Sensitivity Analysis: Quantifying What Matters

Women in Data Science 2020

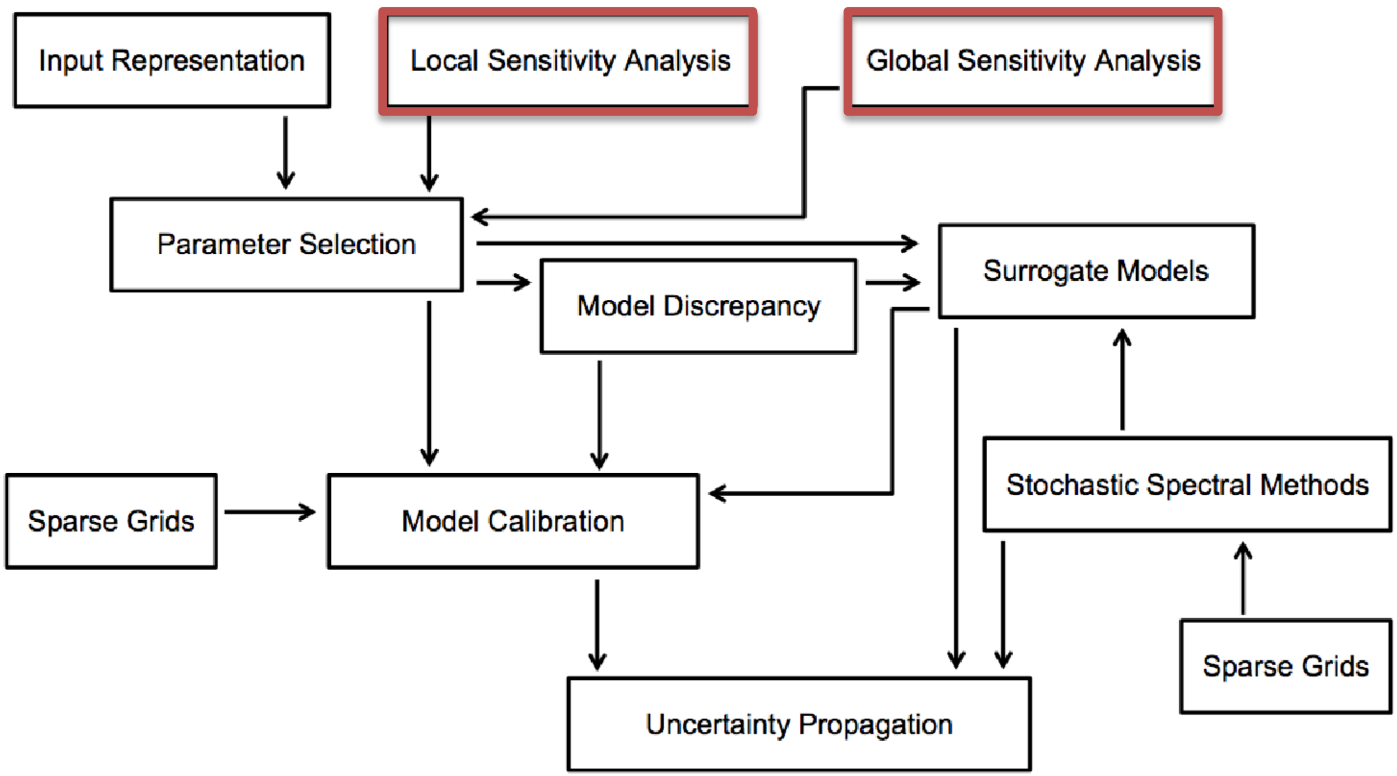
Kathleen Schmidt

March 2, 2020

LLNL-PRES-805661
This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC

Lawrence Livermore National Laboratory

Introduction to Uncertainty Quantification



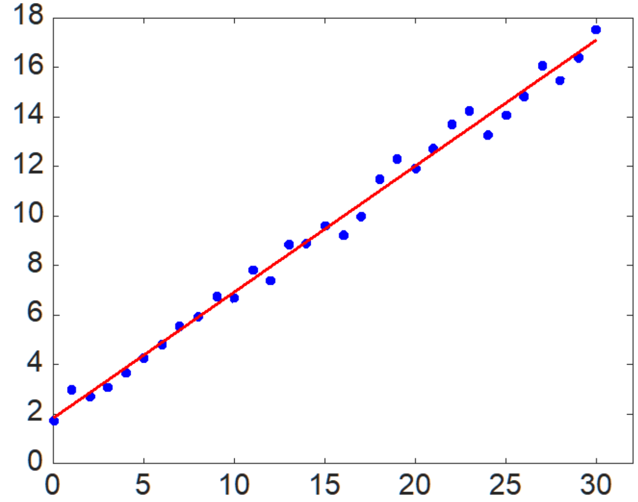
From (Smith, 2014)

Model Calibration

$$y_j = f(x_j;\theta) + \varepsilon_j \ , \ \varepsilon_j \sim N(0,\sigma^2)$$

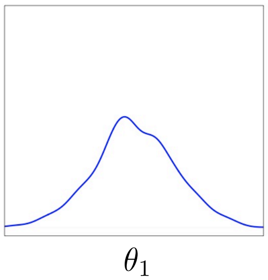
- Frequentist (classical) framework
  - Fixed but unknown parameters
  - Use estimators to approximate values

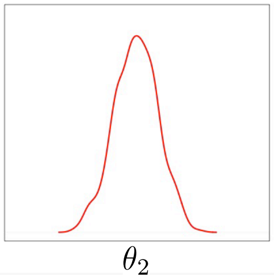
