Data driven inference of model discrepancies in Zika virus dynamics

Michel Tosin michel.tosin@uerj.br

Americo Cunha Jr americo@ime.uerj.br

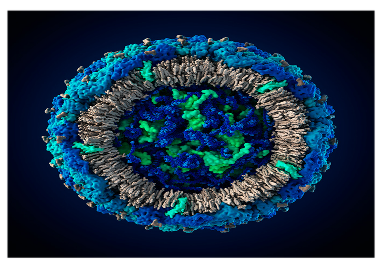
Flávio Coelho fccoelho@fgv.br

NUMERICO – Nucleus of Modeling and Experimentation with Computers

Introduction

- Zika virus: global widespread and connection with congenital diseases
- 2016: Zika becomes a public health emergency of international concern
- Main vector: *Aedes* mosquitoes
- A validated model can reveal new characteristics of the disease
- Prediction results are affected by model errors and lack of data
- Relations of model parameters are also of interest

Aedes aegypti



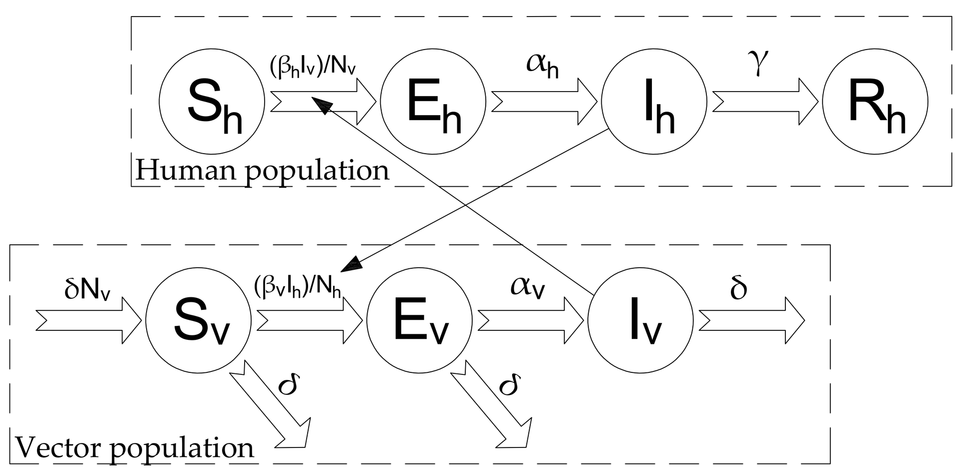
Zika virus

Objectives

- Perform sensitivity analysis to compare the parameters' global effect for a set of hierarchical models
- Improve a calibration result using a bigger data set obtained from a hierarchical superior model
- Develop a statistical framework using Bayesian Inference and Polynomial Chaos Expansion to quantify epidemic model discrepancies

Computational Model

SEIR-SEI Compartmental Model



Dynamical System

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

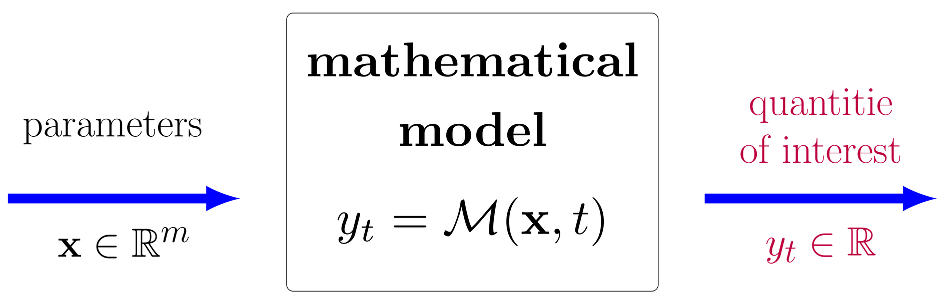
$$\frac{dE_h}{dt} = \beta_h\, S_h\, \frac{I_v}{N_v} - \alpha_h\, E_h\,, \qquad \frac{dE_v}{dt} = \beta_v\, S_v\, \frac{I_h}{N_h} - (\alpha_v + \delta)\, E_v\,,$$

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

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

+ Initial Conditions

Quantities of Interest (QoI)



- Cumulative cases of infectious: $C(t) = \int_{\tau=0}^{t} \alpha_h\, E_h(\tau)\, d\tau$
- New cases per week: $\mathcal{N}_w = C_w - C_{w-1}\,,\ w = 1 \ldots 52\,,\ \mathcal{N}_1 = C_1$

Sensitivity Analysis

The Hoeffding-Sobol' decomposition for n *iid* inputs $X_i \sim \mathcal{U}(0,1)$ gives

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

$$\mathcal{M}_0 = \mathbb{E}[Y_t]\,,\ \mathcal{M}_i(X_i) = \mathbb{E}\big[Y_t | X_i\big] - \mathcal{M}_0\,,\ \mathcal{M}_{ij}(X_i, X_j) = \mathbb{E}\big[Y_t | X_i, X_j\big] - \mathcal{M}_0 - \mathcal{M}_i - \mathcal{M}_j.$$

Sobol' Indices: interaction effect of inputs in $\mathbf{u}$

$$S_{\mathbf{u}} = \mathrm{Var}\left[\mathcal{M}_{\mathbf{u}}(X_{\mathbf{u}})\right] / \mathrm{Var}\left[\mathcal{M}(\mathbf{X})\right]$$

Metamodelling: Polynomial Chaos

The Polynomial Chaos Expansion of model $Y = \mathcal{M}(\mathbf{X})$, for a multivariate orthonormal polynomial family $\Phi_{\boldsymbol{\alpha}}$ with coefficients $y_{\boldsymbol{\alpha}}$,

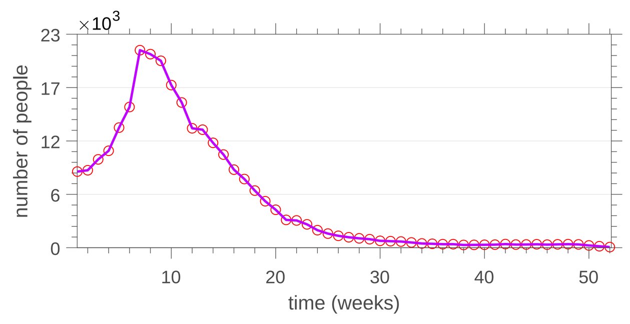
$$Y_t = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^k} y_{\boldsymbol{\alpha}}(t)\, \Phi_{\boldsymbol{\alpha}}(\mathbf{X})\,,$$

enables **analytic** computation of Sobol Indices:

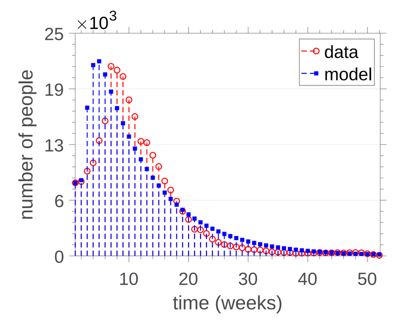
$$S_{\mathbf{u}} = \sum_{\boldsymbol{\alpha} \in \mathcal{A}_{\mathbf{u}}} y_{\boldsymbol{\alpha}}^2 \Bigg/ \sum_{\boldsymbol{\alpha} \in \mathcal{A} \setminus 0} y_{\boldsymbol{\alpha}}^2\,, \qquad \mathcal{A}_{\mathbf{u}} = \{\boldsymbol{\alpha} \in \mathcal{A} : i \in \mathbf{u} \Longleftrightarrow \alpha_i \neq 0\}$$

Hierarchical Calibration

⇒ Real data set: 52 values of New Cases of infectious humans for Zika by epidemiological week of 2016 from Brazil's Health Organizations



New cases of Zika in Brazil, 2016



SEIR-SEI calibration result

⇒ New data set: 358 values of Cumulative Cases from SEIR-SEI model

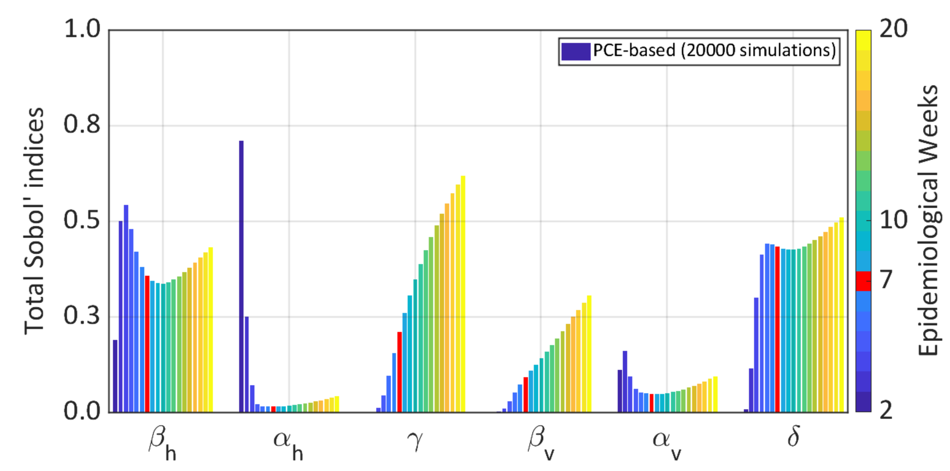
Use bigger data set from a calibrated model to improve the calibration results for other models and test some hypotheses of modeling

Models of Interest

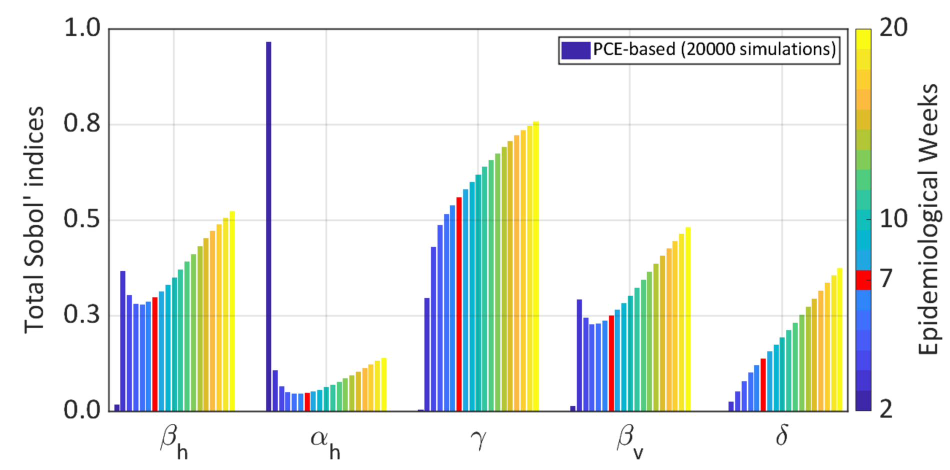
- SIR-SEI model
- SEIR-SI model
- SIR-SI model
- SEIR model

Results

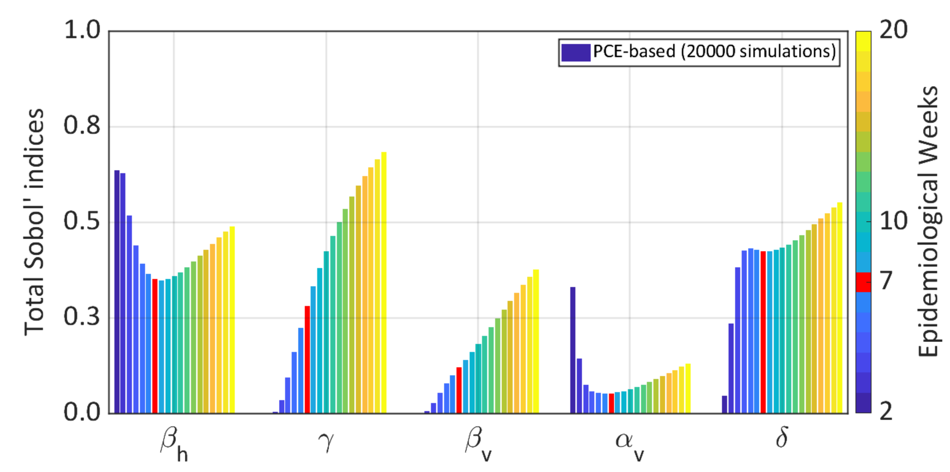
Sobol' Indices



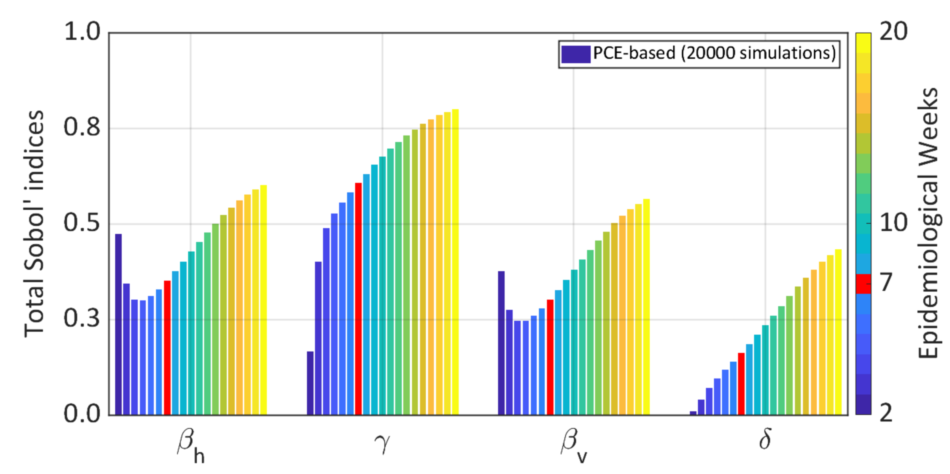
SEIR-SEI model



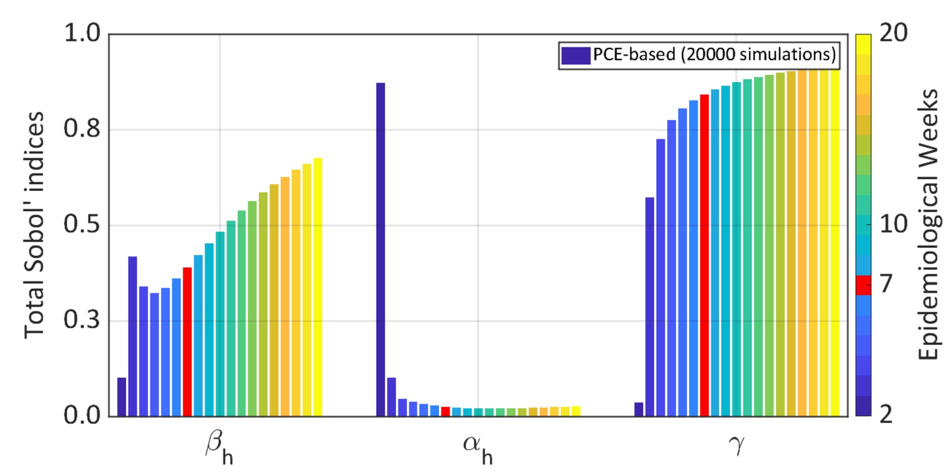
SEIR-SI model



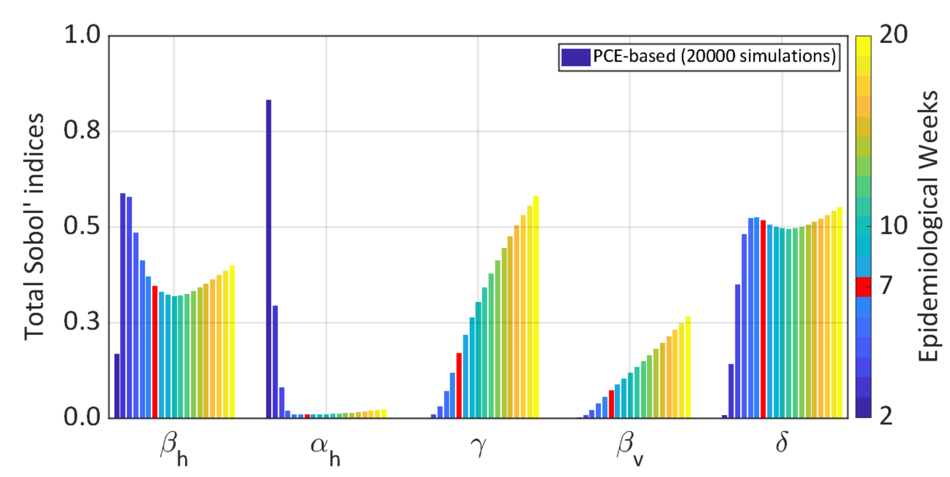
SIR-SEI model



SIR-SI model

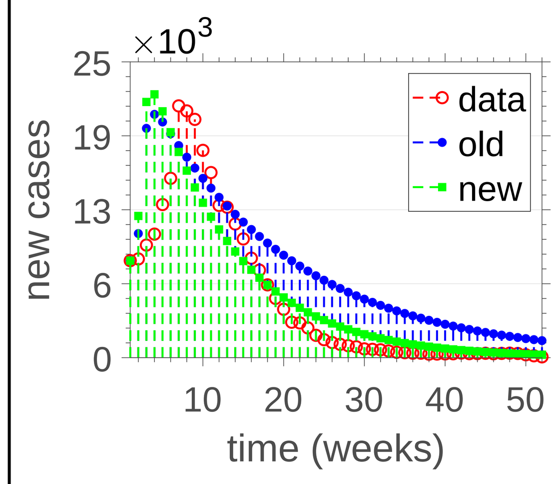


SEIR model

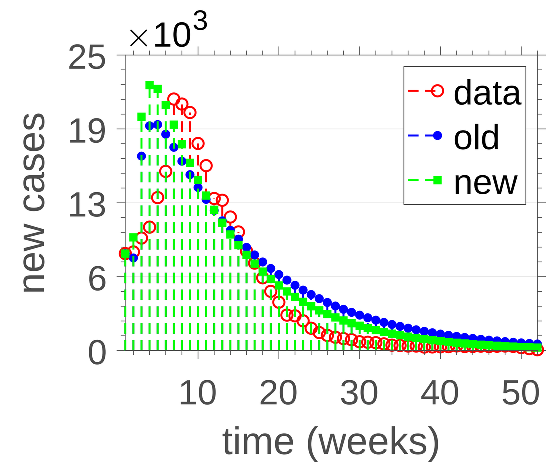


SEIR-SEI model with α_v constant

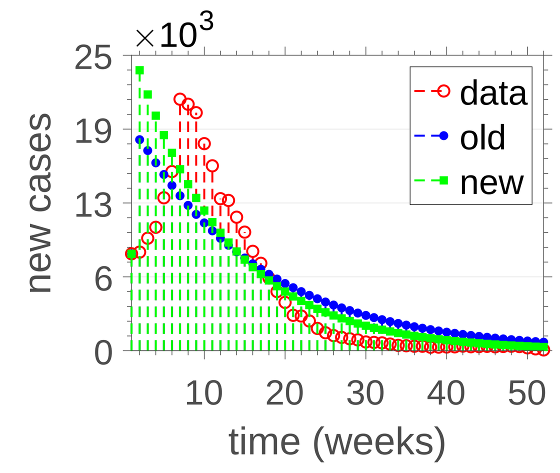
Calibration Results



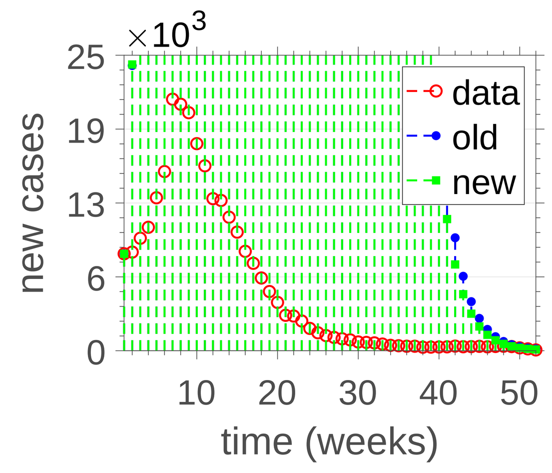
SIR-SEI model



SEIR-SI model



SIR-SI model



SEIR model

Statistical Inference (ongoing research)

Discrepancy Calculation

Suppose a data set $\mathcal{D} = (t_1, y_1^{dat}), (t_2, y_2^{dat}), \ldots, (t_{N_d}, y_{N_d}^{dat})$ of measures of the y_t. The i-th observation is given by

$$\underbrace{y_i^{dat}}_{reality} = \underbrace{\mathcal{M}(\mathbf{x}, t_i)}_{model} + \underbrace{\varepsilon_i}_{error}\,.$$

Sargsyan, Najm and Ghanem's [4] novel approach to deal with the model discrepancies is to adopt a metamodel structure which lumps the error into the parameters

