Interval analysis for the treatment of uncertainty in epidemiological models based on ODEs

Diana Paola Lizarralde-Bejarano Ph.D. Student in Mathematical Engineering

> Supervisor: María Eugenia Puerta Co-supervisor: Sair Arboleda-Sánchez

> > August 12, 2019



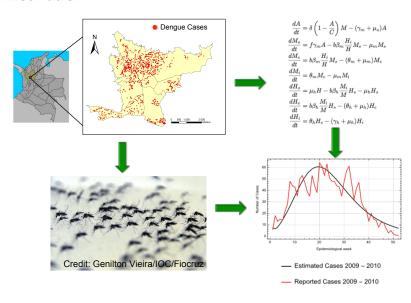
1 Motivation

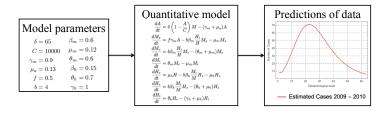
2 Problem statement

3 Framework

4 Conclusions

Motivation



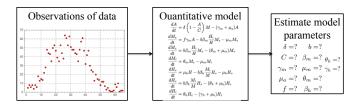


Get a solution of the system of equations given by

$$\frac{d\mathbf{y}}{dt} = \mathbf{f}(t; \mathbf{y}; \theta), \qquad \mathbf{y}(t_0) = x_0,$$

where $\mathbf{x} = (x_1, \dots, x_n)$ is the vector of n variables, $\theta = (\theta_1, \dots, \theta_p) \in \mathbb{R}^p$ is the vector of p parameters, $x_0 = (x_{1_0}, \dots, x_{n_0})$ is the vector of initial conditions, and t represents time.

Inverse problem



The optimization problem that we want to solve is given by

$$\min_{\theta \in \Omega \subseteq \mathbb{R}^p} d(y_i(t;\theta), x_i(t))$$
s.t
$$\frac{d\mathbf{y}}{dt} = \mathbf{f}(t;\mathbf{y};\theta), \ t \in [t_0, T]$$

$$\mathbf{y}(t_0) = y_0$$

where θ is the parameter vector, x_0 is the vector of initial conditions, and x_i is the data observed.

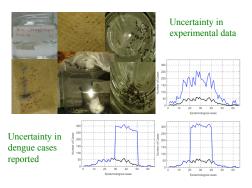
Experimental assays

	Temperature	Stage	Bello
Development	23°C	Larva 1+2	2.3±0.6
time (days)			
		Larva 3	1±0.0
		Larva 4	1±0.0
		Total larva	4.3±0.6
		Pupa	3.3±0.6
		Larva + Pupa	7.7±0.6
	27°C	Larva 1+2	2±0.0
		Larva 3	1±0.0
		Larva 4	1±0.0
		Total larva	4±0.0
		Pupa	1.7±0.6
		Larva + Pupa	5.7±0.6
Emerging time (days)	23°C	Male	10±1.0
		Female	10.3±0.6
	27°C	Male	7.3±0.6
		Female	8±0.0

	Temperature	Stage	Bello
Egg hatching (%)	_	Stage	62.6±4.7
	27°C		51.7±1.5
Immature survival (%)	23°C	Larva	85.2±8.4
		Pupa	98.1±1.7
	27°C	Larva	88.2±5.4
		Pupa	94.2±5.1
Sex ratio (%)	23°C	Male	48.9±4.2
		Female	51.1±4.2
	27°C	Male	55.2±2.7
		Female	44.8±2.7

Number of reported cases

The actual numbers of dengue cases are under-reported and many cases are misclassified. One recent estimate indicates 390 million dengue infections per year (95% confidence interval, 284 - 528 million), of which 96 million (67 - 136 million) manifest clinically (with any severity of disease)(Bhatt et al. 2013).



Uncertainty

Uncertainty is present in any process of measuring and obtaining information that is required to explain a real phenomenon. In many sciences it is possible to conduct experiments to obtain information and test hypotheses. Experiments on the spread of infectious diseases in human populations are often impossible, unethical or expensive (Hethcote, 2009).

