End-to-end Statistical Model Checking for ODEs

David JULIEN
david.julien@univ-nantes.fr

ENS Lyon / Nantes Université - LS2N (VELO)
B. DELAHAYE, G. CANTIN

Oct. 20, 2022

ENS DE LYON

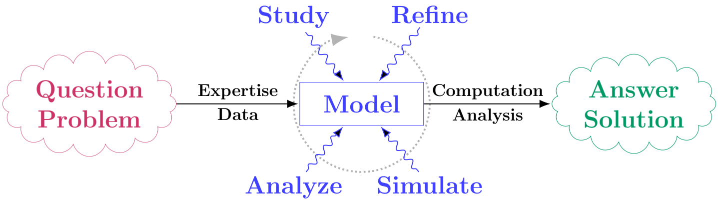
LABORATOIRE DES SCIENCES DU NUMÉRIQUE DE NANTES

Nantes Université

Outline

Outline

Science of Models

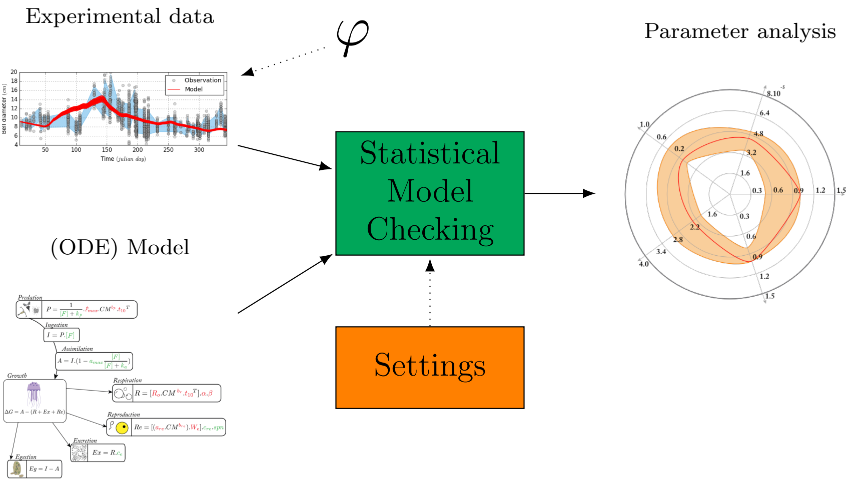


Motivations

Provide mathematical **guarantees** and **tools** for building and analyzing models

- usable by any scientist;
- answering real-life questions.

Example: Parameterization of a jellyfish model [1]



[1] [Ramondec et al. 2020] Probabilistic modeling to estimate jellyfish ecophysiological properties and size distribution. *Scientific Reports*

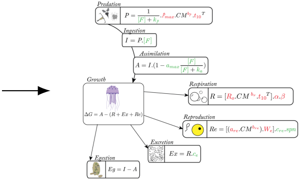
Formal verification of ODE models

System

Formal verification of ODE models

System

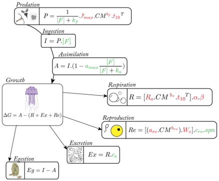
(ODE) Model



Formal verification of ODE models

System

(ODE) Model

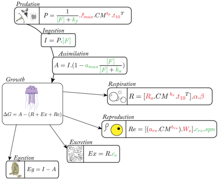


φ

Formal verification of ODE models

System

(ODE) Model



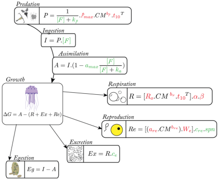
Some models may be verified directly (automata, graphs...).

