Variance-based sensitivity of Bayesian inverse problems to the prior distribution

John Darges

Department of Mathematics
North Carolina State University

October 27, 2023

Joint work with Alen Alexanderian and Pierre Gremaud

Supported by NSF through awards DMS-1745654 and DMS 1953271.

Parameter estimation with SEIR model

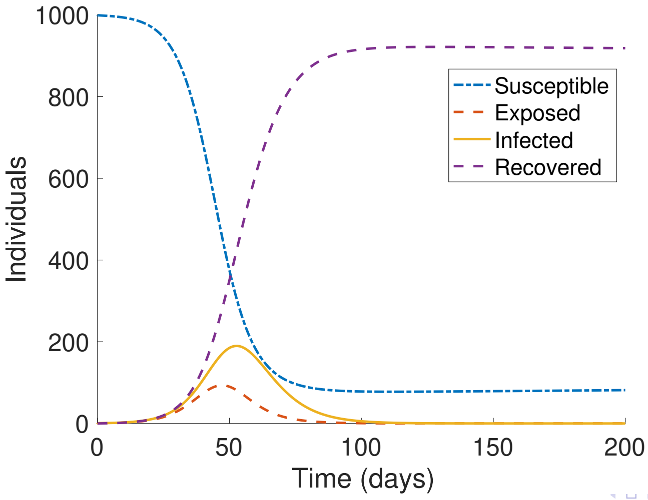
- The SEIR model[1] is a compartmental epidemiology model that describes the outbreak of an epidemic in a population

$$\begin{aligned}
\dot{S} &= \boldsymbol{\mu} N - \boldsymbol{\beta} SI/N - \boldsymbol{\mu} S, && \text{(Susceptible)}\\
\dot{E} &= \boldsymbol{\beta} SI/N - (\boldsymbol{\sigma} + \boldsymbol{\mu})E, && \text{(Exposed)},\\
\dot{I} &= \boldsymbol{\sigma} E - (\boldsymbol{\gamma} + \boldsymbol{\mu})I, && \text{(Infected)},\\
\dot{R} &= \boldsymbol{\gamma} I - \boldsymbol{\mu} R, && \text{(Recovered)}.
\end{aligned} \tag{1}$$

- The total population size stays constant at $N = S(t) + E(t) + I(t) + R(t)$
- Model parameters $\boldsymbol{\theta} = [\ \mu \quad \beta \quad \sigma \quad \gamma\ ]^\top$

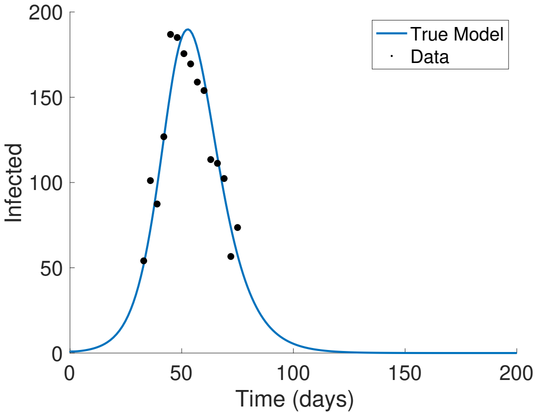
[1] H.W. Hethcote. The Mathematics of Infectious Diseases. 2000.

SEIR dynamics



Goal-oriented parameter estimation

- Suppose we have noisy data on infections
- Goal: find basic reproductive number $R_0 = \frac{\beta}{\gamma+\mu}\frac{\sigma}{\sigma+\mu}$
- R_0 measures how many secondary infection each infection causes on average.
- First we must estimate model parameters $\boldsymbol{\theta} = [\ \mu \quad \beta \quad \sigma \quad \gamma\ ]^\top$



Inverse problem

A deterministic inverse problem[2] has the form

$$\min_{\boldsymbol{\theta}} J(\boldsymbol{\theta}) := \|\mathbf{B}\boldsymbol{y}(\boldsymbol{\theta}) - \mathbf{d}\|^2, \tag{2}$$

where $\boldsymbol{y}$ is the solution to

$$\begin{cases} \boldsymbol{y}' = f(\boldsymbol{y}; \boldsymbol{\theta}) \\ \boldsymbol{y}(t_0) = \boldsymbol{y}_0 \end{cases}, \quad \boldsymbol{y} \in \mathbb{R}^d. \tag{3}$$

Observation operator $\mathbf{B}$ at each measurement time selects only some of the responses from $\boldsymbol{y}(t_i)$ corresponding to data measurements available in array $\mathbf{d}$

[2] A. Tarantola. Inverse Problem Theory and Methods for Model Parameter Estimation. 2005.

