STOCHASTIC INVERSION WITH MAXIMAL UPDATED DENSITIES FOR STORM SURGE WIND DRAG PARAMETER ESTIMATION

Carlos del-Castillo-Negrete[1], Michael Pilosov[2], Troy Butler[2], Clint Dawson[1], and Tian Yu Yen[3]

[1]University of Texas at Austin
[2]University of Colorado
[3]Sandia National Laboratories

June 6, 2023

Stochastic Inversion with Maximal Updated Densities for Storm Surge Wind Drag Parameter Estimation

Carlos del-Castillo-Negrete[1], Michael Pilosov[2], Troy Butler[2], Clint Dawson[1], and Tian Yu Yen[3]

[1]University of Texas at Austin [2]University of Colorado Denver [3]Sandia National Laboratories

Background: Data-Consistent Inversion

Stochastic Inverse Problem (SIP)

When model parameters possess aleatoric uncertainties, e.g. due to naturally occurring variability in system inputs, then a **Stochastic Inverse Problem** can be formulated.

- $\lambda \in \Lambda \subset \mathbb{R}^n$ - parameter(s) of interest
- $Q(\lambda) : \mathbb{R}^n \to \mathbb{R}^m$ - parameter to observable **Quantity of Interest (QoI) Map**
- $d \in D := Q(\Lambda) \subset \mathbb{R}^m$ - observable(s)
- Probability measure spaces $\mathbb{P}_\Lambda := (\Lambda, \mathcal{B}_\Lambda, \mu_\Lambda)$ and $\mathbb{P}_\mathcal{D} := (\mathcal{D}, \mathcal{B}_\mathcal{D}, \mu_\mathcal{D})$, with associated densities π_Λ and $\pi_\mathcal{D}$

Definition. *Given an observed probability measure $\mathbb{P}^{obs}_\mathcal{D}$, the* ***Stochastic Inverse Problem (SIP)*** *is to determine a pullback probability measure $\mathbb{P}_\Lambda$ which is* ***data-consistent*** *in the sense that*

$$\mathbb{P}_\Lambda(Q^{-1}(E)) = \mathbb{P}^{obs}_\mathcal{D}(E), \forall E \subset \mathcal{D}$$

Density Based Solution and Diagnostic

Solution given by a form of Bayes's rule that incorporates the push-forward of the initial π^{in}_Λ through the QoI map, which we indicate as $\pi^{pred}_\mathcal{D}$

$$\pi^{up}_\Lambda(\lambda) = \pi^{in}_\Lambda(\lambda) r(Q(\lambda)), \quad r(Q(\lambda)) = \frac{\pi^{obs}_\mathcal{D}(Q(\lambda))}{\pi^{pred}_\mathcal{D}(Q(\lambda))}$$

Predictability Assumption: $\exists\, C > 0$ such that $\pi^{obs}_\mathcal{D}(q) \leq C \pi^{pred}_\mathcal{D}(q)$ for a.e. $q \in \mathcal{D}$.
Note that predictability assumption satisfied implies: $\int_\Lambda \pi^{up}_\Lambda(\lambda)\, d\mu_\Lambda = 1$, so

$$1 = \int_\Lambda \pi^{in}_\Lambda(\lambda) r(Q(\lambda))\, d\mu_\Lambda = \int_\Lambda r(Q(\lambda))\, dP^{in}_\mathcal{D} = \mathbb{E}(r(Q(\lambda)))$$

Diagnostic for Verifying Predictability Assumption: Sample mean of $r(Q(\lambda)) \approx 1$ (up to finite sampling errors or errors in approximating $\pi^{pred}_\mathcal{D}$).

Maximal Updated Density (MUD) Points

Parameter Identification Problem

While the SIP framework was formulated to deal with aleatoric uncertainty, we show that the framework can also be used to treat problems regarding epistemic uncertainty.

