

Uncertainty quantification and inference of Manning's friction coefficients using DART buoy data during the Tōhoku tsunami

Ihab Sraj ^a, Kyle T. Mandli ^b, Omar M. Knio ^{c,d}, Clint N. Dawson ^b, Ibrahim Hoteit ^{a,d,*}

^aDivision of Physical Sciences and Engineering, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

^bInstitute for Computational Engineering and Science, University of Texas at Austin, 201 E 24th ST, Stop C0200, Austin, TX 78712-1229, USA

^cDepartment of Mechanical Engineering and Materials Science, Duke University, 144 Hudson Hall, Durham, NC 27708, USA

^dDivision of Computer, Electrical and Mathematical Sciences and Engineering, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

ARTICLE INFO

Article history:

Received 10 February 2014

Received in revised form 13 August 2014

Accepted 7 September 2014

Available online 26 September 2014

Keywords:

Tsunami

Manning's n friction coefficient

Sensitivity analysis

Polynomial Chaos

Bayesian inference

ABSTRACT

Tsunami computational models are employed to explore multiple flooding scenarios and to predict water elevations. However, accurate estimation of water elevations requires accurate estimation of many model parameters including the Manning's n friction parameterization. Our objective is to develop an efficient approach for the uncertainty quantification and inference of the Manning's n coefficient which we characterize here by three different parameters set to be constant in the on-shore, near-shore and deep-water regions as defined using iso-baths. We use Polynomial Chaos (PC) to build an inexpensive surrogate for the GEOCLAW model and employ Bayesian inference to estimate and quantify uncertainties related to relevant parameters using the DART buoy data collected during the Tōhoku tsunami. The surrogate model significantly reduces the computational burden of the Markov Chain Monte-Carlo (MCMC) sampling of the Bayesian inference. The PC surrogate is also used to perform a sensitivity analysis.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Tsunamis in the past decade have been responsible for some of the most deadly and costly natural disasters ever recorded. Coastal communities have faced these hazards by assessing the risk they pose and by attempting to make informed decisions about the likelihood that such an event would occur. Computational models of tsunami events are often employed to explore various tsunami disaster scenarios and their ensuing inundation levels. The accurate prediction of water elevations, however, requires accurate estimation of many model parameters that are either measured directly, defined empirically, or estimated from a collection of observed data. Unfortunately, since tsunami events are relatively rare, there is a large uncertainty in the input data for these computational models ranging from effects of the domain, such as bathymetry and friction parameterizations, to the earthquake source. In this study, a methodology for quantifying this uncertainty is presented and applied to the Manning's n friction coefficient, a parameterization of the effect of bottom friction commonly used in tsunami models. We aim at quantifying the

uncertainty in the predicted water elevation and employ a Bayesian inverse modeling approach to estimate the Manning's n coefficient using observations of water elevation measured during a tsunami event.

Previous work concerning uncertainty quantification in tsunami modeling has been primarily focused on constraining the likely earthquake source models (see for example MacInnes et al., 2013). A similar approach for landslide-generated tsunami was presented in Sarri et al. (2012), by building a statistical surrogate using an emulator. This emulator was based on a Gaussian process that requires using a combination of prior knowledge about the simulator and appropriate choices of functions and parameters. Their work was presented as a proof-of-concept case study where they only performed basic statistical and sensitivity analysis. Changes to the friction parameterization has been shown in the past to lead to significant changes in predicted inundation levels (Myers et al., 2001; Jakeman et al., 2010), but these previous studies were limited in range and spatial variation of the values used in the parameterization. Work has also been carried out in a number of studies attempting to invert for input to tidal simulations in both two and three dimensions using adjoint or Kalman filtering based approaches (Das and Lardner, 1992; Lardner and Song, 1995; Verlaan and Heemink, 1997; Heemink et al., 2002). In a recent work Mayo et al. (2014) reformulated an ensemble Kalman

* Corresponding author at: Division of Computer, Electrical and Mathematical Sciences and Engineering, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia.

E-mail address: ibrahim.hoteit@kaust.edu.sa (I. Hoteit).

filtering data assimilation method generally used in the estimation of model states to estimate the Manning's coefficient in a tidal setting as well. They used a low-dimensional representation of the Manning's coefficients and recovered it by assimilating synthetic water elevation data using tidal simulations.

In this work, we focus our attention on the uncertainty in the friction parameterization used in tsunami simulations. We present a three-parameter representation of the Manning's coefficient and an efficient method to estimate this representation using measured water elevation data. It is important to note that the friction parameterization may not be the largest source of uncertainty in tsunami modeling. Here we use the friction parameterization as a test-bed for the methodology proposed recognizing this and relying on future work to untangle the effects of other sources of uncertainty. The method itself is based on recent developments in uncertainty quantification (UQ) methods that allow probing of the sensitivity of realistic tsunami models to uncertain parameters without modifying the forward model and inferring those parameters from a number of observations. In particular, we implement an inverse modeling approach that relies on a Bayesian inference technique. We use Polynomial Chaos (PC) expansions to build a surrogate model (Najm et al., 2009; Alexanderian et al., 2012; Elsheikh et al., 2014; Young and Grace, 2013) of the response of the tsunami predicted using the GeoCLAW model. The PC surrogate enables us to avoid using the forward model in the Bayesian inference, thus dramatically reducing the cost of sampling the posterior distribution.

PC expansions have been developed and applied in the engineering community to quantify uncertainties in numerical simulations. Their main advantage is the efficient propagation of input uncertainties through large, complex and non-linear models to calculate the corresponding output uncertainties. Several studies already investigated the efficiency of PC expansions for UQ in oceanic simulations (Thacker et al., 2012; Srinivasan et al., 2010; Alexanderian et al., 2012; Sraj et al., 2013a). Recently, PC was combined with Bayesian inference to estimate the wind drag coefficient using temperature data collected during Typhoon Fanapi 2010. The same problem was solved using two different techniques: Markov Chain Monte Carlo (MCMC) (Sraj et al., 2013a) and gradient based search method (Sraj et al., 2013b).

In this work, we follow a similar procedure as in Sraj et al. (2013a,b) to estimate Manning's n coefficient using the Tōhoku tsunami event as a case-study. The quantity of interest is taken to be the water surface elevation and is compared to the available NOAA Deep-ocean Assessment and Reporting Tsunamis (DART) buoys. The forward model runs are performed using the GeoCLAW package following the setup in Maclnnes et al. (2013).

The structure of this manuscript is as follows. First, we describe our case-study, the Tōhoku tsunami, summarize the forward simulations setup, and present the DART buoys observational data in Section 2. Section 3 discusses the uncertainties of the Manning's n coefficient to motivate the UQ problem and propose a three-parameter representation of n . Section 4 presents the two main methods employed in the UQ analysis: Bayesian inference that brings together observations and model results (Section 4.1) and PC expansions that are used to build the surrogate model (Section 4.2). Section 5 describes the main results in three subsections: first, we perform error and convergence studies as evidences that the PC surrogate is a faithful representation of the GeoCLAW model; second, we use the PC surrogate to perform a statistical analysis and compute the sensitivity of simulated water elevations to the uncertain input parameters; and third, we present the results of the posterior distributions obtained using Bayesian inference by MCMC. A discussion of our findings and methodology is finally presented in Section 6.

2. Simulating the Tōhoku tsunami

The Tōhoku earthquake and tsunami of 2011 have been the subject of a number of studies due to the wealth of observational evidence and severity of the tsunami (Hiraishi et al., 2011). The earthquake itself had an estimated magnitude of M_w 9.0 and caused massive devastation across Japan. The epicenter was located approximately 72 km east of Tōhoku as indicated in Fig. 1(a). In our case study, the simulation used is based on work done in Maclnnes et al. (2013) and modified to use the specified variable friction field. Hereafter, we briefly describe the forward computational model, GeoCLAW, the simulation setup used in the uncertainty analysis, and finally present the observations from NOAA DART buoys.

2.1. GeoCLAW

The GeoCLAW software package uses a finite volume, wave-propagation approach as described in LeVeque (1997) to solve the two-dimensional shallow water equations (SWE):

$$\begin{aligned} \frac{\partial}{\partial t} h + \frac{\partial}{\partial x}(hu) + \frac{\partial}{\partial y}(hv) &= 0, \\ \frac{\partial}{\partial t}(hu) + \frac{\partial}{\partial x}\left(hu^2 + \frac{1}{2}gh^2\right) + \frac{\partial}{\partial y}(huv) &= fhv - gh\frac{\partial}{\partial x}b - C_f|\mathbf{u}|hu, \\ \frac{\partial}{\partial t}(hv) + \frac{\partial}{\partial x}(huv) + \frac{\partial}{\partial y}\left(hv^2 + \frac{1}{2}gh^2\right) &= -fhu - gh\frac{\partial}{\partial y}b - C_f|\mathbf{u}|hv, \end{aligned} \quad (1)$$

where h is the depth of the water column, u and v the velocities in the longitudinal and latitudinal directions respectively, g the acceleration due to gravity, b the bathymetry, f the Coriolis parameter, and C_f the bottom friction coefficient. The sea-surface anomaly η , the difference between a specified datum, such as mean tide level, and the modeled sea-surface, is $\eta = h + b$.

GeoCLAW originated from the software package CLAWPACK which solves systems of hyperbolic equations where the primary computational kernel is the solution to the Riemann problem at each grid cell interface. The Riemann solver in GeoCLAW contains a number of features relevant to tsunami modeling including: the ability to handle inundation (wet-dry interfaces), implementation of a well-balanced scheme to suppress spurious motion induced by topographic variations even when the momentum is non-zero, and entropy corrections (George et al., 2008). One of the unique features of GeoCLAW is the ability to adaptively refine the computational grid to resolve particular dynamical features as they evolve in time (the surface height disturbance in the present case). GeoCLAW implements the adaptive schemes in Berger and Oliger (1984, 1998) which use overlapping and properly nested grids of different levels of refinement to follow features in the solution of interest. This is accomplished by flagging cells that need refinement and constructing rectangular patches that contain these cells. For tsunamis, a threshold τ_η is defined such that a cell is marked for refinement if the sea-surface $\eta > \tau_\eta$. Coarsening occurs when a cell does not meet this criteria. Regions of refinement can also be specified such that the specified region is forced or restricted to a particular level of refinement.

For the simulations presented, refinement thresholds were used that matched what was presented in Maclnnes et al. (2013). In this study, five levels of refinement with refinement ratios of 4, 6, 5, and 4 were used starting with a resolution of 1° in both the latitudinal and longitudinal directions down to $75''$ resolution (approximately 2000 m) located around the observation locations. The tolerance for the refinement criteria for sea-surface anomaly was 0.02 m. Over the entire domain refinement was restricted to only

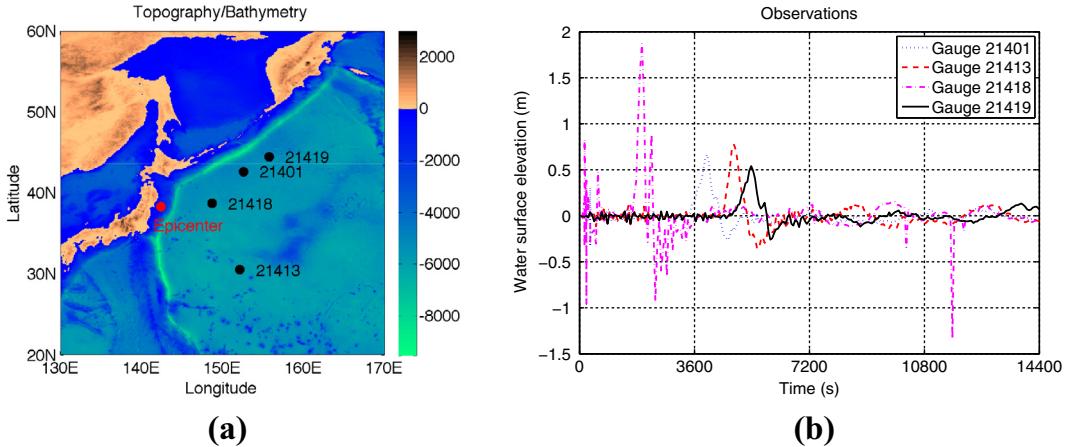


Fig. 1. (a) The topography, bathymetry and gauge locations used in the simulation. (b) Observed de-tided water surface height at all the DART buoys used with $t = 0$ set to initiation time of the earthquake. The buoys during normal operations records sea-surface height every 15 min; once an earthquake is reported they record the sea-surface height every 15 s. As observed in MacInnes et al. (2013), the large spikes reported by gauge 21418 at later times are data anomalies and not subsequent tsunami waves.

the third level except for the region between 140° and 146° east longitude and 35° and 41° north latitude where the simulation was forced to refine to either level 4 or 5 for the first 1000 s of simulation time.

