{\tau,\alpha} \left(t - \sum_{\alpha \in \mathscr{I}_m} a_\alpha \psi_\alpha(\xi) \right) \psi_\alpha(\xi), \end{aligned}$$

where

$$M_{\tau,\alpha}(\tau) = \frac{1}{\mathbb{E}(\psi_\alpha^2)} \sum_{z \in \mathscr{Z}} w_z M(t - a(\xi_z); \xi_z).$$

Note that the same PC basis is used for expanding both the arrival time a and prediction M. As before, we use bases $(\psi_\alpha)_{\alpha \in \mathscr{I}_m}$ corresponding to tensorized Legendre polynomials that are orthogonal with respect to the inner product associated to the uniform prior distributions, with associated fully tensorized Gauss-Legendre quadrature rules. The PNISP method also requires the introduction of an additional time mesh in the scaled time τ and interpolation procedures between discrete times (see [2] for more details).

The effect of the preconditioning can be appreciated in Figure 11 where the PNISP approximation is shown for the same ξ value as for the original prediction and direct NISP. It is seen that the pre-arrival time spurious oscillations have completely been removed, while the subsequent time oscillations remains of similar quality as for the direct NISP approximation. For a more qualitative comparison we introduce the

following relative L^2 error measure $\epsilon(t_f)$, between the original prediction $M(t,\xi)$ and its PC approximation $\widetilde{M}(t,\xi)$ for the NISP and PNISP methods. This error is defined for $t_f > 3$ h as

$$\epsilon^2(t_f) = \frac{\displaystyle\int_{2.5\text{ h}}^{t_f} \mathbb{E}\left(|M(\xi,t) - \widetilde{M}(\xi,t)|^2\right) dt}{\displaystyle\int_{2.5\text{ h}}^{t_f} \mathbb{E}\left(|M(\xi,t)|^2\right) dt}, \tag{6}$$

where the expectations are with respect to the prior distribution of the parameter ξ (uniform). In practice a Monte Carlo estimation is performed to estimate the relative error, using 10,000 sample points.

The results are reported in Figure 13 as a function of the analysis times t_f and for different PC expansion orders. We consider here the convergence with respect to the partial degree of the multivariate polynomial defined as the maximum degree of the univariate polynomials used in the tensorization. It is first observed that the approximation error decays for both NISP and PNISP expansions and all t_f, though it tends to level as the polynomial degree increases for the largest t_f shown. Also, the improvement carried by the preconditioning of the PC projection is seen to be more significant for low expansion order, while the two approaches tend to be similarly accurate at higher orders. Furthermore, the improvement is more pronounced for the lowest t_f, that is close to the waves arrival time, than for longer analysis times as synchronicity is progressively lost.

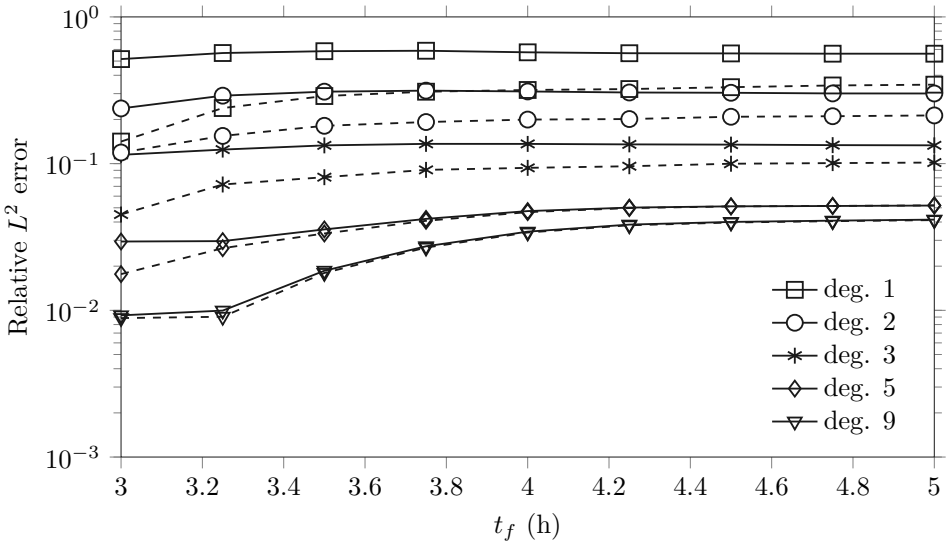

Figure 13: Relative L^2 error norm $\epsilon(t_f)$ (see (6)) for the NISP (solid lines) and PNISP (dashed lines) approximations, for different maximum partial degree of the tensorized Legendre polynomials basis.

In order to analyze the convergence of the truncated PCE $\widetilde{a}(\xi)$ computed by NISP to the arrival time $a(\xi)$, we introduce the relative L^2 error ϵ_a defined by

$$\epsilon_a^2 = \frac{\mathbb{E}\left((a(\xi) - \widetilde{a}(\xi))^2\right)}{\mathbb{E}\left(a(\xi)^2\right)}.$$

The convergence of the L^2 error with respect to the maximum partial degree of the polynomial is illustrated in Figure 14, where the expectations are estimated from a 10,000 random samples drawn according to the prior distribution. An order 1 approximation of the arrival time yields a relative L^2 error of 3.03×10^{-3}. This indicates that the dependence of the arrival time on the parameter is mostly affine. Moreover, the empirical mean and standard deviation of the arrival time are respectively 2 h 56 min and 7 min, which suggests that the arrival time is not constant and the linear terms are important.

