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
Outline

1. Context
2. SMC for ODE models
3. Application to parameterization
4. Perspectives
Outline

1. Context
2. SMC for ODE models
3. Application to parameterization
4. Perspectives
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

---

[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

\[ P = \frac{1}{1 + e^{-5 \tau_{\text{p}} \cdot t \cdot \tau_{\text{c}}}} \]

\[ f = P' \]

\[ S = f \cdot (1 - \frac{1}{1 + e^{-5 \tau_{\text{p}} \cdot t \cdot \tau_{\text{c}}}}) \]

\[ \Delta \Gamma = A \cdot (S + E \cdot R) \]

\[ E = R \cdot r \]

\[ R = [R_{\text{p}, CM^{p}, t \cdot \tau_{\text{c}}}] \cdot \alpha \cdot \beta \]

\[ P = [\alpha_{\text{cm}}, CM^{p}, t \cdot \tau_{\text{c}}], \alpha \cdot \beta \]

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

\( a = 0.19, b = 5.57 \)

\( a = 0.19, b = 5.52 \)

\( a = 2.25, b = 6.15 \)

\( a = 2.66, b = 8.52 \)
Formal verification of ODE models

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

\[ \text{System} \quad \xrightarrow{\text{ODE Model}} \quad \phi \]

Set of traces

If \( \text{Model} \sim \text{Traces} \Rightarrow \text{Checking} \ \phi \) on the traces is equivalent to checking it on the model.

\[ a = 0.19, b = 5.57 \]

\[ a = 0.19, b = 5.52 \]

\[ a = 2.25, b = 6.15 \]

\[ a = 2.66, b = 8.52 \]
Formal verification of ODE models

Some models may be verified directly (automata, graphs...).
Formal verification of ODE models

System

(ODE) Model

Set of traces

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

\[ a = 0.19, b = 5.57 \]
\[ a = 0.19, b = 5.52 \]
\[ a = 2.25, b = 6.15 \]
\[ a = 2.66, b = 8.52 \]

\[ \varphi \]
Formal verification of ODE models

(Fixed) Model

\[
\phi
\]

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

\[
\begin{align*}
a &= 0.19, \quad b = 5.57 \\
a &= 0.19, \quad b = 5.52 \\
a &= 2.25, \quad b = 6.15 \\
a &= 2.66, \quad b = 8.52
\end{align*}
\]

Set of traces

If \( \text{Model} \sim \text{Traces} \Rightarrow \text{Checking } \phi \text{ on the traces is equivalent to checking it on the model.} \)

\[
\psi
\]
Formal verification of ODE models

System

(ODE) Model

Set of traces

\[ \phi \]

\[ \begin{align*}
  a &= 0.19, b = 5.57 \\
  a &= 0.19, b = 5.52 \\
  a &= 2.25, b = 6.15 \\
  a &= 2.66, b = 8.52
\end{align*} \]
Formal verification of ODE models

System

(ODE) Model

Set of traces

If Model $\sim$ Traces
$\Rightarrow$ Checking $\varphi$ on the traces is equivalent to checking it on the model.
Probabilistic models

Uncertainty and variability

- Several experiments \( \Rightarrow \) data uncertainty.
- Family of systems \( \Rightarrow \) parameters variability.
Probabilistic models

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

⇒ Parametric model with probabilistic parameter values.
Outline

1 Context

2 SMC for ODE models

3 Application to parameterization

4 Perspectives
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 $\varphi$.
   
   $X_i = 1 \Leftrightarrow \sigma_i \models \varphi$

3. Compute the estimator $\hat{p} = \frac{\sum X_i}{N}$.
   
   $\Rightarrow P(p \in [\hat{p} - \alpha, \hat{p} + \alpha]) \geq 1 - \theta$

Central Limit Theorem

$\hat{p} \sim \mathbb{E}(X)$

$\Rightarrow \hat{p}$ is a good estimation of $P(M | = \varphi)$.

Application to ODE models


Parameter analysis of ODE models with variability.
SMC for ODE models

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 $\varphi$.
   
   $X_i = 1 \Leftrightarrow \sigma_i \models \varphi$

3. Compute the estimator $\hat{p} = \frac{\sum X_i}{N}$.
   
   $\Rightarrow P(p \in [\hat{p} - \alpha, \hat{p} + \alpha]) \geq 1 - \theta$

Central Limit Theorem

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

Application to ODE models

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

Automated Reasoning for Systems Biology and Medicine

Parameter analysis of ODE models with variability.
Problem: Approximations

Model $\approx$ 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 $\sim$ 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 $\sim$ 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 $\varepsilon$ on the parameter space.
- Define new properties $\varphi_1$ and $\varphi_2$.

$\Rightarrow$ Given a trace $\sigma$ generated from the model $M$,
Solution: Safety margins

- Bound approximation error $\varepsilon$ on the parameter space.
- Define new properties $\varphi_1$ and $\varphi_2$.

$$\Rightarrow \text{Given a trace } \sigma \text{ generated from the model } M,$$

$$\sigma \models \varphi_1 \Rightarrow M \models \varphi$$
**Solution: Safety margins**

- Bound approximation error $\varepsilon$ on the parameter space.
- Define new properties $\varphi_1$ and $\varphi_2$.

