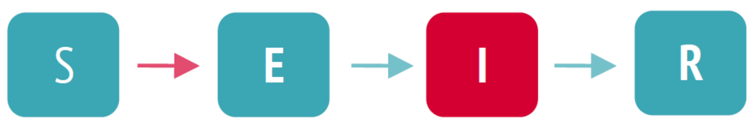
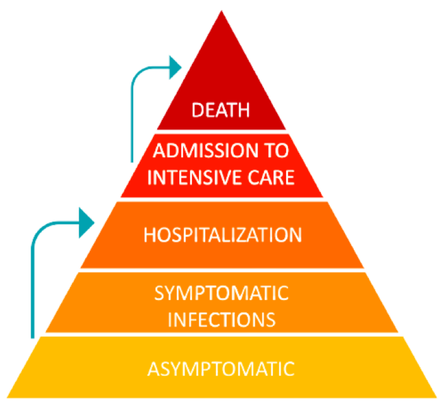
Inferring epidemics from multiple dependent data via pseudo-marginal methods

A Corbella, AM Presanis, PJ Birrell, and D De Angelis
Bayes4Health, WARWICK, MRC Biostatistics Unit, UNIVERSITY OF CAMBRIDGE

Introduction: processes in epidemics

Multiple processes **jointly** contribute to generate epidemic data





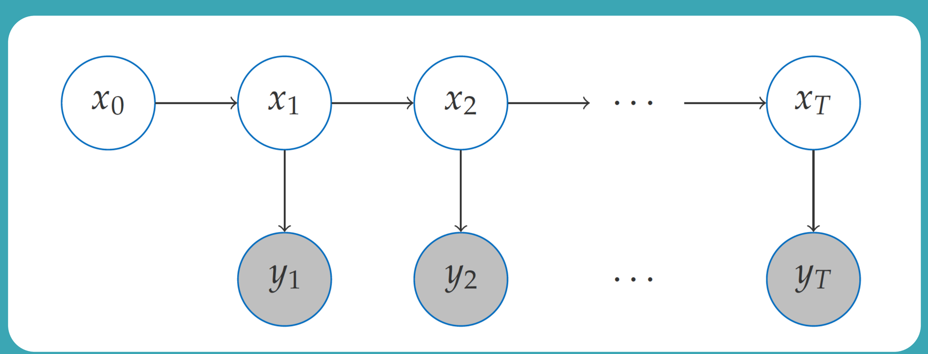
Transmission dynamics in the population

Severity process for disease progression

Other sources of randomness

Which and how much **stochasticity** should we introduce?

SSMs and Pseudo-Marginal methods



State Process

$$X_0|\theta \sim p(x_0|\theta)$$

$$X_t|(X_{t-1},\theta) \sim p(x_t|x_{t-1},\theta)$$

Observation process

$$Y_t|(X_t,\theta) \sim p(y_t|x_t,\theta)$$

The likelihood can be computed **sequentially**

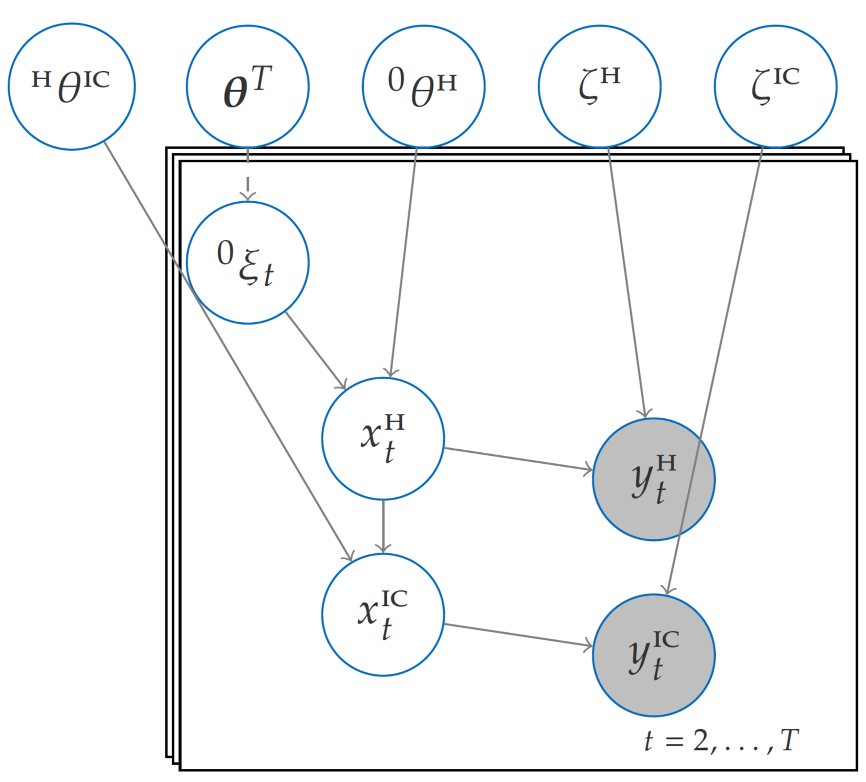
$$p(y_{1:T}|\theta) = \prod_{t=1}^{T} p(y_t|y_{1:t-1},\theta) = \prod_{t=1}^{T} \int_{X_{0:t}} p(y_t, x_{0:t}|y_{1:t-1},\theta)\, dx_{0:t}$$

