Infectious disease modelling: Move beyond deterministic models?

Katie O’Brien, Senior Statistician
Health Service Executive – Health Protection Surveillance Centre

Stochastica Directions Workshop · UCC Cork · April 2026

Background

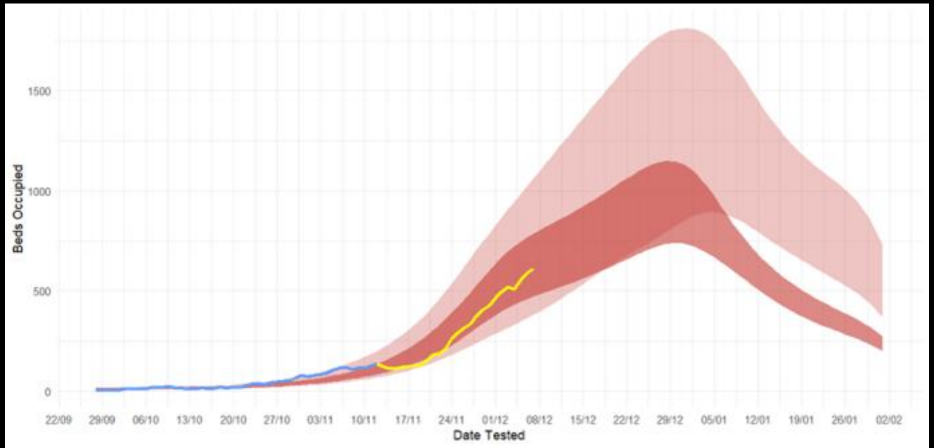
Biostatistics and Modelling Unit @HPSC

- Provide quantitative support for policy and operational decision
- Areas of focus
  - Infectious disease analytics
  - Infectious disease scenario forecasts
  - Environment and one health

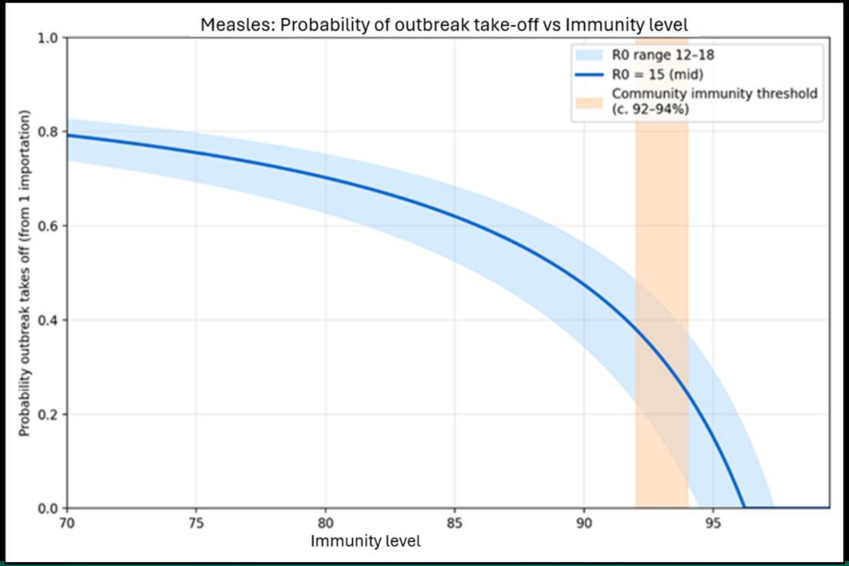
- Working environment
  - Time sensitive decisions
  - Imperfect, delayed and noisy data

Some of our current modelling tools

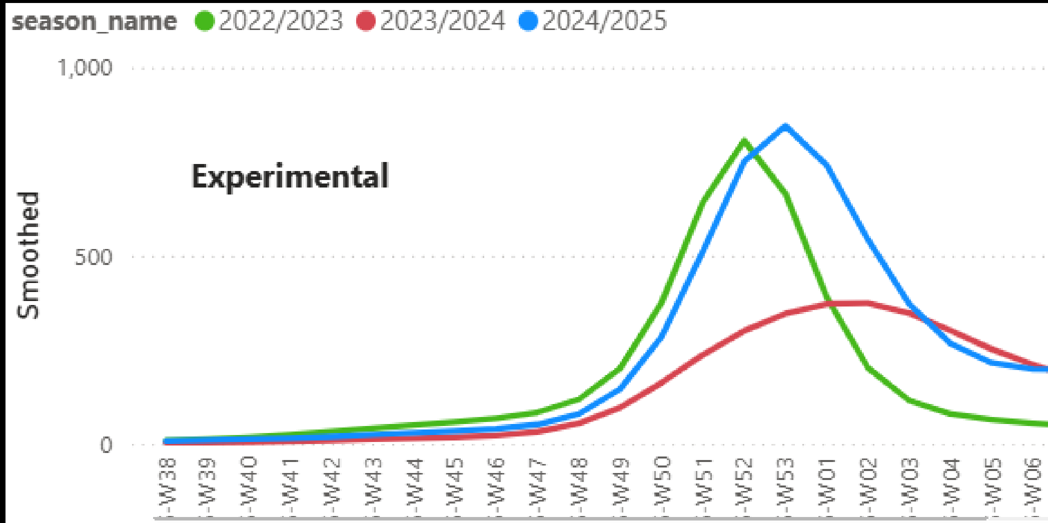
Compartmental models (SEIHR – type)