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

$$\sigma \models \varphi_1 \Rightarrow M \models \varphi$$
Solution: Safety margins

- Bound approximation error $\varepsilon$ on the parameter space.
- Define new properties $\varphi_1$ and $\varphi_2$.

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

$$\sigma \models \varphi_1 \Rightarrow M \models \varphi$$
Solution: Safety margins

- Bound approximation error $\varepsilon$ on the parameter space.
- Define new properties $\varphi_1$ and $\varphi_2$.

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

$$\sigma \models \varphi_1 \Rightarrow M \models \varphi$$
Solution: Safety margins

- Bound approximation error $\varepsilon$ on the parameter space.
- Define new properties $\varphi_1$ and $\varphi_2$.

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

$$M \models \varphi \Rightarrow \sigma \models \varphi_2$$
Solution: Safety margins

- Bound approximation error $\varepsilon$ on the parameter space.
- Define new properties $\varphi_1$ and $\varphi_2$.

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

$$M \models \varphi \Rightarrow \sigma \models \varphi_2$$
Solution: Safety margins

- Bound approximation error $\varepsilon$ on the parameter space.
- Define new properties $\varphi_1$ and $\varphi_2$.

$\implies$ Given a trace $\sigma$ generated from the model $M$,

$$M \models \varphi \implies \sigma \models \varphi_2$$
Solution: Safety margins

- Bound approximation error $\varepsilon$ on the parameter space.
- Define new properties $\varphi_1$ and $\varphi_2$.

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

$$\sigma \models \varphi_1 \Rightarrow M \models \varphi \Rightarrow \sigma \models \varphi_2 \Rightarrow p_1 \leq p \leq p_2$$
Guarantees: global risk $\xi$ and precision $\alpha$

**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

$\Rightarrow \mathbb{P}(p_1 \in [\hat{p}_1 - \alpha, \hat{p}_1 + \alpha]) \geq 1 - \theta$, $\mathbb{P}(p_2 \in [\hat{p}_2 - \alpha, \hat{p}_2 + \alpha]) \geq 1 - \theta$
Guarantees: global risk $\xi$ and precision $\alpha$

**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

$\Rightarrow \mathbb{P}(p_1 \in [\hat{p}_1 - \alpha, \hat{p}_1 + \alpha]) \geq 1 - \theta$, $\mathbb{P}(p_2 \in [\hat{p}_2 - \alpha, \hat{p}_2 + \alpha]) \geq 1 - \theta$

**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

1. Context
2. SMC for ODE models
3. Application to parameterization
4. Perspectives
Parameterization: Find parameter values $\lambda$ for a generic model.

Goal: Find good values for $\lambda$ w.r.t. score ($\mathbb{E}(r)$) of $\varphi$-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 $\varepsilon$!
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 $\varepsilon$!

Proposition
When solving an ODE, the error $\varepsilon$ 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 $\varepsilon$!

Lemma 1.
For any arbitrary $\varepsilon > 0$, there exists an integration step $h$ such that

$$0 < \varepsilon h < \varepsilon, \quad \forall \lambda.$$
Aurelia Aurita

Jellyfish species from the Adriatic Sea.

\[
x'(t) = a \cdot x(t) \cdot \left(1 - \frac{x(t)}{b}\right)
\]

\[\varphi = \text{data} \pm \text{standard error}\]

---

\[^{2}\text{[Melica et al. 2014] Logistic density-dependent growth of an aurelia aurita polyps population. Ecological Modeling}\]
Aurelia Aurita

Jellyfish species from the Adriatic Sea.

- \( x'(t) = a \cdot x(t) \cdot \left(1 - \frac{x(t)}{b}\right) \)
- \( \varphi = \text{data} \pm \text{standard error} \)
- \( \varphi_1 = "\varphi - \varepsilon" \)
- \( \varphi_2 = "\varphi + \varepsilon" \)

\[ \begin{array}{|c|c|c|c|}
\hline
\text{days} & \varphi & \varphi_1 & \varphi_2 \\
\hline
0 & \text{data} & \text{data} - \varepsilon & \text{data} + \varepsilon \\
10 & \text{data} & \text{data} - \varepsilon & \text{data} + \varepsilon \\
20 & \text{data} & \text{data} - \varepsilon & \text{data} + \varepsilon \\
30 & \text{data} & \text{data} - \varepsilon & \text{data} + \varepsilon \\
40 & \text{data} & \text{data} - \varepsilon & \text{data} + \varepsilon \\
\hline
\end{array} \]

\[ \text{density (cm}^2\text{)} \]

---

\(^2\text{[Melica et al. 2014] Logistic density-dependent growth of an aurelia aurita polyps population. Ecological Modeling} \)
Parameter analysis

Probability of staying in the tunnel.

Expected distance to data.
Outline

1. Context
2. SMC for ODE models
3. Application to parameterization
4. Perspectives
Mathematics

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

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

D. JULIEN
End-to-end Statistical Model Checking for ODEs
Perspectives

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

Modeling
- **Structural properties** of ODE models.
  $\rightarrow$ stability, attraction, cycles ...
- **Dynamic** discretization of the parameter grid.
  $\rightarrow$ variance, score

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

D. JULIEN
End-to-end Statistical Model Checking for ODEs
Thank you for your attention!