φ

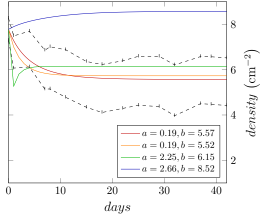
Formal verification of ODE models

System

(ODE) Model



Set of traces

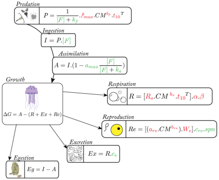


φ

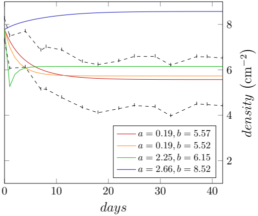
Formal verification of ODE models

System

(ODE) Model



Set of traces

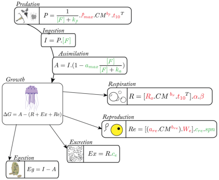


φ

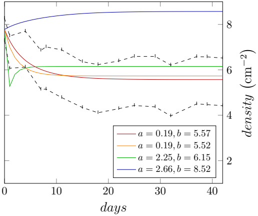
Formal verification of ODE models

System

(ODE) Model



Set of traces

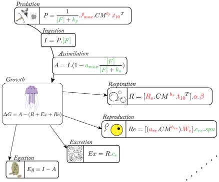


φ

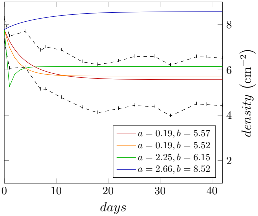
Formal verification of ODE models

System

(ODE) Model



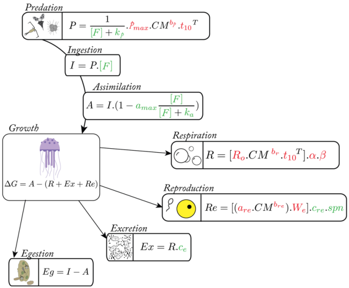
Set of traces



If Model ∼ Traces
⇒ Checking φ on the traces is equivalent to checking it on the model.

φ

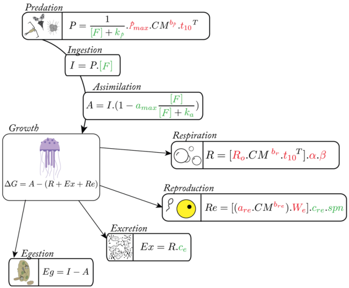
Probabilistic models



Uncertainty and variability

- Several experiments
  ⇒ data uncertainty.
- Family of systems
  ⇒ parameters variability.

Probabilistic models



Uncertainty and variability

- Several experiments
  $\Rightarrow$ data uncertainty.
- Family of systems
  $\Rightarrow$ parameters variability.

$\Rightarrow$ Parametric model with probabilistic parameter values.

Outline

SMC: answering "how good is a model?"

The Monte-Carlo procedure:

1. Randomly generate N samples $(\sigma_1, \ldots, \sigma_N)$ from the model.
2. Check whether the sample i satisfies the property φ.
   $X_i = 1 \Leftrightarrow \sigma_i \models \varphi$
3. Compute the estimator $\hat{p} = \frac{\sum X_i}{N}$.
   $\Rightarrow \mathbb{P}(p \in [\hat{p} - \alpha, \hat{p} + \alpha]) \geq 1 - \theta$

SMC: answering “how good is a model?”

The Monte-Carlo procedure:

1. Randomly generate N samples $(\sigma_1, \ldots, \sigma_N)$ from the model.
2. Check whether the sample i satisfies the property φ.
   $X_i = 1 \Leftrightarrow \sigma_i \models \varphi$
3. Compute the estimator $\hat{p} = \frac{\sum X_i}{N}$.
   $\Rightarrow \mathbb{P}(p \in [\hat{p} - \alpha, \hat{p} + \alpha]) \geq 1 - \theta$

Central Limit Theorem

- $\hat{p} \sim \mathbb{E}(X)$
  $\Rightarrow \hat{p}$ is a good estimation of $\mathbb{P}(M \models \varphi)$.
- **precision, error.**

Application to ODE models

[Liu et al. 2019] Statistical Model Checking-Based Analysis of Biological Networks.

