*STOCHASTICA Directions Workshop*
*University College Cork, 27th – 29th April 2026*

# Opportunities and barriers in application of stochastic differential equations to agriculture and ecology

*Glenn Marion (Biomathematics and Statistics Scotland, UK)*

*glenn.marion@bioss.ac.uk*


**This talk features joint work with many colleagues including:**
Chris Pooley (BioSS), Stephen Bishop (Roslin Inst.), Nikos Alexandridis (Lund), Sonia Mitchell (BioSS), Lucio Marcello (BioSS) Lesley Smith (SRUC), Mike Hutchings (SRUC), Alex Corbishley (Edinburgh), Andrea Doeschl-Wilson (Edinburgh), Kokouvi Gamado (BioSS), Thibuad Porphyre (Lyon), Stephen Catterall (BioSS), Siu Yin 'Max' Lau (Emory), George Streftaris (Heriot-Watt), Gavin Gibson (Heriot-Watt), Martin Knight (BioSS), Piran White (York) and Ross Davidson (SRUC).

# Aims –

**Review our applications in agriculture and ecology**

**Highlight challenges & opportunities for STOCHASTICA**

## Population modelling
Demographic stochasticity and realistic noise terms

## Modelling heterogeneity
The need to represent heterogeneity

## Spatial models
Spatial dynamics and spread

## Individual and agent-based models
Individual behaviours shape system dynamics

## Assessing model fit to data
Challenges for modelling fitting

Inference for spatial models

Assessing model fit

## Comment –

**To consider in terms of identifying challenges & opportunities for STOCHASTICA**

**In our applications typically make more use of discrete state-space stochastic processes**

**Why is this?**

- **Biological realism?**
- **Lack of appropriate results?**
- **Lack of suitable tools?**
- **Lack of awareness of suitable tools and results?**
- **Perceived difficulty in applying tools and results that are available?**

# Background concepts

*glenn.marion@bioss.ac.uk*

**Discrete state stochastic processes**

**Bayesian inference for latent stochastic processes**

# Discrete State Space Markov Processes

*DCTMPs*

- N compartments, integer populations n(t)
- E event types, each alters state via Q
- Rates r(t) depend only on current state
- Simulated via Doob–Gillespie algorithm

Pooley, Bishop & Marion
J. R. Soc. Interface (2015)

## Section 2 — General Framework & SIS Illustration

### General DCTMP (Eq. 2.1)

$$n_c = n_c + Q_{ce}$$

### State vector n(t)

N compartments, each holding an integer population count.
Space of configs is discrete.

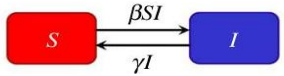
### Transition matrix Q

Fixed matrix. row e=event gives the change to each compartment when event type e fires.

### Rate vector r(θ, n)

Average rate of each event.
Depends on current state n(t) and parameters θ only (Markov property).

### SIS Example (Eq. 2.2)




**n** — $(S, I)^T$ — susceptible & infected counts

**θ** — $(\beta, \gamma)$ — infection & recovery rates

**r** — $(\beta SI, \gamma I)$ — event rates

**Q** — $\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$ — infect: S↓, I↑ recover: I↓, S↑

**Markov property:** rates r depend only on current state n(t), not history. Enables exact simulation via Doob–Gillespie and analytical expressions for event-sequence probability.

# Bayesian Inference Framework

*Combining Parameters, Event Sequences & Data*

**Posterior:**

$$\pi(\theta, \xi \mid y) \propto \pi(y|\xi) \cdot \pi(\xi|\theta) \cdot \pi(\theta)$$

- $\pi(y|\xi)$ Observation model
- $\pi(\xi|\theta)$ Latent process likelihood
- $\pi(\theta)$ Prior on parameters

## Bayesian Framework for stochastic process models

**Prior $\pi(\theta)$**

Encodes prior knowledge about parameters θ.
Flat (uninformative) prior assumed here.

⇒

**Latent Process $\pi(\xi|\theta)$**

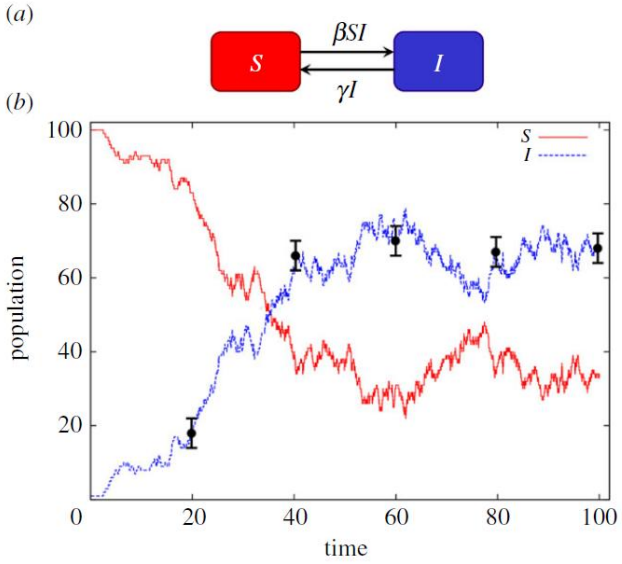
Event sequence ξ simulated via e.g. Doob–Gillespie.
Probability gives weight to sequences consistent with θ.

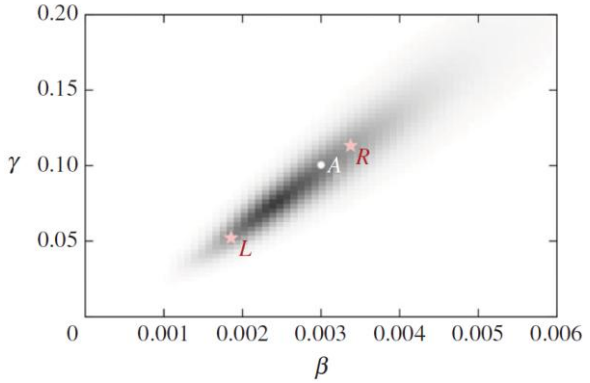
⇒

**Observation model $\pi(y|\xi)$**

Noisy measurements y compared to model state ξ.
Normal errors $N(v; n, \sigma^2)$.
Upweights sequences near data.

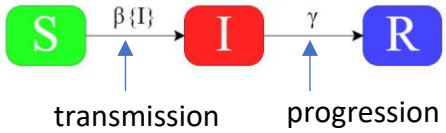
**Posterior $\pi(\theta, \xi \mid y)$ — probability distribution over parameters and event sequences, given the data**




**Key issue**: efficient sampling of joint space of parameters and events

# Computational stats and epi-models




**epi-models**: disease transmission and progression

**data**: real world data provides (multiple) part(s) of the picture e.g. case reports when symptoms show

**Bayesian approach**: allows use of prior knowledge

**computational statistics**: provide practical way to combine the above to estimate uncertainties involved and predict the unseen e.g. transmission and progression events

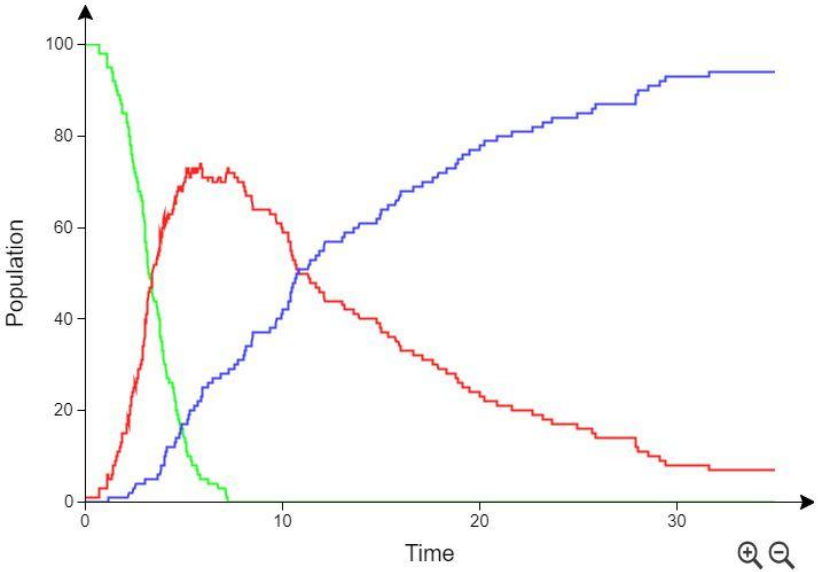
**We talk about *inference* because data is incomplete and noisy and do doesn't provide a complete picture**

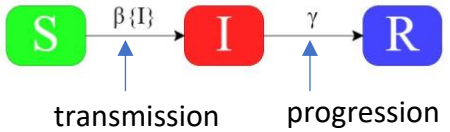
**Models + statistical tools applied to data F I L L I N T H E G A P S**

**Uncertainty in our knowledge is better quantified but remains**

# Computational stats and epi-models

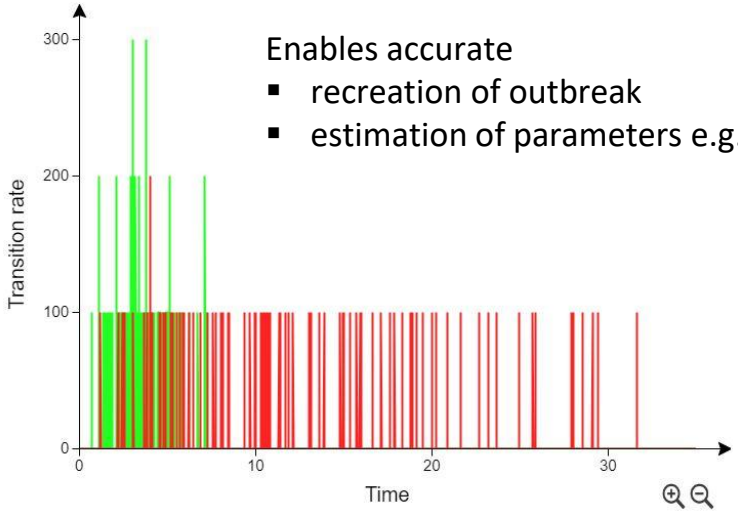
**Intuition:** frequent high quality observation




**epi-model**

Translates into knowledge of (almost) all events

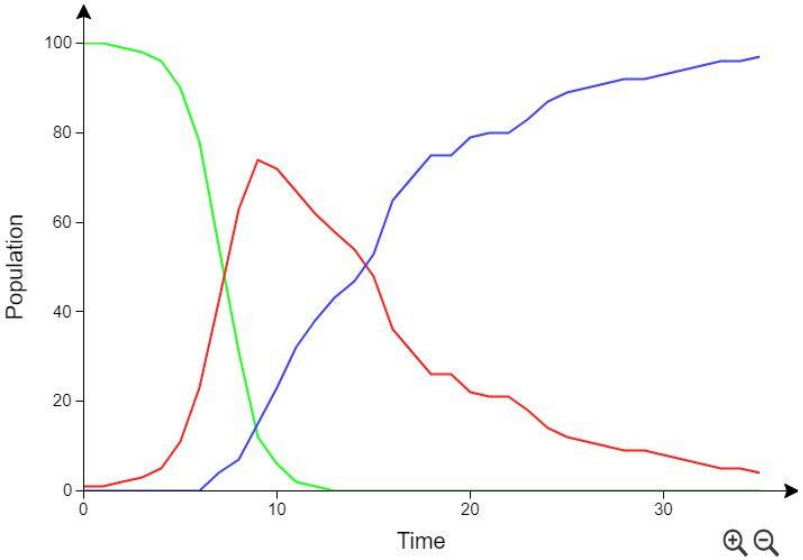



Enables accurate

- recreation of outbreak
- estimation of parameters e.g. β , γ

# Computational stats and epi-models

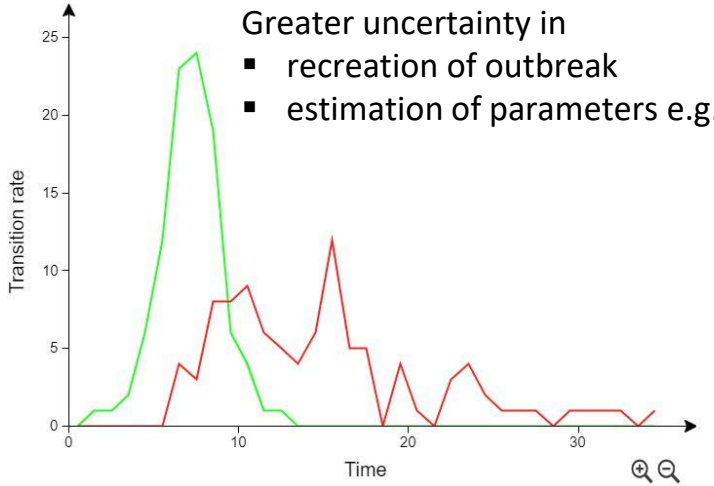
**Intuition:** less frequent observation





