

An uncertainty quantification framework for a Zika virus epidemic model

Americo Cunha Jr

Rio de Janeiro State University – UERJ

NUMERICO – Nucleus of Modeling and Experimentation with Computers
<http://numerico.ime.uerj.br>

In collaboration with: Eber Dantas (UERJ)
Michel Tosin (UERJ)

CCIS 2019
March 19–22, 2019
Georgia Tech, Atlanta, USA



Outline

1 Introduction

2 Dynamic Model

3 Sensitivity Analysis

4 Uncertainty Quantification

5 Final Remarks

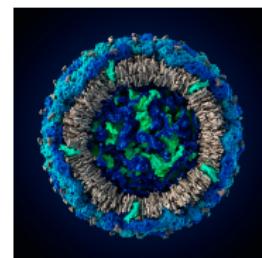


Section 1

Introduction

Zika virus (ZIKV)

- Member of *Flaviviridae* virus family
- First isolated in 1947 at Uganda, Africa
- Mainly spread by *Aedes* mosquitoes
- W.H.O declared it a public health emergency of international concern
- More than 140,000 confirmed cases in Brazil since 2015
- Associated diseases:
 - Congenital Zika syndrome
17,000 probable cases of infants
 - Guillain–Barré syndrome



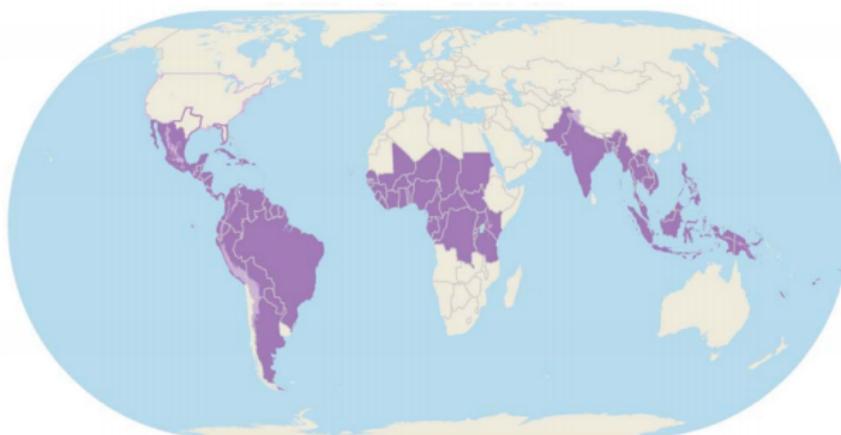
Zika virus



Aedes aegypti

Global outbreak of Zika virus

World Map of Areas with Risk of Zika



International areas and US territories

- [Dark Purple] Areas with risk of Zika infection (below 6,500 feet)
- [Medium Purple] Areas with low likelihood of Zika infection (above 6,500 feet)
- [Light Beige] Areas with no known risk of Zika infection

United States areas

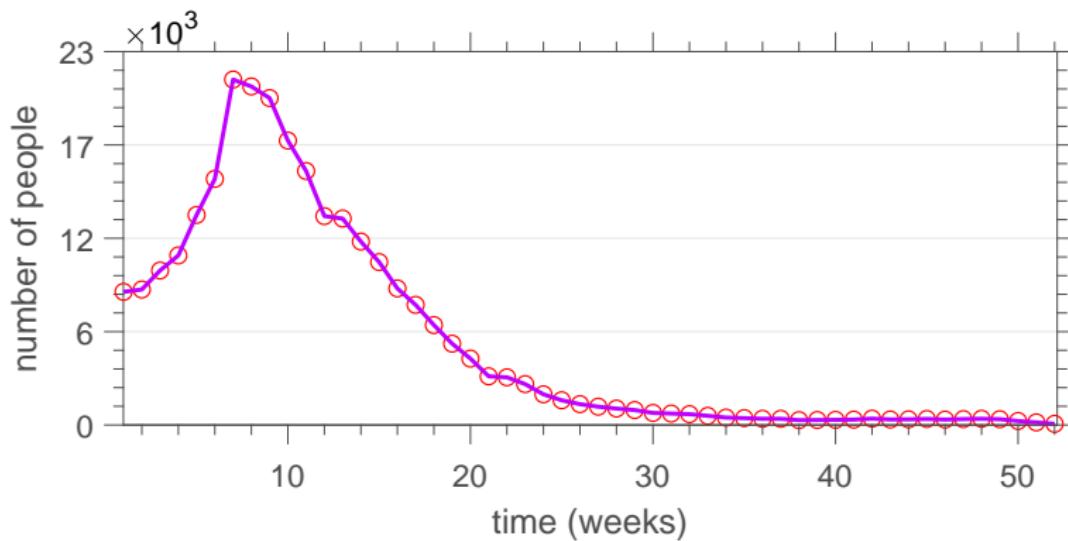
- [Pink] State Reporting Zika
- [Light Beige] No Known Zika



Centers for Disease Control and Prevention, *World Map of Areas with Risk of Zika, March 2018*.

Zika virus outbreak in Brazil

New cases in Brazil by epidemiological week of 2016



Ministério da Saúde. Obtenção de número de casos confirmados de zika, por município e semana epidemiológica. <https://bit.ly/20VgGGt>

Research objectives

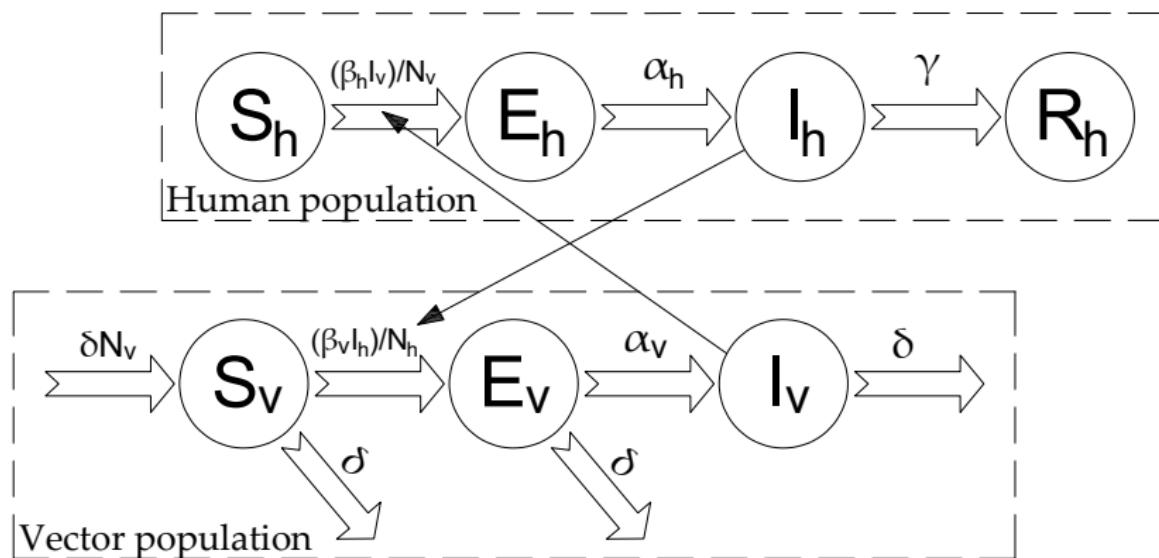
- Develop an epidemic model to describe the recent outbreak of Zika virus in Brazil
- Verify (qualitatively and quantitatively) the epidemic model capacity of prediction
- Calibrate this epidemic model with real data to obtain reliable predictions
- Construct a stochastic model to deal with data uncertainties and made more robust predictions



Section 2

Dynamic Model

SEIR-SEI model for Zika virus dynamics



A. J. Kucharski et al. *Transmission Dynamics of Zika Virus in Island Populations: A Modelling Analysis of the 2013–14 French Polynesia Outbreak*. PLOS Neglected Tropical Diseases, 2016.

Associated dynamical system

$$\frac{dS_h}{dt} = -\beta_h S_h \frac{I_v}{N_v}$$

$$\frac{dS_v}{dt} = \delta - \beta_v S_v \frac{I_h}{N_h} - \delta S_v$$

$$\frac{dE_h}{dt} = \beta_h S_h \frac{I_v}{N_v} - \alpha_h E_h$$

$$\frac{dE_v}{dt} = \beta_v S_v \frac{I_h}{N_h} - (\delta + \alpha_v) E_v$$

$$\frac{dI_h}{dt} = \alpha_h E_h - \gamma I_h$$

$$\frac{dI_v}{dt} = \alpha_v E_v - \delta I_v$$

$$\frac{dR_h}{dt} = \gamma I_h$$

$$\frac{dC}{dt} = \alpha_h E_h$$

+ initial conditions

S - Population of susceptible

C - Infected humans cumulative

h - Human-related

E - Population of exposed

α - Incubation ratio

v - Vector-related

I - Population of infected

δ - Vector lifespan ratio

R - Population of recovered

β - Transmission rate

N - Total population

γ - Recovery rate



A. J. Kucharski et al. *Transmission Dynamics of Zika Virus in Island Populations: A Modelling*

Analysis of the 2013–14 French Polynesia Outbreak. PLOS Neglected Tropical Diseases, 2016.

