Adaptive Frameworks for Epidemic Dynamics: From Bayesian Retrospective Assessment to Reinforcement Learning and Forecasting

Petros Barmpounakis

Department of Oncology, University of Cambridge,
Cambridge, United Kingdom

31 March 2026, AUEB, Department of Statistics seminars

Acknowledgements

This presentation covers work spanning several projects. I would like to sincerely thank my co-authors, with a special acknowledgement to **Nikos Demiris** (Athens University of Economics and Business) who collaborated on all three of these works.

Reinforcement Learning for Epidemic Control:

- Giacomo Iannucci (University College London)
- Alex Beskos (University College London)

Exchangeable Gaussian Processes:

- Kostas Kalogeropoulos (London School of Economics)
- Lampros Bouranis (formerly at Athens University of Economics and Business)

Overview

The (not so) recent global COVID-19 pandemic highlighted the necessity for flexible, data-driven frameworks to guide policy decisions under uncertainty.

This presentation explores three methodological works:

- **Epidemic Modelling Foundations:** Stochastic SEIR models, Poisson contact processes, and multi-type renewal equations.
- **Retrospective Analysis:** Multiphasic models with piece-wise constant reproduction numbers to capture discrete epidemic phases.
- **Dynamic Policy Optimisation:** Combining sequential Bayesian inference with Reinforcement Learning to balance ICU load vs. socio-economic costs.
- **Continuous Forecasting:** Multi-type Exchangeable Gaussian Processes for temporal shrinkage across groups.

Contents

The Stochastic SEIR Compartmental Model

- We model the epidemic as a stochastic process tracking individuals across mutually exclusive disease states ($S \to E \to I \to R$).
- Contacts between individuals occur at the points of a **Poisson process** with time-varying intensity $\frac{\lambda_t}{N}$.
- The number of new exposures (incident cases) c_{t+1} at day $t+1$ follows a Poisson distribution:

$$c_{t+1} \sim \text{Poisson}\left(S_t \frac{\lambda_t}{N} I_t\right)$$

- Expected new infections:

$$E[c_{t+1}] = S_t \frac{\lambda_t}{N} I_t \tag{1}$$

Connection: Compartmental to Renewal Models (1/2)

- Let X_j and Y_j denote the exposed and the infectious periods of the j-th individual infected at time s, respectively.
- I_t denotes the active set of infectious individuals at time t:

$$I_t = \sum_{s=0}^{t} \sum_{j=1}^{c_s} \mathcal{I}_{\{Y_j + X_j > t-s > X_j\}} \tag{2}$$

- Substituting (2) into our expected incidence equation (1):

$$\begin{aligned} E[c_{t+1}] &= S_t \frac{\lambda_t}{N} \sum_{s=0}^{t} \sum_{j=1}^{c_s} \mathcal{I}_{\{Y_j + X_j > t-s > X_j\}} \\ &\approx \frac{S_t}{N} \lambda_t \sum_{s=0}^{t} c_s E[\mathcal{I}_{\{Y_j + X_j > t-s > X_j\}}] \end{aligned} \tag{3}$$

Connection: Compartmental to Renewal Models (2/2)

- The expected number of individuals still infectious is:

$$E[\mathcal{I}_{\{Y_j+X_j>t-s>X_j\}}] = P(Y_j + X_j > t - s > X_j)$$

- We define the serial interval distribution $g_s(t)$ using the mean infectious period $E[Y]$:

$$g_s(t) = \frac{P(Y_j + X_j > t - s > X_j)}{E[Y]}$$

- Substituting this back, and defining the reproduction number $R_t = \lambda_t E[Y]$:

$$E[c_{t+1}] \approx \frac{S_t}{N} R_t \sum_{s=0}^{t} c_s g_s(t) \tag{4}$$

- This mathematically bridges the compartmental stochastic SEIR directly to the aggregate renewal formulation.

Expansion to Multi-Type Epidemic Dynamics

- Real-world epidemics have heterogeneous dynamics across subpopulations (e.g., age groups, geographic regions, vaccination status).
- We partition the population into K types.
- Let $\mathcal{C}_{ij}$ be the elements of a contact matrix $\mathcal{C}$ representing the interaction intensity between type i and type j.
- Expanding the renewal formulation (4), the expected incidence for type i is:

$$E[c_{t+1}^{(i)}] \approx \frac{S_t^{(i)}}{N^{(i)}} \beta_t^{(i)} \sum_{j=1}^{K} \mathcal{C}_{ij} E[Y^{(j)}] \sum_{s=0}^{t} c_s^{(j)} g_s^{(j)}(t) \tag{5}$$

- Instead of a single pooled reproduction number, new infections in group i are driven by its own parameter $\beta_t^{(i)}$, which can be seen as the probability of infection given contact in an infinite population, and the infectious pressure from all groups j scaled by the contact matrix $\mathcal{C}$.