2.2. DART buoy observations

The observations used in this study are from the Deep-ocean Assessment and Reporting of Tsunamis (DART) buoy system developed and maintained by the National Oceanic and Atmospheric Administration (NOAA). The purpose of the network is to provide early-warning detection and forecasting of tsunami propagation in the Pacific Ocean (Milburn et al., 1996). The DART buoys closest to the earthquake source of the Tōhoku tsunami were buoys 21401, 21413, 21418, and 21419 whose locations are shown in Fig. 1(a)) and whose de-tided water surface elevation data for the event are shown in Fig. 1(b) (see Mungov et al., 2012 for details on the data processing methodologies used for the DART buoy data).

As a validation of GEOCLAW's ability to realistically simulate water surface elevation in the context of the Tōhoku tsunami, a comparison between the (DART) buoy observations and their GEOCLAW model counterparts was made with the Manning's n coefficients set to their default values of $n_1 = n_2 = n_3 = 0.025$. Fig. 2 shows a scatter plot that compares the observed water surface elevation at the different gauge locations and the GEOCLAW model counterparts (data points are colored differently for the different gauges). The plot shows a reasonable agreement between the simulations and the observations at gauges 21401, 21413 and 21419. This agreement is reflected in the low standard deviation of the error σ_i calculated at those gauges as indicated in the legend. The high amount of scatter, however, from gauge 21418 can probably be attributed to its proximity to the epicenter of the earthquake and shore region. The differences between the simulation and observations overall can likely be attributed to uncertainties in the input data such as the Manning's n coefficients, errors in the earthquake rupture model, insufficiently accurate bathymetry in the near-shore region, and to model errors such as unresolved effects and approximations inherent in the shallow water model.

2.3. Bathymetry and earthquake source model

The bathymetry used in the simulation is a combination of ETOPO 1' and 4' accurate bathymetry for the region (Amante et al.,

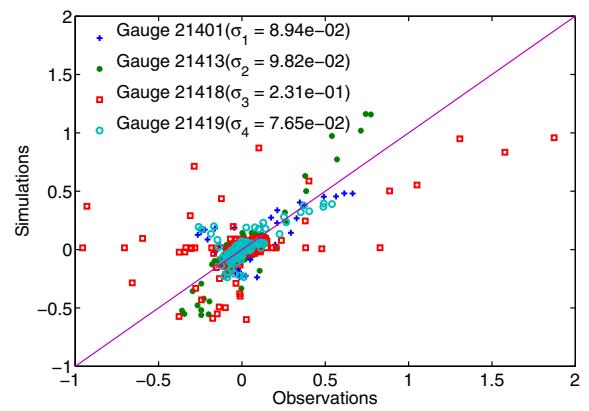


Fig. 2. Scatter plot of measured water surface elevation against their GEOCLAW model counterparts during tsunami Tōhoku. The data points are colored differently for the different gauges and standard deviation of the error σ_i is indicated in the legend.

2009). Finer bathymetry was used in the study of MacInnes et al. (2013) in order to match the on-shore inundation levels. In the current study this bathymetry was excluded since we are only concentrating on the data from the DART buoys. The bathymetry and topography of the considered regions are shown in Fig. 1(a).

The earthquake source model that was used is due to Ammon (2011) who used inversion of the seismic waves to reconstruct the rupture. This model was chosen as it provides a rupture reconstruction based on established techniques and did not include tsunami inversion data. The latter may introduce bias into the reported results of this study and was consequently ignored. As discussed in MacInnes et al., the source models used do not include rupture timing which could have an impact on the accuracy of the simulation. The analysis was also tested with an additional source model due to Saito et al. (2011) whose inversion was based purely on a tsunami inversion methodology.

3. Manning's n friction coefficient

Manning's n law provides a relationship between the dimensional friction coefficient C_f appearing in the SWE governing tsunamis (Eq. (1)) and the empirically derived roughness parameter n :

$$C_f = \frac{gn^2}{h^{4/3}}, \quad (2)$$

where n has units of $s \cdot m^{-1/3}$. The value of n may vary both spatially and temporally due to changes in the bottom roughness from wave action and other processes. In the context of tsunami modeling, it is perhaps unclear that Manning's friction is appropriate in all regimes being modeled, as it is derived on the assumption that a fully developed turbulent state exists in the bottom boundary layer. This is further called into question as it is a common practice to assume a single value of Manning's n for the areas that are initially below sea-level; while satellite images and land-use data are sometimes used to determine the coefficient in areas above sea-level. In either case, this process introduces uncertainty in model output, most importantly in the forecasted water elevation and more critically in the near-shore. MacInnes et al., for instance, found that adopting a uniform value of $n = 0.025$ everywhere leads to poor performance in predicting inundation levels. To address this they used $n = 0.035$ in the Sendai plane to account for land-uses such as pasture, farmland, and rice paddies. We aim to address such uncertainties systematically by quantifying their impact on water elevations. We seek to quantify the parametric uncertainty in Manning's friction coefficient by characterizing its spatial variability and its depth-dependence; we thus define distinct regions in the domain that are characterized by a single coefficient.

In this study three regions are considered, an on-shore region (initially above sea-level), a near-shore region (between sea-level and the 200 meter iso-bath), and the deep-water region (deeper than 200 meters), whose Manning's n coefficient is denoted by n_1 , n_2 , and n_3 , respectively as shown in Fig. 3(a). These regions were chosen in an attempt to capture three different regimes of tsunami propagation recognized as being characterized with different physically relevant processes. In particular each of these regions is thought to have different dependence on the friction parameterization. A choice of more regions or a more localized characterization of the friction to specific parts of the domain were not undertaken in an effort to reduce the number of free-parameters in the inversion.

To quantify the uncertainties in these three parameters we employ Polynomial Chaos (PC) expansions as used in Sraj et al. (2013a,b). The method and its application to our problem is described in Section 4.2. To represent the initial uncertainty in the Manning's n coefficients an uninformative uniform distribution was assumed as the prior with physically relevant values of n sampled from the interval $n \in [n_{\min} - n_{\max}]$, where $n_{\min} = 0.005$ and $n_{\max} = 0.2$.

4. Formulation

In this section, we first present the Bayesian approach to inverse problems and its application to our problem of inferring the Man-

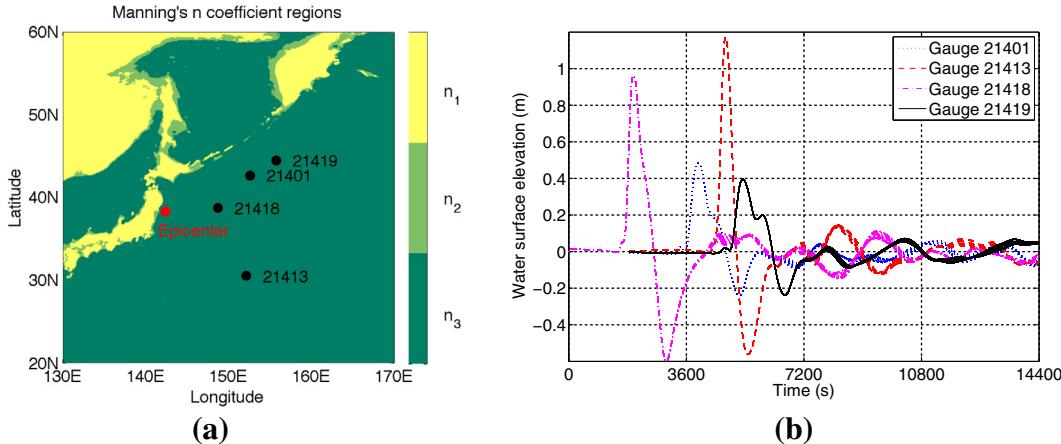


Fig. 3. (a) Manning's n coefficients at three regions: n_1 on-shore, n_2 near-shore, n_3 deep-water. (b) 125 GEOCLAW realizations at different gauge locations.

ning's n coefficients. We also introduce a Polynomial Chaos method used for forward uncertainty propagation, and for accelerating the Bayesian inference.

4.1. Bayesian inference

Bayesian inference is a statistical approach to inverse problems that has recently gained great interest in different applications, including ocean (Alexanderian et al., 2011; Zedler et al., 2012; Sraj et al., 2013a) climate (Olson et al., 2012) and geophysical (Malinverno et al., 2002) modeling. The key idea in this approach is to express all forms of uncertainty in terms of canonical random variables. The first step in Bayesian inference is to formulate the forward problem using a suitable likelihood function and a product of conditional probability densities. The prior distributions over the unknown parameters of the model are then formulated using the best available knowledge of the parameters. Given some observation data, Bayes rule is then used to compute the posterior distribution for these unknown parameters (Sivia et al., 2006). We briefly review this approach in this section.

Let η^a be a vector of observed water surface elevation from the DART buoys and $\mathbf{n} = (n_1, n_2, n_3)^T$ be a vector of model control parameters, in our case the three friction n coefficients. We consider the forward model \mathbf{G} represented by GEOCLAW that predicts the data as function of the parameters such that:

$$\eta^a \approx \mathbf{G}(\mathbf{n}). \quad (3)$$

Applying Bayes' rule yields:

$$\Pi(\mathbf{n}|\eta^a) \propto L(\eta^a|\mathbf{n}) q(\mathbf{n}), \quad (4)$$

where $q(\mathbf{n})$ is the prior of \mathbf{n} , representing the *a priori* knowledge about the parameters; $L(\eta^a|\mathbf{n})$ is the likelihood function representing the probability of obtaining the data given the set of parameters \mathbf{n} ; and finally $\Pi(\mathbf{n}|\eta^a)$ is the posterior, representing the probability of occurrence of \mathbf{n} given the data η^a .

To formulate the likelihood function, we let $\epsilon = \eta^a - \mathbf{G}$ represent the discrepancy between the model and observations. Here, the components of ϵ are assumed to be i.i.d. random variables with density p_ϵ . The likelihood function can thus be written as:

$$L(\eta^a|\mathbf{n}) = \prod_i p_\epsilon(\eta_i^a - G_i(\mathbf{n})). \quad (5)$$

In our application, we assume that the errors ϵ_i are independent and normally distributed with mean zero and variance σ^2 , i.e. $\epsilon_i \sim N(0, \sigma^2)$ resulting in the following likelihood function:

$$L(\eta^a|\mathbf{n}) = \frac{1}{\sqrt{2\pi\sigma^2}} \prod_i \exp \left\{ \frac{-(\eta_i^a - G_i(\mathbf{n}))^2}{2\sigma^2} \right\}. \quad (6)$$

While in general σ^2 (the variance of the noise in the measured data) depends on the observations, in cases where the error amplitude is generally small and does not change throughout space and time one may assume a single σ^2 value in Eq. (6). In our case, the measurement may vary significantly from one gauge to another and the data collected may be exposed to different measurement errors; therefore, it is reasonable to use different variances for each gauge of data collection as indicated in the results section.

The joint posterior from Bayes' rule is then given as:

$$\Pi(\mathbf{n}|\eta_i^a) \propto \frac{1}{\sqrt{2\pi\sigma^2}} \prod_i \exp \left\{ \frac{-(\eta_i^a - G_i(\mathbf{n}))^2}{2\sigma^2} \right\} q(\mathbf{n}). \quad (7)$$

Since σ^2 is unknown *a priori* we treat it as a hyper-parameter. In other words, σ^2 becomes an additional parameter for Bayesian inference and therefore endowed with a prior which is updated based on available observations. In this case the joint posterior is finally expressed as:

$$\Pi(n_1, n_2, n_3, \sigma^2 | \eta_i^a) \propto \frac{1}{\sqrt{2\pi\sigma^2}} \prod_i \exp \left\{ \frac{-(\eta_i^a - G_i)^2}{2\sigma^2} \right\} q(\sigma^2) q(n_1) q(n_2) q(n_3). \quad (8)$$

To determine the posterior, we need to choose proper priors that should be based on some *a priori* knowledge about the parameters. In our case, a uniform prior for all three Manning's n parameters is assumed, with n_i in the range $[n_{min} - n_{max}]$; therefore:

$$q(n_i) = \begin{cases} \frac{1}{n_{max} - n_{min}} & \text{for } n_{min} < n_i \leq n_{max}, \\ 0 & \text{otherwise.} \end{cases} \quad (9)$$

Regarding the noise variance, the only information we know is that σ^2 is always positive. We thus assume a Jeffreys prior (Sivia et al., 2006) for σ^2 , expressed as:

$$q(\sigma^2) = \begin{cases} \frac{1}{\sigma^2} & \text{for } \sigma^2 > 0, \\ 0 & \text{otherwise.} \end{cases} \quad (10)$$

Inferring the coefficients amounts to sampling the posterior. In general, when the space of the unknown parameters is multidimensional, a suitable computational strategy is the Markov Chain Monte Carlo (MCMC) method. We rely on an adaptive Metropolis MCMC algorithm (Roberts and Rosenthal, 2009; Haario et al., 2001) to accurately and efficiently sample the posterior distribution.

4.2. Accelerating Bayesian inference

The four-dimensional posterior in Eq. (8) can be directly explored via MCMC; this requires repeated simulations (tens of thousands) of the forward GeoClaw model, once for every proposed set of parameters of the Markov chain (Marzouk and Najm, 2009; Malinverno et al., 2002). While a single GeoClaw simulation takes about 15 min, depending on the details of the MCMC algorithm used, it is desirable to avoid running the forward model at every realization of the MCMC. This is achieved by constructing a surrogate model that requires a much smaller ensemble of GeoClaw runs, and that can be used instead at a significantly reduced computational cost. Here, we rely on Polynomial Chaos expansions to build a surrogate model, which, in addition, can efficiently provide statistical properties, such as the mean, variance and sensitivities.

4.2.1. Polynomial Chaos

Polynomial Chaos (PC) is a probabilistic methodology that expresses the dependencies of model outputs on the uncertain model inputs as a truncated polynomial expansion (Villegas et al., 2012; Lin et al., 2009; Xiu and Tartakovsky, 2004). We briefly

describe the PC method below; for more details the reader is referred to Le Maître and Knio (2010).

Let $U = U(\mathbf{x}, t, \xi)$ denote a quantity of interest that is the output of a computational model. U is function of space, \mathbf{x} , and time, t , and also depends on the canonical vector of random variables $\xi = (\xi_1, \dots, \xi_n)$. This stochastic vector ξ parameterizes the uncertain inputs (Manning's coefficients) depending on its probability density function. In the case of uniform distributions, the canonical vector is calculated as follows: $\xi_i = \frac{2n_i - (n_{max} + n_{min})}{(n_{max} - n_{min})}$ (Le Maître and Knio, 2010). PC expresses U in the form:

$$U(\mathbf{x}, t, \xi) \doteq \sum_{k=0}^P U_k(\mathbf{x}, t) \Psi_k(\xi), \quad (11)$$

where $U_k(\mathbf{x}, t)$ are the polynomial coefficients, and $\Psi_k(\xi)$ are elements of an orthogonal basis of an underlying probability space. The total number of terms in the truncated PC expansion is $P+1 = \frac{(n+p)!}{n! p!}$ where n is the number of stochastic dimensions and p is the highest order polynomial retained.

The choice of the basis is dictated by the probability density function of the stochastic vector ξ , which appears as a weight function in the probability space's inner product:

$$\langle \Psi_i, \Psi_j \rangle = \int \Psi_i(\xi) \Psi_j(\xi) \rho(\xi) d\xi = \delta_{ij} \langle \Psi_i^2 \rangle, \quad (12)$$

where δ_{ij} is the Kronecker delta. For uniform distributions, the basis functions are scaled Legendre polynomials. For multi-dimensional problems the basis functions are tensor products of 1D basis functions (Le Maître and Knio, 2010).

The series representation (11) can be viewed as a spectral expansion of U along the stochastic dimensions. It can also be seen as a combination of approximation and probabilistic frameworks; this has proven extremely useful in solving UQ problems (Xiu and Xiu, 2003; Lin et al., 2009). The existence and convergence of this series is asserted by the Cameron–Martin theorem (Cameron and Martin, 1947) with the condition that U has a finite variance. The series rate of convergence, and hence the number of terms to retain, depends on the smoothness of U with respect to ξ . The series converges spectrally fast with P when U is smooth; the convergence rate becomes algebraic when U has finite smoothness (Canuto et al., 2006). In practice the series convergence is monitored via various error metrics as discussed in the results section.

4.2.2. Non-Intrusive Spectral Projection (NISP)

The computation of the coefficients of the PC expansions U_k can be done using a number of procedures. Here we adopt a non-intrusive approach that allows the use of the forward model GeoClaw as a black box with no code modifications required. PC expansion coefficients are determined based on a set of response GeoClaw simulations at a specified set of the uncertain parameters. Specifically, we rely on the Non-Intrusive Spectral Projection (NISP) method that exploits the orthogonality of the basis and applies the Galerkin projection to find the PC expansion coefficients as follows:

$$U_k(\mathbf{x}, t) = \frac{\langle U, \Psi_k \rangle}{\langle \Psi_k, \Psi_k \rangle} = \frac{1}{\langle \Psi_k, \Psi_k \rangle} \int U(\mathbf{x}, t, \xi) \Psi_k(\xi) \rho(\xi) d\xi. \quad (13)$$

This orthogonal projection minimizes the L_2 error on the space spanned by the basis. The stochastic integrals are then approximated using a numerical quadrature to obtain:

$$\langle U, \Psi_k \rangle \approx \langle U, \Psi_k \rangle_Q = \sum_{q=1}^Q U(\xi_q) \Psi_k(\xi_q) \omega_q, \quad (14)$$

where the subscript Q refers to approximating the inner product integral with quadrature, and ξ_q and ω_q are respectively multi-

dimensional quadrature points and weights, and Q is the number of nodes in the multi-dimensional quadrature. The quadrature order should be commensurate with the truncation order, and should be high enough to avoid aliasing artifacts. The choice of quadrature rule is hence critical to the performance of the PC (in its NISP version at least).

The computation of the U_k can thus be expressed as a matrix–vector product of the form:

$$U_k(\boldsymbol{x}, t) = \sum_q \mathcal{P}_{kq} U(\boldsymbol{x}, t, \xi_q), \quad \mathcal{P}_{kq} = \frac{\Psi_k(\xi_q) \omega_q}{\langle \Psi_k, \Psi_k \rangle}, \quad (15)$$

where \mathcal{P}_{kq} is the projection matrix and $U(\boldsymbol{x}, t, \xi_q)$ is obtained from an ensemble of the deterministic model realizations with the uncertain parameters set at the quadrature value ξ_q .

4.2.3. Statistical moments and sensitivity analysis

The identification of the inner product weight function with the probability distribution of ξ simplifies the calculations of the statistical moments of U . Noting that since $\Psi_0(\xi)$ is a constant that is normalized so that $\langle \Psi_0, \Psi_0 \rangle = 1$, the expectation and variance of U can be computed as:

$$E[U] = \int U \rho(\xi) d\xi \approx \langle U, \Psi_0 \rangle_Q = U_0 \quad (16)$$

and

$$E[(U - E[U])^2] = \int (U - E[U])^2 \rho(\xi) d\xi \approx \sum_{k=1}^P U_k^2 \langle \Psi_k, \Psi_k \rangle. \quad (17)$$

PC representations also enable efficient global sensitivity analysis that quantify the contributions of different random input parameters to the variance in the output. This can be done by computing the so-called *total* sensitivity index T_i that measures the contribution of the i th random input to total model variability by computing the fraction of the total variance due to all the terms in the PC expansion that involve ξ_i (Le Maître and Knio, 2010; Crestaux et al., 2009; Sudret, 2008) as follows:

$$T_i = \frac{\sum_{k \in K_i} U_k^2 \langle \Psi_k^2 \rangle}{\sum_{k=1}^P U_k^2 \langle \Psi_k^2 \rangle}, \quad (18)$$

where

$$K_i = \{k \in \{1, \dots, P\} : \alpha_i^k > 0\}$$

and α^k is the multi-index associated with the k th term in the PC expansion (Le Maître and Knio, 2010).

5. Results

We begin by performing an error and convergence analysis of the constructed surrogate using the PC expansions to establish its validity. We then present a statistical analysis to quantify the uncertainty in the predicted water surface elevation, as well as a sensitivity analysis that enables us to rank the impact of the different Manning's n coefficients on the uncertainty. Finally, we present the results of the inverse problem where we determine the posterior distributions of the uncertain input parameters and analyze them in light of the available gauge data.

5.1. Error and convergence study

In the current work, we employ a tensorized Gaussian quadrature to construct a 4th order PC surrogate model. To this end, 125 GeoCLAW simulations were used to compute the PC coefficients using Eq. (15). The computational cost of running this ensemble is low compared to the cost of the thousands of runs of the forward

model that would have been required if a standard Monte-Carlo uncertainty propagation technique was to be adopted. Fig. 3(b) plots the evolution of the water surface elevation predicted at the four different gauges using GeoCLAW for the 125 different realizations required to compute the PC expansions. We notice that the variability in water surface elevation is insignificant in the first hour at all gauges as the plots of the different realizations superimpose. Later in time, the variability starts to increase for all gauges as indicated by the thickness of the bands formed by the curves of different realizations. This variability appears to be significant at $t \sim 5400, 6000, 3600$ and 7200 s for gauges 21401, 21413, 21418, 21419, respectively and corresponds roughly with the arrival time of the second set of waves (see Table 2). The arrival time is consistent with the distance from the gauge to the epicenter of the earthquake located approximately 72 km east of Japan where gauge 21418 is the closest and gauge 21419 is the farthest as shown in Fig. 1(a). The uncertainty in the prediction of water surface elevation persists till the end of the simulations at all gauges.

In order to check the consistency of the PC approximation, we compare water surface elevation from the realizations with those obtained from the PC surrogate. The different curves (not shown) reveal excellent agreement for all times and gauge locations. To quantify this agreement, we define an error metric that measures the relative normalized root mean-square error between the left hand side function in Eq. (11) and its PC representation at the sampling points:

$$E = \frac{\left(\sum_{\xi \in \mathcal{S}} |U(\xi) - \sum_{k=0}^P U_k \Psi_k(\xi)|^2 \right)^{1/2}}{\left(\sum_{\xi \in \mathcal{S}} |U(\xi)|^2 \right)^{1/2}}, \quad (19)$$

where \mathcal{S} is the 125-member ensemble used to construct the PC surrogate. This error metric calculated at the different gauge locations, is shown in Fig. 4; the largest relative normalized error for water surface elevation is about 1%. The spikes present in the error curves correspond to times when $\eta = 0$ which leads to round-off error when calculating the relative error.

Another check for the validity of the PC approximation consists of verifying whether the probability density functions (*pdfs*) of water surface elevation at the different gauge locations converges with increased order of the PC expansion (Alexanderian et al., 2012; Sraj et al., 2013a). Sample water surface elevation *pdfs* are shown in Fig. 5 at $t = 7200$ s at different gauges where the different curves correspond to increased order of PC ($p = 1 - 4$). The plots indicate double peaked distributions that are well-resolved with PC order $p = 3$ and becomes insensitive with additional refinement ($p = 4$). We therefore used PC order $p = 4$ in all computations below. This test and the various error metrics presented above provide confidence that the PC expansion is a faithful model surrogate that can be used in both the forward and inverse problems.

5.2. Forward propagation of uncertainty

The PC expansion created using an ensemble of 125 GeoCLAW simulations is now used as a surrogate to propagate prior input parameter uncertainty through the forward model. Here we exploit the surrogate to study the statistics of the water surface elevation, and to conduct a global sensitivity analysis of the impact of the uncertain input parameters.

