The zero problem: Gaussian process emulators for range constrained computer models

Elaine T. Spiller*, Robert L. Wolpert†, Pablo Tierz‡, and Taylor G. Asher §

Abstract. We introduce a zero-censored Gaussian process as a systematic, model-based approach to building Gaussian process emulators for range-constrained simulator output. This approach avoids many pitfalls associated with modeling range-constrained data with Gaussian processes. Further, it is flexible enough to be used in conjunction with statistical emulator advancements such as emulators that model high-dimensional vector-valued simulator output. The zero-censored Gaussian process is then applied to two examples of geophysical flow inundation which have the constraint of non-negativity.

Key words. Gaussian processes, statistical surrogates, censored data, geophysical flows

AMS subject classifications. 60G15, 86-08, 62N01

1. Introduction.

Gaussian process based surrogates of computationally intensive models have become an essential class of tools for uncertainty quantification since the seminal papers led by Currin, Sacks, and Welch (Currin et al., 1988; Sacks et al., 1989b,a; Welch et al., 1992). The flexibility of Gaussian processes to model computationally intensive problems from a wide breadth of applications is remarkable. One challenging class of problems are computer models whose output range is constrained by minimum and/or maximum values. A common subset of these problems are computer models whose output is positive or zero. This "zero problem" poses great challenges in fitting Gaussian process emulators (GPs). To start, data with large numbers of zeros are not naturally modeled by Gaussian probability density functions due to their full support. Yet it is advantageous to leverage the vast body of work over the last few decades – both theory and techniques – on emulating simulators with Gaussian processes. As such we introduce a simulation based strategy to model bounded computer model output that addresses the semi-binary nature of the data and results in a GP model with full support. For the case of nonnegative data taking the value zero with positive probability, our approach begins by modeling the data as the maximum of zero and a latent Gaussian process. The challenge remaining is to find or approximate the intractable posterior distribution of that latent GP given the data.

An interesting and important class of models that suffer from the "zero problem" are geophysical flows. Consider inundation from tsunamis, volcanic flows, storm surge, etc. A given computer model run, representing one possible scenario, of any of these processes, will output the depth of inundation over a spatial region of interest. Such simulations are computationally intensive, taking minutes to days to complete a single simulator run on a super computer.

*Department of Mathematical and Statistical Sciences, Marquette University, Milwaukee, WI, USA (elaine.spiller@marquettet.edu).

†Department of Statistical Science, Duke University, Durham, NC, USA.

‡British Geological Survey, The Lyell Centre, Edinburgh, UK, and The School of Geosciences, University of Edinburgh, Edinburgh, UK.

§Department of Earth, Marine, and Environmental Sciences, University of North Carolina, Chapel Hill, NC, USA.

Hazard analysis or hazard forecasting typically relies on Bayesian simulation-based inference methods that require thousands to millions of simulation runs. Given these constraints, hazard analysis is nearly infeasible using full model evaluation of the simulator. Likewise exploring hazard analyses under various potential aleatory scenarios and/or quantifying epistemic uncertainties in such analyses with direct computer model evaluations is intractable. As such, computationally efficient surrogate models that can address the "zero-problem" have the potential to greatly advance the field of geophysical hazard analysis.

For simulators with vector-valued outputs that are range constrained, the full support of Gaussian processes is not the only challenge for emulation. In particular, the regions of input space that lead zero-output can (and often do) differ for each element of the output vector. In the context of geophysical flows, the boundary in scenario space that leads to zero output or positive output, varies spatially among output map nodes (point of interest inside the hazard domain). Consider a batch of simulator runs covering a wide range of potential scenarios, here each element of the vector-valued output element represents a map node. Further, each will have its own set of runs resulting in positive inundation and set of runs resulting in no inundation. Clearly this kind of model output data is non-stationary, but has the added challenge that the non-stationarity is indicated by a discontinuity in the derivative of the GP.

Our group and others have made significant advances in GP-based probabilistic hazard assessment, probabilistic hazard forecasting, and probabilistic hazard mapping over the last decade (Bayarri et al., 2009, 2015; Beck and Guillas, 2016; Jia et al., 2016; Liu and Guillas, 2017; Rutarindwa et al., 2019). These various works address the large-dimensional spatial nature of the output by fitting emulators independently, by applying partial-parallel emulation (PPE), or by fitting emulators to coefficients of basis functions or principal components (GP-PCA) (Spiller et al., 2014; Gu and Berger, 2016; Higdon et al., 2008). In this work, we do not advocate for a particular choice of handling high-dimensional output, but instead provide a solution to the zero problem that will be suitable to work with any of these techniques. Various previous approaches to the zero problem in the works cited in this paragraph include: trying to ignore it; focusing on spatial regions that are inundated under every scenario; inputing missing (zero) data via spatial interpolation; including only a subset of zeros that are nearest in design space to simulations resulting in positive output at a given node. All of these approaches are rather ad-hoc (although some work quite well) and this particular form of non-stationarity remains a significant challenge for GP emulator-based geophysical hazard analysis.

Several GP emulation methods have been proposed to handle non-stationarity and/or discontinuous data. Many of these approaches are based on partitioning the input space and then either fitting separate GPs to the different regions or taking mixtures of input-region specific kernels to fit the GP (Gramacy and Lee, 2008; Pope et al., 2019; Volodina and Williamson, 2020). Yet for the zero problem, such a partition of input space would necessarily differ for each map node as the set of zero outputs varies spatially. Even if one could automate map node specific partitions, it is not clear how global emulator approaches – like parallel partial emulation or GPs fit to PCA modes – could be applied. Instead, we model the data as the maximum of zero and a latent GP and then, for each map node, we consider imputing negative GP values at design points whose output is zero, from a conditional distribution consistent with the simulator data. Once this preprocessing step is complete, the new partially-imputed model design and response set will fit assumptions needed for any of the high-dimensional GP output emulator approaches. As such the imputation approach employed by the zero-constrained Gaussian process is an *enabling technology* – it allows GP emulation, and variations to fit large-dimensional spatial output, of geophysical models that GP emulators are otherwise poorly suited to model.

There are several recent approaches to developing range constrained GPs in the computer modeling community which are largely inspired by the geostatistics paper on kriging with inequality constraints (Abrahamsen and Benth, 2001). There are two general approaches taken, the first of which relies on choosing constrained basis functions or constrained splines and modeling the associated coefficients with (truncated) Gaussian processes (Ben Salem et al., 2019; López-Lopera et al., 2018; Maatouk and Bay, 2017; Swiler et al., 2020). The common thread of the second approach is to fit all available model data and impute a set of "artificial data" throughout the input space points that maintain the constraint. These auxiliary data are subsequently used for fitting Gaussian processes (Agrell, 2019; Wang and Berger, 2016; Da Veiga and Marrel, 2012, 2020). One other recent work sets up the constrained optimization problem to optimize range parameters under a slightly-relaxed constraint that the predictive GP obeys the range constraint at untested inputs with high probability (Pensoneault et al., 2020). Some ideas of censored GPs are explored in (Kyzyurova, 2017), but are undeveloped. In this work we propose an approach that addresses the non-stationary nature of semi-binary data and that can be readily "plugged-in" to existing GP approaches that handle high-dimensional output.

Because of the non-negativity constraint we cannot take a surrogate to have a multivariate Normal distribution, but we can still leverage the vast development of Gaussian Process technology by constructing a surrogate of computer model output that takes on the maximum of zero and a GP that is constrained to fit the positive output data. Again, we refer to such a process as a *zero-censored Gaussian Process*, or "zGP." After introducing notation and GP basics, we go through the zGP construction noting important details for successful and efficient algorithm implementation including the choice of mean trend, initialization, and zGP parameterization that uses "zero" information and captures uncertainty in the modeling due to imputation. We then demonstrate the zGP's efficacy by applying it to two different hazardous geophysical flows: storm surge and granular volcanic flows.

2. Background.

2.1. Gaussian Process Emulation. In the simplest sense, Gaussian process emulation can be thought of as a statistical model of a complicated and computationally intensive physical model. The idea is to treat the computer model response as coming from a random function in the class of weakly stationary Gaussian processes. To do so, we will only consider random functions that are conditioned on going through (or near) the computer model output data. Determining parameters of a GP that are consistent with the computer model response is described as "fitting" the GP. Once the GP model is determined, one can replace the computationally intensive computer model simulations with a function evaluation (Welch et al., 1992; Santner et al., 2018).

Starting with notation, let $\mathbf{x}$ be a p-dimensional vector of inputs to the computer model, lying in a domain $\mathcal{X} \subseteq \mathbb{R}^p$ of possible values – so $\mathbf{x} = (x_1, \ldots, x_p)^\mathsf{T} \in \mathcal{X} \subseteq \mathbb{R}^p$. This vector is typically comprised of initial conditions, parameters, and/or boundary conditions needed to specify completely a single computer model run. In the context of inundation from geophysical flows, the input vector would represent one possible scenario. Likewise, we will denote the computer model output as $y^M(\mathbf{x})$ – for the applications explored in this work, that is the (necessarily nonnegative) maximum depth of flow inundation from a geophysical simulation for the scenario characterized by $\mathbf{x}$. Consider n space-filling computer model runs, *i.e.*, n scenarios (indexed by $j \in J$) typically called the *design*, and denote that design as $\mathcal{D} = \{\mathbf{x}_j : j \in J\}$. The output from all design runs is taken together as $\mathbf{y}^M = (y_1^M, \ldots, y_n^M)^\mathsf{T} \in \mathbb{R}^n$. Lastly, we will denote the resulting design input-output pairs as $\mathcal{D}^M = \{(\mathbf{x}_j, y_j^M) : j \in J\}$.

Now we will treat this computer model output data as a random vector with components $y_j^M = Z_j$, with $\{Z_j \sim \mathsf{No}(\mu, \Sigma) : j \in J\}$ where $\mu_j = \mu(\mathbf{x}_j)$ is a known mean trend function which may implicitly depend on uncertain parameters. The matrix, $\Sigma = \sigma^2\mathbf{R}$, is an $n \times n$ covariance matrix comparing the design points in $\mathcal{D}$. One can calculate $\big(\mathbf{R}\big)_{i,j} = c(\mathbf{x}_i, \mathbf{x}_j)$ using a covariance function $C(\cdot,\cdot) = \sigma^2 c(\cdot,\cdot)$ with scalar variance σ^2. Throughout this work, we will utilize a separable Matérn 5/2 correlation function (see Stein (1999, §2.10) for arguments supporting this choice). For two inputs $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^\mathsf{T}$ and $\mathbf{x}_j = (x_{j1}, \ldots, x_{jp})^\mathsf{T}$, the standardized distance and correlation are:

$$
\begin{aligned}
d_k &= \Big(\frac{|x_{ik} - x_{jk}|^2}{\rho_k^2}\Big)^{1/2} \\
c(\mathbf{x}_i, \mathbf{x}_j) &= \prod_{k=1}^{p} \Big(1 + \sqrt{5}d_k + \frac{5}{3}d_k^2\Big) \exp\big(-\sqrt{5}d_k\big).
\end{aligned} \tag{2.1}
$$

The *range parameters* $\{\rho_k : k = 1, \ldots, p\}$, along with parameters describing the mean function $\mu(\cdot)$ comprise the set of parameters needed to define a GP, and we call these parameters θ. With an estimate $\hat\theta \approx \theta$ in hand (note hatted quantities represent estimates), we can generate predictions of the computer model output at untried points (indexed by $i \in I$) with

$$Z_I \sim \mathsf{GP}\big(\mu_I, \Sigma_{II} \mid \mathcal{D}^M, \theta\big) = \mathsf{GP}\big(m_{I|J}, V_{I|J}\big) \tag{2.2}$$

with conditional mean vector and covariance matrix given by the usual Gaussian formulas:

$$m_{I|J} = \mathsf{E}[Z_I \mid \mathcal{D}^M, \theta] = \mu_I + \Sigma_{IJ}\Sigma_{JJ}^{-1}(Z_J - \mu_J) \tag{2.3}$$

$$V_{I|J} = \mathsf{E}\big[(Z_I - m_{I|J})(Z_I - m_{I|J})^\mathsf{T} \mid \mathcal{D}^M, \theta\big] = \Sigma_{II} - \Sigma_{IJ}\Sigma_{JJ}^{-1}\Sigma_{JI} \tag{2.4}$$

In practice we must use an estimate $\hat\theta \approx \theta$. Going forward, we will suppress the dependence on θ in our notation, and will sometimes let the $I|J$ be implicit where no confusion arises.

The crux of this paper is adapting and applying this modeling strategy when the computer model output data, $\mathbf{y}^M$, has range constraints. In particular, we will focus on the constraint that the output data is non-negative, but the methodology we develop here would also apply to other minimum and/or maximum value restrictions on the output data.