Definition. *Given, finite amount of data d on a QoI map obtained for a fixed, but unknown, parameter $\lambda^\dagger$, populated with random noise ξ, $d = Q(\lambda^\dagger) + \xi$, the* ***Paremeter Identification Problem (PIP)*** *is to estimate $\lambda^\dagger$.*

The **Maximal Updated Density (MUD)** point can be used to solve the PIP.

$$\lambda^{\text{MUD}} := \arg\max \pi^{up}_\Lambda(\lambda). \tag{1}$$

A Simple Example - MUD vs Bayesian Maximum a Posterior (MAP) Points

- $\Lambda = [-1, 1], Q(\lambda) = \lambda^5 \to \mathcal{D} = [-1, 1], \pi^{in}_\Lambda = \pi^{prior}_\Lambda = \mathcal{U}([-1, 1])$
- Observed singular value $d = 0.25$, with Gaussian error $\to \pi^{obs}_\mathcal{D} = \pi^{like}_\mathcal{D} = N(0.25, 0.1^2)$
- π^{pred}_Λ - Gaussian kernel density estimator on $N = 1000$ samples of π^{in}_Λ pushed through Q.

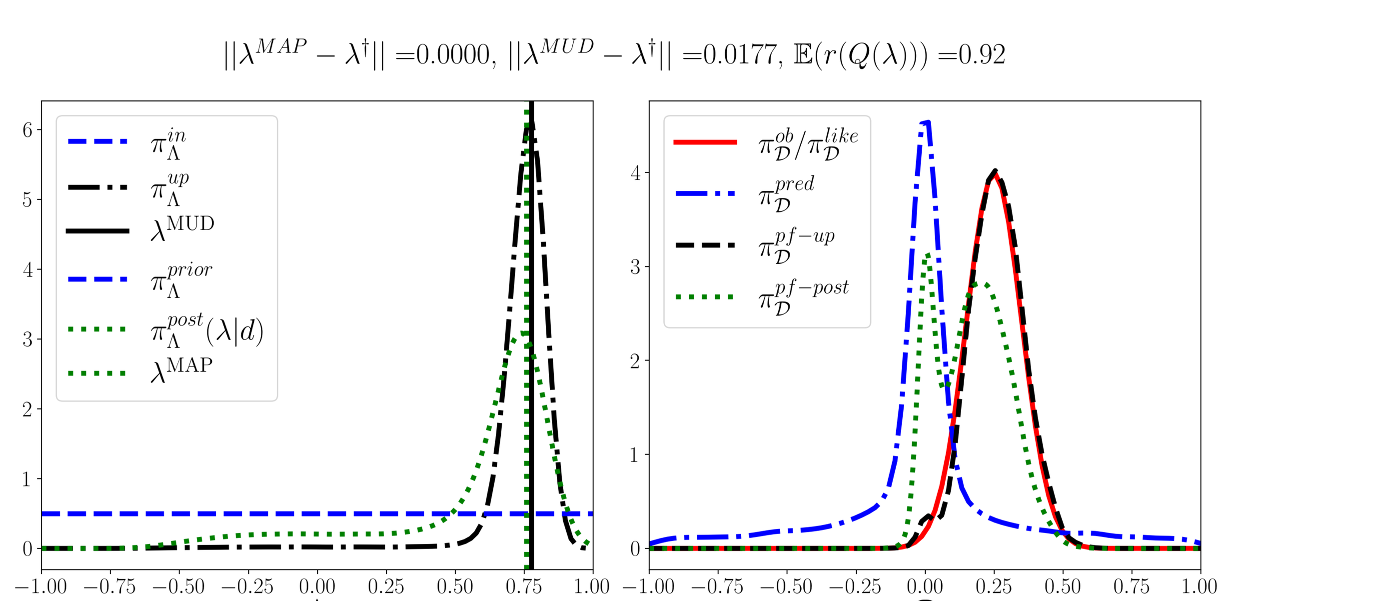

Fig. 1: Comparing Bayesian and Stochastic Inversion. Note how the push-forward of the updated distribution matches the observed distribution (dotted black lines) for the Stochastic Inversion framework - hence why the stochastic inversion solution is often referred to as the **Data Consistent Solution**

Data-Constructed Maps

The Q_{PCA} Map

The main problem with using the SIP framework to solve PIPs is data-assimilation - how do we incorporate new data to solve the PIP and reduce uncertainty (i.e. variance) in our parameter estimate? We propose the usage of **Data-Constructed QoI Maps** to functionally assimilate data and reduce variance as more data is collected. Assume we have:

- $\mathcal{M}_j$ devices collecting data over space and time, each taking N_j measurements.
- Arbitrary ordering of these $n = \sum_{j=1}^m N_j$ data points $\{z_i\}_{i=1}^n$ - d_i equals the ith measurement datum, $\mathcal{M}_{k,i} = \mathcal{M}(\lambda_k; z_i)$ is the ith measurement for the kth simulated sample.