Probability approximation

- This method is based on probability distributions of the parameters with uncertainty.
- Sufficient information on the uncertainty is not always available or sometimes expensive for many practical problems.
- There are works indicating that even a small deviation of the probability distribution is likely to cause a large error of the reliability analysis (Ben-Haim and Elishakoff, 2013).

Interval-valued approximation

- In the last two decades, the interval method in which *interval* is employed to model the uncertainty has been attracting more and more attention (Moore, Bierbaum, and Schwiertz, 1979; Braems et al. 2005; Hijazi et al. 2008).
- We only have to establish the bounds of the uncertainty of a parameter.
- Interval method has been successfully applied to uncertainty optimization problems (Jiang, Liu, and Han, 2008; Gallego-Posada and Puerta-Yepes, 2017).

How has uncertainty been considered in epidemiological models?

- (i) Probability theory
 - In (Luz et al. 2003) it is assumed that parameters as duration of infectious period in humans, biting rate, mosquito to human transmission and human to mosquito transmission follow a uniform distribution, while the extrinsic incubation period follows a triangular distribution. This information was obtained from several works that conducted experiments with different vector populations.
 - In (Britton and Lindenstrand, 2009) it is assumed that latent and infection periods are random and independent with the gamma distribution.

How has uncertainty been considered in epidemiological models?

- (i) Probability theory
 - In (Luz et al. 2003) it is assumed that parameters as duration of infectious period in humans, biting rate, mosquito to human transmission and human to mosquito transmission follow a uniform distribution, while the extrinsic incubation period follows a triangular distribution. This information was obtained from several works that conducted experiments with different vector populations.
 - In (Britton and Lindenstrand, 2009) it is assumed that latent and infection periods are random and independent with the gamma distribution.
- (ii) Fuzzy theory: In (Barros et al. 2001) an SIS model was formulated, where the transmission and recovery rate are given by functions that depend on the amount of virus.

Epidemiological data

The parameters used in the model, their biological descriptions, and their ranges of values.

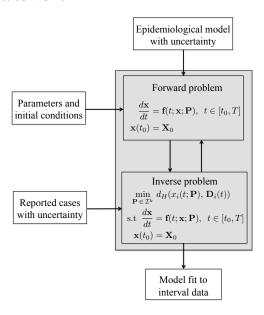
Param.	Meaning	V. / day	V. / week
ь	Biting rate	[0,1]	[0,4]
δ	Per capita oviposition rate	[8, 24]	[55, 165]
γ_m	Transition rate from the aquatic phase to the adult phase	[0.125, 0.2]	[0.875, 1.4]
μ_a	Mortality rate in the aquatic phase	[0.001, 0.5]	[0.007, 0.3]
μ_m	Mortality rate in the adult phase	[0.008, 0.03]	[0.06, 0.20]
f	Fraction of female mosquitoes hatched from all eggs	[0.42, 0.55]	[0.42, 0.55]
С	Carrying capacity of the environment	[6400,95000]	[6400,95000]
μ_h	Birth and death rate of the human population	0.00006	0.0004
β_h	Transmission probability from mosquito to human	[0, 1]	[0,1]
β_m	Transmission probability from human to mosquito	[0, 1]	[0,1]
θ_m	Transition rate from exposed to infectious mosquitoes	[0.08, 0.13]	[0.58, 0.88]
θ_h	Transition rate from exposed to infectious humans	[0.1, 0.25]	[0.7, 1.75]
γ_h	Recovery rate	[0.07, 0.25]	[0.5, 1.75]

Initial conditions

The initial conditions used in the model, their descriptions, and their ranges of values.