3. Methodology.

3.1. Motivation. Our two motivating applications are both geophysical flows that can result in hazardous inundation, namely inundation due to storm surge and inundation due to rapid granular volcanic flows known as pyroclastic density currents (PDCs). Both phenomena are modeled by hyperbolic partial differential equations (PDEs) numerically solved over digital elevation models (DEMs). Such computer models are computationally intensive, and a typical simulation – depending on the scenario considered along with the desired accuracy of the solver – can take hours to days to run on a high performance computing system (further details on these computer models will be given in section 4). Another commonality between these simulators is the complicated spatial footprints of inundation heights that result as output. In Fig. 1 (left) we see the simulated spatial extent and maximum PDC flow depth (color) of two different but typical simulations. Likewise in Fig. 1 (right) we see maximum storm surge inundation for four different simulations (*i.e.*, four differently parameterized storms) at a set of over-water and over-land map nodes. Of the 908 map nodes where storm surge depth is reported, simulated storms labeled (a)–(d) in Fig. 1 yielded 382, 370, 237, and 290 zero-output (or "dry") nodes, respectively.

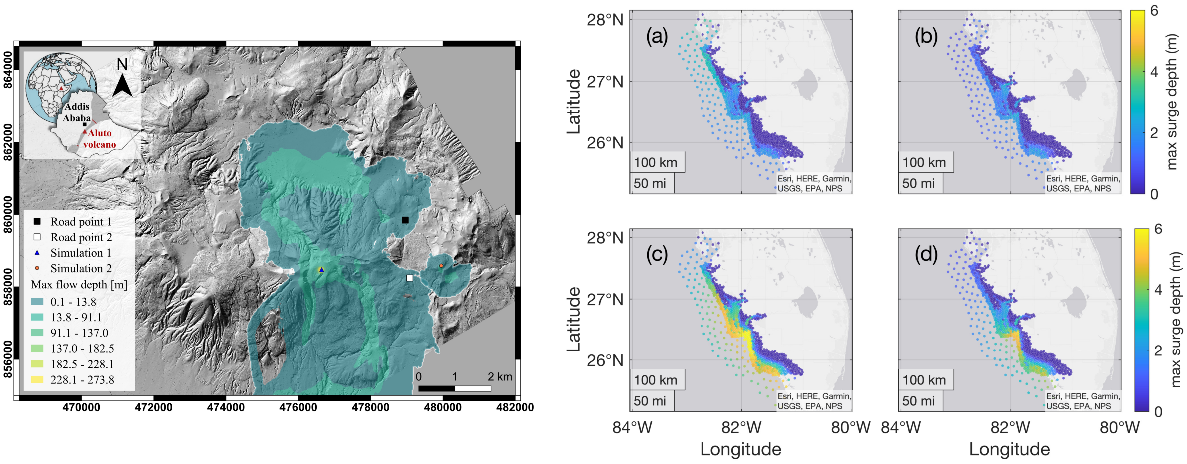

Figure 1: Left: Two simulated max PDC flow depths from flows that originate at different vent locations (blue triangle and orange circle) at Aluto volcano, Ethiopia (see simplified geographical context in the top-left corner of inset.) Note how the PDC simulation that originated at the blue triangle inundates both road points of interest (white and black squares) while the PDC simulation that originated at the orange circle almost inundates the white square road point, but does not come close to inundating the black square road point. Right: Four storms surge simulated max inundation depths on a grid of map nodes both overland and over water on the Southwest coast of Florida, USA. The darkest blue color indicate no inundation at those nodes.

Our strategy is to impute negative values for the zero-outputs that are consistent with GPs fit to the positive model response. In particular, this approach readily distinguishes between simulations that *almost inundate* a given node from those that do not. To elucidate the zGP approach, we will explain and apply it to a scalar output illustrative example as we introduce it.

3.1.1. An illustrative example.. We begin with a pedagogical example to illustrate the approach and introduce the necessary notation. We specify a deterministic function h on the domain $\mathcal{X} = [0,1]^2 \subset \mathbb{R}^2$, playing the role of a deterministic computer model with input space $\mathcal{X}$, and try to reconstruct it from a design set $\mathcal{D}^M = \{\mathbf{x}_j, y_j : j \in J\}$ with $y_j = h(\mathbf{x}_j)$. We begin with a slightly modified toy function of Bastos and O'Hagan (2009) shifted vertically, given as $h(\cdot) = 0 \vee f(\cdot)$, where with $\mathbf{x} = (x_1, x_2)$ and

$$f(\mathbf{x}) = \Big(1 - \exp\big(-\frac{1}{2x_2}\big)\Big)\Big(\frac{2300x_1^3 + 199x_1^2 + 2092x_1 + 60}{100x_1^3 + 500x_1^2 + 4x_1 + 20}\Big) - 6. \tag{3.1}$$

The toy function, h, along with $n = 50$ Latin hypercube (LHC) design-response pairs, are plotted in Fig. 2. Note for the design used in this example, there are 26 design points that lead to a zero response and 24 that lead to a positive response.

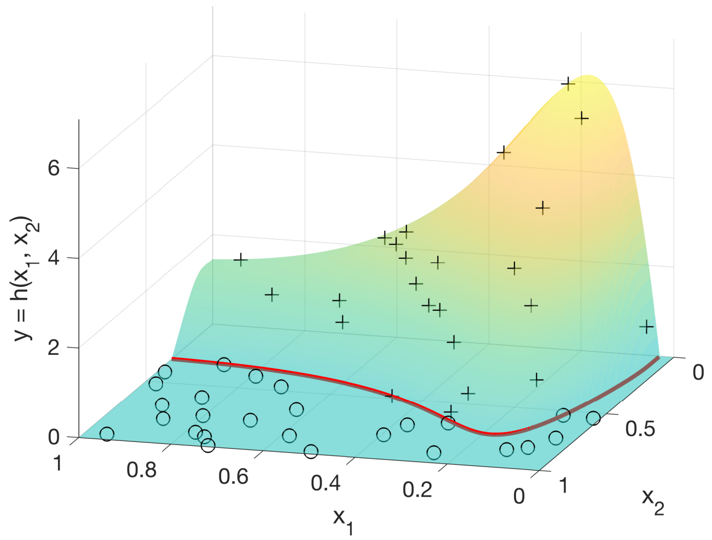

Figure 2: The non-negative function, $h = 0 \vee f$, plotted along with design points/responses that resulted in positive (+) and zero (∘) model output. To add some contrast to the visualization, we have also included a red line indicating the zero-contour of f.

3.2. Zero-censored Gaussian Process. Again, our design consists of a finite set $\mathcal{D} = \{(\mathbf{x}_j) : j \in J\}$, but we now consider the case where $\mathcal{D}^M = \{(\mathbf{x}_j, y_j^M) : j \in J\}$ are ordered pairs of observed nonnegative scalar output values, $y_j^M \in \mathbb{R}_{\geq 0}$, of a computer simulator at model input vectors $\mathbf{x}_j \in \mathcal{X} \subset \mathbb{R}^p$ ($p = 2$ in the illustrative example), all indexed by a finite set J. We can think of each input vector, $\mathbf{x}_j$, representing a distinct *model scenario* or one choice of model inputs that parameterizes a particular realization of the simulator. The model output is *strictly* positive for some number $n_+ := |J_+|$ of indices $J_+ := \{j \in J : y_j^M > 0\}$ ($n_+ = 24$ in the illustrative example), but may take on the exact value $y_j^M = 0$ at some number $n_- := |J_-|$ of indices $J_- := \{j \in J : y_j^M = 0\}$ ($n_- = 26$ in the illustrative example), for a total of $n = n_- + n_+ = |J|$ (here $n = 50$) design points. Note, if the simulator output is vector-valued, we will proceed with this imputation approach by treating each output vector element independently. This choice is motivated by the fact that each vector element will have its own set of design points that lead to positive outputs and to zero outputs. In other words, each output vector element will have its own J_+ and J_-. Obviously, this will add some computational burden to the imputation, but that burden is somewhat alleviated by noting that the imputations can be done for each element of the vector in parallel as a preprocessing step. Through the rest of this section, we will describe the zGP imputation for scalar valued output.

We construct a random field stochastic model $\{\mathbf{x} \rightsquigarrow Z(\mathbf{x}) : \ \mathbf{x} \in \mathcal{X}\}$ which we view as a joint prior distribution for the model outputs $\{y^M\}$ at all possible input points $\{\mathbf{x} \in \mathcal{X}\}$, and then seek the posterior distribution of $\{y^M\}$ at all locations $\{\mathbf{x} \in \mathcal{X}\}$, *conditional* on Z agreeing with the design, $Z(\mathbf{x}_j) = y_j^M$ for $j \in J$. Because of the nonnegativity constraint we cannot take $\{Z(\mathbf{x})\}$ to have a multivariate Normal distribution, but we can still leverage the vast development of Gaussian Process technology by modeling $Z := 0 \vee \zeta$ as the maximum of zero and a GP $\zeta \sim \mathsf{GP}(\mu, \Sigma)$ with some mean function $\mu(\mathbf{x})$ and covariance function $\Sigma(\mathbf{x}, \mathbf{x}')$ on $\mathcal{X}$ and $\mathcal{X}^2 = \mathcal{X} \times \mathcal{X}$, respectively. This is the aformentioned *zero-censored Gaussian Process*, or more succinctly, the zGP. In practice we take the mean function, $\mu(\mathbf{x})$ to be of very simple form— usually either a constant (possibly zero) or a linear function— and take $\Sigma(\mathbf{x}, \mathbf{x}')$ to be from the Matérn class with smoothness parameter 5/2 (see Eqn(2.1)).

The conditional distribution (and even the conditional mean) for $\zeta(\mathbf{x}_i)$ at unobserved locations in input space $\{\mathbf{x}_i \in \mathcal{X} : \ i \in I\}$, given $Z(\mathbf{x}_j) \equiv 0 \vee \zeta(\mathbf{x}_j) = y^M(\mathbf{x}_j)$ for $j \in J$, are unavailable in closed form. To facilitate inference we propose to draw simulations of $\zeta(\mathbf{x}_I) := \{\zeta(\mathbf{x}_i) : \ i \in I\}$ of the GP ζ at finite sets I of new input vectors $\mathbf{x}_i$, given $Z(\mathcal{D}) = y^M(\mathcal{D})$. We can then estimate posterior expectations of $Z(\mathbf{x}_I)$ itself or of functions of $Z(\mathbf{x}_I)$ with ergodic sample averages from these simulations. Even this task is challenging, since the conditional distribution of $\zeta(\mathbf{x}_I)$ constrained to go through non-negative output-design pairs – *i.e.*, given $\zeta(\mathbf{x}_{J_+}) = y_{J_+}^M$ and the condition $\{(\forall j \in J_-)\ \zeta(\mathbf{x}_j) \le 0\}$ – is intractable.

We address this in two steps. First, we use a *substitution sampling* scheme to make a series of imputed independent draws from the conditional distribution of $\zeta(\mathbf{x}_{J_-})$, given $\zeta(\mathbf{x}_{J_+}) = Z(\mathbf{x}_{J_+})$ and the event $\zeta(\mathbf{x}_{J_-}) \le 0$ (*i.e.*, given $Z(\mathbf{x}_J) = y_J^M$). We can then view $\zeta(\mathbf{x}_J)$ as a fully observed draw from the $\mathsf{GP}(\mu, \Sigma)$ distribution, with a known n-variate Normal distribution. For each of those imputed draws we draw $\zeta(\mathbf{x}_I)$ from its conditional distribution (using the usual Gaussian formulas) or, if only the mean and variance of some $\zeta(\mathbf{x}_i)$ are of interest, evaluate those in closed form. Algorithm 1 implements this approach. For the reader unfamiliar with substitution sampling, we preface each step with a brief explanation in italics. In this algorithm, we assume that the estimated GP parameter vector, $\hat{\theta} \approx \theta$ is known. A natural first approach is to use $\hat{\theta}$ obtained from fitting a Gaussian process to $(\mathbf{x}_{J_+}, y_{J_+}^M)$. In Section 3.3 we explore an approach to incorporate information from "nearby" zeros in estimating $\hat{\theta}$.

Algorithm 1: zGP substitution sampling. To construct a zGP sample of size $K \in \mathbb{N}$, for each index $1 \le k \le K$ we:

0) *Begin with an initial sample of output response values that are identical to positive output from the simulator for inputs* $\mathbf{x}_{J_+}$ *and are negative for inputs* $\mathbf{x}_{J_-}$*. A systematic way to achieve this initial sample is described in Algorithm 2, but as long as the constraints are met, any initial sample should work.*

   Begin with an initial set $\zeta^{(0)}$ at step $t = 0$ of candidate imputed values at locations $\mathbf{x}_J$, with $\zeta^{(0)}(\mathbf{x}_{J_+}) = y^M_{J_+}$ and with $\zeta^{(0)}(\mathbf{x}_{J_-}) < 0$.