In the Bayesian context

- Assume data measurements contain uncertainty as i.i.d. random noise

$$y_i(\boldsymbol{\theta}) - d_i \sim \pi_{\text{noise}}$$

- Bayesian inverse problem treats $\boldsymbol{\theta}$ as a random variable (RV)
- Noise model gives likelihood $\pi_{\text{like}}(\mathbf{d}|\boldsymbol{\theta})$ incorporates uncertainty in data
- Prior distribution $\pi_{\text{pr}}(\boldsymbol{\theta})$ summarizes our beliefs or assumptions about parameters before measuring data
- Combine prior beliefs/assumptions with data to create a posterior distribution for $\boldsymbol{\theta}$

$$\textbf{Bayes' rule: } \pi_{\text{post}}(\boldsymbol{\theta}|\mathbf{d}) = \frac{\pi_{\text{like}}(\mathbf{d}|\boldsymbol{\theta})\pi_{\text{pr}}(\boldsymbol{\theta})}{\int \pi_{\text{like}}(\mathbf{d}|\boldsymbol{\theta})\pi_{\text{pr}}(\boldsymbol{\theta})d\boldsymbol{\theta}}$$

- Note: A quantity of interest $q(\boldsymbol{\theta})$ (e.g. R_0) becomes a RV with a posterior

Impact of uncertainty in the prior

- Uncertainty does not just come from the data
- Suppose some hyperparameters $\boldsymbol{\xi}$ parameterize the prior distribution $\pi_{\text{pr}} = \pi_{\text{pr}}^{\boldsymbol{\xi}}$
- E.g. if prior is log-normal, $\log(\boldsymbol{\theta}) \sim \mathcal{N}(m, s^2)$, then $\boldsymbol{\xi} = \begin{bmatrix} m & s \end{bmatrix}^\top$
- Typically, $\boldsymbol{\xi}$ is uncertain

Question: How does uncertainty in the prior assumptions affect our estimation of the QoI?

Related Work

- Robust Bayesian analysis[3] studies robustness to different priors in inference problems
- Hyper-differential sensitivity analysis (HDSA) has been used for Bayesian inverse problems to study measures of posterior uncertainty in[4] and[5]
- Derivative-based global sensitivity measures (DGSM) has been used to study the sensitivity of information gain to uncertain model parameters[6]
- Prior and model hyperparameter sensitity in variance-based framework. Emulated by Gaussian process, training data computed by MCMC[7]

---

[3] J. O. Berger, D. R. Insua, F. Ruggeri. Bayesian Robustness. 2000.

[4] I. Sunseri, A. Alexanderian, J. Hart, B. van Bloemen Waanders. Hyper-differential sensitivity analysis for nonlinear Bayesian inverse problems. 2022.

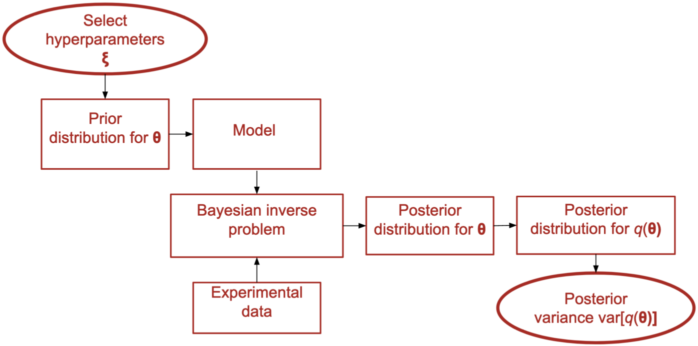
[5] W. Reese, A. Saibaba, J. Hart, B. van Bloeman Waanders, M. Perego, J. Jakeman. Hyper-differential sensitivity analysis in the context of Bayesian inference applied to ice-sheet problems. 2022.