Model parameters and outbreak data

- open scientific literature



- Brazilian health system

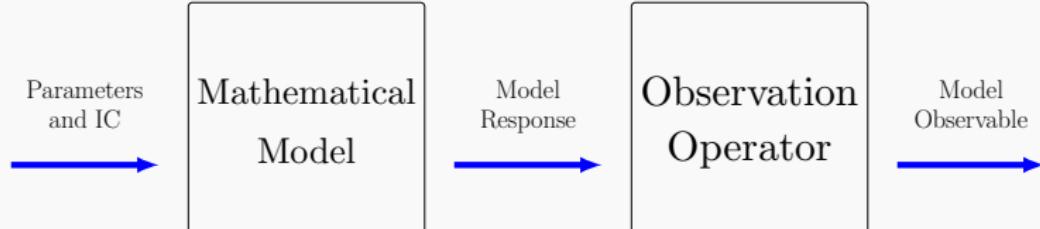


FIOCRUZ
Fundação Oswaldo Cruz

parameter	value	unit
α_h	1/5.9	days ⁻¹
α_v	1/9.1	days ⁻¹
γ	1/7.9	days ⁻¹
δ	1/11	days ⁻¹
β_h	1/11.3	days ⁻¹
β_v	1/8.6	days ⁻¹
N	206×10^6	people



Quantities of interest (QoI)



QoI 1: cumulative number of infectious

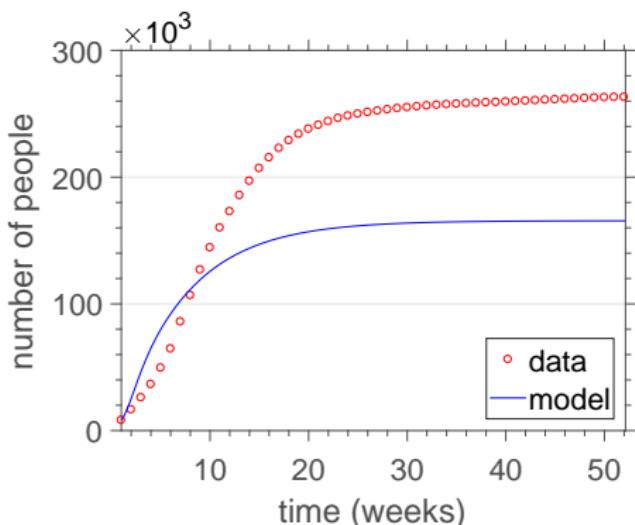
$$C_t = \int_{\tau=0}^t \alpha_h E_h(\tau) d\tau$$

QoI 2: new infectious cases

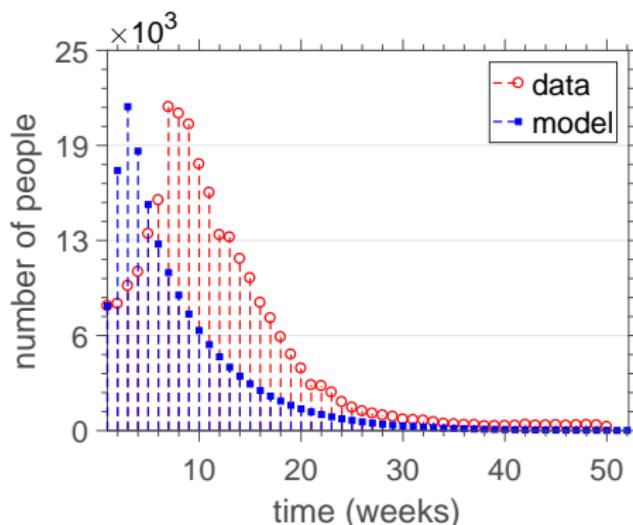
$$\mathcal{N}_w = C_w - C_{w-1}, \quad (w = 2, 3, \dots, 52)$$

$$\mathcal{N}_1 = C_1$$

Time series for QoL's



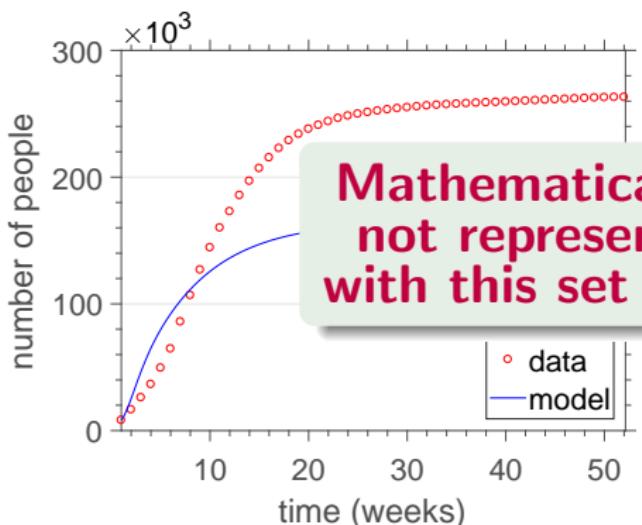
cumulative number of infectious



new infectious cases

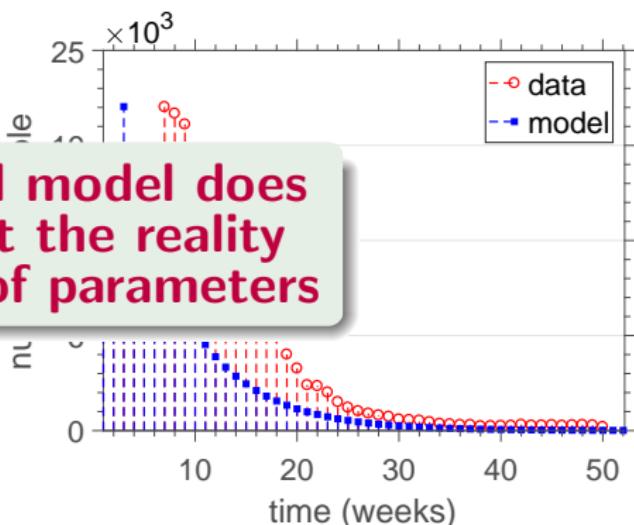


Time series for QoL's



Mathematical model does
not represent the reality
with this set of parameters

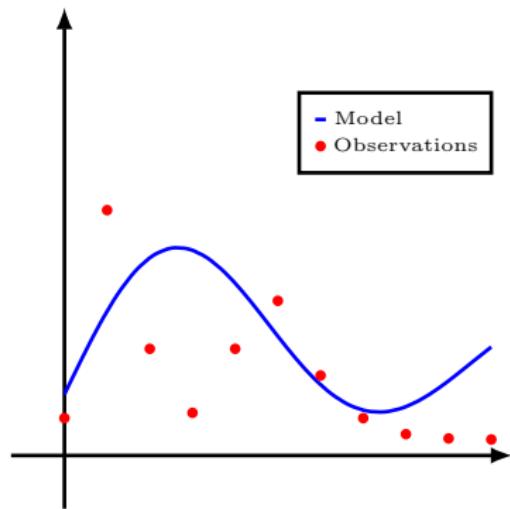
cumulative number of infectious



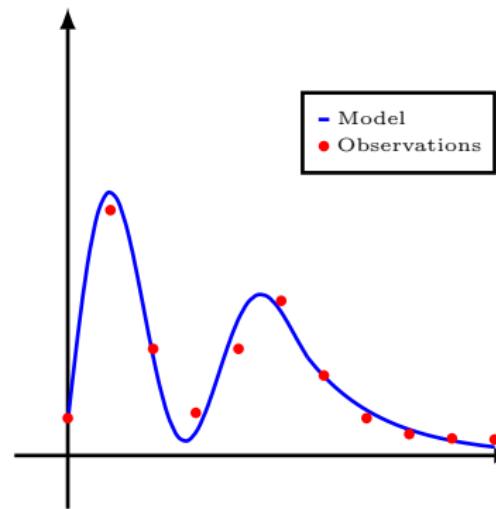
new infectious cases

Calibration of the model

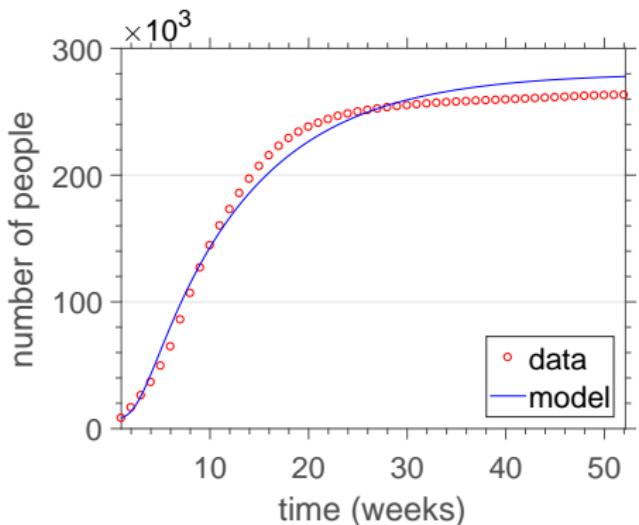
Uncalibrated Model



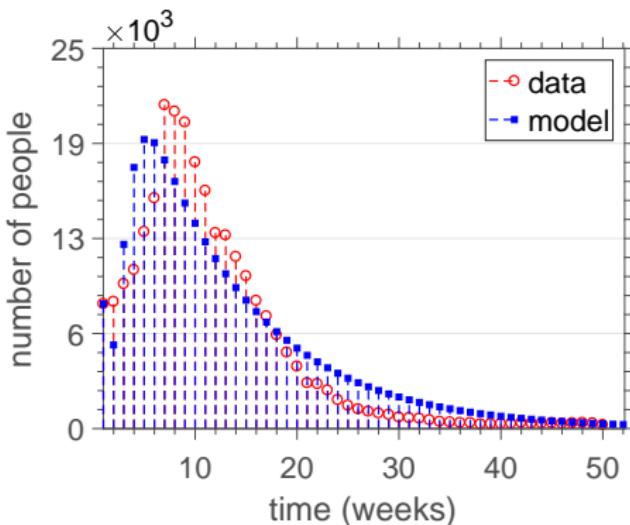
Calibrated Model



Calibrated model response

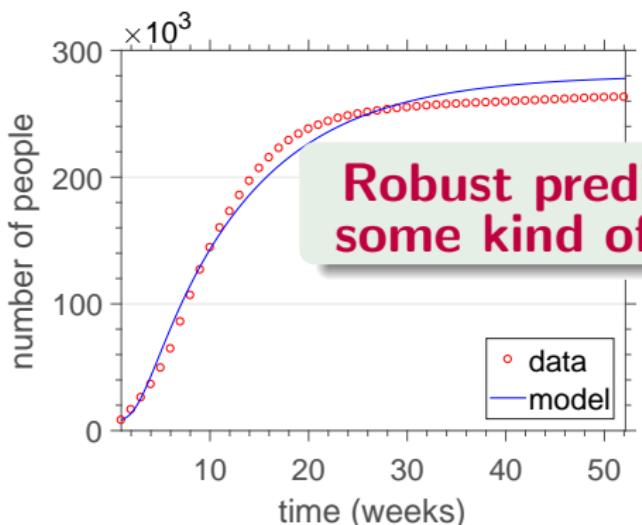


cumulative number of infectious

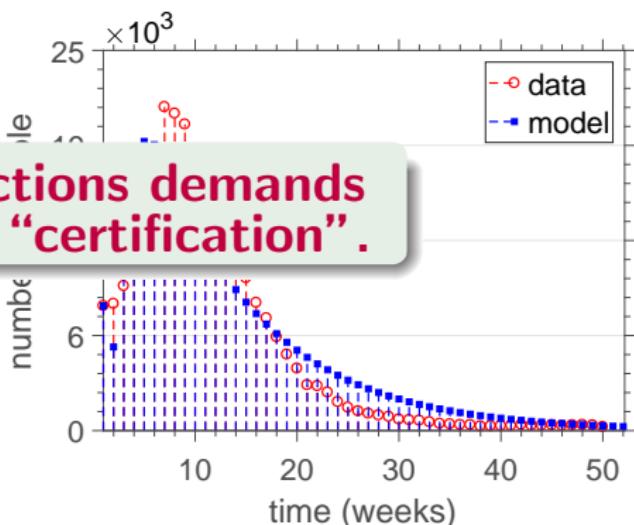


new infectious cases

Calibrated model response



cumulative number of infectious



new infectious cases

Section 3

Sensitivity Analysis

Variance-based sensitivity analysis

Mathematical model:

$$Y = \mathcal{M}(\mathbf{X}), \quad X_i \sim \mathcal{U}(0, 1), \quad (\text{i.i.d.})$$