Observation Model: Connecting Latent Infections to Data

- True daily infections c_t are practically unobserved. To calibrate our frameworks, we link the latent epidemic states to measurable downstream indicators (e.g., deaths or ICU admissions).
- Expected events μ_t are modeled as a convolution of past incidence with a delay distribution $f(\tau)$, scaled by an outcome probability p (e.g., Infection Fatality Rate or ICU-admission rate):

$$\mu_t = p \sum_{\tau=0}^{t} c_{t-\tau} f(\tau)$$

- To account for severe overdispersion in real-world public health data (e.g., weekend lags, administrative undercounting, and batch reporting), we use a **Negative-Binomial** likelihood:

$$Y_t \sim \text{NegBin}(\mu_t, \phi)$$

Multiphasic Epidemic Models

- The implementation of NPIs to combat the rapid expansion of COVID-19 resulted in multiple distinct epidemic phases.
- Classical SEIR models assume a single, fixed transmission rate. This is not suitable for real outbreaks.
- We develop hierarchical models with piece-wise constant R_t that capture these shifting phases.
- The model complexity is directly determined by the data without any prior assumptions regarding the beginning or end of an epidemic phase.
- The number of phases can be either:
  - **Deterministic**: selected via Information Criteria (WAIC, LOO-CV)
  - **Stochastic**: inferred from data using Poisson Point Process or Dirichlet Process priors

Deterministic Number of Phases

$$R_t = \begin{cases} & r_1, \quad t \leq T_1 \\ \ldots & \\ & r_{j+1}, \quad T_j < t \leq T_{j+1} \\ \ldots & \\ & r_K, \quad T_{K-1} < t \leq T \end{cases}$$

$$\begin{aligned} r_j &\sim f(\cdot), \quad r_j \in (0, \infty), \quad j = 1, ..., K \\ T_{i+1} &= T_i + e_i \\ T_1 &\sim \text{Uniform}\,(3, T) \\ e_i &\sim \text{Uniform}\,(0, 100)\,, \quad i = 1, ..., K-1 \end{aligned} \tag{6}$$

The number of phases K is selected using WAIC and approximate leave-one-out CV.

Stochastic Number of Phases: Non-Parametric Priors

Poisson Point Process

Uses the stick-breaking representation of a Poisson process for phase arrival.

$$
\begin{aligned}
R_t &= r_{z_t} \\
r_j &\sim f(\cdot), \quad j = 1, \ldots, K \\
z_t &\sim \text{Categorical}(\pi_{1:K}) \\
\pi_K &= 1 - \sum_{k=1}^{K} \pi_k \\
K &= \min \left\{ j : \sum_{i=1}^{j} T_i \geq T \right\} \\
\pi_k &= \frac{T_k}{T}, \quad k = 1, \ldots, K-1 \\
T_i &\sim \text{Exponential}(\lambda_{\text{phase}}) \\
\lambda_{\text{phase}} &\sim \text{Gamma}(0.02, 1)
\end{aligned}
$$

Dirichlet Process

Alternatively, uses the stick-breaking representation of a Dirichlet process.

$$
\begin{aligned}
R_t &= r_{z_t} \\
r_j &\sim f(\cdot), \quad j = 1, \ldots, L \\
z_t &\sim \text{Categorical}(w_{1:L}) \\
w_L &= \prod_{k<L} (1 - v_k) \\
K &= \sum_{k=1}^{L} \mathbb{I}\{w_k \geq 0\} \\
w_l &= v_l \prod_{j=1}^{l-1} (1 - v_j) \\
w_1 &= v_1, \quad v_i \sim \text{Beta}(1, \theta), \quad \theta \sim \text{Gamma}(1, 1)
\end{aligned}
$$

Real world applications-Results for Deterministic number of phases

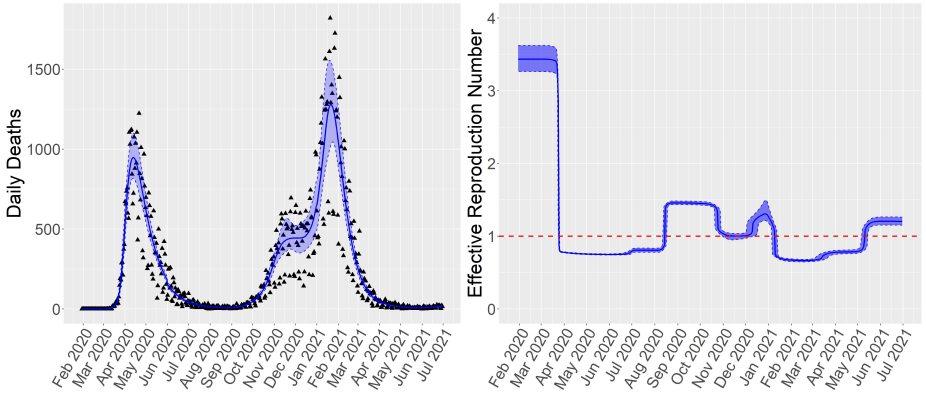
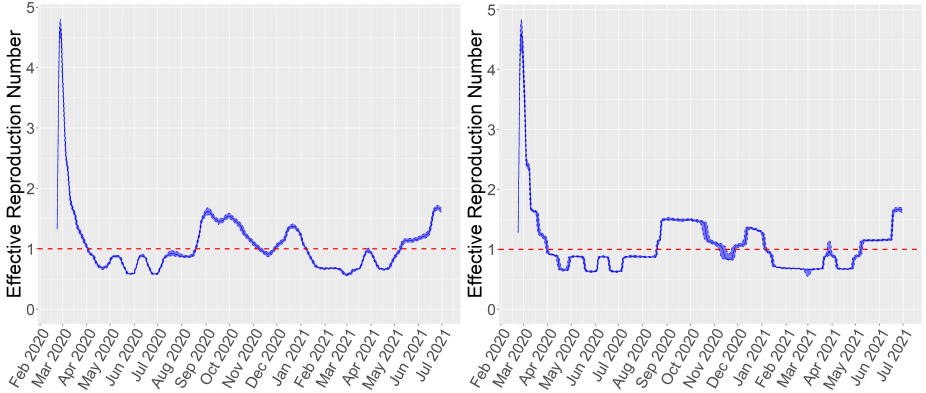


Figure: The United Kingdom, results from observing deaths.

Real world applications-Results for Stochastic number of phases



(a) The United Kingdom-Poisson Process (b) The United Kingdom-Dirichlet Process

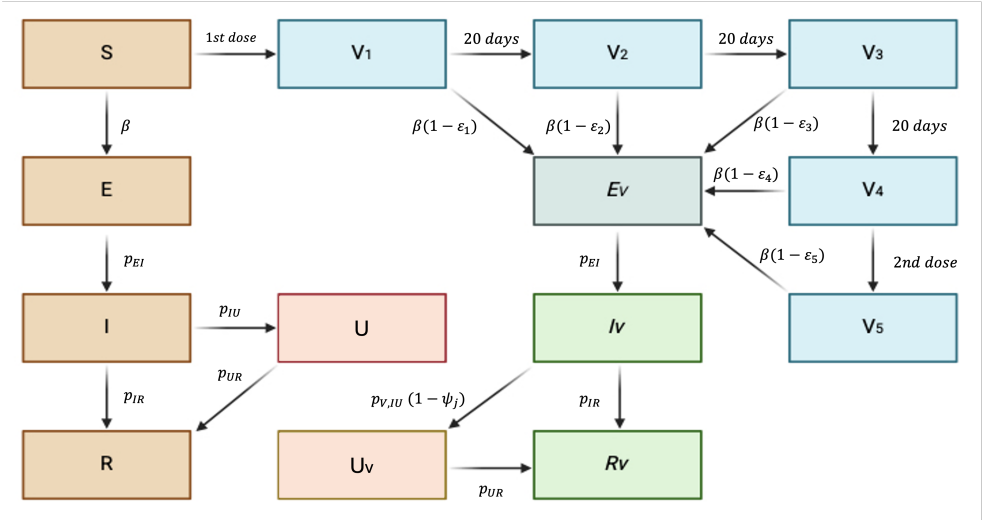
Key takeways

- All the proposed models seem to agree on the inferred reproduction number.
- The reproduction number is directly inferred from the data, without any assumptions for the duration or the level of the effect of the imposed NPIs.
- Thus, retrospectively, we can assess the NPIs effectiveness directly from data without introducing bias of when or what the effect of each intervention is.
- In applications to real-world data, school closures and restrictions of movement were consistently followed by an identifiable decrease in the effective reproduction number. When these measures were relaxed, the reproduction number rose again above the critical value of 1.

Extending the Framework

- Multiphasic models allow us to **retrospectively** assess the magnitude of transmission changes and NPI effectiveness.
- **Question:** Can we use these Bayesian models **prospectively** to actively design optimal intervention policies in real-time?
- **Solution:** We incorporate the compartmental models into a **Reinforcement Learning (RL)** framework.
- We transition from simply estimating R_t to adaptively controlling it via dynamic policy.

The Extended Epidemic Environment



The State-Space Model for Inference

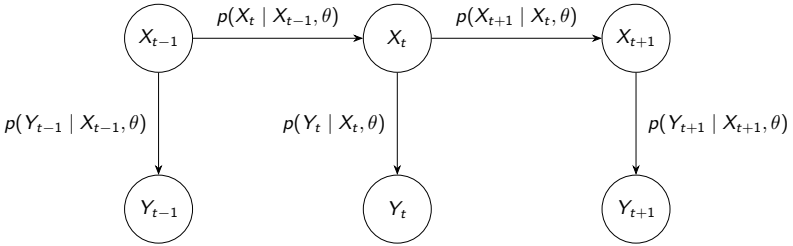


Figure: State-space model over three time steps. Latent compartmental states X evolve according to Markov dynamics, each giving rise to a noisy real-world observation Y (e.g., ICU reports).