5.2.1. Statistical analysis

The mean of the sea-surface elevation and its standard deviation is computed from the PC coefficients as indicated in Eqs. (16) and (17). In Fig. 6 we plot the evolution of the mean water surface elevation along with two standard deviations bounds at the

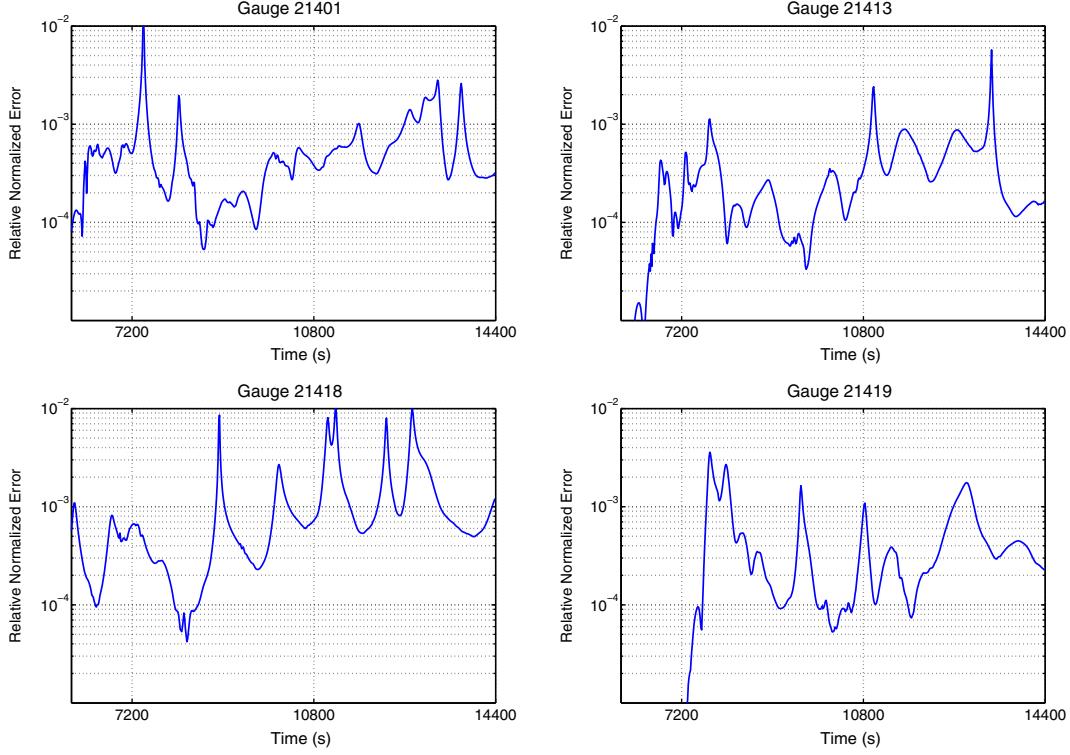


Fig. 4. Relative normalized error between realizations and the corresponding PC surrogates at different gauge locations calculated using Eq. (19).

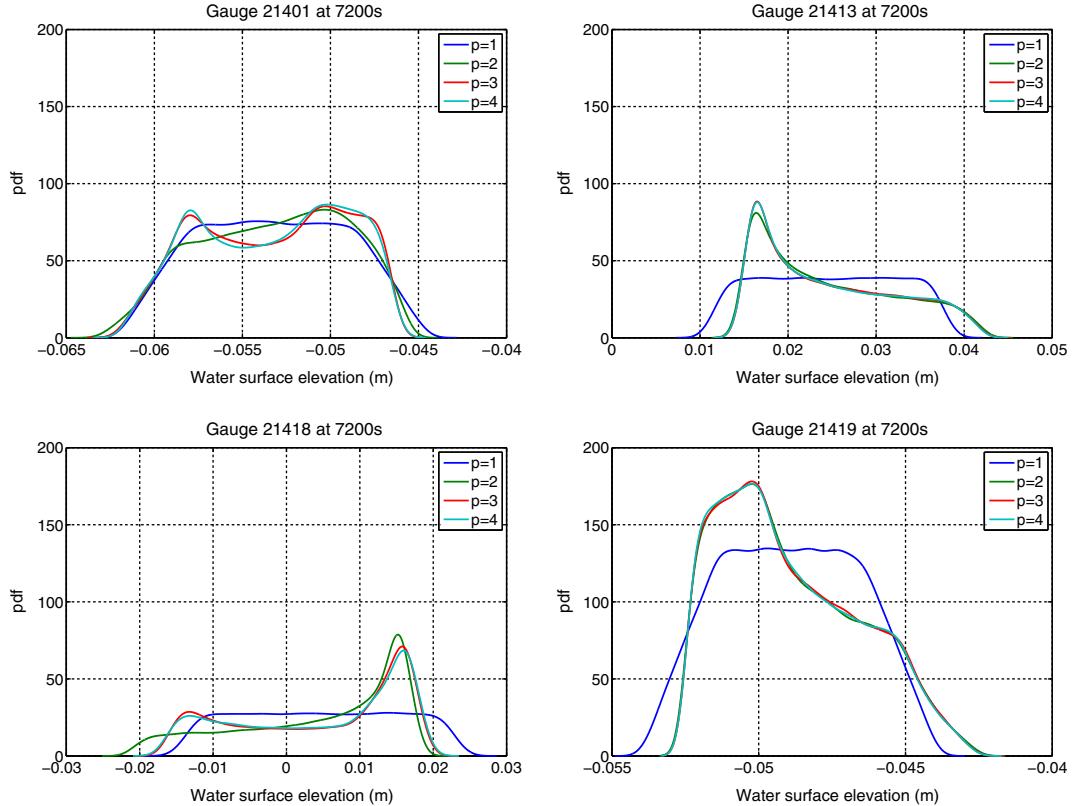


Fig. 5. pdf of water surface elevation at the different gauge locations at $t = 7200$ s.

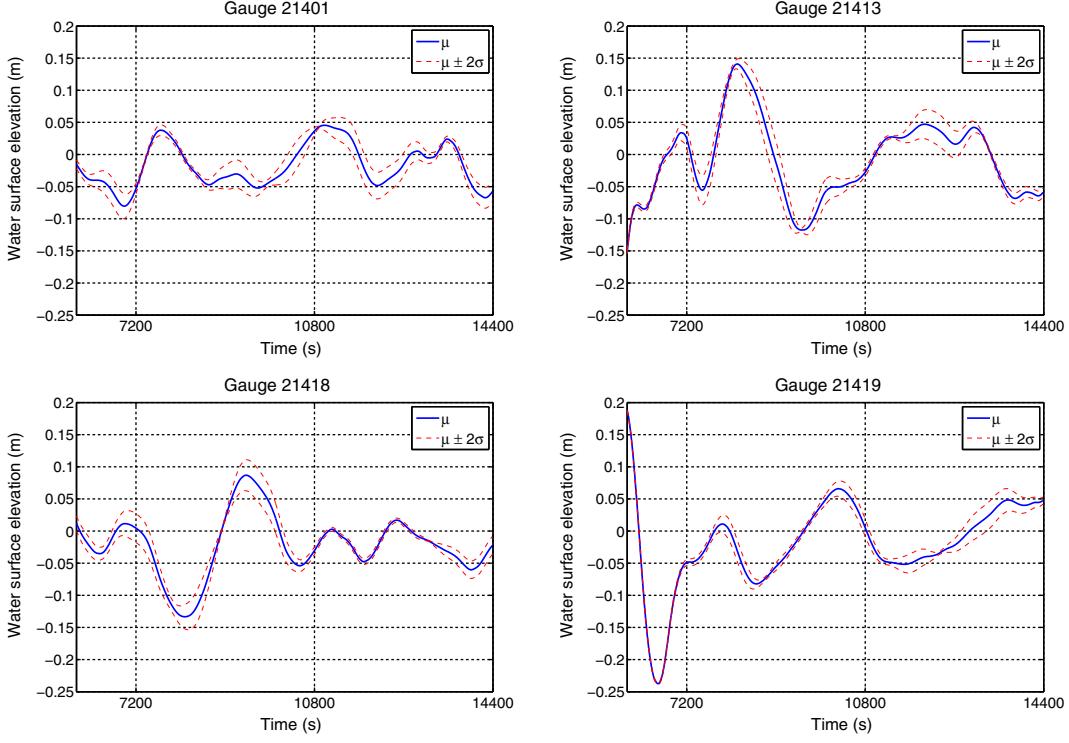


Fig. 6. Evolution of PC mean water surface elevation at different gauge locations.

four gauges. Note that the evolution shown starts at $t = 6000$ s when the uncertainty becomes significant. An interesting observation is that the standard deviation in water surface elevations waxes and wanes as the tsunami evolves. The narrowing of the variance at these instances is possibly associated with the waves that arrive due to reflections from a single source and then move

away from the gauge location, and thus cause no variance in the water surface elevation.

The predicted mean water surface elevation and its standard deviation can be used for comparison with the (DART) buoy observations collected during the tsunami event. Fig. 7 plots the PC mean water surface elevation with the observed water surface ele-

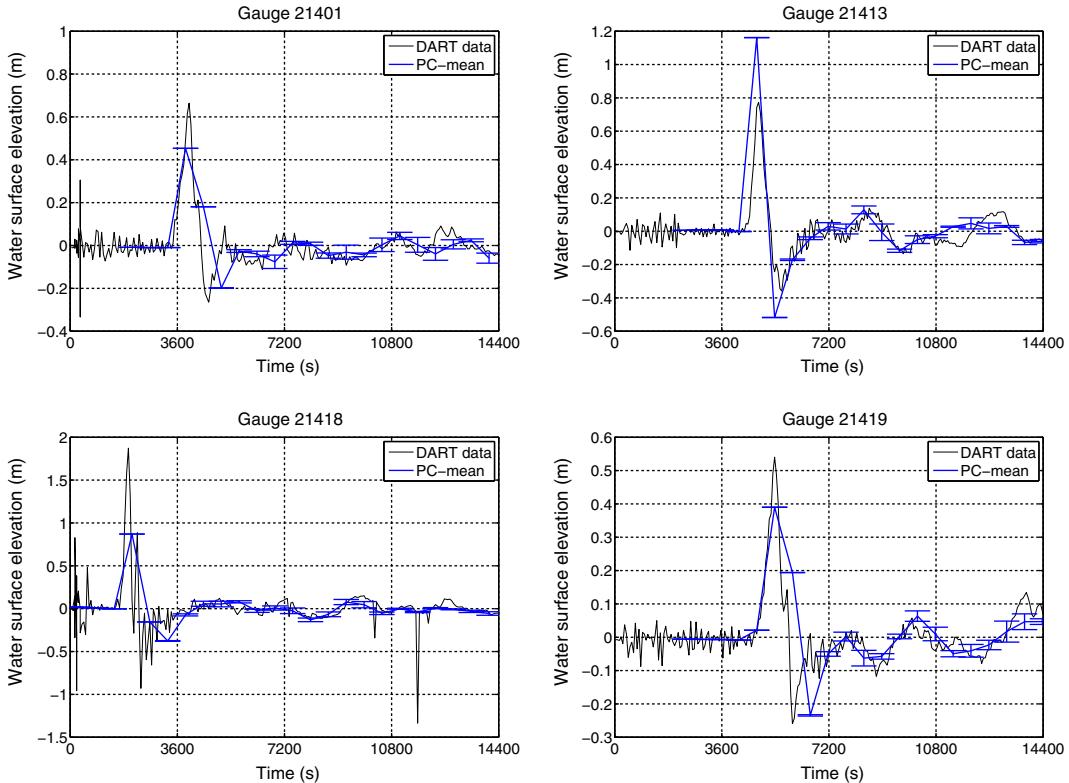


Fig. 7. Comparison of PC mean and observed DART data of water surface elevation with time at the four gauges.

Table 1

Locations of the numerical tide gauges placed in the near-shore region.

Gauge	Longitude	Latitude
1	140.846971	36.351141
2	141.115000	37.420000
3	141.133576	42.133955
4	143.739806	42.535828

vation at the different gauges locations with variances indicated at few time instances. The plots show a good agreement between the PC simulations and the observations at different locations and at different times.

PC expansions can be helpful in risk assessment as it can predict inundation at practically any location. This can be shown by predicting the tsunami run-up once it hits the shore. To this end, we compute the mean water surface elevation at four virtual gauges near the shoreline of Japan. The location of these near-shore gauges are found in Table 1. In Fig. 8, we plot the evolution of the mean water surface elevation along with two standard deviation bounds at the different near-shore gauges. At these locations the uncertainty in water surface elevation varies significantly from one location to another with gauge 3 having nearly no variance until 2.5 h into the simulation while gauge 2, placed just off-shore of the Fukushima Daiichi power plant, has a higher overall amplitude, reaching more than 4 m, due to the first wave hitting the shore as well as a higher variance that becomes significant much earlier. This significant variation in the near-shore gauges supports the idea that studying the uncertainty in the friction parameterization is worthwhile.

The same statistical analysis can be performed for the entire domain in 2D. Fig. 9 (top row) shows the PC mean water surface elevation for the considered computational domain at three differ-

Table 2

Estimated arrival times of the first and second tsunami wave at the DART buoy locations determined from the GeoClaw simulations.