Hoeffding-Sobol' decomposition:

$$Y = \mathcal{M}_0 + \sum_{1 \leq i \leq n} \mathcal{M}_i(X_i) + \sum_{1 \leq i < j \leq n} \mathcal{M}_{ij}(X_i, X_j) + \cdots + \mathcal{M}_{1\dots n}(X_1 \cdots X_n)$$

An **orthogonal decomposition** in terms of conditional expectations:

- $\mathcal{M}_0 = \mathbb{E}\{Y\}$
- $\mathcal{M}_i(X_i) = \mathbb{E}\{Y|X_i\} - \mathcal{M}_0$
- $\mathcal{M}_{ij}(X_i, X_j) = \mathbb{E}\{Y|X_i, X_j\} - \mathcal{M}_0 - \mathcal{M}_i - \mathcal{M}_j$
- etc

Sobol' indices

Total variance:

$$D = \text{Var} [\mathcal{M}(\mathbf{X})] = \sum_{\mathbf{u} \subset \{1, \dots, k\}} \text{Var} [\mathcal{M}_{\mathbf{u}}(\mathbf{X}_u)]$$

First order Sobol' indices:

$$S_i = \text{Var} [\mathcal{M}_i(X_i)] / D$$

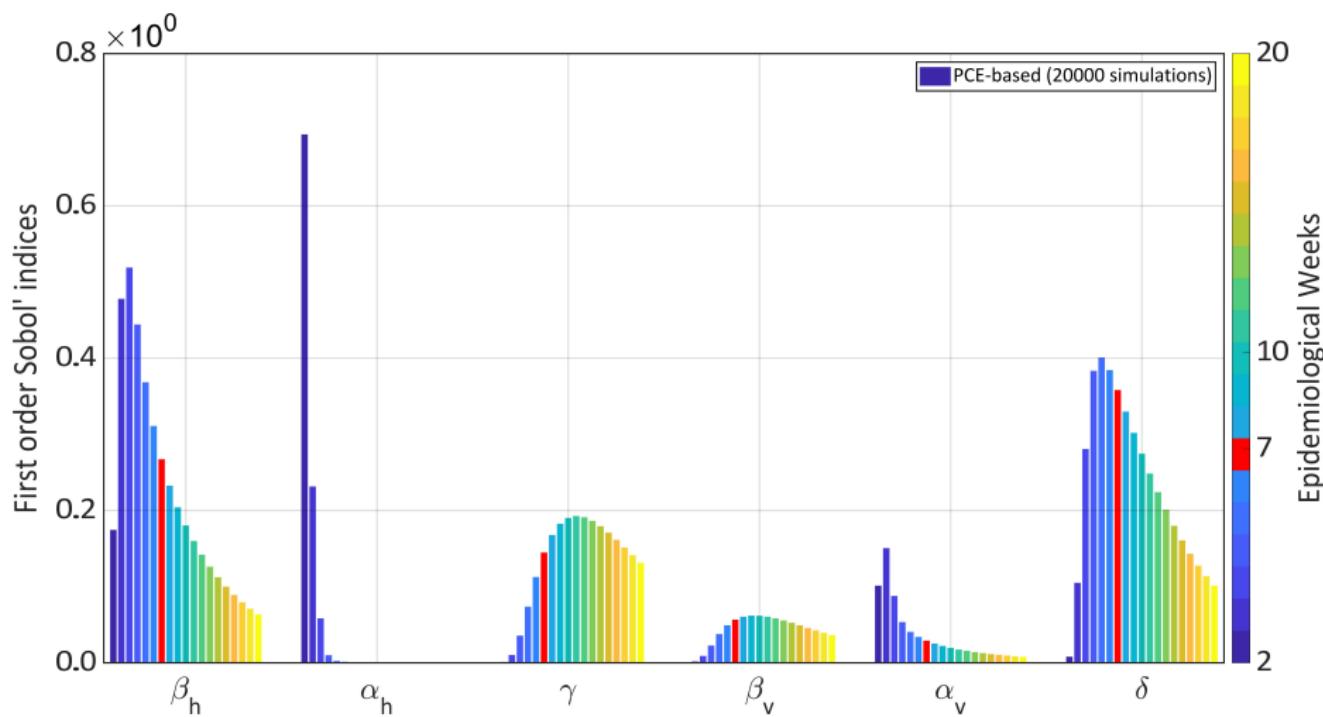
(quantify the additive effect of each input separately)

Second order Sobol' indices:

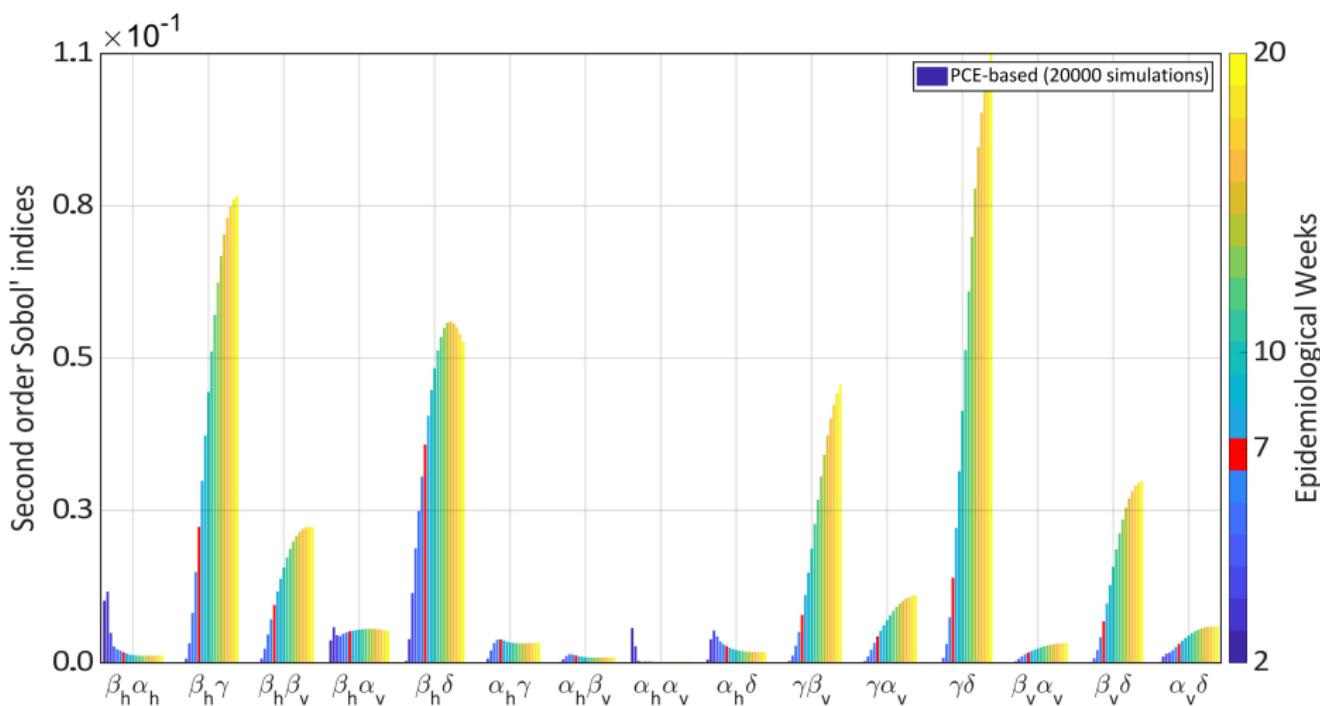
$$S_{ij} = \text{Var} [\mathcal{M}_{ij}(X_i, X_j)] / D$$

(quantify interaction effect of inputs X_i and X_j)

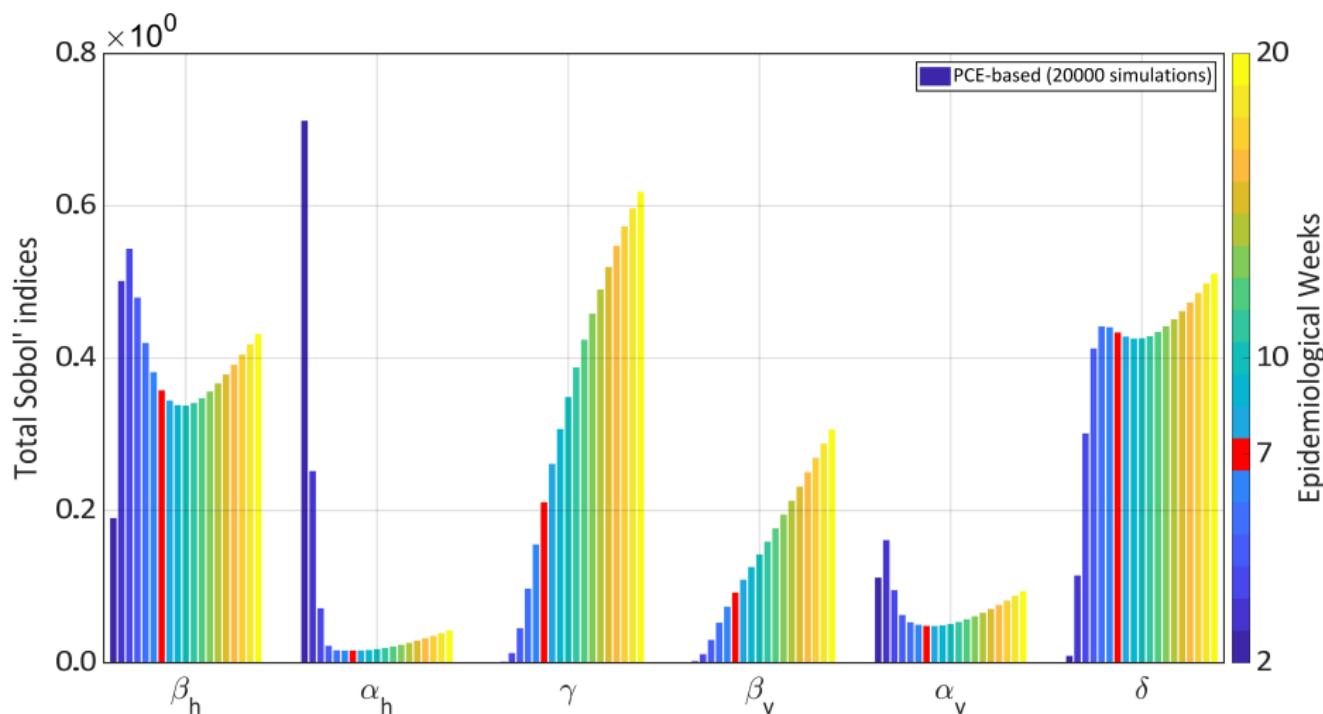
Global sensitivity analysis: first order