Pseudo-Marginal methods consists of using a MC approximation of the likelihood $\hat{p}(y_{1:T}|\theta)$ within an MCMC routine.

Multiple data on epidemic

For large epidemics/pandemics a **semi-stochastic model** with deterministic transmission and stochastic severity is realistic

$$\begin{aligned}
&({}^{O}_{t}X^{H}) \sim \text{Pois}({}^{O}\xi \cdot {}^{O}\theta^{H}) \\
&({}^{O}_{t}X^{H}_{t:t+D}|{}^{O}_{t}X^{H} = {}^{O}_{t}x^{H}) \sim \text{Multi}({}^{O}_{t}x^{H}, {}^{O}f^{H}_{0:D}), \qquad X^{H}_{t} = \sum_{s=0}^{S} {}_{t-s}{}^{O}X^{H}_{t} \\
&({}^{H}_{t}X^{IC}|X^{H}_{t} = x^{H}_{t}) \sim \text{Bin}(x^{H}_{t}, {}^{H}\theta^{IC}) \\
&({}^{H}_{t}X^{IC}_{t:t+D}|{}^{H}_{t}X^{IC} = {}^{H}_{t}x^{IC}) \sim \text{Multi}({}^{H}_{t}x^{IC}, {}^{H}f^{IC}_{0:D}), \qquad X^{IC}_{t} = \sum_{s=0}^{S} {}_{t-s}{}^{H}X^{IC}_{t} \\
&(Y^{H}_{t}|X^{H}_{t} = x^{H}_{t}) \sim \text{Bin}(x^{H}_{t}, \zeta^{H}_{t}) \\
&(Y^{IC}_{t}|X^{IC}_{t} = x^{IC}_{t}) \sim \text{Bin}(x^{IC}_{t}, \zeta^{IC}_{t})
\end{aligned} \tag{1}$$



Even if the temporal dependence is broken there is still **dependence in the severity states** so that:

$$p(y^{H}_{1:T}, y^{IC}_{1:T}|\theta) \neq p(y^{H}_{1:T}|\theta)p(y^{IC}_{1:T}|\theta)$$

The joint likelihood factorises as:

$$\begin{aligned}
p(y^{H}_{1:T}, y^{IC}_{1:T}|\theta) &= p(y^{H}_{1:T}|y^{IC}_{1:T},\theta)p(y^{IC}_{1:T}|\theta) \\
&= p(y^{IC}_{1:T}|y^{H}_{1:T},\theta)p(y^{H}_{1:T}|\theta)
\end{aligned}$$

where one of the two components can be **approximated via MC integration** and embedded in a Pseudo-Marginal routine.

