A stochastic nonlinear dynamic model for Zika virus outbreak in Brazil

Americo Cunha Jr

Rio de Janeiro State University – UERJ
NUMERICO – **Nu**cleus of **M**odeling and **E**xpe**ri**mentation with **Co**mputers
http://numerico.ime.uerj.br

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

DDays LAC 2018
November 26 – 30, 2018
Punta del Este, Uruguay

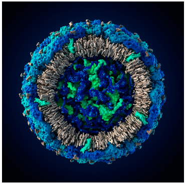
Outline

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
- International consensus that ZIKV is a cause of:
  - Guillain–Barré syndrome
  - Microcephaly

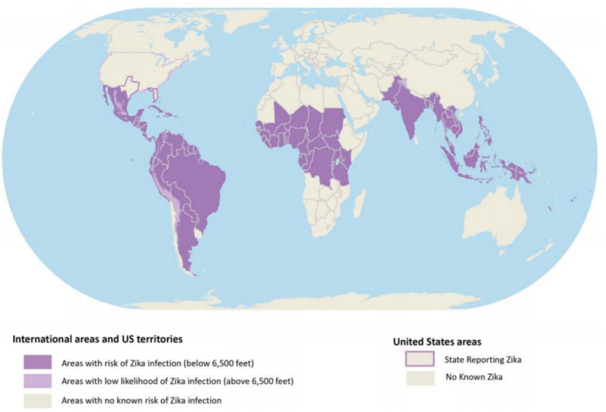


Zika virus

Aedes aegypti

Global outbreak of Zika virus

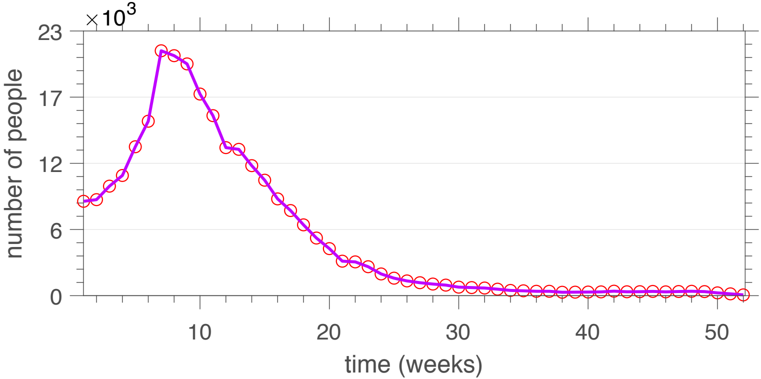
World Map of Areas with Risk of Zika



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/2OVgGGt`

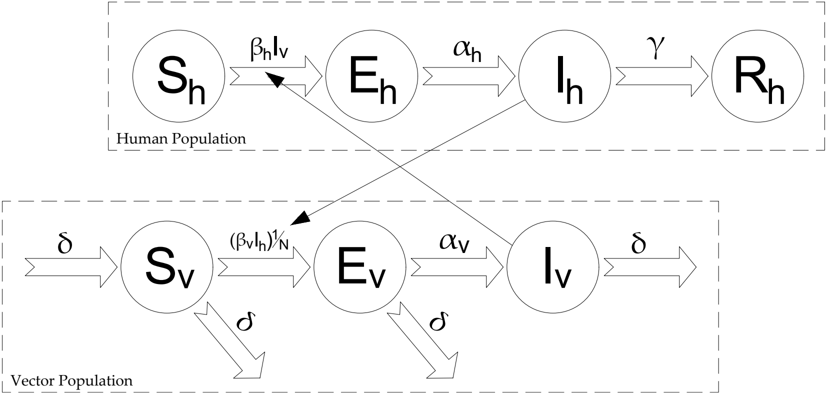
Research objectives

- 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

$$\frac{dS_h}{dt} = -\beta_h\, S_h\, I_v \qquad \frac{dS_v}{dt} = \delta - \beta_v\, S_v\, \frac{I_h}{N} - \delta\, S_v$$

$$\frac{dE_h}{dt} = \beta_h\, S_h\, I_v - \alpha_h\, E_h \qquad \frac{dE_v}{dt} = \beta_v\, S_v\, \frac{I_h}{N} - (\delta + \alpha_v)\, E_v$$

$$\frac{dI_h}{dt} = \alpha_h\, E_h - \gamma\, I_h \qquad \frac{dI_v}{dt} = \alpha_v\, E_v - \delta\, I_v$$

$$\frac{dR_h}{dt} = \gamma\, I_h \qquad \frac{dC}{dt} = \alpha_h\, E_h$$

$+$ initial conditions

S - Population of susceptible
E - Population of exposed
I - Population of infected
R - Population of recovered
N - Population of humans

C - Infected humans cumulative
α - Incubation ratio
δ - Vector lifespan ratio
β - Transmition rate
γ - Recovery rate

h - Human-related
v - Vector-related

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

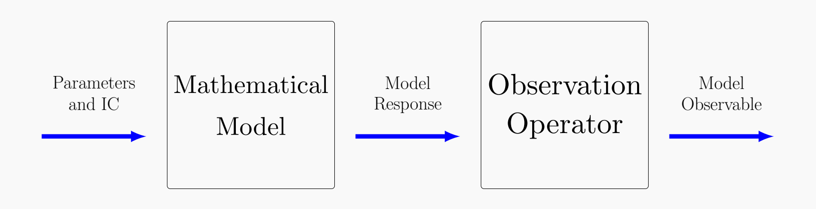
- Brazilian health system

MINISTÉRIO
DA SAÚDE

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)