Data $\Rightarrow$ Estimator $\Rightarrow$ $\hat{\theta}$



- Bayesian framework
  - Parameters are random variables with associated distributions

Prior → Posterior ← Data





$$\pi(\theta|\nu) \propto \pi(\nu|\theta)\pi_0(\theta)$$

Sensitivity Analysis

How much does changing a parameter affect the model output?

Model: $y = f(x;\theta)$

$f(x;\tilde{\theta})$ → Big Change in Output — **Sensitive Parameter**

$[\theta_1 + \delta, \theta_2, \ldots, \theta_n]$

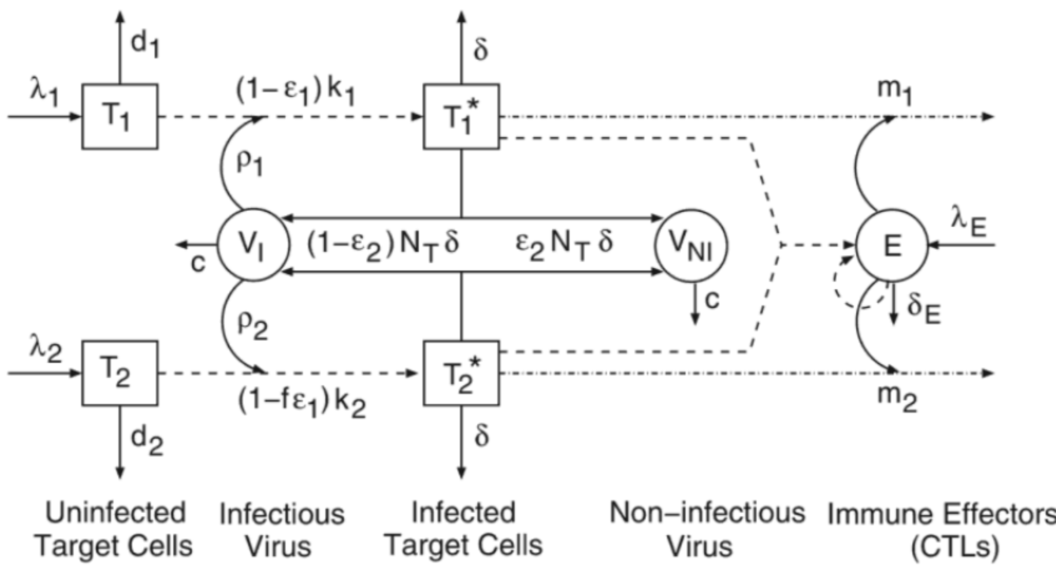
$f(x;\tilde{\theta})$ → Small Change in Output — **Insensitive Parameter**

$[\theta_1 + \delta, \theta_2, \ldots, \theta_n]$

How does this help with calibration?