Relevance of the dependence

Simulation study in a

- small dependence scenario
- big dependence scenario

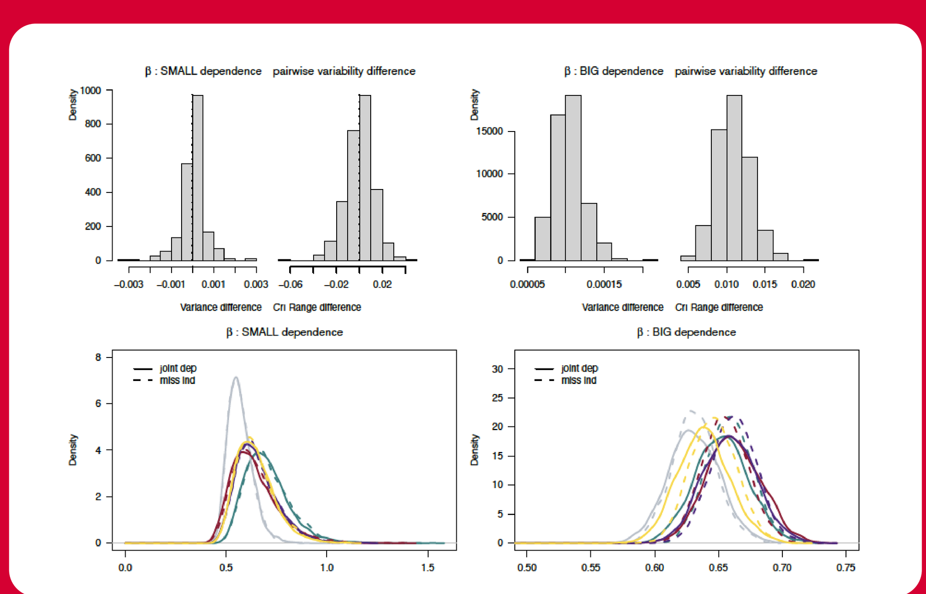
Compare the inference obtained with algorithms that assume

Misspecified independent

vs

Joint dependent

data

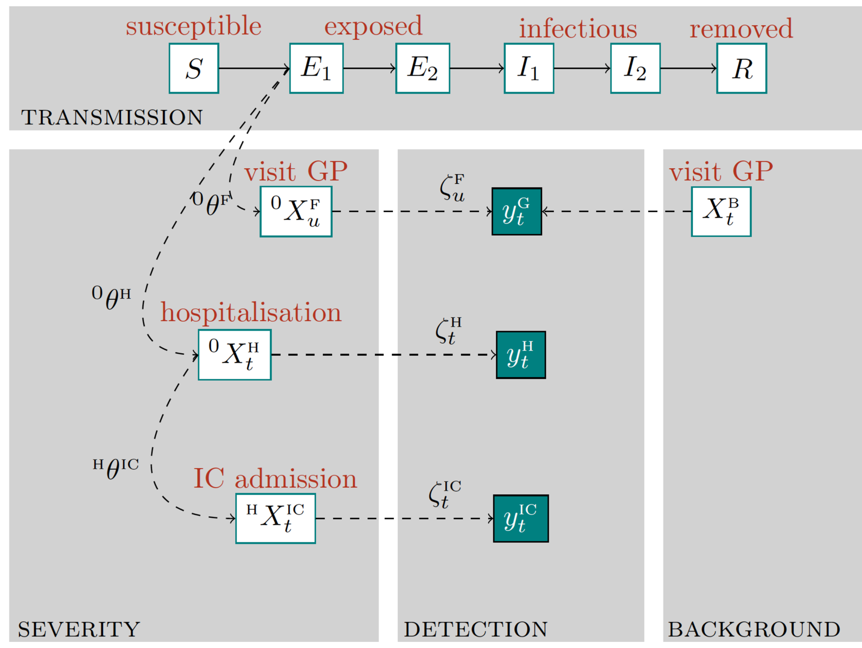


A model for the 2017/18 Flu epidemic

Fuse data on

- weekly hospital and intensive care (IC) admissions
- daily GP consultations
- virology and serology

We assume a deterministic transmission model with **stochastic severity** as (1)