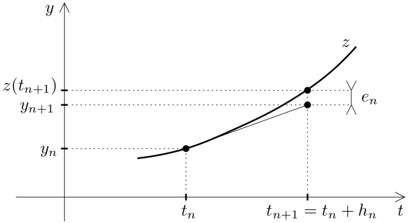
Automated Reasoning for Systems Biology and Medecine

Parameter analysis of ODE models with variability.

Problem: Approximations

Model $\not\approx$ Traces

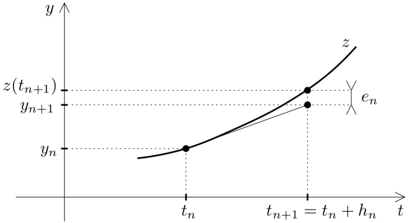
Approximation error stacks up at each step: $\varepsilon = \max_n e_n$
$\Rightarrow$ SMC estimation does not apply to the original ODE model.

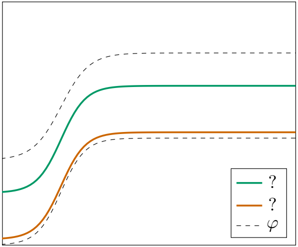


Problem: Approximations

Model ≁ Traces

Approximation error stacks up at each step: $\varepsilon = \max_n e_n$
$\Rightarrow$ SMC estimation does not apply to the original ODE model.

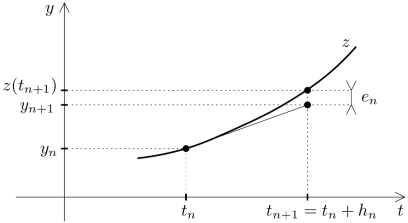


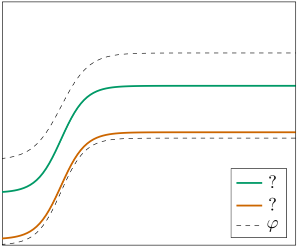


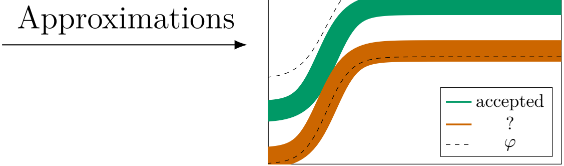
Problem: Approximations

Model ≁ Traces

Approximation error stacks up at each step: $\varepsilon = \max_n e_n$
$\Rightarrow$ SMC estimation does not apply to the original ODE model.







Solution: Safety margins

- Bound approximation error ε on the parameter space.
- Define new properties φ_1 and φ_2.

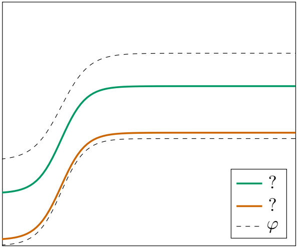
$\Rightarrow$ Given a trace σ generated from the model M,

Solution: Safety margins

- Bound approximation error ε on the parameter space.
- Define new properties φ_1 and φ_2.

$\Rightarrow$ Given a trace σ generated from the model M,

$$\sigma \models \varphi_1 \Rightarrow M \models \varphi$$

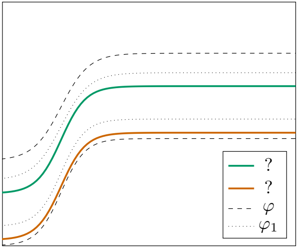


Solution: Safety margins

- Bound approximation error ε on the parameter space.
- Define new properties φ_1 and φ_2.

$\Rightarrow$ Given a trace σ generated from the model M,

$$\sigma \models \varphi_1 \Rightarrow M \models \varphi$$

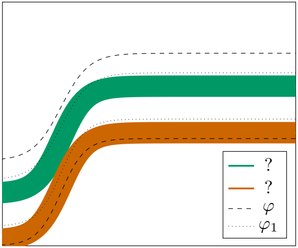


Solution: Safety margins

- Bound approximation error ε on the parameter space.
- Define new properties φ_1 and φ_2.