**epi-model**

Translates into a (very) incomplete picture




Greater uncertainty in

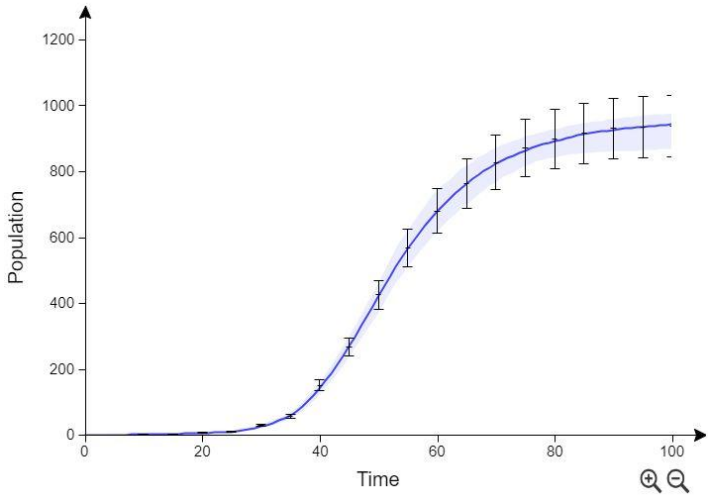
- recreation of outbreak
- estimation of parameters e.g. β , γ

# Computational stats and epi-models

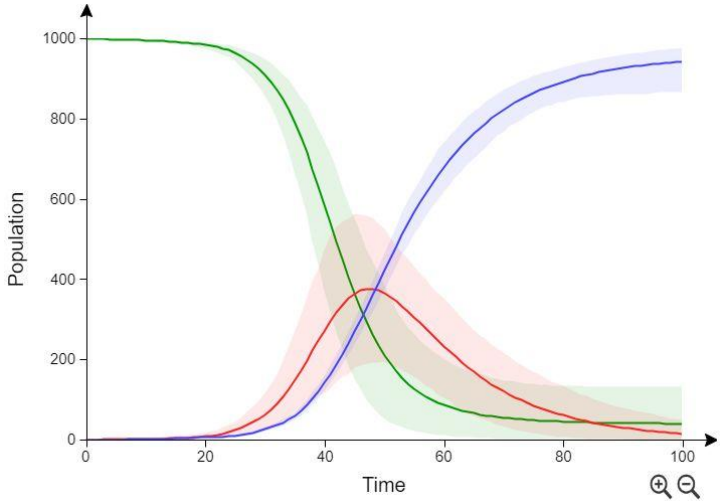
In reality have **imperfect and partial** data


**epi-model**

**data**: periodic observation of 'recovered' individuals




**inference:** estimated uncertainties



**Developments and capabilities in inference for epidemic models**

- Estimate hard to measure quantities
- Quantify disease risk from outbreak data to inform control
- Methods for faster computation: larger problems and integration of multiple data types
- Novel insights into disease dynamics – potential to inform control
- Prospects – current and future work

# Opportunities and barriers in application of stochastic differential equations to agriculture and ecology

*glenn.marion@bioss.ac.uk*

## Population modelling
Demographic stochasticity and realistic noise terms

## Modelling heterogeneity
The need to represent heterogeneity

## Spatial models
Spatial dynamics and spread

## Individual and agent-based models
Individual behaviours shape system dynamics

## Assessing model fit to data
Challenges for modelling fitting

Inference for spatial models

Assessing model fit

# Opportunities and barriers in application of stochastic differential equations to agriculture and ecology

*glenn.marion@bioss.ac.uk*

## Population modelling
Demographic stochasticity and realistic noise terms

## Modelling heterogeneity
The need to represent heterogeneity

## Spatial models
Spatial dynamics and spread

## Individual and agent-based models
Individual behaviours shape system dynamics

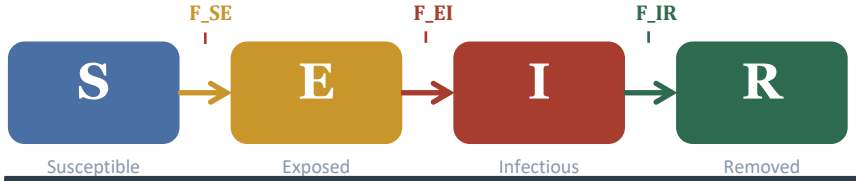
## Assessing model fit to data
Challenges for modelling fitting

Inference for spatial models

Assessing model fit

# The SDE-SEIR (non-spatial/population level) Model

*Demographic stochasticity via Itô SDEs — three independent Wiener processes drive three reaction fluxes*




**Parameters** $\theta = (\beta, \sigma, \gamma, N, \rho, \phi, S_0, E_0, I_0)$

| | |
|---|---|
| $\beta$ — transmission rate (day$^{-1}$) | $N$ — total (closed) population |
| $\sigma$ — E→I transition rate (day$^{-1}$) | $\rho$ — reporting probability $\in (0,1)$ |

**Three reaction fluxes (Chemical Langevin Equation)**

$$F_{SE} = \beta \cdot S \cdot I/N \cdot dt + \sqrt{(\beta \cdot S \cdot I/N)} \cdot dW_{SE}$$

S → E: density-dependent infection; variance ∝ flux magnitude

$$F_{EI} = \sigma \cdot E \cdot dt + \sqrt{(\sigma \cdot E)} \cdot dW_{EI}$$

E → I: linear transition; noise collapses to zero when E = 0

$$F_{IR} = \gamma \cdot I \cdot dt + \sqrt{(\gamma \cdot I)} \cdot dW_{IR}$$

I → R: recovery; mirrors E → I structure

**State equations (Itô form, closed population N = S+E+I+R)**

$$dS = -F_{SE}$$

$$dE = +F_{SE} - F_{EI}$$

$$dI = +F_{EI} - F_{IR}$$

$$dR = +F_{IR}$$

**Euler-Maruyama discretisation:**

Brownian increments $dW \sim N(0, \sqrt{\Delta t})$ drawn independently. Positivity enforced by reflection (clamp to zero); Conservation S+E+I+R = N maintained exactly by setting R = N−S−E−I.

# Natural Pest Control

Context, Challenge & the Need for Archetype Models

*Alexandridis et al. (2022)*
*Ecological Applications*

## Why Natural Pest Control Matters

Arthropod pests cause major crop losses worldwide. Predators and parasitoids suppress pest populations naturally — a sustainable alternative to pesticides.

## Landscape Context Drives Outcomes

Landscape composition (noncrop habitat) and configuration (field edge density) regulate how pests and natural enemies move and interact across agricultural systems.

## The Modeling Challenge — Enter Archetypes

Pest responses to landscape change are highly variable across systems — existing models are either too system-specific or too generic to be actionable.

Archetypes group systems sharing key biological traits into mechanistic models of intermediate generality, bridging this gap.

# Defining Archetypes

*Resident Pest (A1) vs. Transient Pest (A2) — key traits driving different landscape responses*

## Archetype A1 — Resident Pest

| | |
|---|---|
| Pest overwinters: | **IN CROP** |
| Enemy diet: | **SPECIALIST** |
| Enemy dispersal: | **PASSIVE (WIND)** |

**System Behaviour:**

More noncrop habitat favours generalists, which outcompete the specialist enemy. Net effect: resident pest abundance is unchanged by landscape complexity.

## Archetype A2 — Transient Pest

| | |
|---|---|
| Pest overwinters: | **IN NONCROP HABITAT** |
| Enemy diet: | **SPECIALIST (FLEXIBLE)** |
| Enemy dispersal: | **PASSIVE (WIND)** |

**System Behaviour:**

Pest migration diversifies the specialist's diet, decoupling it from pest density. Higher field edge density boosts generalists and suppresses the transient pest; noncrop habitat proportion alone has no net effect.

# Stochastic Differential Equation Models

*Full SDE for Archetype A1 — deterministic dynamics (black) plus stochastic noise terms (gold) | A2: set $\varepsilon_s = 0$*

*Archetype A1 — full stochastic differential equations (Alexandridis et al. 2022, Fig. 2c)* ■ deterministic ■ stochastic noise

**Pests**

$$\mathbf{dP(t) =} \left( aP - \mu_P P - f_s PS/(P+K_s) - f_g PG/(P+K_g) + \upsilon_p P_0 - \upsilon_p P \right) dt$$
$$+ \sqrt{(aP)}\, dB_1(t) - \sqrt{(\mu_P P)}\, dB_2(t) - \sqrt{(f_s PS/(P+K_s))}\, dB_3(t) - \sqrt{(f_g PG/(P+K_g))}\, dB_4(t) + \sqrt{(\upsilon_p P_0)}\, dB_5(t) - \sqrt{(\upsilon_p P)}\, dB_6(t)$$

**Specialist predators**

$$\mathbf{dS(t) =} \left( -\mu_S S + \varepsilon_s f_s PS/(P+K_s) + \upsilon_s S_0 - \upsilon_s S \right) dt$$
$$- \sqrt{(\mu_S S)}\, dB_7(t) + \sqrt{(\varepsilon_s f_s PS/(P+K_s))}\, dB_8(t) + \sqrt{(\upsilon_s S_0)}\, dB_9(t) - \sqrt{(\upsilon_s S)}\, dB_{10}(t)$$

**Generalist predators**

$$\mathbf{dG(t) =} \left( -\mu_G G + \upsilon_g G_0 - \upsilon_g G \right) dt$$
$$- \sqrt{(\mu_G G)}\, dB_{10}(t) + \sqrt{(\upsilon_g G_0)}\, dB_{11}(t) - \sqrt{(\upsilon_g G)}\, dB_{12}(t)$$

**Demographic Noise terms:** $dB_i(t)$ — independent Gaussian white noise, drawn from $N(0, \sqrt{dt})$ for each t and i = 1…12

## Parameter Glossary

| | | | | | |
|---|---|---|---|---|---|
| $P$, $S$, $G$ | Pest, specialist & generalist enemy density in crop | $f_s$, $f_g$ | Functional response attack rates (type II) | $\upsilon_p$, $\upsilon_s$, $\upsilon_g$ | Crop–noncrop connectivity rates |
| $a$ | Pest intrinsic growth rate | $K_s$, $K_g$ | Half-saturation constants (½ max attack rate) | $P_0$, $S_0$, $G_0$ | Noncrop habitat carrying capacities |
| $\mu_P$, $\mu_S$, $\mu_G$ | Natural mortality rates | $\varepsilon_s$ | Specialist conversion efficiency (0 in A2) | $dB_i(t)$ | Independent Gaussian white noise increments |

Estimating dynamics of AMR horizontal gene transfer & demography

# Understanding population-level persistence of anti-microbial resistance: the need to quantify the dynamics of mobile genetic elements

Sonia Mitchell, Glenn Marion, Lucio Marcello (BioSS)

Lesley Smith, Mike Hutchings (SRUC)

Alex Corbishley (University of Edinburgh)

# Estimating dynamics of AMR

- Modelling population dynamics and horizontal transfer of AMR genes
- Estimation of key characteristics from timeseries data
- Application to experimental data

# AMR gene demography and HGT

## A simple model – demography and horizontal gene transfer

**Resistant bacteria**

$$\frac{dR}{dt} = \alpha e^{-\phi} R\left(1 - \frac{R+S}{K}\right) + \frac{\beta_H}{K} SR - \gamma R - \mu(1+\epsilon)R$$