[6] A. Chowdhary, A. Alexanderian, S. Tong, G. Stadler. Sensitivity Analysis of Information Gain in Infinite-Dimensional Bayesian Linear Inverse Problems. 2023.

[7] I. Vernon, J. P. Gosling. A Bayesian Computer Model Analysis of Robust Bayesian Analyses. 2022.

Hyperparameter-to-statistic mapping

- Let $F(\xi)$ map prior hyperparameters to posterior statistic of QoI $q(\boldsymbol{\theta})$
- E.g. $F_{\mathrm{var}}(\xi) = \mathrm{var}^{\xi}_{\mathrm{post}}(q)$, $F_{\mathrm{mean}}(\xi) = \mathbb{E}^{\xi}_{\mathrm{post}}(q)$



- **Goal:** Study sensitivity of $F(\xi)$ to prior hyperparameters

Variance-based global sensitivity analysis

- Let $Y = F(\boldsymbol{\xi})$ where $Y \in \mathbb{R}$, and $\boldsymbol{\xi} \sim \mathcal{D}$ has independently distributed entries
- Sobol' indices[8] quantify the contribution of each input to variance in model output:

$$S_k := \frac{\mathrm{var}[F_k(\xi_k)]}{\mathrm{var}[F(\boldsymbol{\xi})]}, \quad S_k^T := 1 - \frac{\mathrm{var}[\mathbb{E}(F(\boldsymbol{\xi})|\xi_j,\ j \neq k)]}{\mathrm{var}[F(\boldsymbol{\xi})]}$$

- $F_k(\xi_k) := \int F(\boldsymbol{\xi}) d\boldsymbol{\xi}_{-k} - \mathbb{E}(F\boldsymbol{\xi})$, where $d\boldsymbol{\xi}_{-k}$ denotes integration over inputs **except** ξ_k
- First order Sobol' index S_k measures influence of ξ_k **outside of** interactions
- Total Sobol' index S_k^T measures influence of ξ_k **including** interactions with other inputs

[8] C. Prieur, S. Tarantola. Variance-based sensitivity analysis: Theory and estimation algorithms. 2017.

Computing Sobol' indices of $F(\xi)$ - Challenges

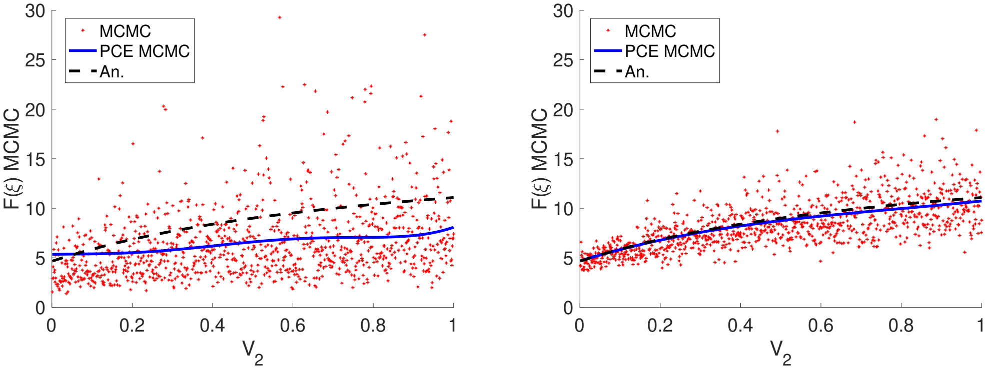
1. Estimating Sobol' indices S_k, S_k^T requires **thousands** of evaluations of $F(\xi)$
2. Evaluating $F(\xi)$ requires **thousands** of evaluations of $\pi(\mathbf{d}|\boldsymbol{\theta})$ (involves e.g. solving ODE model)
   - Evaluating $F(\xi)$ requires sampling from π_{post}^{ξ} with Markov Chain Monte Carlo (MCMC) run
   - Each MCMC run can require thousands of evaluations of $\pi_{\text{like}}(\mathbf{d}|\boldsymbol{\theta})$

- **Thousands** $\times$ **Thousands** $=$ **Too expensive!**

- **Idea:** Likelihood does not change when ξ changes
- Can we recycle likelihood evaluations when we look at different ξ's?