1) *Select, at random, one of input design points indexed by* J_-*. In other words, select one input design point from those that led to an output of zero. Using all of the other design points except the one selected (*i.e.*, other points indexed by* J_- *and all points indexed by* J_+*), construct a GP conditioned to go through these design/response pairs, using the current value of the negative imputed responses corresponding to the remaining* $\mathbf{x}_{J_-}$*. Sample this GP at the selected design point from its (tractable) truncated Gaussian distribution and replace its current imputed response value with this new, negative sample.*

   Select $j^* \in J_-$ uniformly at random. Construct $\zeta^{(t+1)}$ by setting $\zeta^{(t+1)}(\mathbf{x}_j) = \zeta^{(t)}(\mathbf{x}_j)$ for $j \in J_c$ where $J_c = J \backslash j^*$ and for $\zeta^{(t+1)}(\mathbf{x}_{j^*})$ take a random draw from the truncated (to the negative half-line $\mathbb{R}_-$) Normal distribution with mean and variance of the conditional $\mathsf{GP}(\mu, \Sigma)$ distribution, given $\{\zeta^{(t+1)}(\mathbf{x}_j) : \ j \neq j^*\}$. Specifically, sample $\zeta^{(t+1)}(\mathbf{x}_{j^*}) \sim \mathsf{TN}\Big(\mathbf{m}_{j^*|J_c}, \mathbf{V}_{j^*j^*|J_c}\Big)$, where

$$\begin{aligned} \mathbf{m}_{j^*|J_c} &= \hat{\mu}(\mathbf{x}_{j^*}) + \mathbf{r}_{J_c}(\mathbf{x}_{j^*})^\mathsf{T} \hat{\mathbf{R}}_{J_c}^{-1} \Big(\zeta^{(t+1)}(\mathbf{x}_{J_c}) - \hat{\mu}(\mathbf{x}_{J_c})\Big) \\ \mathbf{V}_{j^*j^*|J_c} &= \hat{\sigma}^2 \Big(1 - \mathbf{r}_{J_c}(\mathbf{x}_{j^*})^\mathsf{T} \hat{\mathbf{R}}_{J_c}^{-1} \mathbf{r}_{J_c}(\mathbf{x}_{j^*})\Big). \end{aligned} \tag{3.2}$$

   where $(\hat{\mathbf{R}}_{J_C})_{j,j'} = c(\mathbf{x}_j, \mathbf{x}_{j'})$ for $j, j' \in J_c$ and the jth component of the vector $(\mathbf{r}_{J_c}(\mathbf{x}_{j^*}))_j = c(\mathbf{x}_{j^*}, \mathbf{x}_j)$ for all $j \in J_C$.

2) *Repeat step 1 several times (a reasonable choice would be as many times as there are design points leading to zero output); think of this as one pass of the substitution sampler. Note, as the index sampling is random in step 1, some zero-output design points may get sampled repeatedly while others may not be sampled on a given pass.*

   Repeat step 1 n_- times and increment $t \leftarrow t + 1$.

3) *Repeat passes (steps 1 and 2) to develop a sequence of of negative imputed samples for* $\mathbf{x}_{J_-}$*. Note the responses corresponding tot* $\mathbf{x}_{J_+}$ *will not change. Repeat until a user-defined stopping criterion is reached.*

   Repeat steps 1, 2 until a convergence criterion is met. Return $\zeta(\mathbf{x}_J) := \zeta^{(t)}(\mathbf{x}_J)$.

This generates a sequence of K iid replicates $\zeta(\mathbf{x}_J)$ with approximately the correct $\mathsf{GP}(\mu, \Sigma)$ conditional distribution, consistent with the observed values of y^M_J. Now, for each of these replicates $\zeta(\mathbf{x}_J)$, draw $\zeta(\mathbf{x}_I)$ from the conditional $\mathsf{GP}(\mu, \Sigma)$ Gaussian distribution, given $\zeta(\mathbf{x}_J)$, and set $Z(\mathbf{x}_I) := \big(0 \vee \zeta(\mathbf{x}_I)\big)$. If the object of interest is the posterior mean or variance of $Z(\mathbf{x}_i)$ for some $i \in I$, those are available in closed form for each particular imputation of $\zeta(\mathbf{x}_{J_-})$.

We fit the zGP to our illustrative example by drawing $N = 100$ sets of correlated imputed negative samples (for $\{\mathbf{x}_j : j \in J_-\}$) with a zero mean trend, $\mu(\cdot) = 0$. For each zero-output design point, we took the mean value of those 100 samples, let us call these $\{y_j^- = \frac{1}{K}\sum_{k=1}^{K} \zeta^{(k)}(\mathbf{x}_j) : j \in J_-\}$. Further we will let $\mathbf{y}^{\mathrm{Imp}} = \{y_j^- : j \in J_-\} \cup \{y_j^M : j \in J_+\}$. Now, we fit a GP with a linear mean trend to $\mathcal{D}^{\mathrm{Imp}} = \{(\mathbf{x}_j, y_j^{\mathrm{Imp}}) : j \in J\}$. This design, along with the resulting mean surface of the GP and zGP are plotted in Fig. 3.

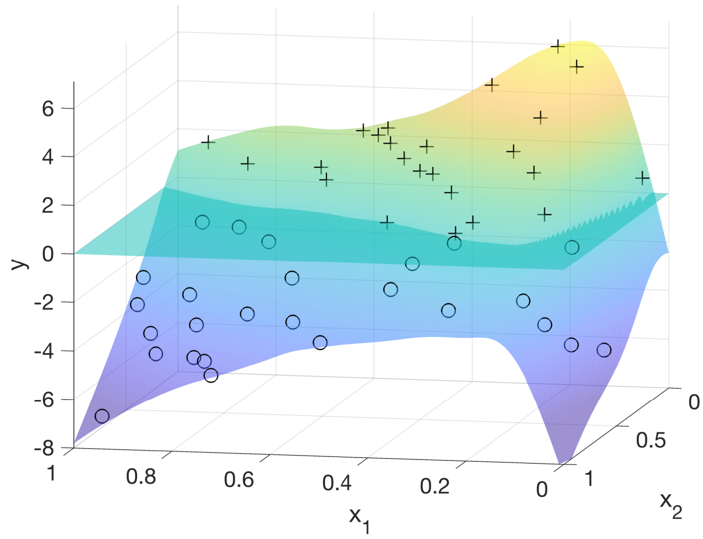

Figure 3: Mean surface of the GP fit to $\mathcal{D}^{\mathrm{Imp}}$ along with the maximum of that surface and zero, *i.e.*, the mean zGP. Design points from $\mathcal{D}^{\mathrm{Imp}}$ are also plotted with $(+)$ corresponding to positive responses, and $(\circ)$ corresponding to negative, imputed responses.

To further illustrate the zGP approach, and its effectiveness at modeling, we sampled the zGP (over the whole computational grid, *i.e.*, for each pixel in input/scenario space). We counted the fraction of times that the true function was zero, but the zGP provided a positive prediction. Likewise, we counted the fraction that the true function was positive, but the zGP predicted a zero. The resulting predicted false positives and false zeros yield a band of uncertainty around the true zero-contour of $f(\cdot,\cdot)$ as can be seen in Fig. 4(a). We also repeated this illustration for smaller designs, with $n = 50, 30$, and 20, also presented in Fig. 4. With a large number of design points, the "transition contour" from zero-predicted output to positive predicted output is very well resolved as indicated by a narrow band of predicted false zeros/false positives in Fig 4(a). The wider bands in Fig 4(b)–4(d) reflect additional uncertainty with fewer design points.

Ultimately, to fully reflect uncertainty using the zGP, one would sample the imputed replicate points $\zeta^{(k)}(\mathbf{x}_J) = \{\zeta^{(k)}(\mathbf{x}_j) : j \in J, k = 1, \ldots, K\}$ and then sample the GP conditioned on equaling $\zeta^{(k)}(\mathbf{x}_J)$. In practice, this may be computationally excessive. With this in mind, we explore the uncertainty in the zGP with the imputed mean, $\mathbf{y}^{\mathrm{Imp}}$, by sampling that zGP. In contrast, we calculate the conditional mean of a zGP fit to each sample set of imputed

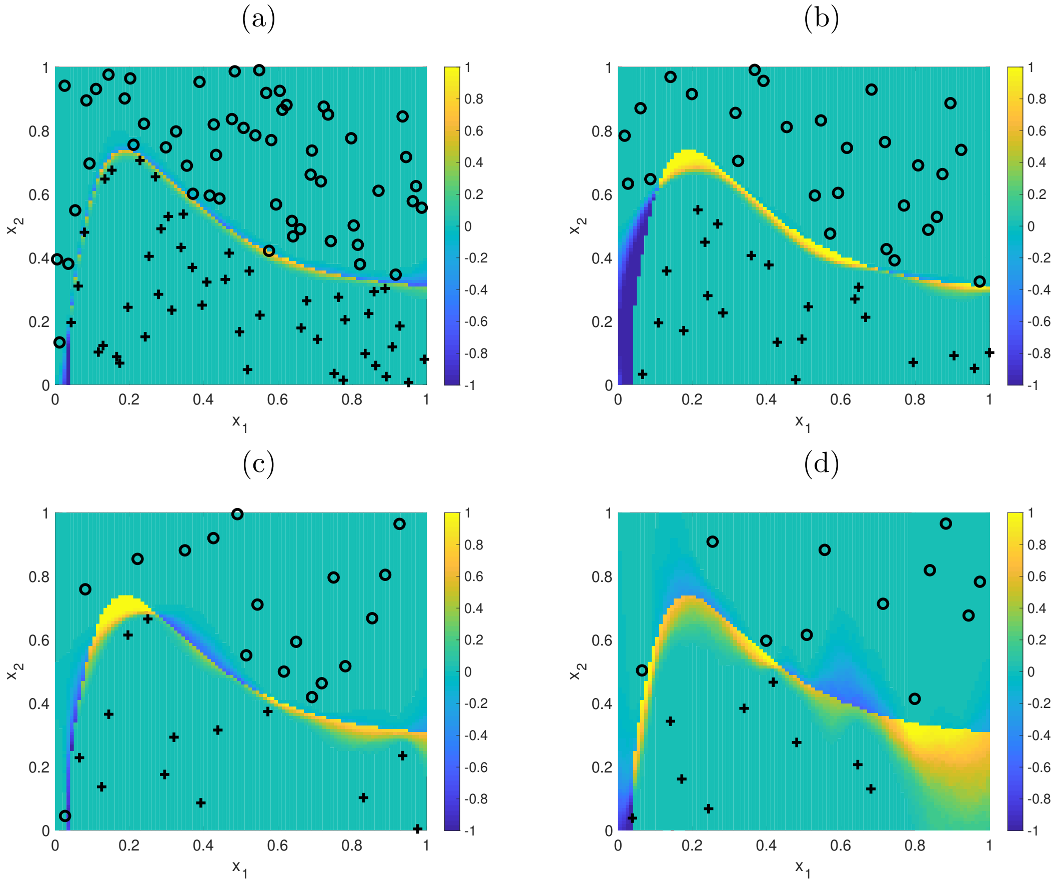

Figure 4: Fraction of samples that indicated false zero response (reported by positive values – toward yellow – on the color scale) and false positive responses (reported by negative values – toward blue – on the color scale). Number of design points, n, from Panels (a)–(d): 100, 50, 30, 20. The symbols $\circ$ and $+$ indicate design points that resulted in a zero response or a positive response, respectively.

points, $\zeta^{(k)}(\mathbf{x}_J)$ (but we do not then sample those GPs, we only evaluate the means.) We compare these two approximations to reflecting zGP uncertainty on the illustrative example in Fig. 5.