Global sensitivity analysis: second order



Global sensitivity analysis: total order



Global sensitivity analysis: general overview

- Two most relevant: δ and β_H (75% variance around 7th EW)
- Third most, γ , mainly by nonlinear interactions with δ and β_H
- Parameters limited to nonlinear interactions have, in general, delayed effects (significant for $EW > 15$)
- (*sparsity-of-effects principle*) Higher order interactions have minor effect: 1st and 2nd are 99.8–96.7% variance on 5–10th EW

Around 7th EW → uncertainty propagation of $\{\beta_h, \delta\}$

Section 4

Uncertainty Quantification

Uncertainty Quantification (UQ) framework

Mathematical model:

$$Y = \mathcal{M}(\mathbf{X})$$

General steps for UQ:

- ① Stochastic modeling
→ characterization of inputs uncertainties
(MaxEnt Principle)
- ② Uncertainty propagation
→ characterization of output uncertainties
(Monte Carlo Method)
- ③ Response certification
→ specification of reliability levels for predictions
(Nonparametric Statistical Inference)

Probabilistic model 1

Random variables: β_h and δ

Available information: support and mean (nominal) value

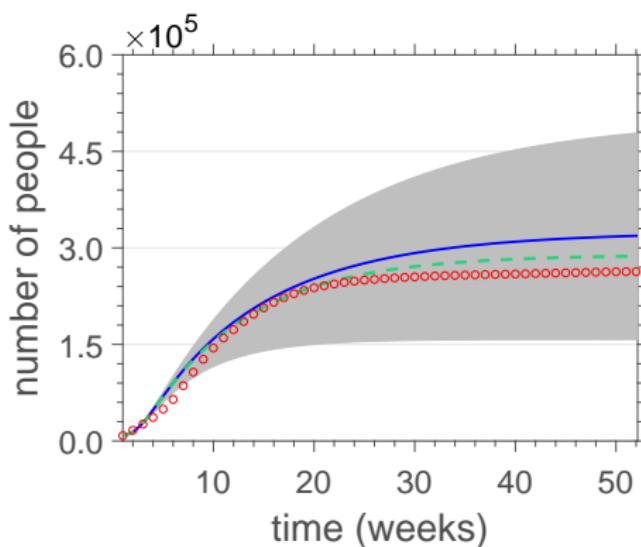
MaxEnt distribution

$$p_X(x) = \mathbb{1}_{[a,b]}(x) \exp(-\lambda_0 - \lambda_1 x)$$

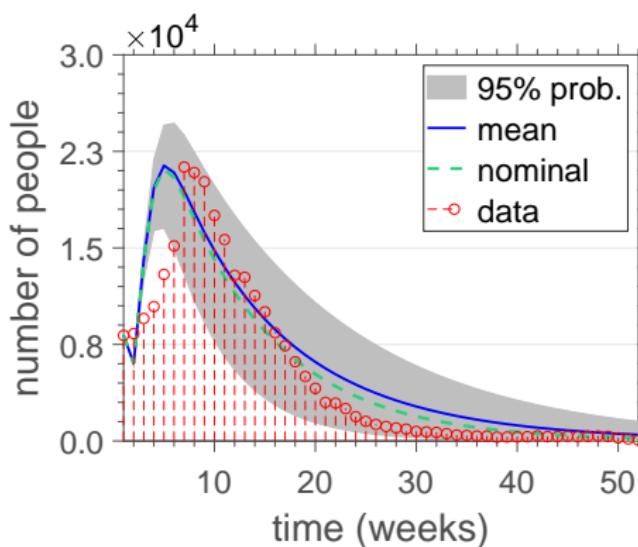
“truncated exponential (2 parameters)”



Confidence band for the Qols



cumulative number of infectious



new infectious cases

Probabilistic model 2

Random variables: β_h and δ

Available information: support, mean (nominal) value and dispersion

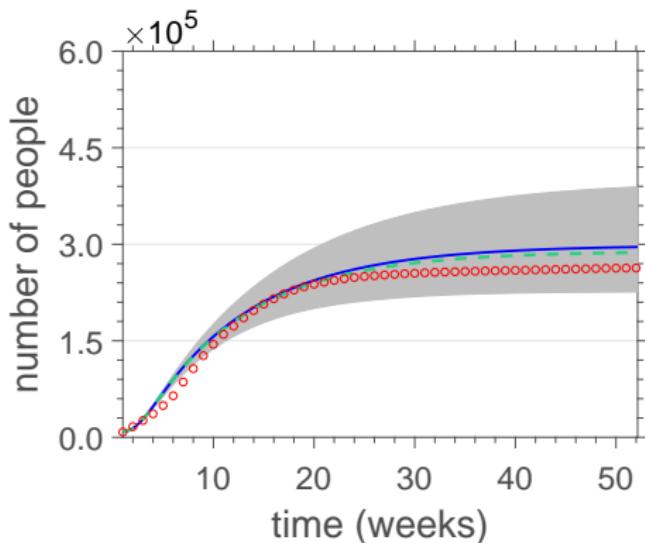
MaxEnt distribution

$$p_X(x) = \mathbb{1}_{[a,b]}(x) \exp\left(-\lambda_0 - \lambda_1 x - \lambda_2 x^2\right)$$

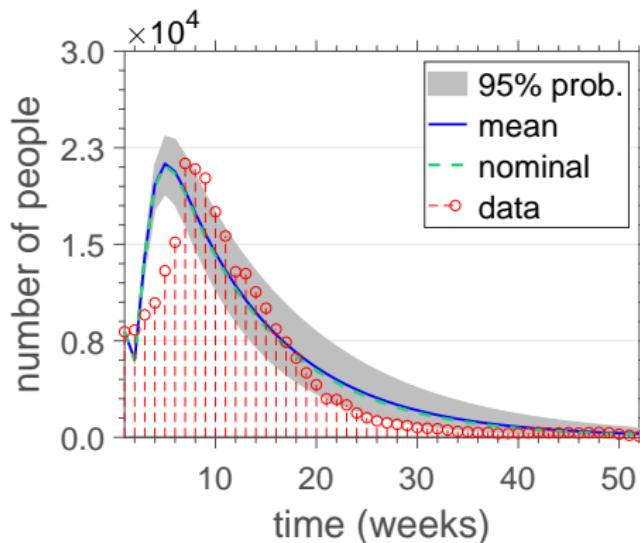
“truncated exponential (3 parameters)”