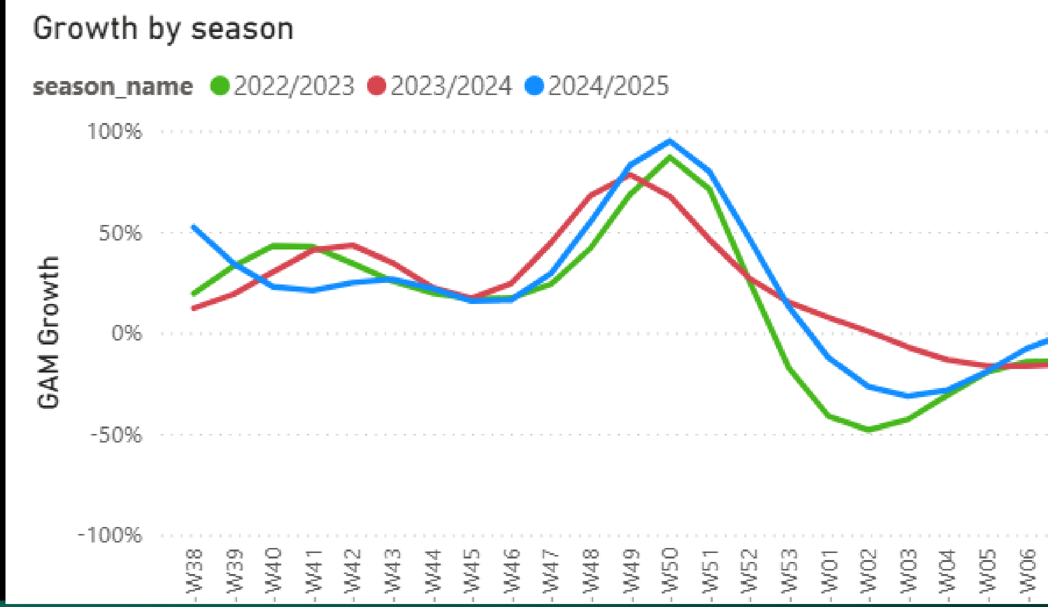


Branching process models



Generalised Additive Models





Why these Models Perform Well in Practice

- Operationally effective & stable
  - Fast to run and update
- Well Aligned with Observed Data
  - While acknowledging data may be sparse, have reporting delays & be biased
- Interpretable
  - Clear outputs for policy makers
- Implicit trade off
  - Simplicity v. realism

Case study: Influenza — what we built

Using **p**artially **o**bserved **M**arkov **p**rocess framework via `pomp` R package:

Latent epidemic process

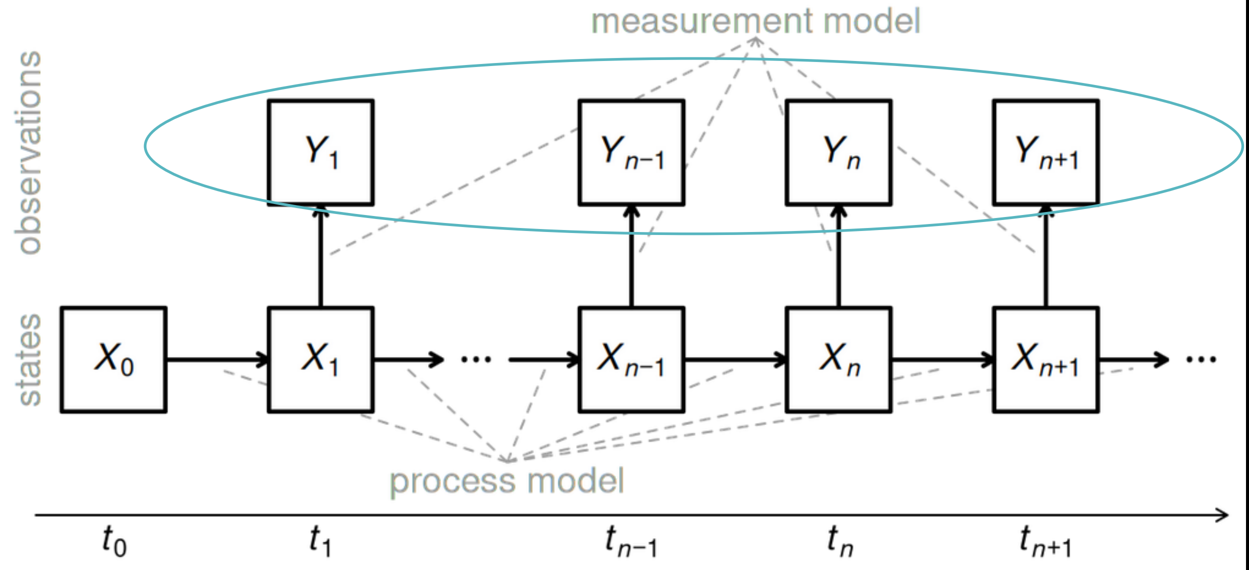
[redacted]