Initial condition	Meaning	Range		
A(0)	Initial condition for the aquatic phase	[5755, 17265]		
$M_s(0)$	Initial condition for susceptible mosquitoes	[0,1200000]		
$M_e(0)$	Initial condition for exposed mosquitoes	[0,100]		
$M_i(0)$	Initial condition for infectious mosquitoes	[0,100]		
$H_s(0)$	Initial condition for susceptible humans	[244402, 321734]		
$H_e(0)$	Initial condition for exposed humans	[18,72]		
$H_i(0)$	Initial condition for infectious humans	[6,24]		
$H_r(0)$	Initial condition for recovered humans	[81405, 158809]		

Problem statement



Consider the initial value problem

$$y'(t) = f(y,\theta)$$

$$y(t_0) = y_0 \in Y_0, \ \theta \in \Theta$$
(1)

where $t \in [t_0, t_m]$, θ is a p-dimensional vector of parameters, y is the n-dimensional vector of state variables, y_0 is the n-dimensional vector of initial values.

- The basis for solving problems of this type is the approximation by Taylor models, $T_f = p_f + r_f$ (Makino and Berz, 2003).
- The algorithms applied to solve (1) are divided into two stages. The first stage validates the existence and uniqueness of the solution, and the second stage computes a tighter enclosure where the solution is found (Lin and M. A. Stadtherr, 2007).
- In (Nedialkov, Jackson, and Corliss, 1999), the authors presented a review of the methods used to solve (1) and difficulties that may arise.

- In (Lin and M. A. Stadtherr, 2007), they considered autonomous systems with initial conditions and parameters given by intervals as the system (1).
- In (Enszer and M. Stadtherr, 2009) they used the VSPODE, for propagating uncertainties through nonlinear ODE models in population epidemiology.
- Softwares: VNODE, COSY INFINITY, VSPODE, among others.

Example taken from (Enszer and M. Stadtherr, 2009)

Consider the SIRS model. We assume a constant total population n = s + i + r.

$$\frac{ds}{dt} = -\beta si + \gamma (n - s - i)$$

$$\frac{di}{dt} = \beta si - \nu i$$

Example taken from (Enszer and M. Stadtherr, 2009)

For this example, we obtain

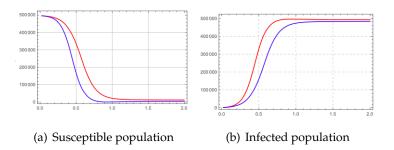


Figure: Initial conditions: n = 500000 (indv), s(0) = 498000 (indv), i(0) = 2000 (indv). Parameters: $\gamma = 50$ (yr⁻¹), $\beta \in [2, 2.5] \times 10^{-5}$ yr⁻¹indv⁻¹, $\nu \in [0.125, 0.250]$ (yr⁻¹).

Example: Dengue model, Bello's case

$$\frac{dA}{dt} = \delta \left(1 - \frac{A}{C}\right) M - (\gamma_m + \mu_a) A$$

$$\frac{dM_s}{dt} = f \gamma_m A - b \beta_m \frac{H_i}{H} M_s - \mu_m M_s$$

$$\frac{dM_e}{dt} = b \beta_m \frac{H_i}{H} M_s - (\theta_m + \mu_m) M_e$$

$$\frac{dM_i}{dt} = \theta_m M_e - \mu_m M_i$$

$$\frac{dH_s}{dt} = \mu_h H - b \beta_h \frac{M_i}{M} H_s - \mu_h H_s$$

$$\frac{dH_e}{dt} = b \beta_h \frac{M_i}{M} H_s - (\theta_h + \mu_h) H_e$$

$$\frac{dH_e}{dt} = \theta_h H_e - (\gamma_h + \mu_h) H_i$$

$$\frac{dH_r}{dt} = \gamma_h H_i - \mu_h H_r$$

Example: Dengue model, Bello's case

Consider $B_m = b\beta_m \in [2.2, 2.5]$ and $B_h = b\beta_h \in [0.55, 0.65]$.

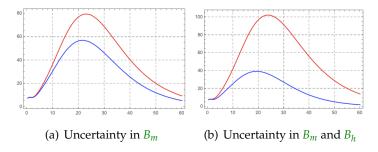


Figure: Parameter values: $\delta=65$, $\gamma_m=0.9$, $\mu_a=0.13$, $\mu_m=0.12$, b=4, $\theta_m=0.6$, f=0.5, $\theta_h=0.7$, C=10000, $\gamma_h=1$, $\beta_h=0.15$, and $\mu_h=0.0004$. Initial conditions: A(0)=9000, $M_s(0)=1199950$, $M_e(0)=40$, $M_i(0)=10$, $H_s(0)=321710$, $H_e(0)=18$, $H_i(0)=6$, and $H_r(0)=81501$.

