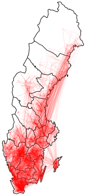
Data-driven Epidemiological Simulations: Verotoxigenic *E. coli O157*



Stefan Engblom

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

Mathematical Biology for Understanding Emerging Infectious Diseases at the Human-Animal-Environment Interface: a "One Health" Approach

Banff, Alberta, Canada, November 20–25, 2016

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 causing enteroheamorrhagic colitis (EHEC) in humans (∼500 cases anually in Sweden, cost per case ∼$2600).

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 causing enteroheamorrhagic colitis (EHEC) in humans (∼500 cases anually in Sweden, cost per case ∼$2600).
- "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
  - longitudal studies of farms

Event data

by European Union law

REPORTER	WHERE	ABATTOIR	DATE	EVENT	ANIMALID	BIRTHDATE
83466	83958	0	2009-10-01	2	SE0834660433	1997-04-04
83958	83466	0	2009-10-01	1	SE0834660433	1997-04-04
83958	83829	0	2012-03-15	2	SE0834660433	1997-04-04
83829	83958	0	2012-03-15	1	SE0834660433	1997-04-04
83829	83958	0	2012-03-15	4	SE0834660433	1997-04-04
54234	83829	0	2012-04-11	1	SE0834660433	1997-04-04
83829	54234	0	2012-04-11	2	SE0834660433	1997-04-04
83829	83958	0	2012-04-11	5	SE0834660433	1997-04-04

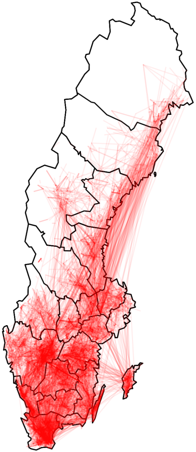
Total: 18 649 921 reports and 37 221 holdings

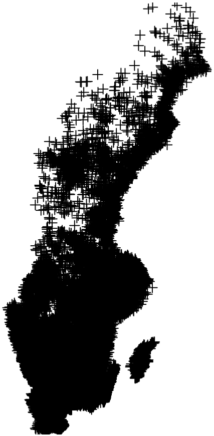
Events

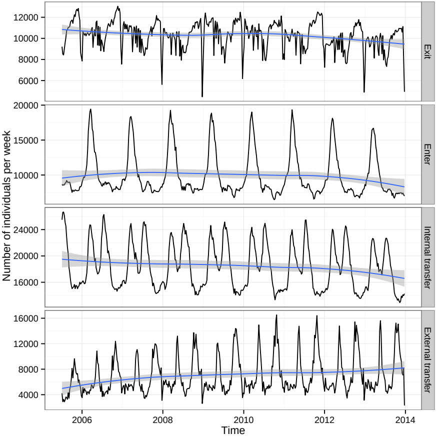
- Exit (death, n=1 438 506)
- Enter (birth, n=3 479 000)
- Internal transfer (ageing, n=6 593 921)
- External transfer (transport between holdings, n=732 292)

Event data

Area Sweden:Alberta is 2:3, population 2:1

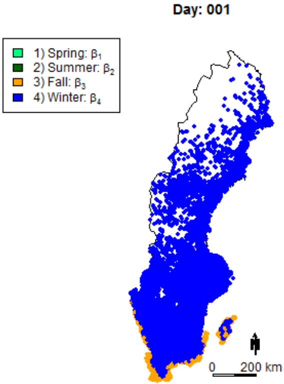






Meterological data

by SMHI



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, *a single* continuous-time mathematical model, a framework

Local model

"SIS_E"

Model states: **S**usceptible, **I**nfected

State transitions

$$I \longrightarrow S \text{ at rate } \propto I(t)$$
$$S \longrightarrow I \text{ at rate } \propto S(t)\varphi(t)$$

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: **S**usceptible, **I**nfected

State transitions

$$I \longrightarrow S \text{ at rate } \propto I(t)$$
$$S \longrightarrow I \text{ at rate } \propto S(t)\varphi(t)$$

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

$$d\mathbb{X}_t^{(i)} = \underbrace{\mathbb{S}\boldsymbol{\mu}^{(i)}(dt)}_{\text{local } SIS_E\text{-model+local events}} \underbrace{- \sum_{j \in C(i)} \mathbb{C}\boldsymbol{\nu}^{(i,j)}(dt) + \sum_{j;\, i \in C(j)} \mathbb{C}\boldsymbol{\nu}^{(j,i)}(dt)}_{\text{global events+physics}}.$$

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;

$$\tilde{\mathbb{X}}^{(i)}_{n+1} = \mathbb{X}^{(i)}_n + \int_{t_n}^{t_{n+1}} \mathbb{S}\boldsymbol{\mu}^{(i)}(\tilde{\mathbb{X}}^{(i)}(s);\ ds),$$

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

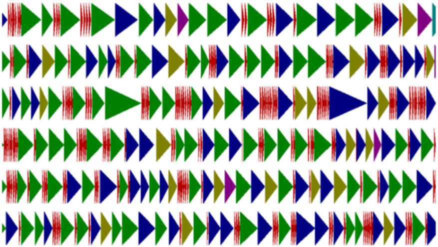
$$+ \int_{t_n}^{t_{n+1}} \sum_{j;\, i \in C(j)} \mathbb{C}\boldsymbol{\nu}^{(j,i)}(\mathbb{X}^{(i)}(s);\ ds)$$