3.3. Notes on fitting the zGP: initialization and range parameters.

3.3.1. Initialization. We will explore a general approach to initializing a set of negative imputed outputs for $\{\mathbf{x}_j : j \in J_-\}$. This strategy is one way to obtain an initialization for substitution sampling (step 0 in Algorithm 1). In summary, start with the set of positive output response and corresponding design points, those indexed by J_+. We then sample a GP fit to only these points, and evaluate that sample at all designs point indexed by J_-. If all of these samples are negative, we are done (typically, unless the input space is one dimensional,

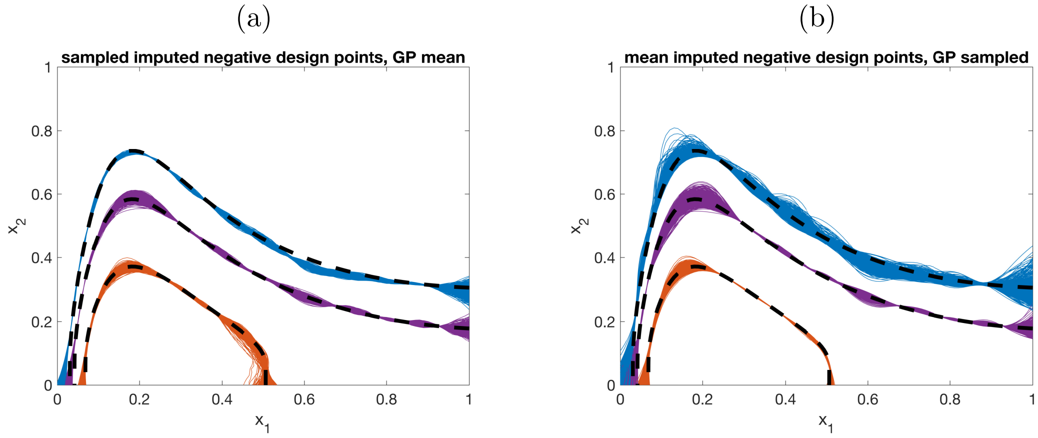

Figure 5: Panel (a): mean of imputed samples, $\mathbf{y}^{\text{Imp}}$ is used to fit a GP with true contours at $y = 0$, $y = 1$, and $y = 3$. This GP is sampled 500 times with the same contours calculated for each sample (0 is blue, 1 is purple, 3 is orange.) Panel (b): a GP fit to each of $K = 500$ sampled sets of imputed design design points, $\zeta^{(k)}(\mathbf{x}_J)$. The mean surface of each GP is calculated and contours plotted for each with the same color scheme indicated level.

this does not happen.) At this point, we collect this round of negatively sampled outputs for $\{\mathbf{x}_j : j \in J_-\}$ along with the positive outputs, fit a GP conditioned to go through all of these points, and then sample the GP at the remaining $\{\mathbf{x}_j : j \in J_-\}$. We repeat this cycle until we have negative samples for all $\mathbf{x}_{J_-}$. Details of this approach follow.

Algorithm 2: Initializing negative imputed samples.

0) Start with a sample $\{\zeta_j^0 : j \in J_-\} \sim \mathsf{No}\big(\mathbf{m}_{J_-|J_+}, \mathbf{V}_{J_-J_-|J_+}\big)$. Here we assume that $\mu(\cdot) = 0$ and the definitions of $\mathbf{m}$ and $\mathbf{V}$ leading to a simpler form of Eqn 3.2:

$$\begin{aligned} \mathbf{m}_{J_-|J_+} &= \hat{\mathbf{R}}_{J_-J_+}\hat{\mathbf{R}}^{-1}_{J_+J_+} y^M(\mathbf{x}_{J_+}) \\ \mathbf{V}_{J_-J_-|J_+} &= \hat{\sigma}^2\Big(\hat{\mathbf{R}}_{J_-J_-} - \hat{\mathbf{R}}_{J_-J_+}\hat{\mathbf{R}}^{-1}_{J_+J_+}\hat{\mathbf{R}}_{J_+J_-}\Big). \end{aligned} \tag{3.3}$$

If all $\zeta_j^0 \le 0$, we are done. Otherwise set $t = 1$.

1) Set $J_*^t = \{j \in J_- : \zeta_j^{t-1} > 0\}$ and set $J_*^{c,t} = J \setminus J_*^t$ (not just $J_- \setminus J_*^t$).
2) If $J_*^t = \emptyset$, set $\{\zeta_j : j \in J\} = \{\zeta_j^{t-1} : j \in J_-\} \cup \{Z_j : j \in J_+\}$ and exit the loop.
3) Draw $\{\zeta_j^t : j \in J_*^t\} \sim \mathsf{No}\big(\mathbf{m}_{J_*^t|J_*^{c,t}}, \mathbf{V}_{J_*^tJ_*^t|J_*^{c,t}}\big)$.
4) Increment $t \leftarrow t + 1$ and repeat steps 1)–4).

Note, $\mathbf{m}$ and $\mathbf{V}$ are updated in step 3) as in Eqn 3.3. Either one can utilize one sample of $\{\zeta_j : j \in J\}$ or repeat this process K times and take the sample average for each $j \in J$ to initialize substitution sampling for Algorithm 1. Our illustrative example and applications proceed with the latter.

3.3.2. Fitting trend and correlation parameters.. With a negative sample for all $j \in J_-$ in hand, before implementing the zGP substitution sampling of Algorithm 1, we select and fit a mean trend for the zGP using these initial imputed points. Often a constant or a linear trend for $\mu(\cdot)$ is appropriate, but a particular application may benefit from a problem-specific mean function as we will see in Section 4.

Until this point we have relied on fitting the GP (*i.e.*, finding reasonable range parameters) using only the design points $\mathcal{D}_+^M = \{(\mathbf{x}_j, y_j^M) : j \in J_+\}$ with strictly positive output $y_j^M > 0$. Surely we lose some information on the range parameters by ignoring the influence of *all* the design points that result in zero outputs. As such we propose to include a subset of the design points that result in zero output for the purpose of fitting range parameters. We focus our search for a prudent selection of these zeros by considering two factors: 1) the minimum distance between each zero-output design point $\mathbf{x}_j$ and the set of positive-output design points, and 2) the probability of obtaining a negative sample at each zero-design point from a GP fit to $\mathcal{D}_+^M$. We posit that the most influential zeros are those that are *both* close to positive-output design points *and* have a small probability of being negative under the original fit to $\mathcal{D}_+^M$. A specific choice of the number of zeros to include and/or thresholds for each metric will be user defined. For the pedagogical example, we sorted the zero-output design points under each metric, considered the smallest $\frac{1}{2}n_-$ design points of each ordered set (*i.e.*, those design points resulting in zero output that are *both* nearest to a design point resulting in positive output *and* those that have the smallest probability of being negative under a GP model fit only to $\mathcal{D}_+^M$.) Then we selected the zero-output design points in the intersection of these two sets. This set of additional design points along with design in J_+ will be indexed by J_+^*. The resulting subset is displayed in Fig. 6(a) along with all of the design/response pairs, including the negative imputed response values, $\mathbf{y}^{\text{Imp}}$, in Fig. 6(b). We then compare three mode-posterior estimates of the range parameters: one set of estimates fit to only positive outputs, $\{(\mathbf{x}_j, \zeta_j^{(0)}) : j \in J_+\}$; a histogram of mode-posterior range parameter estimates fit to positive outputs and "closest" imputed outputs, $\{(\mathbf{x}_j, \zeta_j^{(t)}) : j \in J_+^*\}$; and one set of estimates fit to positive outputs and "closest" imputed outputs, $\{(\mathbf{x}_j, y_j^{\text{Imp}}) : j \in J_+^*\}$. In the pedagogical example (with $n = 50$) it is worth noting that dominant input variable (*i.e.*, the one with the smallest estimated correlation length) swaps roles when fit to design points indexed by J_+ versus those indexed by J_+^*. In particular, for θ_2, the mode estimate found by fitting a GP fto $\mathcal{D}_+^M$ does not even fall in the support of histogram for θ_2 when influential zero-output (and then negative imputed-output) design points are included in the GP model. This indicates that a GP fit to only positive-output designs points may not be an optimal model for the zGP.

4. Applications.

We apply the zGP to two geophysical flow applications, namely computer models of storm surge from tropical storms and of volcanic flows known as pyroclastic density currents. In each case, the inundation footprint is spatially complex and the set of map nodes (spatial pixels on a map) that result in no-inundation (*i.e.*, zero outputs) varies when the computer models are run at different (storm or volcanic) scenarios. We first apply the zGP to storm surge simulations and compare the resulting zGP model to using a convention GP that does not account for the semi-binary nature of the computer model output. We then do a more in-depth application of the zGP to pyroclastic density current simulations to demosntrate how the zGP could be used in a probabilistic analysis of hazards.

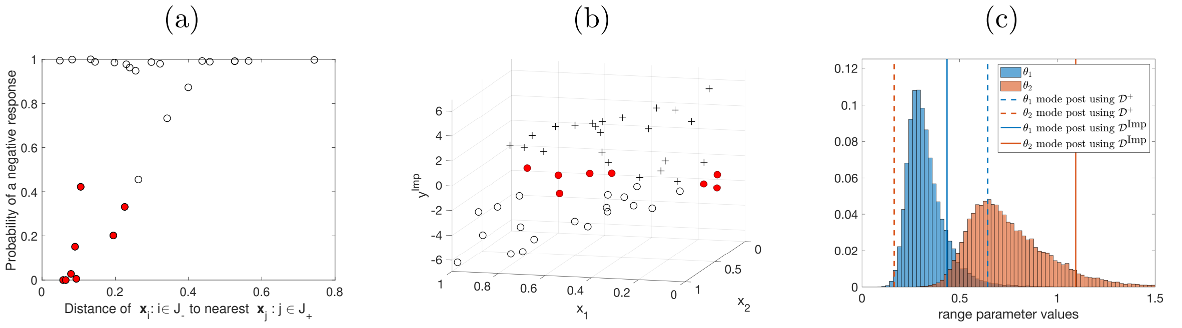

Figure 6: Pedagogical example. Panel (a): For each zero-output design point, the probability of a negative response at that input predicted by a GP model fit to only positive-output design points is plotted against the Euclidean distance (in input space) to its nearest positive-output design point. Red interiors indicate the design points that were chosen to be included in the set to fit range parameters for the zGP. Panel (b): Positive response (+) and negative imputed response (∘) plotted against the corresponding design points. Again, red filled points correspond to the additional points considered to fit range parameters for the zGP. Panel (c): Mode posterior estimates of range parameters (θ_1 in blue, θ_2 in ochre). The histograms of range parameter values are those computed during the replacement sampling imputation algorithm and fit to $\{(\mathbf{x}_j, \zeta_j^{(t)}) : j \in J_+^*\}$. Dashed lines are fit only to $\mathcal{D}_+^M$. The solid lines are range parameter values fit to $\{(\mathbf{x}_j, y_j^{\mathrm{Imp}}) : j \in J_+^*\}$.

4.1. zGP for computational models of storm surge. Several threats are associated with hurricanes and tropical cyclones. In addition to persistent high winds and torrential rainfall, storm surge — flooding due to, effectively, a hurricane pushing ocean water onto land — is often responsible for severe property damage and loss of life associated with hurricanes. In fact, roughly half of the deaths in North America from Atlantic hurricanes in the late 20$^{\text{th}}$ century/early 21$^{\text{st}}$ century are attributed to storm surge (Rappaport, 2014).

Storm surge simulators are numerically implemented models of ocean circulation that commonly solve barotropic, depth-averaged shallow water equations over realistic bathymetry. Such models are forced by atmospheric conditions, notably wind and atmospheric pressure, as well as bottom drag. ADCIRC is the storm surge simulator we explore in this example (Luettich and Westerink, 2004; Westerink et al., 2008). It employs Galerkin methods in combination with finite elements over an unstructured mesh that is amenable to dealing with geometrically complicated domains like coastlines.

The skill of storm surge simulators has increased markedly over the last few decades (Resio and Irish, 2015), leaving the aleatory variability of storms as the major sources of uncertainty – how big, how strong, landfalling location etc. Several recent studies apply GP-based surrogate methods to output from storm surge simulations that vary storm parameterizations as inputs (Jia and Taflanidis, 2013; Jia et al., 2016; Zhang et al., 2018; Yang et al., 2019; Taflanidis et al., 2020; Plumlee et al., 2021). Some studies ignore the zero-problem by focusing on "all wet" map nodes while others use an ad-hoc spatial interpolation for imputing replacement values for zeros. Here we apply a principled, model-based approach to imputation that can be used in conjunction with ad-hoc approaches, or to replace imputation for problematic map nodes, or when detailed spatial information is not available. In this study we focus on storms that threaten southwest Florida, USA. We consider a latin hypercube design of 200 storms. These are parameterized at landfall by: latitude of the storm's center, a storm's central pressure deficit (dp – indicates a storm's intensity), radius of maximum wind speed (r_{mw} – indicates a storm's size), storm forward speed (v_f), storm heading (θ – angle of incidence, measured in degrees clockwise from 0 at due North), and Holland's B (a shape parameter to the radial wind and pressure fields).