$\Rightarrow$ Given a trace σ generated from the model M,

$$\sigma \models \varphi_1 \Rightarrow M \models \varphi$$

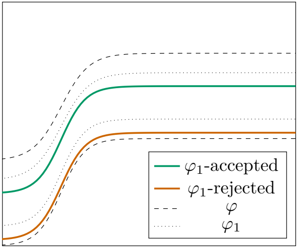


Solution: Safety margins

- Bound approximation error ε on the parameter space.
- Define new properties φ_1 and φ_2.

$\Rightarrow$ Given a trace σ generated from the model M,

$$\sigma \models \varphi_1 \Rightarrow M \models \varphi$$

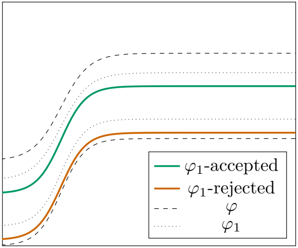


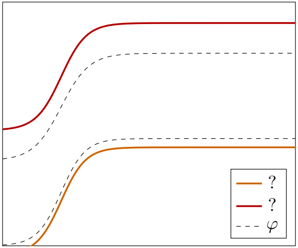
Solution: Safety margins

- Bound approximation error ε on the parameter space.
- Define new properties φ_1 and φ_2.

$\Rightarrow$ Given a trace σ generated from the model M,

$$M \models \varphi \Rightarrow \sigma \models \varphi_2$$



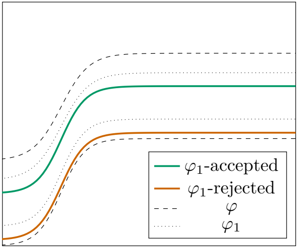


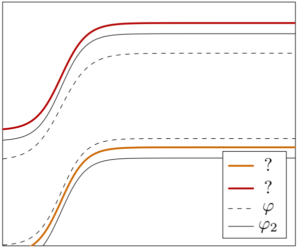
Solution: Safety margins

- Bound approximation error ε on the parameter space.
- Define new properties φ_1 and φ_2.

$\Rightarrow$ Given a trace σ generated from the model M,

$$M \models \varphi \Rightarrow \sigma \models \varphi_2$$



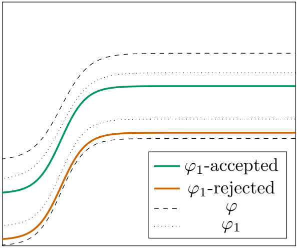


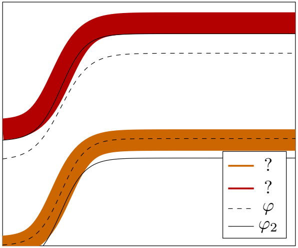
Solution: Safety margins

- Bound approximation error ε on the parameter space.
- Define new properties φ_1 and φ_2.

$\Rightarrow$ Given a trace σ generated from the model M,

$$M \models \varphi \Rightarrow \sigma \models \varphi_2$$



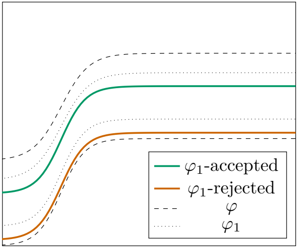


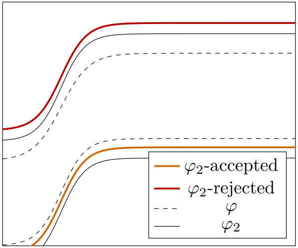
Solution: Safety margins

- Bound approximation error ε on the parameter space.
- Define new properties φ_1 and φ_2.

$\Rightarrow$ Given a trace σ generated from the model M,

$$\sigma \models \varphi_1 \Rightarrow M \models \varphi \Rightarrow \sigma \models \varphi_2 \quad \Rightarrow \quad p_1 \leq p \leq p_2$$





Guarantees: global risk ξ and precision α

Usual SMC

- $n = \frac{\log(2/\xi)}{2\alpha^2}$ simulations.