Example: Dengue model, Bello's case

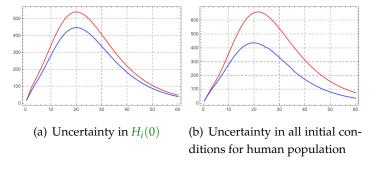
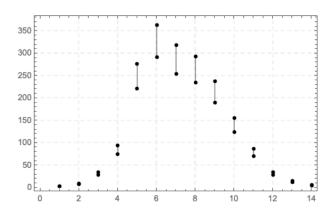
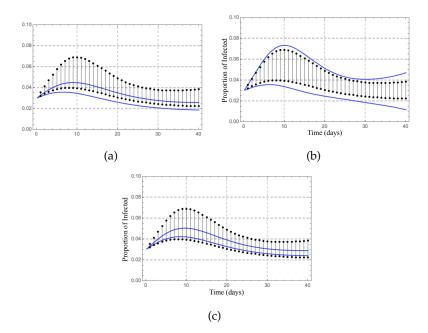


Figure: Parameter values: $\delta = 65$, $\gamma_m = 0.9$, $\mu_a = 0.13$, $\mu_m = 0.12$, b = 4, $\theta_m = 0.6$, f = 0.5, $\theta_h = 0.7$, C = 10000, $\gamma_h = 1$, $\beta_h = 0.15$, and $\mu_h = 0.0004$. Initial conditions: A(0) = 9000, $M_s(0) = 1199950$, $M_e(0) = 40$, and $M_i(0) = 10$. In (a) $H_s(0) = 321710$, $H_e(0) = 18$, $H_i(0) \in [6,12]$. In (b) $H_s(0) \in [321710,369967]$, $H_e(0) \in [18,20.7]$, $H_i(0) \in [6,6.9]$.

Inverse problem

If we have interval data as input. What is that meaning that the output of the model "fit" or be compatible with the interval data?





Notation

- Let be $Z = (Y_0, \Theta)$ a p-dimensional vector where its first m coordinates correspond to the interval initial conditions of the model in (1) and its last p-m coordinates correspond to interval parameters of the aforementioned model.
- Consider the vector $X_i = (X_{i_1}, X_{i_2}, \dots, X_{i_k})$ as the interval experimental data regarding the model in (1) where i is the number of variables in the model ranging from 1 to n, and the number of measurements at different times is denoted by k.
- The vector $Y_i = (Y_{i_1}, Y_{i_2}, ..., Y_{i_k})$ is the enclosure of the system solution in (1) for the variable i at the time t_k .

Definition

The vector Z is **compatible** with an interval experimental data $X_{i_1}, X_{i_2}, \ldots, X_{i_k}$, if for each measurement k, there are representatives $x_{i_1} \in X_{i_1}, \ldots, x_{i_k} \in X_{i_k}$ such that $x_{i_1} \in Y_{i_1}, \ldots, x_{i_k} \in Y_{i_k}$.

This set of vectors is called *united solution set* of the model (1).

$$E_{unit}(\mathbf{X}, \mathbf{Y}) = \{ Z \in \mathcal{I}^p \mid \text{there exist } \mathbf{x} \in \mathbf{X} \text{ such that } \mathbf{x} \in \mathbf{Y} \}$$

where **X** and **Y** are interval matrices, and **x** is a real matrix. All these matrices have the same dimension, $k \times n$.

Definition

The vector Z is **strongly compatible** with the interval experimental data $X_{i_1}, X_{i_2}, \ldots, X_{i_k}$ if for each measurement k, and for any representatives $x_{i_1} \in X_{i_1}, \ldots, x_{i_k} \in X_{i_k}$ there holds that $x_{i_1} \in Y_{i_1}, \ldots, x_{i_k} \in Y_{i_k}$.