Hidden Markov structure

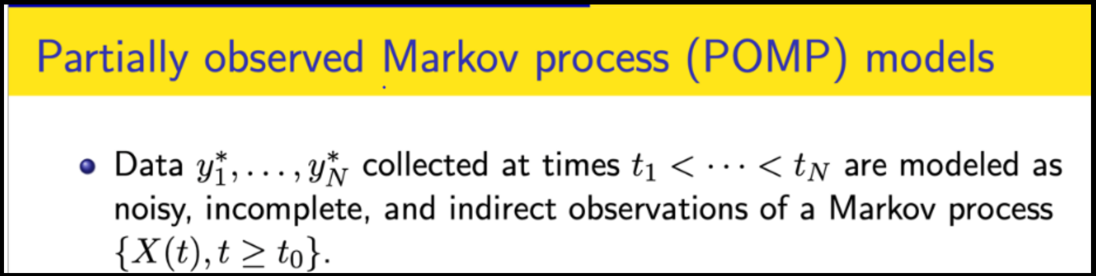
[redacted]

Observation model

[redacted]



[redacted]



Partially observed Markov process (POMP) models

- Data $y_1^*, \ldots, y_N^*$ collected at times $t_1 < \cdots < t_N$ are modeled as noisy, incomplete, and indirect observations of a Markov process $\{X(t), t \geq t_0\}$.

Case study: Influenza — what we built

Using **p**artially **o**bserved **M**arkov **p**rocess framework via `pomp` R package:

Latent epidemic process

This is the unobserved epidemic dynamics governing the true numbers of individuals in each disease state over continuous time
State vector at time t for the flu model:

$$X_t = (S_t, E_t, I_t, H_t^0, H_t^1, R_t)$$

Hidden Markov structure

The latent state X_t evolves over time according to Markov dynamics:

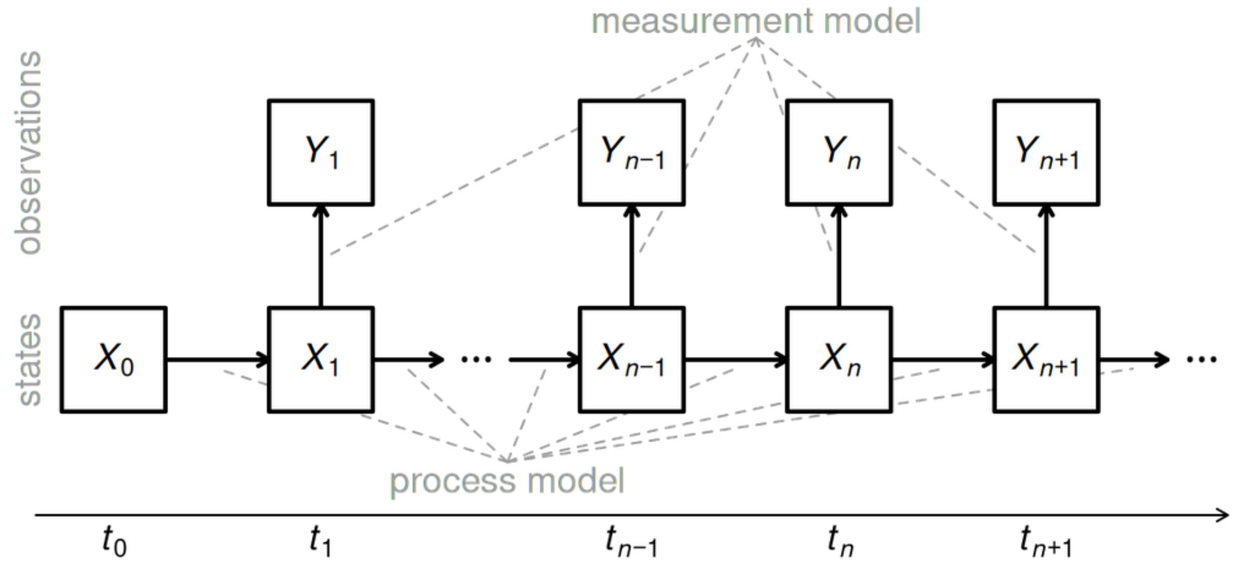
$$\Pr(X_t \mid X_{t-1}, X_{t-2}, \dots) = \Pr(X_t \mid X_{t-1}).$$

