Identifying individual disease dynamics in a stochastic multi-pathogen model from aggregated reports and laboratory data

Oksana A. Chkrebtii

Department of Statistics, The Ohio State University

Yury E. García

Universidad de Antioquia, Medellin, Colombia

Marcos A. Capistrán

Centro de Investigación en Matemáticas, Guanajuato, México

Daniel E. Noyola

Department of Microbiology, Faculty of Medicine, Universidad Autónoma de San Luis Potosí, México

BIRS Workshop 18w5144, November 14, 2018

MBI Rules of Life summit

mbi About Participate People Programs Resources Education Visiting Donate Search Go

ORGANIZERS

Janet Best
Mathematics, The Ohio State University

Catherine Calder
Department of Statistics, The Ohio State University

Oksana Chkrebtii
Department of Statistics, Ohio State University

Cassandra Extavour
Department of Organismic and Evolutionary Biology, Harvard University

Avner Friedman
Department of Mathematics, The Ohio State University

Richard Lenski
Department of Microbiology & Molecular Genetics, Michigan State University

Michael Mackey
Applied Mathematics in Bioscience and Medicine

Frederik Nijhout
Biology, Duke University

PARTICIPATE

Apply for Event

SHARE

Twitter
Facebook
Google+
LinkedIn
Email

https://mbi.osu.edu/event/?id=1251.

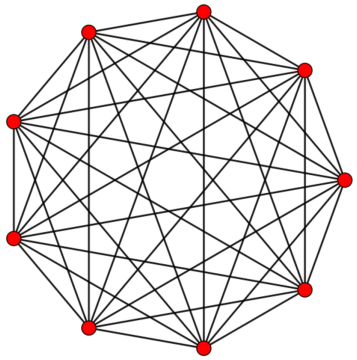
Acute respiratory disease

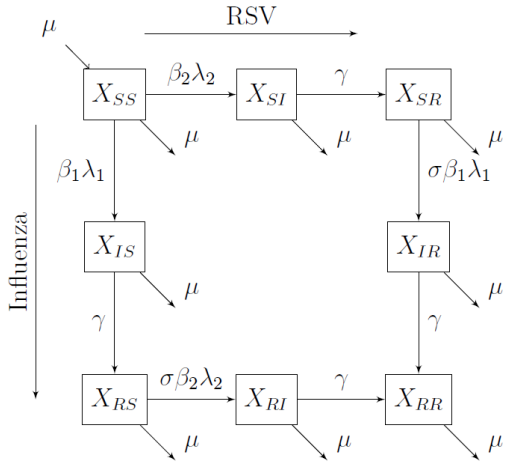
What are ARIs?

- Acute respiratory infections (ARI) are infections of the respiratory tract caused by viruses such as Adenovirus, Influenza A and B, Parainfluenza, Respiratory Syncytial Virus (RSV), and Rhinovirus
- Responsible for mortality and morbidity worldwide, mainly affecting children under 5 and adults above 65 years of age
- Influenza and Respiratory Syncytial Virus (RSV) are the leading etiologic agents of seasonal Acute Respiratory Infections (ARI)
- Understanding the mechanisms of these diseases and the impact of control measures helps public health to make decisions

Some challenges in statistical inference for epidemics

- Realistic mathematical models of epidemics are often stochastic with unknown transition probabilities
- Numerical methods for simulating these models (e.g. Euler - Maruyama, Gillespie) are prohibitively expensive
- State space lies on a low-dimensional manifold that is difficult to explore