The design for this study along with a grid of 908 map nodes where simulated max storm surge output is recorded are shown in Fig. 7. In this simulated storm surge data set, 559 of

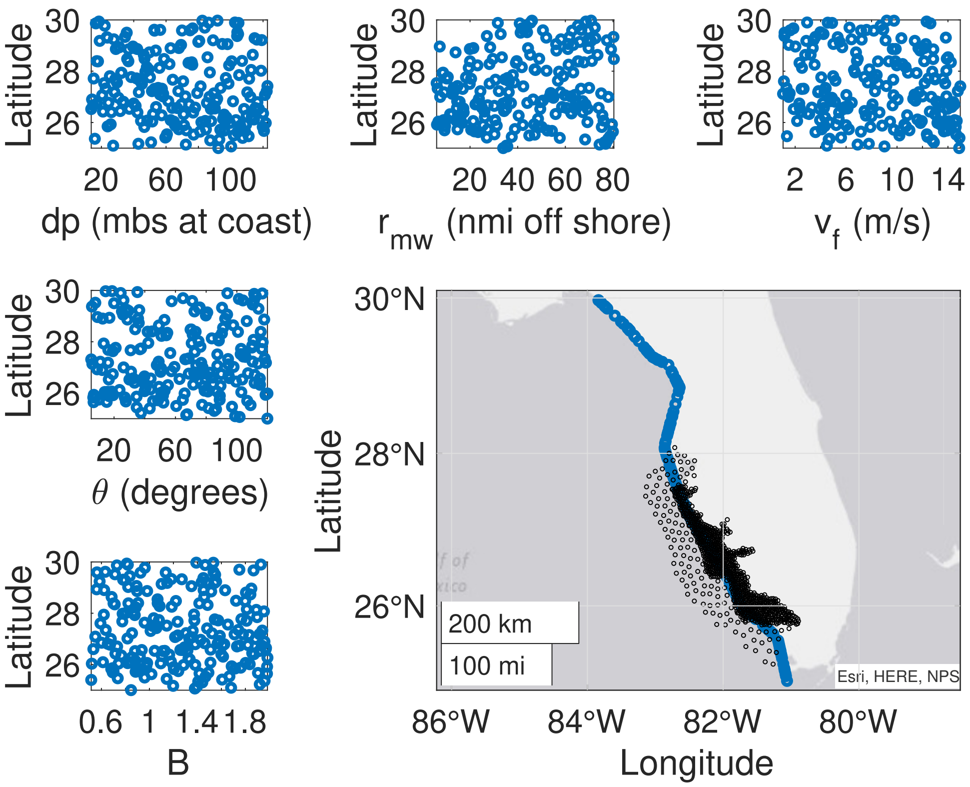

Figure 7: Storm surge simulator design. The lower right plot shows the landfall location of 200 simulated storms (blue circles) along with an unstructured grid of map nodes under consideration for storm surge inundation. Each of the scatter plots is Latitude of the storm's center at landfall vs one of the other storm parameters at landfall, clockwise from lower left: Holland's B, angle of incidence, central pressure deficit (millibars), radius of maximum wind speed (nautical miles), and forward speed.

the 908 map nodes have some "dry" storms (zeros recorded as output at that node) ranging from one dry storm to 193 dry storms of the 200 simulated storms. We fit the zGP to the storm surge output for each of 559 nodes and impute negative values to replace the zero-valued outputs. Then we apply PCA to the full data set of storm surge inundation and negative imputed storm surges to perform dimension reduction on over the 908 spatial modes. Keeping 10 PCA modes, we then fit GPs to each of the 10 associated PCA loadings as output with the input design described in Fig. 7. Then we construct predicted surges by computing loadings given by the GP predictive mean evaluated at the left-out storm parameter inputs. Finally, we take the predicted surge at each node to be the maximum of that given by the GP+PCA reconstruction and zero.

To demonstrate the efficacy of the zGP in this case, we perform leave-out experiments and predict storm surge inundation depths for cases not used to fit the emulator. In Fig. 8 we leave out four representative storms, and use the zGP emulator as just described to estimate the output of the four left-out ADCIRC storm surge simulations. We also show the signed differences which, for the storms under consideration, range ± 1 meter. We also consider a full leave-one-out experiment and calculate predicted errors for each storm at each node ($200 \times 908 = 161,800$ errors.) For comparison, we build two PCA-based emulators – one on the original data set including all of the zeros, and one on the zGP imputed negatives-for-zeros data set. Fig. 9 shows normalized histograms of error magnitudes for each of these two cases. The zGP-imputed error histogram has more mass for small errors (say, ≤ 0.2m) which one might anticipate as the imputation adds information for storms that are "near misses" vs "far off." We also found that the zGP has many fewer large errors (say ≥ 2m) which is a somewhat surprising result.

4.2. zGP for volcanic hazard analysis. Pyroclastic density currents (PDCs) are hot, fast-moving flows made of gas and volcanic particles of very different sizes (Sulpizio et al., 2014). Their destructive potential is extremely high and they have caused the greatest number of fatalities related to volcanic activity over the last centuries (Brown et al., 2017). PDC generation mechanisms and initial conditions, including the spatial location of the eruptive vent, are quite complex and can vary significantly from one eruption to another, or even within a single eruptive episode. Additionally, understanding and hence forecasting the spatio-temporal propagation of PDCs, which is largely influenced by the topography at a given volcanic system, stands as an arduous challenge in modern volcanology (*e.g.*, Dufek (2016)). PDC initiation can either be modeled as one or more piles of material that collapse under their own weight, or one or more fluxes of material that collapse back to the ground after losing their vertical momentum (*e.g.*, Charbonnier and Gertisser (2012); Esposti Ongaro et al. (2007); Valentine and Sweeney (2018)). The flows then propagate under the action of gravity and lose momentum due to frictional forces acting both within the flow and at the interface between the flow and the basal surface (Pitman et al. (2003); Patra et al. (2005); see also https://vhub.org/resources/4057/download/Titan2D_User_Guide.pdf).

In order to quantify aleatory and epistemic uncertainties related to PDC generation and propagation, and therefore fully quantify a PDC hazard, several modeling strategies have been recently adopted (Dalbey et al., 2008; Neri et al., 2015; Sandri et al., 2018; Tierz et al., 2018). One such strategy is to build GP emulators of the computer model outputs from the widely

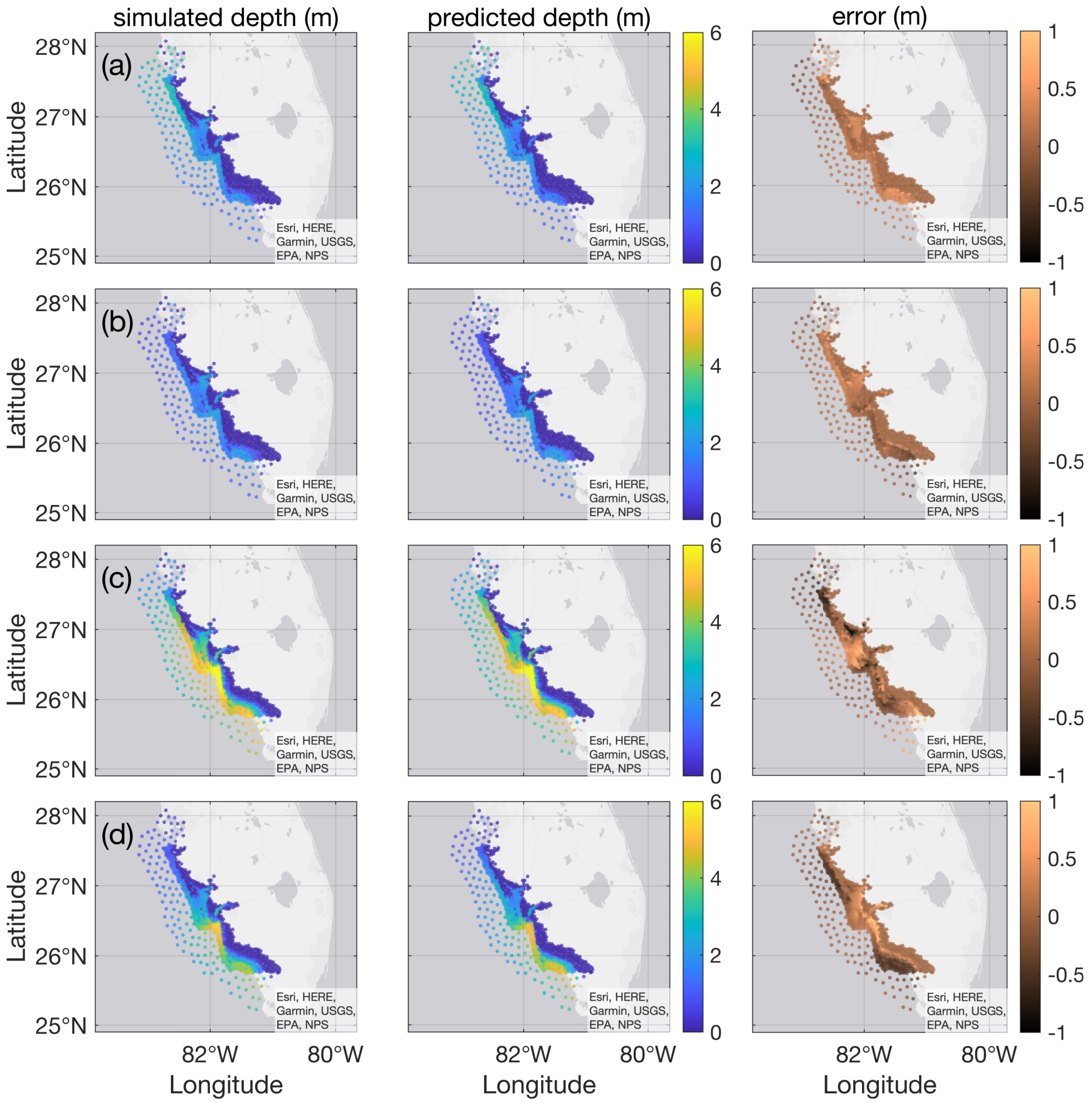

Figure 8: Left column: storm surge depths from four simulated storms labeled (a)–(d) (note these are the same simulated storms as in Fig. 1.) For visualization purposes, the surge depth color scale is set from 0 to 6m although a few nodes exceed surge depths of 6m for storms labeled c and d. Middle column: estimated storm surge depth utilizing emulators with zGP imputation for the parameterized storms (a)–(d). Right column: signed error in storm surge estimation defined as the difference between simulation depth and estimated depth at each node. Note here that the color scale varies from -1m to 1m.

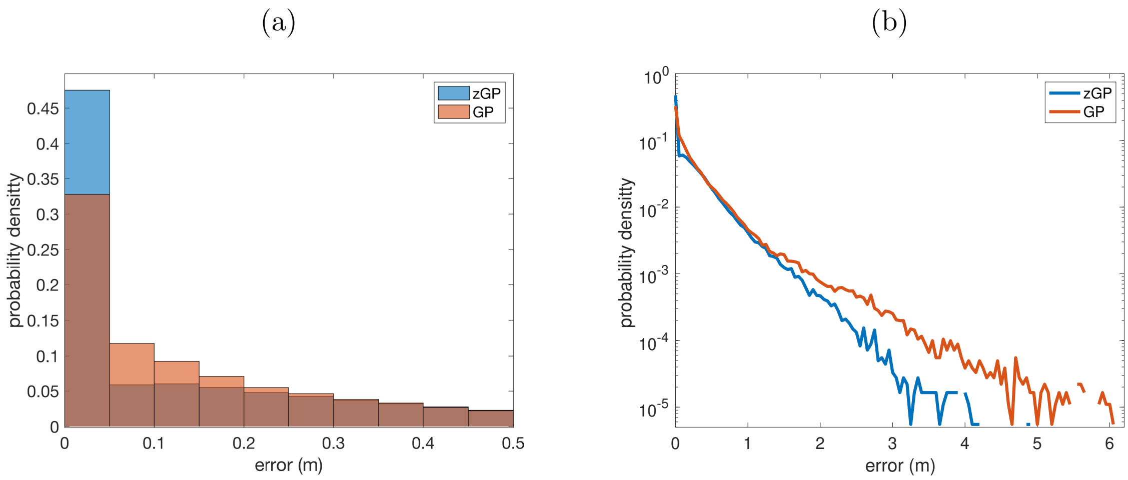

Figure 9: Normalized histograms of the magnitude error between simulated and emulated storm surge depths. Blue corresponds to emulators fit with zGP imputed values for zeros while ochre corresponds to emulators fit to output including zeros. Panel (a): Truncated histogram to compare the mass of the two cases for small amplitude errors. Panel (b): Histograms heights plotted on a logarithmic scale against error in order to visualize the relative frequency of large predicted storm surge errors for the two emulators.

used and freely available software TITAN2D (Patra et al., 2005). TITAN2D offers numerical approximations to a hyperbolic system of PDEs, solved over a digital elevation model (DEM), for modeling dry granular flows as "shallow-water" along with constitutive friction terms to account for the granular nature of the flowing mass. The TITAN2D-GP strategy to quantify PDC hazards has been successfully implemented at a few volcanic systems (Bayarri et al., 2015; Rutarindwa et al., 2019; Spiller et al., 2020), but with the zero-censoring handled in an ad-hoc manner. In this manuscript, we illustrate how the zGP emulator can be used in conjunction with TITAN2D, and applied to probabilistic volcanic hazard assessment of PDCs.