- Models can have (very!) large parameter sets, resulting in
  - High computational cost
  - Identifiability problems

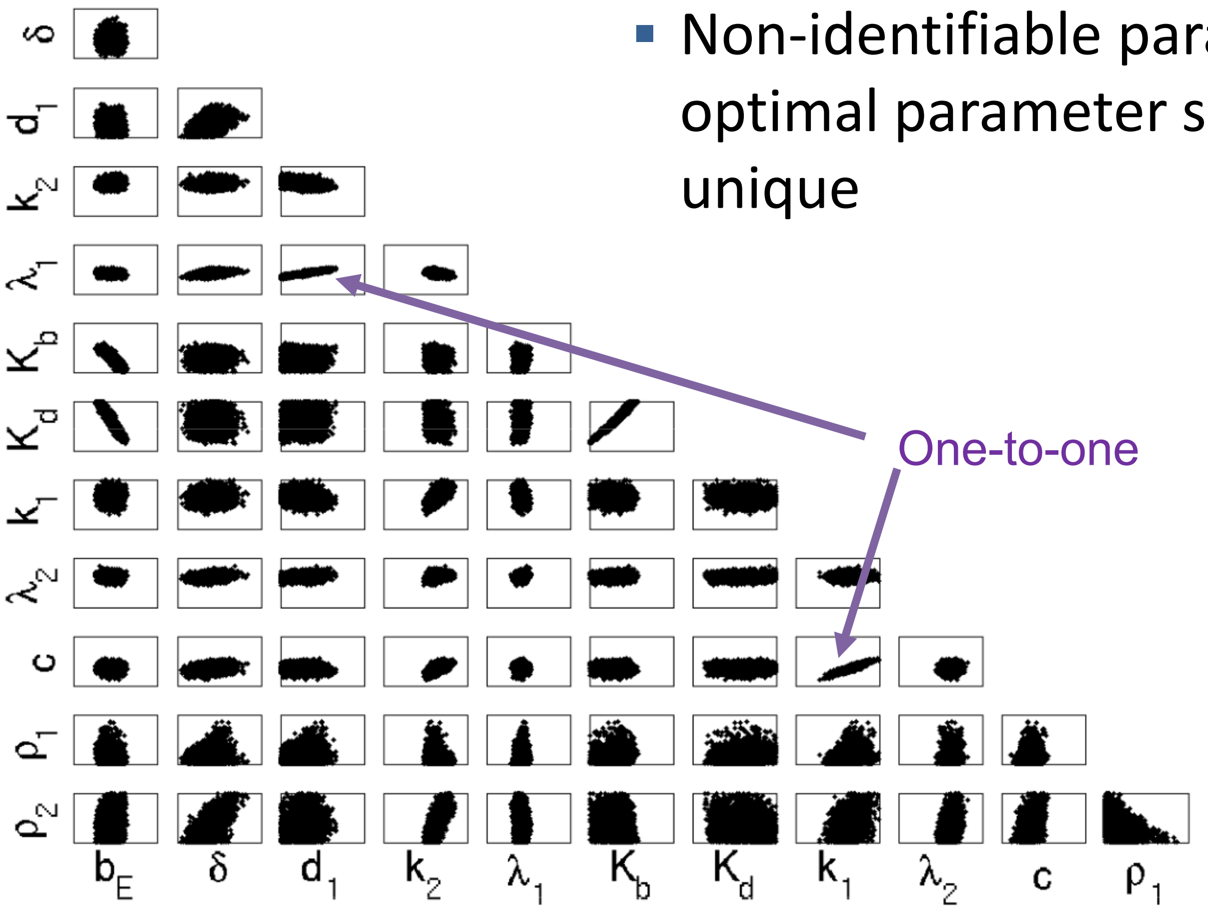
HIV model with 30+ parameters



From (Smith, 2014)