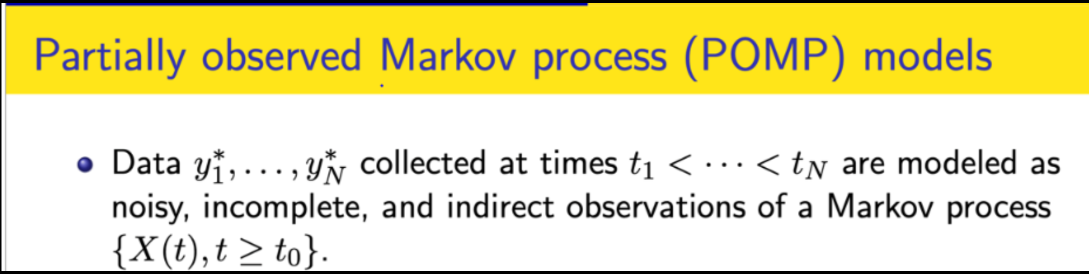
This links the unobserved true epidemic dynamics with observed reported flu data

Observation model

Influenza hospitalisations as observed data $\sim$ Negative binomial

Hospitalised cases (7 day moving average of new hospitalisations) are linked to the latent epidemic state via an observation model.



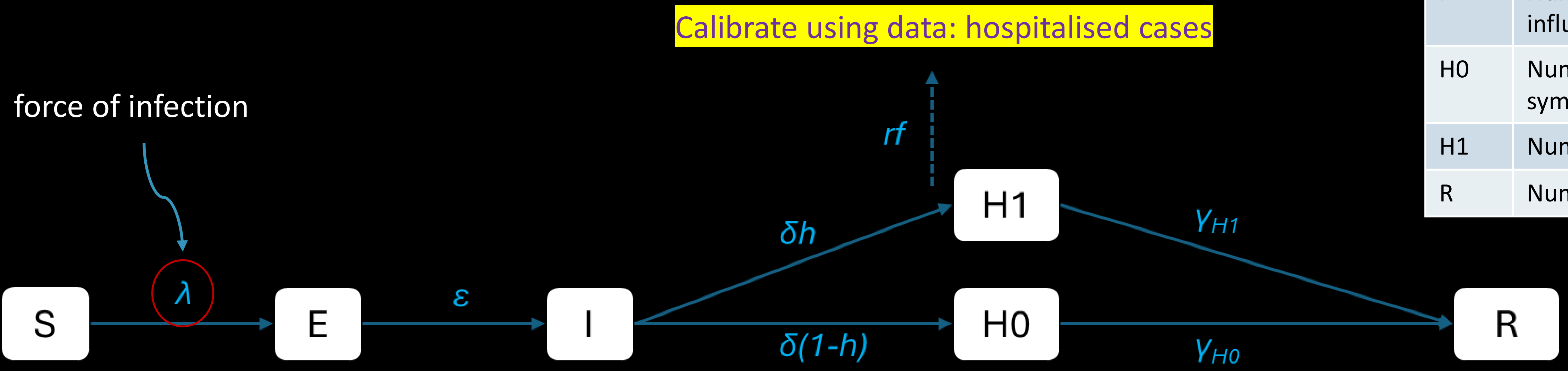


Partially observed Markov process (POMP) models

- Data $y_1^*, \dots, y_N^*$ collected at times $t_1 < \dots < t_N$ are modeled as noisy, incomplete, and indirect observations of a Markov process $\{X(t), t \geq t_0\}$.