**Births** **HGT** **R-gene loss** **Deaths**

**Susceptible bacteria**

$$\frac{dS}{dt} = \alpha S\left(1 - \frac{R+S}{K}\right) - \frac{\beta_H}{K} SR + \gamma R - \mu S$$

# AMR gene demography and HGT

## A simple model – demography and horizontal gene transfer

Costs of resistance

Reduced growth rate

higher death rate

Resistant bacteria

$$\frac{dR}{dt} = \alpha e^{-\phi} R\left(1 - \frac{R+S}{K}\right) + \frac{\beta_H}{K} SR - \gamma R - \mu(1+\epsilon)R$$

Births

HGT

R-gene loss

Deaths

Susceptible bacteria

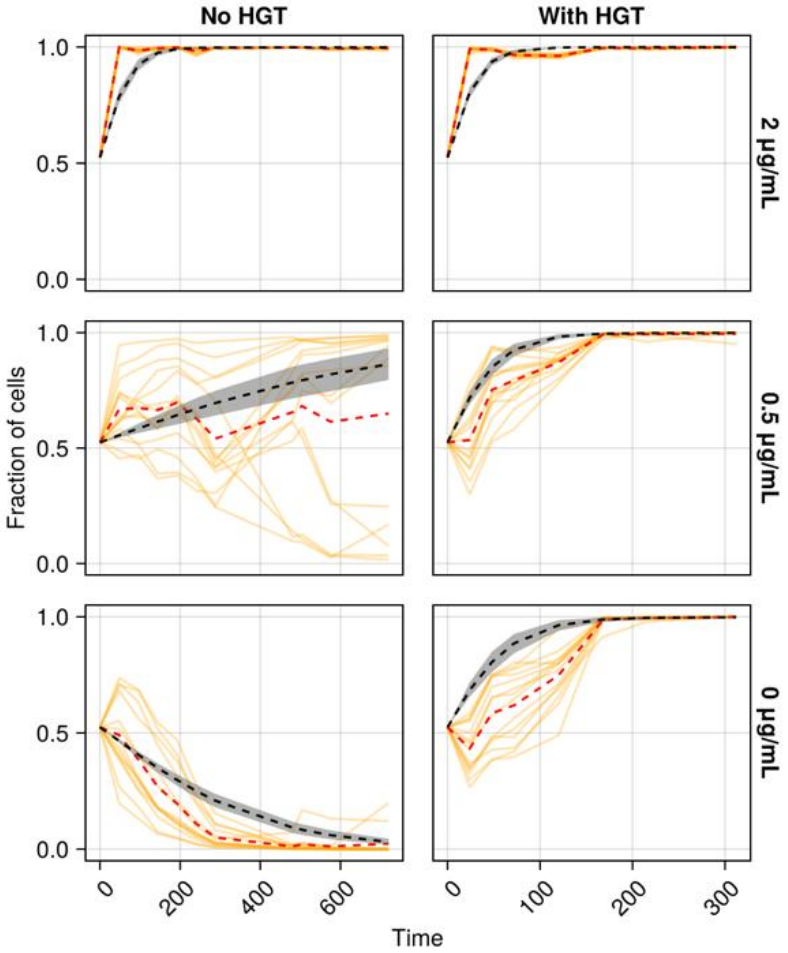
$$\frac{dS}{dt} = \alpha S\left(1 - \frac{R+S}{K}\right) - \frac{\beta_H}{K} SR + \gamma R - \mu S$$

## Can estimate parameters from experimental data

**But estimates do not account for variation in experimental outcomes**

**Demographic stochasticity very small**

**Need to explore environmental noise on different aspects of the model**



# Opportunities and barriers in application of stochastic differential equations to agriculture and ecology

*glenn.marion@bioss.ac.uk*

## Population modelling
Demographic stochasticity and realistic noise terms

## Modelling heterogeneity
The need to represent heterogeneity

## Spatial models
Spatial dynamics and spread

## Individual and agent-based models
Individual behaviours shape system dynamics

## Assessing model fit to data
Challenges for modelling fitting

Inference for spatial models

Assessing model fit

# Age-Stratified COVID-19 Model

*9 Compartments*
*Across 18 Age Groups*

**S → E**
Force of infection $\lambda_{a,t}$ drives exposure

**E branches**
Asymptomatic (A) or Infectious (I) — age-dependent

**I → C → H → D**
Clinical pathway to hospitalisation & death

**Surveys**
PCR & seroprevalence link model states to data
Pooley, Doeschl-Wilson & Marion
Phil. Trans. R. Soc. A (2022)

## Age stratified extended SEIR model for SARS-CoV-2 transmission

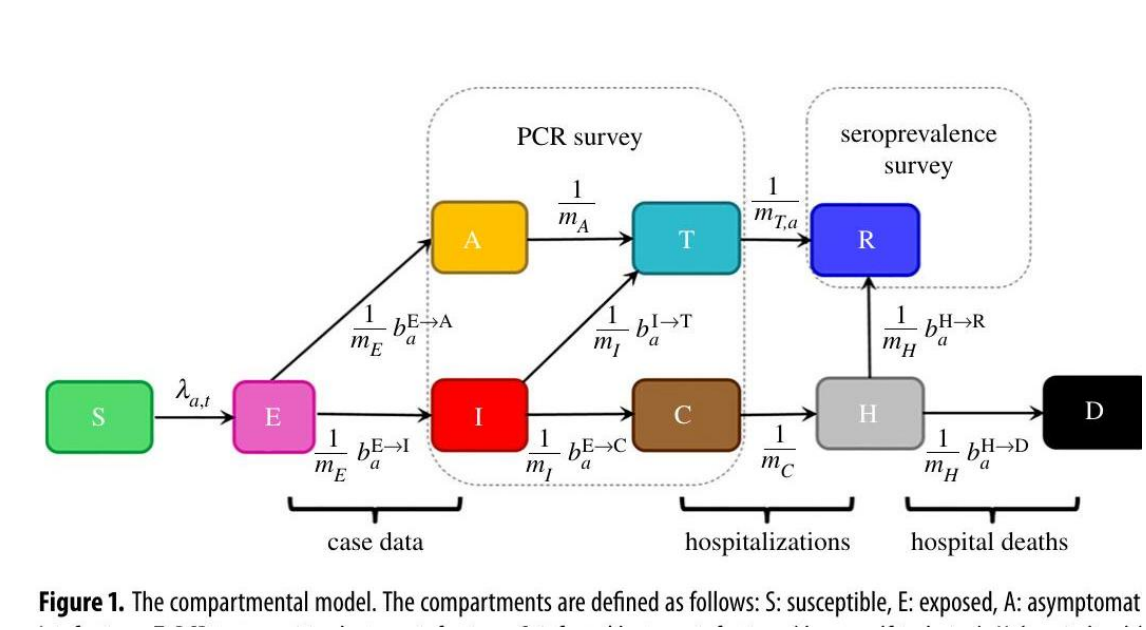


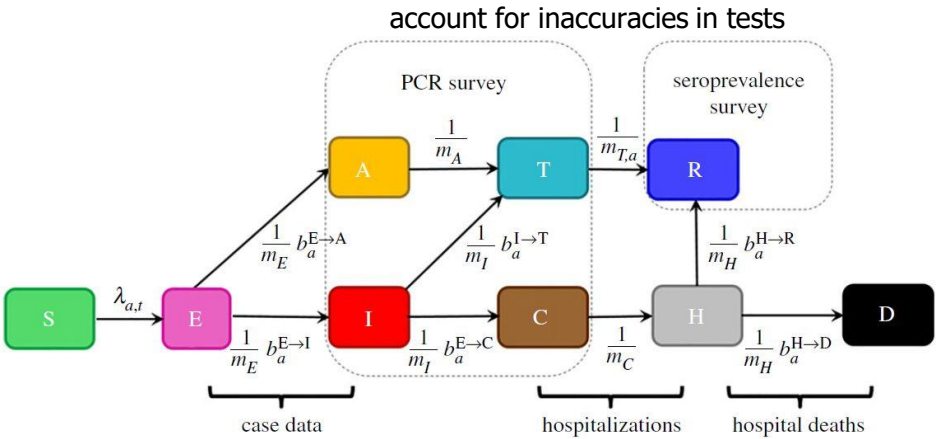

**Figure 1.** The compartmental model. The compartments are defined as follows: S: susceptible, E: exposed, A: asymptomat

| S | E | A | I | T | C | H | R | D |
|---|---|---|---|---|---|---|---|---|
| Susceptible | Exposed | Asymptomatic | Infectious | PCR+, not infectious | Isolated (non-hosp.) | Hospitalised | Recovered | Dead |

**Key insight:** All branching probabilities and transition rates are **age-stratified (subscript a)** — 18 age bands from 0–4 to 80+ plus care home residents.
**Data linkage:** PCR surveys → A, I, T, C states | Seroprevalence → R state | Case / hospital / death records → E→I, C→H, H→D transitions.

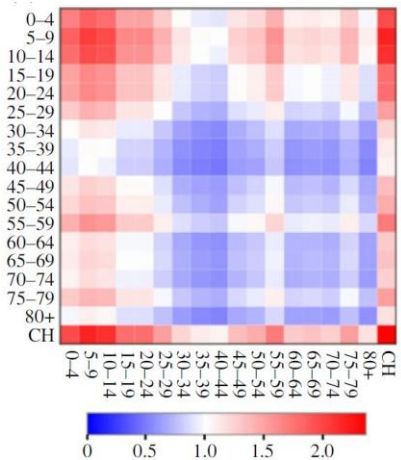
## Methods for faster computation: larger problems and integration of multiple data types

- National-scale COVID models combining multiple data




- Estimation of parameters: age-related contacts

Pre-pandemic based on the BBC Pandemic study




Changes relative to pre-pandemic contact patterns

**Contact patterns critical to effective modelling**

Pooley CM, Doeschl-Wilson AB, Marion G. 2022 Estimation of age-stratified contact rates during the COVID-19 pandemic using a novel inference algorithm. *Phil. Trans. R. Soc. A* **380**: 20210298.

## Many other cases where need to represent heterogeneity

**Natural Pest control**

- many different species and/or
- variation across and between landscapes

**Antimicrobial Resistance (AMR)**

- hundreds of species of bacteria and strains
- many resistance genes
- multiple antimicrobial agents

# Opportunities and barriers in application of stochastic differential equations to agriculture and ecology

*glenn.marion@bioss.ac.uk*

**Population modelling**
Demographic stochasticity and realistic noise terms

**Modelling heterogeneity**
The need to represent heterogeneity

**Spatial models**
Spatial dynamics and spread

**Individual and agent-based models**
Individual behaviours shape system dynamics

**Assessing model fit to data**
Challenges for modelling fitting
Inference for spatial models
Assessing model fit

# Spatial Risk from Small Outbreaks

*Classical Swine Fever*
*East Anglia, UK (2000)*

- Only 16 farms infected
- 1,703 farms at risk in region
- Detection times only (no infection times)
- Goal: spatial risk maps for control

Gamado, Marion & Porphyre
Front. Vet. Sci. (2017)

## The Challenge — Reliable Inference from Limited Data SIR model

### Why small outbreaks matter

Large outbreaks (FMD 2001: £8bn; CSF Netherlands 1997: £1.1bn) are rare. Emerging pathogens typically cause small, repeated, localised incursions — but these are what we observe first.

### The inference problem

Infection times are unobserved. Only farm detection times and locations are available. Data-augmentation MCMC treats missing infection times as latent variables and samples them jointly with model parameters θ.

### Why kernel choice matters

The spatial transmission kernel h(d) determines how infection risk decays with distance. Four candidates ($K_1$–$K_4$) are fitted — and predicted risk maps differ dramatically between them.