We choose Aluto volcano, in central Ethiopia, as an illustrative volcanological example of hazard analysis utilizing the zGP emulator for three reasons: (1) like other volcanic systems worldwide (Connor and Hill, 1995; Selva et al., 2012; Bebbington, 2012), Aluto has shown significant spatial variability in the location of its eruptive vents (Hutchison et al., 2014; Clarke et al., 2020); (2) evidence from geological fieldwork from the most recent eruptive period at Aluto suggests that new PDCs may be relatively small in volume (Clarke, 2020); and (3) the topography at Aluto volcano (Fig. 1-left) is more complicated than many other volcanoes (Branney and Acocella, 2015; Davidson and de Silva, 2000; Grosse et al., 2009; Clarke et al., 2020). The combination of factors (2) and (3) above implies that many of the (real and simulated) PDC events at Aluto are expected to result in complex, but relatively small inundation footprints across the hazard domain. In other words, many points of interest will not be inundated by typical PDCs and hence TITAN2D output there will present GP emulation with the "zero problem". Hence, Aluto volcano represents an interesting volcanological example

TITAN2D parameter	minimum value	maximum value
x_1: Flux-source (vent) radius, r [m]	1.0	148.3
x_2: Flux rate , h [m/s]	20.0	148.4
x_3: Bed friction angle [deg]	6.1	26.8
x_4: Vent location, UTM Easting [m]	475260	480930
x_5: Vent location, UTM Northing [m]	855190	862860
(fixed parameters)	value	
Internal friction angle [deg]	30.0	
Flux-source duration, (d) [s]	240	
Stopping time [s]	400	
(calculated quantity: $v_{\text{PDC}} = \pi x_1^2 x_2 d/4$)	minimum value	maximum value
PDC volume [M m^3]	0.053	500

Table 1: TITAN2D parameter values under consideration in this illustrative study of PDC hazard analysis at Aluto volcano (Ethiopia.)

for the use of zGP emulators for probabilistic hazard quantification.

We are aiming to model column-collapse PDCs (Sulpizio et al., 2014) with TITAN2D, so we adopt a different and more realistic approach to scenario modeling (*e.g.*, the choice the input/scenario space for our simulation design that more closely mimics the physical initiation processes) than taken in previous approaches (Tierz et al., 2018; Rutarindwa et al., 2019). In total, we explore five uncertain TITAN2D inputs: vent radius, influx rate, bed friction angle, and Easting and Northing Universal Transverse Mercator (UTM) coordinates of the vent location. In terms of vent locations, vents could open over a large area (about 300 km^2) across the volcanic edifice of Aluto and its surroundings. Here, we illustrate our results by focusing on two nearby map points located on the SE area of the volcano (Fig. 1-left). The area covered by the TITAN2D simulations that are relevant to potential inundation at those map points is approximately 30 km^2. That is, given the parameter ranges we are considering, no PDCs are able to inundate the locations of interest if they initiate from a vent location outside this 30 km^2 zone. For each map location, we use a subdesign of 250 simulations, which is a subsample of a Latin hypercube design that covers the entire hazard domain. The subdesign points are chosen to include all runs that lead to inundation at the location of interest along with the simulations resulting in zero output that are nearest in design space to scenarios leading to inundation (as in Rutarindwa et al. (2019).) The subdesign along with indication of resulting inundation (or not) at one or both locations of interest is shown below in Fig. 12 and ranges of input design values are given in Table 1.

To demonstrate the efficacy of the zGP for analyzing inundation hazards of PDCs at Aluto, we compare the predictive mean of the zGP to that of a GP fit only to design points resulting in positive flows, and to a GP that expands on that set to include selected zero-output design points as in Spiller et al. (2014). It is clear that the zGP can readily define the boundary between inundation and no inundation while the GPs that ignore most or all of the zero-outputs struggle to do so. Figure 11-a is particularly revealing of the benefits of the

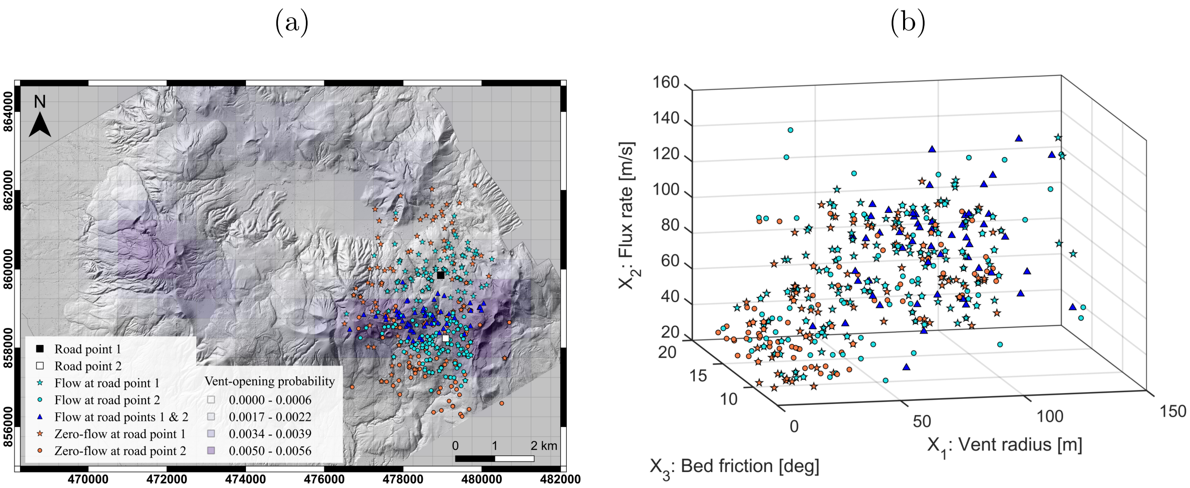

Figure 10: Summary of TITAN2D input subdesign points and corresponding outputs used to build zGP emulators for quantifying hazard probabilities at two locations of interest (road points) at Aluto volcano, Ethiopia (see simplified geographical context in the top-left corner of Fig. 1-left.) Panel (a): Spatial vent locations subdesign points plotted on a base map that is a 2-meter-resolution LiDAR Digital Elevation Model (DEM) (Hutchison et al., 2014). For reference, the vent opening probability density function from (Clarke et al., 2020) is shaded in purple with darker shades representing higher probability. Likewise, the two map points of interest (road points) are plotted along with all of the subdesign vent locations. Note, the symbols to mark these points also reflect if the resulting TITAN2D simulation inundated one or both point road points, and whether it is included as a zero in the design data set for that road point. Panel (b): a 3-D scatter plot of the other design variables (vent radius, flux rate and bed friction) marked with symbols corresponding to the vent location design and legend in Panel (a).

zGP. The zGP transition to zero follows the intuitive boundary of the caldera rim, *i.e.*, flows originating at vents outside of the caldera rim (except those just to the south), will not result inundation at road point 1, and only the zGP captures that behavior. Further, figure 11-a demonstrates a "rebound" of the GP mean predictions back to positive inundation in regions where no flow simulations result in inundation (see top panel in figure 11-a, toward the north side of caldera rim.) As the zGP includes all of those zero-outputs, it does not suffer such issues which would be highly problematic if used in a hazard analysis.

To perform the hazard analysis, we build a zGP emulator $\tilde{y}$ using TITAN2D output at each of the map points of interest (indexed by k) to approximate the maximum PDC flow height $\tilde{y}_k(\mathbf{x}) \approx y_k(\mathbf{x})$ where $\mathbf{x} =$[vent radius, flux rate, bed friction angle, UTM Easting, UTM Northing]. We define the hazard scenario domain $\mathcal{D}$ to be the five dimensional hypercube with vertices in each of the j dimensions varying from $\min(x_j)$ to $\max(x_j)$ with those values given in Table 1. We further define PDC inundation to be a maximum inundation height, y_k, of at

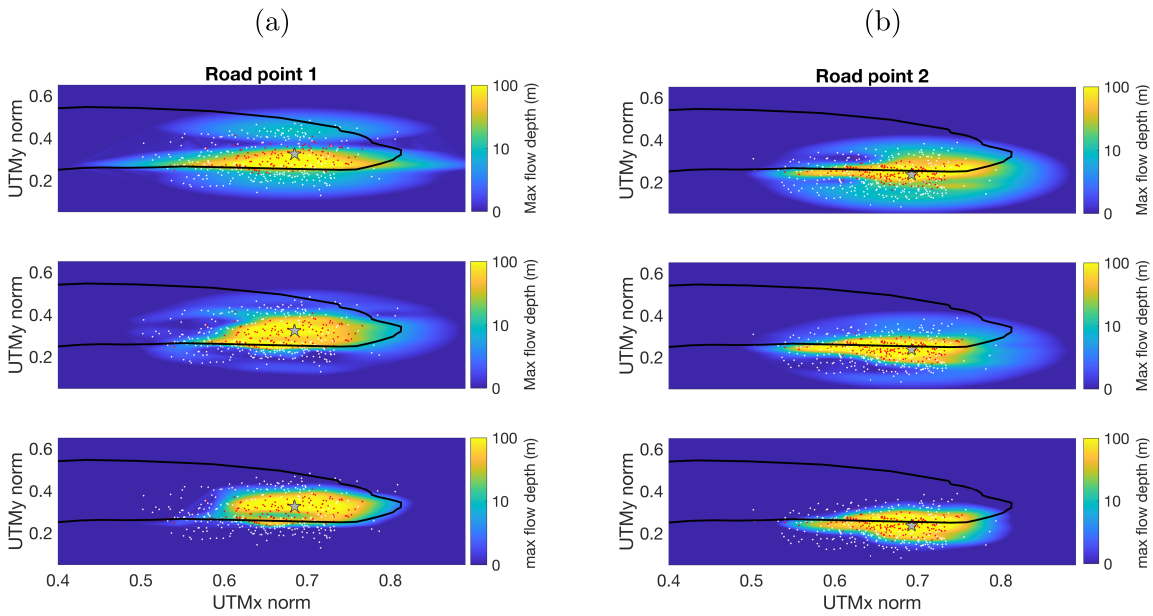

Figure 11: Panel (a): Emulator mean evaluations at road point 1. Panel (b): Emulator mean evaluations at road point 2. For each figure, UTMx and UTMy coordinates of the design points are plotted in red if flows originating at those coordinates led to positive inundation at the respective road point (labeled with a star), and in white if they led to no inundation. A black contour representing the caldera rim is plotted in each figure for reference. Blue-yellow pixels in each figure represents the mean of a GP prediction evaluated at each (UTMx, UTMy) coordinate for a fixed volume and basal friction (with color applied on a log scale in meters.) Top row: mean evaluations of a GP fit only to design points with positive (red) output. Middle row: same as top row with a few additional design points with zero-output. Bottom row: zGP fit to all design points.

least $h_{\text{crit}} = 0.1$m, and define the probability of inundation for location k as

$$
\begin{aligned}
&(4.1) \qquad P_k(\text{inundation} \mid \text{PCD occurs}) = \int_{\mathcal{D}} \mathbf{1}_{\{y_k(\mathbf{x}) \geq h_{\text{crit}}\}} p(\mathbf{x}) d\mathbf{x} \\
&(4.2) \qquad \approx \frac{1}{M} \sum_{i=1}^{M} \mathbf{1}_{\{\tilde{y}_k(\mathbf{X}_i) \geq h_{\text{crit}}\}}, \qquad \mathbf{X}_i \sim p,
\end{aligned}
$$

where $p(\cdot)$ is the probability density function describing the aleatory variability of potential hazard scenarios and $\mathbf{1}_{\{\text{Event}\}}$ is an indicator function that takes on one if the event happens and zero otherwise. In our MC computations, we take $M = 10^5$ replicates. To explore the effects of aleatory uncertainty on vent opening, we compare two vent opening models over a 100 km^2 region encompassing the hazard domain: $p(x_4, x_5)$ as uniform, and $p(x_4, x_5)$ as the vent opening model developed by Clarke et al. (2020). In our exploration we fix the bed friction at 15°, *i.e.*, set $p(x_3) = \delta(x - 15)$. Vent radius and flux are treated differently in each of our two analyses as described below.