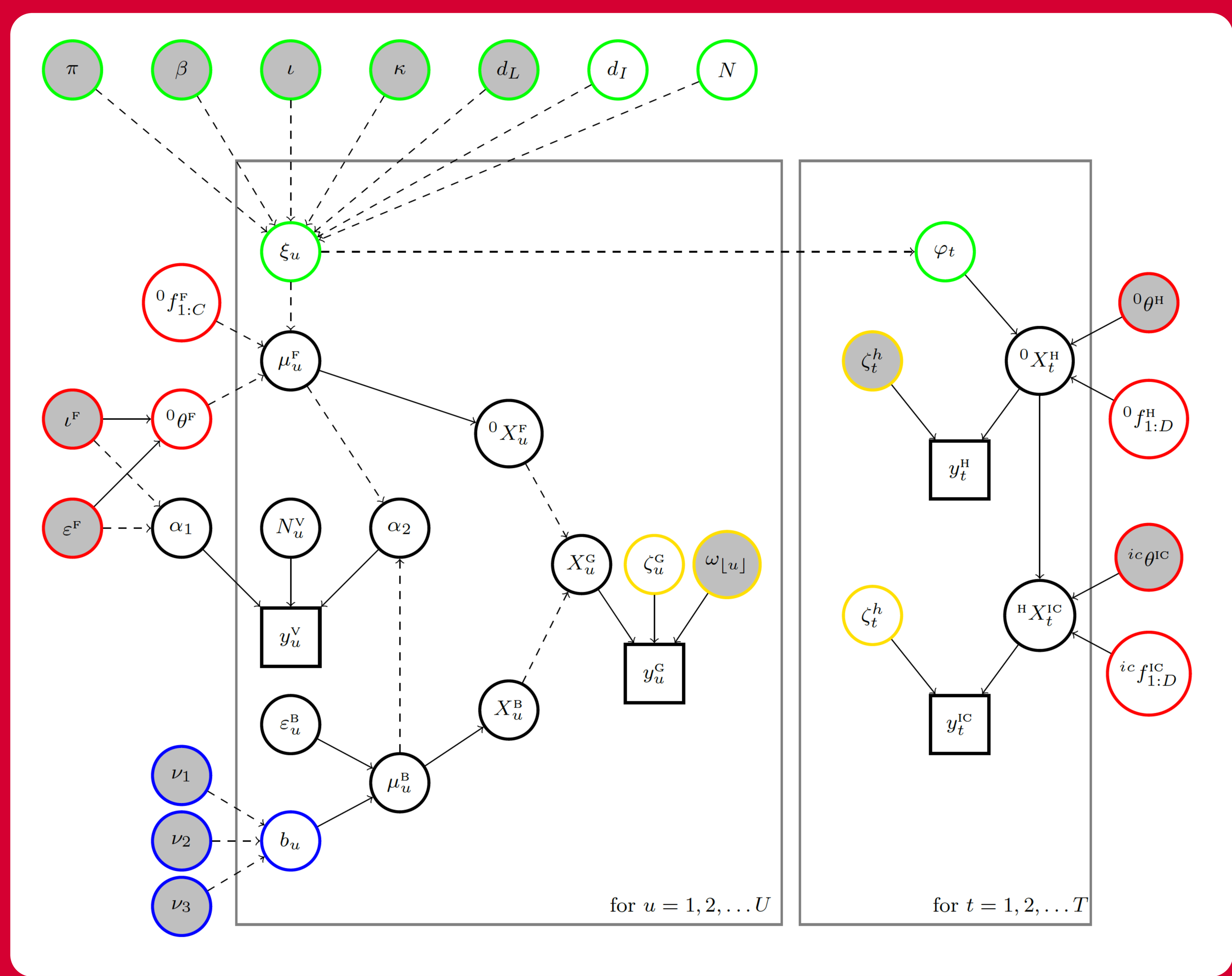
- background non flu ILI cases:

$$\left(X^{B}_{u}\middle|\mu^{B}_{u}\right) \sim \text{Pois}\left(\mu^{B}_{u}\right), \text{ with } \mu^{B}_{u} \text{ r.v. centred in } b_u = e^{\left\{\nu_1+\nu_2\cos\left(\frac{2\pi t_u}{52}\right)+\nu_3\sin\left(\frac{2\pi t_u}{52}\right)\right\}}$$