We can define a matrix $X \in \mathbb{R}^{s \times n}$ of Z-scored residuals for a sample set of s samples component-wise as

$$X_{k,i} = \frac{\mathcal{M}_{k,i} - d_i}{\sigma_i}.$$

Letting $p^{(\ell)}$ be the ℓth principle component of X, we define Q_{PCA} component-wise as

$$(Q_{PCA})_\ell(\lambda_k) = \sum_{i=1}^n p_i^{(\ell)} \frac{\mathcal{M}(\lambda_k; z_i) - d_i}{\sigma_i}, \quad 1 \leq \ell \leq n,$$

Properties of Q_{PCA} Map

- Ordering of measurements in X is irrelevant, since PCA does not depend on column order.
- It can be shown that for the Q_{PCA} map the observed distribution $\pi^{obs}_\mathcal{D}$ is always a stationary $N(0,1)$ distribution no matter how many data-points collected.
- We only take up to the first m components that capture a user-specified percentage of variance in the original data set X. We expect the number of components, m, to be equal to the dimension of our parameter space, p, however if data is not sensitive to all parameters present in our inverse problem, this may not be the case. In these examples, we turn to the diagnostic $\mathbb{E}(r)$ as an important measure of the quality of the updated density and thus the reliability of λ^{MUD}.

ADCIRC Problem Set-Up

Shinnecock Inlet Simulated Extreme Event

- Used well known test mesh based on the Shinnecock Inlet on the Outer Barrier of Long Island, NY, USA.
- External forcing using tides, constant air pressure of 1013 millibars, and winds computed from a $0.25°$ hourly CFSv2 10-m wind fields for a period of 16 days (29 December 2017 - 31 January 2018).
- Winds are modified for the purposes of the numerical experiment to simulate a more extreme (Category 4) event, with winds scaled radially down to zero from the point of interest.
- Modified ADCIRC to include a parameterized form of the Garratt wind drag law, with a slope (λ_1) and a cut-off (λ_2) parameter:
$$C_d = \min\left[10^{-3}(.75 + \lambda_1 u), \lambda_2\right]$$
- Constructed a "true" signal by collected water elevation data at an artificial recording station and populating each measurement with i.i.d. $N(0, \sigma^2)$ noise, using $\sigma = 0.05$.
- **Goal: Estimate wind drag parameter values that produced the "true" signal.**