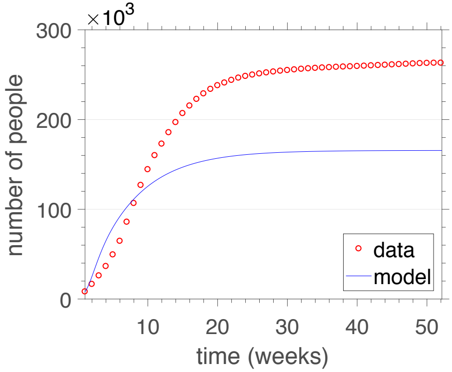
QoI 1: cumulative number of infectious

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

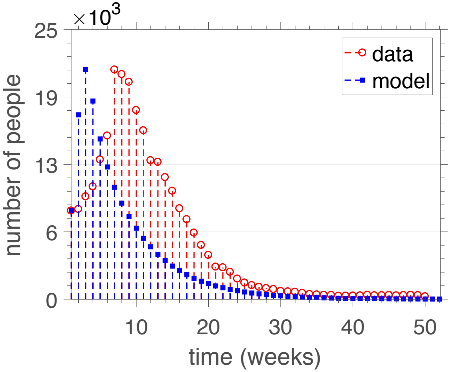
QoI 2: new infectious cases

$$\mathcal{N}_w = C_w - C_{w-1}, \quad (w = 2, 3, \cdots, 52)$$
$$\mathcal{N}_1 = C_1$$

Time series for QoI's

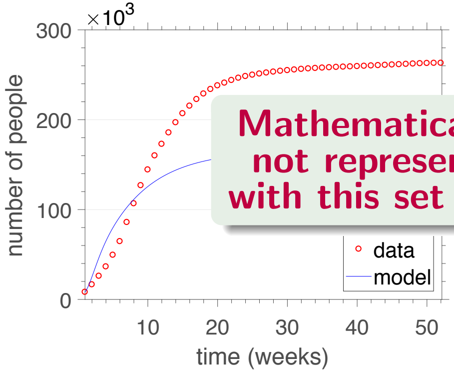


cumulative number of infectious

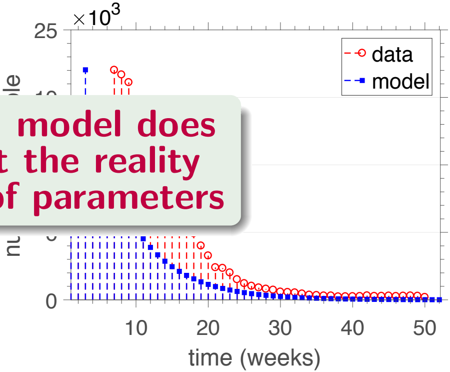


new infectious cases

Time series for QoI's



cumulative number of infectious



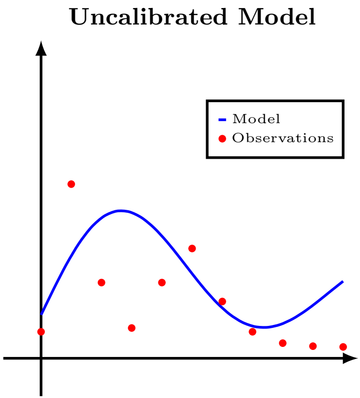
new infectious cases

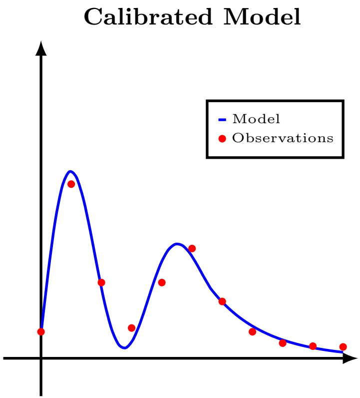
Mathematical model does not represent the reality with this set of parameters

Section 3

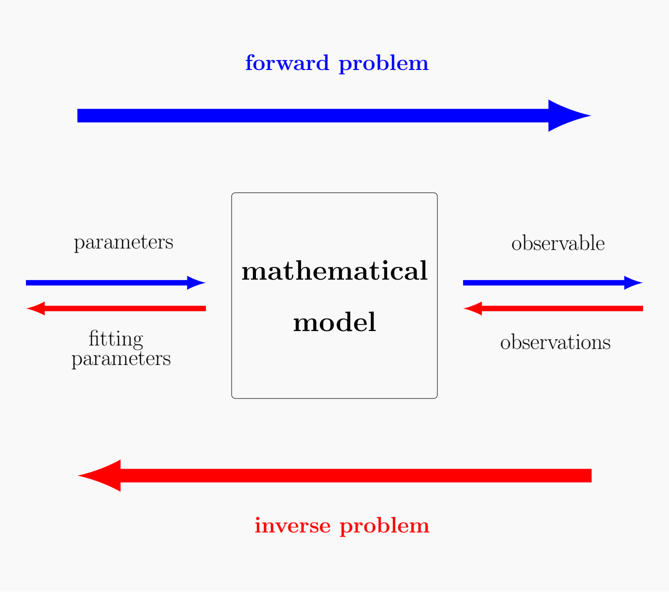
Model Calibration

Calibration of the model





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: $\boldsymbol{y} = (y_1, y_2, \cdots, y_M) \in F$
- prediction vector: $\boldsymbol{\phi}(\boldsymbol{\alpha}) = (\phi_1, \phi_2, \cdots, \phi_M) \in F$
- misfit function:

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

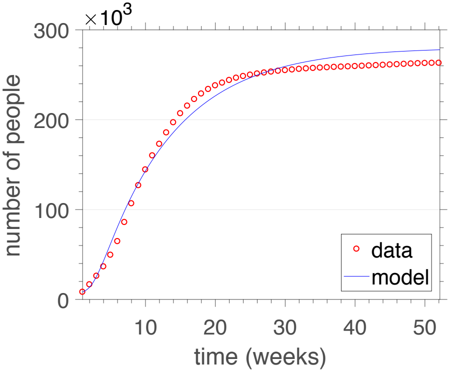
Find a vector of parameters such that

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

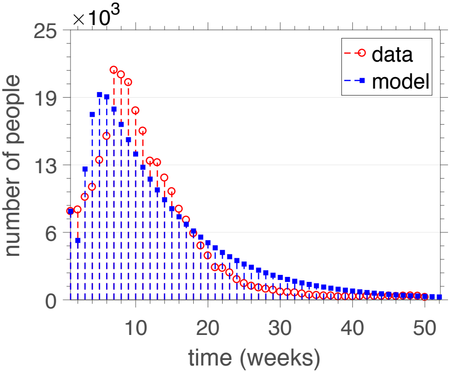
$\Longrightarrow$ Q-wellposed: existence, uniqueness, unimodality and local stability

$\Longrightarrow$ Solution algorithm: bounded trust-region-reflective

Calibrated model response

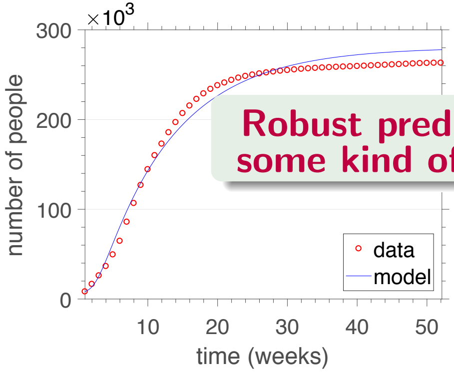


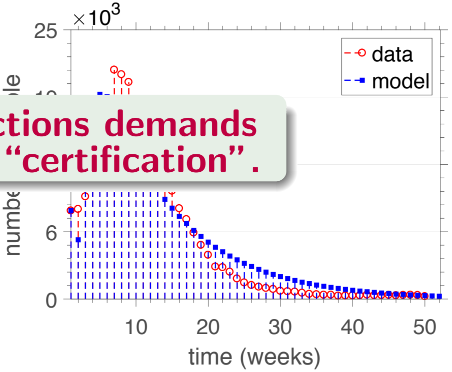
cumulative number of infectious



new infectious cases

Calibrated model response





Robust predictions demands some kind of "certification".

cumulative number of infectious

new infectious cases

Section 4

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:

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

An orthogonal decomposition in terms of conditional expectations:

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

Sobol' indices

Total variance:

$$D = Var\left[\mathcal{M}(\boldsymbol{X})\right] = \sum_{\boldsymbol{u} \subset \{1,\cdots,k\}} Var\left[\mathcal{M}_{\boldsymbol{u}}(\boldsymbol{X}_u)\right]$$

First order Sobol' indices:

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

(quantify the additive effect of each input separately)

Second order Sobol' indices:

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

(quantify interaction effect of inputs X_i and X_j)

Metamodelling via Polynomial Chaos

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

$$Y = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^k} y_{\boldsymbol{\alpha}} \, \boldsymbol{\Phi}_{\boldsymbol{\alpha}}$$

where

- $\boldsymbol{\Phi}_{\boldsymbol{\alpha}}$: multivariate orthonormal polynomials
- $y_{\boldsymbol{\alpha}}$: real-valued coeficients 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

$$Y \approx \sum_{\boldsymbol{\alpha} \in \mathcal{A}} y_{\boldsymbol{\alpha}} \, \boldsymbol{\Phi}_{\boldsymbol{\alpha}}$$

Thus, Sobol' indices are given by

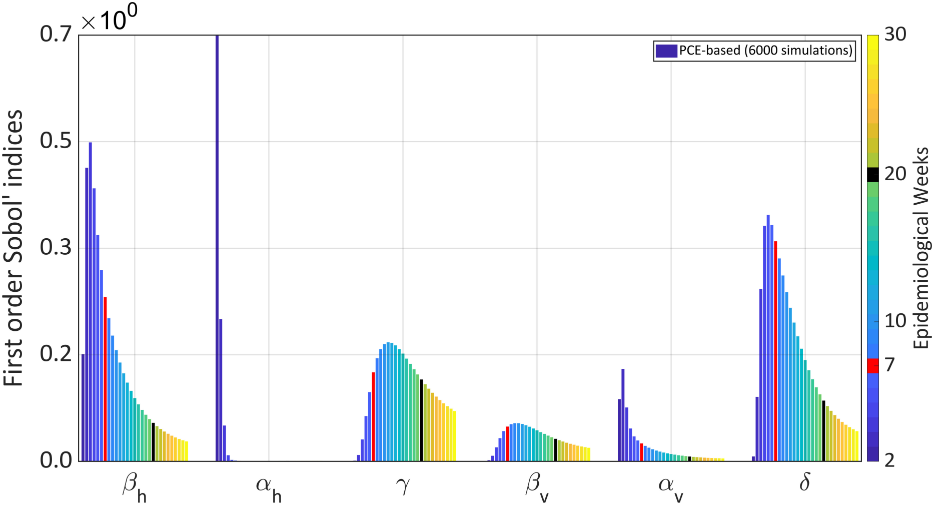
$$S_{\boldsymbol{u}} = D_{\boldsymbol{u}} / D = \sum_{\boldsymbol{\alpha} \in \mathcal{A}_{\boldsymbol{u}}} y_{\boldsymbol{\alpha}}^2 \Bigg/ \sum_{\boldsymbol{\alpha} \in \mathcal{A} \backslash 0} y_{\boldsymbol{\alpha}}^2$$

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

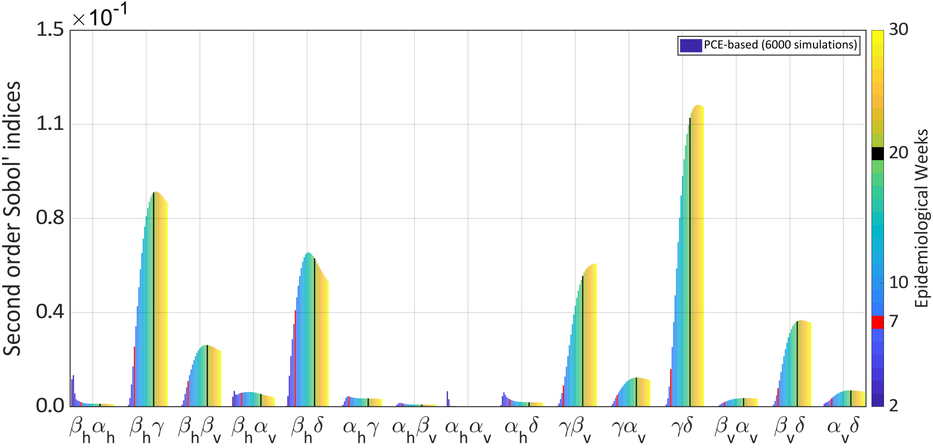
Sobol' indices of any order can be obtained, analytically, from the coefficients of the PC expansion!

B. Sudret, *Global sensitivity analysis using polynomial chaos expansions.* **Reliability Engineering & System Safety, 2016**, 93(7): 964–979, 2008.

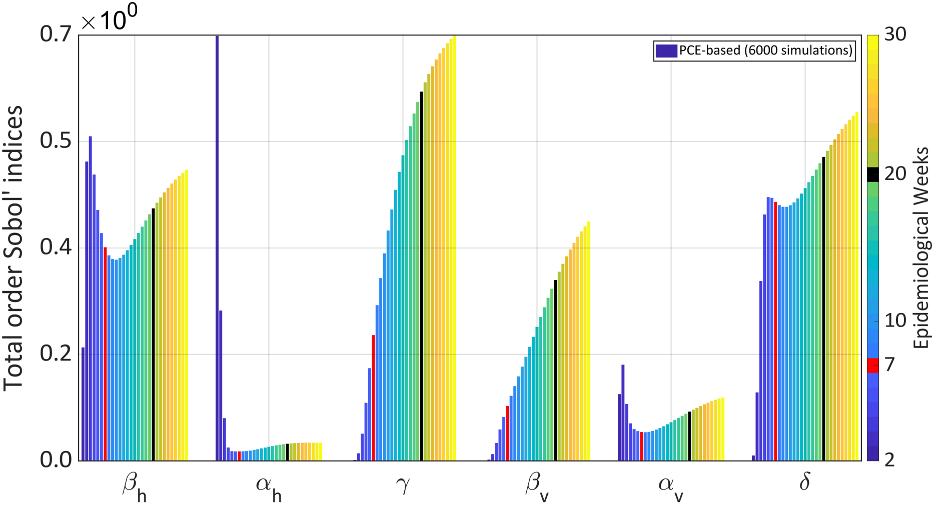
Global sensitivity analysis: first order



Global sensitivity analysis: second order



Global sensitivity analysis: total order



Global sensitivity analysis: general overview

$\rightarrow$ Two most relevant: δ and β_H (75% variance around 7th *EW*)

$\rightarrow$ Third most, γ, mainly by nonlinear interactions with δ and β_H

$\rightarrow$ Parameters limited to nonlinear interactions have, in general, delayed effects (significant for $EW > 15$)

$\rightarrow$ (*sparsity-of-effects principle*) Higher order interactions have minor effect: 1st and 2nd are 98.5%–86% variance on 7–15th *EW*

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

Section 5

Uncertainty Quantification

Uncertainty Quantification (UQ) framework

Mathematical model:

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

General steps for UQ:

1. Stochastic modeling
   $\longrightarrow$ characterization of inputs uncertainties
2. Uncertainty propagation
   $\longrightarrow$ characterization of output uncertainties
3. Response certification
   $\longrightarrow$ specification of reliability levels for predictions

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

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

"measure for the level of uncertainty"

MaxEnt optimization problem

Maximize

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

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

$$\int_{\mathbb{R}} g_k\left(X\right) p_X(x)\, dx = m_k, \qquad k = 0, \cdots, N,$$

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

MaxEnt optimization problem

Maximize

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

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

$$\int_{\mathbb{R}} g_k\left(X\right) p_X(x)\, dx = m_k, \qquad k = 0, \cdots, N,$$

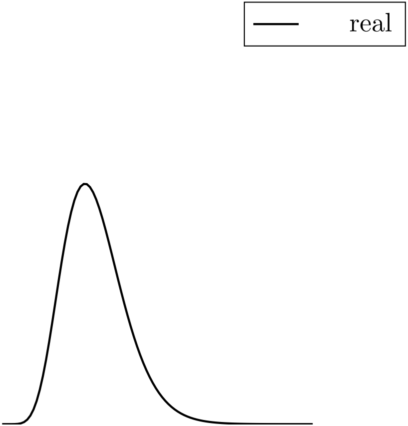
where the g_k are known real functions, with $g_0(x) = 1$.