Assume (certain assumptions). Then

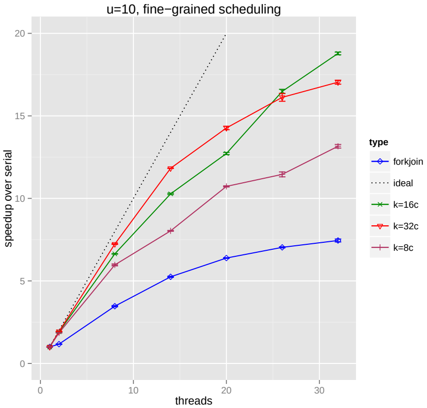
- $\mathbb{E}[\sup_{t_n \in [0,t]} \|\mathbb{X}_n\|_{\boldsymbol{l}}^p]$ bounded, any $p \geq 1$ (stability)
- $\mathbb{E}[\|\mathbb{X}_n - \mathbb{X}(t_n)\|^2] = O(h)$, $h = max_n(t_{n+1} - t_n)$ (convergence)

Parallel implementation

Dependency-aware scheduling via task-based framework

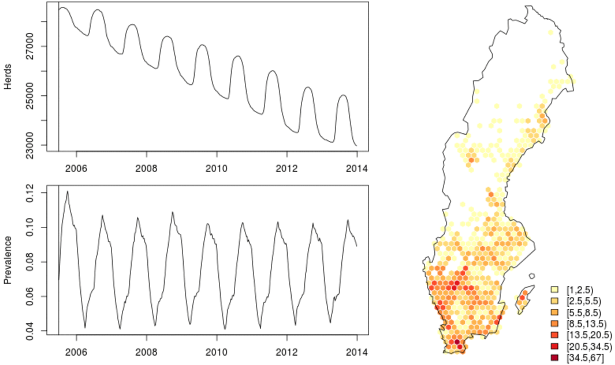


6 core task execution trace; red tasks are dependent steps (requiring thread synchronization).



Sample simulation

∼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")

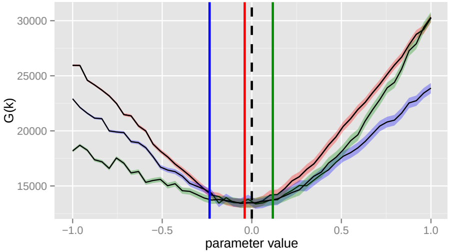
Setup: determine $\hat{k} = \arg\min_k G(k)$,

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

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

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

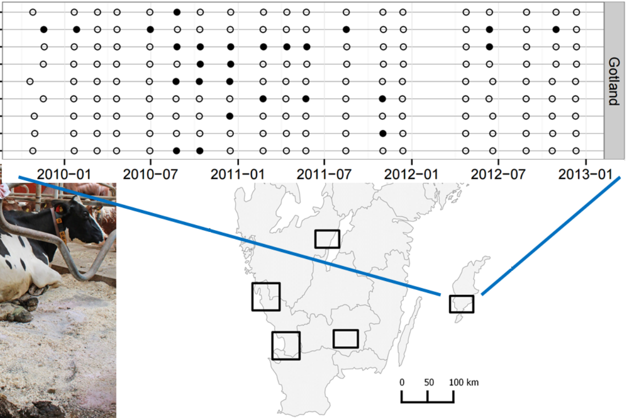
M	Residual	12 cores	32 cores
10	0.174	46.6 min	30.2 min
20	0.090	94.2 min	61.5 min
40	0.036	189.3 min	123.7 min



Parameter estimation

Real data

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



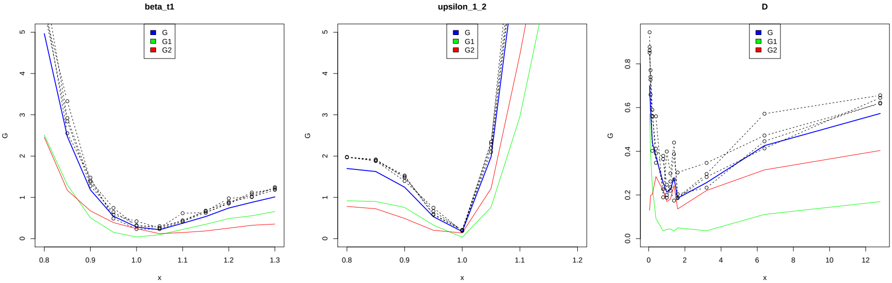
Parameter estimation

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

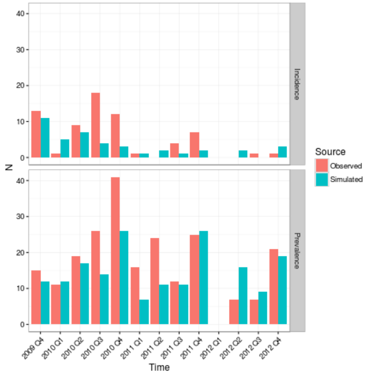
Setup: determine $\hat{k} = \arg\min_k G(k)$,

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

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

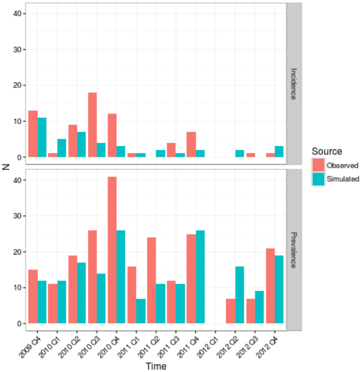


Outcome



- 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...

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)

Summary

- Case of national-scale computational modeling in Epidemics, incorporating large amounts of data (data bases, internet)
- Consistent modeling in continuous-time (*here:* Markov chain, ODE); clear what is the intended mathematical "truth", what is a numerical error, errors due to uncertainties in parameters, data errors...
- 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)
- $\Longrightarrow$ Parametrization of a national-scale model solved in SimInf (GitHub), interesting findings when attempting to fit parameters to data
- Ongoing: modeling of ASF in the wild boar-domestic pigs population (so freely moving animals), modeling of AMR *on top* of our VTEC animal-model

Thanks!

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