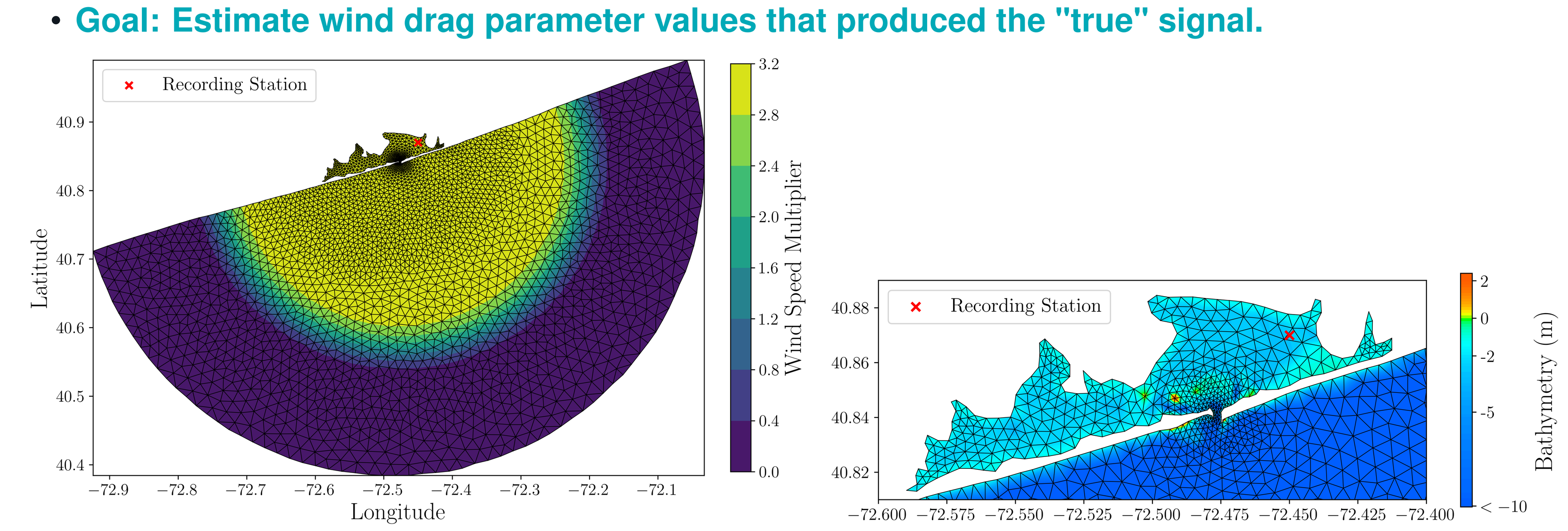

Fig. 2: (left) Shinnecock Inlet Mesh containing 5780 triangular elements. Contours show value of wind multiplier applied to scale winds up artificially near the inlet, tapering them off to zero at the boundaries.(Right) Bathymetry of inlet.

Results: Estimating Wind Drag Coefficients in ADCIRC

Solving the PIP

- Assume that the uncertain parameters (λ_1, λ_2) lie within $\pm 50\%$ of commonly used default values of $(0.067, 0.0025) \to \Lambda = [0.0335, 0.1105] \times [0.00125, 0.00375] \subset \mathbb{R}^2$.
- 1000 samples are generated from a uniform distribution over Λ and pushed through our forward model, ADCIRC, recording data at 3 hour intervals for each sample.
- Three different time windows of data are used to construct Q_{PCA} - One with low winds (T_1), one with high winds (T_2), and one with both (T_3).
- Used diagnostic $\mathbb{E}(r)$ to compare using one ($\mathbb{E}(r_1)$) vs. two ($\mathbb{E}(r_2)$) principal components.

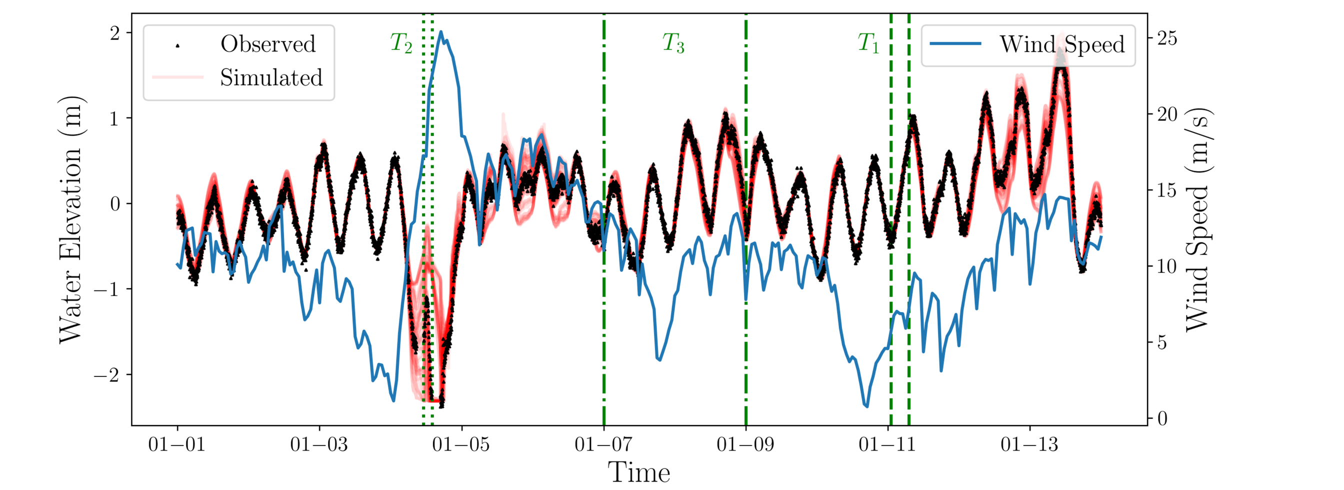

Fig. 3: Water elevations (left axis) for "observed" data (black triangles) and simulated data (faded red lines) along with wind forcing (right axis, blue line). Time windows of data used are indicated in the vertical dashed (T_1), dotted (T_2), and dashed-dotted (T_3) green lines.

