How to predict an epidemic of Zika virus?
A challenge in nonlinear stochastic dynamics

Americo Cunha Jr

Universidade do Estado do Rio de Janeiro – UERJ
NUMERICO – Núcleo de Modelagem e Experimentação Computacional
http://numerico.ime.uerj.br

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

COLMEA 2019 - IM-UFRJ
March 28, 2019
Rio de Janeiro, Brazil
Outline

1 Introduction
2 Dynamic Model
3 Inverse Problem
4 Sensitivity Analysis
5 Uncertainty Quantification
6 Ongoing
7 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
- International consensus that ZIKV is a cause of:
  - Guillain–Barré syndrome
  - Microcephaly
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

Dengue virus (DENV)

- Member of *Flaviviridae* virus family
- Mainly spread by *Aedes* mosquitoes, as in the case for Zika virus
- Probable cases in Brazil:
  - > 170,000 in 2018
  - > 250,000 in 2017
  - > 3 million in 2016 and 2015
Other Arbovirus

- **Arthropod-Borne virus**
- **Yellow Fever**: South America and Africa  
  (261 deaths in Brazil in 2017)
- **Chikungunya**: worldwide  
  (> 204,000 confirmed cases in Brazil since 2015)
- **Rift Valley fever**: Africa and Arabian Peninsula  
  (ongoing outbreak in Kenya by June 2018)

Chikungunya cases (May 2018)
Other Arbovirus

- **Japanese encephalitis**: Southeast Asia, Western Pacific
- **West Nile virus**: widely established from Canada to Venezuela
- Both transmitted by the *Culex* mosquitoes

![West Nile virus activity in USA (July 2018)](image-url)
Typical questions to be answered

- How many people will the outbreak potentially infect?
- How far and how quickly will the disease spread?
- What areas and people are at highest risk, and when are they most at risk?
- How can we best make use of limited resources?
- How can we best slow or prevent the outbreak and protect vulnerable populations?

Typical questions to be answered

- How many people will the outbreak potentially infect?
- How far and how quickly will the disease spread?
- What areas and people are at highest risk, and when are they most at risk?
- How can we best make use of limited resources?
- How can we best slow or prevent the outbreak and protect vulnerable populations?

Mathematical models to simulate Zika virus spread can provide important guidance and insight to these questions.

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
SIS model

- Population of susceptible
- Population of infected
- Total population
- Transmission rate
- Recovery rate
- Birth rate
- Mortality rate

\[
\frac{dS}{dt} = bN - \beta I S + \gamma I - \mu S
\]

\[
\frac{dI}{dt} = \beta I S - \gamma I - \mu I
\]
SIS model

\[ \frac{dS}{dt} = bN - \beta \frac{I}{N} S - \mu S \]

Rate of change of \( S \) = Input of \( S \) - Output of \( S \)

- \( S \) - Population of susceptible
- \( I \) - Population of infected
- \( N \) - Total population
- \( \beta \) - Transmission rate
- \( \gamma \) - Recovery rate
- \( b \) - Birth rate
- \( \mu \) - Mortality rate

\[ S = \text{Population of susceptible} \]
\[ I = \text{Population of infected} \]
\[ N = \text{Total population} \]
\[ \beta = \text{Transmission rate} \]
\[ \gamma = \text{Recovery rate} \]
\[ b = \text{Birth rate} \]
\[ \mu = \text{Mortality rate} \]
**SIS model**

- Population of susceptible ($S$)
- Population of infected ($I$)
- Total population ($N$)
- Transmission rate ($\beta$)
- Recovery rate ($\gamma$)
- Birth rate ($b$)
- Mortality rate ($\mu$)

Rate of change of $S = \text{Input of } S - \text{Output of } S$

$$\frac{dS}{dt} = \left( \text{Births} + \text{Recovery} \right) - \left( \text{Infections} + \text{Mortality} \right)$$

- $\text{Births} = bN$
- $\text{Recovery} = \gamma I$
- $\text{Infections} = \beta I/N$
- $\text{Mortality} = \mu S$
**SIS model**

- **S** - Population of susceptible
- **I** - Population of infected
- **N** - Total population
- **β** - Transmission rate
- **γ** - Recovery rate
- **b** - Birth rate
- **μ** - Mortality rate

\[
\begin{align*}
\frac{dS}{dt} &= bN - \frac{\beta I}{N} S - \mu S \\
\frac{dI}{dt} &= \frac{\beta I}{N} S - \gamma I - \mu I
\end{align*}
\]
SIS model

- Population of susceptible
- Population of infected
- Total population
- Transmission rate
- Recovery rate
- Birth rate
- Mortality rate

Rate of change of $I = \text{Input of } I - \text{Output of } I$
SIS model

\[ S - \text{Population of susceptible} \]
\[ I - \text{Population of infected} \]
\[ N - \text{Total population} \]
\[ \beta - \text{Transmission rate} \]
\[ \gamma - \text{Recovery rate} \]
\[ b - \text{Birth rate} \]
\[ \mu - \text{Mortality rate} \]

Rate of change of \( I = \) Input of \( I \) - Output of \( I \)

\[
\frac{dI}{dt} = \beta \frac{S}{N} I - \left( \gamma I + \mu I \right)
\]

Infections
Recovery
Mortality
SIS model dynamical system

\[
\begin{align*}
\frac{dS}{dt} &= bN + \gamma I - \left( \beta \frac{I}{N} + \mu \right) S \\
\frac{dI}{dt} &= \beta \frac{I}{N} S - (\gamma + \mu) I
\end{align*}
\]

+ initial conditions

\( S \) - Population of susceptible
\( I \) - Population of infected
\( N \) - Total population

\( \beta \) - Transmission rate
\( \gamma \) - Recovery rate
\( b \) - Birth rate
\( \mu \) - Mortality rate
SIR model

\[ bN \rightarrow S \quad \beta I/N \rightarrow I \quad \gamma \rightarrow R \]

\[ \mu \rightarrow S \quad \mu \rightarrow I \quad \mu \rightarrow R \]
SIR model dynamical system

\[
\begin{align*}
\frac{dS}{dt} &= bN + \gamma I - \left( \beta \frac{I}{N} + \mu \right) S \\
\frac{dI}{dt} &= \beta \frac{I}{N} S - (\gamma + \mu) I \\
\frac{dR}{dt} &= \gamma I - \mu R
\end{align*}
\]

+ initial conditions

\[S - \text{Population of susceptible}\]
\[I - \text{Population of infected}\]
\[R - \text{Population of recovered}\]
\[N - \text{Total population}\]

\[\beta - \text{Transmission rate}\]
\[\gamma - \text{Recovery rate}\]
\[b - \text{Birth rate}\]
\[\mu - \text{Mortality rate}\]
SEIR model

\[ bN \quad \rightarrow \quad S \quad \quad \quad \mu \quad \rightarrow \quad E \quad \quad \quad \mu \quad \rightarrow \quad I \quad \quad \quad \mu \quad \rightarrow \quad R \]

\[ \beta I / N \quad \rightarrow \quad E \quad \quad \quad \alpha \quad \rightarrow \quad I \quad \quad \quad \gamma \quad \rightarrow \quad R \]
SEIR model dynamical system

\[
\frac{dS}{dt} = bN - \beta \frac{I}{N} S - \mu S
\]

\[
\frac{dE}{dt} = \beta \frac{I}{N} S - (\alpha + \mu) E
\]

\[
\frac{dI}{dt} = \alpha E - (\gamma + \mu) I
\]

\[
\frac{dR}{dt} = \gamma I - \mu R
\]

+ initial conditions

- $S$ - Population of susceptible
- $E$ - Population of exposed
- $I$ - Population of infectious
- $R$ - Population of recovered
- $N$ - Total population

- $\alpha$ - Incubation ratio
- $\beta$ - Transmission rate
- $\gamma$ - Recovery rate
- $b$ - Birth rate
- $\mu$ - Mortality rate
SEIR-SEI model for Zika virus dynamics

\[
\begin{align*}
S_h &\xrightarrow{(\beta_h I_h)/N_v} E_h \\
E_h &\xrightarrow{\alpha_h} I_h \xrightarrow{\gamma} R_h
\end{align*}
\]

\[
\begin{align*}
S_v &\xrightarrow{N_v \delta} E_v \xrightarrow{\delta} S_v \\
E_v &\xrightarrow{\delta} I_v \xrightarrow{\delta} I_v
\end{align*}
\]

Associated dynamical system

\[
\begin{align*}
\frac{dS_h}{dt} &= -\beta_h \frac{I_v}{N_v} S_h \\
\frac{dE_h}{dt} &= \beta_h \frac{I_v}{N_v} S_h - \alpha_h E_h \\
\frac{dl_h}{dt} &= \alpha_h E_h - \gamma I_h \\
\frac{dR_h}{dt} &= \gamma I_h \\
\frac{dS_v}{dt} &= \delta - \beta_v S_v \frac{I_h}{N_h} - \delta S_v \\
\frac{dE_v}{dt} &= \beta_v S_v \frac{I_h}{N_h} - (\delta + \alpha_v) E_v \\
\frac{dl_v}{dt} &= \alpha_v E_v - \delta I_v \\
\frac{dC}{dt} &= \alpha_h E_h
\end{align*}
\]

+ initial conditions

\[ S - \text{Population of susceptible} \]
\[ V - \text{Population of vaccinated} \]
\[ E - \text{Population of exposed} \]
\[ I - \text{Population of infectious} \]
\[ R - \text{Population of recovered} \]

\[ C - \text{Infected humans cumulative} \]
\[ N - \text{Total population} \]
\[ \alpha - \text{Incubation ratio} \]
\[ \beta - \text{Transmission ratio} \]
\[ \gamma - \text{Recovery rate} \]

\[ \delta - \text{Vector lifespan ratio} \]
\[ \sigma - \text{Infection rate of vaccinated} \]
\[ \nu - \text{Fraction of vaccinated} \]
\[ h - \text{Human-related} \]
\[ v - \text{Vector-related} \]
Model parameters and outbreak data

- open scientific literature
- Brazilian health system

<table>
<thead>
<tr>
<th>parameter</th>
<th>value</th>
<th>unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_h$</td>
<td>1/5.9</td>
<td>days$^{-1}$</td>
</tr>
<tr>
<td>$\alpha_v$</td>
<td>1/9.1</td>
<td>days$^{-1}$</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>1/7.9</td>
<td>days$^{-1}$</td>
</tr>
<tr>
<td>$\delta$</td>
<td>1/11</td>
<td>days$^{-1}$</td>
</tr>
<tr>
<td>$\beta_h$</td>
<td>1/11.3</td>
<td>days$^{-1}$</td>
</tr>
<tr>
<td>$\beta_v$</td>
<td>1/8.6</td>
<td>days$^{-1}$</td>
</tr>
<tr>
<td>$N$</td>
<td>$206 \times 10^6$</td>
<td>people</td>
</tr>
</tbody>
</table>
Time series of susceptible humans

**SIS model**

**SEIR model**

**SIR model**

**SEIR-SEI model**
Time series of infectious humans

- **SIS model**
- **SIR model**
- **SEIR model**
- **SEIR-SEI model**
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**

\[ \mathcal{N}_w = C_w - C_{w-1}, \quad (w = 2, 3, \ldots, 52) \]
\[ \mathcal{N}_1 = C_1 \]
Time series for QoI’s (SEIR-SEI model)

- Cumulative number of infectious cases
- New infectious cases
Time series for QoI’s (SEIR-SEI model)

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

cumulative number of infectious
new infectious cases
Section 3

Inverse Problem
Calibration of the model

Uncalibrated Model

Calibrated Model
Forward and inverse problem

- **Forward problem**
  - mathematical model
  - observable
  - parameters
  - fitting parameters

- **Inverse problem**
  - observable
  - parameters
  - fitting parameters
Inverse problem formulation

- data space: $F = \mathbb{R}^M$
- parameter space: $C = \left\{ \alpha \in \mathbb{R}^{12} \mid \alpha_{\text{min}} \leq \alpha \leq \alpha_{\text{max}} \right\}$
- observation vector: $y = (y_1, y_2, \cdots, y_M) \in F$
- prediction vector: $\phi(\alpha) = (\phi_1, \phi_2, \cdots, \phi_M) \in F$
- misfit function:
  \[
  J(\alpha) = \| y - \phi(\alpha) \|^2_F = \sum_{m=1}^{M} \left| y_m - \phi_m(\alpha) \right|^2
  \]

Find a vector of parameters such that

\[
\alpha^* = \arg \min_{\alpha \in C} J(\alpha).
\]
Inverse problem formulation

- data space: \( F = \mathbb{R}^M \)
- parameter space: \( C = \{ \alpha \in \mathbb{R}^{12} \mid \alpha_{\text{min}} \leq \alpha \leq \alpha_{\text{max}} \} \)
- observation vector: \( y = (y_1, y_2, \cdots, y_M) \in F \)
- prediction vector: \( \phi(\alpha) = (\phi_1, \phi_2, \cdots, \phi_M) \in F \)
- misfit function:

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

Find a vector of parameters such that

\[
\alpha^* = \arg\min_{\alpha \in C} J(\alpha).
\]

\[\Rightarrow \text{Q-wellposed: existence, uniqueness, unimodality and local stability}\]
Inverse problem formulation

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

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

Find a **vector of parameters** such that

\[
\alpha^* = \arg \min_{\alpha \in C} J(\alpha).
\]

- \( \Rightarrow \) **Q-wellposed**: existence, uniqueness, unimodality and local stability
- \( \Rightarrow \) **Solution algorithm**: bounded trust-region-reflective
Calibrated model response

Cumulative number of infectious cases

New infectious cases
Calibrated model response

Robust predictions demands some kind of "certification".

Cumulative number of infectious cases

New infectious cases
Section 4

Sensitivity Analysis
Parametric study for $\beta_h$

**cumulative number of infectious**

**new infectious cases**
Parametric study for $\alpha_h$

- Cumulative number of infectious cases
- New infectious cases
Parametric study for $\gamma$

- Cumulative number of infectious cases
- New infectious cases

---

- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
- $\times 10^3$
Parametric study for $\beta_v$
Parametric study for $\alpha_V$

Cumulative number of infectious cases vs. time (weeks)

New infectious cases vs. time (weeks)
Parametric study for $\delta$

Cumulative number of infectious cases

New infectious cases
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:

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

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:

\[
D = \text{Var} \left[ M(X) \right] = \sum_{u \subset \{1, \ldots, k\}} \text{Var} \left[ M_u(X_u) \right]
\]

First order Sobol’ indices:

\[
S_i = \frac{\text{Var} \left[ M_i(X_i) \right]}{D}
\]

(quantify the additive effect of each input separately)

Second order Sobol’ indices:

\[
S_{ij} = \frac{\text{Var} \left[ M_{ij}(X_i, X_j) \right]}{D}
\]

(quantify interaction effect of inputs \( X_i \) and \( X_j \))
Assuming $Y = \mathcal{M}(X)$ has finite variance, then it admits a Polynomial Chaos expansion

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

where

- $\Phi_\alpha(X)$: multivariate orthonormal polynomials
- $y_\alpha$: real-valued coefficients to be determined

PC-based Sobol’ indices

For computational purposes, a truncated PCE is employed

\[ Y \approx \sum_{\alpha \in A} y_{\alpha} \Phi_{\alpha}(X) \]

Thus, Sobol’ indices are given by

\[ S_u = D_u/D = \frac{\sum_{\alpha \in A_u} y_{\alpha}^2}{\sum_{\alpha \in A \setminus 0} y_{\alpha}^2} \]

\[ A_u = \{ \alpha \in A : i \in u \iff \alpha_i \neq 0 \} \]

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

Global sensitivity analysis: first order
Global sensitivity analysis: second order
Global sensitivity analysis: total order
→ Two most relevant: $\delta$ and $\beta_H$ (75% variance around 7th $EW$)

→ Third most, $\gamma$, mainly by nonlinear interactions with $\delta$ and $\beta_H$

→ 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 5

Uncertainty Quantification
Uncertainty Quantification (UQ) framework

Mathematical model:

\[ Y = \mathcal{M}(X) \]

General steps for UQ:

1. **Stochastic modeling**
   \[ \rightarrow \text{characterization of inputs uncertainties} \]
   
   (MaxEnt Principle)

2. **Uncertainty propagation**
   \[ \rightarrow \text{characterization of output uncertainties} \]
   
   (Monte Carlo Method)

3. **Response certification**
   \[ \rightarrow \text{specification of reliability levels for predictions} \]
   
   (Nonparametric Statistical Inference)

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

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

“measure for the level of uncertainty”
MaxEnt optimization problem

Maximize

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

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

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

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

Maximize

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

requiring \( N + 1 \) constraints (known information) given by

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

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

MaxEnt general solution

\[ p_X(x) = 1_K(x) \exp(-\lambda_0) \exp \left( -\sum_{k=1}^{N} \lambda_k g_k(x) \right) \]
Philosophy of MaxEnt Principle

- The parameter of interest has a unknown distribution
The parameter of interest has an 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.
Monte Carlo Method

**pre-processing**
- generation of scenarios
- known $F_X$
- generator of random vector $X$

**processing**
- solution of model equations
- deterministic solver of $u = h(x)$

**post-processing**
- computation of statistics
- estimated $F_U$
- statistical inference to estimate convergence and distribution of $U$

Uncertainty propagation through the model
Random variables: $\beta_h$ and $\delta$

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 QoIs

cumulative number of infectious

new infectious cases
Probabilistic model 2

Random variables: $\beta_h$ and $\delta$

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

MaxEnt distribution

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

“truncated exponential (3 parameters)”
Confidence band for the QoIs

$\beta_h$ dispersion $= 5\%$, $\delta$ dispersion $= 5\%$

**cumulative number of infectious**

**new infectious cases**
Confidence band for the QoIs

\[ \beta_h \text{ dispersion } = 10\% , \ \delta \text{ dispersion } = 5\% \]

**Cumulative Number of Infectious**

**New Infectious Cases**
Confidence band for the QoIs

\[ \beta_h \text{ dispersion } = 10\% , \; \delta \text{ dispersion } = 10\% \]

- **Cumulative number of infectious cases**
  - Graph 1: 
    - X-axis: time (weeks)
    - Y-axis: cumulative number of people
    - Legend: 95% prob., mean, nominal, data

- **New infectious cases**
  - Graph 2: 
    - X-axis: time (weeks)
    - Y-axis: number of people
    - Legend: 95% prob., mean, nominal, data
Random variables: $\beta_h$, $\delta$ and $\sigma$

Available information for $\beta_h$ and $\delta$: support, mean (nominal) value

**Distribution for $\beta_h$ and $\beta_v$**

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

Available information for $\sigma$: support

**MaxEnt distribution for $\sigma$**

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

“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**
- $\sigma = \{5\%,5\%\}$
- $\sigma = \{10\%,5\%\}$
- $\sigma = \{10\%,10\%\}$
- $\sigma \sim U(5\%,10\%)$
Evolution of QoIs PDFs

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

cumulative number of infectious
case

new infectious cases
Time-averaged cumulative infectious

- 

- no dispersion

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

- $\sigma \sim U(5\%, 10\%)$
(mean) Cumulative infectious CDF until EW 20

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

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

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

Statistics of $N_w$

- mean $= 1.57 \times 10^4$
- std. dev. $= 1.35 \times 10^3$
- skewness $= -0.032$
- kurtosis $= 2.656$
- $P(N_w \geq NC^*) = 83.40\%$

$NC^* = 14,440$

average $NC$ (data) until EW 10
Section 6

Ongoing
Investigation of control strategies
SVEIR-SEI model for Zika virus dynamics

Associated dynamical system

\[
\begin{align*}
\frac{dS_h}{dt} &= - \left( \beta_h \frac{I_v}{N_v} + \nu \right) S_h \\
\frac{dV_h}{dt} &= \nu S_h - \sigma \beta_h \frac{I_v}{N_v} V_h \\
\frac{dE_h}{dt} &= \beta_h (S_h + \sigma V_h) \frac{I_v}{N_v} - \alpha_h E_h \\
\frac{dI_h}{dt} &= \alpha_h E_h - \gamma I_h \\
\frac{dR_h}{dt} &= \gamma I_h \\
\frac{dS_v}{dt} &= \delta - \beta_v S_v \frac{I_h}{N_h} - \delta S_v \\
\frac{dE_v}{dt} &= \beta_v S_v \frac{I_h}{N_h} - (\delta + \alpha_v) E_v \\
\frac{dI_v}{dt} &= \alpha_v E_v - \delta I_v \\
\frac{dC}{dt} &= \alpha_h E_h
\end{align*}
\]

+ initial conditions

\[
\begin{align*}
S & - \text{Population of susceptible} & C & - \text{Infected humans cumulative} & \delta & - \text{Vector lifespan ratio} \\
V & - \text{Population of vaccinated} & N & - \text{Total population} & \sigma & - \text{Infection rate of vaccinated} \\
E & - \text{Population of exposed} & \alpha & - \text{Incubation ratio} & \nu & - \text{Fraction of vaccinated} \\
I & - \text{Population of infectious} & \beta & - \text{Transmission rate} & h & - \text{Human-related} \\
R & - \text{Population of recovered} & \gamma & - \text{Recovery rate} & v & - \text{Vector-related}
\end{align*}
\]

Time series for QoI’s (SVEIR-SEI model)

cumulative number of infectious

new infectious cases
Quantification of model discrepancy
Calculation of model discrepancy

Conventional statistical calibration:
\[ y_{\text{truth}} = f(x, p) + \varepsilon_{\text{model}} \]

Novel approach:
\[ y_{\text{truth}} \approx f(x, p_\varepsilon), \quad p_\varepsilon = \sum_k \alpha_k \Phi_k(\xi) \]

Bayesian inversion to identify $\alpha$
\[ \pi(\text{model} | \text{data}) \propto \pi(\text{data} | \text{model}) \times \pi(\text{model}) \]


Section 7

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 a parametric probabilistic model of uncertainties

Ongoing research:

- Investigate the effectiveness of different control strategies
- Quantify model discrepancy in a nonparametric way

Future directions:

- Scenarios exploration with active subspace method
- Data-driven identification of epidemiological models
Acknowledgments

Invitation for the talk:

- Profª. Maria Eulalia Vares
- Prof. Leandro Pimentel

Financial support:
Thank you for your attention!

americo@ime.uerj.br

www.americocunha.org

E. Dantas, M. Tosin and A. Cunha Jr,
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, 2019 (in preparation).
https://hal.archives-ouvertes.fr/hal-02005320
nominal parameters
Nominal parameters and initial conditions

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_h$</td>
<td>$1/5.9$</td>
<td>days$^{-1}$</td>
</tr>
<tr>
<td>$\alpha_v$</td>
<td>$1/9.1$</td>
<td>days$^{-1}$</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>$1/7.9$</td>
<td>days$^{-1}$</td>
</tr>
<tr>
<td>$\delta$</td>
<td>$1/11$</td>
<td>days$^{-1}$</td>
</tr>
<tr>
<td>$\beta_h$</td>
<td>$1/11.3$</td>
<td>days$^{-1}$</td>
</tr>
<tr>
<td>$\beta_v$</td>
<td>$1/8.6$</td>
<td>days$^{-1}$</td>
</tr>
<tr>
<td>$N$</td>
<td>$206 \times 10^6$</td>
<td>people</td>
</tr>
<tr>
<td>$S_h^i$</td>
<td>$205,953,959$</td>
<td>people</td>
</tr>
<tr>
<td>$E_h^i$</td>
<td>$8,201$</td>
<td>people</td>
</tr>
<tr>
<td>$I_h^i$</td>
<td>$8,201$</td>
<td>people</td>
</tr>
<tr>
<td>$R_h^i$</td>
<td>$29,639$</td>
<td>people</td>
</tr>
<tr>
<td>$S_v^i$</td>
<td>$0.99956$</td>
<td>______</td>
</tr>
<tr>
<td>$E_v^i$</td>
<td>$2.2 \times 10^{-4}$</td>
<td>______</td>
</tr>
<tr>
<td>$I_v^i$</td>
<td>$2.2 \times 10^{-4}$</td>
<td>______</td>
</tr>
</tbody>
</table>
Model response with nominal parameters

- **Susceptible humans**
  - Number of people decreases over time, approaching zero.
  - Time (weeks): 0 to 50

- **Exposed humans**
  - Number of people increases initially, then stabilizes.
  - Time (weeks): 0 to 50

- **Infectious humans**
  - Number of people increases rapidly initially, then decreases.
  - Time (weeks): 0 to 50

- **Recovered humans**
  - Number of people increases from zero.
  - Time (weeks): 0 to 50

- **Cumulative infectious**
  - Proportion of vectors increases over time, approaching a peak.
  - Time (weeks): 0 to 50

- **Susceptible vectors**
  - Proportion of vectors increases from zero.
  - Time (weeks): 0 to 50

- **Exposed vectors**
  - Proportion of vectors increases from zero.
  - Time (weeks): 0 to 50

- **Infectious vectors**
  - Proportion of vectors decreases over time, approaching zero.
  - Time (weeks): 0 to 50

- **New cases**
  - Data and model comparison.
  - Time (weeks): 0 to 50
Inverse Problem
Well-posedness

Let the forward map \( \phi : E \rightarrow F \) associates to each parameter vector \( x \), restricted to be on the set of admissible values \( C \) in the parameter space \( E \), an observable vector in the data space \( F \). The NLS problem is Quadratically (Q-) wellposed if, and only if, \( \phi(C) \) possesses an open neighborhood \( \vartheta \) such that

1. **Existence and uniqueness:** for every \( z \in \vartheta \), the inverse problem has a unique solution \( \hat{x} \)
2. **Unimodality:** for every \( z \in \vartheta \), the objective function \( x \mapsto J(x) \) has no parasitic stationary point
3. **Local stability:** the mapping \( z \mapsto \hat{x} \) is locally Lipschitz continuous from \( (\vartheta, \| \cdot \|_F) \) to \( (C, \| \cdot \|_E) \).

Theorem

Let the follow finite dimension minimum set of hypothesis hold:

\[
\begin{align*}
E &= \text{finite dimensional vector space, with norm } || \cdot ||_E, \\
C &= \text{closed, convex subset of } E, \\
C_\eta &= \text{convex open neighborhood of } C \text{ in } E, \\
F &= \text{Hilbert space, with norm } || \cdot ||_F, \\
z &\in F, \\
\phi : C_\eta &\rightsquigarrow F \text{ is twice differentiable along segments of } C_\eta, \\
\text{and: } V &= \frac{\partial}{\partial t} \phi((1 - t)x_0 + tx_1), \\
A &= \frac{\partial^2}{\partial t^2} \phi((1 - t)x_0 + tx_1) \\
\text{are continuous functions of } x_0, x_1 &\in C_\eta \text{ and } t \in [0, 1].
\end{align*}
\]

Then, if moreover \( C \) is small enough for the deflection condition \( \theta \leq \pi/2 \) to hold, \( x \) is OLS-identifiable on \( C \), or equivalently: the NLS problem is Q-wellposed on \( C \).

### Calibrated parameters and initial conditions

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>TRR input</th>
<th>$lb$</th>
<th>$ub$</th>
<th>TRR output</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_h$</td>
<td>1/5.9</td>
<td>1/12</td>
<td>1/3</td>
<td>1/12</td>
</tr>
<tr>
<td>$\alpha_v$</td>
<td>1/9.1</td>
<td>1/10</td>
<td>1/5</td>
<td>1/10</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>1/7.9</td>
<td>1/8.8</td>
<td>1/3</td>
<td>1/3</td>
</tr>
<tr>
<td>$\delta$</td>
<td>1/11</td>
<td>1/21</td>
<td>1/11</td>
<td>1/21</td>
</tr>
<tr>
<td>$\beta_h$</td>
<td>1/11.3</td>
<td>1/16.3</td>
<td>1/8</td>
<td>1/10.40</td>
</tr>
<tr>
<td>$\beta_v$</td>
<td>1/8.6</td>
<td>1/11.6</td>
<td>1/6.2</td>
<td>1/7.77</td>
</tr>
<tr>
<td>$S_v^{i_h}$</td>
<td>205,953,959</td>
<td>0.9 $\times$ N</td>
<td>N</td>
<td>205,953,534</td>
</tr>
<tr>
<td>$E_h^{i_h}$</td>
<td>8,201</td>
<td>0</td>
<td>10,000</td>
<td>6,827</td>
</tr>
<tr>
<td>$I_h^{i_h}$</td>
<td>8,201</td>
<td>0</td>
<td>10,000</td>
<td>10,000</td>
</tr>
<tr>
<td>$S_v^{i_v}$</td>
<td>0.9996</td>
<td>0.99</td>
<td>0.999</td>
<td>0.999</td>
</tr>
<tr>
<td>$E_v^{i_v}$</td>
<td>$2.2 \times 10^{-4}$</td>
<td>0</td>
<td>1</td>
<td>$4.14 \times 10^{-4}$</td>
</tr>
<tr>
<td>$I_v^{i_v}$</td>
<td>$2.2 \times 10^{-4}$</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Remarks on the calibration

- Reasonable parameters
- Cumulative number of infectious overshoots data by only 5.74%
- Initial infectious humans is approximately 10,000 individuals
- Peak value of new infectious cases differs from the data maximum by 10.57%
- Peak of new infectious cases occurs two weeks before the peak of the data
Comparison of infectious humans curves

First calibration

Second calibration
Comparison of infectious humans curves

Curves for various initial infectious humans values

Zoom in the local peak region of the image to the left
Comparison of cumulative and new infectious curves

cumulative number of infectious cases

new infectious cases
Calibrated model response

- Susceptible humans
- Exposed humans
- Infectious humans
- Recovered humans
- Cumulative infectious
- New cases

Model vs. data comparison for different populations and cumulative infectious.
Monte Carlo convergence
Study of convergence for MC simulation

Stochastic dynamic model:

\[ \dot{U}(t, \omega) = f(U(\omega, t)) \]

Convergence metric for Monte Carlo simulation:

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

Study of convergence for MC simulation

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