## Four Spatial Transmission Kernels h(d) — where d = Euclidean distance between farms

| $K_1$ — $\exp\{-\tau d\}$ | $K_2$ — $1 / (1 + (d/\delta))^{\wedge}$ | $K_3$ — $1 / (1 + d/\delta)$ | $K_4$ — $1 - \exp(-(d/\delta))^{\wedge}$ |
|---|---|---|---|
| Exponential — fastest decay | Power-law (Cauchy-type) — moderate tail | Cauchy — heaviest tail ($K_2$ with $\tau$=1) | Bounded — emphasises short-range |

## Inference pipeline

Farm locations + detection times R → Data-augmented MCMC samples θ, I → Posterior predictive simulations (Gillespie) → Farm-level risk maps

**Key finding:** Latent residuals (Infectious Link Residuals, ILRs) outperform DIC for kernel selection, especially for small outbreaks. For ≥20 cases, ILRs select the correct kernel ≥90% of the time. Risk maps are highly sensitive to which kernel is chosen — making reliable model selection critical for disease control decisions.

# Predicted Infection Risk Maps

*Figure 4 — CSF East Anglia at t = 90 days*

**Dot colour = probability of infection**

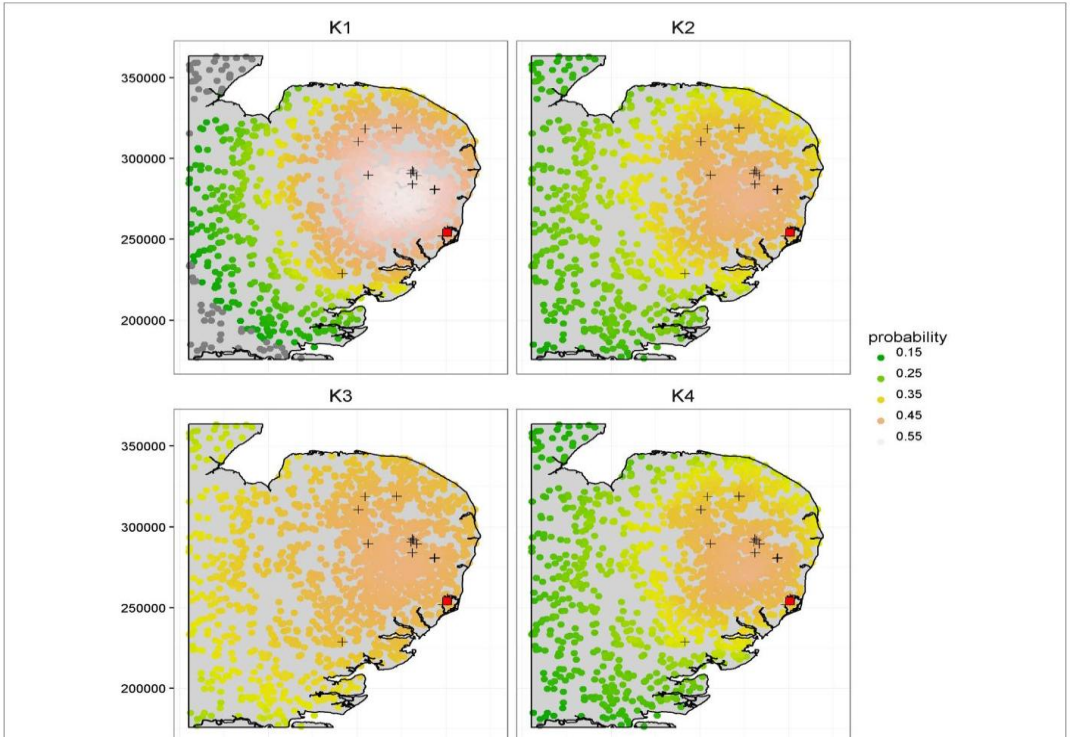
- 0.15 — low risk
- 0.25
- 0.35
- 0.45
- 0.55 — high risk

+ = detected case
■ = index case (red)

Gamado, Marion & Porphyre
Front. Vet. Sci. (2017)

## Kernel Choice Dramatically Alters Spatial Risk Profile






| $K_1$ | $K_2$ | $K_3$ | $K_4$ |
|---|---|---|---|
| Fastest decay — high risk tightly localised around index | Moderate tail — ILRs prefer this; similar to $K_4$ | Heaviest tail — predicts highest risk at all distances | Short-range emphasis — similar risk profile to $K_2$ |

# Inferring Disease Spread with Unknown Populations at Risk

*iPAR framework — Catterall, Porphyre & Marion (2025) PLoS Comput. Biol.*

## The Problem

Spatial epidemic models require knowledge of where the host population lives — but this is often unavailable.

Wildlife populations (e.g. wild boar) are poorly mapped. Livestock records may be incomplete or private.

Case-only data: disease reports exist, but the distribution of susceptible individuals does not.

Without population data, existing transmission models cannot be reliably applied.

## The iPAR Solution

**Landscape covariates**
Land use and other spatial data proxy the unknown host distribution.

**Patch-based SI model**
Estonia divided into 575 patches (10 km²). Each patch is susceptible or infected.

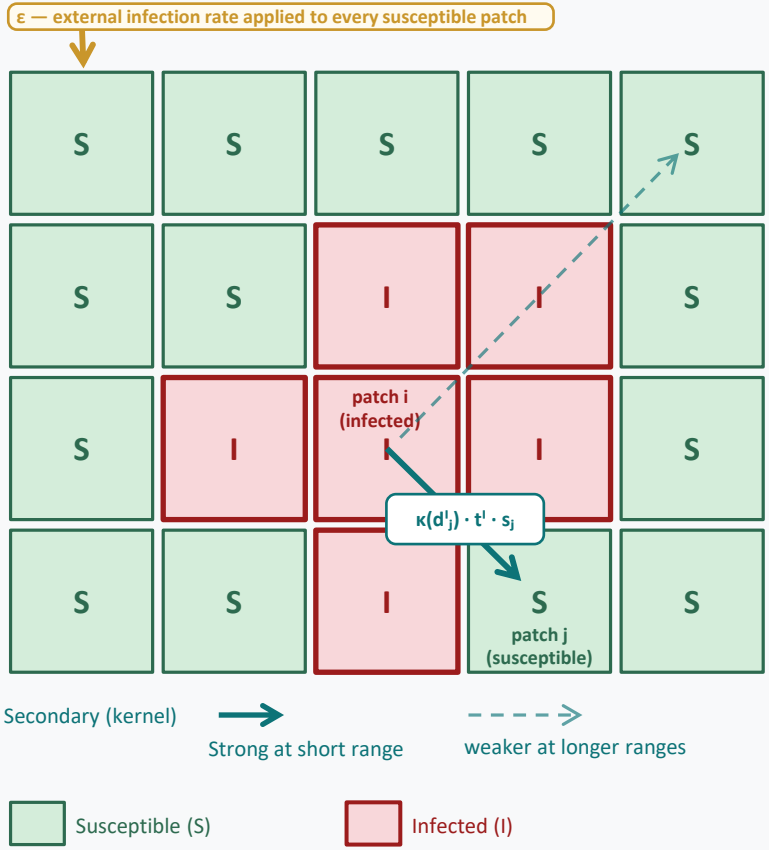
**Inferred surfaces**
Susceptibility and infectivity surfaces estimated from case data, not assumed.

**Bayesian MCMC**
Data-augmentation MCMC fits the model to spatio-temporal case reports.

*Catterall, Porphyre & Marion (2025) PLoS Comput. Biol. 21(6): e1012622 | https://doi.org/10.1371/journal.pcbi.1012622*

# Spatial Model Structure: Patches, Transmission & Covariate Surfaces

*Force of infection · Susceptibility · Infectivity — Catterall et al. (2025)*




## ① Force of Infection on susceptible patch i

$$f_i = \varepsilon s_i + \rho \Sigma_j s_i \cdot t_j \cdot \kappa(d_{ij})$$

ε — background infection rate

ρ — secondary transmission rate

$\kappa(d^{I}_{j})$ — decays with distance $d^{I}_{j}$

## ② Patch Susceptibility $s^{I}$ — weighted by land-use composition

$$s_i = \Sigma_l h_{il} \sigma_l \quad \text{where } \Sigma_l \sigma_l = 1 \text{ (for identifiability)}$$

| $\sigma_1$ | $\sigma_2$ | $\sigma_3$ | $\sigma_4$ | $\sigma_5$ | $\sigma_6$ |
|---|---|---|---|---|---|
| Urban | Agri | Broadleaf | Conifer | Semi-nat | Wetland |

## ③ Patch Infectivity $t^{I}$ — same covariate structure as susceptibility

$$t_i = \Sigma_l h_{il} \gamma_l \quad \text{where } \Sigma_l \gamma_l = 1 \text{ (for identifiability)}$$

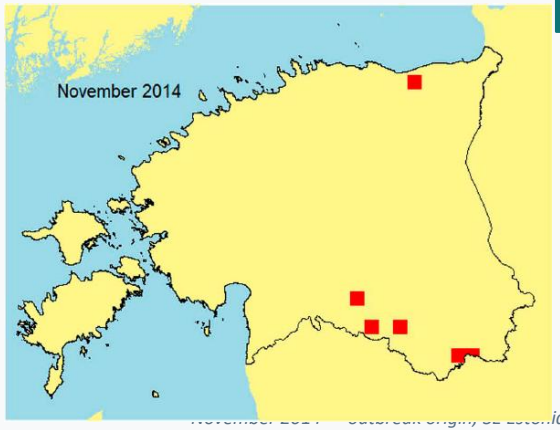
$h_{il}$ — proportion of patch i with land use l

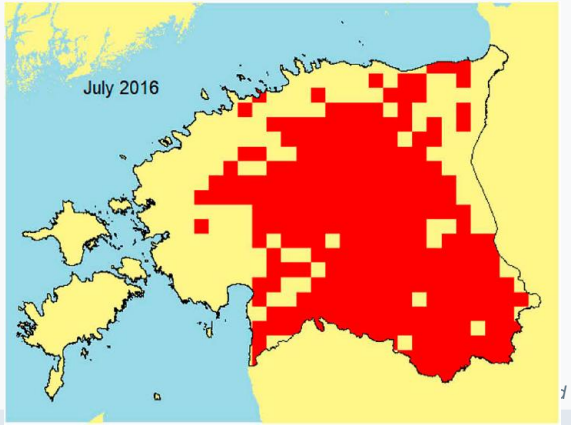
$\gamma_l$ — infectivity weight of land use l (estimated from outbreak data)

*Catterall, Porphyre & Marion (2025) PLoS Comput. Biol. 21(6): e1012622 | https://doi.org/10.1371/journal.pcbi.1012622*

# Case Study: African Swine Fever in Estonian Wild Boar (2014–2019)

*Spatial spread of ASF — Fig. 6, Catterall et al. (2025)*




## Outbreak Context

**Disease:** African Swine Fever (ASF) — near 100% fatal in wild boar; no vaccine available.

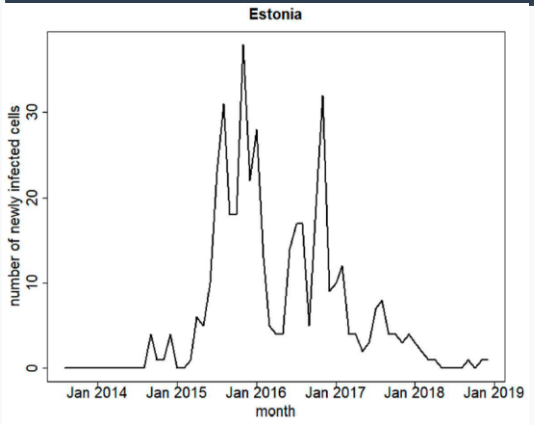
**Entry:** First case detected in SE Estonia (Nov 2014), introduced from Russia via Latvia.

**Scale:** 2,045 wild boar cases reported 2014–2019 across 575 × 10 km patches covering Estonia.

**Control:** Intensive hunting campaigns reduced wild boar density; decline apparent in later outbreak data.

**Data:** Monthly case reports with locations.

## Outbreak Incidence Curve (Fig. 6, bottom right panel)

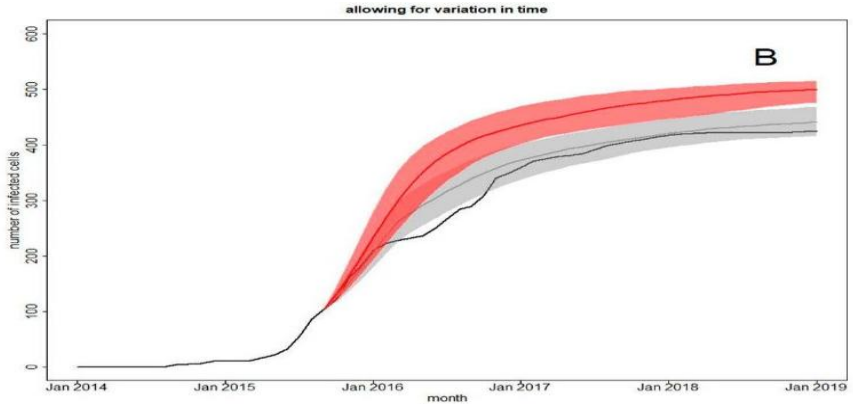



*Monthly case reports 2014–2019; y-axis: new cases per month; sharp decline after 2017 reflects control*

*Catterall, Porphyre & Marion (2025) PLoS Comput. Biol. 21(6): e1012622 | https://doi.org/10.1371/journal.pcbi.1012622*

# Results: Model Predictions

## Model predictions – non-spatial



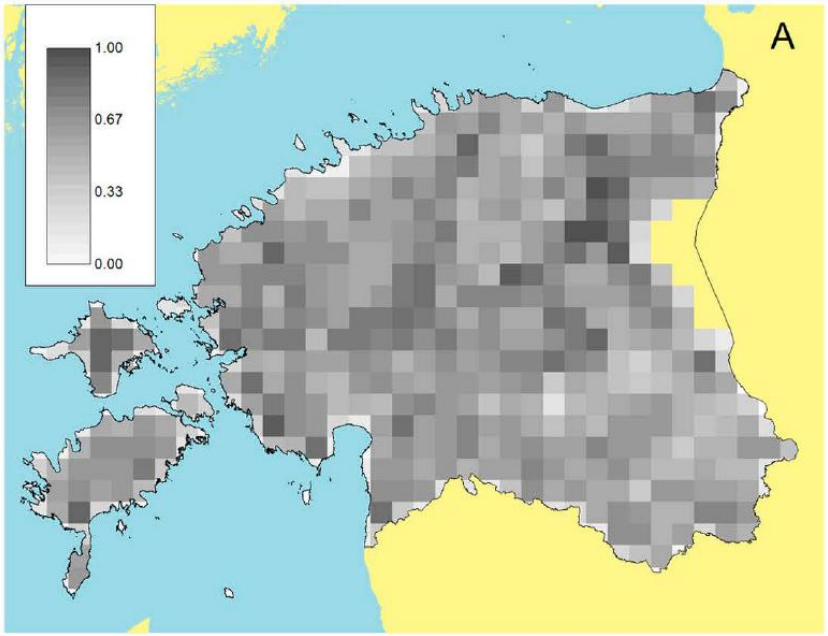

*Varying-in-time model fitted to full Estonian ASF outbreak – accounts for hunting campaign*


**Model fit:** Heterogeneous iPAR model (fitted to early data up to Sep 2015) projects cumulative patch incidence with 95% credible intervals. The varying-in-time model fitted to the full data captures the post-2017 decline in spread rate — evidence of successful population control.

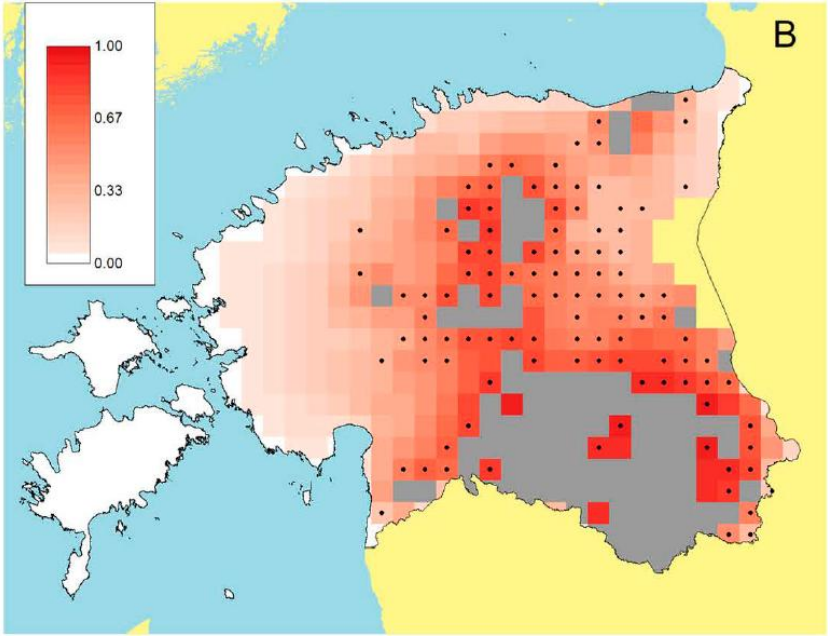
**Spread rate:** β(t) declines markedly from 2017 onwards — coinciding with intensified hunting campaigns. This provides statistical evidence that population reduction through culling reduced disease transmission rates.

# Results: Model Predictions & Susceptibility Map

## Estimated susceptibility map and model output



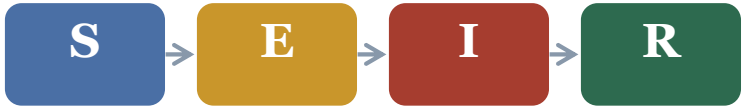

*Estimated susceptibility surface across Estonia; higher values = greater risk of infection incursion.*




*Predicted probability of spread (red) & observed spread (dots) from known infections (grey)*

*Catterall, Porphyre & Marion (2025) PLoS Comput. Biol. 21(6): e1012622 | https://doi.org/10.1371/journal.pcbi.1012622*

# The Spatial SEIR Transmission Model (semi-Markov)




**Probability of exposure during [t, t+dt):**

$$p(j,t) = \{\alpha + \beta \Sigma_{i \in I(t)} K(d_{ij}, \kappa)\}\, dt$$

**α** primary rate · **β** contact rate · $\mathbf{K(d_{ij}, \kappa)}$ spatial kernel

**Three competing spatial kernels K(d, κ):**

**A — Exponential**

$$K(d, \kappa_1) = \exp(-\kappa_1 d)$$

Locally bounded; probability of long-range transmission decays rapidly.

**B — Cauchy-type**

$$K(d, \kappa_2) = 1 / (1 + d/\kappa_2)$$

Heavy tail; long-range transmission remains substantial. Rejected for hogweed (§4).

**C — Power-law**

$$K(d, \kappa_3) = d^{-\kappa_3}$$

Scale-free; intermediate tail behaviour. Comparable fit to A for hogweed.

## Latent Period (time in class E)

$$x \sim \text{Gamma}(\mu, \sigma^2) \quad f_E(x; \mu, \sigma^2) = x^{\mu^2/\sigma^2 - 1} e^{-x\mu/\sigma^2} / [(\sigma^2/\mu)^{\mu^2/\sigma^2} \Gamma(\mu^2/\sigma^2)]$$

μ = mean · σ² = variance · parameterised for interpretability

## Infectious Period (time in class I)

$$x \sim \text{Weibull}(\gamma, \eta) \quad f_I(x; \gamma, \eta) = (\eta/\gamma) \cdot (x/\gamma)^{\eta - 1} \cdot e^{-(x/\gamma)^\eta}$$

γ = shape · η = scale · fitted in all model variants

## Case Study: Giant Hogweed SI Model with Site Suitability

$$p(j,t) = c_j \{\alpha + \beta \Sigma K(d_{ij}, \kappa)\}\, dt$$

$c_j \in [0,1]$ — habitat suitability score (temperature, altitude, other covariates; Catterall et al. 2012). E and I merged into one class (no recovery R). Six model variants: 3 kernels × ±suitability.

**Inference:** Data augmentation MCMC samples from the joint posterior $p(\theta, Z \mid y)$, where Z is the complete (partially unobserved) epidemic trajectory and y the interval-censored observations. Sojourn times are interval-censored; infection times are latent variables sampled alongside model parameters.

# Opportunities and barriers in application of stochastic differential equations to agriculture and ecology

*glenn.marion@bioss.ac.uk*

**Population modelling**
Demographic stochasticity and realistic noise terms

**Modelling heterogeneity**
The need to represent heterogeneity

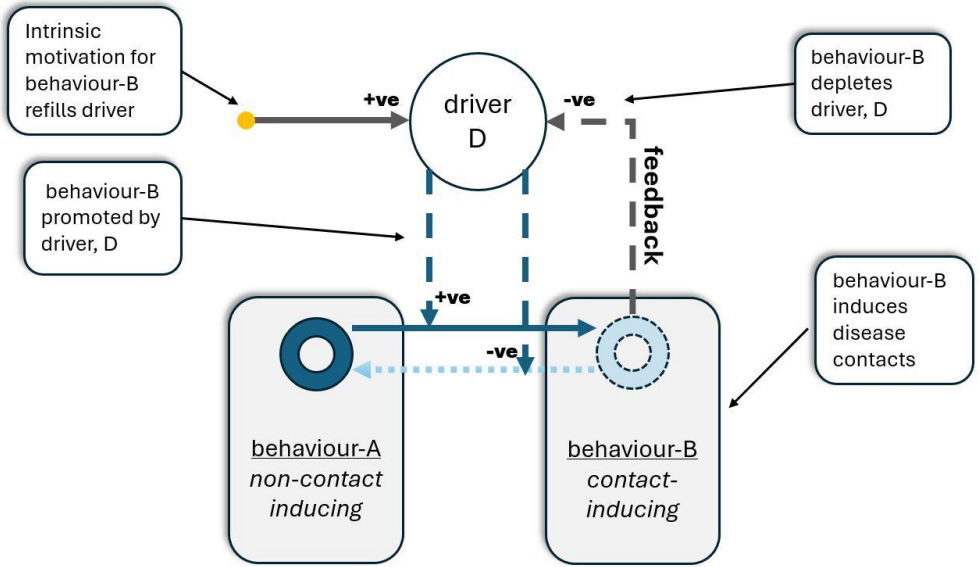
**Spatial models**
Spatial dynamics and spread