Choose a smaller set of parameters to estimate

Parameter Identifiability



From (Wentworth, 2015)

- Non-identifiable parameter set: optimal parameter set is non-unique

Global vs Local Sensitivity Analysis

- Local sensitivity
  - How does the output change when parameters are perturbed around a nominal value?
- Global sensitivity
  - How do input parameters sampled from the entire admissible parameter space affect the model output?

Local Sensitivity Analysis: Partial Derivatives

- Partial derivatives inform how the function changes with respect to parameters *in the neighborhood of where it is evaluated*

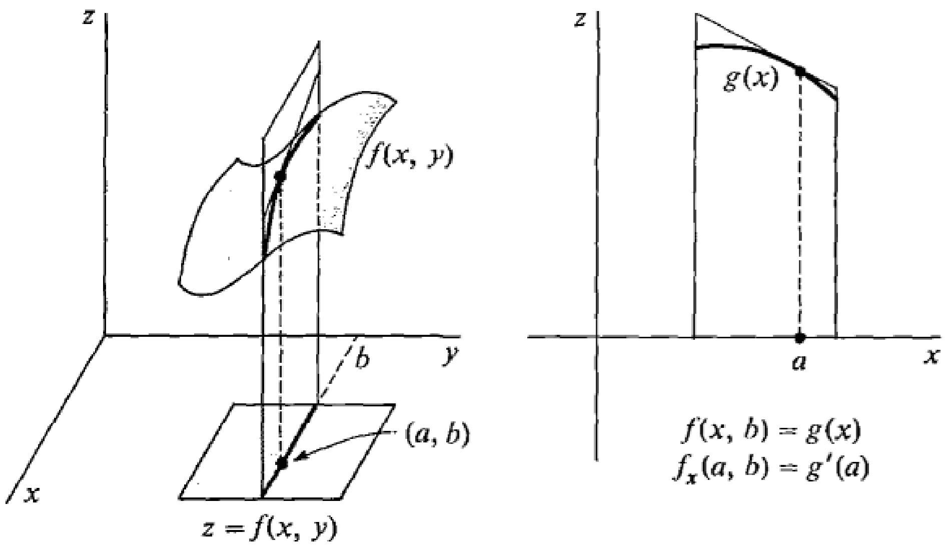


Image from http://www.vias.org/calculus/11_partial_differentiation_03_01.html

Global Sensitivity Analysis: Sobol' Decomposition

Decompose variance:

$$\mathrm{Var}(Y) = \sum_{i=1}^{p} V_i + \sum_{i<j}^{p} V_{ij} + \cdots + V_{1,2,\ldots,p}$$

Sobol' Indices:

$$S_i = \frac{V_i}{\mathrm{Var}(Y)}, \ S_{ij} = \frac{V_{ij}}{\mathrm{Var}(Y)}, \ldots, S_{1,2,\ldots,p} = \frac{V_{1,2,\ldots,p}}{\mathrm{Var}(Y)}$$

1st Order | **2nd Order** | **pth Order**

Total Sensitivity:

$$S_{T_i} = S_i + \sum_{i \neq j} S_{ij} + \sum_{i \neq j \neq k} S_{ijk} + \ldots$$

Note: Analytic calculation of Sobol' indices requires evaluation of high-dimensional integrals