$$Y^{dat} \approx \mathcal{M}(\mathbf{X}_\varepsilon, t)\,, \quad \mathbf{X}_\epsilon = \sum_{\boldsymbol{\alpha} \in \mathcal{I}} \mathbf{X}_{\boldsymbol{\alpha}}(t)\, \boldsymbol{\Psi}_{\boldsymbol{\alpha}}(\boldsymbol{\xi}),$$

where $\mathbf{X}_{\boldsymbol{\alpha}}$ coefficients are defined as random to be able to be identified by using Bayesian Inference

Bayesian Inference

- Inference problem become use data information to update the *prior* probability density function(PDF), defined for $\mathbf{X}_{\boldsymbol{\alpha}}$. The solution corresponds *posterior* PDF
- From Bayes' rule,

$$\pi(\mathbf{X}_{\boldsymbol{\alpha}} | \mathcal{D}) = \frac{\pi(\mathcal{D} | \mathbf{X}_{\boldsymbol{\alpha}})\, \pi(\mathbf{X}_{\boldsymbol{\alpha}})}{\pi(\mathcal{D})}\,.$$

→ $\pi(\mathbf{X}_{\boldsymbol{\alpha}} | \mathcal{D})$: posterior distribution
→ $\pi(\mathbf{X}_{\boldsymbol{\alpha}})$: prior distribution
→ $\pi(\mathcal{D} | \mathbf{X}_{\boldsymbol{\alpha}})$: likelihood function
→ $\pi(\mathcal{D})$: evidence

To define a good point of start, the Maximum Entropy Principle is applied to construct the most informative prior distribution

Final Remarks

- Comparative results of global Sobol' Indices show how the lack of some parameters can change the sensibility effect of the others
- With a bigger data set, the limitations in the prediction capacity of the hierarchically inferior models become more evident
- A framework for statistical inference exploring Polynomial Chaos to measure the model discrepancies was presented
- In future works, the authors intend explore this new framework to quantify model discrepancy and then improve its predictions

Acknowledgements

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

CNPq Conselho Nacional de Desenvolvimento Científico e Tecnológico

CAPES

References

[1] 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

[2] E. Dantas, M. Tosin and A. Cunha Jr. Uncertainty quantification in the nonlinear dynamics of Zika virus, 2019. hal.archives-ouvertes.fr/hal-02005320

[3] I. M. Sobol. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55(1-3): 271-280, 2001. https://doi.org/10.1016/S0378-4754(00)00270-6

[4] K. Sargsyan, H. N. Najm and R. Ghanem. On the statistical calibration of physical models. *International Journal of Chemical Kinetics*, 47(4): 246-276, 2015. https://doi.org/10.1002/kin.20906