**Individual and agent-based models**
Individual behaviours shape system dynamics

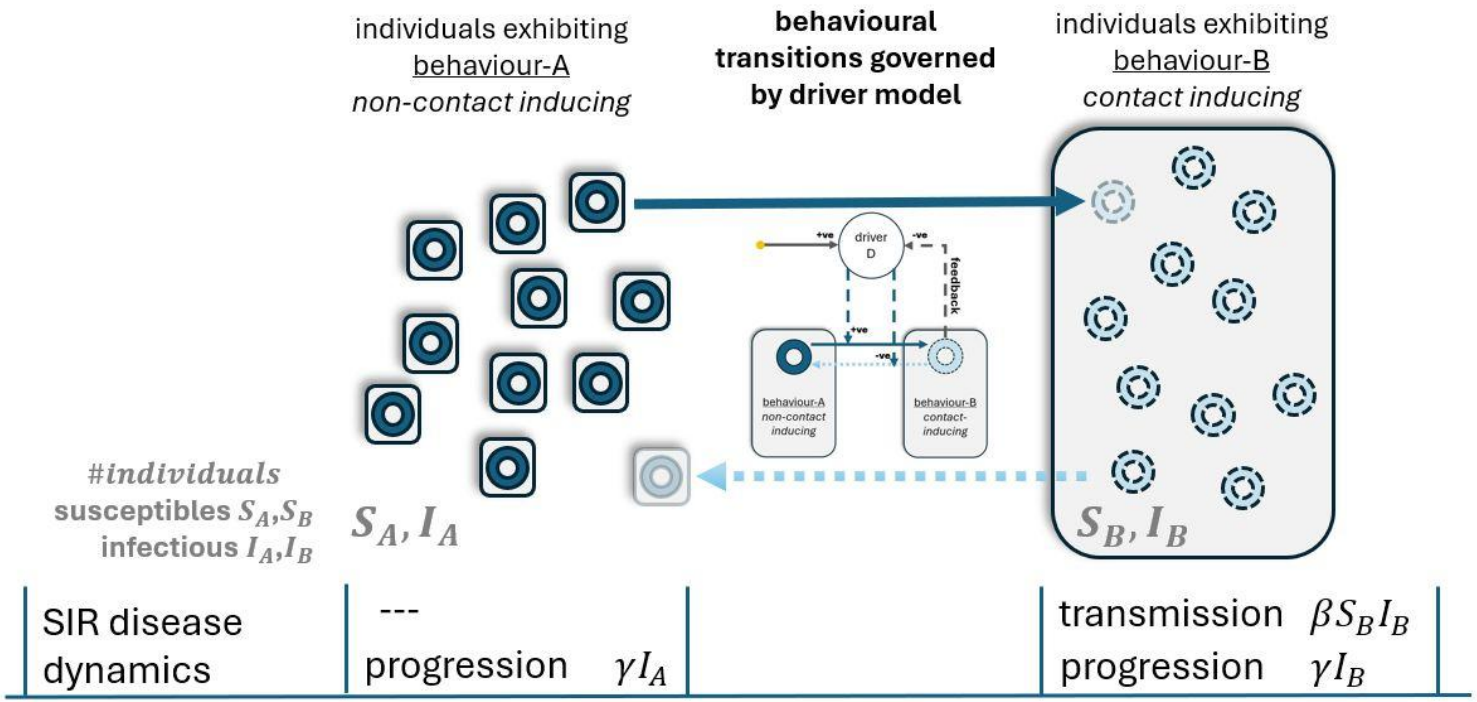
**Assessing model fit to data**
Challenges for modelling fitting
Inference for spatial models
Assessing model fit

# Individual behavioural drivers for disease contact processes

- Disease controls typically aim to reduce contact/contact behaviours
- **But** these behaviours done for a reason
- **So** expect pushback and compensatory behaviours in response
- Agent-based modelling of systems allows for exploration of these dynamics and the potential of different control measures
- Ideas have been applied to livestock disease spread via cattle trading
  - Showed impact of changing trading behaviour
  - Strong benefit of sharing information on disease risk and how this changes trading behaviours



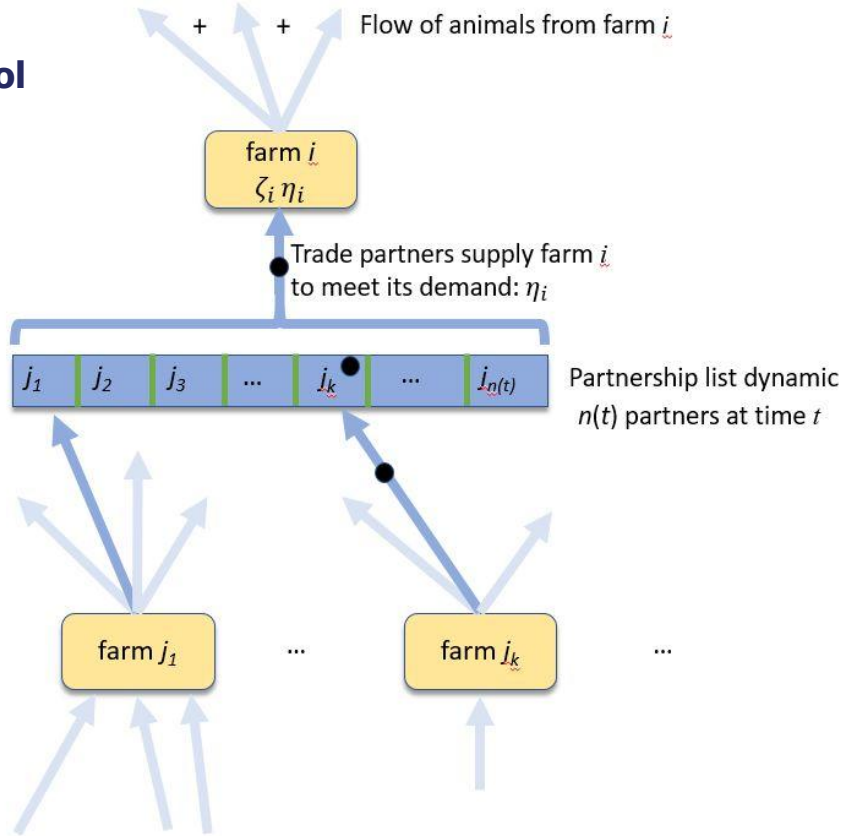

**individual behavioural states, transitions and driver dynamic**


individuals exhibiting
behaviour-A
non-contact inducing
behavioural
transitions governed
by driver model
individuals exhibiting
behaviour-B
contact inducing
driver
D
+ve
-ve
feedback
behaviour-A
non-contact
inducing
behaviour-B
contact-
inducing
#individuals
susceptibles $S_A$,$S_B$
infectious $I_A$,$I_B$
$S_A, I_A$
$S_B, I_B$
SIR disease
dynamics
---
progression $\gamma I_A$
transmission $\beta S_B I_B$
progression $\gamma I_B$

## Initial attempt at systems approach to disease control

Generative model of cattle trade between farms

- given farm generates supply and/or demand
- those with demand satisfy this by trade with farms with available supply
- rates determined from CTS data – currently for Scotland – *digital twin*
- get power-law like distributions – a few large players with many smaller

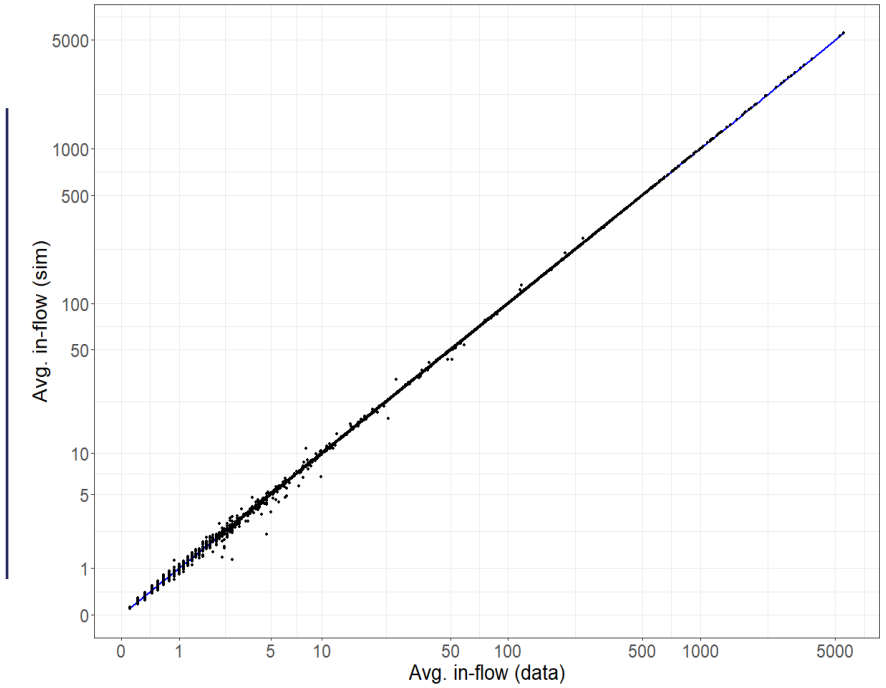



Knight, M.A., White, P.C.L., Hutchings, M.R., Davidson, R.S. and Marion, G.(2021). Generative models of network dynamics provide insight into the effects of trade on endemic livestock disease *Royal Society Open Science* **8**, 201715

## Initial attempt at systems approach to disease control

Generative model of cattle trade between farms

- given farm generates supply and/or demand
- those with demand satisfy this by trade with farms with available supply
- rates determined from CTS data – currently for Scotland – *digital twin*
- get power-law like distributions – a few large players with many smaller
- **approach captures key features**

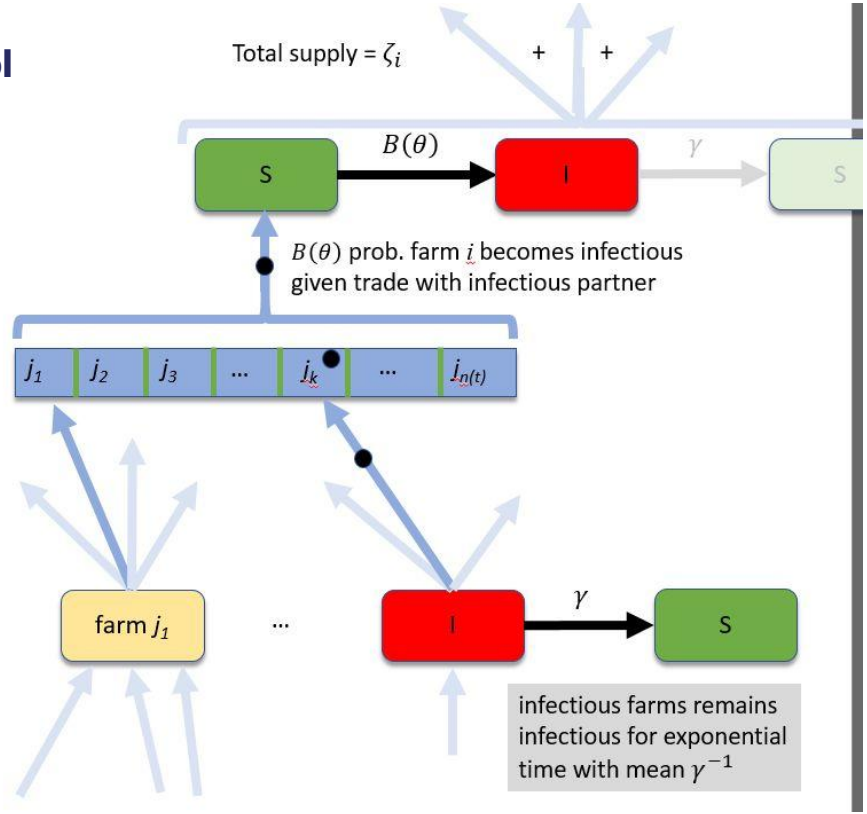



Knight, M.A., Marion, G., White, P.C.L., Hutchings, M.R. and Davidson, R.S.(2022). A mechanistic model captures livestock trading, disease dynamics, and compensatory behaviour in response to control measures J*ournal of Theoretical Biology* **539**, 111059

## Initial attempt at systems approach to disease control

Now just add disease dynamics!

- SIS dynamic – recovery rate $\gamma$
- Trade with infected farms can lead to transmission – dependent on batch size B
- Possible controls
  - Biosecurity – increase recovery rate
  - Test (and cull) on movement
  - Change trading behaviour
  - Combinations on the above




Knight, M.A., Marion, G., White, P.C.L., Hutchings, M.R. and Davidson, R.S.(2022). A mechanistic model captures livestock trading, disease dynamics, and compensatory behaviour in response to control measures J*ournal of Theoretical Biology* **539**, 111059

# Opportunities and barriers in application of stochastic differential equations to agriculture and ecology

*glenn.marion@bioss.ac.uk*

## Population modelling
Demographic stochasticity and realistic noise terms

## Modelling heterogeneity
The need to represent heterogeneity

## Spatial models
Spatial dynamics and spread

## Individual and agent-based models
Individual behaviours shape system dynamics

## Assessing model fit to data
Challenges for modelling fitting
Inference for spatial models
Assessing model fit

# Latent Residuals for Fitted Model Diagnostics

*The core idea: embed classical residuals within Bayesian inference*

*Lau, Marion, Streftaris & Gibson (2014)*
*J. R. Soc. Interface 11: 20131093*

## The Assessment Gap

Bayes factors are sensitive to priors and computationally prohibitive. The DIC is not uniquely defined for partially observed processes. Posterior predictive checks can be insensitive to model choice — even when kernels are very different (§4.2).