Sequential Monte Carlo (SMC^2): Algorithm Architecture

- SMC^2 continuously processes data as it arrives, maintaining a joint posterior $p(\theta, X_{1:t} \mid Y_{1:t})$ using a mathematically rigorous nested structure:
- **Outer Filter (Parameters θ):**
  - Maintains M parameter particles $\theta^{(m)}$ (e.g., transmission rate, vaccine efficacy).
  - Outer weights are updated using the incremental marginal likelihood $p(Y_t \mid Y_{1:t-1}, \theta^{(m)})$ provided by the inner filter.
- **Inner Filter (Latent States X_t):**
  - For *each* parameter particle $\theta^{(m)}$, the algorithm runs a standard particle filter maintaining N state trajectories.
  - It pushes hidden compartments forward via stochastic transitions.
  - It evaluates proposed states against the Negative-Binomial observation model to generate the marginal likelihood score.

Sequential Monte Carlo (SMC^2): PMCMC Rejuvenation

- Because the parameters θ are static, continuously resampling the outer particles inevitably leads to **particle degeneracy** (all particles collapsing to identical values).
- **The "Rejuvenation" Step:** When the Effective Sample Size (ESS) of outer particles drops below a threshold, the algorithm triggers a Particle Marginal Metropolis-Hastings (PMCMC) rejuvenation:
  1. **Propose:** Generate new parameters θ^* using an adaptive Gaussian random walk scaled by the empirical covariance of surviving particles.
  2. **Evaluate:** Run the inner state filter entirely from scratch (Time 0 to t) to compute the new marginal likelihood for θ^*.
  3. **Accept/Reject:** Apply the Metropolis-Hastings ratio to accept or discard the new parameters.
- **Result:** Ensures the algorithm continually explores the parameter space, passing accurate Bayesian uncertainty to the downstream RL controller.

The RL Framework (Markov Decision Process)

- At each time step (e.g., weekly), the RL controller observes the state and takes an action.
- **State (S_t):** The current posterior prediction of the epidemic (primarily ICU load, discretised into bins).
- **Action (A_t):** The chosen intervention level (e.g., 1 = Open, 2 = Mild closures, 3 = Strict closures, 4 = Full Lockdown).
- **Reward/Cost (C_t):** A scalar function penalising two competing outcomes.

$$C(S_t, A_t) = \text{Penalty}_{health}(S_t) + \text{Penalty}_{econ}(A_t)$$

$$\text{Penalty}_{health}(S_t) = w_1 \max(0, \text{ICU}_t - \text{Capacity})^2$$

$$\text{Penalty}_{econ}(A_t) = w_2 \cdot \text{Stringency}(A_t)$$

Policy A: ICU-Threshold Rule via Monte Carlo Grid Search

- **Concept:** The decision-making structure is strictly **predefined**. We force the policy into a rigid parametric step-function mapping the ICU load to 4 available intervention levels.
- **The Rule:** The algorithm must learn exactly **three threshold values** ($\tau_1 < \tau_2 < \tau_3$) to navigate the four states:
  - Level 1 (Open): ICU $< \tau_1$
  - Level 2 (Mild): $\tau_1 \leq$ ICU $< \tau_2$
  - Level 3 (Strict): $\tau_2 \leq$ ICU $< \tau_3$
  - Level 4 (Lockdown): ICU $\geq \tau_3$
- **Optimization:** The algorithm does not organically map states to actions; it tests a massive, fixed grid of (τ_1, τ_2, τ_3) combinations.
- **Handling Uncertainty:** For every triplet, it runs Monte Carlo simulated roll-outs using parameters drawn directly from the Bayesian (SMC^2) posterior, selecting the one with the lowest average cost.

Policy A: Advantages & Disadvantages

Advantages (The Political Reality)

- **Highly Interpretable:** Politicians and the public easily understand traffic-light systems (e.g., "Green, Yellow, Red" zones).
- **Predictability:** Citizens know exactly what triggers a lockdown.

Disadvantages

- **Structurally Constrained:** It is mathematically forced to be monotonic (interventions must strictly step upwards as beds fill).
- **Sub-optimal:** The total combined cost is consistently higher than a truly unconstrained approach.

Policy B: Posterior-Averaged Q-Learning

- **Concept:** A "blank slate" dynamic programming approach. The state-to-action mapping is an unconstrained discrete lookup table (Q-matrix), rather than a predefined staircase.
- **Optimization via Q-Learning:** The RL controller organically maps any discretized ICU state to any of the 4 actions by solving the Bellman equation to minimize long-term expected cost.

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha[C_{t+1} + \gamma \min_a Q(S_{t+1}, a) - Q(S_t, A_t)]$$

- **The Core Innovation (Posterior Averaging):**
  - Standard Q-learning assumes a single, fixed environment.
  - Here, every single training episode samples a fresh parameter particle θ directly from the SMC^2 posterior.
  - The controller receives penalties based on the full distribution of possible epidemic realities.
- **Result:** The final Q-table mathematically marginalizes out the epidemiological uncertainty.

Policy B: Advantages & Disadvantages

Advantages

- **True Optimality:** Finds the absolute lowest combined cost without human-imposed straightjackets.
- **Non-Linear Tactics:** Can discover counter-intuitive strategies, like the "Sharp Shock" (applying a brutal Level 4 lockdown when the ICU is low to crush momentum early, preventing long lockdowns later).
- **Robustness:** Inherently hedges against worst-case scenarios hiding in the tails of the Bayesian posterior.