$\Rightarrow \mathbb{P}(p \in [\hat{p} - \alpha, \hat{p} + \alpha]) \geq 1 - \xi$

In our case (for each property)

- SMC risk $\theta = 1 - \sqrt{1 - \xi} < \xi$
- $n' = \frac{\log(2/\theta)}{2\alpha^2} > n$ simulations

$$\Rightarrow \mathbb{P}(p_1 \in [\hat{p_1} - \alpha, \hat{p_1} + \alpha]) \geq 1 - \theta, \quad \mathbb{P}(p_2 \in [\hat{p_2} - \alpha, \hat{p_2} + \alpha]) \geq 1 - \theta$$

Guarantees: global risk ξ and precision α

Usual SMC

- $n = \frac{\log(2/\xi)}{2\alpha^2}$ simulations.

$\Rightarrow \mathbb{P}(p \in [\hat{p} - \alpha, \hat{p} + \alpha]) \geq 1 - \xi$

In our case (for each property)

- SMC risk $\theta = 1 - \sqrt{1 - \xi} < \xi$
- $n' = \frac{\log(2/\theta)}{2\alpha^2} > n$ simulations

$$\Rightarrow \mathbb{P}(p_1 \in [\hat{p_1} - \alpha, \hat{p_1} + \alpha]) \geq 1 - \theta, \quad \mathbb{P}(p_2 \in [\hat{p_2} - \alpha, \hat{p_2} + \alpha]) \geq 1 - \theta$$

Theorem 2 (Main theorem).

After performing $N = 2 \times n'$ simulations, the following statements hold:

- $\mathbb{P}(p \in [\hat{p_1} - \alpha, \hat{p_2} + \alpha]) \geq 1 - \xi$;
- $\mathbb{P}(|\hat{p_1} - \hat{p_2}| \leq 3\alpha) \geq 1 - \xi$.

Bonus: extension to reward functions.

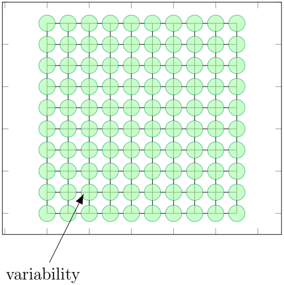
Outline

Generalities

- Parameterization: Find parameter values λ for a generic model.
- Goal: Find good values for λ w.r.t. score ($\mathbb{E}(r)$) of φ-satisfaction.
- Any algorithm:
  - Local search: low execution time / superficial search;
  - **Global** search: more informative / higher execution time;
  - ...

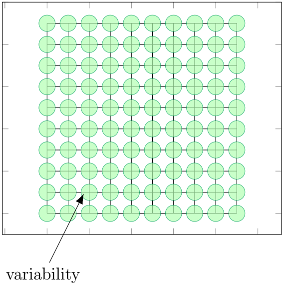
What we do

1. Compute the grid of parameters.
2. Compute the score of each value.
3. Select the best value.



What we do

1. Compute the grid of parameters.
2. Compute the score of each value.
3. Select the best value.

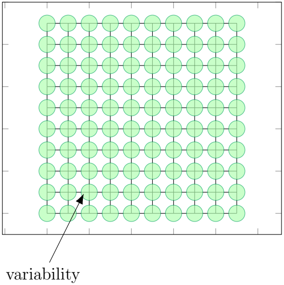


Reminder

It only works if we can bound the error ε!

What we do

1. Compute the grid of parameters.
2. Compute the score of each value.
3. Select the best value.



Reminder

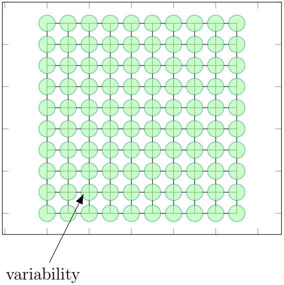
It only works if we can bound the error ε!

Proposition

When solving an ODE, the error ε is bounded by a function of the integration step h.

What we do

1. Compute the grid of parameters.
2. Compute the score of each value.
3. Select the best value.



Reminder

It only works if we can bound the error ε!