## The Functional Model (Cox & Snell 1968)

$$Z = h_\theta(\tilde{r}) \quad \text{where } \tilde{r} \sim U(0,1)$$

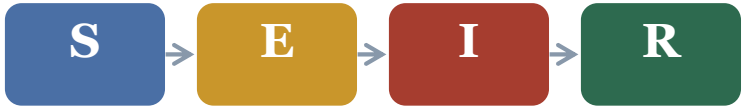
$Z$ — complete epidemic data $\theta$ — model parameters

Any stochastic model realisation can be written as a deterministic function of uniform random variables. The residuals $\tilde{r}$ have known sampling properties — which can be tested.

## Link to Sellke Thresholds

Sellke (1983) showed that each exposure event can be triggered by a threshold drawn from U(0,1) applied to the integrated infectious challenge A(t). This is precisely the ETR construction used here (eq. 2.5). The earlier use of Sellke thresholds as Bayesian residuals (Gibson et al. 2006) required imputation even for uninfected individuals — a problem for small epidemics. The present construction avoids this: residuals are imputed only at each infection event.

# The Spatial SEIR Transmission Model




**Probability of exposure during [t, t+dt):**

$$p(j,t) = \{\alpha + \beta \Sigma_{i \in I(t)} K(d_{ij}, \kappa)\}\, dt$$

$\alpha$ primary rate · $\beta$ contact rate · $K(d_{ij}, \kappa)$ spatial kernel

**Three competing spatial kernels K(d, κ):**

**A — Exponential**

$$K(d, \kappa_1) = \exp(-\kappa_1 d)$$

Locally bounded; probability of long-range transmission decays rapidly.

**B — Cauchy-type**

$$K(d, \kappa_2) = 1 / (1 + d/\kappa_2)$$

Heavy tail; long-range transmission remains substantial. Rejected for hogweed (§4).

**C — Power-law**

$$K(d, \kappa_3) = d^{-\kappa_3}$$

Scale-free; intermediate tail behaviour. Comparable fit to A for hogweed.

**Latent Period (time in class E)**

$$x \sim \text{Gamma}(\mu, \sigma^2) \quad f_E(x; \mu, \sigma^2) = x^{\mu^2/\sigma^2 - 1} e^{-x\mu/\sigma^2} / [(\sigma^2/\mu)^{\mu^2/\sigma^2} \Gamma(\mu^2/\sigma^2)]$$

$\mu$ = mean · $\sigma^2$ = variance · parameterised for interpretability

**Infectious Period (time in class I)**

$$x \sim \text{Weibull}(\gamma, \eta) \quad f_I(x; \gamma, \eta) = (\eta/\gamma) \cdot (x/\gamma)^{\eta - 1} \cdot e^{-(x/\gamma)\eta}$$

$\gamma$ = shape · $\eta$ = scale · fitted in all model variants

**Case Study: Giant Hogweed SI Model with Site Suitability**

$$p(j,t) = c_j \{\alpha + \beta \Sigma K(d_{ij}, \kappa)\}\, dt$$

$c_j \in [0,1]$ — habitat suitability score (temperature, altitude, other covariates; Catterall et al. 2012). E and I merged into one class (no recovery R). Six model variants: 3 kernels × ±suitability.

**Inference:** Data augmentation MCMC samples from the joint posterior $p(\theta, Z \mid y)$, where Z is the complete (partially unobserved) epidemic trajectory and y the interval-censored observations. Sojourn times are interval-censored; infection times are latent variables sampled alongside model parameters.

# Three Latent Residual Types — Construction & Diagnostic Power

## $\tilde{r}_1$ Exposure Time Residuals (ETR)

**Construction:**

For the kth exposure, $r_{1k}$ is set against the integrated infectious challenge A(t):

$$1 - e^{-A(t)} > r_{1k} \rightarrow \text{exposure at } t_k$$

**Targets:**

Overall timing of exposures

**Diagnostic intuition:**

If the force of infection is correctly specified, intervals between exposures integrate to Exp(1) — the threshold $r_{1k}$ captures this. Based directly on Sellke (1983).

## $\tilde{r}_2$ Infection-Link Residuals (ILR)

**Construction:**

Links are ranked by kernel weight $p(ij) \propto K(d_{ij})$. The rank of the observed link among all S×I links gives $r_{2k}$:

$$r_{2k} \sim U(\Sigma\, p(\alpha), \Sigma\, p(\alpha) + p(s'))$$

**Targets:**

The spatial kernel $K(d_{ij}, \kappa)$

**Diagnostic intuition:**

Ordering links by kernel weight means a too-rapidly-decaying kernel concentrates residuals at extremes; a too-slowly-decaying kernel shows a scarcity at extremes — opposite, diagnosable patterns (Fig. 2).

★ Key diagnostic for spatial kernel choice

## $\tilde{r}_3$ Latent Time Residuals (LTR)

**Construction:**

For exposed individual k, $r_{3k}$ is set against the accumulated pressure of becoming infectious Q(t):

$$1 - e^{-Q(t)} > r_{3k} \rightarrow \text{infectious at } t'_k$$

**Targets:**

Latent period distribution f(·)

**Diagnostic intuition:**

If the latent-period distribution is mis-specified, the LTRs depart from U(0,1). Correctly specified, they are exactly uniform — a cleanly targeted diagnostic.

# The General Principle: Every Stochastic Realisation Built from r.v. with known distribution(s)

Any stochastic model realisation $\mathbf{Z} = \mathbf{h}_\theta(\tilde{\mathbf{r}})$ where $\tilde{\mathbf{r}} \sim U(0,1)$ **but other options may be preferred**

Under the correct model, the imputed residuals $\tilde{\mathbf{r}}$ are exactly uniform.
Mis-specification → systematic departure from U(0,1) → detectable by classical tests.

## ① Any Stochastic Process or Model

- Epidemic model (SEIR, SI, SIR)
- Ecological invasion model
- Animal movement model
- Plant dispersal / colonisation
- Any Markov / point process
- Branching processes, renewal models

*Discrete or continuous states and or time*

## ② Design Targeted Residuals r̃

- Partition r̃ so each component corresponds to one model aspect
- Exposure times → ETR (Sellke)
- Spatial links → ILR (kernel)
- Sojourn times → LTR (periods)
- Suitability / covariates → $\tilde{r}_4$, ...
- Any unknown latent process

*Design drives sensitivity: partition matches structure.*

## ③ Test Against U(0,1) Classically

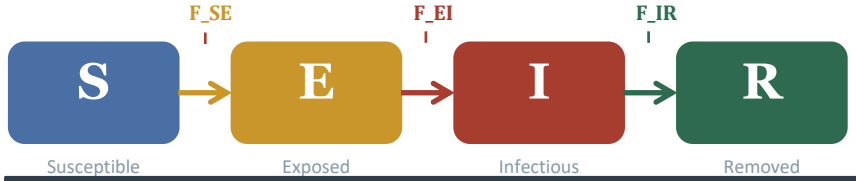
- Impute r̃ from MCMC posterior
- Apply Anderson–Darling test
- Get posterior distribution of p-values
- p(P(r̃) < 0.05 | y) → evidence measure
- Histogram shape reveals mis-spec type
- Extreme concentration or scarcity

*Addendum to existing MCMC — no new algorithm needed.*

**Key insight:** the principle is universal. Any stochastic model realisation is built on a set of underlying uniform random numbers r̃. Under the correct model these are exactly U(0,1). Mis-specification is visible as a systematic, interpretable departure from uniformity.

# The SDE-SEIR (non-spatial/population level) Model

*Demographic stochasticity via Itô SDEs — three independent Wiener processes drive three reaction fluxes*




**Parameters** $\theta = (\beta, \sigma, \gamma, N, \rho, \phi, S_0, E_0, I_0)$

| | |
|---|---|
| $\beta$ — transmission rate (day$^{-1}$) | $N$ — total (closed) population |
| $\sigma$ — E→I transition rate (day$^{-1}$) | $\rho$ — reporting probability $\in (0,1)$ |

**Three reaction fluxes (Chemical Langevin Equation)**

$$F_{SE} = \beta \cdot S \cdot I/N \cdot dt + \sqrt{(\beta \cdot S \cdot I/N)} \cdot dW_{SE}$$

S → E: density-dependent infection; variance ∝ flux magnitude

$$F_{EI} = \sigma \cdot E \cdot dt + \sqrt{(\sigma \cdot E)} \cdot dW_{EI}$$

E → I: linear transition; noise collapses to zero when E = 0

$$F_{IR} = \gamma \cdot I \cdot dt + \sqrt{(\gamma \cdot I)} \cdot dW_{IR}$$

I → R: recovery; mirrors E → I structure

**State equations (Itô form, closed population N = S+E+I+R)**

$$dS = -F_{SE}$$

$$dE = +F_{SE} - F_{EI}$$

$$dI = +F_{EI} - F_{IR}$$

$$dR = +F_{IR}$$

**Euler-Maruyama discretisation:**

Brownian increments $dW \sim N(0, \sqrt{\Delta t})$ drawn independently. Positivity enforced by reflection (clamp to zero); Conservation S+E+I+R = N maintained exactly by setting R = N−S−E−I.

# Data and model fitting – fitted model matches data generation

*Synthetic SDE trajectories → reported cases via Negative-Binomial*

## Data generation & observation model

### ① Observed only Recovered cases periodically

The CLE adds Wiener-process fluctuations to each reaction flux — demographic stochasticity. Even with identical parameters, every simulated epidemic trajectory is different. Variance ∝ propensity (square-root diffusion).

### ② Observation noise

Only a fraction ρ of true recoveries ΔR are reported, and reporting itself is overdispersed (e.g. day-of-week effects, test backlogs). Modelled as NegBin(ρ·ΔR, ϕ).

**Observation Model: Negative-Binomial reporting**

$$y_k \sim \text{NegBin}(\mu_k = \rho \cdot \Delta R_k,\ \varphi)$$

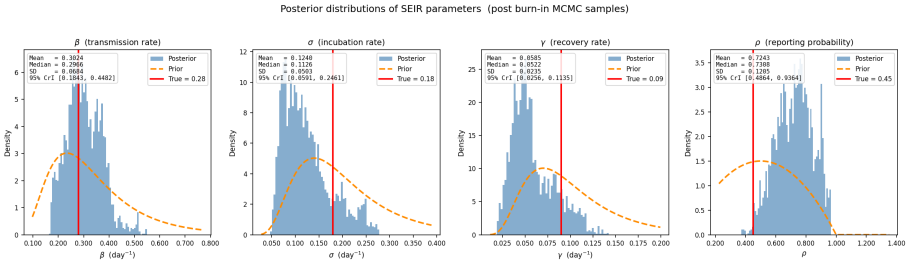
**ρ** — reporting probability ·
**φ** — overdispersion (ϕ→∞ gives Poisson)

## Fit to data via particle MCMC

① Estimate likelihood with bootstrap particle filter

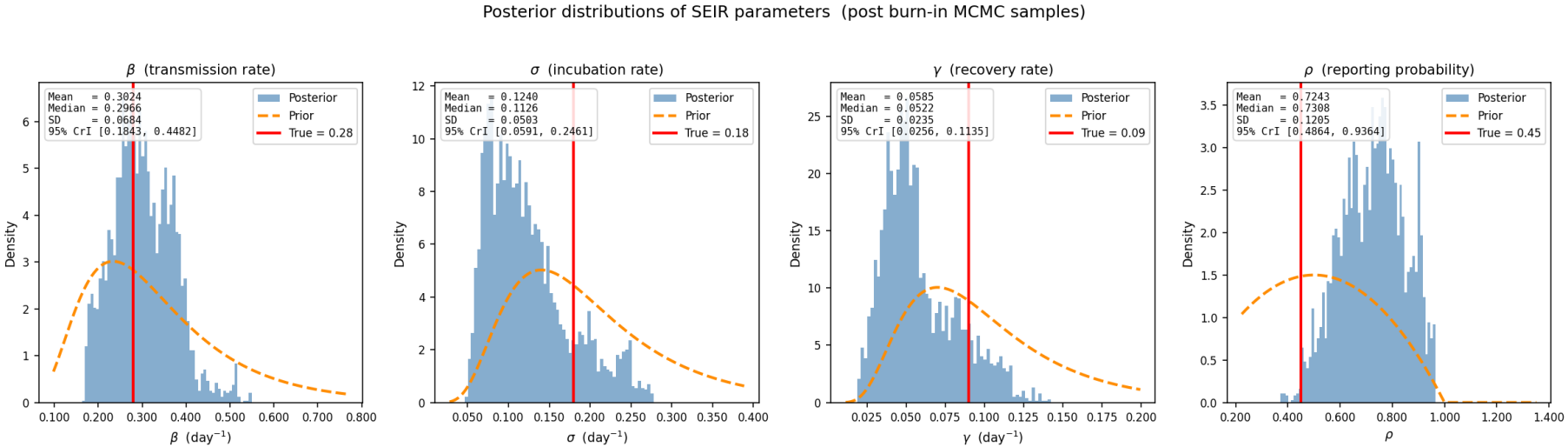
① Update parameters in via standard M-H MCMC