The set composed of Z vectors that satisfy the previous definition is called *tolerable solution set* of the model (1).

$$E_{tol}(\mathbf{X}, \mathbf{Y}) = \{ Z \in \mathcal{I}^p \mid \text{for any } \mathbf{x} \in \mathbf{X}, \ \mathbf{x} \in \mathbf{Y} \}$$
 (2)

Observation

$$E_{tol}(\mathbf{X}, \mathbf{Y}) \subseteq E_{unit}(\mathbf{X}, \mathbf{Y})$$

How to know if a vector Z belongs to the *tolerable* solution set?

We have to check that $X_{i_1} \subseteq Y_{i_1}, X_{i_2} \subseteq Y_{i_2}, \ldots, X_{i_k} \subseteq Y_{i_k}$ for each variable i.

Let $X_{i_j} = [\underline{X_{i_j}}, \overline{X_{i_j}}]$ and $Y_{i_j} = [\underline{Y_{i_j}}, \overline{Y_{i_j}}]$. We said X_{i_j} is subset of Y_{i_j} if

$$\min_{1 \le i \le n} \min_{1 \le j \le k} \left\{ \left(\underline{X_{i_j}} - \underline{Y_{i_j}} \right), \left(\overline{Y_{i_j}} - \overline{X_{i_j}} \right) \right\} \ge 0 \tag{3}$$

or equivalently if

$$\min_{1 \leq i \leq n} \min_{1 \leq j \leq k} \left\{ \operatorname{rad} Y_{i_j} - \operatorname{rad} X_{i_j} - \left| \operatorname{mid} X_{i_j} - \operatorname{mid} Y_{i_j} \right| \right\} \geq 0 \quad (4)$$

where,
$$X = [\underline{X}, \overline{X}] = mid X + [-rad X, rad X]$$

Example

Consider the sir model normalized, i.e., $0 \le s \le 1$, $0 \le i \le 1$ and $0 \le r \le 1$, where μ , β , and γ are mortality rate, transmission probability and recovery rate, respectively. Also the population is constant, s+i+r=1.

$$\frac{ds}{dt} = \mu - \beta si - \mu s$$

$$\frac{di}{dt} = \beta si - (\mu + \gamma)i$$

$$\frac{dr}{dt} = \gamma i - \mu r$$

this system can be reduced to the two first equations

$$\frac{ds}{dt} = \mu - \beta si - \mu s$$

$$\frac{di}{dt} = \beta si - (\mu + \gamma)i$$
(5)

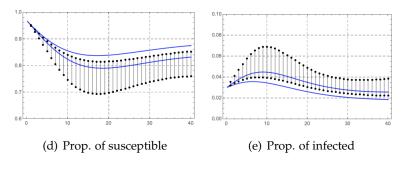


Figure: Initial conditions: s(0) = 0.97, i(0) = 0.03. Parameters: $\mu = 0.1, \gamma = 0.5, \beta \in [0.68, 0.72]$.

For (4),

$$\min_{1 \le i \le 2} \{ -0.0974893, -0.0243558 \} = -0.0974893 \le 0$$

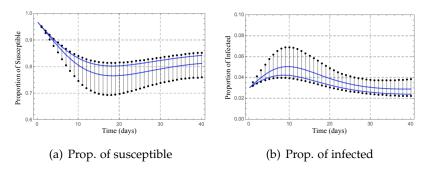


Figure: Initial conditions: s(0)=0.97, i(0)=0.03. Parameters: $\mu=0.1, \gamma=0.5, \beta\in[0.71$, 0.74].