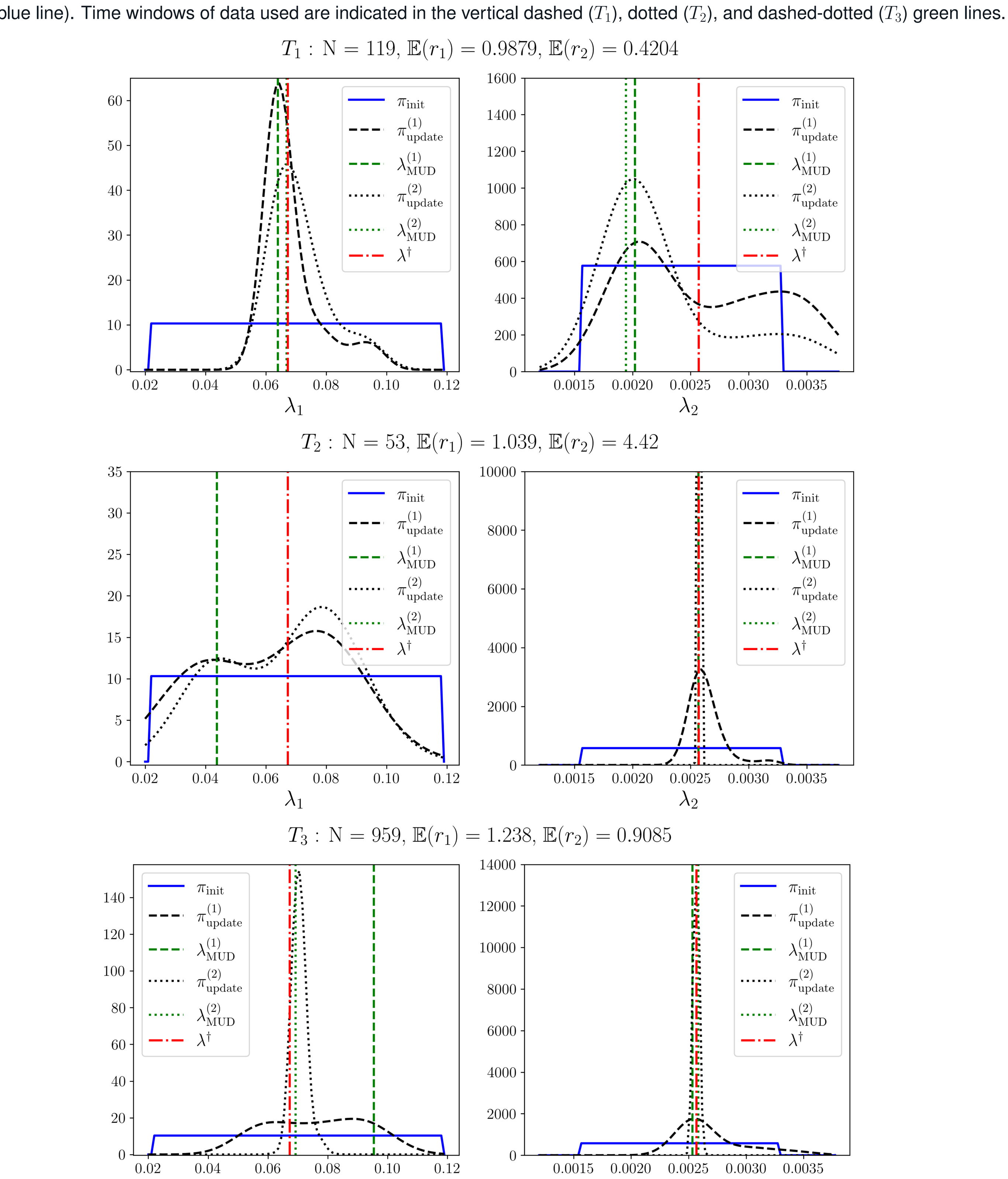

Fig. 4: Solution plots for each parameter containing updated distributions (dotted black line) and mud estimates (dotted green line)for each time window. (Top) T_1 estimates λ_1 well. (Middle) T_2 estimate λ_2 well using 1 component. (Bottom) T_3 estimates both parameters well using 2 (dotted black line) principal components instead of 1 (dashed black line).

Conclusions

- Q_{PCA} map can effectively estimate both parameters provided the data used in the map exhibit sensitivity to those parameters.
- **The diagnostic $\mathbb{E}(r)$ gives us a specific metric to determine the quality of reconstructed distributions for each parameter.**

Data: Code: Preprint: