A case study of
Data-driven computational modeling in Epidemics:
bringing the dirt to the classroom

Stefan Engblom

Div of Scientific Computing, Dept of Information Technology, Uppsala University, Sweden

LTH Seminar

Lund, March 6th, 2017
Outline

1. The Case: national-scale epidemics
   VTEC
   Computational modeling
   Parameter estimation and outcomes

2. Outlook
   The role of the Lax principle
   Model of models: the role of test equations

3. Bringing holistic computing into the classroom
   Why design
   A worked example: Applied finite elements

Summary
The Case: national-scale epidemics

- Ongoing research to better understand the spread of verotoxinogenic *E. coli* O157:H7 (VTEC O157:H7) in the Swedish cattle population
- **Zoonotic pathogen** (animal → human) of great public health interest, causing enteroheamorrhagic colitis (EHEC) in humans (∼500 cases annually in Sweden, cost per case ∼24kSEK)
The Case: national-scale epidemics

- Ongoing research to better understand the spread of verotoxinogenic *E. coli* O157:H7 (VTEC O157:H7) in the Swedish cattle population

- *Zoonotic pathogen* (animal → human) of great public health interest, causing enterohemorrhagic colitis (EHEC) in humans (∼500 cases annually in Sweden, cost per case ∼24kSEK)

- “Understand” means to determine the dominating mechanisms in the dynamics, evaluate the effect of counter measures, investigate “what ifs”...

- Substantial amount of data available:
  - individual-level cattle data from 2005 and onwards (“events”)
  - geographical and meteorological data
  - longitudinal studies of farms
VTEC epidemics

Flera barn sjuka i nytt utbrott bakteriesmitta

Sexton nya fall av ehec-smitta har konstaterats i Sverige. Sex av dem är barn och tre har fått den allvarliga följdsjukdomen hus. Merparten av fallna är i Stockholmsområdet, men det är ännu oklart vad som orsakat utbrottet – Den här sorten kan ge mer allvarliga komplikationer, säger Andra Folkhälsomyndigheten.

Av Ninna Bengtsson
📅 28 okt, 2016
📑 Spara artikel
VTEC epidemics

in short

Infected animals show no signs of the disease!
Event data
by European Union law

<table>
<thead>
<tr>
<th>REPORTER</th>
<th>WHERE</th>
<th>ABATTOIR</th>
<th>DATE</th>
<th>EVENT</th>
<th>ANIMALID</th>
<th>BIRTHDATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>83466</td>
<td>83958</td>
<td>0</td>
<td>2009-10-01</td>
<td>2</td>
<td>SE0834660433</td>
<td>1997-04-04</td>
</tr>
<tr>
<td>83958</td>
<td>83466</td>
<td>0</td>
<td>2009-10-01</td>
<td>1</td>
<td>SE0834660433</td>
<td>1997-04-04</td>
</tr>
<tr>
<td>83958</td>
<td>83829</td>
<td>0</td>
<td>2012-03-15</td>
<td>2</td>
<td>SE0834660433</td>
<td>1997-04-04</td>
</tr>
<tr>
<td>83829</td>
<td>83958</td>
<td>0</td>
<td>2012-03-15</td>
<td>1</td>
<td>SE0834660433</td>
<td>1997-04-04</td>
</tr>
<tr>
<td>83829</td>
<td>83958</td>
<td>0</td>
<td>2012-03-15</td>
<td>4</td>
<td>SE0834660433</td>
<td>1997-04-04</td>
</tr>
<tr>
<td>54234</td>
<td>83829</td>
<td>0</td>
<td>2012-04-11</td>
<td>1</td>
<td>SE0834660433</td>
<td>1997-04-04</td>
</tr>
<tr>
<td>83829</td>
<td>54234</td>
<td>0</td>
<td>2012-04-11</td>
<td>2</td>
<td>SE0834660433</td>
<td>1997-04-04</td>
</tr>
<tr>
<td>83829</td>
<td>83958</td>
<td>0</td>
<td>2012-04-11</td>
<td>5</td>
<td>SE0834660433</td>
<td>1997-04-04</td>
</tr>
</tbody>
</table>

Total: 18 649 921 reports and 37 221 holdings

Events

▶ Exit (death, n=1 438 506)
▶ Enter (birth, n=3 479 000)
▶ Internal transfer (ageing, n=6 593 921)
▶ External transfer (transport between holdings, n=732 292)
Event data
Meterological data

by SMHI
Best practise
Modeling and parametrization in epidemics

Typically:

1. Highly coarse-grained models, e.g. “mosaic” ODEs, combined with rule of thumbs, and various types of data averaged and understood as fluxes and source terms
2. For a parameter proposal, a single model run is used to obtain a residual of some kind wrt some measured data
3. What parameter combination minimizes $\|\text{residual}\|_{\text{some norm}}$?
Best practise
Modeling and parametrization in epidemics

Typically:

1. Highly coarse-grained models, e.g. “mosaic” ODEs, combined with rule of thumbs, and various types of data averaged and understood as fluxes and source terms

2. For a parameter proposal, a *single* model run is used to obtain a residual of some kind wrt some measured data

3. What parameter combination minimizes $\|\text{residual}\|_{\text{some norm}}$?

-On the one hand, what is the *uncertainty* of the model so obtained?

-On the other hand, the problem is not easy! The *topology* of the model might be the research question. Data is scarce and expensive to collect...
Forming a model

*a priori* thoughts

The dynamics/epidemics is quite likely stochastic, nonlinear, spatially inhomogeneous...

Designing/understanding computational models: either we do

- “mosaic approach” relying on fingerspitzengefühl...
- or, relying on the **Lax principle**: *if the numerical physics \( \approx \) the wanted “true” physics *(consistency)*, then the numerical solution \( \rightarrow \) the true solution *(convergence)* IFF the numerical physics is stable
Local model

"SIS_E"

Model states: Susceptible, Infected

State transitions

\[ I \rightarrow S \text{ at rate } \propto I(t) \]
\[ S \rightarrow I \text{ at rate } \propto S(t) \varphi(t) \]

80% of the holdings consist of <100 individuals. A suitable model for \((S, I)\) is therefore a continuous-time Markov chain.
Local model

“$SIS_E$”

Model states: Susceptible, Infected

State transitions

\[ I \rightarrow S \text{ at rate } \propto I(t) \]
\[ S \rightarrow I \text{ at rate } \propto S(t)\varphi(t) \]

80% of the holdings consist of <100 individuals. A suitable model for $(S, I)$ is therefore a continuous-time Markov chain.

Environmental infectious pressure (plain ODE)

\[
\frac{d\varphi}{dt} = \frac{I(t)}{S(t) + I(t)} - \beta(t)\varphi(t) + (\ldots)
\]
Global model

Stochastic reaction-transport framework

Put $X_t^{(i)} = [S_{ij}, I_{ij}, \varphi_i]^T$ for $j \in \{\text{calves, young stock, adults}\}$ and $i = 1, \ldots, \sim 40,000$ holdings.

$$dX_t^{(i)} = S\mu^{(i)}(dt) - \sum_{j \in C(i)} C\nu^{(i,j)}(dt) + \sum_{j; i \in C(j)} C\nu^{(j,i)}(dt).$$

Data now goes into all these forward operators.

The above general framework is implemented in SimInf (GitHub).
Numerical split-step method

Set-up

Local physics first, then global;

\[
\tilde{X}_{n+1}^{(i)} = X_n^{(i)} + \int_{t_n}^{t_{n+1}} S\mu^{(i)}(\tilde{X}(i)(s); \ ds),
\]

\[
X_{n+1}^{(i)} = \tilde{X}_{n+1}^{(i)} - \int_{t_n}^{t_{n+1}} \sum_{j \in C(i)} C\nu^{(i,j)}(X(i)(s); \ ds)
\]

\[+ \int_{t_n}^{t_{n+1}} \sum_{j; i \in C(j)} C\nu^{(j,i)}(X(i)(s); \ ds)\]

Assume (certain assumptions). Then

- \(E[\sup_{t_n \in [0,t]} \|X_n\|_p^p] \) bounded, any \( p \geq 1 \) (stability)
- \(E[\|X_n - X(t_n)\|^2] = O(h), \ h = \max_n(t_{n+1} - t_n) \) (convergence)
Parallel implementation
Dependency-aware scheduling via task-based framework

6 core task execution trace; red tasks are dependent steps (requiring thread synchronization).
Sample simulation

$\sim 9$ years of actual data

$\sim 10^8$ data base events plus $\sim 10^9$ infectious events during 9 years simulated in 15s on a desktop
Feasibility of parameter estimation

Synthetic data ("inverse crime")

Setup: determine $\hat{k} = \arg \min_k G(k)$,

$$G(k)^2 = M^{-1} \sum_{i=1}^{M} \| \mathcal{F} \circ X_{\text{simulated}}(k) - \mathcal{F} \circ X_{\text{input}}(k^*) \|^2,$$

$\mathcal{F}$ a "summary statistics"/"measurement filter" (…)

Using $M \in \{10, 20, 40\}$ simulations for $G$ and $N = 20$ iterations of an optimization routine:

<table>
<thead>
<tr>
<th>$M$</th>
<th>Residual</th>
<th>12 cores</th>
<th>32 cores</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.174</td>
<td>46.6 min</td>
<td>30.2 min</td>
</tr>
<tr>
<td>20</td>
<td>0.090</td>
<td>94.2 min</td>
<td>61.5 min</td>
</tr>
<tr>
<td>40</td>
<td>0.036</td>
<td>189.3 min</td>
<td>123.7 min</td>
</tr>
</tbody>
</table>
1. The Case: national-scale epidemics
Parameter estimation and outcomes

Parameter estimation

Real data

126 holdings sampled regularly during 38 months; ∼17 swipe samples per group of 3 animals. Probability(test positive|\(n\) individuals infected), \(n \in \{0, 1, 2, 3\}\) estimated via detailed studies \textit{a priori}.

![Diagram of data collection and analysis over time and across locations.]
Parameter estimation

Real data, but after testing the equivalent synthetic situation first!

Setup: determine $\hat{k} = \arg \min_k G(k),

\[
G(k)^2 = M^{-1} \sum_{i=1}^{M} \| \mathcal{F} \circ \mathbf{X}_{\text{simulated}}(k) - \mathcal{F}_{\text{measured}} \|^2,
\]

$\mathcal{F}$ is now the probabilistic map from state $\mathbf{X}$ to sample $\{0, 1\}$. 

\begin{figure}
\centering
\begin{subfigure}{0.3\textwidth}
\centering
\includegraphics[width=\textwidth]{beta_t1}
\caption{beta_t1}
\end{subfigure}
\begin{subfigure}{0.3\textwidth}
\centering
\includegraphics[width=\textwidth]{upsilon_1_2}
\caption{upsilon_1_2}
\end{subfigure}
\begin{subfigure}{0.3\textwidth}
\centering
\includegraphics[width=\textwidth]{D}
\caption{D}
\end{subfigure}
\end{figure}
1. The Case: national-scale epidemics

Parameter estimation and outcomes

Outcome

- On the one hand, “an answer”, a parametrized model
- More importantly, and usually from mistakes/misfits: a better understanding of the dynamics, of the interplay between parameters, an efficient procedure to find optimal models among suggestions...

Finding #1: \( \beta(t) = \beta \) required in the Swedish climate.

Finding #2: a mathematical analysis reveals a finite-time extinction in the stochastic model, contrary to a corresponding deterministic model.

"The purpose of computing is insight, not numbers." (R. Hamming)
1. The Case: national-scale epidemics

Parameter estimation and outcomes

**Outcome**

- On the one hand, “an answer”, a parametrized model
- More importantly, and usually from mistakes/misfits: a better understanding of the dynamics, of the interplay between parameters, an efficient procedure to find optimal models among suggestions...

**Finding #1:** decay $\beta = \beta(t)$ required in the Swedish climate.

**Finding #2:** a mathematical analysis reveals a finite-time extinction in the stochastic model, contrary to a corresponding deterministic model.

"The purpose of computing is insight, not numbers." (R. Hamming)
The role of the Lax principle

Complex modeling situations

On the one hand, convergence to a well-defined “truth” from consistency and stability is necessary...

...but it’s not enough. Parameters need to be observable too. And observable from data that can actually be collected!
The role of the Lax principle

Complex modeling situations

On the one hand, convergence to a well-defined “truth” from consistency and stability is necessary...

...but it’s not enough. Parameters need to be observable too. And observable from data that can actually be collected!

⇒ Understanding the limits is important, knowing what you cannot do is a good thing! Negative reasoning is a good take-away.

We are all faced with a series of great opportunities brilliantly disguised as unsolvable problems.

(John W. Gardner)
The role of test equations

Model of models!

Classical model $y' = \lambda y$, one parameter $\lambda \in \mathbb{C}$

Study convergence $y_h \rightarrow y$ as $h \rightarrow 0$, but what about $y_{h,N} \rightarrow y$ with $N$ the # observations to estimate $\lambda_N \approx \lambda$?
The role of test equations

Model of models!

Classical model $y' = \lambda y$, one parameter $\lambda \in \mathbb{C}$

Study convergence $y_h \to y$ as $h \to 0$, but what about $y_{h,N} \to y$ with $N$ the # observations to estimate $\lambda_N \approx \lambda$?