Example of evaluating every $F(\xi)$ with MCMC

Simple case where one prior hyperparameter changes



(a) 100 iterations per MCMC run

(b) 1000 iterations per MCMC run

Importance sampling

Importance sampling[9] uses re-weighted samples from an auxiliary distribution π_{IS} to integrate over $\pi_{\text{post}}^{\boldsymbol{\xi}}$ without sampling from it

$$
\begin{aligned}
F_{\text{mean}}(\boldsymbol{\xi}) = \int_{\mathbb{R}^n} q(\boldsymbol{\theta}) \pi_{\text{post}}^{\boldsymbol{\xi}}(\boldsymbol{\theta}) \; d\boldsymbol{\theta} = \int_{\mathbb{R}^n} q(\boldsymbol{\theta}) \frac{\pi_{\text{post}}^{\boldsymbol{\xi}}(\boldsymbol{\theta})}{\pi_{\text{IS}}(\boldsymbol{\theta})} \pi_{\text{IS}}(\boldsymbol{\theta}) \; d\boldsymbol{\theta} \\
\approx \sum_{j=1}^{N} q(\boldsymbol{\theta}_i) \frac{\pi_{\text{post}}^{\boldsymbol{\xi}}(\boldsymbol{\theta}_j)}{\pi_{\text{IS}}(\boldsymbol{\theta}_j)}, \quad \theta_j \sim \pi_{\text{IS}}
\end{aligned}
\tag{4}
$$

We can reuse the same sample set to approximate $F(\boldsymbol{\xi})$ for every choice of $\boldsymbol{\xi}$
Can we avoid re-evaluating $\pi_{\text{like}}(\mathbf{d}|\boldsymbol{\theta})$ on the sample set when we change $\boldsymbol{\xi}$?

[9] S. Tokdar, R. Kass, Importance sampling: A review. 2010.

Importance sampling tailored to Bayesian inverse problems

Let the importance sampling distribution be

$$\pi_{\mathrm{IS}}(\boldsymbol{\theta}) := \pi_{\mathrm{post}}^{\mathrm{IS}}(\boldsymbol{\theta}|\mathbf{d}) \propto \pi_{\mathrm{like}}(\mathbf{d}|\boldsymbol{\theta}) \cdot \pi_{\mathrm{pr}}^{\mathrm{IS}}(\boldsymbol{\theta}), \tag{5}$$

We let $\pi_{\mathrm{pr}}^{\mathrm{IS}}$ belong to the same parameterized family as $\pi_{\mathrm{pr}}^{\boldsymbol{\xi}}$

Because $\pi_{\mathrm{like}}(\mathbf{d}|\boldsymbol{\theta})$ appears in the expressions of $\pi_{\mathrm{post}}^{\boldsymbol{\xi}}$ and $\pi_{\mathrm{post}}^{\mathrm{IS}}$, we get convenient cancellation

$$\frac{\pi_{\mathrm{post}}^{\boldsymbol{\xi}}}{\pi_{\mathrm{post}}^{\mathrm{IS}}} \propto \frac{\pi_{\mathrm{like}} \cdot \pi_{\mathrm{pr}}^{\boldsymbol{\xi}}}{\pi_{\mathrm{like}} \cdot \pi_{\mathrm{pr}}^{\mathrm{IS}}} = \frac{\pi_{\mathrm{pr}}^{\boldsymbol{\xi}}}{\pi_{\mathrm{pr}}^{\mathrm{IS}}} \tag{6}$$