To compute the results displayed in Fig. 12-a, we assume the vent radius and flux are distributed uniformly from across their respective domains. For each sample of $p(\mathbf{x})$, we calculate the resulting volume $V_{\text{PDC}} = \pi X_1^2 X_2 d$, and compute the estimated probability of inundation as function of the PDC volume, v_{PDC}. Additionally, we sample both vent opening models as described above over the vent-opening domain shown in Fig. 12-b as a red outlined rectangle. Our assumption is that this domain covers all vent locations that can – in a volcanologically plausible sense – result in PDC inundation at map points of interest. This choice is both consistent with the results presented here (Fig. 1-left) as well as estimates of maximum flow runout from our exploratory study of TITAN2D simulations at Aluto. From this hazard analysis we see that the probability of inundation at both road points assuming the Clarke model of vent opening is roughly double that of assuming a uniform model of vent opening. Interestingly under the uniform model, the probability of PDC inundation for road point two is less than the probability of inundation at road point one, but under the Clarke model the probability of inundation at road point two is greater than at road point one. Use of the zGP in such hazard analysis enables this kind of rapid comparison of uncertain modeling assumptions. In Fig. 12-b, the values of conditional probability of PDC inundation obtained by building zGP emulators on each of a grid of map points over a small hazard domain ($\sim$4 km^2 in area). In this calculation, the volume is fixed at ≈ 0.01 km^3 by taking $p(x_1)p(x_2) = \delta(x_1 - 30)\delta(x_2 - 60)$ (*i.e.*, the emulator is evaluated at $x_1 = 30$, and $x_2 = 60$) and the Clarke vent opening distribution is sampled. The latter analysis serves to illustrate how our approach could be expanded to a full probabilistic volcanic hazard assessment via construction of probabilistic hazard maps (Clarke et al., 2020; Spiller et al., 2014, 2020; Tierz et al., 2018, 2020; Rutarindwa et al., 2019).

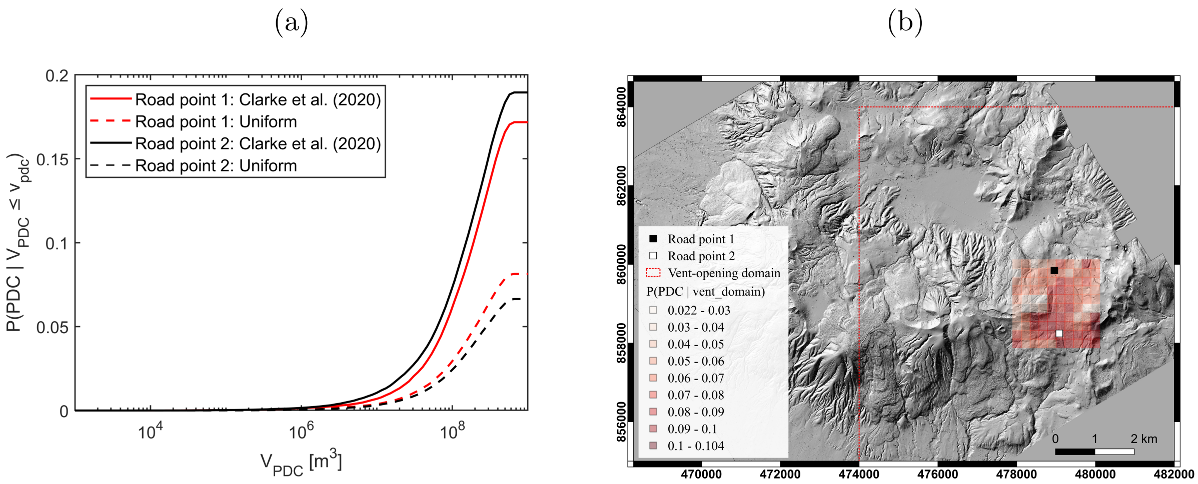

Figure 12: Summary of the illustrative probabilistic hazard analysis utilizing the zGP for example locations at Aluto volcano, Ethiopia. Panel (a): Conditional probability of PDC inundation (given PDC volume) at road points 1 and 2, for different PDC volume thresholds, calculated by Monte Carlo evaluation of the zGP emulators fitted at these points (see text for more details). Two different hazard models in terms of the aleatory variability in vent opening are explored: the model presented in (Clarke et al., 2020) and an equal (*i.e.*, Uniform)-vent-opening-probability model. Panel (b): Conditional probability of PDC inundation (given vent locations within a given spatial domain: red dashed line) over a hazard grid composed of 100 points, covering an area of approximately 4 km^2, calculated by Monte Carlo evaluation of the zGP emulators fitted at these map points (see text for more details). Road points 1 and 2 are shown for reference in Fig. 1 as well.

5. Discussion and conclusions. In this work, we have introduced a zero-censored Gaussian process as a systematic, model-based approach to apply GPs to range-constrained simulator output. This approach relies on imputing replacement computer model runs resulting in zero output (or, attaining the max/min of a range constraint) that intentionally violate the constraint of non-negativity. Then a GP is constructed utilizing the negative imputed data in place of zero-output data, and zGP predictions at untested inputs are taken to be the maximum of the GP and zero. Moreover, the zGP can be applied as a pre-processing step to then be used in conjunction with other GP advances. In Section 4 we applied the zGP before implementing two common approaches to handling large-dimensional output data, namely the parallel-partial emulator and GPs on PCA loadings.

The zGP approach overcomes several challenges associated with range-constrained output. By construction, the GP utilized in the zGP has full support. The imputed data also allows us to avoid the (nearly ubiquitous) non-stationarity that arises in models fit directly to range constrained model output – flat over some regions of input space and varying over others. This non-stationarity offers a particular challenge for vector-valued output (*e.g.*, storm surge and PDC models) as the sets of design points that result in zero outputs change as we consider different components of the vector-valued output (*e.g.*, different map nodes in geophysical flows have different inputs in the design that lead to no flow.) This issue is a formidable challenge for approaches that partition the input space and utilize different kernels on different partitions to handle non-stationarity. Further, the transition of the computer model output from positive values to zero may not be smooth, and most likely will not occur exactly at design points. The zGP can readily estimate these transitions without assumptions on the geometry of the input space. Lastly, there is some computational overhead in fitting a zGP for vector-valued outputs, but those computations are a "distributable" preprocessing step.

We applied the zGP to a pedagogical example, and to two geophysical flow examples. Yet, like many new methodologies, the potential of the zGP lies in ease of implementation and wide applicability. For storm surge hazard analysis, the zGP may prove useful for map nodes (subsets of the vector-valued output) where imputation based on topographic interpolation (Kyprioti et al., 2021) is unsuccessful. It will likely prove quite useful for spatial processes with nearly no topographic influences, or those that do not have "easily modeled" topographic influences. For example, an interesting application of the zGP is a systematic study to understand the influence topography on pyroclastic flows where the topography has complex features (*e.g.*, more in depth studies on volcanoes like Aluto which was examined in Section 4.) Spatially-varying dynamic infectious disease models offer another example where the zGP may prove a powerful tool for validation and uncertainty quantification. Of course, there are a wide array of vector-valued outputs without spatial dependence – lengths, volumes, etc – that must be positive or bounded, and the zGP has the potential to enable GP surrogate modeling for such problems. Additionally, one could imagine using the zGP in conjunction with derivative constrained GP construction as in (Wang and Berger, 2016) to meet monotonicity constraints.

Acknowledgements. We would like to acknowledge support for ETS from NSF-DMS 2053872, 1821338; for RLW from NSF-DMS 1821289. We would like to thank Ben Clarke, Eliza Calder, Bruce Pitman, Sarah Ogburn, Firawalin Dessalegn, Gezahegn Yirgu, Pierre Barbillon, Susan Loughlin and Luigi Passarelli for useful discussions.

References.

P. Abrahamsen and F. E. Benth. Kriging with inequality constraints. *Mathematical Geology*, 33(6):719–744, 2001. doi: 10.1198/TECH.2009.08019.

C. Agrell. Gaussian processes with linear operator inequality constraints. *Journal of Machine Learning Research*, 20(135):1–36, 2019. URL http://jmlr.org/papers/v20/19-065.html.

L. S. Bastos and A. O'Hagan. Diagnostics for Gaussian process emulators. *Technometrics*, 51 (4):425–438, 2009. doi: 10.1198/TECH.2009.08019.

M. J. Bayarri, J. O. Berger, E. S. Calder, K. Dalbey, S. Lunagómez, A. K. Patra, E. B. Pitman, E. Spiller, and R. L. Wolpert. Using statistical and computer models to quantify volcanic hazards. *Technometrics*, 51(4):402–413, 2009. doi: 10.1198/TECH.2009.08018.

M. J. Bayarri, J. O. Berger, E. S. Calder, A. K. Patra, E. B. Pitman, E. T. Spiller, and R. L. Wolpert. A methodology for quantifying volcanic hazards. *International Journal of Uncertainty Quantification*, 5(4):297–325, 2015. doi: 10.1615/Int.J.UncertaintyQuantification.2015011451.

M. S. Bebbington. Models for temporal volcanic hazard. *Statistics in Volcanology*, 1:1–24, 2012. doi: 10.5038/2163-338X.1.1.

J. Beck and S. Guillas. Sequential design with mutual information for computer experiments (MICE): Emulation of a tsunami model. *SIAM/ASA Journal of Uncertainty Quantification*, 4(1):739–766, 2016. doi: 10.1137/140989613.

M. Ben Salem, F. Bachoc, O. Roustant, F. Gamboa, and L. Tomaso. Gaussian process-based dimension reduction for goal-oriented sequential design. *SIAM/ASA Journal of Uncertainty Quantification*, 7(4):1369–1397, 2019. doi: 10.1137/18M1167930.

M. Branney and V. Acocella. Calderas. In H. Sigurdsson, B. Houghton, S. R. McNutt, H. Rymer, and J. Stix, editors, *The Encyclopedia of Volcanoes*, chapter 16, pages 299–315. Elsevier, second edition, 2015. doi: 10.1016/B978-0-12-385938-9.00016-X.

S. K. Brown, S. F. Jenkins, R. S. J. Sparks, H. Odbert, and M. R. Auker. Volcanic fatalities database: analysis of volcanic threat with distance and victim classification. *Journal of Applied Volcanology*, 6(1):1–20, 2017. doi: 10.1186/s13617-017-0067-4.

S. J. Charbonnier and R. Gertisser. Evaluation of geophysical mass flow models using the 2006 block-and-ash flows of Merapi Volcano, Java, Indonesia: Towards a short-term hazard assessment tool. *Journal of Volcanology and Geothermal Research*, 231:87–108, 2012. doi: 10.1016/j.jvolgeores.2012.02.015.

B. A. Clarke. *Post-caldera eruptions and pyroclastic density current hazard in the Main Ethiopian Rift*. PhD thesis, University of Edinburgh, 2020.

B. A. Clarke, P. Tierz, E. Calder, and G. Yirgu. Probabilistic volcanic hazard assessment for pyroclastic density currents from pumice cone eruptions at Aluto volcano, Ethiopia. *Frontiers in Earth Science: Geohazards and Georisks*, 8, 2020. doi: 10.3389/feart.2020.00348.

C. B. Connor and B. E. Hill. Three nonhomogeneous Poisson models for the probability of basaltic volcanism: application to the Yucca Mountain region, Nevada. *Journal of Geophysical Research: Solid Earth*, 100(B6):10107–10125, 1995. doi: 10.1029/95JB01055.

C. Currin, T. Mitchell, M. Morris, and D. Ylvisaker. A Bayesian approach to the design and analysis of computer experiments. Technical report, Oak Ridge National Laboratory, Oak Ridge, TN (USA), 1988. URL https://www.osti.gov/biblio/6734087.

S. Da Veiga and A. Marrel. Gaussian process modeling with inequality constraints. *Annales de la Faculté des sciences de Toulouse: Mathématiques*, Série 6, 21(3):529–555, 2012. doi: 10.5802/afst.1344.

S. Da Veiga and A. Marrel. Gaussian process regression with linear inequality constraints. *Reliability Engineering & System Safety*, 195:1–13, 2020. doi: 10.1016/j.ress.2019.106732.

K. Dalbey, A. . Patra, E. B. Pitman, M. I. Bursik, and M. F. Sheridan. Input uncertainty propagation methods and hazard mapping of geophysical mass flows. *Journal of Geophysical Research: Solid Earth*, 113(B5):1–16, 2008. doi: 10.1029/2006JB004471.

J. Davidson and S. de Silva. Composite volcanoes. In H. Sigurdsson, B. Houghton, S. R. McNutt, H. Rymer, and J. Stix, editors, *The Encyclopedia of Volcanoes*, chapter 43, pages 663–681. Academic Press, San Diego, CA, first edition, 2000. URL https://booksite.elsevier.com/9780126431407/netscape4/pdfs/CH43.pdf.