Gauge	First wave	Second (reflected) wave
21418	0.5 h (1800 s)	1.25 h (2700 s)
21401	1 h (3600 s)	1.75 h (6300 s)
21413	1.25 h (2700 s)	2 h (7200 s)
21419	1.5 h (5400 s)	2 h (7200 s)

ent times as indicated in the title of each panel. The standard deviation is also shown in Fig. 9 (bottom row) at different times. We clearly notice the propagation of the variance due to the parametric uncertainty with time and along the tsunami's reflections.

5.2.2. Sensitivity analysis

A global sensitivity analysis is performed to quantify the contribution of each uncertain parameter to the variance in water surface elevation. To this end, we calculate the total sensitivity index using the PC coefficients as shown in Eq. (18) (Alexanderian et al., 2012; Sudret, 2008; Crestaux et al., 2009). The evolution of the total sensitivity index of each of the uncertain parameters is shown in Fig. 10 at the four gauges. The Manning's n coefficient near-shore n_2 is clearly dominant and contributes the most to the variance in the water surface elevation compared to the other two Manning's n coefficients n_1 and n_3 ; this is true for almost the entire simulation time and at the four gauges. The Manning's n coefficient in the deep ocean n_3 at gauge number 21419 exhibits a higher sensitivity index during the first hour of simulation however, as noted previously, none of the gauge locations exhibit significant variance between simulations until after the first hour of simulation. As a result, the high sensitivity of n_3 is not reliable. The on-shore Manning's n coefficient, n_1 , has no effect on the variance in the water surface elevation for the entire simu-

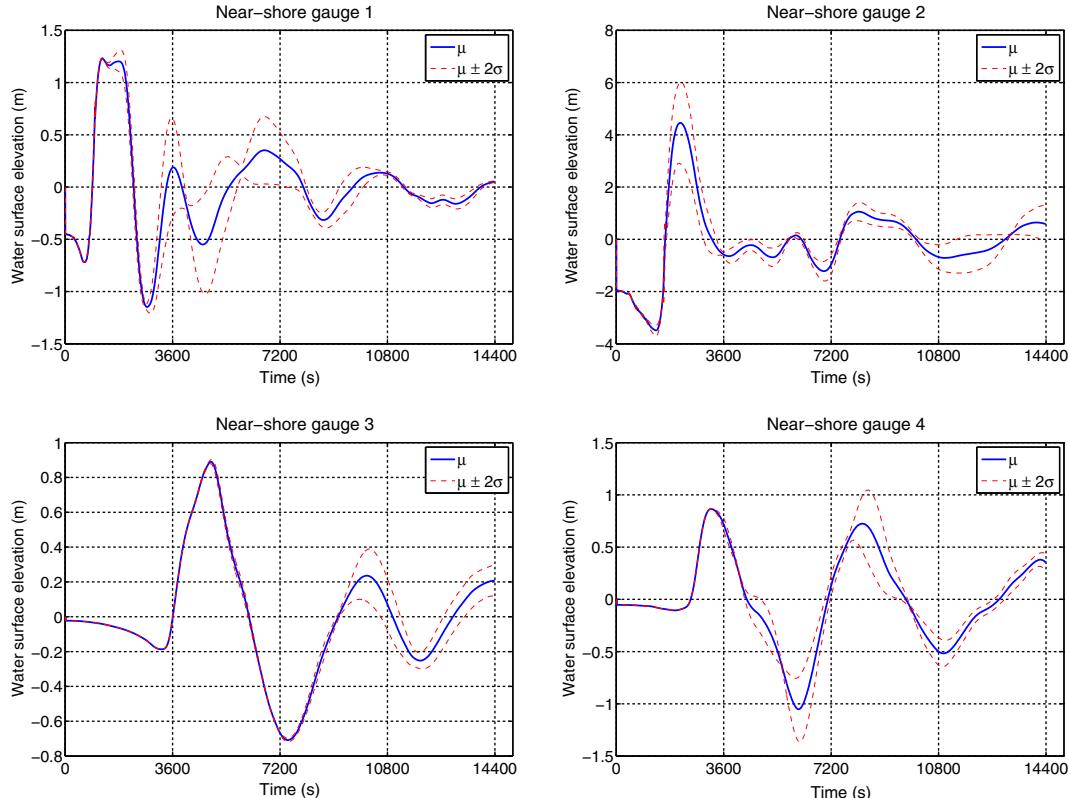


Fig. 8. Evolution of PC mean water surface elevation at different near-shore gauges.

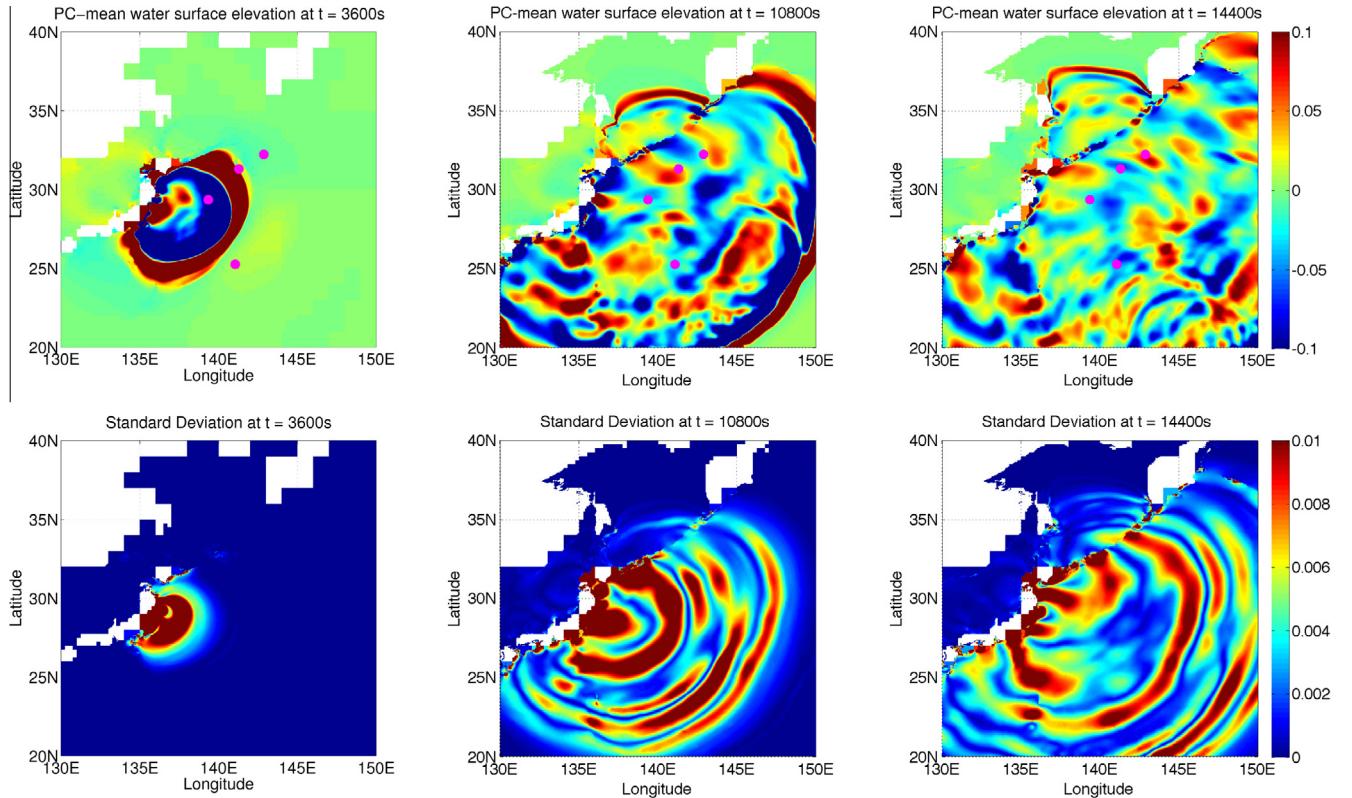


Fig. 9. PC mean (top row) and standard deviation (bottom row) of the water surface elevation at different times.

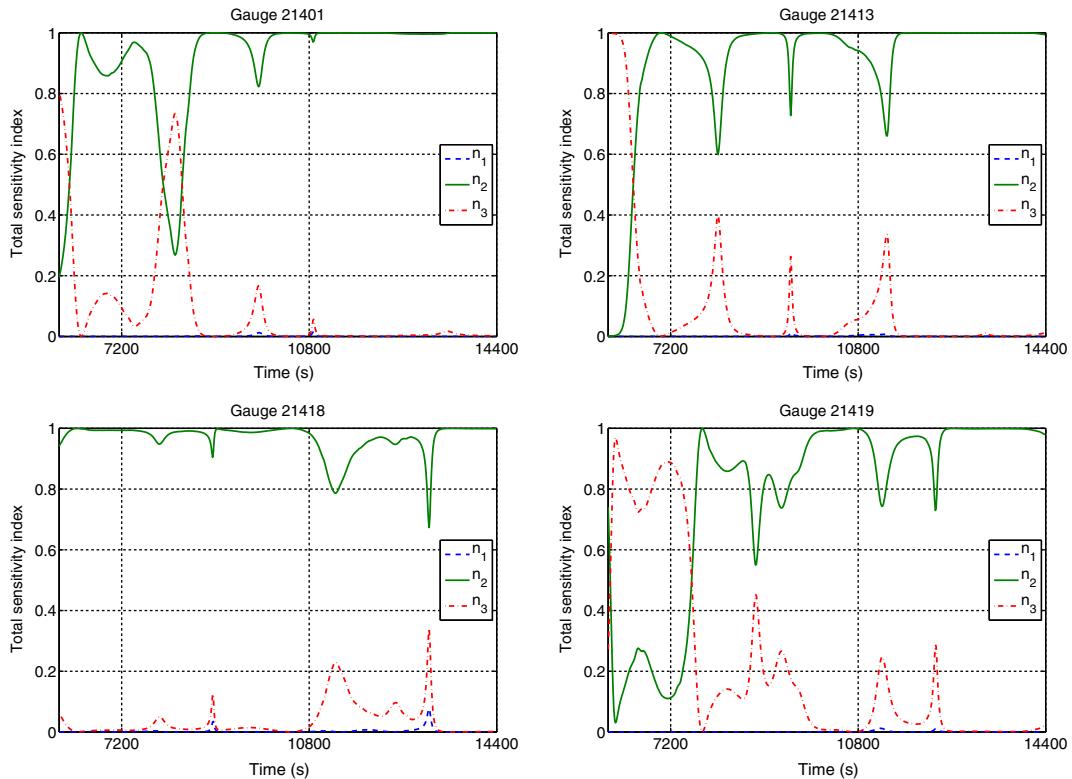


Fig. 10. Total sensitivity index of different input parameters.

lation time at all gauges. We note that we performed the same sensitivity analysis at the virtual near-shore gauges and found the same results affirming the dominance of n_2 over the other parameters.

In 2D, the sensitivity analysis shows also that n_2 is dominant at the same regions where we observe significant variance in water surface elevation. This is indicated from Fig. 11 that shows the total sensitivity index for n_1 (top row) n_2 (center row) and n_3 (bottom row) at different times. Note that the high sensitivity indices of both n_1 and n_3 are not reliable due to the corresponding insignificant variance in water surface elevation as shown in Fig. 9 (bottom row).

5.2.3. Response surface

In addition to the statistical and sensitivity analysis, the PC surrogate can be used to construct a response surface for the uncertain input parameters. This is achieved by sampling the PC surrogate for different values of the canonical vector of random variables ξ

within the prior range. Since n_1 shows insignificant contribution to the uncertainty in the model output, we only consider the response surface of n_2 versus n_3 for a fixed value of $n_1 = 0.035$ as illustrated in Fig. 12 at $t = 7200$ seconds. The most striking features are the relatively flat horizontal contours in the n_3 direction suggesting that water surface elevation depends only mildly on n_3 even during peak tsunami events. This is true for gauges 21401, 21413 and 21418. However, at gauge 21419, the horizontal contours are observed in the n_2 direction meaning that water surface elevation depends only mildly on n_2 .

5.3. Inverse problem

Bayesian inference is now used to estimate the Manning's n coefficient's ranges that minimize the scatter between observed and modeled water surface elevations, and to estimate the variance of the noise in the measured data. To this end, an adaptive MCMC method is implemented to sample the posterior distribu-

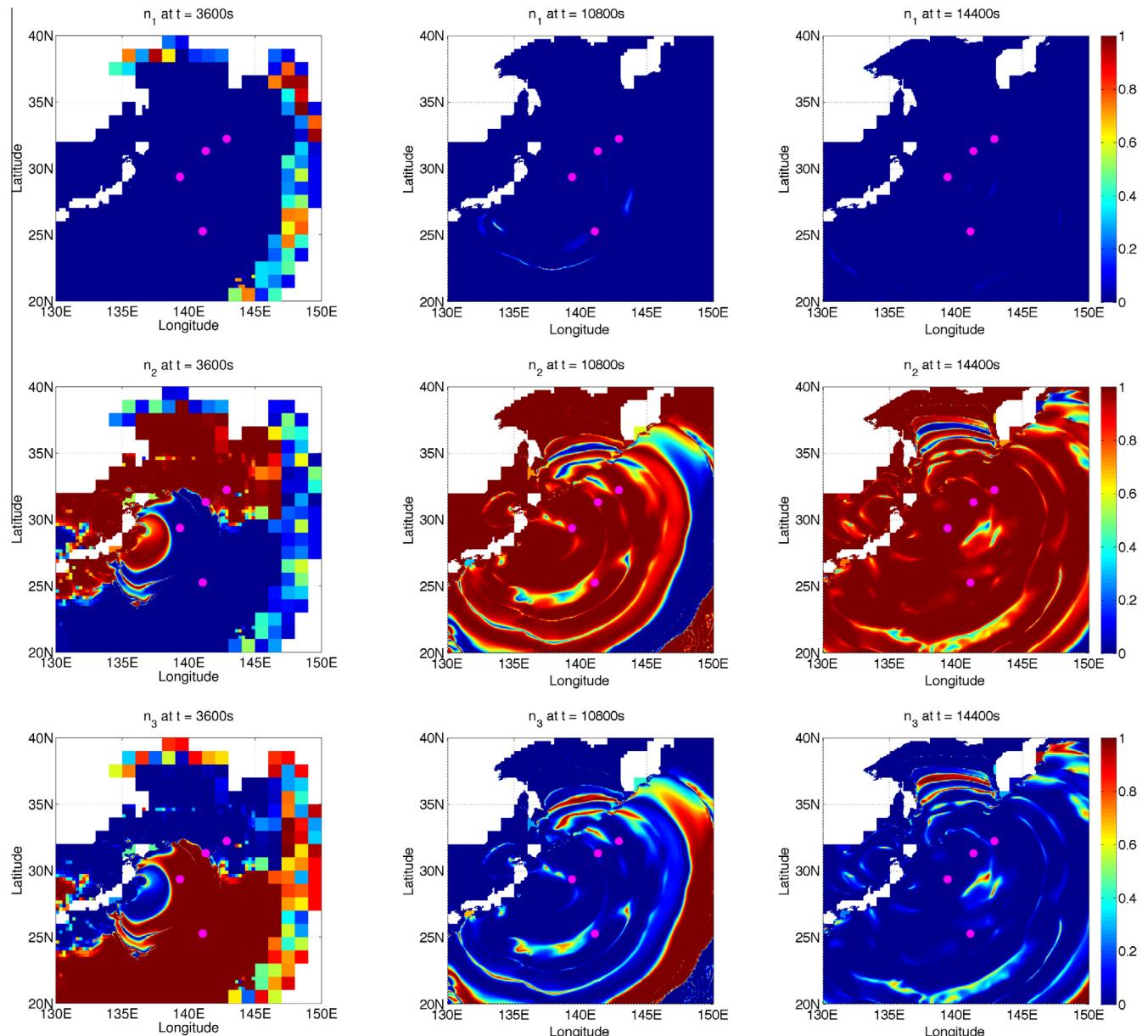


Fig. 11. Total sensitivity index for n_1 (top row) n_2 (center row) and n_3 (bottom row) at different times as indicated.

tions (Roberts and Rosenthal, 2009; Haario et al., 2001) and consequently update the Manning's n coefficients distributions in light of the observed data. This sampling requires tens of thousands of GeoCLAW model runs that are prohibitively expensive as each MCMC sample requires an independent GeoCLAW realization. Instead, the surrogate model created using PC expansions provides a computationally efficient alternative that requires only evaluating the PC expansion for different values of the canonical vector of random variables ξ .

All of the results presented here are based on 10^6 MCMC samples; we find negligible change in the obtained posteriors of the Manning's n coefficients: n_1 , n_2 and n_3 as well as for the noise variance σ_i^2 with further iterations. Fig. 13 (top row) shows the sample chains for the input parameters for different iterations of the

MCMC algorithm. The different panels show well-mixed chains for all input parameters where the chain of n_1 spans the entire range of the prior, meaning that the observations are not informative concerning this input. In contrast, the n_2 chain appears to be concentrated in the lower end of the parameter range, with values between 0.005 and 0.1. For n_3 , the chain appears to be concentrated in the upper end of the parameter range. These values align well with what is commonly considered physically relevant for near-shore and deep-water regions. Finally, the chain for the noise variances ($\sigma_1^2, \sigma_2^2, \sigma_3^2$) are shown in Fig. 13 (bottom row) for selected gauges and appear to be well mixed with a well defined posterior range. The maximum variance appears to be at gauge 21418 and its range between 0.04 and 0.08.

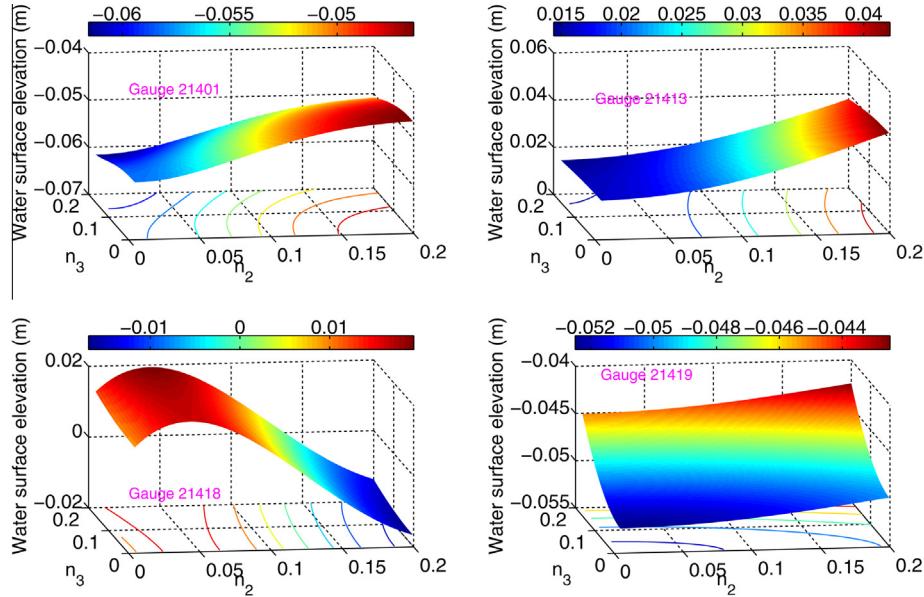


Fig. 12. Response surface of water surface elevation at the different gauge locations at $t = 7200$ s as function of n_2 and n_3 , with fixed $n_1 = 0.035$.

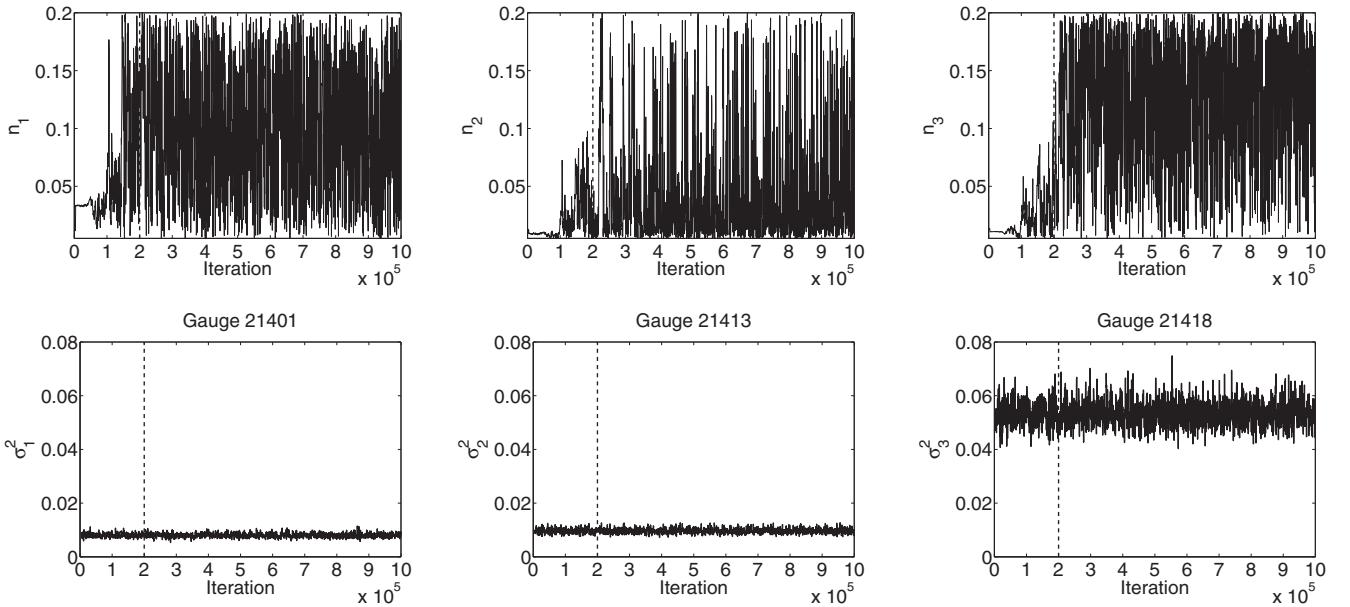


Fig. 13. Chain samples for the three Manning's n coefficients n_1 , n_2 , n_3 (top row) and the noise variance σ_1^2 , σ_2^2 , σ_3^2 at gauges 21401, 21413 and 21418 respectively (bottom row). The vertical dotted lines corresponds to the burn-in iterations.

Next, the computed MCMC chains are used to determine the marginalized posterior distributions using kernel density estimation (KDE) (Parzen et al., 1962; Silverman et al., 1986). The resulting marginalized posterior *pdfs* of the three Manning's n coefficients n_1, n_2, n_3 are shown in Fig. 14 (top row). Note that the first 2×10^5 iterations, associated with the burn-in period, were discarded. As expected from the chains shown in Fig. 13, the marginalized posterior *pdfs* of n_1 appear to be fairly flat, and

similar to the uniform prior; an indication that the observed data were not informative to refine our prior knowledge for n_1 . In contrast, the posterior *pdf* of n_2 exhibits a well-defined peak, with a Maximum A Posteriori (MAP) estimate of around 0.011 and an extended tail towards the larger Manning's n coefficient values. For n_3 , we observe a posterior that has a well defined peak of 0.18 but no clear *pdf* shape. The posterior distributions of the noise variances are also shown in Fig. 14 (bottom row) at selected

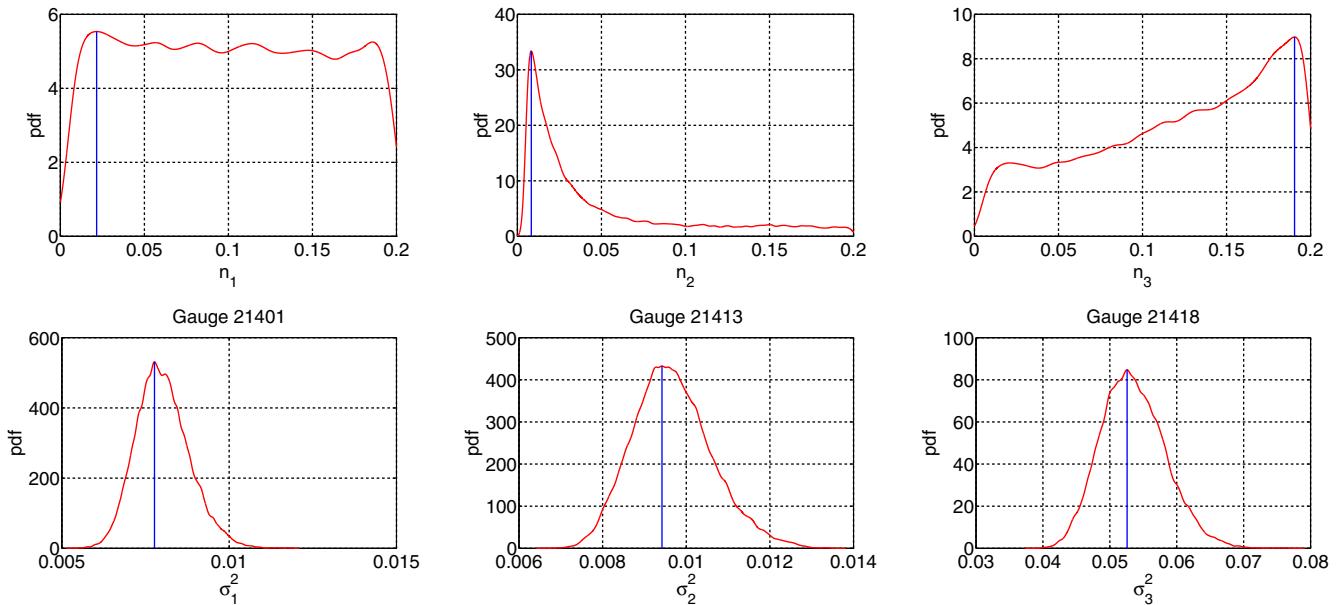


Fig. 14. KDE of the marginalized posterior distributions for the three Manning's n coefficients $\Pi(n_1), \Pi(n_2)$ and $\Pi(n_3)$ (top row), and for the noise variance $\Pi(\sigma_1^2), \Pi(\sigma_2^2), \Pi(\sigma_3^2)$ at gauges 21401, 21413 and 21418, respectively (bottom row).

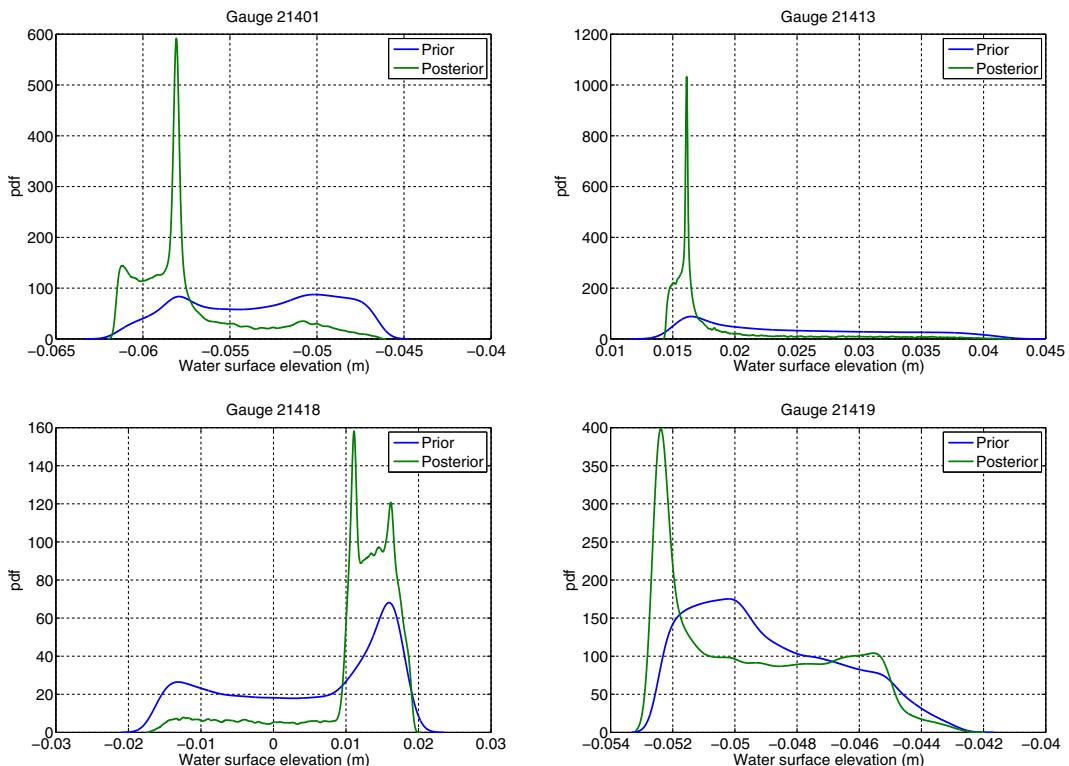


Fig. 15. Comparison of prior and posterior water surface elevation at different gauges.

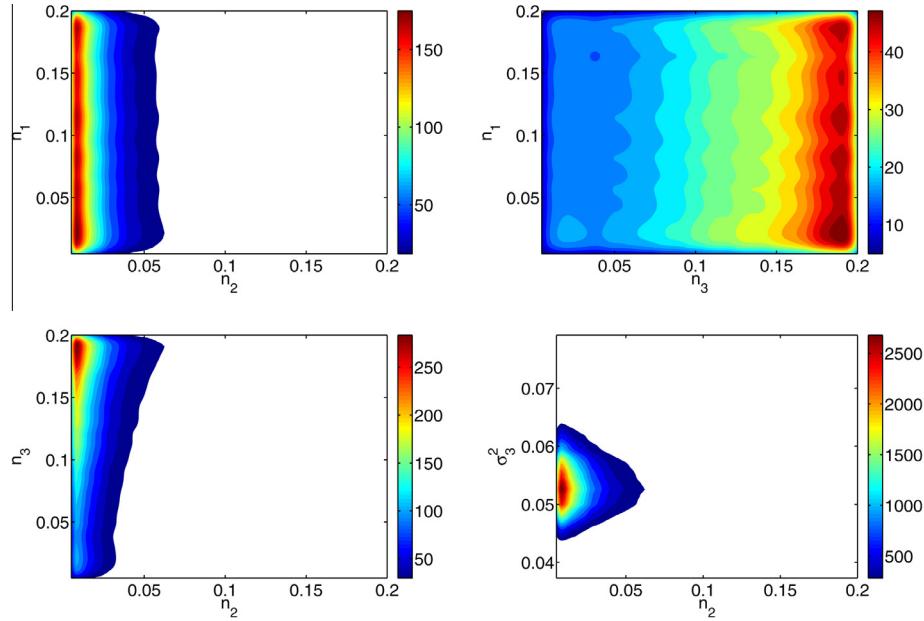


Fig. 16. Contour plots of the marginalized joint posterior distribution $\Pi(n_1, n_2)$, $\Pi(n_1, n_3)$, $\Pi(n_2, n_3)$ and $\Pi(n_2, \sigma_3^2)$. Contours of greater than 10% of the maximum probability are shown.

gauges. The *pdfs* appear to be well-defined of Gaussian-like shape with a clear MAP values. These MAP values can be used to estimate the maximum water surface elevation standard deviation that was found to be $\sigma_3 = 0.229m$ at gauge 21418. This value is a reflection of the mismatch between the model and observed data. The σ_i estimates are noticeably lower than those obtained with GEOCLAW default values (shown in Fig. 2): 0.0880 versus 0.0894, 0.0970 versus 0.0982, 0.2293 versus 0.2310, and 0.0759 versus 0.0765 at gauges 20401, 21413, 21418 and 21419 respectively. Thus, the parameters MAP values have reduced the discrepancies between simulated water surface elevation and DART buoy data. This comparison can be seen as an evaluation of the *a posteriori* goodness-of-fit. Additionally, to evaluate the goodness-of-fit we can generate the posterior distribution of the model output using the corresponding posterior distribution of the model parameters. In Fig. 15, we plot the prior and posterior distribution of water surface elevation at the DART gauges at $t = 7200$ s for comparison. The different panels show the improvement demonstrated by the leveling off of the uncertainties due to the prior distribution of the parameters n_i . Although the MAP values of the posteriors are not dramatically different from those of the prior the uncertainty bounds are minimized as the posterior distributions narrow. This is mainly due to the consequent narrowing in the posterior distribution of n_2 .

We finally investigate the existence of any correlation arising between the inferred parameters. It is indeed instructive to analyze the correlation between the various n parameters to gain some insight into their mutual dependence. To this end, we compute the two-dimensional marginalized joint posteriors for each possible combination of the parameters n_1 , n_2 and n_3 , namely $\Pi(n_1, n_2)$, $\Pi(n_1, n_3)$ and $\Pi(n_2, n_3)$ as well as for $\Pi(n_2, \sigma_3^2)$. The joint density for each combination is constructed from the chain samples of the marginalized posterior excluding the burn-in iterations via KDE. Fig. 16 shows the contour plots of the marginalized joint posteriors $\Pi(n_1, n_2)$, $\Pi(n_1, n_3)$ and $\Pi(n_2, n_3)$ in addition to $\Pi(n_2, \sigma_3^2)$ that are greater than 10% of the maximum probability. The plots reveal no clear correlation in all three cases.

To quantify the degree of no-correlation between each possible pair of parameters, we compute the corresponding correlation coefficient which is a unit-less scalar that falls in the range

(−1, 1), with 1 indicating a perfect positive correlation, −1 a perfect negative correlation and 0 for a no-correlation case. The correlation coefficient, $R(X, Y)$, between two random variables X and Y is defined as:

$$R(X, Y) = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}, \quad (20)$$

where σ_X and σ_Y are the standard deviations of X and Y , μ_X and μ_Y are the means of X and Y respectively; and E is the expectation operator. In our case, the correlation coefficient was in the range (−0.002, 0.06) for all possible combinations of parameters; again indicating no correlation between the inferred parameters.

6. Discussion and conclusions

The present study attempted to estimate the Manning's n friction coefficient which plays an important role in the accurate prediction of water surface elevation in tsunami modeling. We proposed a three-parameter representation of Manning's n friction coefficient that was characterized using iso-baths to define three distinct regions, on-shore, near-shore, and deep-water, in the domain and represented each region with a single Manning's n coefficient. The estimation relied on a Bayesian inference approach that sharpens the initial estimates of the three uncertain parameters using real observations. In our test case, the Tōhoku tsunami, we used four DART buoy gauges that provide surface elevation information in deep-water.

To accelerate the Bayesian inference, we relied on the Polynomial Chaos expansions to produce a faithful and efficient surrogate of the forward model GEOCLAW. This PC surrogate model was then used to quantify the uncertainties in the predicted water surface elevations due to the uncertainties in the Manning's n coefficient. This included the mean and standard deviation of water surface elevations which were also compared to the measured buoy data. A global sensitivity analysis was also performed in order to quantify the contribution of each uncertain parameter to the variance in surface elevation at the DART gauges. It was found that the Manning's n coefficient in the near-shore region n_2 contributed the

most to the variance in the surface elevations compared to the other two Manning's n coefficients in the on-shore and deep-water regions. This is expected as the deep-water friction has little effect on the overall column. On the other hand, the sensitivity of the analysis to the on-shore value being less than that in the deep-water was unexpected. We suspect that this is primarily due to the location of the DART buoy gauges being too far from land regions to have a significant impact, and to the nature of the deep-water gauge measurements. This might not be the case if data from inundation zones were used. The built PC model was also used to perform the same statistical and sensitivity analysis on several near-shore virtual gauges to assess the uncertainty in tsunami run up. We thus calculated and reported the mean and standard deviation of water surface elevation at locations near the shoreline of Japan. The sensitivity analysis at the near-shore gauges revealed the same aforementioned results.

Using Bayesian inference, MAP estimates were computed for the three region's n parameters using MCMC. These values were (excluding n_1 where no meaningful MAP value was found), $n_2 = 0.011$ and $n_3 = 0.180$. The values and their corresponding marginalized pdfs tell perhaps the most complete story in this analysis. As mentioned earlier, the surface elevations are most sensitive to n_2 which has the most peaked MAP value. Although the value was lower than expected, the tail includes the most commonly used values (0.022–0.025). The deep-water value n_3 has a peak that is much higher than expected but a tail that does not taper off particularly quickly. This is not altogether unexpected due to the low sensitivity of η to this region's friction. The on-shore value n_1 shows again how insensitive the simulations and observations were to on-shore values in our particular setting. In conclusion, it is clear that the use of only the DART buoy observations and the bathymetry resolution used here is not sufficiently informative to invert for the Manning's n values with any strong confidence, especially outside of the near-shore region.

Finally, it should be noted that the friction source term is only one of a number of other sources of uncertainty. In particular, the rupture model, bathymetry and model resolution can lead to large uncertainties. Several tests have been performed to assess the robustness of our PC methodology and the sensitivity of the inference results to the rupture model and model resolution. The impact of the rupture model was tested by adopting the [Saito et al. model \(2011\)](#) whose inversion was based purely on a tsunami inversion methodology. The PC surrogate was re-constructed and global sensitivity analysis re-performed revealing the same aforementioned results, namely that n_2 was the dominant uncertain parameter. The inference analysis was also repeated and the posterior distributions of n_i built using the same DART data; no significant differences in the results was observed. Increased resolution near the coastline was also applied and compared to model runs using less resolution; again no significant differences in the inference were observed. We believe that this is primarily due to the insensitivity of the deep-water DART gauges to the on-shore region. It is also important to recognize that the relatively small improvement in accuracy indicated in the small decrease in error as measured from the DART buoy data before and after the parameter inference leads to the conclusion that Manning's n friction has a relatively small impact on the off-shore gauges.

In the future, this type of analysis would be more illustrative if propagation of uncertainties in both the bathymetry and source models were included in each Manning's n estimate. The present study focused on formulating and estimating a low-dimensional representation of the Manning's n coefficient using UQ techniques, namely Bayesian inference and PC expansions. A high-dimensional representation of the Manning's n coefficient would require a large number of forward runs and this remains computationally expensive. We will work on alternative methods to reduce the

dimensionality of the problem and this will be the objective of a future study.

Acknowledgment

Research reported in this publication was supported by the King Abdullah University of Science and Technology (KAUST).

References

- Alexanderian, A., Le Maître, O., Najm, H., Iskandarani, M., Knio, O., 2011. Multiscale stochastic preconditioners in non-intrusive spectral projection. *J. Sci. Comput.*, 1–35. <http://dx.doi.org/10.1007/s10915-011-9486-2>.
- Alexanderian, A., Winokur, J., Sraj, I., Srinivasan, A., Iskandarani, M., Thacker, W., Knio, O., 2012. Global sensitivity analysis in an ocean general circulation model: a sparse spectral projection approach. *Comput. Geosci.* 16, 757–778. [<http://dx.doi.org/10.1007/s10596-012-9286-2>](http://dx.doi.org/10.1007/s10596-012-9286-2).
- Amante C., Eakins, B.W., ETOPO1 1 arc-minute global relief model: procedures, data sources and analysis, *Tech. Rep. NGDC-24*, NOAA, Mar. 2009.
- Ammon, 2011. A rupture model of the 2011 off the Pacific coast of Tohoku earthquake. *Earth Planets Space* 63 (7), 693–696.
- Berger, M.J., LeVeque, R.J., 1998. Adaptive mesh refinement using wave-propagation algorithms for hyperbolic systems. *SIAM J. Numer. Anal.* 35, 2298–2316.
- Berger, M.J., Oliger, J., 1984. Adaptive mesh refinement for hyperbolic partial differential equations. *J. Comput. Phys.* 52, 484–512.
- Cameron, R.H., Martin, W.T., 1947. The orthogonal development of non-linear functionals in series of Fourier-Hermite functionals. *Ann. Math.* 48, 385–392.
- Canuto, C.G., Hussaini, M.Y., Quarteroni, A., Zang, T.A., 2006. *Spectral Methods: Fundamentals in Single Domains*, Scientific Computation. Springer-Verlag.
- Crestaux, T., Maitre, O.L., Martinez, J.-M., 2009. Polynomial chaos expansion for sensitivity analysis. *Reliab. Eng. Syst. Saf.* 94 (7), 1161–1172 (Special Issue on Sensitivity Analysis).
- Das, S.K., Lardner, R.W., 1992. Variational parameter estimation for a two dimensional numerical tidal model. *Int. J. Numer. Methods Fluids* 15 (3), 313–327.
- Elsheikh, A.H., Hoteit, I., Wheeler, M.F., 2014. Efficient Bayesian inference of subsurface flow models using nested sampling and sparse polynomial chaos surrogates. *Comput. Methods Appl. Mech. Eng.* 269, 515–537. [<http://www.sciencedirect.com/science/article/pii/S004578251300296X>](http://dx.doi.org/10.1016/j.cma.2013.11.001).
- George, D.L., 2008. Augmented Riemann solvers for the shallow water equations over variable topography with steady states and inundation. *J. Comput. Phys.* 227 (6), 3089–3113.
- Haario, H., Saksman, E., Tamminen, J., 2001. An adaptive metropolis algorithm. *Bernoulli* 7 (2), 223–242 <<http://www.jstor.org/stable/3318737>>.
- Heemink, A.W., Moutahaan, E., Roest, M., Vollebregt, E., Robaczewska, K.B., Verlaan, M., 2002. Inverse 3D shallow water flow modelling of the continental shelf. *Cont. Shelf Res.* 22 (3), 465–484.
- Hiraishi, T., Yoneyama, N., Baba, Y., Azuma, R., 2011. Field survey of the damage caused by the 2011 off the Pacific coast of Tohoku earthquake tsunami. In: Kawase, H. (Ed.), *Studies on the 2011 Off the Pacific Coast of Tohoku Earthquake, Natural Disaster Science and Mitigation Engineering: DPRI reports*. Springer, Japan, pp. 37–48. <http://dx.doi.org/10.1007/978-4-431-54418-04>.
- Jakeman, J.D., Nielsen, O.M., Putten, K.V., Mleczko, R., Burbidge, D., Horspool, N., 2010. Towards spatially distributed quantitative assessment of tsunami inundation models. *Ocean Dyn.* 60 (5), 1115–1138.
- Lardner, R.W., Song, Y., 1995. Optimal estimation of Eddy viscosity and friction coefficients for a Quasi-three-dimensional numerical tidal model. *Atmos. Ocean* 33 (3), 581–611.
- Le Maître, O.P., Knio, O.M., 2010. *Spectral Methods for Uncertainty Quantification*. Springer-Verlag.
- LeVeque, R.J., 1997. Wave propagation algorithms for multidimensional hyperbolic systems. *J. Comput. Phys.* 131 (2), 327–353.
- Lin, G., Karniadakis, G.E., 2009. Sensitivity analysis and stochastic simulations of non-equilibrium plasma flow. *Int. J. Numer. Methods Eng.* 80 (6–7), 738–766. <http://dx.doi.org/10.1002/nme.2582>.
- MacInnes, B.T., Gusman, A.R., LeVeque, R.J., Tanioka, Y., 2013. Comparison of earthquake source models for the 2011 Tohoku event using tsunami simulations and near-field observations. *Bull. Seismol. Soc. Am.* 103 (2B), 1256–1274.
- Malinverno, A., 2002. Parsimonious Bayesian markov chain monte carlo inversion in a nonlinear geophysical problem. *Geophys. J. Int.* 151 (3), 675–688. [<http://dx.doi.org/10.1046/j.1365-246X.2002.01847.x>](http://dx.doi.org/10.1046/j.1365-246X.2002.01847.x).
- Marzouk, Y.M., Najm, H.N., 2009. Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems. *J. Comput. Phys.* 228 (6), 1862–1902. <http://dx.doi.org/10.1016/j.jcp.2008.11.024>.
- Mayo, T., Butler, T., Dawson, C., Hoteit, I., 2014. Data assimilation within the advanced circulation (adcirc) modeling framework for the estimation of Mannings friction coefficient. *Ocean Model.* 76, 43–58.

- Milburn, H.B., Nakamura, A.I., González, F.I., 1996. Real-time tsunami reporting from the deep ocean. In: Oceans 96 MTS/IEEE Conference, Fort Lauderdale, FL, pp. 390–394.
- Mungov, G., Eblé, M., Bouchard, R., 2012. DART tsunami retrospective and real-time data: a reflection on 10 years of processing in support of tsunami research and operations. *Pure Appl. Geophys.* 170 (9–10), 1369–1384.
- Myers, E.P., Baptista, A.M., 2001. Analysis of factors influencing simulations of the 1993 Hokkaido Nansei-Oki and 1964 Alaska tsunamis. *Nat. Hazards* 23 (1), 1–28.
- Najm, H., 2009. Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics. *Ann. Rev. Fluid Mech.* 41, 35–52.
- Olson, R., Srivastava, R., Goes, M., Urban, N.M., Matthews, H.D., Haran, M., Keller, K., 2012. A climate sensitivity estimate using Bayesian fusion of instrumental observations and an earth system model. *J. Geophys. Res.* 117, D04103.
- Parzen, E., 1962. On estimation of a probability density function and mode. *Ann. Math. Stat.* 33 (3), 1065–1076 <<http://www.jstor.org/stable/223788>>.
- Roberts, G.O., Rosenthal, J.S., 2009. Examples of adaptive MCMC. *J. Comput. Graph. Stat.* 18 (2), 349–367. <<http://dx.doi.org/10.1198/jcgs.2009.06134>>, <arXiv:10.1198/jcgs.2009.06134>.
- Saito, T., Ito, Y., Inazu, D., Hino, R., 2011. Tsunami source of the 2011 Tohoku-Oki earthquake, Japan: inversion analysis based on dispersive tsunami simulations. *Geophys. Res. Lett.* 38.
- Sarri, A., Guillou, S., Dias, F., 2012. Statistical emulation of a tsunami model for sensitivity analysis and uncertainty quantification. *Nat. Hazards Earth Syst. Sci.* 12 (6), 2003–2018. <<http://dx.doi.org/10.5194/nhess-12-2003-2012>> <<http://www.nat-hazards-earth-syst-sci.net/12/2003/2012/>>.
- Silverman, B.W., 1986. Density Estimation: For Statistics and Data Analysis. Chapman and Hall, London.
- Sivia, D.S., 2006. Data Analysis – A Bayesian Tutorial. Oxford Science Publications, Oxford.
- Sraj, I., Iskandarani, M., Srinivasan, A., Thacker, W.C., Winokur, J., Alexanderian, A., Lee, C.-Y., Chen, S.S., Knio, O.M., 2013a. Bayesian inference of drag parameters using Fanapi AXBT data. *Mon. Weather Rev.* 141, 2347–2367. <<http://dx.doi.org/10.1175/MWR-D-12-00228.1>>.
- Sraj, I., Iskandarani, M., Srinivasan, A., Thacker, W.C., Knio, O.M., 2013b. Drag parameter estimation using gradients and hessian from a polynomial chaos model surrogate. *Mon. Weather Rev.* 142, 933–941.
- Srinivasan, A., Helgers, J., Paris, C.B., LeHenaff, M., Kang, H., Kourafalou, V., Iskandarani, M., Thacker, W.C., Zysman, J.P., Tsinoremas, N.F., Knio, O.M., 2010. Many task computing for modeling the fate of oil discharged from the deep water horizon well blowout. In: 2010 IEEE Workshop on Many-Task Computing on Grids and Supercomputers (MTAGS). IEEE, pp. 1–7. <<http://dx.doi.org/10.1109/MTAGS.2010.5699424>> <<http://dx.doi.org/10.1109/MTAGS.2010.5699424>>.
- Sudret, B., 2008. Global sensitivity analysis using polynomial chaos expansions. *Reliab. Eng. Syst. Saf.* 93 (7), 964–979. <<http://dx.doi.org/10.1016/j.ress.2007.04.002>> (Bayesian Networks in Dependability).
- Thacker, W.C., Srinivasan, A., Iskandarani, M., Knio, O.M., Henaff, M.L., 2012. Propagating boundary uncertainties using polynomial expansions. *Ocean Model.* 43–44, 52–63. <<http://dx.doi.org/10.1016/j.ocemod.2011.11.011>>.
- Verlaan, M., Heemink, A.W., 1997. Tidal flow forecasting using reduced rank square root filters. *Stochastic Hydraul. Hydraul.* 11 (5), 349–368.
- Villegas, M., Augustin, F., Gilg, A., Hmaidi, A., Wever, U., 2012. Application of the polynomial chaos expansion to the simulation of chemical reactors with uncertainties. *Math. Comput. Simul.* 82 (5), 805–817. <<http://dx.doi.org/10.1016/j.matcom.2011.12.001>>.
- Xiu, D., Tartakovsky, D., 2004. Uncertainty quantification for flow in highly heterogeneous porous media. In: Cass, W.G.C., Miller, T., Farthing, Matthew W., Pinder, G.F. (Eds.), Computational Methods in Water Resources: Volume 1, Part 1 of Developments in Water Science, vol. 55. Elsevier, pp. 695–703.
- Xiu, D., Xiu, G., 2003. Modeling uncertainty in flow simulations via generalized Polynomial Chaos. *J. Comput. Phys.* 187, 137–167.
- Young, K.C., Grace, M.D., 2013. Simulation of stochastic quantum systems using polynomial chaos expansions. *Phys. Rev. Lett.* 110, 110402. <<http://dx.doi.org/10.1103/PhysRevLett.110.110402>> <<http://link.aps.org/doi/10.1103/PhysRevLett.110.110402>>.
- Zedler, S., Kanschat, G., Korty, R., Hoteit, I., 2012. A new approach for the determination of the drag coefficient from the upper ocean response to a tropical cyclone: a feasibility study. *J. Oceanogr.* 68 (2), 227–241.