Importance sampling estimator

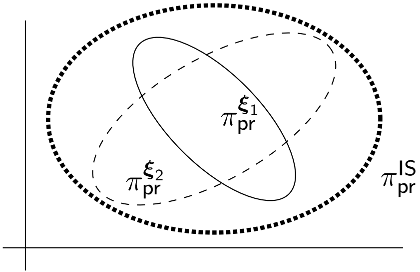
Estimate $F_{\text{mean}}(\boldsymbol{\xi})$ by Monte Carlo integration:

$$\begin{aligned} F_{\text{mean}}(\boldsymbol{\xi}) &= \int_{\mathbb{R}^n} q(\boldsymbol{\theta}) \frac{\pi_{\text{post}}^{\boldsymbol{\xi}}(\boldsymbol{\theta})}{\pi_{\text{post}}^{\text{IS}}(\boldsymbol{\theta})} \pi_{\text{IS}}(\boldsymbol{\theta})\ d\boldsymbol{\theta} \\ &= \frac{1}{C_P} \int_{\mathbb{R}^n} q(\boldsymbol{\theta}) \frac{\pi_{\text{pr}}^{\boldsymbol{\xi}}(\boldsymbol{\theta})}{\pi_{\text{pr}}^{\text{IS}}(\boldsymbol{\theta})} \pi_{\text{IS}}(\boldsymbol{\theta})\ d\boldsymbol{\theta} \\ &\approx \frac{1}{\widehat{C_P}} \sum_{j=1}^{N} q(\boldsymbol{\theta}_i) \frac{\pi_{\text{pr}}^{\boldsymbol{\xi}}(\boldsymbol{\theta}_j)}{\pi_{\text{pr}}^{\text{IS}}(\boldsymbol{\theta}_j)}, \quad \theta_j \sim \pi_{\text{post}}^{\text{IS}}, \end{aligned} \tag{7}$$

where $\widehat{C_P} = \sum_{j=1}^{N} \frac{\pi_{\text{pr}}^{\boldsymbol{\xi}}(\boldsymbol{\theta}_j)}{\pi_{\text{pr}}^{\text{IS}}(\boldsymbol{\theta}_j)}$ estimates the ratio C_P of normalization constants of $\pi_{\text{post}}^{\text{IS}}$ and $\pi_{\text{post}}^{\boldsymbol{\xi}}$

How to choose the importance sampling prior

How do we choose $\pi_{\mathrm{pr}}^{\mathrm{IS}}$ so that $\pi_{\mathrm{post}}^{\mathrm{IS}}$ is a good for all our choices of $\boldsymbol{\xi}$?



We want the “high-density region” of $\pi_{\mathsf{pr}}^{\mathsf{IS}}$ to cover the high-density regions of every $\pi_{\mathsf{pr}}^{\boldsymbol{\xi}}$

Using a diagnostic to choose the importance sampling prior

- A common diagnostic[10] for importance sampling is seeing if the following quantity is small

$$\int \left(\frac{\pi_{\text{post}}^{\boldsymbol{\xi}}(\boldsymbol{\theta})}{\pi_{\text{post}}^{\text{IS}(\boldsymbol{\theta})}} - 1 \right)^2 \pi_{\text{post}}^{\text{IS}}(\boldsymbol{\theta})\ d\boldsymbol{\theta} = \int \frac{(\pi_{\text{post}}^{\boldsymbol{\xi}}(\boldsymbol{\theta}))^2}{\pi_{\text{post}}^{\text{IS}}(\boldsymbol{\theta})} d\boldsymbol{\theta} - 1 \tag{8}$$

  Note the following bound on this

$$\int \frac{(\pi_{\text{post}}^{\boldsymbol{\xi}}(\boldsymbol{\theta}))^2}{\pi_{\text{post}}^{\text{IS}}(\boldsymbol{\theta})} d\boldsymbol{\theta} \leq \left(\frac{\mathcal{M}}{C^{\boldsymbol{\xi}}} \right)^2 \int \frac{(\pi_{\text{pr}}^{\boldsymbol{\xi}}(\boldsymbol{\theta}))^2}{\pi_{\text{pr}}^{\text{IS}}(\boldsymbol{\theta})} d\boldsymbol{\theta}, \quad \mathcal{M} = \max_{\boldsymbol{\theta}} \pi_{\text{like}}(\mathbf{d}|\boldsymbol{\theta}). \tag{9}$$

- This suggests that $\pi_{\text{pr}}^{\text{IS}}$ close to $\pi_{\text{pr}}^{\boldsymbol{\xi}}$ will mean $\pi_{\text{post}}^{\text{IS}}$ close to $\pi_{\text{post}}^{\boldsymbol{\xi}}$

[10] A.B. Owen. Monte Carlo theory, methods and examples. 2013.

