Bayesian analysis of diffusion-driven multi-type epidemic models with application to COVID-19

Lampros Bouranis

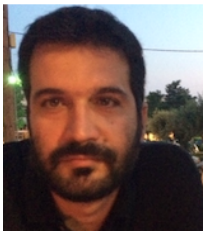
Marie Sklodowska-Curie Research Fellow
Department of Statistics
Athens University of Economics and Business, Greece

bouranis@aueb.gr

May 2023

The BERNADETTE MSCA project

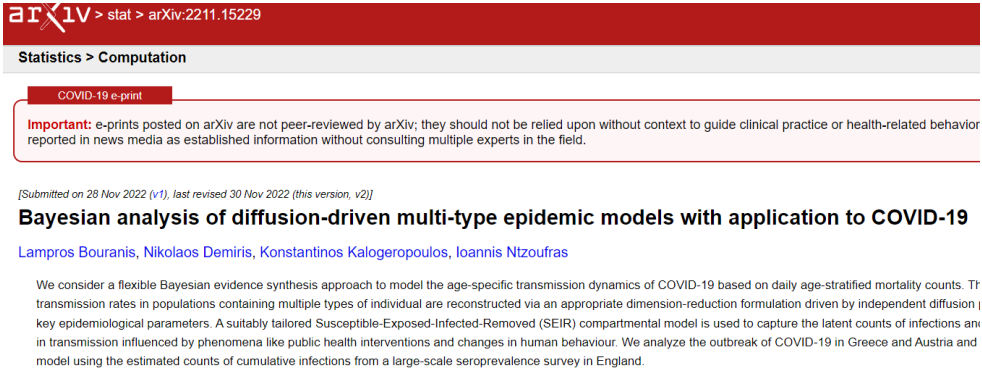


- **B**ayesian inf**ER**e**N**ce **A**nd mo**D**el s**E**lec**T**ion for s**T**ochastic **E**pidemics
- Collaborators: Demiris, Ntzoufras (AUEB), Kalogeropoulos (LSE)
- Webpage & R library
  - *https://bernadette-eu.github.io/*
  - *https://github.com/bernadette-eu/Bernadette*
- Funding
  - Supported by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 101027218.

Preprint details

- arXiv e-print: *https://arxiv.org/abs/2211.15229*.



Motivation

- Large-scale seroprevalence studies to estimate actual number of infections
  - Severe under-ascertainment (Ward et al., 2021), varying in time and across countries.
  - Level of under-ascertainment depends on national testing and tracing policies, testing capacities and impact of false positives under different regimes.
  - Only a proportion of infections detected and reported at the early stages of the pandemic (Li et al., 2020).
- Methods that rely on reported counts of infections expected to yield biases in the inferred rates of transmission.

Motivation

- Challenges of under-ascertainment of COVID-19 infections and presence of heterogeneity in type, relevance, and granularity of the data
  - Seminal paper by Flaxman et al. (2020) and its extension to multiple age groups by Monod et al. (2021).
- A Bayesian evidence synthesis approach for the analysis of COVID-19 data
  - Infer the true number of infections using age-stratified daily COVID-19 attributable mortality counts.
  - Learn the age-specific transmission rates.
  - Reconstruct the epidemic drivers from publicly available data sources.