Case study: Influenza (underlying model structure)

S	number of susceptibles in Population
E	Number exposed to influenza
I	Number infected by influenza
H0	Number with mild symptoms
H1	Number hospitalised
R	Number Recovered



$$\lambda = \frac{(I + H_0)\beta_t}{N}$$

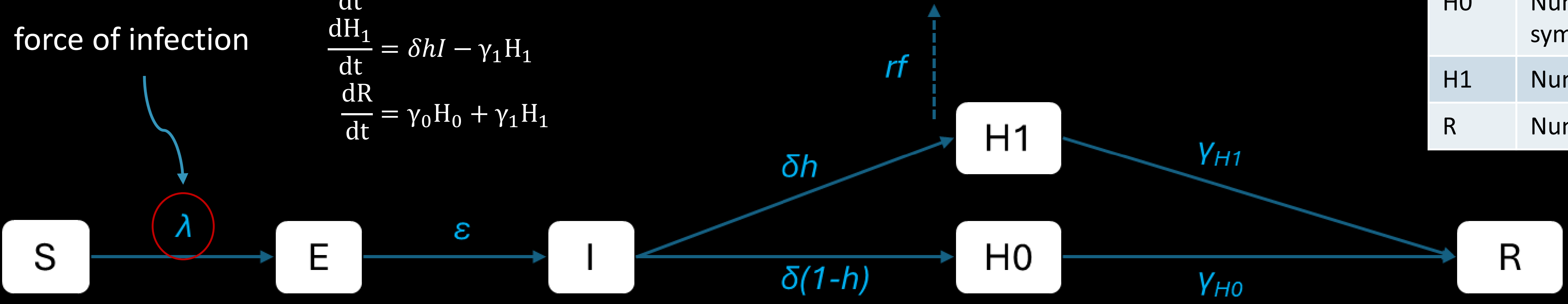
The parameter β_t is the **time-varying disease transmission rate,** the estimation of which is one of the objectives of the modelling.

Case study: Influenza (underlying model structure)

$$\frac{dS}{dt} = -\lambda S$$
$$\frac{dE}{dt} = \lambda S - \epsilon E$$
$$\frac{dI}{dt} = \epsilon E - \delta I$$
$$\frac{dH_0}{dt} = \delta(1-h)I - \gamma_0 H_0$$
$$\frac{dH_1}{dt} = \delta h I - \gamma_1 H_1$$
$$\frac{dR}{dt} = \gamma_0 H_0 + \gamma_1 H_1$$

S	number of susceptibles in Population
E	Number exposed to influenza
I	Number infected by influenza
H0	Number with mild symptoms
H1	Number hospitalised
R	Number Recovered

Calibrate using data: hospitalised cases



$$\lambda = \frac{(I + H_0)\beta_t}{N}$$

The parameter β_t is the **time-varying disease transmission rate**, the estimation of which is one of the objectives of the modelling.

Case study: Influenza

$\beta\_(t)$ is the time-varying disease transmission rate

We treat it as piece-wise constant.

PNAS

Detecting changepoints in dynamical systems: Modelling time-varying transmission of seasonal influenza

Ajay N. Oza[1], Katie M. O'Brien[1*], James P. Gleeson[1,2]

[1]HSE-Health Protection Surveillance Centre, Dublin, D01 A4A3, Ireland

[2]Mathematics Applications Consortium for Science and Industry (MACSI), Department of Mathematics and Statistics, University of Limerick, V94 T9PX, Ireland

*Corresponding author: Katie M O'Brien

Email: katie.obrien@hpsc.ie

hpsc

Case study: Influenza — what we built

Using **p**artially **o**bserved **M**arkov **p**rocess framework via `pomp` R package:

Latent epidemic process

This is the unobserved epidemic dynamics governing the true numbers of individuals in each disease state

State variables typically include:

$$X_t = (S_t, E_t, I_t, H_t^0, H_t^1, R_t)$$

Hidden Markov structure

The latent state X_t evolves over time according to Markov dynamics:

$$\Pr(X_t \mid X_{t-1}, X_{t-2}, \ldots) = \Pr(X_t \mid X_{t-1}).$$

This links the unobserved true epidemic dynamics with observed reported flu data

Observation model

Influenza hospitalisations as observed data $\sim$ Negative binomial

Observations (7 day moving average of new hospitalisations) are linked to the latent epidemic state via an observation model.

What works well

Captures seasonal epidemic shape and peak timing for flu reasonably well

Hospitalisation data gives cleaner signal than case counts - less affected by testing rate variation

Current limitation

Deterministic skeleton cannot reproduce stochastic extinction or variability in outbreak size between seasons with similar parameters

Underestimates uncertainty in early-season forecasts and end-of-season fade-out

The natural extension ?

Replace deterministic skeleton with an SDE or stochastic process

POMP framework already supports stochastic skeletons—iterated filtering is available

Open Q: which SDE formulation is most tractable for inference here?