Disadvantages

- **Black Box:** Much harder to explain to policymakers why the RL controller recommends a strict lockdown when hospitals are seemingly empty.
- **State Aliasing Risk:** Requires meticulous discretization of the ICU state space; too few bins causes mistakes, while too many prevents the Q-table from converging.

Full RL Method Workflow

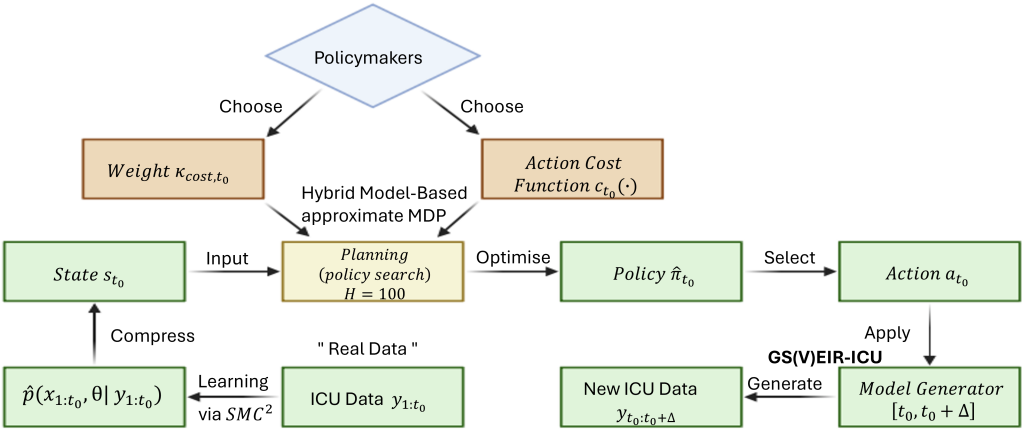
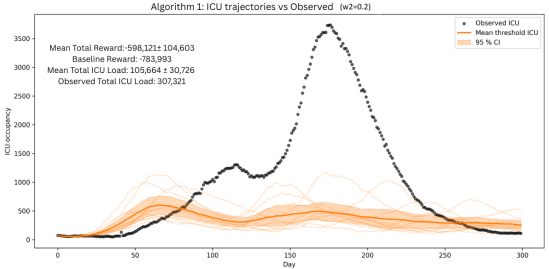


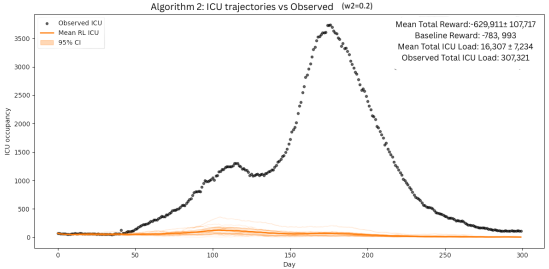
Figure: The complete sequential decision-making pipeline.

Results: RL Policies vs. Historical Government Actions (w2=0.2)

- We evaluated the frameworks using a counterfactual simulation based on real COVID-19 data from England.
- Both RL controllers reduced the combined socio-economic and health costs compared to the historical interventions.



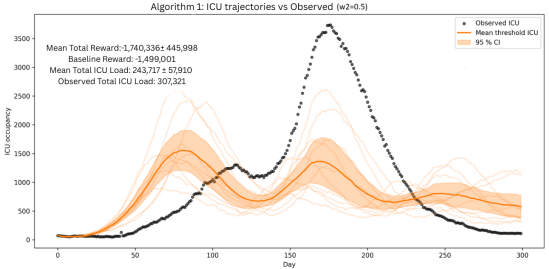
(a) Policy A



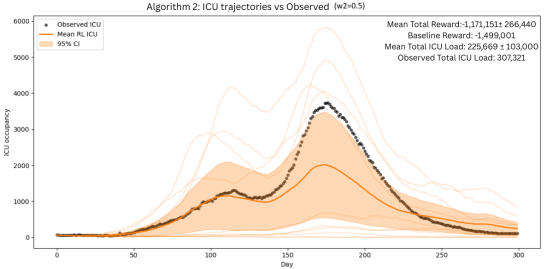
(b) Policy B

Results: RL Policies vs. Historical Government Actions (w2=0.5)

- We evaluated the frameworks using a counterfactual simulation based on real COVID-19 data from England.
- Both RL controllers reduced the combined socio-economic and health costs compared to the historical interventions.



(a) Policy A



(b) Policy B

Transition: Continuous Forecasting via Multi-Task GPs

- **The Challenge:** Multiphasic models capture discrete phase changes, but continuous forecasting across multiple distinct population groups yields high variance when estimated independently.
- **The Solution:** We integrate **Multi-task Exchangeable Gaussian Processes (GPs)** into the transmission framework to share temporal information and model cross-dependence between groups.
- **Broader Applicability:** While applied here to infectious diseases, this exchangeable multi-output framework is highly versatile for modelling correlated time series in other domains:
  - **Macroeconomics & Finance:** Modelling multivariate financial time series (e.g., exchangeable stochastic volatility or yield curve forecasting across interrelated markets).
  - **Energy Systems:** Probabilistic regional grid forecasting, where different locations share underlying seasonal trends but possess local noise.