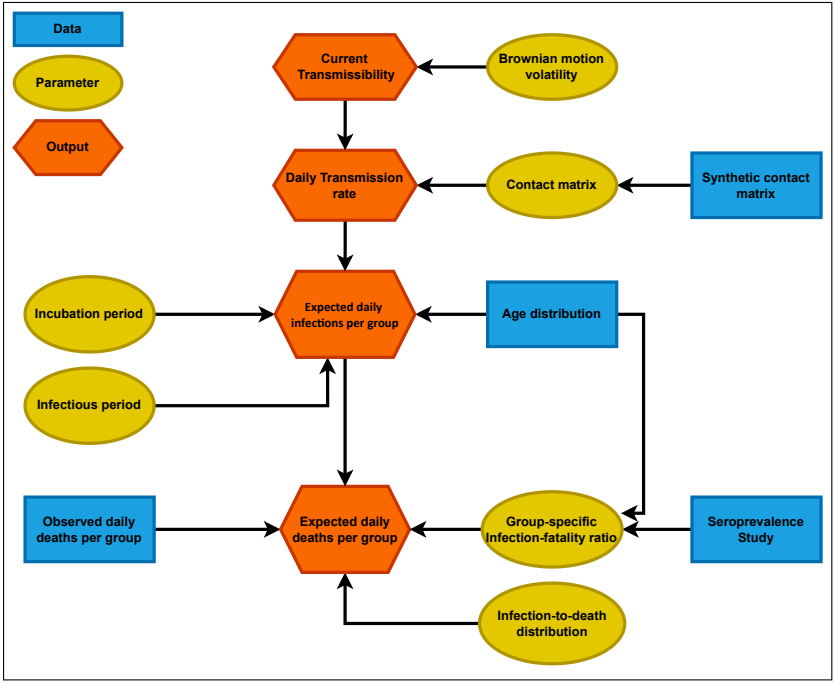
Introduction

- Allow for indirectly observed infections (Chatzilena et al., 2022).
- Age-stratified transmission model accounting for presence of social structures.
- Target the transmission rate matrix process:
  - Dimension increases quadratically with # of age groups.
  - Offer a dimension-reduction formulation projecting to latent diffusion processes.
- Desirable characteristics
  - Natural decomposition into underlying biological and social components.
  - Further evidence synthesis utilising information from contact surveys.
  - Driven by diffusion processes – capture temporal evolution of transmission rates & extrinsic environmental factors (NPIs and climatic changes).
  - Facilitates model determination at a latent level, performed here by appropriate model expansion.

Introduction

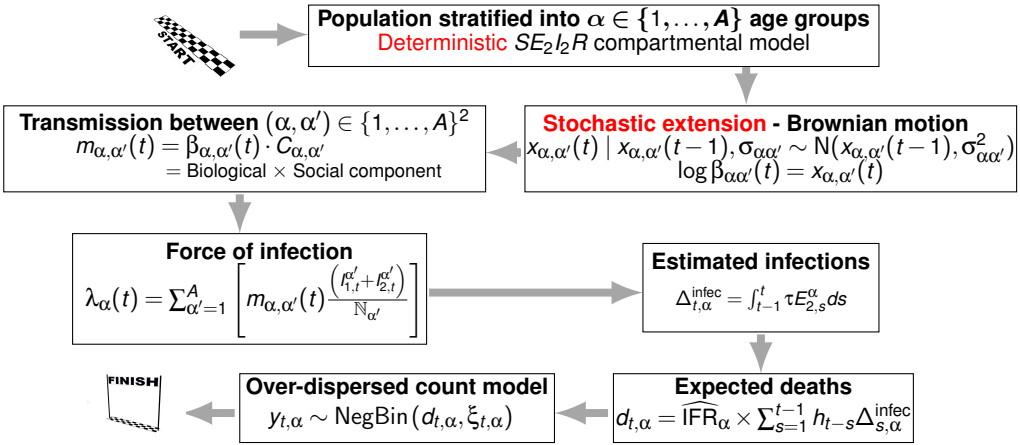
- Integration of multiple data sources, publicly available across countries:
  - Age-stratified daily COVID-19 attributable mortality counts and laboratory-confirmed COVID-19 infection counts.
  - Contact surveys.
  - Age distribution of the population.
- Contact survey data used to delineate potential identifiability issues (Britton, 1998) at the unobserved infection rate level when those rates are decomposed in their social and biological component.
- Uncertainty in contact structure is expressed via suitable prior distributions.

Model overview



Diffusion-driven multi-type transmission model

- Let $y_{t,\alpha}$ be the number of observed deaths on day $t = 1, \ldots, T$ in age group $\alpha \in \{1, \ldots, A\}$. A given infection may lead to observation events (i.e deaths) in the future.



Transmission rate matrix process

- Transmission between different age groups $(\alpha, \alpha') \in \{1, \ldots, A\}^2$:

$$\begin{aligned} m_{\alpha,\alpha'}(t) &= \beta_t^{\alpha\alpha'} \cdot C_{\alpha,\alpha'} \\ &= \text{Biological} \times \text{Social} \end{aligned}$$

- Diffusion process for modeling $\beta_t^{\alpha\alpha'}$ – Key advantages:
  - A data-driven approach; no need to specify the shape of change.
  - Tackles non-stationarity in the data.
  - Capture features like behavior changes, preventive measures, seasonal effects, holidays, etc.
  - Assess & interpret the evolution of transmission wrt mitigation strategies.