SDEs in This Context (High-Level View)

Natural stochastic extension introduces process noise:

$$dS_t = -\beta S_t I_t \, dt \; + \; \sigma_S(S_t, I_t) \, dW_t^{(S)},$$

$$dI_t = (\beta S_t I_t - \gamma I_t) \, dt \; + \; \sigma_I(S_t, I_t) \, dW_t^{(I)}.$$

Drift term → deterministic epidemic dynamics

Diffusion term → intrinsic randomness / unobserved variability (**magnitude of process noise**, allowing for stochastic fluctuations due to unobserved heterogeneity, random contact patterns, or other sources of intrinsic variability) . How should σ be specified?

Wiener processes $W_t^{(S)}$ and $W_t^{(I)}$ represent continuous-time stochastic shocks

Current workflows often incorporate uncertainty via:

- Parameter uncertainty
- Observation noise

SDEs instead model uncertainty at the process level

Possibilities for diffusion term

- **Diffusion approximation to infection events**
- $\boxed{\sigma_{\text{inf}}(S_t, I_t) = \sqrt{\beta_t S_t I_t}}$
- Used in:
- $$dS_t = -\beta_t S_t I_t \, dt - \sqrt{\beta_t S_t I_t} \, dW_t^{(\ \ )}$$
  $$dI_t = (\beta_t S_t I_t - \gamma I_t) \, dt + \sqrt{\beta_t S_t I_t} \, dW_t^{(\ \ )}$$
- **Interpretation**
- Noise is shared between S and I, preserving mass balance.

Hospitalisation-driven process noise

$$\boxed{\sigma_H(I_t) = \sqrt{\delta h I_t}}$$

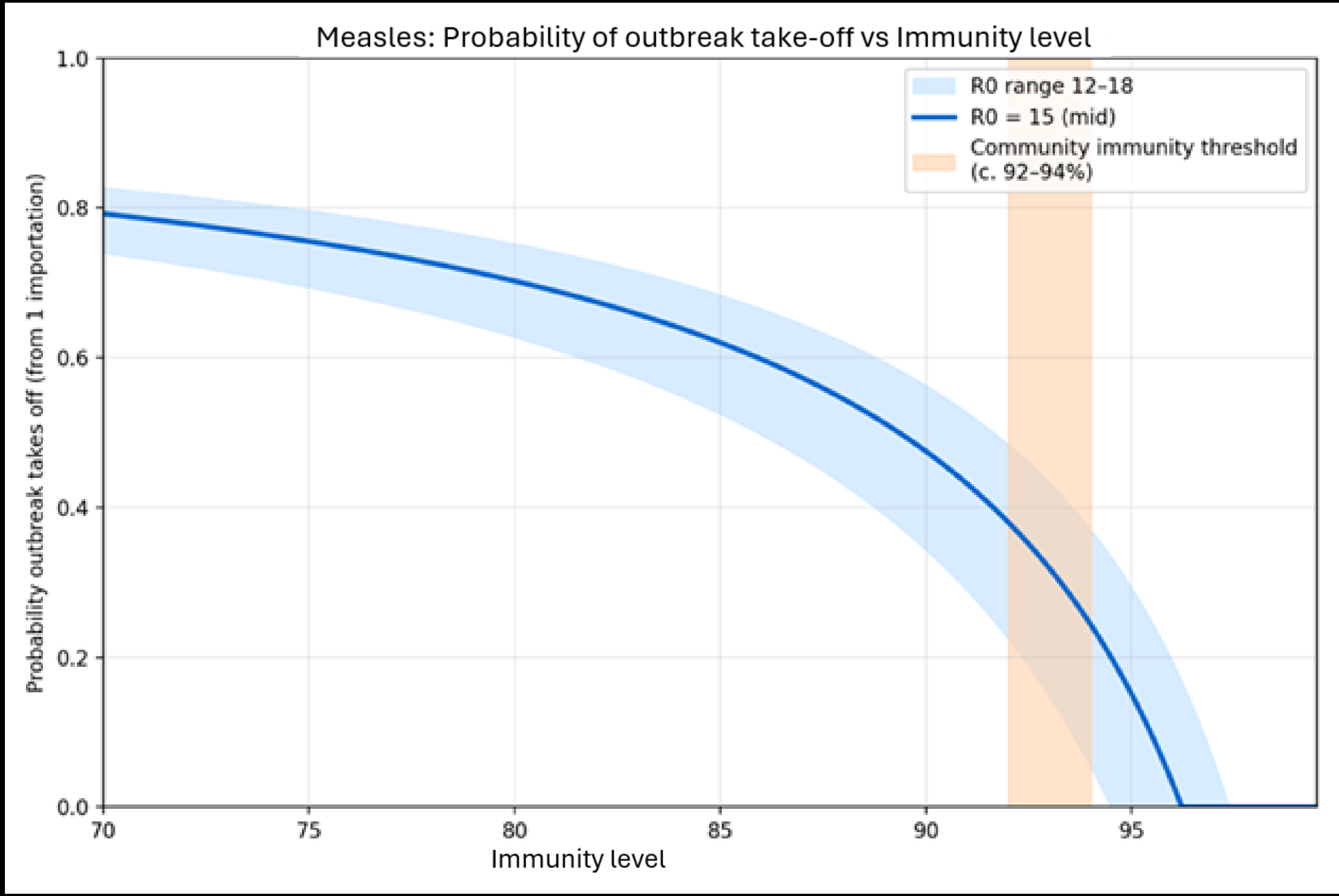
For the hospitalised compartment:

$$dH_{1,t} = \delta h I_t \, dt + \sqrt{\delta h I_t} \, dW_t^{(H)} - \gamma_1 H_{1,t} \, dt$$

Interpretation

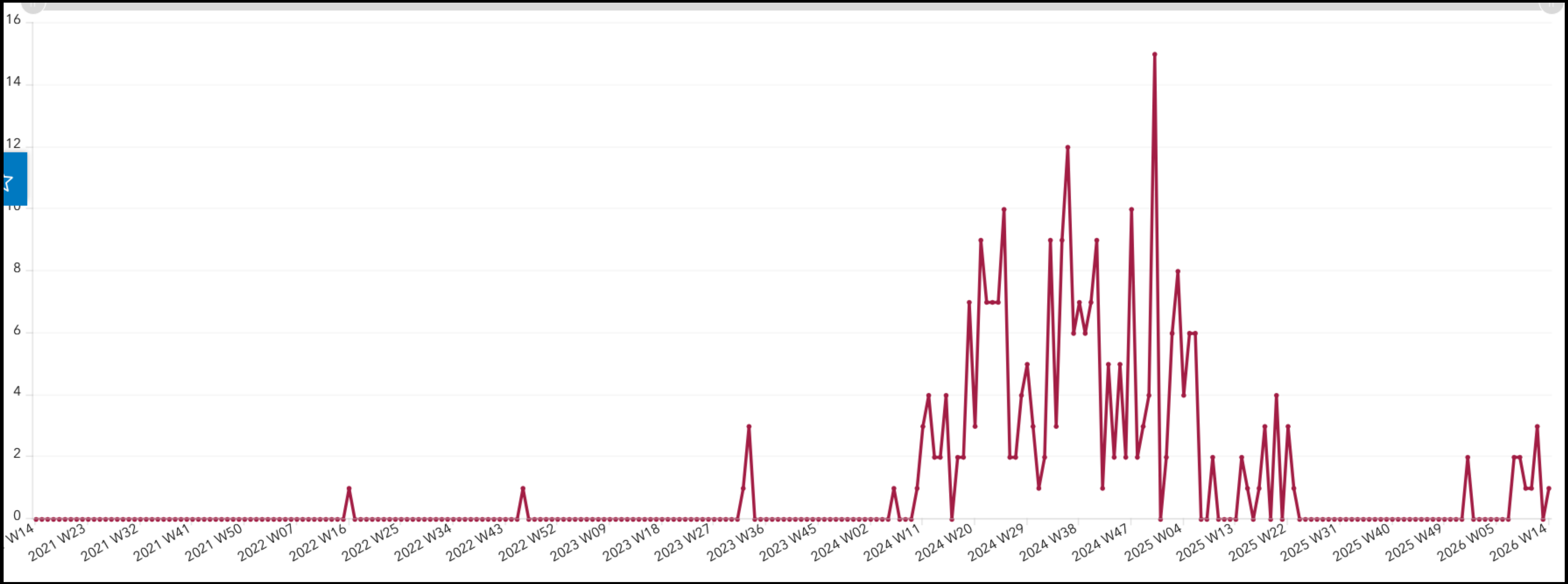
- Captures randomness in which infected individuals get hospitalised
- Reflects:
  - triage variability,
  - Potential outbreak clustering in nursing homes
  - random severe-case occurrence.

Case study: Measles - in a highly (but unevenly) vaccinated population



Branching process model to examine the probability of an outbreak take off, given one imported case