Confidence band for the Qols

β_h dispersion = 5% , δ dispersion = 5%



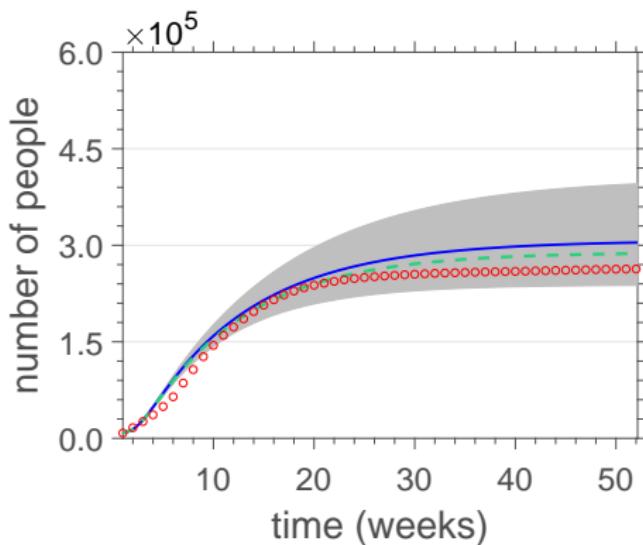
cumulative number of infectious



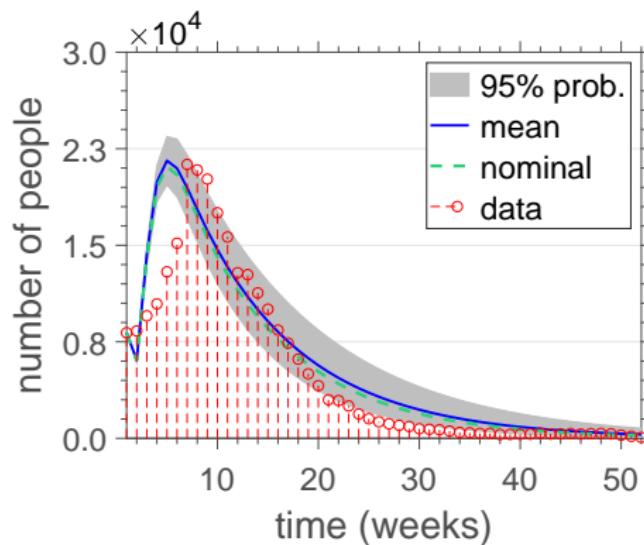
new infectious cases

Confidence band for the Qols

β_h dispersion = 10% , δ dispersion = 5%



cumulative number of infectious

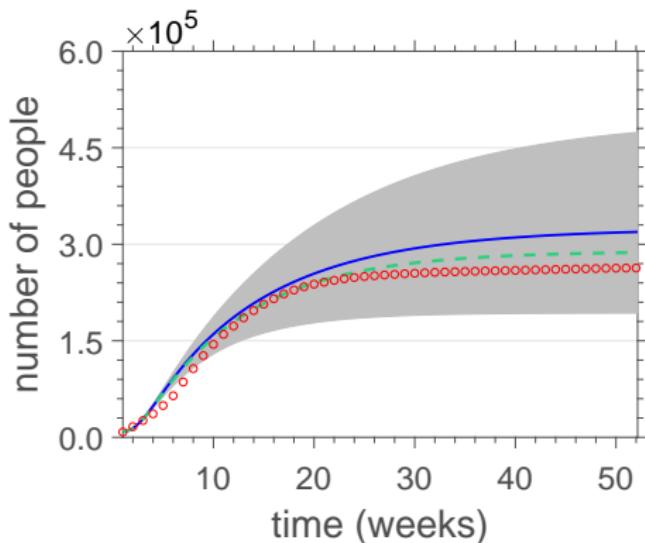


new infectious cases

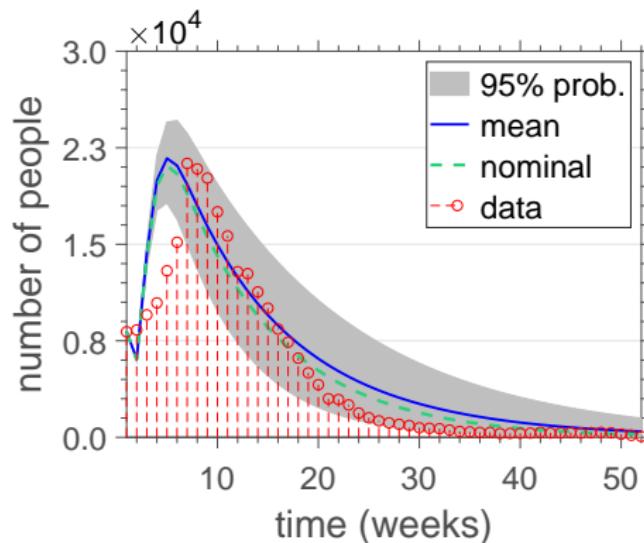


Confidence band for the Qols

β_h dispersion = 10% , δ dispersion = 10%



cumulative number of infectious



new infectious cases



Probabilistic model 3

Random variables: β_h , δ and σ

Available information for β_h and δ : support, mean (nominal) value

Distribution for β_h and β_v

$$p_X(x) = \mathbb{1}_{[a,b]}(x) \exp\left(-\lambda_0 - \lambda_1 x - \lambda_2 x^2\right)$$

Available information for σ : support

MaxEnt distribution for σ

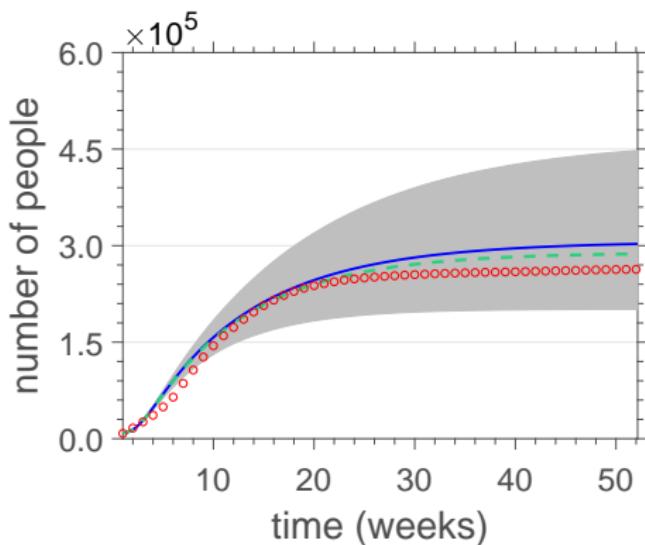
$$p_X(x) = \mathbb{1}_{[a,b]}(x) \frac{1}{b-a}$$

“uniform”