MaxEnt general solution

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

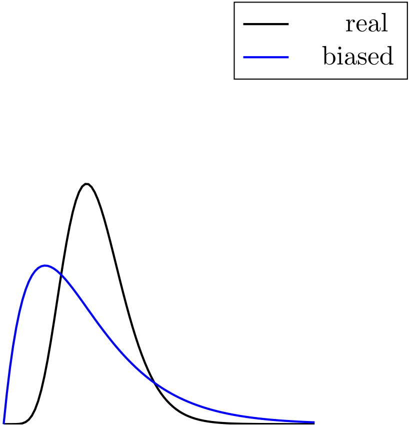
Philosophy of MaxEnt Principle

- The parameter of interest has a unknown distribution



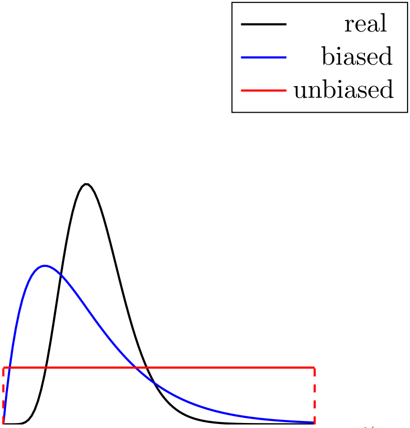
Philosophy of MaxEnt Principle

- 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	solution of model equations	computation of statistics
$\boldsymbol{X}_1$ $\vdots$ $\boldsymbol{X}_M$	$\boldsymbol{U} = h(\boldsymbol{X})$ **computational model**	$\boldsymbol{U}_1 = h(\boldsymbol{X}_1)$ $\vdots$ $\boldsymbol{U}_M = h(\boldsymbol{X}_M)$
$\longrightarrow$ known $F_{\boldsymbol{X}}$		$\longrightarrow$ 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 $\boldsymbol{U}$

Probabilistic model 1

Random variables: β_h and δ

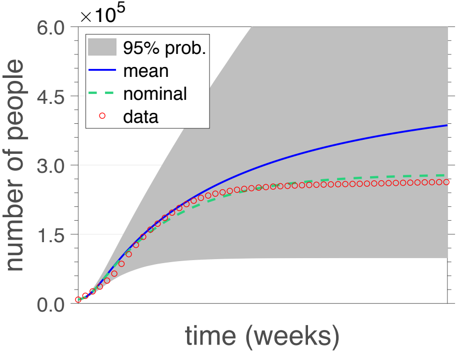
Available information: support and mean (nominal) value

MaxEnt distribution

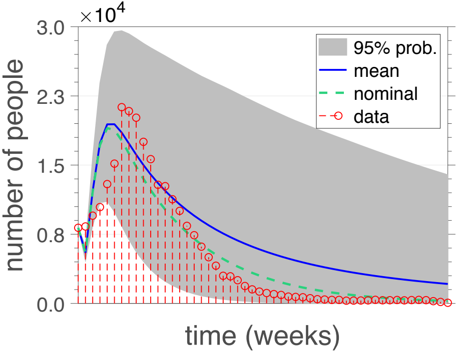
$$p_X(x) = \mathbb{1}_{[a,b]}(x)\, \exp\left(-\lambda_0 - \lambda_1\, x\right)$$

"truncated exponential (2 parameters)"

Confidence band for the QoIs



cumulative number of infectious



new infectious cases

Probabilistic model 2

Random variables: β_h and δ

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

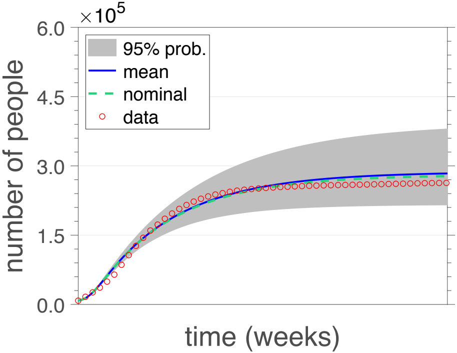
MaxEnt distribution

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

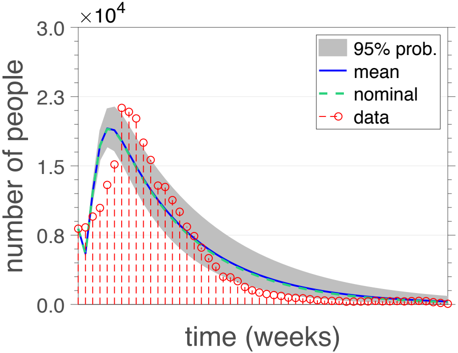
"truncated exponential (3 parameters)"