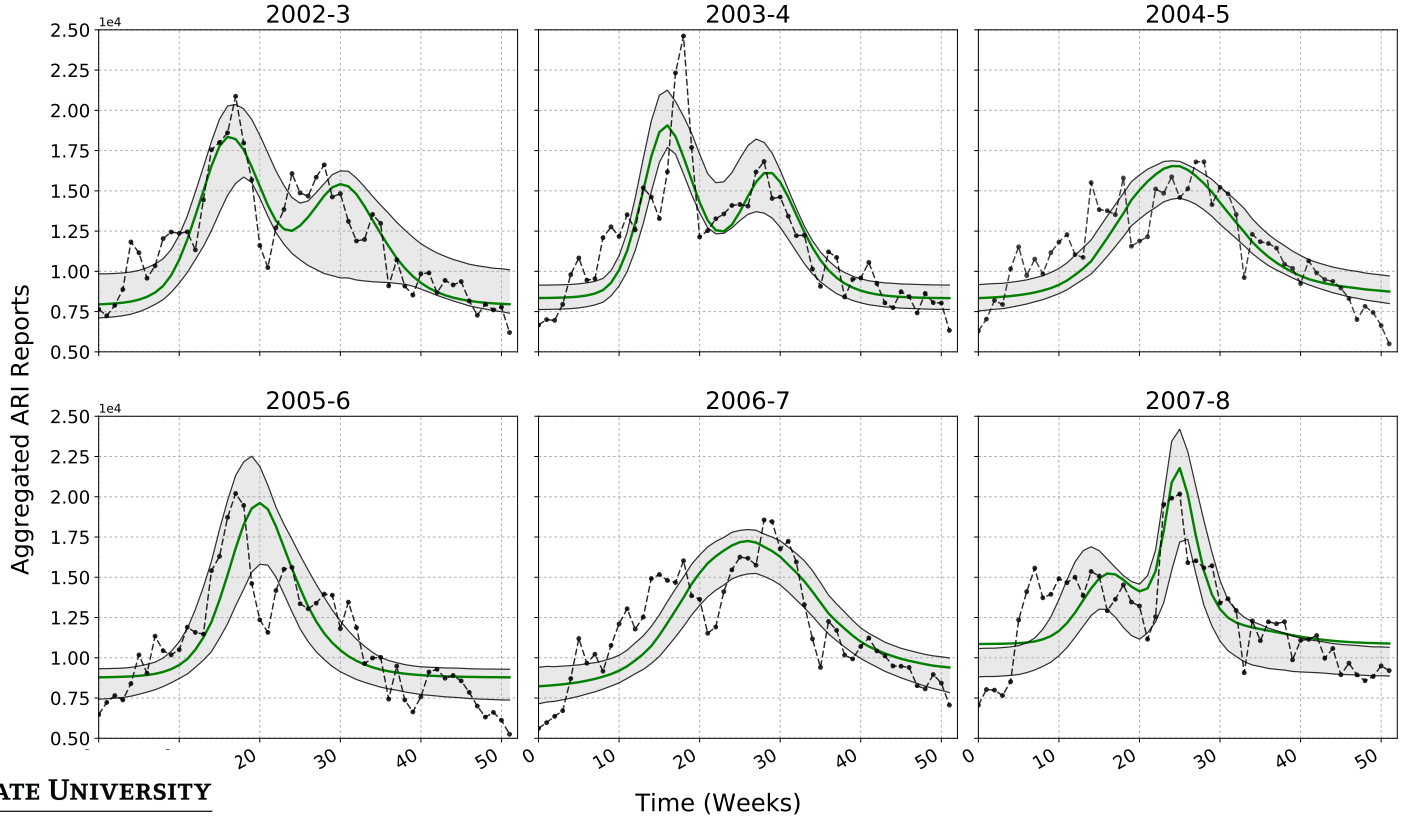




Left: states on the 8-simplex obey conservation laws; Right: two-pathogen SIR model.