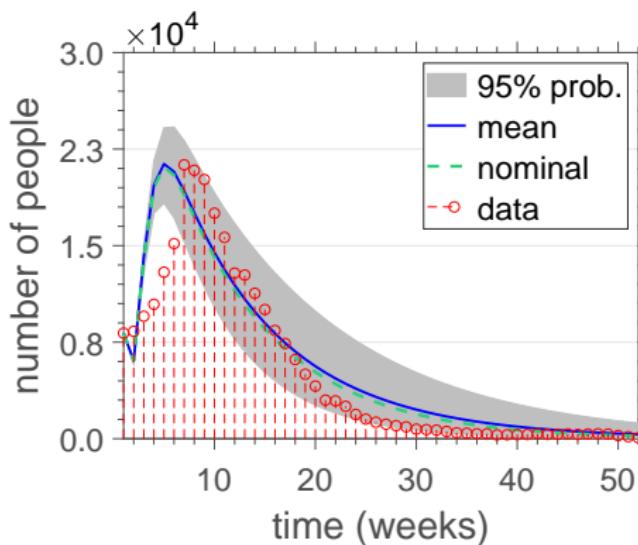


Confidence band for the Qols

random dispersion $\sim U(5\%, 10\%)$

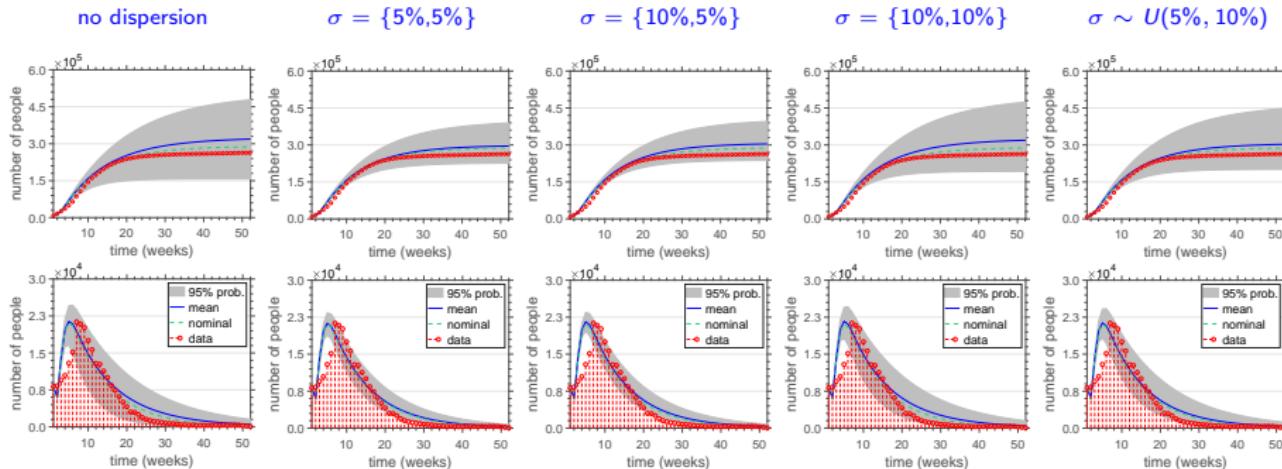


cumulative number of infectious



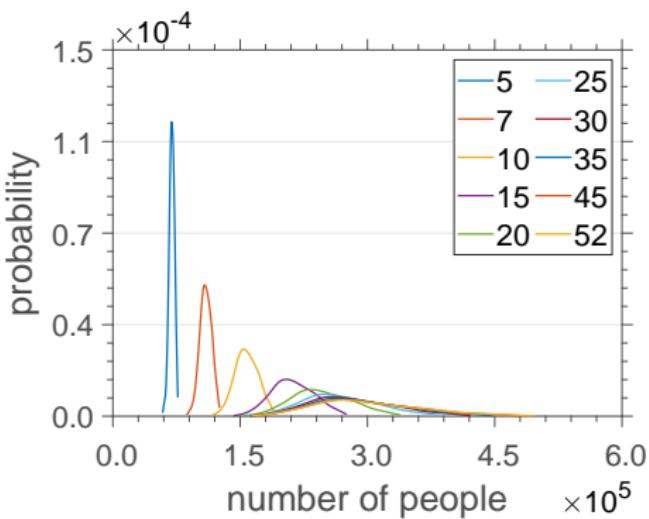
new infectious cases

Confidence band for the Qols

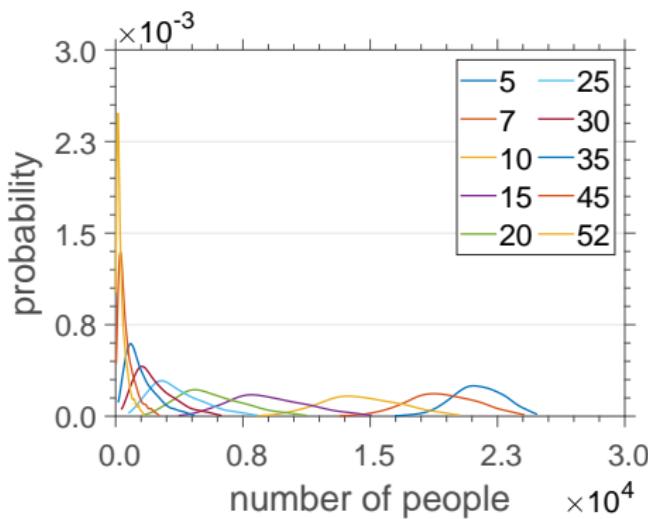


Evolution of Qols PDFs

random dispersion $\sim U(5\%, 10\%)$

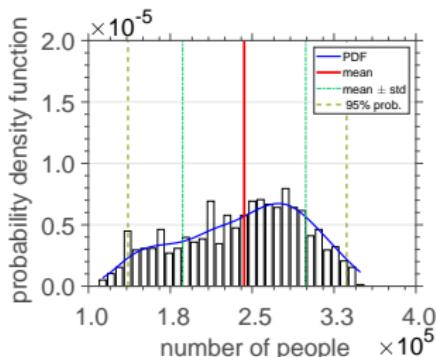


cumulative number of infectious

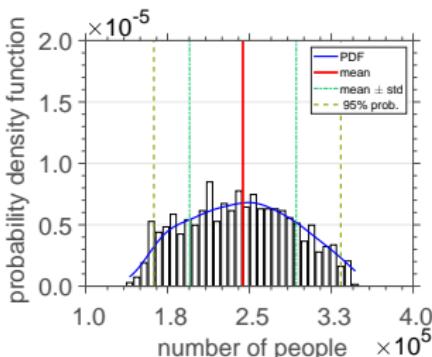


new infectious cases

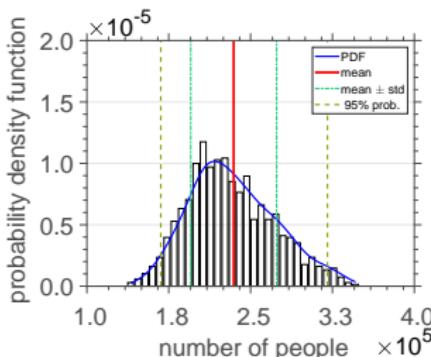
Time-averaged cumulative infectious



no dispersion

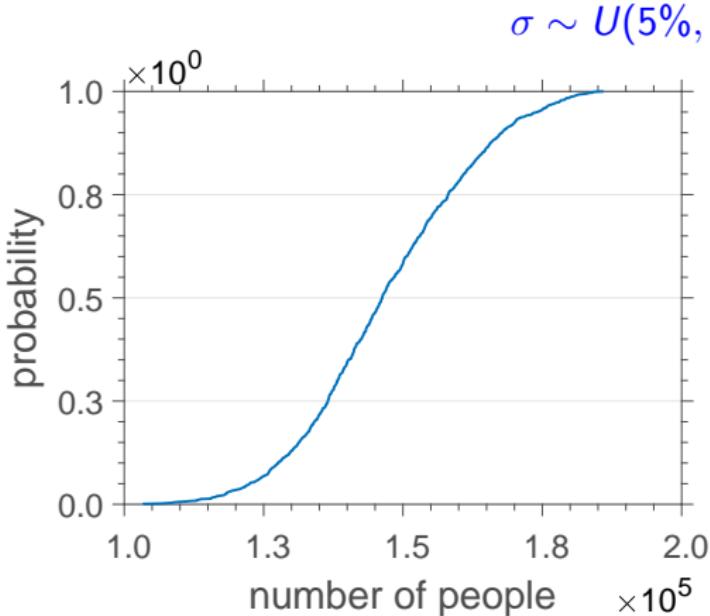


$$\sigma = \{10\%, 10\%\}$$



$$\sigma \sim U(5\%, 10\%)$$

(mean) Cumulative infectious CDF until EW 20

Statistics of C

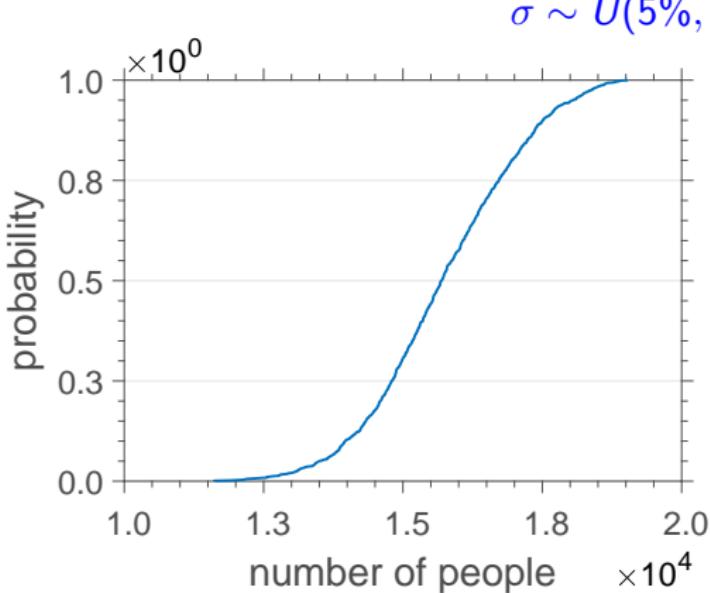
mean	=	$1,47 \times 10^5$
std. dev.	=	$1,53 \times 10^4$
skewness	=	0.084
kurtosis	=	2.605
$P(C \geq c^*)$	=	87.10%

$c^* = 130,000$

Half the maximum C (data)



(mean) New cases CDF until 10th EW



Statistics of \mathcal{N}_w

mean	=	$1,57 \times 10^4$
std. dev.	=	$1,35 \times 10^3$
skewness	=	-0.032
kurtosis	=	2.656
$P(\mathcal{N}_w \geq NC^*)$	=	83.40%

$NC^* = 14,440$
average NC (data) until EW 10

Section 5

Final Remarks

Concluding remarks

Contributions:

- Development of an epidemic model to describe Brazilian outbreak of Zika virus
- Calibration of this model with real epidemic data
- Construction of parametric probabilistic model of uncertainties

Ongoing research:

- Bayesian updating to improve the model calibration
- Quantify model discrepancy in a nonparametric way

Future directions:

- Investigate the effectiveness of different control strategies
- Scenarios exploration with active subspace method
- Data-driven identification of epidemiological models

Acknowledgments

Financial support:



Thank you for your attention!

americo@ime.uerj.br

www.americocunha.org



E. Dantas, M. Tosin and A. Cunha Jr,

Calibration of a SEIR–SEI epidemic model to describe Zika virus outbreak in Brazil,
Applied Mathematics and Computation, 338: 249–259, 2018.

<https://doi.org/10.1016/j.amc.2018.06.024>



E. Dantas, M. Tosin and A. Cunha Jr,

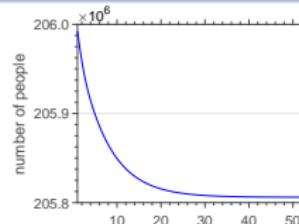
Uncertainty quantification in the nonlinear dynamics of Zika virus, 2018
(in preparation).

nominal parameters

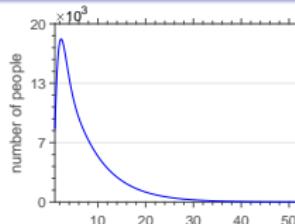
Nominal parameters and initial conditions

α	value	unit
α_h	$1/5.9$	days^{-1}
α_v	$1/9.1$	days^{-1}
γ	$1/7.9$	days^{-1}
δ	$1/11$	days^{-1}
β_h	$1/11.3$	days^{-1}
β_v	$1/8.6$	days^{-1}
N	206×10^6	people
S_h^i	205,953,959	people
E_h^i	8,201	people
I_h^i	8,201	people
R_h^i	29,639	people
S_v^i	0.99956	—
E_v^i	2.2×10^{-4}	—
I_v^i	2.2×10^{-4}	—

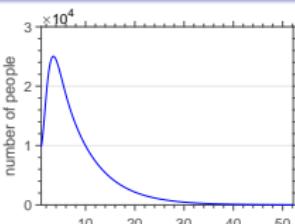
Model response with nominal parameters



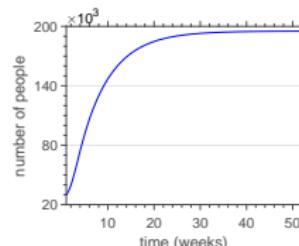
Susceptible humans



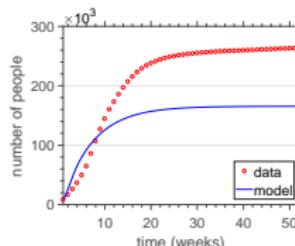
Exposed humans



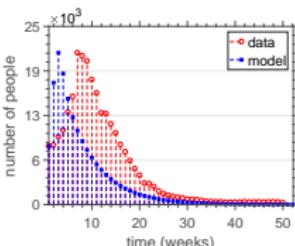
Infectious humans



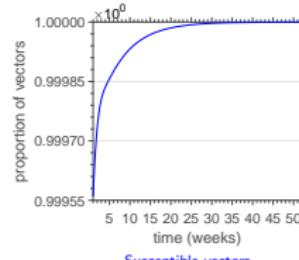
Recovered humans



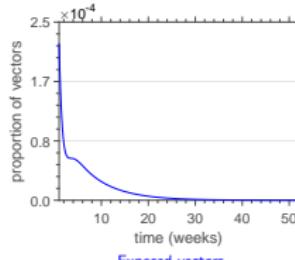
Cumulative infectious



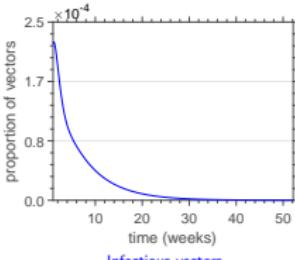
New cases



Susceptible vectors



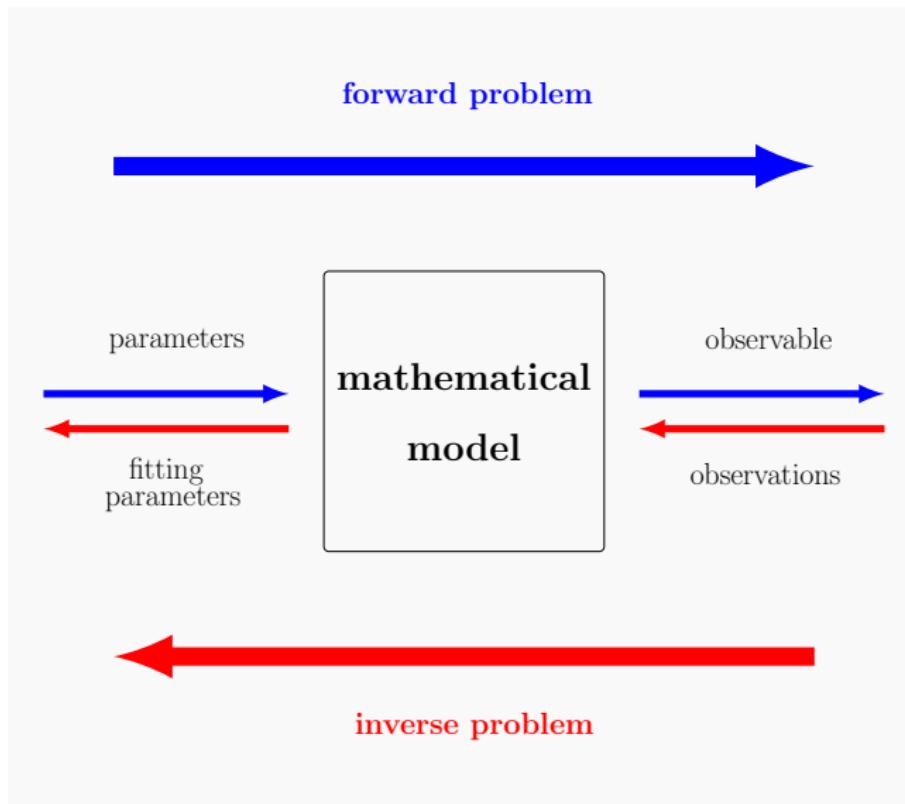
Exposed vectors



Infectious vectors

calibration

Forward and inverse problem



Inverse problem formulation

- data space: $F = \mathbb{R}^M$
- parameter space: $C = \left\{ \boldsymbol{\alpha} \in \mathbb{R}^{12} \mid \boldsymbol{\alpha}_{min} \leq \boldsymbol{\alpha} \leq \boldsymbol{\alpha}_{max} \right\}$
- observation vector: $\mathbf{y} = (y_1, y_2, \dots, y_M) \in F$
- prediction vector: $\phi(\boldsymbol{\alpha}) = (\phi_1, \phi_2, \dots, \phi_M) \in F$
- misfit function:

$$J(\boldsymbol{\alpha}) = \|\mathbf{y} - \phi(\boldsymbol{\alpha})\|_F^2 = \sum_{m=1}^M |y_m - \phi_m(\boldsymbol{\alpha})|^2$$