Confidence band for the QoIs

β_h dispersion $= 5\%$, δ dispersion $= 5\%$



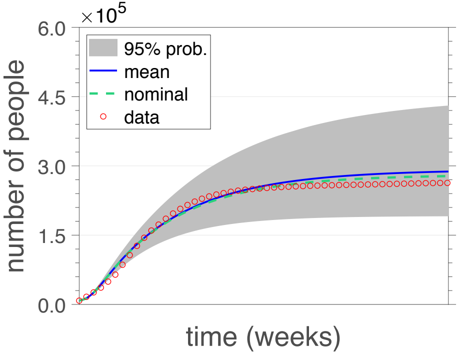
cumulative number of infectious



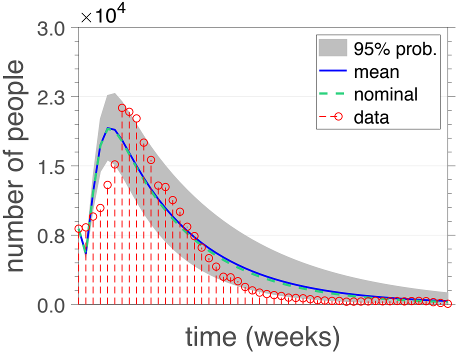
new infectious cases

Confidence band for the QoIs

β_h dispersion = 10% , δ dispersion = 5%



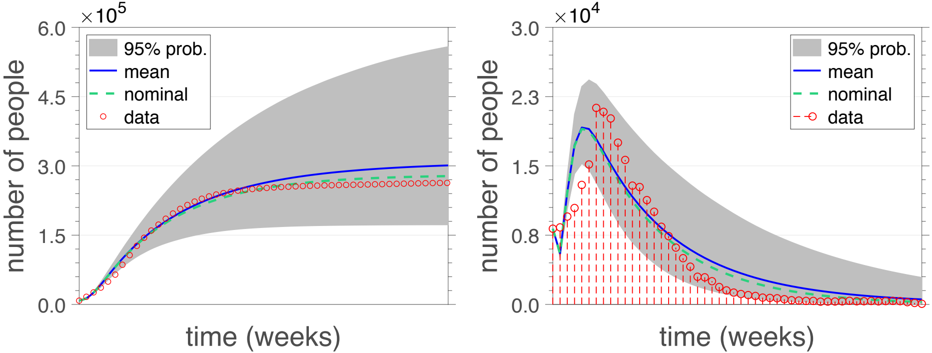
cumulative number of infectious



new infectious cases

Confidence band for the QoIs

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

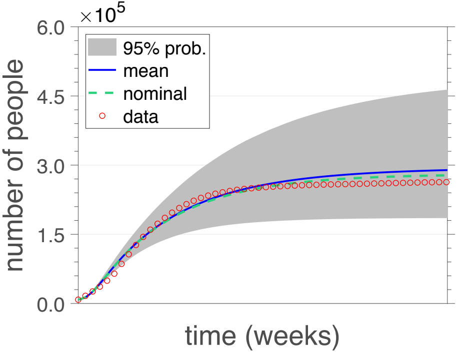
MaxEnt distribution for σ

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

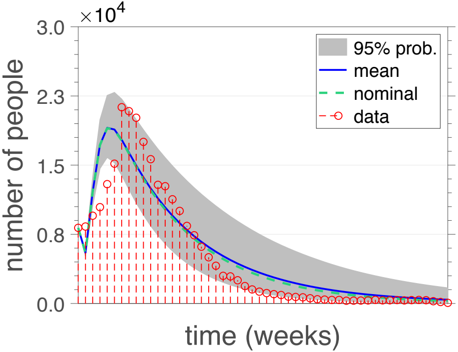
"uniform"

Confidence band for the QoIs

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



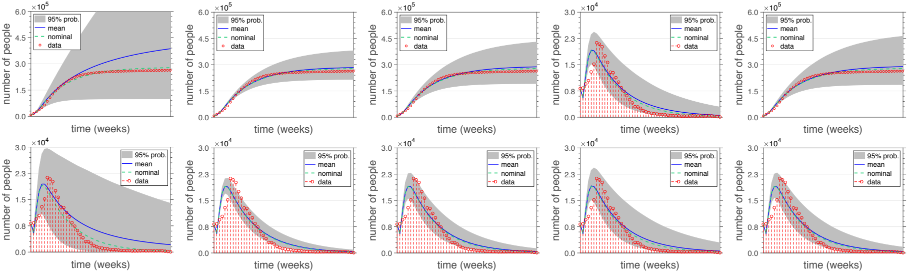
cumulative number of infectious



new infectious cases

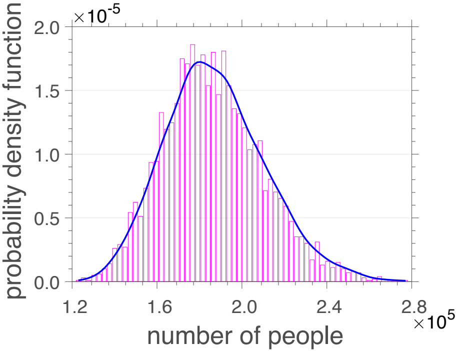
Confidence band for the QoIs

no dispersion | $D = \{5\%,5\%\}$ | $\sigma = \{10\%,5\%\}$ | $\sigma = \{10\%,10\%\}$ | $\sigma \sim U(5\%, 10\%)$