$\implies$ Linear birth-death model $y' = k - \mu y$, a model of a source and a sink. Such sources/sinks typically occur in more than one place, e.g.,

$$S' \propto I - \varphi S,$$
$$\varphi' \propto I - \varphi,$$

(hence only one of the birth-constants, the “$ks$”, can be non-dimensionalized away).
The role of test equations

Model of models!

Classical model $y' = \lambda y$, one parameter $\lambda \in \mathbb{C}$

Study convergence $y_h \rightarrow y$ as $h \rightarrow 0$, but what about $y_{h,N} \rightarrow y$ with $N$ the # observations to estimate $\lambda_N \approx \lambda$?

$\implies$ Linear birth-death model $y' = k - \mu y$, a model of a source and a sink.

Such sources/sinks typically occur in more than one place, e.g.,

$$S' \propto I - \varphi S,$$

$$\varphi' \propto I - \varphi,$$

(hence only one of the birth-constants, the “$k$s”, can be non-dimensionalized away).

- Is it doable to get $(k, \mu)$ from observations?

-Negative reasoning: if it won’t work when almost everything is known, it won’t work when confronted with a more realistic situation...
Data

\( N = 50 \) trajectories, sampled 10 times each
Posterior

MCMC using exact likelihood
Holistic computing to the classroom

It’s all about design!

*Design, transitive verb:*

to create, fashion, execute, or construct according to plan

Design as a task requires:

1. a working “forward model” — *(I tend to stop here!)*
2. a way to find the parameters
3. a method to make it plausible that the design is sound
4. communication of the result
Project – Design a Medical Torus

Applied Finite Element Methods 1TF056 (5.0 hp)

Murtazo Nazarov, Hanna Holmgren

October 28, 2016

INTRODUCTION

The overall grand project is to design a medical torus which emits a hormone through diffusion. The design variables are the shape of the torus and the initial concentration of hormone in the torus. The design constraints are the time until a certain amount of hormone has been emitted. Additional requirements are robustness to manufacturing errors and to model errors.

In this project we consider a medical torus of outer radius $R$ and inner radius $r$ that may be used in certain hormone treatments (see Fig. 1). It is inserted just under the skin on the inside of the upper arm where the conditions are such that the hormone will slowly diffuse out in the surrounding tissue.
A huge task
...solved in pieces

“The design constraints are the time until a certain amount of hormone has been emitted. Additional requirements are robustness to manufacturing errors and to model errors.”

“Your task is to investigate the physics and the numerical modeling associated with this problem, design the torus so that it fulfills certain conditions and requirements, and communicate the design and the qualities of the design in a convincing way. Your results should support an expert committee in reaching a decision concerning the design of a product.”

Part A 1D simplification, show convergence of adaptive FEM
Part B 2D simplification, implement in Matlab (assembly, solving)
Part C 3D using FEniCS, final design
Evaluation comments

Student voices:

► “The project was very enjoyable, and I liked very much that the course was so project-driven. Having a large “final project” that you build upon continually, rather than three disjoint assignments, was great and I think more courses should follow that recipe.”

► “I gotta say, the assignments were rather hard. But I feel like I’ve learned so much and I think the assignments probably are the most important part of the course.”

► “I really liked working with the project. It really sparked an interest for finite element methods for me, and I feel like I learned a lot.”

(On the teacher’s side: helps to motivate what goes in or out of a course…)
Summary

- Case of national-scale computational modeling in Epidemics, incorporating large amounts of data (data bases, internet)
- **Consistent** modeling and the Lax principle $\Rightarrow$ well-posedness, stability, consistency, convergence

- Efficient simulation, numerical method designed in order to expose parallelism ($\sim 10^8$ data base events plus $\sim 10^9$ infectious events during 9 years simulated in 15s on a desktop)

- $\Rightarrow$ Parametrization of a national-scale model solved in **SimInf** (GitHub), interesting findings when fitting parameters to data

- At the meta-level: the actual role of stability, consistence, observability, test equations...

- Discrepancy wrt what we tend to teach

- $\Rightarrow$ A worked example of “PBL Light”
Acknowledgment

Joint work with:

- Pavol Bauer (PhD-student, Uppsala university)
- Augustin Chevallier (MSc-student, ENS Cachan/INRIA Sophia Antipolis)
- Stefan Widgren (National Veterinary Institute)
- Murtazo Nazarov and Hanna Holmgren (Uppsala university)
Thanks!

Programs, Papers, and Preprints are available from my web-page.
Thank you for the attention!