## Can estimate parameters



# Data and model fitting – fitted model matches data generation

*Synthetic SDE trajectories → reported cases via Negative-Binomial*

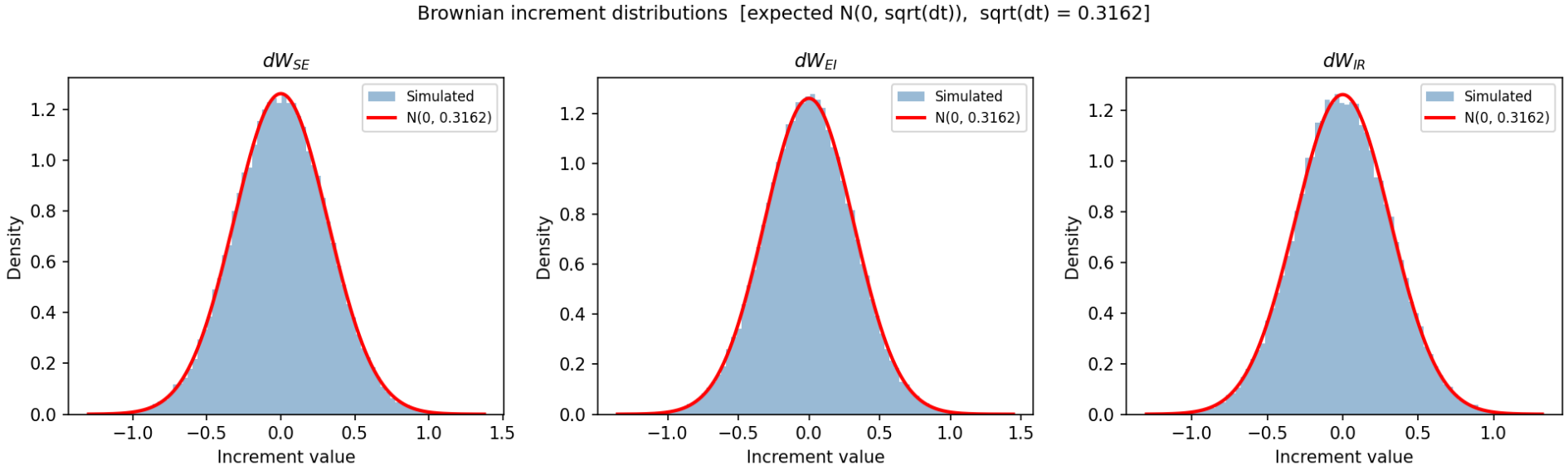


# Data and model fitting – fitted model matches data generation

*Synthetic SDE trajectories → reported cases via Negative-Binomial*

**Brownian increments are Latent residuals in this context**

## Once model is fitted to data do increments vary from expectation?

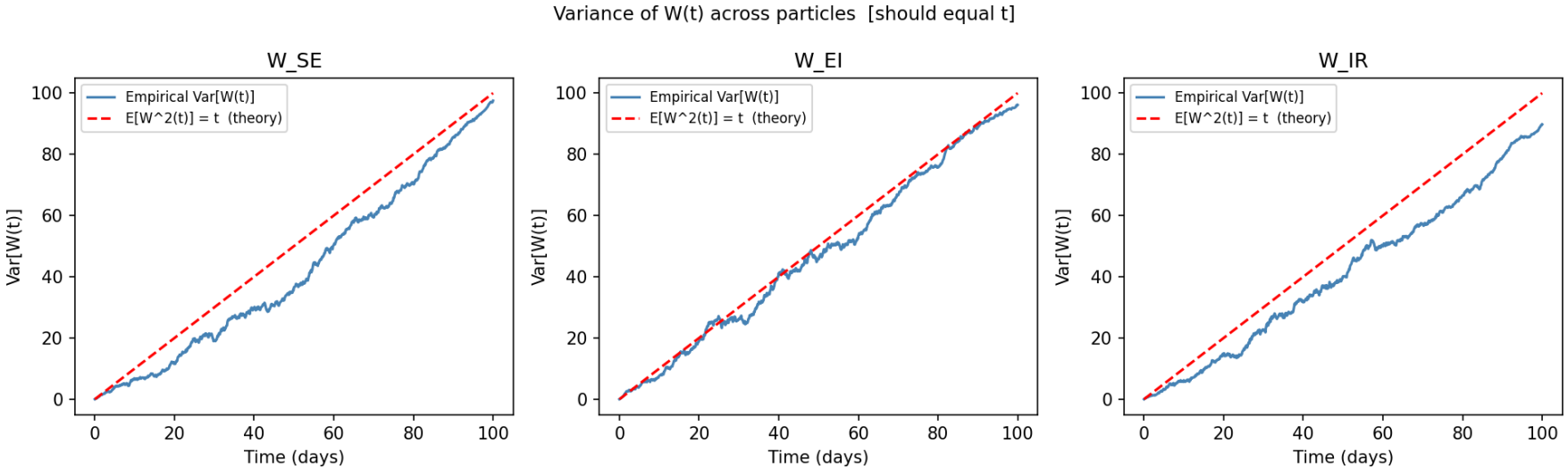


# Data and model fitting – fitted model matches data generation

*Synthetic SDE trajectories → reported cases via Negative-Binomial*

**Brownian increments are Latent residuals in this context**

## Once model is fitted to data do increments vary from expectation?



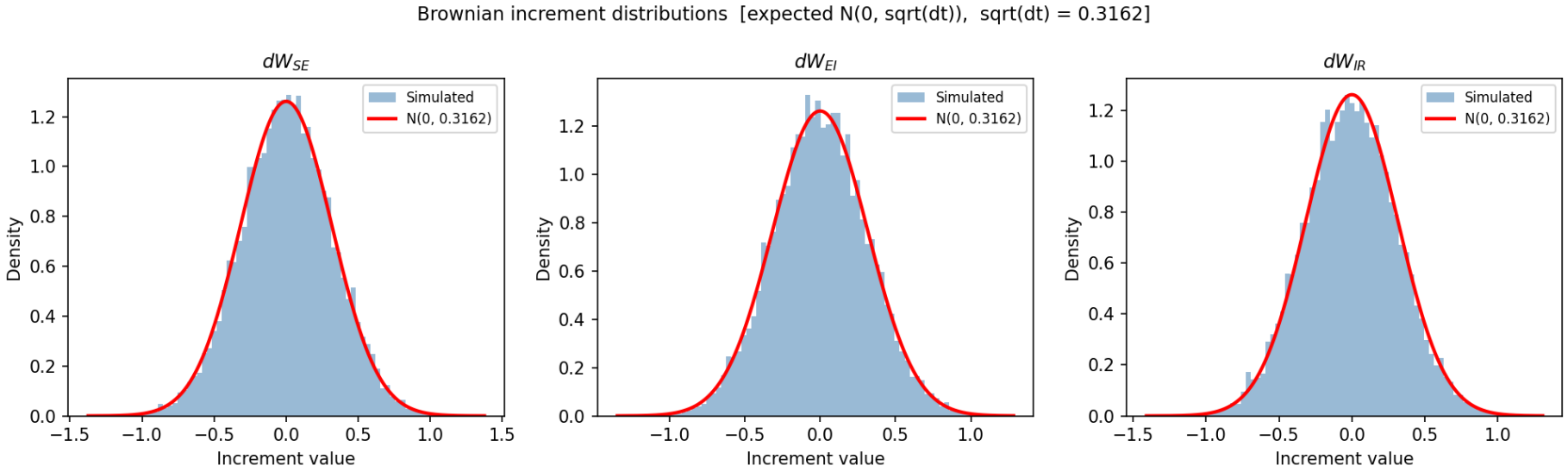
# Model mismatch – SEIR fitted to data generated by SIR model

*Synthetic SDE trajectories → reported cases via Negative-Binomial*

**Brownian increments are Latent residuals in this context**

Results (very) preliminary

## Once model is fitted to data do increments vary from expectation?



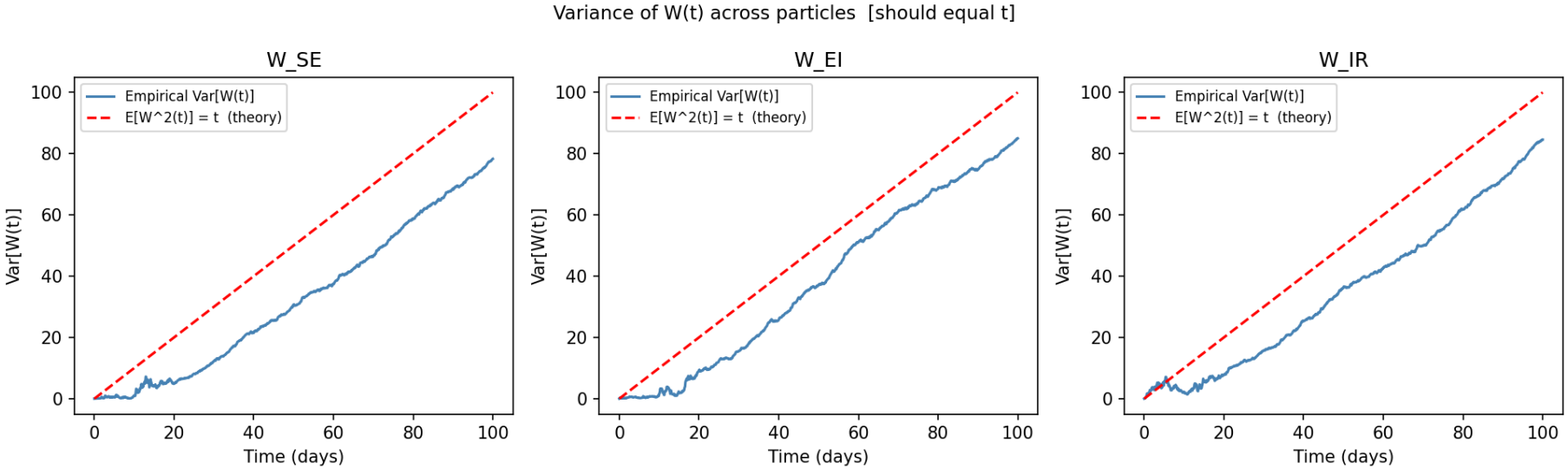
# Model mismatch – SEIR fitted to data generated by SIR model

*Synthetic SDE trajectories → reported cases via Negative-Binomial*

**Brownian increments are Latent residuals in this context**

Results (very) preliminary

Once model is fitted to data do increments vary from expectation?




Suggests mismatch early on in epidemic … perhaps indicative of lack of E state?

# Opportunities and barriers in application of stochastic differential equations to agriculture and ecology

*glenn.marion@bioss.ac.uk*

## Population modelling

Demographic stochasticity and realistic noise terms

What tools and results are available for complex noise terms

## Modelling heterogeneity

What tools & results carry over to heterogeneous systems

## Spatial models

Spatial dynamics and spread – how can we use SDEs/PDEs

## Individual and agent-based models

What is the role of SDEs?

## Assessing model fit to data

Methods and tools for SDEs/SPDEs

Dealing with spatial and other heterogeneity

Assessing model fit