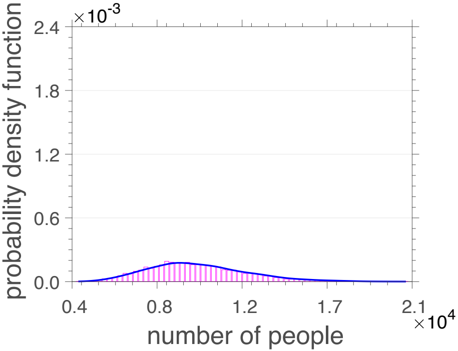


Evolution of QoIs PDFs

Epidemiological Week 13

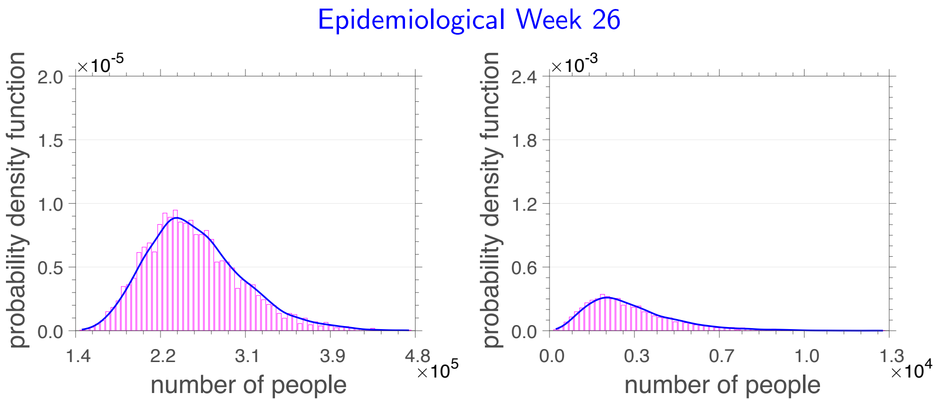


cumulative number of infectious



new infectious cases

Evolution of QoIs PDFs

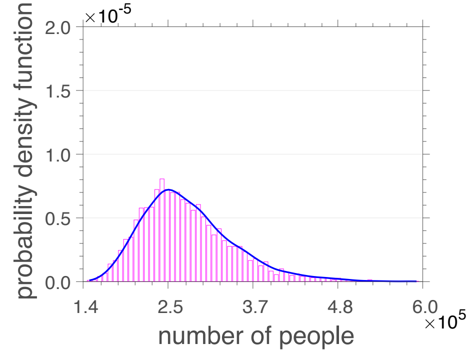


cumulative number of infectious

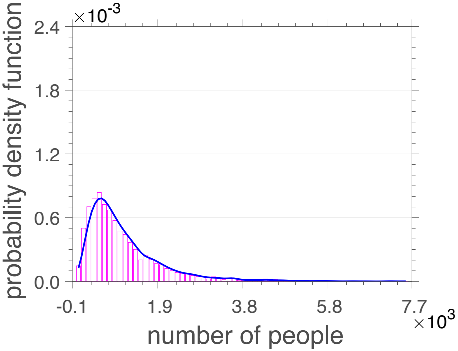
new infectious cases

Evolution of QoIs PDFs

Epidemiological Week 39



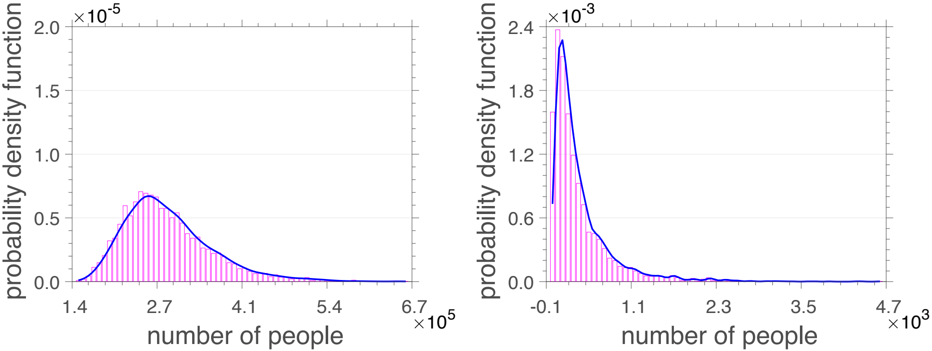
cumulative number of infectious



new infectious cases

Evolution of QoIs PDFs

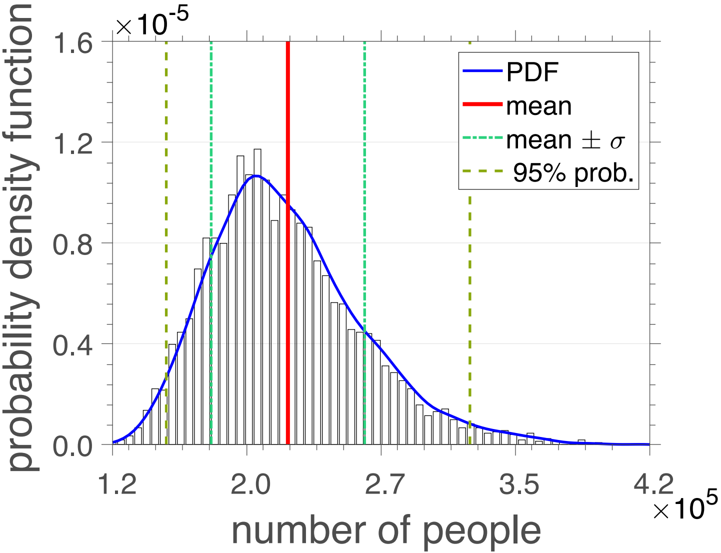
Epidemiological Week 52



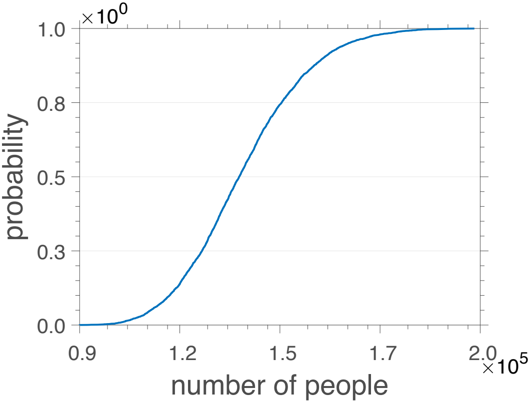
cumulative number of infectious

new infectious cases

Time-averaged cumulative infectious



Cumulative infectious CDF until EW 20

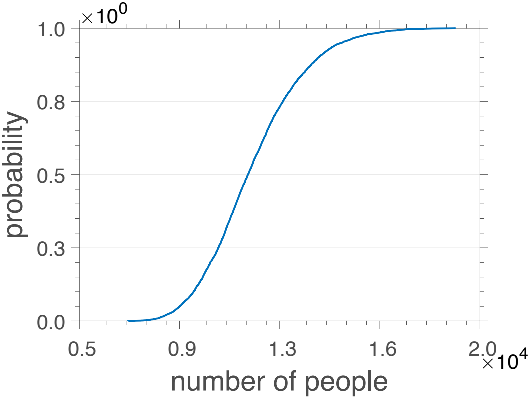


Statistics of C

$$\begin{aligned} \text{mean} &= 135{,}410 \\ \text{std. dev.} &= 16{,}366 \\ \text{skewness} &= 0.3894 \\ \text{kurtosis} &= 3.1135 \\ P(C \geq c^*) &= 60.33\% \end{aligned}$$

$c^* = 130{,}000$

New cases CDF until 20th EW



Statistics of $\mathcal{N}_w$