For (4),

$$\min_{1 \le i \le 2} \left\{ -0.0733967, \, -0.0186444 \right\} = -0.0733967 \le 0$$

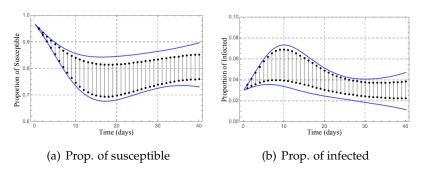
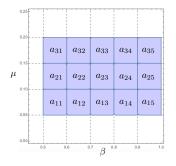


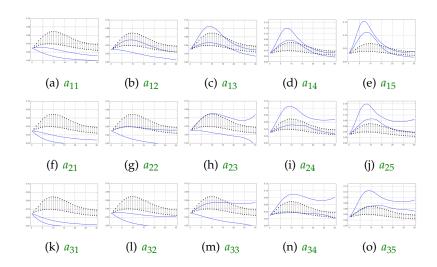
Figure: Initial conditions: s(0) = 0.97, i(0) = 0.03. Parameters: $\mu = 0.1, \gamma = 0.5, \beta \in [0.68, 0.81].$

$$\min_{1 \le i \le 2} \{ 0.000414505, 0.000336667 \} = 0.000336667 \ge 0$$

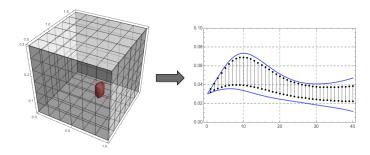
If we want to explore all the parameter space? How can we do that?



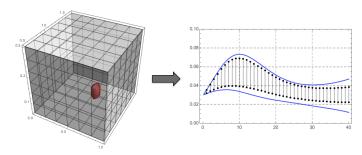
Region	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}	a_{21}	a ₂₂	a ₂₃	a_{24}	a_{25}	a_{31}	a ₃₂	a_{33}	a_{34}	a ₃₅
Susceptibles	-0.132	-0.018	-0.066	-0.191	-0.297	-0.218	-0.119	0	-0.021	-0.132	-0.265	-0.199	-0.092	0.001	-0.051
Infected	-0.042	-0.017	-0.009	-0.038	-0.072	-0.050	-0.030	0	-0.016	-0.047	-0.058	-0.041	-0.015	-0.005	-0.035



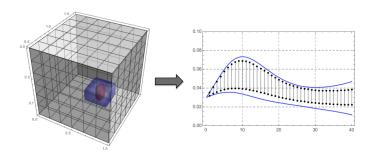
First stage



First stage



Second stage



Input:

- \blacksquare *M*, the model given by (1).
- **2** $X \in \mathcal{I}^{s \times k}$, the interval experimental data matrix where s corresponds to the number of variables and k corresponds to the number of measurements for each variable.
- 3 $Z = (Y_0, \Theta) \in \mathcal{I}^p$, where its first m coordinates correspond to the interval initial conditions and its last p m coordinates correspond to interval parameters of the model.
- 4 *m*, the number of intervals for each component of the interval vector *Z*.

Output:

■ *W* is the set of vectors *Z* which are strongly compatible with the interval data **X**.

Algorithm 1 GSearch of strongly compatible vectors with IData

```
procedure TOLERABLE SOLUTION SET
      W \leftarrow \emptyset
      for k = 1 to m do
             for i = 1 to p do
                  Z_{i_k} \leftarrow \left[ \underline{Z_i} + k \left( \frac{\overline{Z_i} - \underline{Z_i}}{m} \right), \underline{Z_i} + (k+1) \left( \frac{\overline{Z_i} - \underline{Z_i}}{m} \right) \right]
             end for
            Z_k \leftarrow [Z_{1_k}, Z_{2_k}, \ldots, Z_{p_k}]
            \mathbf{Y}_k \leftarrow \mathsf{VSPODE}(Z_k, M)
            if Tol(Z_k, X, Y_k) \geq 0 then
                   W \leftarrow W \cup \{Z_k\}
             end if
      end for
end procedure
```

■ This work provides an original method to determine the type of compatibility between the interval-data and the vector of initial conditions and parameters of ODEs which can only be solved numerically.

- This work provides an original method to determine the type of compatibility between the interval-data and the vector of initial conditions and parameters of ODEs which can only be solved numerically.
- We propose an algorithm to estimate interval parameters in order to fit the model given by (1) to interval-data. The proposed framework, based on a measure of compatibility (4) is generic, reliable and simple to use.