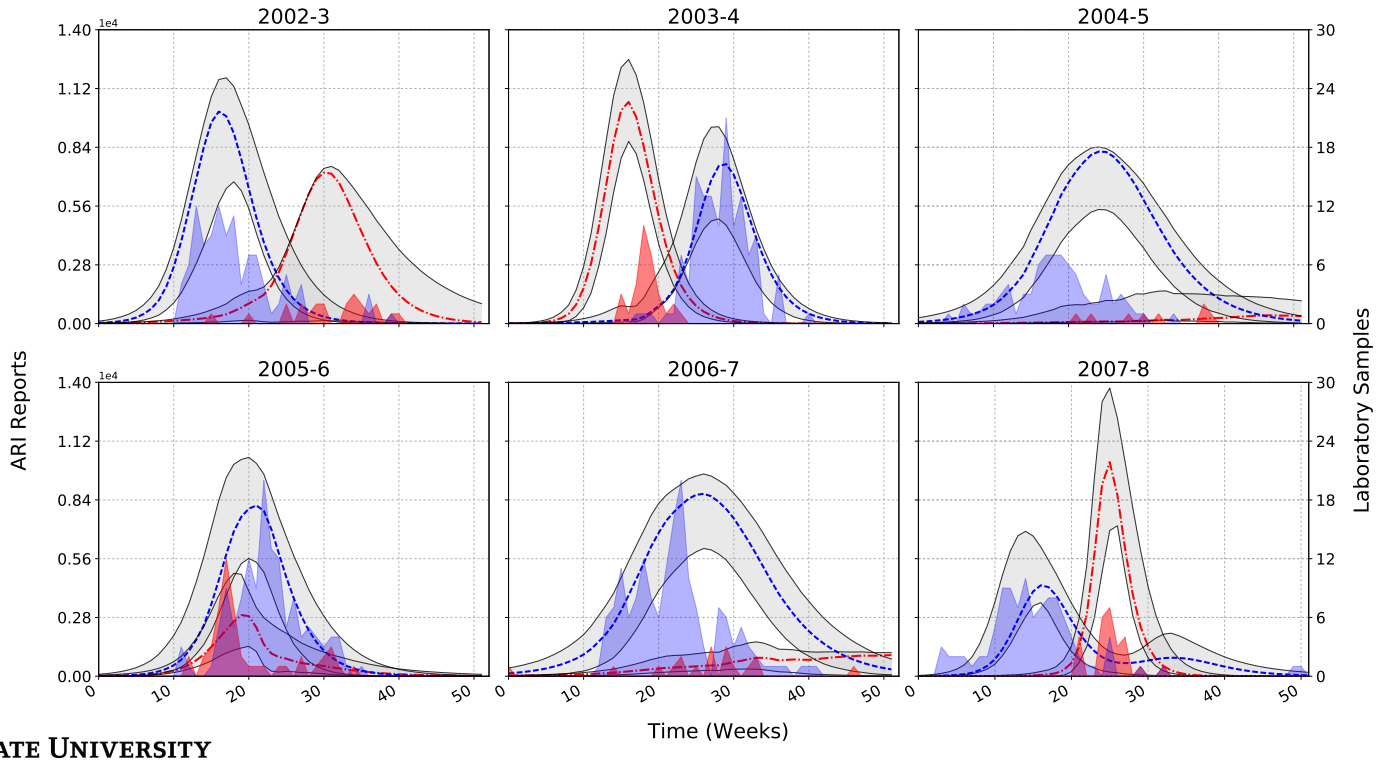
Some challenges in statistical inference for ARIs

- ARIs typically exhibit similar symptoms and physician visit data does not differentiate disease type, although additional genetic testing data may be available.



Inference from aggregate data on epidemic counts

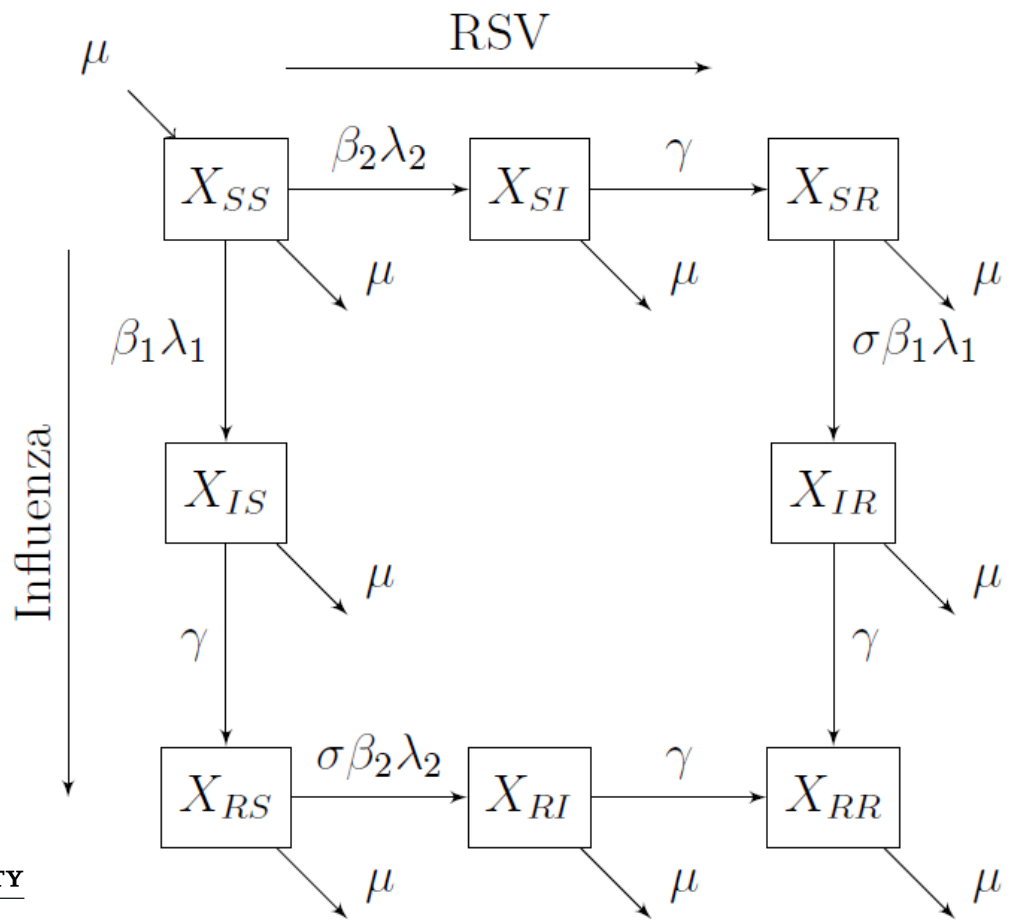
A Bayesian hierarchical modeling approach incorporating a Linear Noise Approximation of the governing equations allows parameter estimation for a multi-pathogen model from a combination of aggregate physician report data and laboratory samples



Stochastic SIR model with two pathogens

Stochastic SIR model with two pathogens

$X_{kl}(t)$ denotes the number of individuals at time t in immunological status $k \in \{S, I, R\}$ for pathogen 1 and immunological status $l \in \{S, I, R\}$ for pathogen 2.



Model parameters

Ω	Average yearly population size
σ	Cross-immunity or enhancement
λ_p	Proportion of people infected with pathogen $p \in \{1, 2\}$
β_p	Baseline transmission rate for pathogen $p \in \{1, 2\}$
μ	Birth/death rate
γ	Recovery rate

Parametres defining the interacting pathogen model

Chemical master equation for the system

Let $a_j(X)$ be transition probabilities and let v_j be stoichiometric vectors corresponding to reaction type $j = 1, \ldots, \mathcal{R}$.

The mechanism is encoded in the Kolmogorov forward equation (chemical master equation) for the system:

$$\frac{d}{dt} P_X(t) = \sum_{j=1}^{\mathcal{R}} \left\{ a_j(X - v_j) P_{X - v_j}(t) - a_j(X) P_X(t) \right\}$$

Average yearly population in San Luis Potosí is $\Omega \approx 2.5$ million people. We assume that the population is reasonably well mixed.

Large-volume approximation for the latent states

For large Ω the system states X can be approximated by the sum of:

1. a deterministic term ϕ
2. a stochastic term ξ

$$X(t) = \Omega\phi(t) + \Omega^{1/2}\xi(t), \quad t \in [0, T]$$

Assuming constant average concentration, the size of the stochastic component will increase as the square root of population size.

This result is known as the van Kampen expansion or Linear Noise Approximation.

Large-volume approximation for the latent states

1. Deterministic component $\phi_i(t) = \lim_{\Omega,X \longrightarrow \infty} X_i/\Omega$, $i = 1, \ldots, \dim\{X(t)\}$ evolves as:

$$\begin{cases} \dfrac{d\phi_i(t)}{dt} = \displaystyle\sum_{j=1}^{\mathcal{R}} S_{ij} a_j(\phi(t)), \quad t \in (0, T] \\ \phi_i(0) = \phi_0, \end{cases}$$

2. Stochastic component ξ is governed by the Itô diffusion equation,

$$d\xi(t) = A(t)\xi(t)dt + \sqrt{B(t)}dW(t), \quad t \in [0, T],$$

with Gaussian initial states, with $A(t) = \dfrac{\partial S\, a(\phi(t))}{\partial \phi(t)}$, $B(t) = S\,\mathrm{diag}\{a(\phi(t))\}\,S^\top$, and $W(t)$ denotes the $\mathcal{R}$ dimensional Wiener process

Large-volume approximation for the latent states

A further approximation characterizes the distribution of the Markov process $X(t), t \in [0, T]$ as,

$$X(t) \mid \theta \sim \mathcal{N}\left(\Omega\phi(t), \Omega C(t,t)\right),$$

where θ are model parameters and C is the solution to a system of ordinary differential equations parameterized by θ.

Bayesian Hierarchical model

Aggregated physician reported ARI counts

Due to the similarity in symptoms, we can only observe the total number of infections of both kinds:

$$G^\top X(t) = X_{IS}(t) + X_{IR}(t) + X_{SI}(t) + X_{RI}(t)$$

That is,

$$G^\top X(t) \mid \theta \sim \mathcal{N}\left(\Omega G^\top \phi(t), \Omega G^\top C(t,t) G\right)$$

Aggregated physician reported ARI counts

Physician reported ARI counts are indirect observations of the Markov process $X(t)$ measured weekly,

$$Y(t_i) \mid X(t_i), \theta, \tau \sim \mathcal{N}\left(rG^\top X(t_i) + r\Omega\alpha, r^2\Omega G^T CG + \Sigma\right),$$
$$i = 1, \ldots, 52,$$

where r is a reporting proportion, $\alpha \in (0, 1)$ is a background term, and Σ is the covariance matrix of the observation error.

Laboratory sample of infants

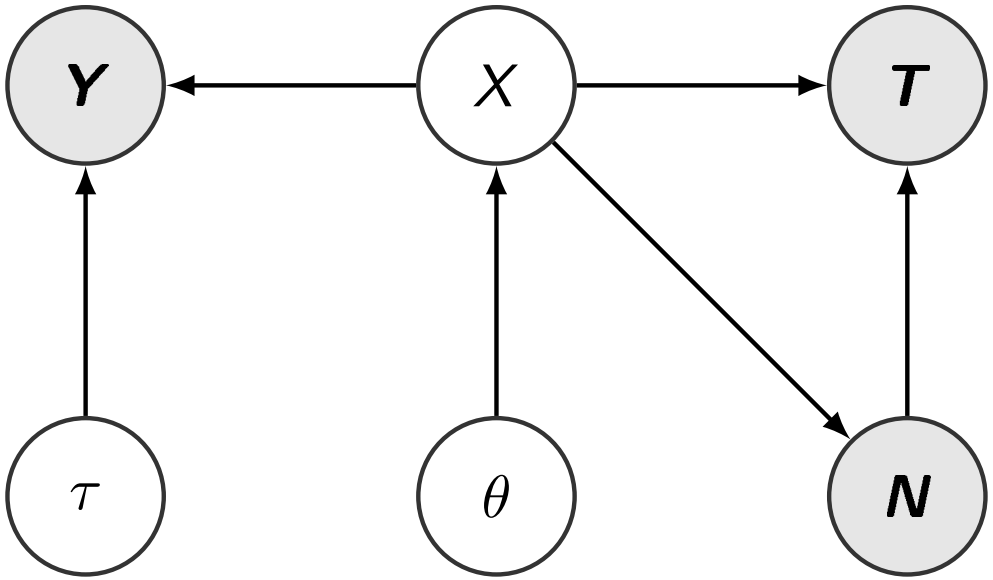
$T(t_i)$ represents the number of infants who were diagnosed with Influenza out of a sample of $N(t_i)$ infants tested.

$$\begin{aligned} T(t_i) \mid N(t_i), X(t_i), \theta &\sim \text{Bin}\left(N(t_i), P(t_i)\right), \; i = 1, \ldots, 52, \\ N(t_i) \mid X(t_i), \theta &\sim \delta(cX(t_i)), \; i = 1, \ldots, 52, \end{aligned}$$

where c denotes the proportion in the population of children under 5 years of age who were tested for Influenza. Dependence on X and θ is through the probability of a subject being diagnosed with Influenza:

$$P(t_i) = \frac{X_{IS}(t_i) + X_{IR}(t_i)}{X_{IS}(t_i) + X_{IR}(t_i) + X_{SI}(t_i) + X_{RI}(t_i)}, \; i = 1, \ldots, 52,$$