Multi-task Scheme with Exchangeable GPs

- We model the log-transmission rate $x_i(t) = log(\lambda_i(t))$ for each task/group $i \in \{1, \ldots, S\}$ as:

$$x_i(t) = \mu(t) + D_i(t)$$

- $\mu(t)$ is the shared latent mean process capturing common epidemic characteristics, modelled as $\mathcal{GP}(0, \sigma_\mu^2 k_\mu(t, t'))$.
- $D_i(t)$ are the cross-sectional differences (type-specific deviations), modelled as independent $\mathcal{GP}(0, \sigma_x^2 k_x(t, t'))$, conditional on the hyper-parameters.
- This hierarchical construction yields a marginal GP with zero mean and cross-covariance kernel:

$$k_{i,j}(t, t') = \mathbb{E}[x_i(t)x_j(t')] = \delta_{i,j}\sigma_x^2 k_x(t, t') + \sigma_\mu^2 k_\mu(t, t')$$

where $\delta_{i,j}$ is the delta function.

Marginalisation of the Latent Mean Process

- For observations at equidistant points $t_1, \ldots, T$, let $X_i = (x_i(1), \ldots, x_i(T))'$.
- We denote $M = (\mu(t_1), \ldots, \mu(T))'$. The conditional distribution for each group is:

$$X_i \mid M \sim \mathcal{N}\Big(M,\, \sigma_x^2 C_x\Big), \quad \forall i = 1, \ldots, S$$

- By stacking all X_i into a single vector X and assigning the prior $M \sim \mathcal{N}\big(0_T, \sigma_\mu^2 C_\mu\big)$, the conditional joint distribution is:

$$X \mid M \sim \mathcal{N}\Big(1_S \otimes M,\, \sigma_x^2 I_S \otimes C_x\Big)$$

- We can exactly integrate out the shared mean M to obtain the marginal distribution:

$$X \sim \mathcal{N}\Big(0_{TS},\, \sigma_\mu^2 (1_S 1_S') \otimes C_\mu + \sigma_x^2 I_S \otimes C_x\Big)$$

Recovering the Shared Mean *Post Hoc*

- **Computational Advantage:** Marginalising M out of the likelihood dramatically reduces the parameter space and breaks the severe posterior correlations between X and M, allowing highly efficient sampling via Hamiltonian Monte Carlo (HMC).
- Because the multi-task model is jointly Gaussian, the unobserved shared mean M can be analytically recovered post-sampling via its exact conditional posterior:

$$M \mid X, \theta \sim \mathcal{N}\left(\Sigma_{M|X}\left[\Sigma_x^{-1}\sum_{i=1}^{S} X_i\right], \Sigma_{M|X}\right)$$

- Where the posterior covariance is given by:

$$\Sigma_{M|X} = \left(\Sigma_\mu^{-1} + S\Sigma_x^{-1}\right)^{-1}$$

Covariance Structures: Exchangeable vs. Multiple Exchangeable

Building directly upon our cross-covariance kernel formulation, we introduce varying levels of structural complexity for the type-specific deviations $D_i(t)$:

1. Exchangeable GP (x)

- Assumes all groups share a uniform variance σ_x^2 for their specific deviations.
- Useful when tasks/groups (e.g., adjacent regions) are expected to exhibit structurally homogeneous volatility.

2. Multiple Exchangeable GP (mx)

- Relaxes the variance assumption, allowing each specific group i to possess its own unique variance σ_i^2, yielding $\text{diag}(\sigma_1^2, \ldots, \sigma_S^2)$ in the marginalisation step.
- Crucial when comparing inherently different subpopulations (e.g., highly volatile younger age groups vs. stable elderly cohorts) while still sharing the common underlying temporal gradient $\mu(t)$.

Application 1: Continuous Forecasting of COVID-19 in Europe

- **Data:** Daily reproduction number (R_t) across Germany, Greece, and the UK.
- **Task:** Prequential analysis (1-week ahead, out-of-sample predictions rolled forward over 8 weeks).

Table: Predictive performance metrics (lower is better).

Metric	iBM	xBM	mxBM	iEQ	xEQ (Best)	mxEQ
LogS	-0.584	-0.587	-0.590	-0.639	**-1.018**	-0.803
CRPS	0.023	0.023	0.023	0.005	**0.003**	0.004
RMSE	0.025	0.025	0.024	0.011	**0.005**	0.008
MAE	0.017	0.017	0.017	0.007	**0.003**	0.005

Conclusion: Borrowing information on disease spread across countries via the shared underlying mean trajectory reduces out-of-sample forecasting errors.

Application 2: Chikungunya in French Polynesia & West Indies

- **Data:** Weekly incidence data across islands in French Polynesia (5 islands) and the West Indies (3 islands), incorporating precipitation covariates.