■ We already implemented the first stage of the proposed algorithm in *Mathematica* software. For this, we constructed a link between the VSPODE and *Mathematica*.

- We already implemented the first stage of the proposed algorithm in *Mathematica* software. For this, we constructed a link between the VSPODE and *Mathematica*.
- We expect with this algorithm to be capable of estimating solutions more robust for models which simulated the transmission of dengue disease in different regions of Colombia.

Thank you!

Questions?

References

Barros, L.C., R.C. Bassanezi, R. Zotin G. Oliveira, and M.B.F Leite (2001). A disease evolution model with uncertain parameters. In: IFSA World Congress and 20th NAFIPS International Conference, 2001. Joint 9th. Vol. 3. IEEE, pp. 1626–1630 (cit. on pp. 11, 12).

Ben-Haim, Yakov and Isaac Elishakoff (2013). Convex models of uncertainty in applied mechanics. Vol. 25. Elsevier (cit. on p. 9).

Bhatt, Samir et al. (2013). The global distribution and burden of dengue. Nature 496.7446, p. 504 (cit. on p. 7).

Braems, I et al. (2005). New set-membership techniques for parameter estimation in presence of model uncertainty. In: Proceedings of the 5th International Conference on Inverse Problems in Engineering: Theory and Practice. Vol. 11, 15th (cit. on p. 10).

Britton, Tom and David Lindenstrand (2009). Epidemic mod-

sciences 222.2, pp. 109–116 (cit. on pp. 11, 12).

Enszer, Joshua and Mark Stadtherr (2009). Verified solution method for population epidemiology models with uncertainty. International Journal of Applied Mathematics and Computer Science 19.3, pp. 501–512 (cit. on pp. 18–20).

elling: aspects where stochasticity matters. Mathematical bio-

Gallego-Posada, Jose Daniel and María Eugenia Puerta-Yepes (2017). Internal analysis and optimization applied to parameter estimation under uncertainty. Boletim da Sociedade Paranaense de Matemática 36.2, pp. 107–124 (cit. on p. 10). Hethcote, Herbert W (2009). The basic epidemiology models: models, expressions for *R*₀, parameter estimation, and appli-

cations. In: Mathematical understanding of infectious disease

- dynamics. World Scientific, pp. 1–61 (cit. on p. 8).

 Hijazi, Younis, Hans Hagen, Charles D Hansen, and Kenneth I Joy (2008). Why interval arithmetic is so useful. Visualization of large and unstructured data sets (cit. on p. 10).
- Jiang, C, GR Liu, and X Han (2008). A novel method for uncertainty inverse problems and application to material characterization of composites. Experimental Mechanics 48.4, pp. 539–548 (cit. on p. 10).

- Lin, Youdong and Mark A Stadtherr (2007). Validated solutions of initial value problems for parametric ODEs. Applied Numerical Mathematics 57.10, pp. 1145–1162 (cit. on pp. 17, 18).

 Luz, Paula Mendes, Cláudia Torres Codeco, Eduardo Massad,
- Luz, Paula Mendes, Cláudia Torres Codeço, Eduardo Massad, and Claudio José Struchiner (2003). Uncertainties regarding dengue modeling in Rio de Janeiro, Brazil. Memórias do Instituto Oswaldo Cruz 98.7, pp. 871–878 (cit. on pp. 11, 12).
- Makino, Kyoko and Martin Berz (2003). Taylor models and other validated functional inclusion methods. International Journal of Pure and Applied Mathematics 4.4, pp. 379–456 (cit. on p. 17).
 - Moore, Ramon E, Fritz Bierbaum, and Klaus-Peter Schwiertz (1979). Methods and applications of interval analysis. Vol. 2. SIAM (cit. on p. 10).

Nedialkov, Nedialko S, Kenneth R Jackson, and George F Corliss (1999). Validated solutions of initial value problems for ordinary differential equations. Applied Mathematics and Computation 105.1, pp. 21–68 (cit. on p. 17).