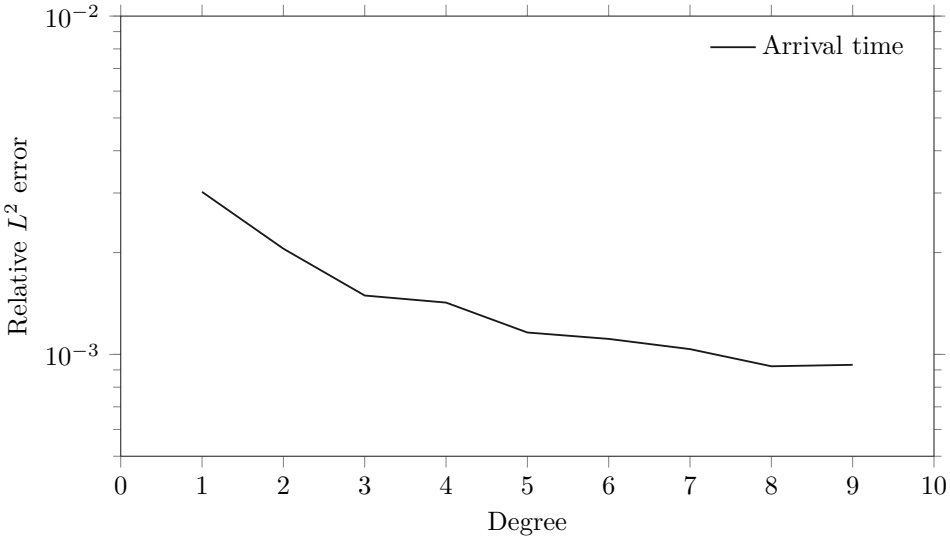

Figure 14: Convergence of the relative L^2 error norm for the NISP of the arrival time with respect to the maximum partial degree of the tensorized Legendre polynomials basis.

5.1.2. Inference results

It can be concluded from the previous numerical results that the preconditioning improves the quality of the PC approximation, principally in the neighborhood of the arrival time where it allows for an eventual significant PC order reduction. For the whole observation range, that is large t_f, the effect can appear disappointing with a less significant improvement reported, but we stress that the PNISP and NISP methods have the same computational complexity as they use the same set of quadrature points. However, we claim that having an accurate PC surrogate of the predictions around the arrival time is crucial for the inference problem. This can be anticipated by the inference results reported in the previous section, where the large level of inferred noise, after the arrival time, indicates that the latest observations are not very informative. To prove this claim on the dominant importance of the model accuracy around the arrival time of the wave, we estimate the Kullback-Leibler divergence (KLD) [18] between the prior and posterior distributions, for the preconditioned and direct NISP PC surrogates and different expansion orders. Given two distributions f_0 and f_1, the KLD $D_{\mathrm{KL}}\left(f_0||f_1\right)$ of f_1 from f_0 measures the information gain achieved by using the distribution f_1 instead of f_0. It is defined by

$$\begin{aligned} D_{\mathrm{KL}}\left(f_0||f_1\right) &= \int_{\Xi}\int_{\Theta} f_0(\xi,\theta)\log\frac{f_0(\xi,\theta)}{f_1(\xi,\theta)}\,\mathrm{d}\xi\,\mathrm{d}\theta \\ &= \mathbb{E}_{f_0}\left(\log\frac{f_0}{f_1}\right), \end{aligned}$$

with the particular property that $D_{\mathrm{KL}}\left(f_0||f_1\right) = 0 \Leftrightarrow f_0 = f_1$.

Given a PC model $\widetilde{M}$, we are interested in the KLD between the corresponding posterior distribution

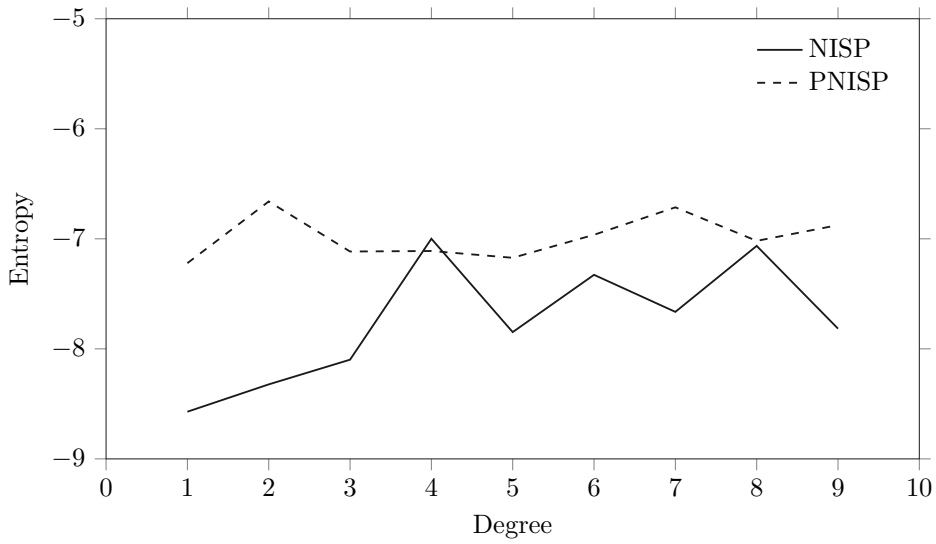

Figure 15: Estimates of the posterior entropy with respect to the maximum partial degree of the PC basis and NISP and PNISP methods as indicated.