Transmission rate matrix process

- What we propose:
  - Independent diffusions to $\log(\beta_t^{11}), \log(\beta_t^{22}), \ldots, \log(\beta_t^{AA})$ with volatility parameters $\sigma_{x,\alpha}, \alpha \in \{1, \ldots, A\}$ [parsimony and interpretability]:

$$m_{\alpha,\alpha'}^{\text{MBM}}(t) = \beta_t^{\alpha\alpha'} \cdot C_{\alpha,\alpha'} \equiv \beta_t^{\alpha\alpha} \cdot C_{\alpha,\alpha'}.$$

  - Viewed as factor analysis, offering dimension reduction from elements of transmission rate matrix process to the Brownian motions $\beta_t^{\alpha\alpha}$ for $\alpha \in \{1, \ldots, A\}$.
  - Appealing feature – helps separate the contact matrix.
- Further reduction of the dimension of $m_{\alpha,\alpha'}^{\text{MBM}}(t)$:

$$m_{\alpha,\alpha'}^{\text{SBM}}(t) = \beta_t^{\alpha\alpha'} \cdot C_{\alpha,\alpha'} \equiv \beta_t \cdot C_{\alpha,\alpha'}.$$

- SBM is a nested model to MBM.

Observation process

- Denote the new COVID-19 attributable deaths occurring in day $t = 1, \ldots, T$ and age group $\alpha = 1, \ldots, A$ by $y_{t,\alpha}$.
- A given infection may lead to observation events (i.e deaths) in the future. Establish a link between $y_{t,\alpha}$ and the expected number of new infections via the function (Flaxman et al., 2020; Monod et al., 2021):

$$d_{t,\alpha} = \widehat{\text{IFR}}_{\alpha} \times \sum_{s=1}^{t-1} h_{t-s} \Delta_{s,\alpha}^{\text{infec}}.$$

- We link $d_{t,\alpha}$ under the proposed model to $y_{t,\alpha}$ through an over-dispersed count model (Birrell et al., 2021). The log-likelihood of the observed deaths is given by

$$\ell^{Deaths}(y \mid \phi) = \sum_{t=1}^{T} \sum_{\alpha=1}^{A} \text{logNegBin}\left(y_{t,\alpha} \mid d_{t,\alpha}, \xi_{t,\alpha}\right),$$

where $\xi_{t,\alpha} = \frac{d_{t,\alpha}}{\phi}$, such that $\mathbb{E}[y_{t,\alpha}] = d_{t,\alpha}$ and $\mathbb{V}[y_{t,\alpha}] = d_{t,\alpha}(1+\phi)$.

Observation process

- Infection-to-death distribution
  - Assumed to be Gamma distributed with mean 24.2 days and coefficient of variation 0.39 (Flaxman et al., 2020; Monod et al., 2021) and is given by

$$h \sim \text{Gamma}(6.29, 0.26).$$

- Infection fatality rate = deaths among all infected individuals
  - REal-time Assessment of Community Transmission-2 (REACT-2) national study of over 100,000 people in England (Ward et al., 2021): estimated total number of SARS-CoV-2 infections since start of the epidemic until mid-July 2020.

Age group	**15-44**	**45-64**	**65-74**	**75+**
IFR (%)	0.03	0.52	3.13	11.64

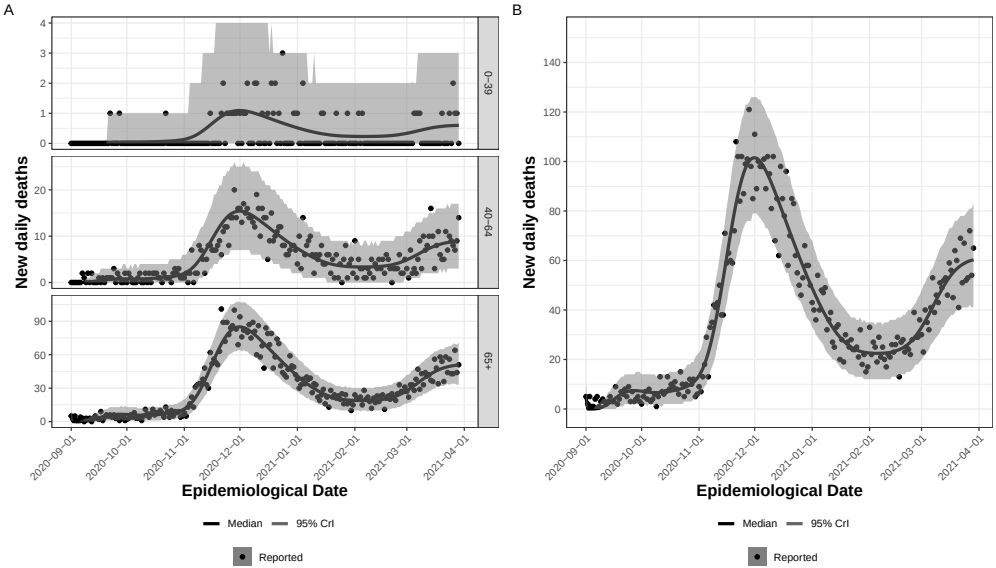
Parameter estimation - challenges

- Non-linear ODE system + SDE = hypo-elliptic diffusion $\rightarrow$ Intractable!
- Data augmentation framework of Dureau et al. (2013):
  - Latent sample path x of the diffusion is infinite-dimensional.
  - Indirectly observed through the time evolution of the disease states.
- Decrease dimensionality stochastic process:
  - Split study period $t = 1, \ldots, T$ into $k = 1, \ldots, K$ batches and model x_{k_t}.
  - Time-discretization via Euler-Maruyama approximation of x_{k_t}.