$$\begin{aligned} \text{mean} &= 11{,}493 \\ \text{std. dev.} &= 1.844 \\ \text{skewness} &= 0.5394 \\ \text{kurtosis} &= 3.3211 \\ \mathrm{P}(\mathcal{N}_w \geq N^*) &= 78.66\% \end{aligned}$$

$N^* = 10{,}000$

Section 6

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:

FAPERJ
Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro

CNPq
Conselho Nacional de Desenvolvimento Científico e Tecnológico

CAPES

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

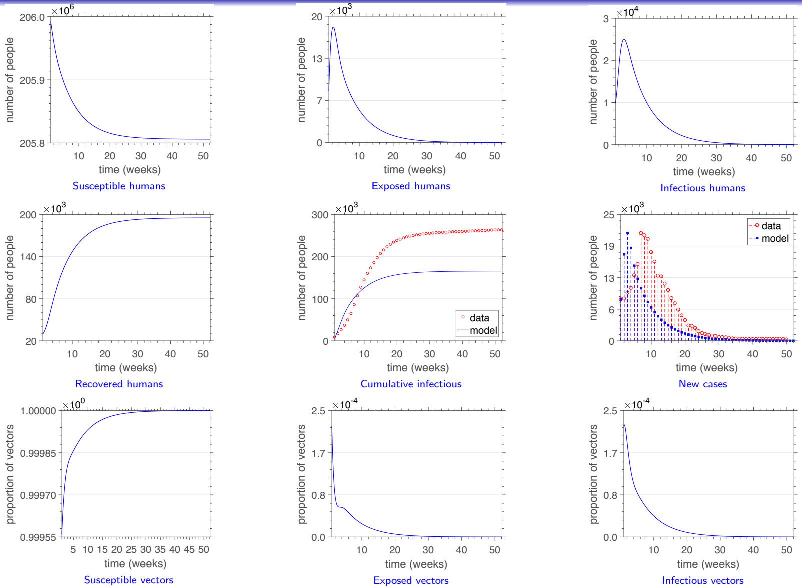
E. Dantas, M. Tosin and A. Cunha Jr,
Uncertainty quantification in the nonlinear dynamics of Zika virus, 2018
(in preparation).

nominal parameters

Nominal parameters and initial conditions

$\boldsymbol{\alpha}$	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

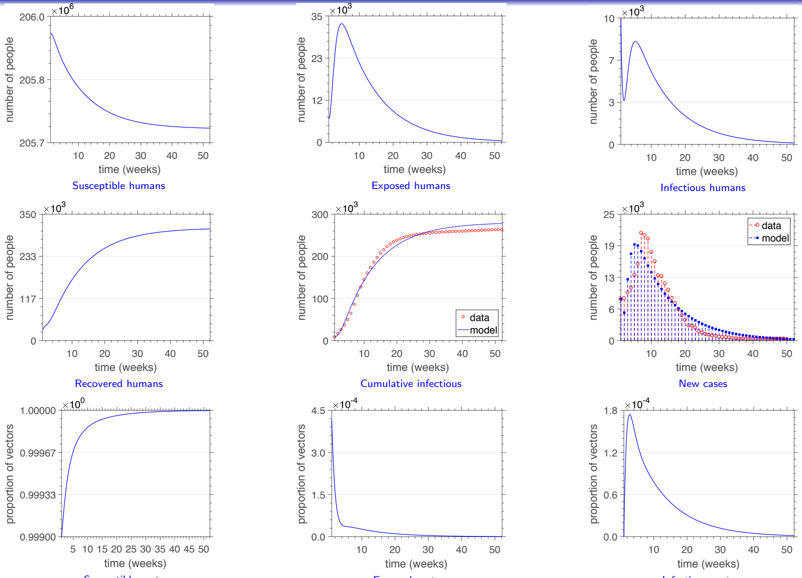


calibration

Calibration parameters and initial conditions

$\boldsymbol{\alpha}$	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

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 of Monte Carlo simulation:

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

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

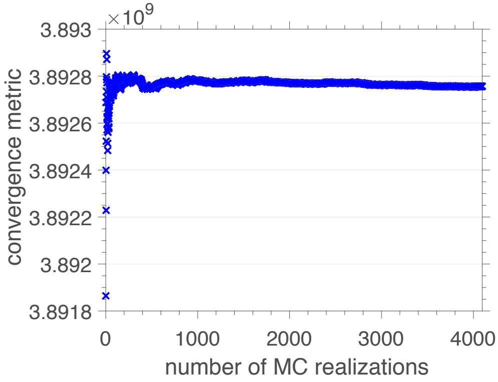


Figure: MC convergence metric as function of the number of realizations.