Find a **vector of parameters** such that

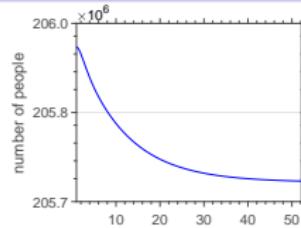
$$\boldsymbol{\alpha}^* = \arg \min_{\boldsymbol{\alpha} \in C} J(\boldsymbol{\alpha}).$$

- ⇒ Q-wellposed: existence, uniqueness, unimodality and local stability
- ⇒ Solution algorithm: bounded trust-region-reflective

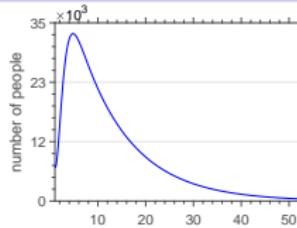
Calibration parameters and initial conditions

α	TRR input	lb	ub	TRR output
α_h	1/5.9	1/12	1/3	1/12
α_v	1/9.1	1/10	1/5	1/10
γ	1/7.9	1/8.8	1/3	1/3
δ	1/11	1/21	1/11	1/21
β_h	1/11.3	1/16.3	1/8	1/10.40
β_v	1/8.6	1/11.6	1/6.2	1/7.77
S_h^i	205,953,959	$0.9 \times N$	N	205,953,534
E_h^i	8,201	0	10,000	6,827
I_h^i	8,201	0	10,000	10,000
S_v^i	0.9996	0.99	0.999	0.999
E_v^i	2.2×10^{-4}	0	1	4.14×10^{-4}
I_v^i	2.2×10^{-4}	0	1	0

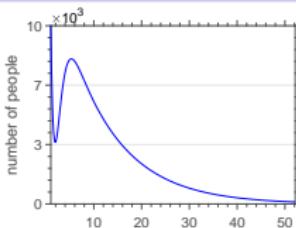
Model response for calibration



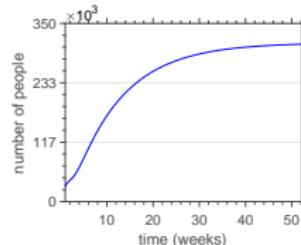
Susceptible humans



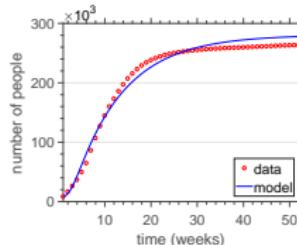
Exposed humans



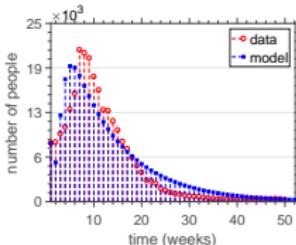
Infectious humans



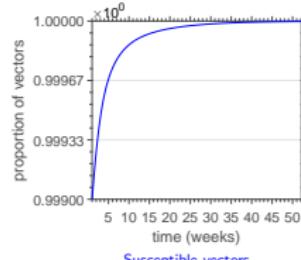
Recovered humans



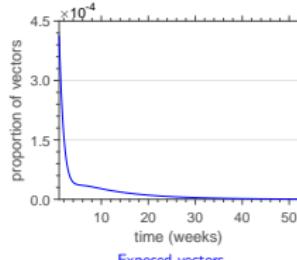
Cumulative infectious



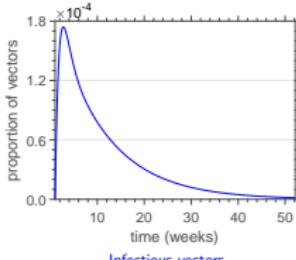
New cases



Susceptible vectors



Exposed vectors



Infectious vectors

sensitivity analysis

Metamodelling via Polynomial Chaos

Assuming $Y = \mathcal{M}(\mathbf{X})$ has finite variance, then it admits a
Polynomial Chaos expansion

$$Y = \sum_{\alpha \in \mathbb{N}^k} y_\alpha \Phi_\alpha$$

where

- Φ_α : multivariate orthonormal polynomials
- y_α : real-valued coefficients to be determined



D. Xiu, and G. Karniadakis, *The Wiener-Askey Polynomial Chaos for Stochastic Differential Equations*. SIAM Journal on Scientific Computing, 24: 619-644, 2002.



PC-based Sobol' indices

For computational purposes, a truncated PCE is employed

$$Y \approx \sum_{\alpha \in \mathcal{A}} y_\alpha \Phi_\alpha$$

Thus, Sobol' indices are given by

$$S_{\mathbf{u}} = D_{\mathbf{u}}/D = \frac{\sum_{\alpha \in \mathcal{A}_{\mathbf{u}}} y_\alpha^2}{\sum_{\alpha \in \mathcal{A} \setminus 0} y_\alpha^2}$$

$$\mathcal{A}_{\mathbf{u}} = \{\alpha \in \mathcal{A} : i \in \mathbf{u} \iff \alpha_i \neq 0\}$$

Sobol' indices of any order can be obtained, analytically, from the coefficients of the PC expansion!



B. Sudret, *Global sensitivity analysis using polynomial chaos expansions. Reliability Engineering &*

System Safety, 2016, 93(7): 964–979, 2008.

uncertainty quantification

Maximum Entropy Principle (MaxEnt)

Among all the probability distributions, consistent with the known information about a random parameter, choose the one which corresponds to the maximum of entropy (MaxEnt).

MaxEnt distribution = most unbiased distribution

Entropy of the random variable X is defined as

$$\mathcal{S}(p_X) = - \int_{\mathbb{R}} p_X(x) \ln(p_X(x)) dx,$$

“measure for the level of uncertainty”

MaxEnt optimization problem

Maximize

$$\mathcal{S}(p_X) = - \int_{\mathbb{R}} p_X(x) \ln(p_X(x)) dx,$$

respecting $N + 1$ constraints (known information) given by

$$\int_{\mathbb{R}} g_k(X) p_X(x) dx = m_k, \quad k = 0, \dots, N,$$

where the g_k are known real functions, with $g_0(x) = 1$.

MaxEnt optimization problem

Maximize

$$\mathcal{S}(p_X) = - \int_{\mathbb{R}} p_X(x) \ln(p_X(x)) dx,$$

respecting $N + 1$ constraints (known information) given by

$$\int_{\mathbb{R}} g_k(X) p_X(x) dx = m_k, \quad k = 0, \dots, N,$$

where the g_k are known real functions, with $g_0(x) = 1$.

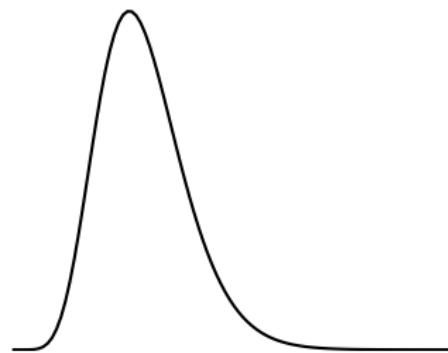
MaxEnt general solution

$$p_X(x) = \mathbb{1}_{\mathcal{K}}(x) \exp(-\lambda_0) \exp\left(-\sum_{k=1}^N \lambda_k g_k(x)\right)$$

Philosophy of MaxEnt Principle

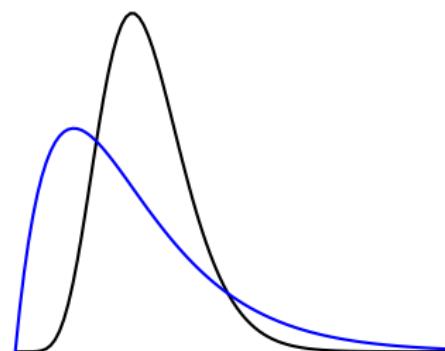
— real

- The parameter of interest has a unknown distribution



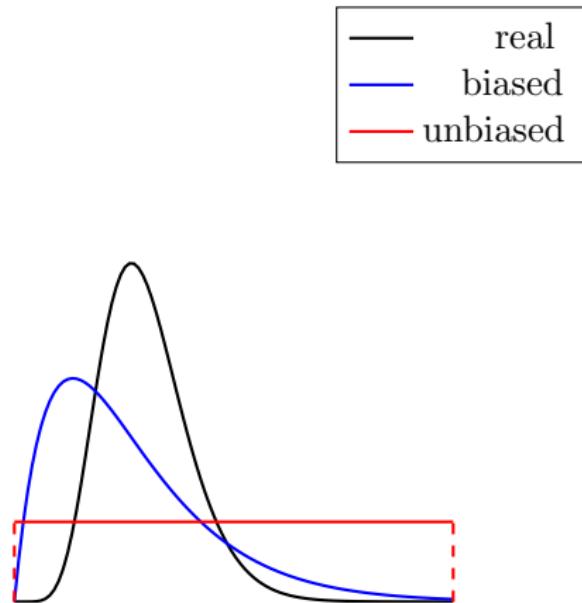
Philosophy of MaxEnt Principle

- The parameter of interest has a unknown distribution
- Distributions arbitrarily chosen can be coarse and biased



Philosophy of MaxEnt Principle

- The parameter of interest has a unknown distribution
- Distributions arbitrarily chosen can be coarse and biased
- A conservative strategy is to use the most unbiased (MaxEnt) distribution



Uncertainty Propagation

Monte Carlo Method

pre-processing processing post-processing

generation
of scenarios

\boldsymbol{X}_1

⋮

\boldsymbol{X}_M

known $F_{\boldsymbol{X}}$

solution of
model equations

$$\boldsymbol{U} = h(\boldsymbol{X})$$

computational
model

computation
of statistics

$$\boldsymbol{U}_1 = h(\boldsymbol{X}_1)$$

⋮

$$\boldsymbol{U}_M = h(\boldsymbol{X}_M)$$

estimated $F_{\boldsymbol{U}}$

generator of
random vector \boldsymbol{X}

deterministic solver
of $\boldsymbol{u} = h(\boldsymbol{x})$

statistical inference
to estimate convergence
and distribution of \boldsymbol{U}

Monte Carlo convergence

Study of convergence for MC simulation

Stochastic dynamic model:

$$\dot{\boldsymbol{U}}(t, \omega) = f(\boldsymbol{U}(\omega, t))$$

Convergence metric for Monte Carlo simulation:

$$\text{conv}(n_s) = \left(\frac{1}{n_s} \sum_{n=1}^{n_s} \int_{t_0}^{t_f} \| \boldsymbol{U}(t, \omega_n) \|^2 dt \right)^{1/2}$$



C. Soize, A comprehensive overview of a non-parametric probabilistic approach of model uncertainties for predictive models in structural dynamics. *Journal of Sound and Vibration*, 288: 623–652, 2005.