Algorithm

1. Find $\pi_{\mathrm{pr}}^{\mathrm{IS}}$ which minimally varies from all $\pi_{\mathrm{pr}}^{\boldsymbol{\xi}}$
2. Sample $\{\boldsymbol{\theta}_j\}_{j=1}^N$ from $\pi_{\mathrm{post}}^{\mathrm{IS}}$ using one MCMC run
3. For $i = 1, \ldots, M$ compute IS estimator of $F(\boldsymbol{\xi}_i)$
4. Build/train surrogate models (PCE[11] and ELM[12]) of $F(\boldsymbol{\xi})$ from $\{\boldsymbol{\xi}_i, F(\boldsymbol{\xi}_i)\}_{i=1}^M$ to estimate Sobol' indices

Use two surrogate methods to validate results

Most expensive step is second step

[11] A. Doostan, J. Hampton. Compressive Sampling Methods for Sparse Polynomial Chaos Expansions. 2017.

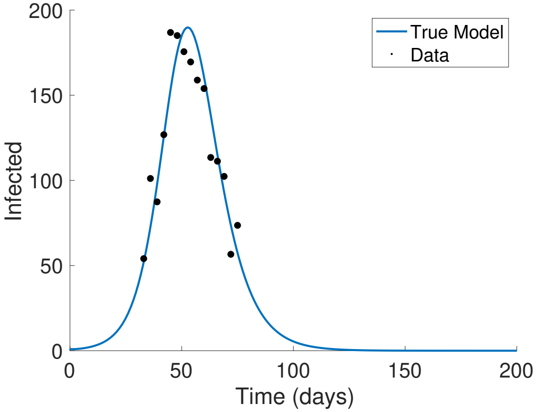
[12] J. Darges, A. Alexanderian, P. Gremaud. Extreme learning machines for variance-based global sensitivity analysis. 2022.

Prior hyperparameter uncertainty in SEIR problem

- Recall basic reproductive number $R_0 = \frac{\beta}{\gamma+\mu}\frac{\sigma}{\sigma+\mu}$
- We assume a log-normal prior on $\boldsymbol{\theta} = [\ \mu \quad \beta \quad \gamma \quad \sigma\ ]^\top$ since parameters are positive
- $\log(\boldsymbol{\theta}) \sim \mathcal{N}(\boldsymbol{m}, \boldsymbol{\Sigma})$ defines $\boldsymbol{\xi}$ by

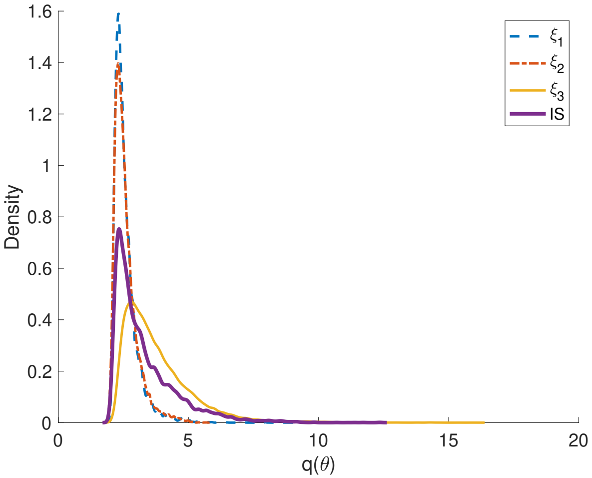
$$\boldsymbol{m} = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \xi_5 & 0 & 0 & 0 \\ 0 & \xi_6 & 0 & 0 \\ 0 & 0 & \xi_7 & 0 \\ 0 & 0 & 0 & \xi_8 \end{bmatrix}$$