J. Dufek. The fluid mechanics of pyroclastic density currents. *Annual Review of Fluid Mechanics*, 48:459–485, 2016. doi: 10.1146/annurev-fluid-122414-034252.

T. Esposti Ongaro, C. Cavazzoni, G. Erbacci, A. Neri, and M.-V. Salvetti. A parallel multiphase flow code for the 3D simulation of explosive volcanic eruptions. *Parallel Computing*, 33(7-8):541–560, 2007. doi: 10.1016/j.parco.2007.04.003.

R. B. Gramacy and H. K. H. Lee. Bayesian treed Gaussian process models with an application to computer modeling. *Journal of the American Statistical Association*, 103(483):1119–1130, 2008. doi: 10.1198/016214508000000689.

P. Grosse, B. van Wyk de Vries, I. A. Petrinovic, P. A. Euillades, and G. E. Alvarado. Morphometry and evolution of arc volcanoes. *Geology*, 37(7):651–654, 2009. doi: 10.1130/G25734A.1.

M. Gu and J. O. Berger. Parallel partial Gaussian process emulation for computer models with massive output. *The Annals of Applied Statistics*, 10(3):1317–1347, 2016. doi: 10.1214/16-AOAS934.

D. Higdon, J. Gattiker, B. Williams, and M. Rightley. Computer model calibration using high-dimensional output. *Journal of the American Statistical Association*, 103(482):570–583, 2008. doi: 10.1198/016214507000000888.

W. Hutchison, D. M. Pyle, T. A. Mather, J. Biggs, and G. Yirgu. *2012 Aluto LiDAR data*, 2014.

G. Jia and A. A. Taflanidis. Kriging metamodeling for approximation of high-dimensional wave and surge responses in real-time storm/hurricane risk assessment. *Computer Methods in Applied Mechanics and Engineering*, 261–262:24–38, 2013. doi: 10.1016/j.cma.2013.03.012.

G. Jia, A. A. Taflanidis, N. C. Nadal-Caraballo, J. A. Melby, A. B. Kennedy, and J. M. Smith. Surrogate modeling for peak or time-dependent storm surge prediction over an extended coastal region using an existing database of synthetic storms. *Natural Hazards*, 81(2):909–938, 2016. doi: 10.1007/s11069-015-2111-1.

A. P. Kyprioti, A. A. Taflanidis, M. Plumlee, T. G. Asher, E. T. Spiller, R. A. Luettich, B. Blanton, T. L. Kijewski-Correa, A. Kennedy, and L. Schmied. Improvements in storm surge surrogate modeling for synthetic storm parameterization, node condition classification and implementation to small size databases. *Natural Hazards*, 109(2):1349–1386, 2021. doi: 10.1007/s11069-021-04881-9.

K. Kyzyurova. *On Uncertainty Quantification for Systems of Computer Models*. PhD thesis, Duke University, 2017.

X. Liu and S. Guillas. Dimension reduction for Gaussian process emulation: An application to the influence of bathymetry on tsunami heights. *SIAM/ASA Journal of Uncertainty Quantification*, 5(1):787–812, 2017. doi: 10.1137/16M1090648.

A. F. López-Lopera, F. Bachoc, N. Durrande, and O. Roustant. Finite-dimensional gaussian approximation with linear inequality constraints. *SIAM/ASA Journal of Uncertainty Quantification*, 6(3):1224–1255, 2018. doi: 10.1137/17M1153157.

R. A. Luettich, Jr. and J. J. Westerink. *Formulation and numerical implementation of the 2D/3D ADCIRC finite element model version 44.XX*. ADCIRC.org, 2004. URL https://adcirc.org/wp-content/uploads/sites/2255/2018/11/2004_Luettich.pdf.

H. Maatouk and X. Bay. Gaussian process emulators for computer experiments with inequality constraints. *Mathematical Geosciences*, 49(5):557–582, 2017. doi: 10.1007/s11004-017-9673-2.

A. Neri, A. Bevilacqua, T. E. Ongaro, R. Isaia, W. P. Aspinall, M. Bisson, F. Flandoli,

P. J. Baxter, A. Bertagnini, E. Iannuzzi, S. Orsucci, M. Pistolesi, M. Rosi, and S. Vitale. Quantifying volcanic hazard at Campi Flegrei caldera (Italy) with uncertainty assessment: 2. Pyroclastic density current invasion maps. *Journal of Geophysical Research: Solid Earth*, 120(4):2330–2349, 2015. doi: 10.1002/2014JB011775.

A. K. Patra, A. C. Bauer, C. C. Nichita, E. B. Pitman, M. F. Sheridan, and M. I. Bursik. Parallel adaptive numerical simulation of dry avalanches over natural terrain. *Journal of Volcanology and Geothermal Research*, 139(1–2):1–21, 2005. doi: 10.1016/j.jvolgeores.2004.06.014.

A. Pensoneault, X. Yang, and X. Zhu. Nonnegativity-enforced Gaussian process regression. *Theoretical and Applied Mechanics Letters*, 10(3):182–187, 2020. doi: 10.1016/j.taml.2020.01.036.

E. B. Pitman, C. C. Nichita, A. K. Patra, A. C. Bauer, M. F. Sheridan, and M. I. Bursik. Computing granular avalanches and landslides. *Physics of Fluids*, 15(12):3638–3646, 2003. doi: 10.1063/1.1614253.

M. Plumlee, T. G. Asher, W. Chang, and M. V. Bilskie. High-fidelity hurricane surge forecasting using emulation and sequential experiments. *The Annals of Applied Statistics*, 15(1):460–480, 2021. doi: 10.1214/20-AOAS1398.

C. A. Pope, J. P. Gosling, S. Barber, J. S. Johnson, T. Yamaguchi, G. Feingold, and P. G. Blackwell. Gaussian process modeling of heterogeneity and discontinuities using Voronoi tessellations. *Technometrics*, pages 1–20, 2019. doi: 10.1080/00401706.2019.1692696.

E. N. Rappaport. Fatalities in the United States from Atlantic tropical cyclones: New data and interpretation. *Bulletin of the American Meteorological Society*, 95(3):341–346, 2014. doi: 10.1175/BAMS-D-12-00074.1.

D. T. Resio and J. L. Irish. Tropical cyclone storm surge risk. *Current Climate Change Reports*, 1(2):74–84, 2015. doi: 10.1007/s40641-015-0011-9.

R. Rutarindwa, E. T. Spiller, A. Bevilacqua, M. I. Bursik, and A. K. Patra. Dynamic probabilistic hazard mapping in the Long Valley Volcanic Region, CA: Integrating vent opening maps and statistical surrogates of physical models of pyroclastic density currents. *Journal of Geophysical Research: Solid Earth*, 124(9):9600–9621, 2019. doi: 10.1029/2019JB017352.

J. Sacks, S. B. Schiller, and W. J. Welch. Designs for computer experiments. *Technometrics*, 31(1):41–47, 1989a. doi: 10.1080/00401706.1989.10488474.

J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn. Design and analysis of computer experiments. *Statistical Science*, 4(4):409–423, 1989b. doi: 10.1214/ss/1177012413.

L. Sandri, P. Tierz, A. Costa, and W. Marzocchi. Probabilistic hazard from pyroclastic density currents in the Neapolitan area (Southern Italy). *Journal of Geophysical Research: Solid Earth*, 123(5):3474–3500, 2018. doi: 10.1002/2017JB014890.

T. J. Santner, B. J. Williams, and W. I. Notz. *The Design and Analysis of Computer Experiments*. Springer Series in Statistics. Springer-Verlag, New York, NY, second edition, 2018. ISBN 978-1-4939-8845-7. doi: 10.1007/978-1-4939-8847-1.

J. Selva, W. Marzocchi, P. Papale, and L. Sandri. Operational eruption forecasting at high-risk volcanoes: the case of Campi Flegrei, Naples. *Journal of Applied Volcanology*, 1(5):1–14, 2012. doi: 10.1186/2191-5040-1-5.

E. T. Spiller, M. J. Bayarri, J. O. Berger, E. S. Calder, A. K. Patra, E. B. Pitman, and R. L. Wolpert. Automating emulator construction for geophysical hazard maps. *SIAM/ASA*

Journal of Uncertainty Quantification, 2(1):126–152, 2014. doi: 10.1137/120899285.

E. T. Spiller, R. L. Wolpert, S. E. Ogburn, E. S. Calder, J. O. Berger, A. K. Patra, and E. B. Pitman. Volcanic hazard assessment for an eruption hiatus, or post-eruption unrest context: Modeling continued dome collapse hazards for Soufrière Hills Volcano. *Frontiers in Earth Science: Geohazards and Georisks*, 8(535567):396, 2020. doi: 10.3389/feart.2020.535567.

M. L. Stein. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer Series in Statistics. Springer Verlag, New York, New York, 1999.

R. Sulpizio, P. Dellino, D. M. Doronzo, and D. Sarocchi. Pyroclastic density currents: state of the art and perspectives. *Journal of Volcanology and Geothermal Research*, 283:36–65, 2014. doi: 10.1016/j.jvolgeores.2014.06.014.

L. P. Swiler, M. Gulian, A. L. Frankel, C. Safta, and J. D. Jakeman. A survey of constrained gaussian process regression: Approaches and implementation challenges. *Journal of Machine Learning for Modeling and Computing*, 1(2):119–156, 2020. doi: 10.1615/JMachLearnModelComput.2020035155.

A. Taflanidis, J. Zhang, A. Kyprioti, A. Kennedy, and T. Kijewksi-Correa. Developments in storm surge estimation using surrogate modeling techniques. *Coastal Engineering Proceedings*, (36v):currents.37, 2020. doi: 10.9753/icce.v36v.currents.37.

P. Tierz, E. R. Stefanescu, L. Sandri, R. Sulpizio, G. A. Valentine, W. Marzocchi, and A. K. Patra. Towards quantitative volcanic risk of pyroclastic density currents: Probabilistic hazard curves and maps around Somma-Vesuvius (Italy). *Journal of Geophysical Research: Solid Earth*, 123(8):6299–6317, 2018. doi: 10.1029/2017JB015383.

P. Tierz, B. Clarke, E. S. Calder, F. Dessalegn, E. Lewi, G. Yirgu, K. Fontijn, J. M. Crummy, Y. Bekele, and S. C. Loughlin. Event trees and epistemic uncertainty in long-term volcanic hazard assessment of rift volcanoes: The example of Aluto (Central Ethiopia). *Geochemistry, Geophysics, Geosystems*, 21(10):e2020GC009219, 2020. doi: 10.1029/2020GC009219.

G. A. Valentine and M. R. Sweeney. Compressible flow phenomena at inception of lateral density currents fed by collapsing gas-particle mixtures. *Journal of Geophysical Research: Solid Earth*, 123(2):1286–1302, 2018. doi: 10.1002/2017JB015129.

V. Volodina and D. B. Williamson. Nonstationary Gaussian process emulators with kernel mixtures. *SIAM/ASA Journal of Uncertainty Quantification*, 8(1):1–26, 2020. doi: 10.1137/19M124438X.

X. Wang and J. O. Berger. Estimating shape constrained functions using Gaussian processes. *SIAM/ASA Journal of Uncertainty Quantification*, 4(1):1–25, 2016. doi: 10.1137/140955033.

W. J. Welch, R. J. Buck, J. Sacks, H. P. Wynn, T. J. Mitchell, and M. D. Morris. Screening, predicting, and computer experiments. *Technometrics*, 34(1):15–25, 1992. doi: 10.2307/1269548. URL http://www.jstor.org/stable/1269548.

J. J. Westerink, R. A. Luettich, J. C. Feyen, J. H. Atkinson, C. Dawson, H. J. Roberts, M. D. Powell, J. P. Dunion, E. J. Kubatko, and H. Pourtaheri. A basin-to channel-scale unstructured grid hurricane storm surge model applied to southern Louisiana. *Monthly Weather Review*, 136(3):833–864, 2008. doi: 10.1175/2007MWR1946.1.

K. Yang, V. Paramygin, and Y. P. Sheng. An objective and efficient method for estimating probabilistic coastal inundation hazards. *Natural Hazards*, 99(2):1105–1130, 2019. doi: 10.1007/s11069-019-03807-w.

J. Zhang, A. A. Taflanidis, N. C. Nadal-Caraballo, J. A. Melby, and F. Diop. Advances in surrogate modeling for storm surge prediction: storm selection and addressing characteristics related to climate change. *Natural Hazards*, 94(3):1225–1253, 2018. doi: 10.1007/s11069-018-3470-1.