$f_{\xi,\theta}(\cdot|Y=y,\widetilde{M})$ and the prior distribution $f_{\xi,\theta}$. The KLD is then

$$D_{\mathrm{KL}}\left(f_{\xi,\theta}(\cdot|Y=y,\widetilde{M})||f_{\xi,\theta}\right) = \int_{\Xi}\int_{\Theta} f_{\xi,\theta}(\xi,\theta|Y=y,\widetilde{M})\log f_{\xi,\theta}(\xi,\theta|Y=y,\widetilde{M})\mathrm{d}\xi\ \mathrm{d}\theta - \int_{\Xi}\int_{\Theta} f_{\xi,\theta}(\xi,\theta|Y=y,\widetilde{M})\log f_{\xi,\theta}(\xi,\theta)\mathrm{d}\xi\ \mathrm{d}\theta = -H(f_{\xi,\theta}(\cdot|Y=y,\widetilde{M})) + C,$$

where $H(f) = -\int_{\Xi} f(\xi)\log f(\xi)\mathrm{d}\xi$ is the entropy of f and C is a constant due to the uniform priors. Hence, we can characterize the information gain between the posterior and the prior distributions by computing the entropy of the posterior distributions.

Concerning the practical estimation of a posterior entropy, a Markov chain of 4×10^5 samples is generated with the corresponding metamodel $\widetilde{M}$, after a burn-in period of 10^5 samples. The samples are used to estimate the posterior distribution by means of a Kernel Density Estimation method [24], and subsequently the expectation $H(f) = -\mathbb{E}_f(\log f)$ from the sample set.

These posterior entropy estimates are shown in Figure 15, for varying maximum partial degree of the PC basis and for the NISP and PNISP methods. It is seen that the preconditioning yields entropy estimates that depends much less on the polynomial degree compared to the NISP-model based estimates. In other words, the inference is more robust at low PC expansion orders for the PNISP method than for the NISP method, and the faster convergence of the entropy for PNISP reflects the importance of the information related to the wave arrival time.

5.2. Inference from reduced models

The previous experiment has evidenced that some features of the model M are more important than others for the purpose of inferring ξ from the observations. In this section, reduced models for M_τ are proposed as alternative to the full model in the inference problem. In particular, we will reduce the discretized model M_τ based on singular value decomposition (SVD). The goal is to compress the model, by means of a decomposition into a few elementary signals, in a way that the inferred posterior distribution is preserved,

i.e. the reduced model still contains meaningful information about the inferred parameter ξ. In all the numerical experiments proposed in this section, the hyperparameter θ is set to θ^{MAP} (see Equation (4)) and we use PC expansions with partial degree 9.

5.2.1. Model reduction

The shifted time τ is discretized on a fixed uniform grid $(\tau_i)_{i=1}^N$. We shall denote $\mathrm{M}(\xi)$ the random vector with components $\mathrm{M}_i(\xi) = M_\tau(\tau_i, \xi)$. Let $A \in \mathbb{R}^{N\times N}$ be a symmetric positive definite matrix, $\langle \cdot, \cdot \rangle_A$ be the associated inner product defined by $\langle u, v\rangle_A = u^T A^{-1} v$ and $\| \cdot \|_A$ be the induced norm. The vector $\mathrm{M}(\xi)$ is assumed to be A-square integrable, meaning that $\mathbb{E}_\xi(\|\mathrm{M}\|_A^2) < \infty$. In this section, we denote $\mathbb{E}_\xi(\cdot)$ the expectation with respect to the prior distribution of ξ, that is $f_\xi(\xi)$.

Now, we would like to approximate M with a rank-r approximation, under the form

$$\mathrm{M}^r(\xi) = \sum_{i=1}^{r} \sigma_i Z_i \lambda_i(\xi),$$

where $\sigma_i \geq 0$, $Z_i \in \mathbb{R}^N$ and $\lambda_i \in L^2(\Xi)$. Given the tensor structure of the space of A-square integrable functions [14], the optimal low-rank approximation with respect to the norm $(\mathbb{E}_\xi(\|\cdot\|_A^2))^{\frac{1}{2}}$ is given by a truncated singular value decomposition.

The dominant left singular vectors $(Z_i)_{i=1}^r$ and singular values $(\sigma_i)_{i=1}^r$ are computed by taking the r dominant eigenpairs (Z, σ^2) of the eigenvalue problem

$$\mathbb{E}_\xi(\mathrm{M} \langle \mathrm{M}, Z\rangle_A) = \sigma^2 Z \quad \Leftrightarrow \quad \mathbb{E}_\xi(\mathrm{M}\mathrm{M}^T) A^{-1} Z = \sigma^2 Z.$$

The matrix A being symmetric and positive definite, the eigenvalue problem can be recast as

$$\mathbb{E}_\xi(\mathrm{M}\mathrm{M}^T) A^{-1} Z = \sigma^2 Z,$$

and subsequently by introducing the change of variable $\widetilde{Z} = A^{-1} Z$, yielding the generalized eigenvalue problem

$$\mathbb{E}_\xi(\mathrm{M}\mathrm{M}^T) \widetilde{Z} = \sigma^2 A \widetilde{Z}, \quad Z = A\widetilde{Z}.$$