"Easy" Case: Scalar Output

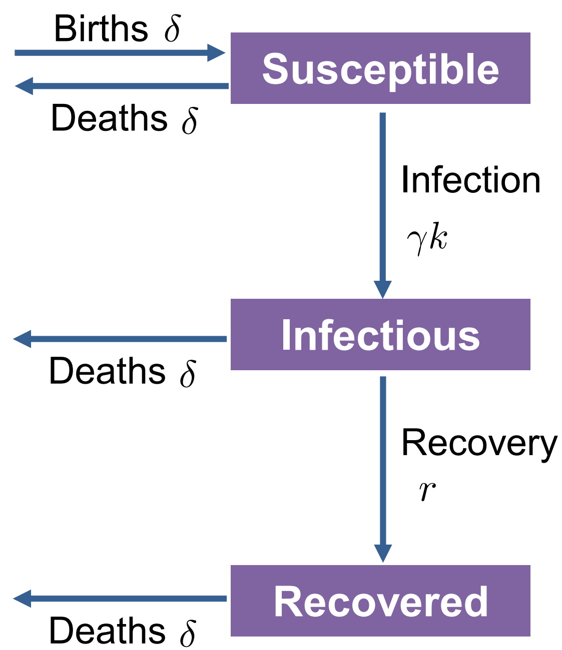
SIR Disease Model:

$$\frac{dS}{dt} = \delta N - \delta S - \gamma k I S \quad , \; S(0) = S_0$$
$$\frac{dI}{dt} = \gamma k I S - (r + \delta) I \quad , \; I(0) = I_0$$
$$\frac{dR}{dt} = rI - \delta R \quad , \; R(0) = R_0$$

Scalar QoI: $y = \int_0^5 R(t,q)\,dt$

Parameter	Meaning	Value
$S(t)$	Susceptible population	$S(0) = 900$
$I(t)$	Infected population	$I(0) = 100$
$R(t)$	Susceptible population	$R(0) = 0$
N	Population size	$\mathrm{N} = 1000$
γ	Infection coefficient	-
k	Interaction coefficient	-
r	Recovery rate	-
δ	Birth rate/Death rate	-

Example from (Smith, 2014)

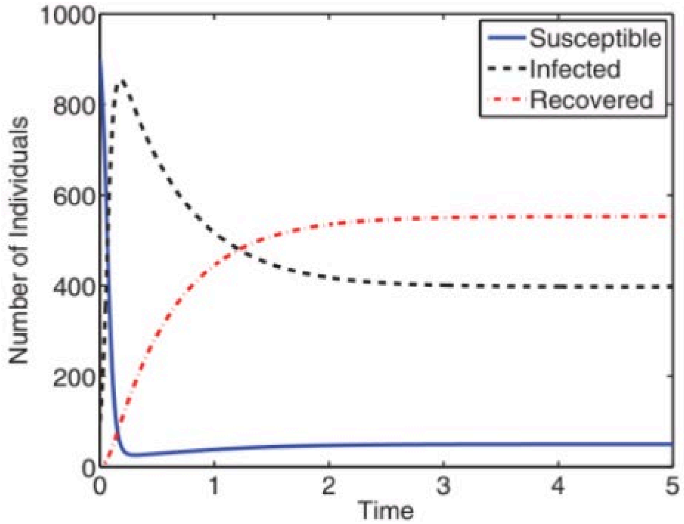


We assume:

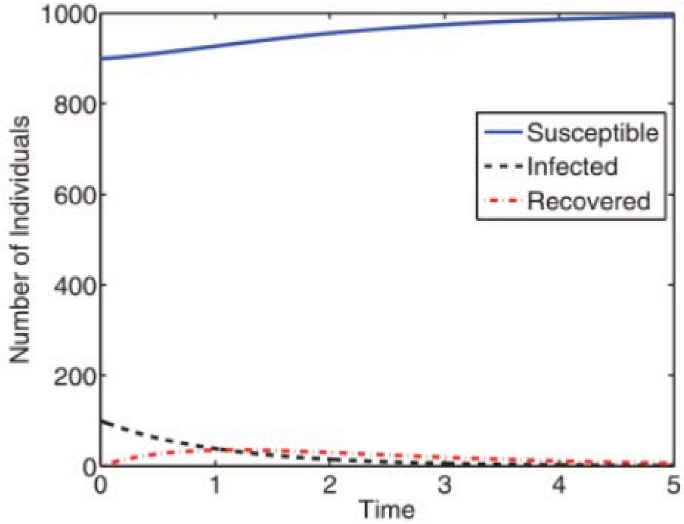
$$\gamma \sim \mathcal{U}(0,1) \; , \; k \sim \mathrm{Beta}(\alpha, \beta) \; , \; r \sim \mathcal{U}(0,1) \; , \; \delta \sim \mathcal{U}(0,1)$$

"Easy" Case: Scalar Output (cont.)