Table: Model selection via approximate leave-one-out information criterion (LOOIC). Lower is better. Standard errors are provided in parentheses.

Model	French Polynesia	French West Indies
SIR with random effects	1028.3 (33.4)	2158.1 (102.2)
xBM	**746.4 (14.8)**	1757.7 (62.3)
mxBM	753.8 (15.1)	1747.1 (71.1)
xEQ	753.1 (15.0)	1757.8 (58.3)
mxEQ	753.3 (14.2)	**1735.2 (65.9)**

All *GP-driven hierarchical models drastically outperformed the static baseline SIR model with random effects.*

Application 3: Age-Stratified COVID-19 in England

- **Model:** Gaussian-process priors on the logit-probability of infection given contact across four age bands.

Table: Model selection via LOOIC. Lower is better; standard errors in parentheses.

Model	LOOIC
iBM	**2867.2 (95.4)**
xBM	2869.8 (94.7)
mxBM	2870.7 (94.7)
iEQ	3180.6 (101.3)
xEQ	3182.8 (101.7)
mxEQ	3181.3 (101.5)

Key findings:

- The Brownian-motion specifications clearly outperform the exponentiated-quadratic alternatives, but within the BM and EQ families, the LOOIC values are very close.
- Nevertheless, the richer exchangeable specifications remain scientifically useful: the additional dependence parameters help quantify shared temporal structure across age groups, beyond what an independent model can reveal.

Summary of Adaptive Frameworks

1. **Epidemic Modelling Foundations:** Stochastic SEIR models driven by Poisson contact processes, leading directly to multi-type renewal equations.
2. **Retrospective Analysis:** Multiphasic models utilising Poisson/Dirichlet processes automatically detect transmission change-points without deterministic assumptions.
3. **Policy Optimisation:** SMC^2 inference combined with Posterior-Averaged Q-learning creates robust intervention strategies that adaptively balance ICU constraints against economic harm under uncertainty.
4. **Continuous Forecasting:** Exchangeable multi-task Gaussian Processes model the temporal dynamics of type-specific transmission rates, enabling probabilistic forecasting with cross-type information sharing.

Integrating Bayesian sequential learning with machine learning decision engines represents a powerful pathway for real-time epidemic control.

References

B.P. and Nikolaos Demiris (Nov. 2024). "Multiphasic stochastic epidemic models". In: *Journal of the Royal Statistical Society Series C: Applied Statistics* 74.2, pp. 491–505. ISSN: 0035-9254. DOI: 10.1093/jrsssc/qlae064. URL: https://doi.org/10.1093/jrsssc/qlae064.

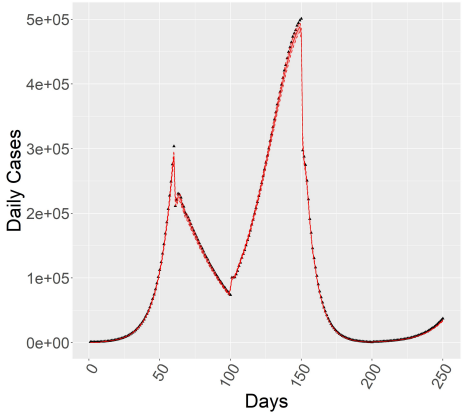
Bouranis, Lampros et al. (2025). *Exchangeable Gaussian Processes with application to epidemics*. arXiv: 2512.05227 [stat.ME]. URL: https://arxiv.org/abs/2512.05227.

Iannucci, Giacomo et al. (2025). *On a Reinforcement Learning Methodology for Epidemic Control, with application to COVID-19*. arXiv: 2511.18035 [stat.ME]. URL: https://arxiv.org/abs/2511.18035.

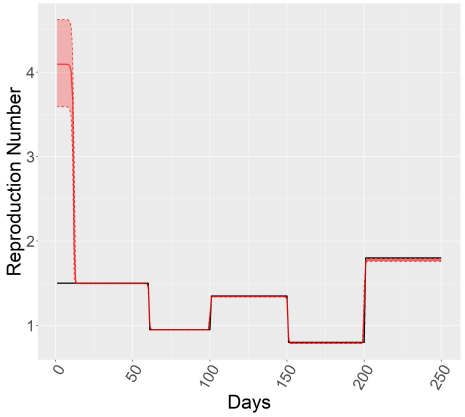
Questions?

Thank you for your attention!

Simulated Experiments-Results for Deterministic number of phases

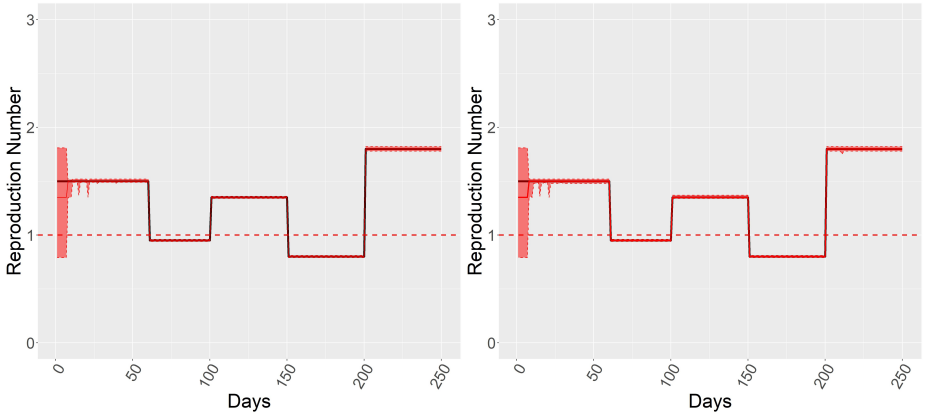


(a) Simulated (triangles) and estimated daily infections with 95% Cr.I. (line).