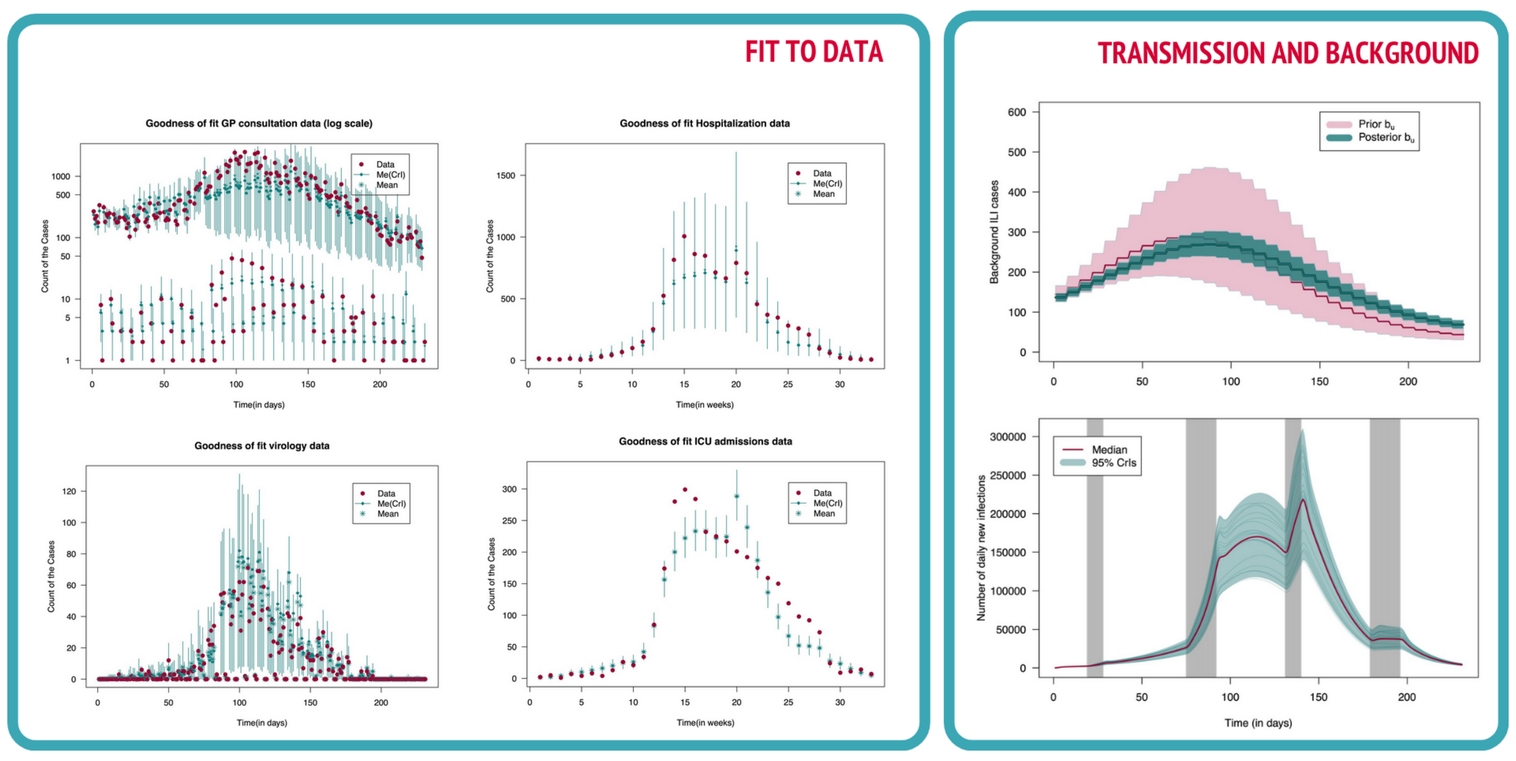
- Accounts for discrete delay between events
- Day-of-the-week effect on the detection parameter $\omega_{\lfloor u \rfloor}$
- Virology data distribution

$$\left(Y^{V}_{u}\middle|N^{V}_{u}, \mu^{F}_{u}, \mu^{B}_{u}\right) \sim \text{Binom}\left(N^{V}_{u}, \frac{\mu^{F}_{u}}{\mu^{F}_{u}+\mu^{B}_{u}}\right)$$

Model DAG



Results



Main Findings and Discussion

Findings

- general set up for epidemic inference as SSMs
- inferential tool for dependent data
- showed that it is key to account for dependence
- application to multiple dependent data on flu in 2017/18

What next?

- Improve the exploration of the parameter space with more efficient MCMCs
- Test the methods on models for prediction

Reference

Corbella A, Birrell PJ, Presanis AM and De Angelis D, **Inferring Epidemics from Multiple Dependent Data via Pseudo-Marginal Methods**. Submitted to *The Annals of Applied Statistics*.