High Interaction $k \sim \text{Beta}(2,7)$



Low Interaction $k \sim \text{Beta}(0.2,15)$



$$\gamma \sim \mathcal{U}(0,1)\ ,\ k \sim \text{Beta}(\alpha,\beta)\ ,\ r \sim \mathcal{U}(0,1)\ ,\ \delta \sim \mathcal{U}(0,1)$$

		γ	k	r	δ
High	S_i	0.0547	0.0199	**0.6213**	0.1972
High	S_{T_i}	0.0923	0.0954	**0.7979**	0.2982
Low	S_i	0.0485	**0.5133**	0.1250	0.0750
Low	S_{T_i}	0.1106	**0.8320**	0.2353	0.1052

Example from (Smith, 2014)

Harder Case: Functional Output

- Model with functional output

$$\text{Model:}\quad \overbrace{y(t)}^{\text{Functional output}} = f(\underset{\text{Parameters}}{\mathbf{x}}, \underset{\text{Input to function}}{t}) + \epsilon\ ,\ \ \epsilon \sim N(0, \sigma^2)$$

- Two options
  - Treat functional input as a parameter—i.e., quantify its sensitivity
  - Examine how sensitivity changes with respect to input parameter(s)

Functional Output (cont.)

Preston-Tonks-Wallace Model:

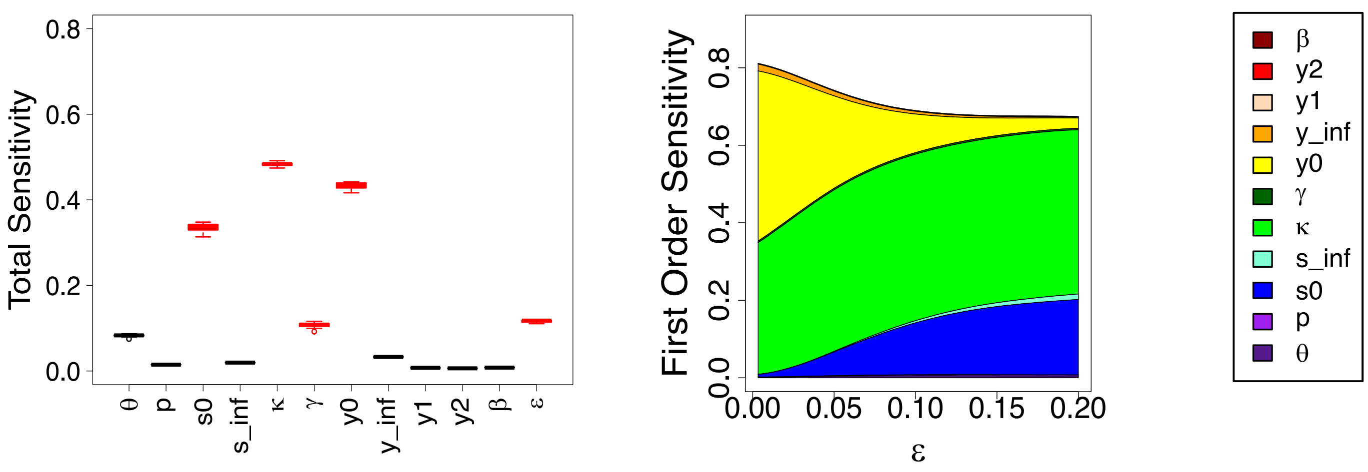
strain

$$\sigma = \sigma_s + \frac{1}{p}(s_0 - \sigma_y)\ln\left[1 - \left(1 - \exp\left[-p\frac{\sigma_s - \sigma_y}{s_0 - \sigma_y}\right]\right)\exp\left(-\frac{p\theta\varepsilon^p}{(s_0 - \sigma_y)\left(\exp\left[p\frac{\sigma_s - \sigma_y}{s_0 - \sigma_y}\right] - 1\right)}\right)\right],$$