Due to the orthogonality of the singular vectors, the dominant right singular vectors $(\lambda_i)_{i=1}^r$ are given for $\sigma_i > 0$ by

$$\lambda_i(\xi) = \frac{1}{\sigma_i} \langle Z_i, \mathrm{M}\rangle_A .$$

The matrix $\mathbb{E}_\xi(\mathrm{M}\mathrm{M}^T)$ is computed from the polynomial chaos expansion $\mathrm{M}(\xi) = \sum_{\alpha\in\mathscr{I}_m} \mathrm{M}_\alpha \psi_\alpha(\xi)$. Given the fact that $(\psi_\alpha)_{\alpha\in\mathscr{I}_m}$ is an orthogonal family for the inner product $\langle \cdot, \cdot \rangle_{L^2}$, we have

$$\begin{aligned} \mathbb{E}_\xi\left(\mathrm{M}\mathrm{M}^T\right) &= \sum_{\alpha\in\mathscr{I}_m} \sum_{\beta\in\mathscr{I}_m} \mathrm{M}_\alpha \mathrm{M}_\beta^T \langle \psi_\alpha, \psi_\beta\rangle_{L^2} \\ &= \sum_{\alpha\in\mathscr{I}_m} \mathrm{M}_\alpha \mathrm{M}_\alpha^T \|\psi_\alpha\|_{L^2}^2 . \end{aligned}$$

Concerning the metric A^{-1}, we consider the following two possibilities. The first one is the identity, $A = \mathrm{I}$, that leads to the standard SVD. In fact, this choice corresponds to the natural discretization of the L^2 inner product in the τ-range, since the discretization of the shifted time is uniform. It means that all discrete times τ_i are given the same importance. The associated basis will be denoted by $(Z_i^{\mathrm{I}})_{i=1}^r$.

The second case corresponds to the Mahalanobis distance and is based on the covariance of the Gaussian noise ε_τ, discretized at the time τ_i. The Gaussian noise ε_τ is related to the previous correlated noise model ε of Section 4.2 by

$$\varepsilon_\tau(\tau) = \varepsilon(\tau + 3\ \mathrm{h}).$$

Thus, the correlated noise with the Matérn covariance starts at $\tau = 0$. The norm induced by this choice favors the components in M that are associated to a low noise level and that can then be properly inferred. Therefore, this is the natural norm to reduce the forward model in view of solving the inference problem, because of the form of the Bayesian model in Equation (1) and the assumed noise structure. Indeed, the posterior distribution is such that

$$f_\xi(\xi|Y = y, \theta = \theta^{\text{MAP}}) \propto \exp(-\frac{1}{2}\|y - M(\xi) - m_\varepsilon\|_\varepsilon^2),$$

where $\|\cdot\|_\varepsilon = \|\cdot\|_{\text{Cov}(\varepsilon_\tau)}$ is the Mahalanobis distance. The basis associated to the Mahalanobis distance will be denoted by $(Z_i^\varepsilon)_{i=1}^r$.

The normalized dominant singular values σ_i^I/σ_1^I and $\sigma_i^\varepsilon/\sigma_1^\varepsilon$ are plotted in Figure 16. They exhibit a similar decay from $i = 1$ to $i = 30$. Beyond that point, the normalized singular values associated with the Mahalanobis norm exhibit a slower decay. It is important to note that the decay of the singular values are not an indicator of the efficiency of the approach since we are interested in the estimation of the posterior distribution, and not the model M.

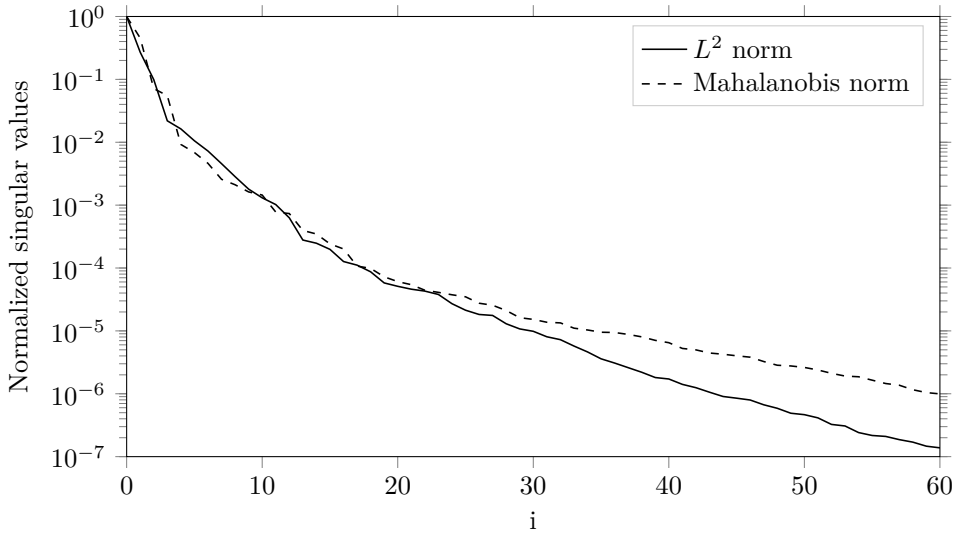

Figure 16: Normalized dominant singular values for the L^2 and Mahalanobis norms.

The dominant eigenvectors Z_i^{I} and Z_i^ε are reported in Figure 17. These eigenvectors have been scaled such that

$$\max_j |Z_i^{\text{I}}(\tau_j)| = \max_j |Z_i^\varepsilon(\tau_j)| = 1.$$

The eigenvectors having arbitrary orientation, we fix the signs of the $(Z_i^\varepsilon)_{i=1}^r$ to maximize the alignment between Z_i^{I} and Z_i^ε. This is achieved by ensuring

$$(Z_i^{\text{I}})^T Z_i^\varepsilon \geq 0.$$

It is seen that the vectors Z_i^{I} and Z_i^ε are similar for $i \leq 13$. Starting from $i = 14$, the vectors present different features, with visible differences in local amplitudes and frequencies.