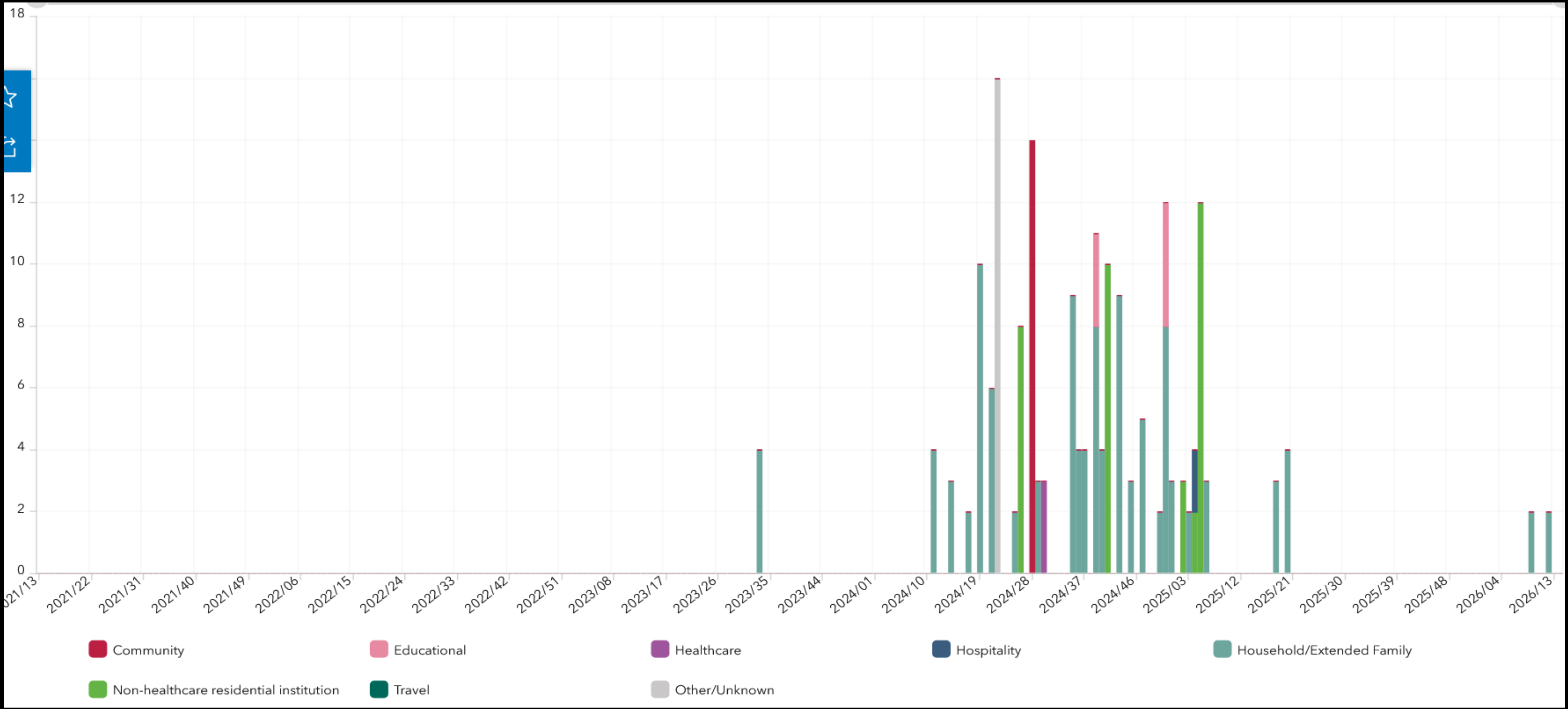
Case study: Measles - in a highly (but unevenly) vaccinated population



https://notifiabledisease.hpsc.ie/

Notified measles cases to the HPSC, Ireland 2021 - 2026

Case study: Measles - in a highly (but unevenly) vaccinated population



https://outbreaks.hpsc.ie/

Measles outbreaks (two or more linked cases) Ireland 2021 - 2026

Challenges and Gaps

- **Where we are**

Toolkit is pragmatic and operationally effective
Focus has been on **delivery and reliability**

- **What we encounter**

Continuous-time processes vs discrete observations
Need for better **uncertainty representation**
Increasing relevance of **small outbreaks / heterogeneity**

- **The question**

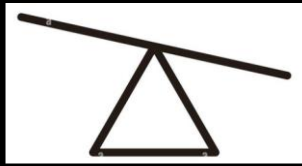
"What modelling approaches are we underutilising or overlooking?"

Practical Lens & Discussion

For new methods (incl. SDEs), key criteria:
- Improves decision-making, not just model fidelity
- Outputs are interpretable and communicable

New methods need to work with real-world data constraints
- partial observability and sparse surveillance data
- Computationally tractable for operational timelines

Trade-off:
- Added realism vs usability



Question:

Where have SDE-based approaches delivered real-world value under these constraints?

Acknowledgements / Buíochas

- Co-authors: Ajay Oza (HPSC) and James Gleeson (University of Limerick)

- Respiratory Virus Unit @HPSC
- Lisa Domegan

- Vaccine Preventable Diseases Team @HPSC
- Sarah Gee and Michael Carton

- Cathal Walsh (Trinity College Dublin)
- Darren Dahly (University College Cork)