stress

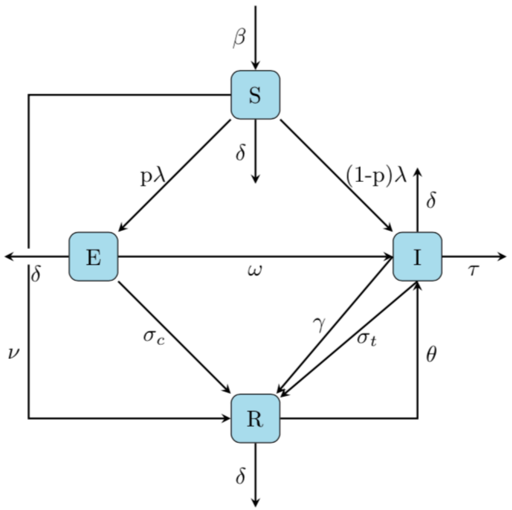
$$\sigma_y = \max\left[y_0 - (y_0 - y_\infty)\mathrm{erf}[\kappa T \ln(\gamma\dot{\xi}/\dot{\varepsilon}^p)], \min\left[y_1(\dot{\varepsilon}^p/\gamma\dot{\xi})^{y_2}, s_{0d}(\dot{\varepsilon}^p/\gamma\dot{\xi})^{\beta}\right]\right]$$

$$\sigma_S = \max\left[s_0 - (s_0 - s_\infty)\mathrm{erf}[\kappa T \ln(\gamma\dot{\xi}/\dot{\varepsilon}^p)], s_{0d}(\dot{\varepsilon}^p/\gamma\dot{\xi})^{\beta}\right]$$