Model visualization



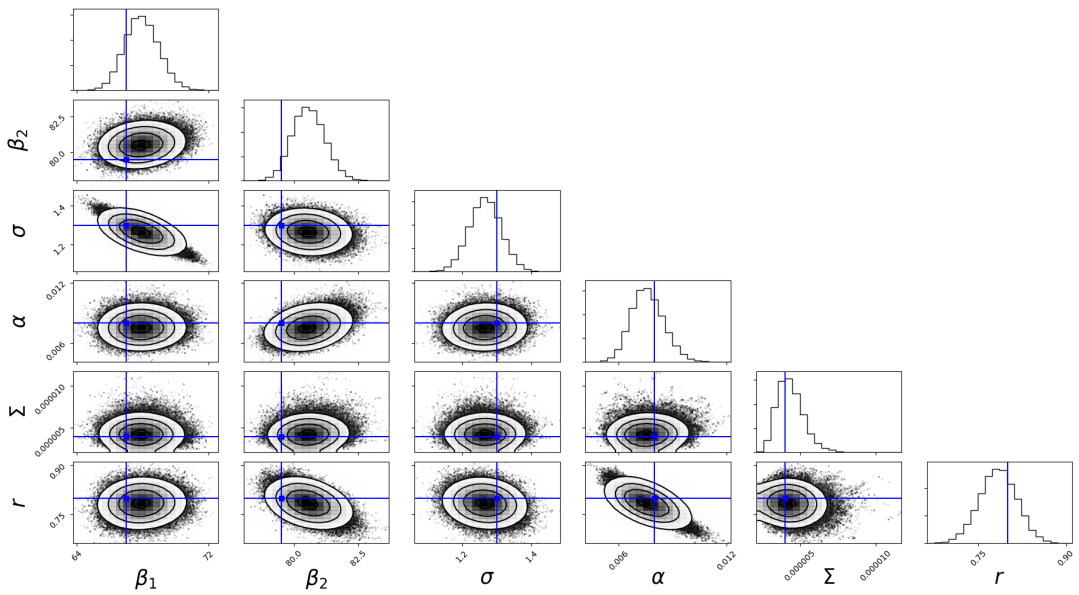
Arrows represent conditional dependence; nodes shaded in gray indicate observed data.

$$\pi\left(\theta, \tau, X(t) \mid \boldsymbol{Y}, \boldsymbol{T}, \boldsymbol{N}\right)$$
$$\propto p\left(\boldsymbol{Y} \mid X(t), \tau\right) p\left(\boldsymbol{T} \mid \boldsymbol{N}, X(t)\right) p\left(\boldsymbol{N} \mid X(t)\right) p\left(X(t) \mid \theta, \tau\right) \pi\left(\theta, \tau\right)$$

Posterior sampling

Simultaneously modeling all years results in a relatively high dimensional parameter space and strong posterior correlation.

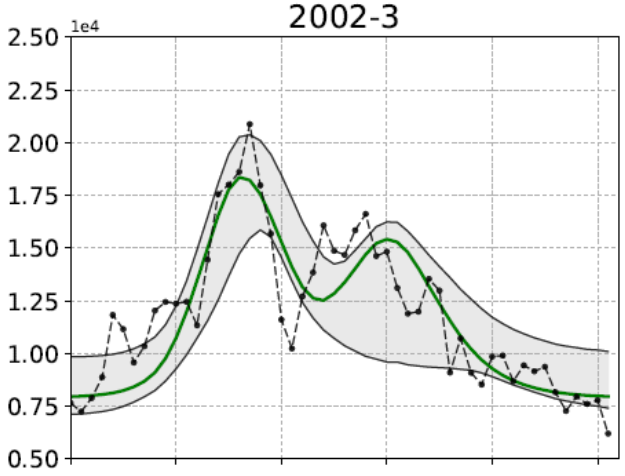
Posterior functionals are estimated from a Markov chain Monte Carlo sample employing parallel tempering.



Posterior samples over SIR model parameters

Future work

- Model parameters may be related across years, which suggests a random effects structure for the SIR model parameters.
- Introduce a more realistic model for the background infections, which could be interpreted as a discrepancy term between the model and the data.
- Better visualization tools (joint work with Xiao Zang and Sebastian Kurtek)



References

1. Paul Fearnhead, Vasilieos Giagos, and Chris Sherlock. *Inference for Reaction Networks Using the Linear Noise Approximation*, Biometrics, 2014
2. Yury E. García, Oksana A. Chkrebtii, Marcos A. Capistrán. *Inference on a Stochastic Multi-pathogen Model of Disease Dynamics with Aggregated Data*. In revision. https://arxiv.org/abs/1710.10346
3. Nicolaas Godfried Van Kampen. *Stochastic Processes in Physics and Chemistry*, volume 1. Elsevier, 1992.
4. Andrew Golightly, Daniel A Henderson, and Christopher Sherlock. *Statistics and Computing*. 2015.
5. D.J. Wilkinson. *Stochastic Modelling for Systems Biology*, Second Edition. Chapman & Hall/CRC Mathematical and Computational Biology. Taylor & Francis, 2011.