(b) Real (solid line) and estimated reproduction number R_t with 95% Cr.I. (dashed line).

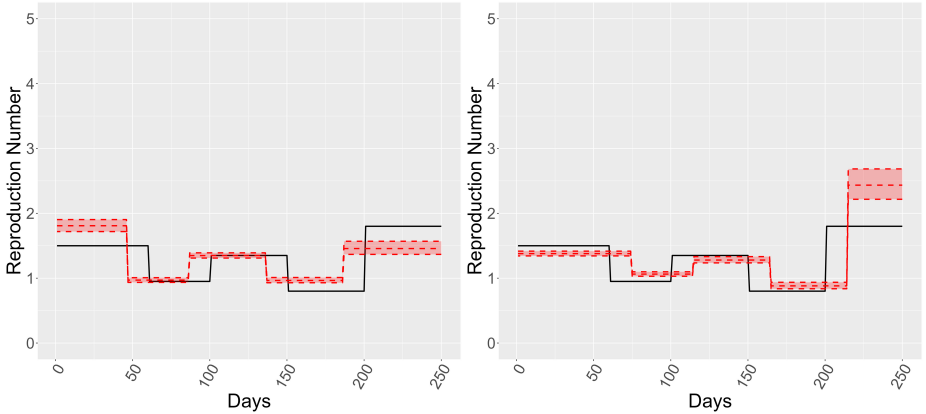
Simulated Experiments-Results for Stochastic number of phases



(a) Dirichlet process model

(b) Poisson Process model

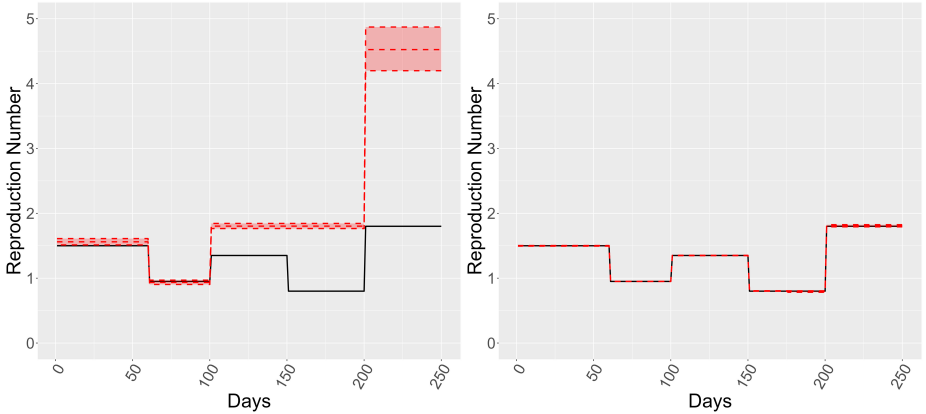
Simulated Experiments-What happens if we have wrong assumptions about the phases?



(a) The beginnings of the epidemic phases are assumed to start 2 weeks earlier.

(b) The beginnings of the epidemic phases are assumed to start 2 weeks later.

Simulated Experiments-What happens if we have wrong assumptions about the phases?



(a) The number of epidemic phases is considered lower than the true one.

(b) The number of epidemic phases is considered higher than the true one.

Toy Example: Tabular Q-Learning Mechanics

To demystify Q-learning, consider a simplified 5-room hospital corridor game.

- **States:** Room 1 to 5. The agent starts in Room 3.
- **Actions:** Move Left or Right.
- **Objective:** Minimise cumulative cost.
- **Rules:**
  - Reaching Room 1 (Crash Room) = Massive Cost of 100.
  - Reaching Room 5 (Discharge) = Cost 0 (Success).
  - Wandering the hallway = Cost 1 per step.

The Bellman Update Equation:

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha[C_{t+1} + \gamma \min_a Q(S_{t+1}, a) - Q(S_t, A_t)]$$

Toy Example: The Learned Policy

After 50 simulated episodes updating the Bellman equation, the Q-table (representing Expected Future Costs) converges:

State	Q(Left)	Q(Right)	Optimal Action
Room 1	0.00	0.00	Terminal
Room 2	51.98	4.47	**Right**
Room 3	16.63	2.52	**Right**
Room 4	4.06	0.00	**Right**
Room 5	0.00	0.00	Terminal

Interpretation: In Room 2, the algorithm learns moving "Left" leads to a massive expected penalty, whilst "Right" is cheap.