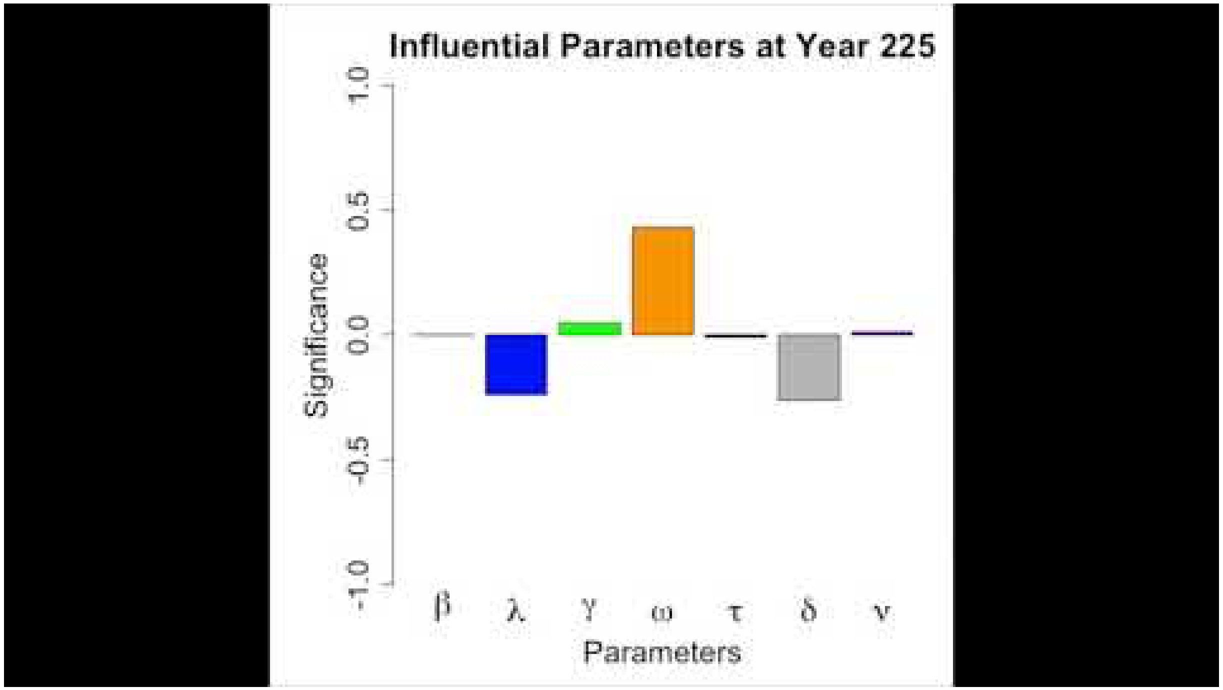


Functional Output: How Sensitivities Change

Tuberculosis SEIR Model:



Parameter	Meaning
β	Birth rate
δ	Natural death rate
τ	TB death rate
ν	Rate of vaccination
λ	Rate of infection
ω	Rate of deterioration
γ	Rate of recovery
σ_c	Latent TB treatment
σ_t	Active TB treatment
θ	Rate of reinfection
p	Slow TB proportion



Example from (Clark, et al., 2019)

Video from
https://www.youtube.com/watch?v=0cUAISFzPuk

Even Harder Case: Image Outputs

Hurricane Model:

Climate Parameters

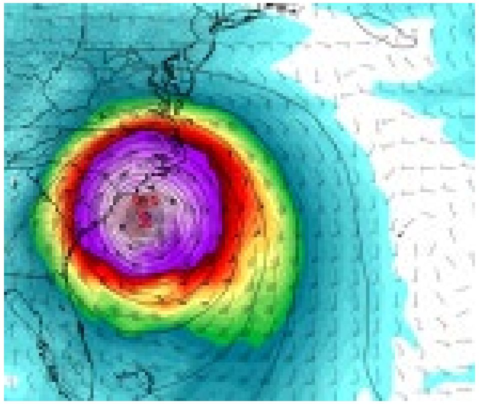


Image of Hurricane

- This type of sensitivity analysis is an open research question
  - How do we assess image variability?
  - Which metrics should we use to assess image similarity?
  - Should we consider adversaries?

Image from https://www.weatherboy.com/odds-hurricane-irmas-impacts-increase/

Conclusions

- Sensitivity analysis apportions the variation of a model response to variation of model inputs
- Sensitivity analysis is frequently used for dimension reduction in mathematical and statistical modeling
- For models with non-scalar outputs, sensitivity analysis is an ongoing area of research

Acknowledgements: The work in the materials science example was funded through the LLNL Laboratory Directed Research and Development (LDRD) Program, (17-ERD-029).

References

- K.B. Clark, M. Cortez, C. Hernandez, B.E. Thomas, and A.L. Lewis, “Combating tuberculosis: Using time-dependent sensitivity analysis to develop strategies for treatment and prevention,” *Spora: A Journal of Biomathematics*, *5*(1), pp.14-23, 2019.

- K. Schmidt, J. Bernstein, N. Barton, J. Florando, and A. Kupresanin, "Sensitivity analysis of strength models using Bayesian adaptive splines," *AIP Conference Proceedings*, 1979(1), pp. 140004, AIP Publishing LLC, 2018.

- R.C. Smith, *Uncertainty Quantification: Theory, Implementation, and Applications*, SIAM, Philadelphia, PA, 2014.

- M.T. Wentworth, *Verification Techniques for Parameter Selection and Bayesian Model Calibration Presented for an HIV Model*, 2015, url: https://repository.lib.ncsu.edu/bitstream/handle/1840.16/10229/etd.pdf

- M.T. Wentworth, R.C. Smith, and B. Williams, “Bayesian model calibration and uncertainty quantification for an HIV model using adaptive Metropolis algorithms,” *Inverse Problems in Science and Engineering*, 26(2), pp.233-256, 2018.