Lemma 1.

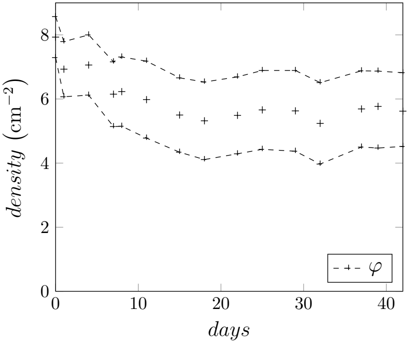
For any arbitrary $\varepsilon > 0$, there exists an integration step h such that

$$0 < \varepsilon_h < \varepsilon, \quad \forall \lambda.$$

Aurelia Aurita[2]

Jellyfish species from the Adriatic Sea.

- $x'(t) = a \cdot x(t) \cdot \left(1 - \frac{x(t)}{b}\right)$
- φ = data $\pm$ standard error

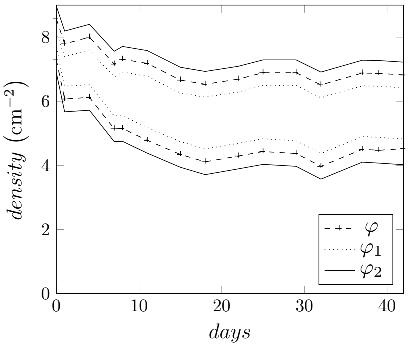


[2] [Melica et al. 2014] Logistic density-dependent growth of an aurelia aurita polyps population. *Ecological Modeling*

Aurelia Aurita[2]

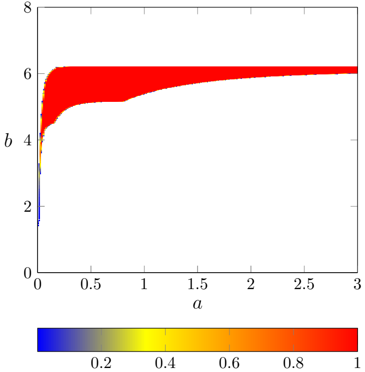
Jellyfish species from the Adriatic Sea.

- $x'(t) = a \cdot x(t) \cdot \left(1 - \frac{x(t)}{b}\right)$
- φ = data $\pm$ standard error
- φ_1 = "$\varphi - \varepsilon$"
- φ_2 = "$\varphi + \varepsilon$"

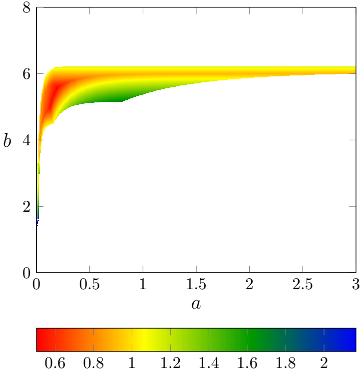


[2] [Melica et al. 2014] Logistic density-dependent growth of an aurelia aurita polyps population. *Ecological Modeling*

Parameter analysis



Probability of staying in the tunnel.



Expected distance to data.

Outline

Perspectives

Mathematics

- Convergence speed of $\hat{p_1} - \hat{p_2}$.
- Suitable values for ε and h in the general case.

Perspectives

Mathematics

- Convergence speed of $\hat{p}_1 - \hat{p}_2$.
- Suitable values for ε and h in the general case.

General scientific community

- Tool.
- Larger case studies.
  $\rightarrow$ mixing models (humans/forest, epidemiology, ...)

Perspectives

Mathematics

- Convergence speed of $\hat{p}_1 - \hat{p}_2$.
- Suitable values for ε and h in the general case.

General scientific community

- Tool.
- Larger case studies.
  → mixing models (humans/forest, epidemiology, ...)

Modeling

- **Structural properties** of ODE models.
  → stability, attraction, cycles ...
- **Dynamic** discretization of the parameter grid.
  → variance, score

Thank you for your attention !