- Uncertain $\xi_i \sim \mathcal{U}([a_i, b_i])$ are uniformly distributed



Effectiveness of importance sampling

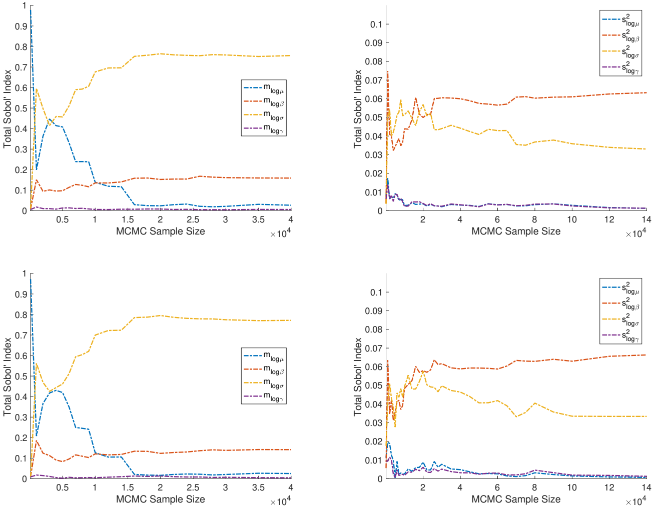
- Optimal $\pi_{\text{pr}}^{\text{IS}}$ takes mean at the nominal mean hyperparameters and wider variance
- High density region $\pi_{\text{post}}^{\text{IS}}$ covers high density regions of every $\pi_{\text{post}}^{\boldsymbol{\xi}}$
- How many MCMC samples are needed for accurate Sobol' indices of $F(\boldsymbol{\xi})$?



Convergence study for $F_{\text{mean}}(\xi) = \mathbb{E}^{\xi}_{\text{post}}(R_0)$

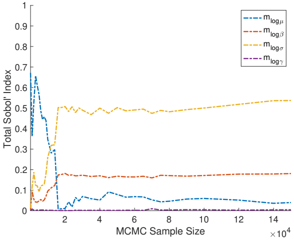
ELM surrogate:

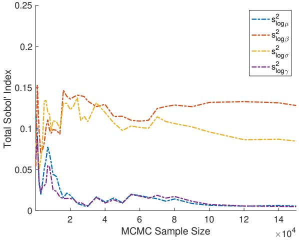
PCE surrogate:



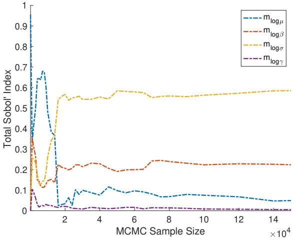
Convergence study for $F_{\mathrm{var}}(\xi) = \mathrm{var}^{\xi}_{\mathrm{post}}(R_0)$

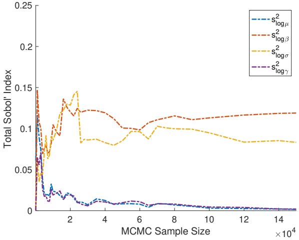
ELM surrogate:





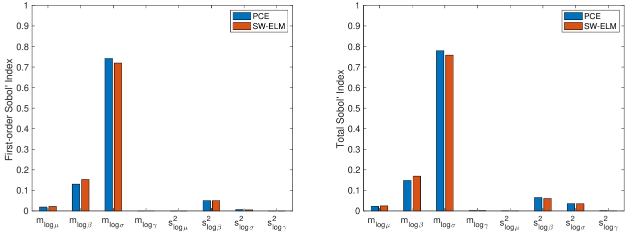
PCE surrogate:



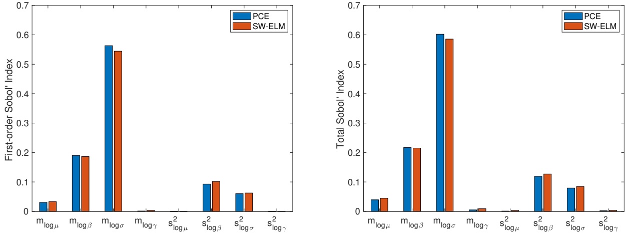


Prior hyperparameter importance

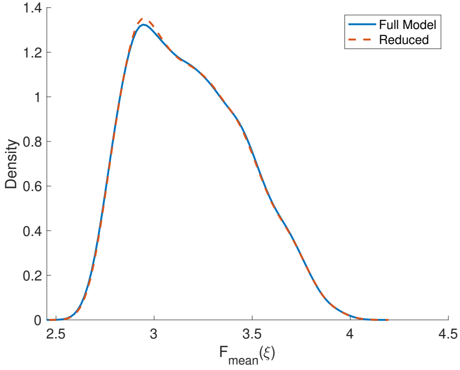
$$F_{\text{mean}}(\xi) = \mathbb{E}^{\xi}_{\text{post}}(R_0):$$

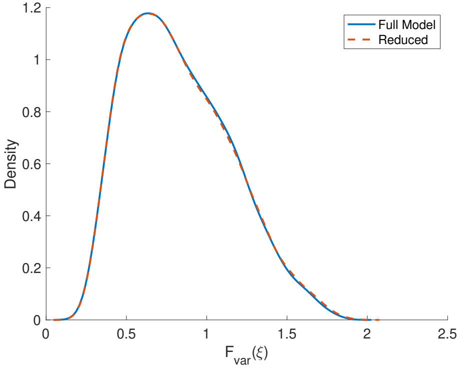


$$F_{\text{var}}(\xi) = \text{var}^{\xi}_{\text{post}}(R_0):$$



Uncertainty caused by influential prior hyperparameters





Conclusion

- Showed that the solution to Bayesian inverse problems has complicated dependence on prior hyperparameters
- Developed feasible method for understanding effects of uncertainty in prior hyperparameters[13]
- For future work, explore high dimensional Bayesian inverse problems
- Explore ways to broaden framework to include other hyerparameters

[13] J. Darges, A. Alexanderian, P. Gremaud. Variance-based sensitivity of Bayesian inverse problems to the prior distribution. 2023.

References I

[1] J. O. Berger, D. R. Insua, and F. Ruggeri. Bayesian Robustness. In D. R. Insua and F. Ruggeri, editors, *Robust Bayesian Analysis*. Springer, 2000.

[2] A. Chowdhary, A. Alexanderian, S. Tong, and G. Stadler. Sensitivity Analysis of the Information Gain in Infinite-Dimensional Bayesian Linear Inverse Problems. *in preparation*, 2023.

[3] J. Darges, A. Alexanderian, and P. Gremaud. Extreme learning machines for variance-based global sensitivity analysis. *preprint*, 2022.

[4] J. E. Darges, A. Alexanderian, and P. A. Gremaud. Variance-based sensitivity of Bayesian inverse problems to the prior distribution. *preprint*, 2023.

[5] H. W. Hethcote. The Mathematics of Infectious Diseases. *SIAM Review*, 42(4), 2000.

[6] L. Le Gratiet, S. Marelli, and B. Sudret. Metamodel-based sensitivity analysis: polynomial chaos expansions and Gaussian processes. In R. Ghanem, D. Higdon, and H. Owhadi, editors, *Handbook of Uncertainty Quantification*. Springer, 2017.

References II

[7] A. B. Owen. Importance sampling. In *Monte Carlo theory, methods and examples*, chapter 9. 2013.

[8] C. Prieur and S. Tarantola. Variance-based sensitivity analysis: Theory and estimation algorithms. In R. Ghanem, D. Higdon, and H. Owhadi, editors, *Handbook of Uncertainty Quantification*. Springer, 2017.

[9] W. Reese, J. Hart, B. van Bloemen Waanders, M. Perego, J. Jakeman, and A. Saibaba. Hyper-differential sensitivity analysis in the context of Bayesian inference applied to ice-sheet problems, 2022.

[10] A. Tarantola. *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics.

[11] S. Tokdar and R. Kass. Importance sampling: A review. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2, 2010.