5.2.2. Inference results

We now perform the Bayesian inference using the reduced models $\text{M}^r(\xi)$ for different ranks r and the two norms of the decomposition. It is clear that for $r < N$, the subspace spanned by the r dominant mode depends in general on the selected norm, and we expect the resulting posterior distribution to be consequently affected.

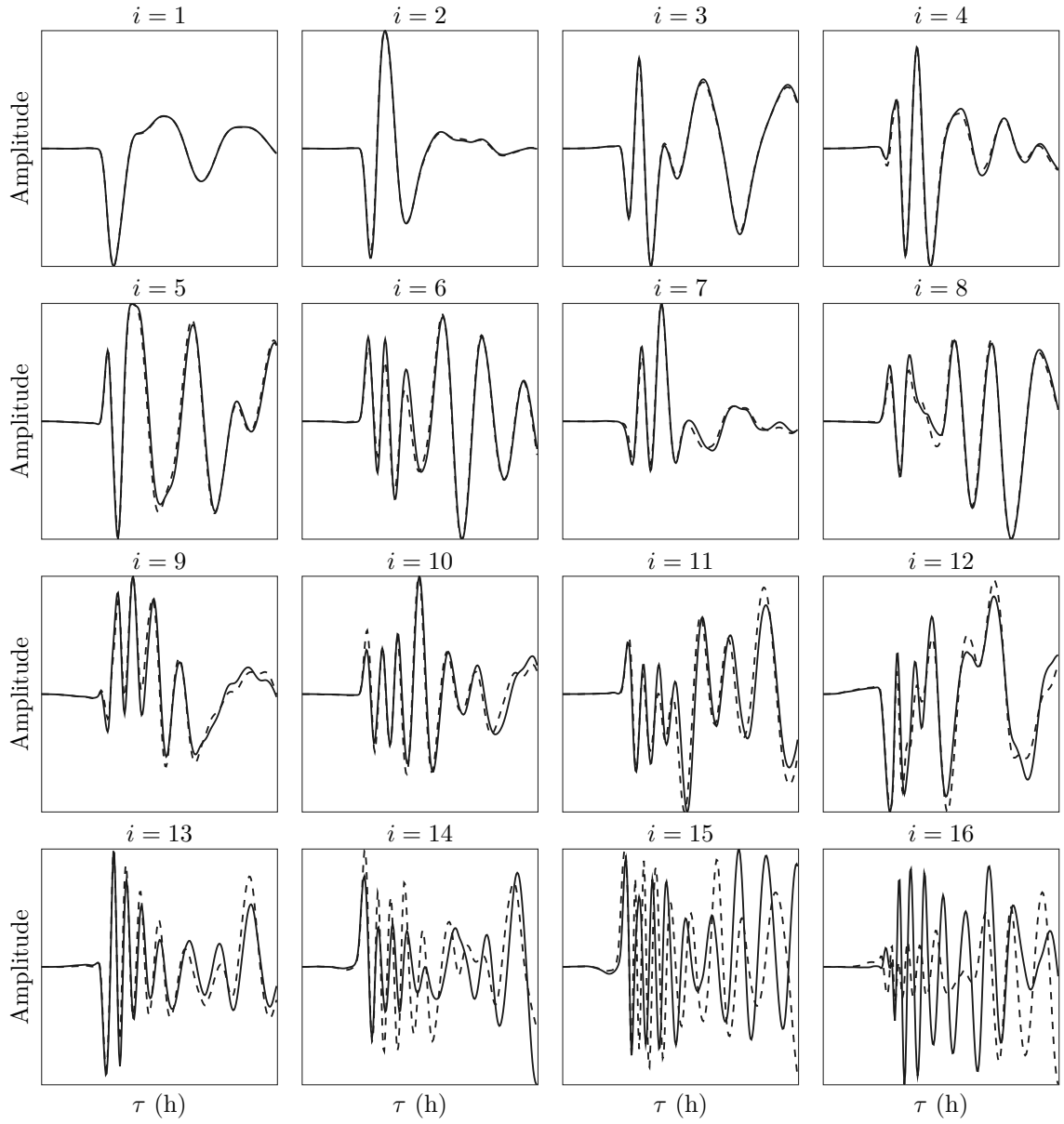

Figure 17: First 16 dominant eigenvectors Z_i^{I} and Z_i^{ε}, corresponding to the classical L^2 norm (solid line) and the Mahalanobis norm (dashed line) respectively.

As before, we start by comparing the posterior distributions for the different reduced models by computing their entropy. The convergence of the entropy with respect to the size r of the truncated bases is shown in Figure 18. The entropies for the two reduction norm quickly converge to the reference entropy, the latter being computed from the full model obtained by the PNISP method with degree 9. This convergence means that the shape of the posterior distribution is well-captured with few tens of eigenvectors for the two norms. It is also remarked that the reduced model for the Mahalanobis distance seems to converge faster with r compared to the classical L^2 norm model. Note that the values of the entropies differ between Figures 15 and 18. This is due to the use of different expectations in the definition of the entropy, the usual one in Section 5.1 and the conditional expectation in this section, i.e. $\mathbb{E}(\cdot\,|\theta = \theta^{\mathrm{MAP}})$.

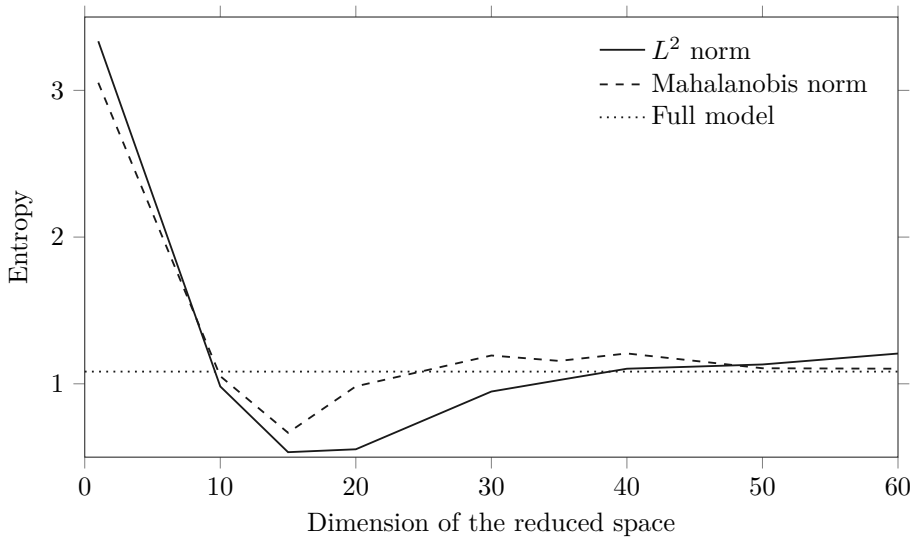

Figure 18: Convergence of the entropy of the posterior distribution with respect to the size r of the reduce models.

The entropy measures the information content of the posterior distribution, and thus only characterize the shape. As an example, the multivariate Gaussian distribution f_η, with values in $\mathbb{R}^N$ and covariance C_η, has an entropy

$$H(f_\eta) = \frac{N}{2}(1 + \log(2\pi)) + \frac{1}{2}\log|C_\eta|,$$

which depends only on the covariance C_η. To further verify the convergence of posterior distributions, with the size r of the reduced model, we also compare the convergence of the (Euclidian) distance between the posterior means. Let $\overline{\xi}$ and $\overline{\xi}_r$ denote the posterior distribution means associated to the full and reduced models respectively, the error in mean is then defined by

$$e_r = \frac{\left\|\overline{\xi}_r - \overline{\xi}\right\|}{\left\|\overline{\xi}\right\|}.$$

The convergence of e_r is illustrated in Figure 19 for the two reduction norms. We can see that the Mahalanobis distance greatly improves the convergence with r of the posterior mean, compared to the L^2 norm. In particular, for $r = 15$, the reduced model based on the standard L^2 norm is not able to locate properly the posterior distribution, with a relative error on the mean $e_r^{\mathrm{I}} = 28.8\%$. In contrast, the error on the mean for the Mahalanobis norm is $e_r^{\varepsilon} = 3.53\%$.

These numerical experiments demonstrate that selecting carefully the norm to construct the reduced model can have an important effect on the dimension of the reduced model needed to achieved a certain reduction error on the inferred parameters, as measured here by the entropy and mean of the posterior distribution. At a more fundamental level, these examples highlight the fact that not all features of a model are useful for the inference task, nor it is important that the model be accurate over the whole range of observations. Instead, our simulations suggest to consider reduced models involving few carefully selected features, with the possibility to alleviate the computational times involved in its construction and evaluation.

6. Conclusion

We investigate a computational approach to propagate uncertainty in the parameters of the Okada model of the 2010 Chile earthquake, and to calibrate the uncertain parameters based on data collected by a DART® station.

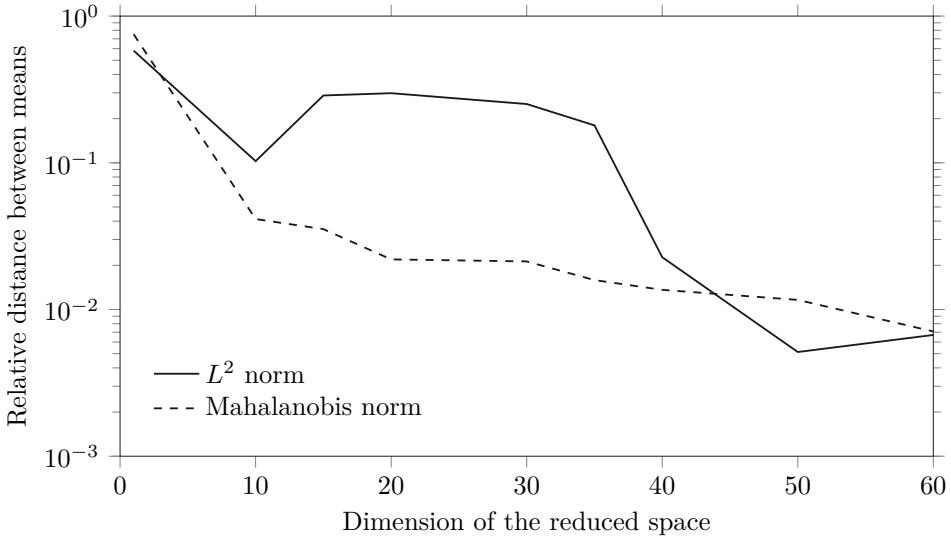

Figure 19: Convergence of the mean of the posterior distribution with respect to the size r of the reduced models.

The methodology consists in first establishing an approximation of the model response using a polynomial chaos expansion. In order to calibrate the uncertain parameters based on the data collected by the DART® station using a Bayesian formalism, we have first explored the use of an additive Gaussian noise model in the likelihood function. A more elaborate likelihood function was also considered by incorporating a non-stationary correlated Gaussian noise model that takes into account the two different regimes observed in the measurements, namely before and after the arrival of the tsunami wave. We show that considering a suitable noise model is crucial to achieve a robust and reliable parameter inference. In particular, the inferred solution obtained using the proposed correlated noise model outperforms the solution obtained using classical white noise assumption in terms of Akaike and Bayesian information criteria; it also stabilizes the MCMC sampling from the corresponding posterior distribution.

A preconditioned non-intrusive spectral projection strategy is also proposed to improve the accuracy of the PC model regarding the arrival time of the tsunami. Specifically, the polynomial chaos expansion of the arrival time of the tsunami wave at the buoy location is first computed. The resulting representation is then used to construct a time transform (shift) that smoothes the representation of the local wave-height signal. A polynomial chaos expansion of the signal is then developed as a function of the shifted time variable. Analysis of computations indicates that for low-order expansions, the preconditioned PC representation is more accurate in term of the L^2 norm than the same-order expansion obtained with direct application of a non-intrusive spectral projection approach. Moreover, using the preconditioned representation, the convergence to the posterior distribution is more stable with respect to the Shannon entropy than is observed with a direct NISP approach.

We finally explored the possibility of performing the Bayesian inference based on reduced PC model representation determined using optimal decompositions. Two approaches were considered. In the first, the decomposition was applied by relying on the L^2 norm (in time) optimality, whereas the second relied on the Mahalanobis distance, i.e. on the metric induced by the inverse of the covariance of the noise. The analysis showed that the second reduction method stabilizes the convergence of the posterior distribution with respect to the entropy and to the mean posterior as well. On the other hand, the use of the standard L^2 norm degrades the convergence rate with the reduced model dimension, especially with respect to the posterior mean.

Future works will focus on a goal-oriented model reduction applied to Bayesian inference. The objective is to be able to compute a reduced forward model that preserves the features of the posterior distribution. In particular, we are interested in the development of variational principles applied to model reduction based

on the information theory, involving the Shannon entropy and the Kullback-Leibler divergence.

Acknowledgments

This work is supported by King Abdullah University of Science and Technology Award CRG3-2156.

[1] Abramowitz, M., Stegun, I.A.: Handbook of mathematical functions: with formulas, graphs, and mathematical tables, vol. 55. Courier Corporation (1964)

[2] Alexanderian, A., Le Maître, O.P., Najm, H.N., Iskandarani, M., Knio, O.M.: Multiscale stochastic preconditioners in non-intrusive spectral projection. Journal of Scientific Computing **50**(2), 306–340 (2012). DOI 10.1007/s10915-011-9486-2. URL http://dx.doi.org/10.1007/s10915-011-9486-2

[3] American Red Cross Multi-Disciplinary Team: Report on the 2010 Chilean earthquake and tsunami response. Tech. rep., USGS Survey (2011). URL http://pubs.usgs.gov/of/2011/1053/

[4] Babuška, I., Nobile, F., Tempone, R.: A stochastic collocation method for elliptic partial differential equations with random input data. SIAM Review **52**(2), 317–355 (2010). DOI 10.1137/100786356. URL http://dx.doi.org/10.1137/100786356

[5] Berger, M.J., George, D.L., LeVeque, R.J., Mandli, K.T.: The GeoClaw software for depth-averaged flows with adaptive refinement. Advances in Water Resources **34**(9), 1195–1206 (2011). DOI 10.1016/j.advwatres.2011.02.016

[6] Blatman, G., Sudret, B.: Sparse polynomial chaos expansions and adaptive stochastic finite elements using a regression approach. Comptes Rendus Mécanique **336**(6), 518–523 (2008). DOI 10.1016/j.crme.2008.02.013. URL http://apps.webofknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=49&SID=P13PgLmH5nFlA54846F&page=1&doc=2&cacheurlFromRightClick=nohttp://linkinghub.elsevier.com/retrieve/pii/S1631072108000582

[7] Burnham, K.P., Anderson, D.R. (eds.): Model Selection and Multimodel Inference. Springer New York, New York, NY (2004). DOI 10.1007/b97636. URL http://link.springer.com/10.1007/b97636

[8] Cameron, R.H., Martin, W.T.: The orthogonal development of non-linear functionals in series of fourier-hermite functionals. The Annals of Mathematics **48**(2), 385–392 (1947)

[9] Das, S.K., Lardner, R.W.: Variational parameter estimation for a two-dimensional numerical tidal model. International Journal for Numerical Methods in Fluids **15**(3), 313–327 (1992)

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

[11] Gelman, A., Rubin, D.B.: Inference from iterative simulation using multiple sequences. Statistical science pp. 457–472 (1992)

[12] George, D.L.: Augmented Riemann solvers for the shallow water equations over variable topography with steady states and inundation. Journal of Computational Physics **227**(6), 3089–3113 (2008). DOI DOI10.1016/j.jcp.2007.10.027

[13] Haario, H., Saksman, E., Tamminen, J.: An adaptive metropolis algorithm. Bernoulli pp. 223–242 (2001)

[14] Hackbusch, W.: Tensor Spaces and Numerical Tensor Calculus, *Springer Series in Computational Mathematics*, vol. 42. Springer-Verlag Berlin Heidelberg (2012). DOI 10.1007/978-3-642-28027-6. URL http://www.springerlink.com/index/10.1007/978-3-642-28027-6

[15] Heemink, A.W., Mouthaan, E., Roest, M., Vollebregt, E., Robaczewska, K.B., Verlaan, M.: Inverse 3D shallow water flow modelling of the continental shelf. Continental Shelf Research **22**(3), 465–484 (2002)

[16] Jones, E., Oliphant, T., Peterson, P., et al.: SciPy: Open source scientific tools for Python (2001–). URL http://www.scipy.org/

[17] Kass, R.E., Raftery, A.E.: Bayes factors. Journal of the american statistical association **90**(430), 773–795 (1995)

[18] Kullback, S., Leibler, R.A.: On information and sufficiency. Ann. Math. Statist. **22**(1), 79–86 (1951). DOI 10.1214/aoms/1177729694. URL http://dx.doi.org/10.1214/aoms/1177729694

[19] Lardner, R.W., Song, Y.: Optimal estimation of Eddy viscosity and friction coefficients for a Quasi-three-dimensional numerical tidal model. Atmosphere-Ocean **33**(3), 581–611 (1995)

[20] MacInnes, B.T., Gusman, A.R., LeVeque, R.J., Tanioka, Y.: Comparison of Earthquake Source Models for the 2011 Tohoku Event Using Tsunami Simulations and Near-Field Observations. Bulletin of the Seismological Society of America **103**(2B), 1256–1274 (2013). DOI 10.1785/0120120121. URL http://www.bssaonline.org/cgi/doi/10.1785/0120120121

[21] Marzouk, Y.M., Najm, H.N.: Dimensionality reduction and polynomial chaos acceleration of bayesian inference in inverse problems. Journal of Computational Physics **228**(6), 1862–1902 (2009). DOI 10.1016/j.jcp.2008.11.024. URL http://www.sciencedirect.com/science/article/pii/S0021999108006062http://linkinghub.elsevier.com/retrieve/pii/S0021999108006062

[22] Mayo, T., Butler, T., Dawson, C., Hoteit, I.: Data assimilation within the advanced circulation (adcirc) modeling framework for the estimation of manning's friction coefficient. Ocean Modelling **76**, 43–58 (2014)

[23] Meinig, C., Stalin, S., Nakamura, A., Milburn, H.: Real-time deep-ocean tsunami measuring, monitoring, and reporting system: The noaa dart ii description and disclosure. Tech. rep. (2005)

[24] Morariu, V.I., Srinivasan, B.V., Raykar, V.C., Duraiswami, R., Davis, L.S.: Automatic online tuning for fast gaussian summation. In: Advances in Neural Information Processing Systems (2008)

[25] Okada, Y.: Surface deformation due to shear and tensile faults in a half-space. Bulletin of the Seismological Society of America **75**(4), 1135–1154 (1985). URL http://www.bssaonline.org/content/75/4/1135.abstract

[26] Rasmussen, C.E., Williams, C.K.I.: Gaussian processes for machine learning. The MIT Press (2006)

[27] Reagan, M.T., Najm, H.N., Ghanem, R.G., Knio, O.M.: Uncertainty quantification in reacting-flow simulations through non-intrusive spectral projection. Combustion and Flame **132**(3), 545–555 (2003). DOI 10.1016/S0010-2180(02)00503-5. URL http://www.sciencedirect.com/science/article/pii/S0010218002005035

[28] Sarri, A., Guillas, S., Dias, F.: Statistical emulation of a tsunami model for sensitivity analysis and uncertainty quantification. Natural Hazards and Earth System Science **12**(6), 2003–2018 (2012). DOI 10.5194/nhess-12-2003-2012. URL http://www.nat-hazards-earth-syst-sci.net/12/2003/2012/

[29] Sraj, I., Mandli, K., Knio, O.M., Hoteit, I.: Uncertainty Quantification and Inference of Manning's Friction Coefficients using DART Buoy Data during the Tohoku Tsunami. Ocean Modelling **83**, 82–97 (2014). DOI 10.1016/j.ocemod.2014.09.001

[30] Sraj, I., Mandli, K., Knio, O.M., Hoteit, I.: Quantifying uncertainties in Fault Slip Distribution During the Tohoku Tsunami using Polynomial Chaos. Submitted to Ocean Modelling (2016)

[31] Storn, R., Price, K.: Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. Journal of Global Optimization **11**(4), 341–359 (1997). DOI 10.1023/A:1008202821328. URL http://dx.doi.org/10.1023/A:1008202821328

[32] Verlaan, M., Heemink, A.W.: Tidal flow forecasting using reduced rank square root filters. Stochastic Hydrology and Hydraulics **11**(5), 349–368 (1997)

[33] Wiener, N.: The homogeneous chaos. American Journal of Mathematics **60**(4), 897 (1938). DOI 10.2307/2371268. URL http://dx.doi.org/10.2307/2371268http://www.jstor.org/discover/10.2307/2371268?uid=3738016&uid=2129&uid=2&uid=70&uid=4&sid=56294222093http://www.jstor.org/stable/2371268?origin=crossref

[34] Xiu, D., Karniadakis, G.E.: The wiener–askey polynomial chaos for stochastic differential equations. SIAM Journal on Scientific Computing **24**(2), 619–644 (2002). DOI 10.1137/S1064827501387826. URL http://epubs.siam.org/doi/abs/10.1137/S1064827501387826