Study of convergence for MC simulation

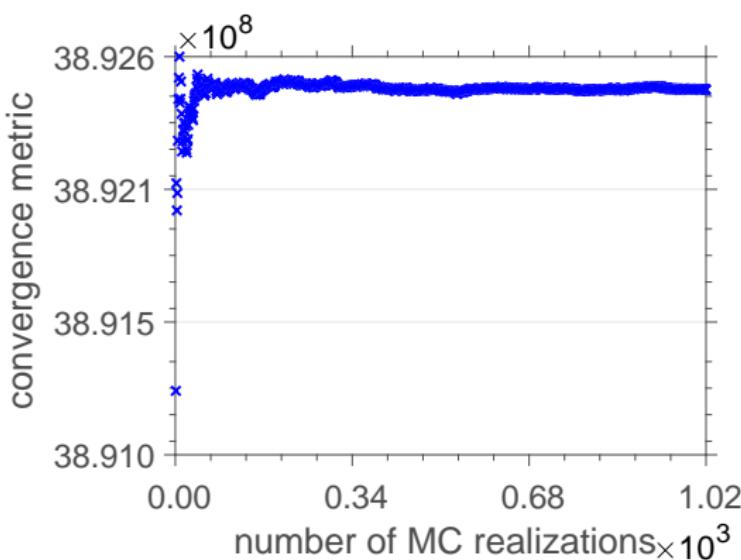
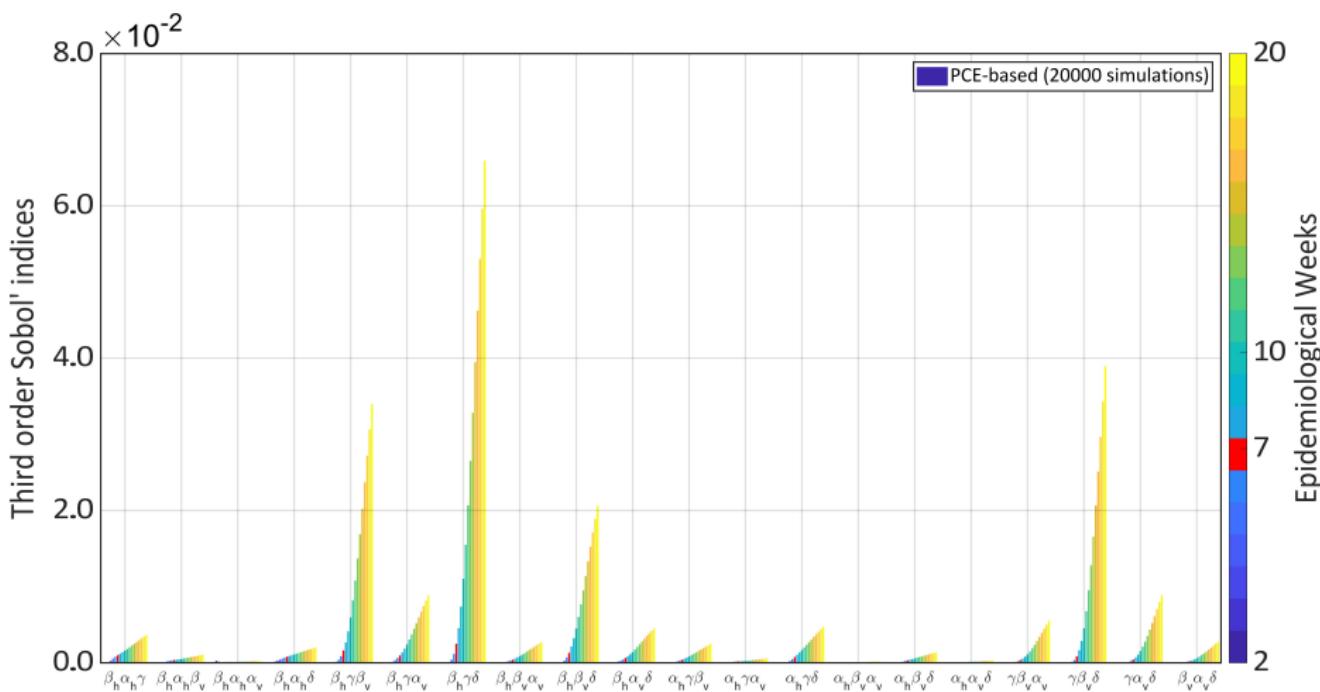


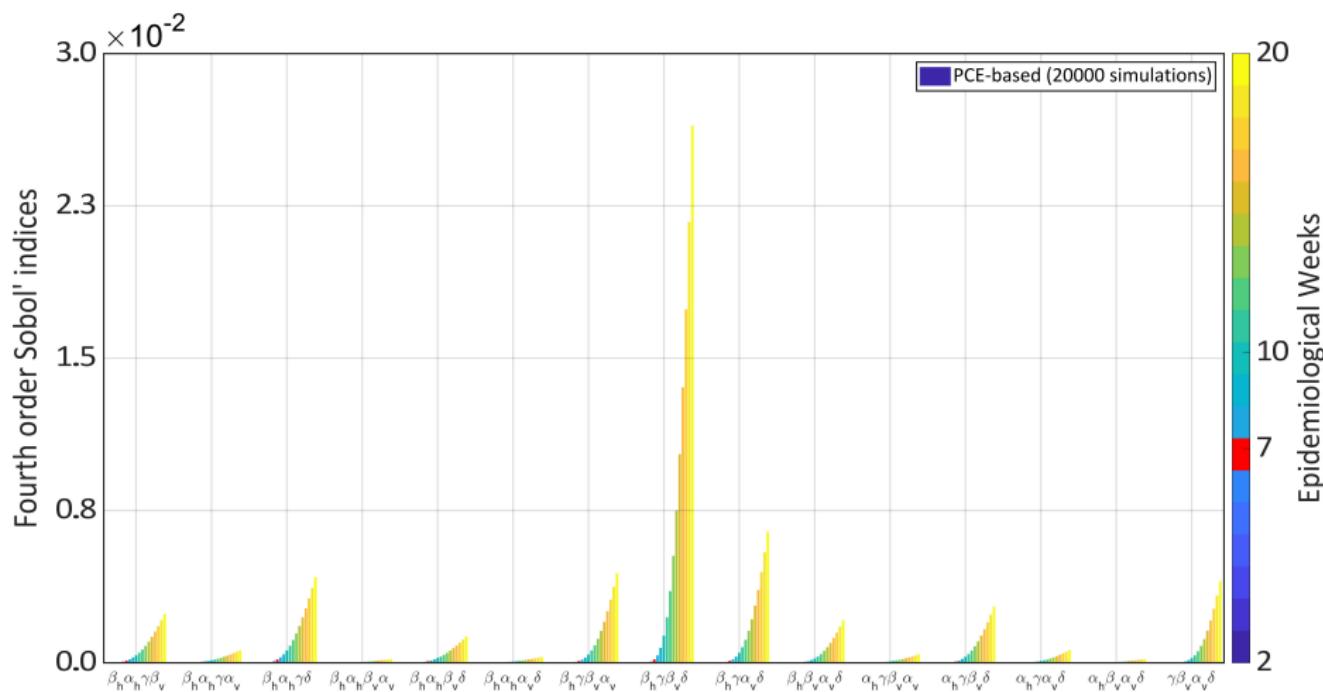
Figure: MC convergence metric as function of the number of realizations.

Sobol' Indices

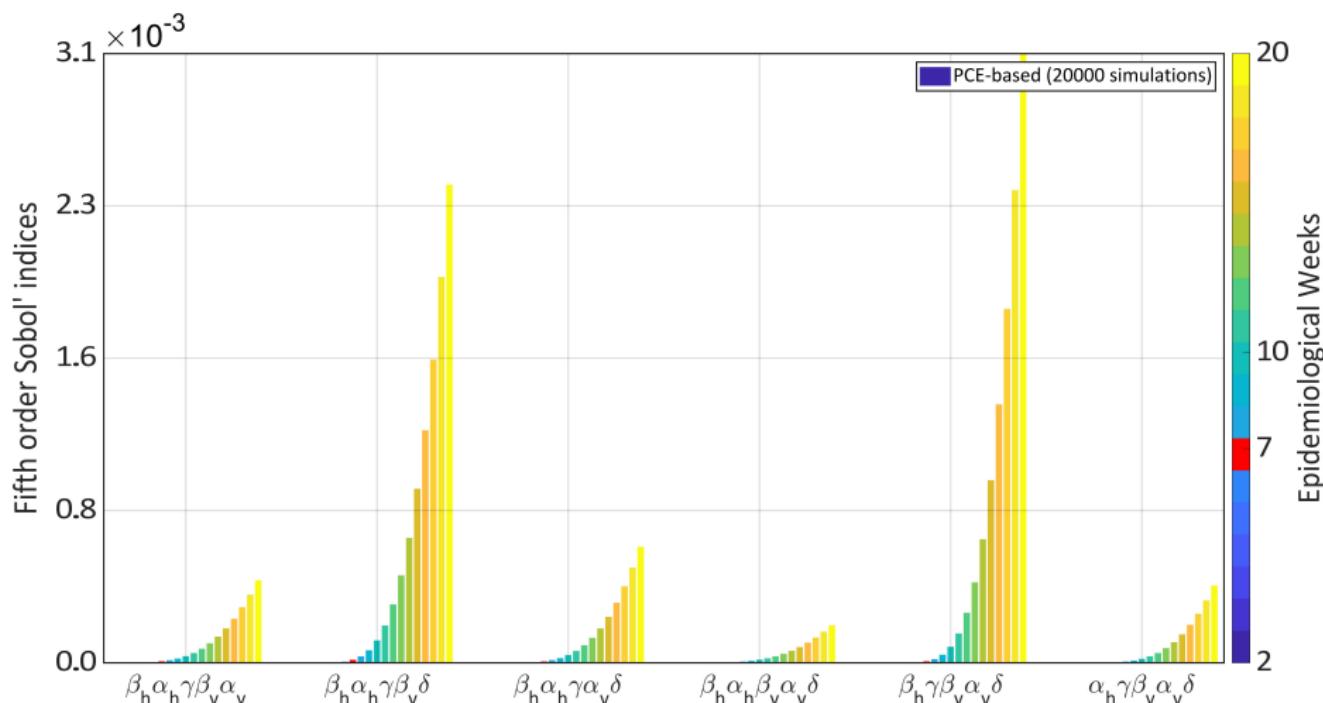
Global sensitivity analysis: third order



Global sensitivity analysis: fourth order



Global sensitivity analysis: fifth order



Global sensitivity analysis: sixth order