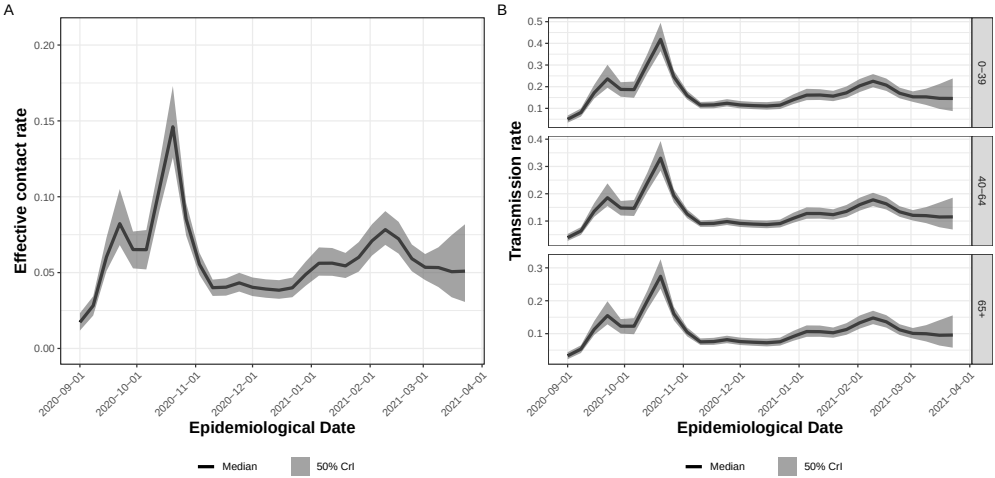
Parameter estimation - challenges

- Estimation of the volatilities $\sigma_{x,\alpha}, \alpha \in \{1, \ldots, A\}$.
- Longer time horizon $\rightarrow$ burden of ODE solver increases:
  - Numerical approximation via the Trapezoidal rule.
- Posterior sampling
  - Dynamic Hamiltonian Monte Carlo algorithm (Betancourt, 2018).
  - Can handle high-dim Θ and posteriors with weird shapes.
  - Volumes of data $\uparrow \Longrightarrow$ cost of solving ODEs $\uparrow$ (Birrell et al., 2021; Gosh et al., 2022): need for further developments.

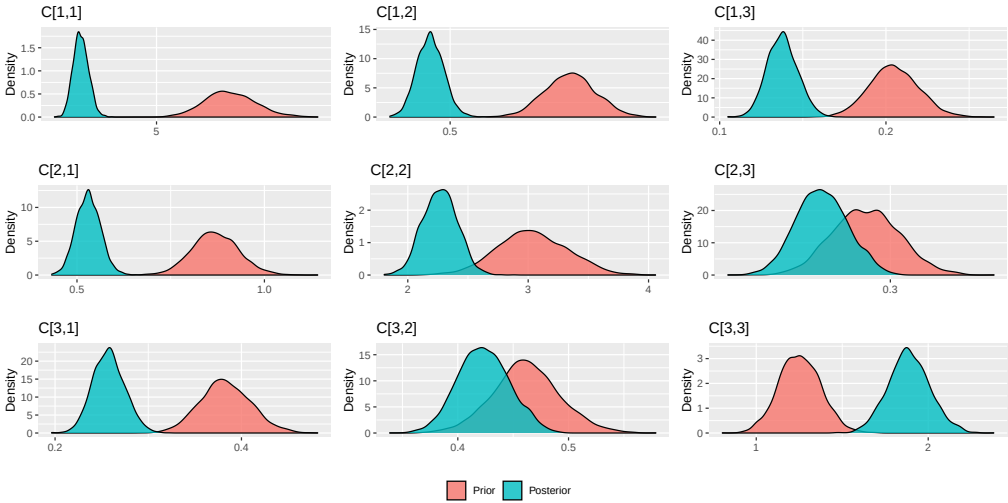
Greece – SBM model



Greece – SBM model



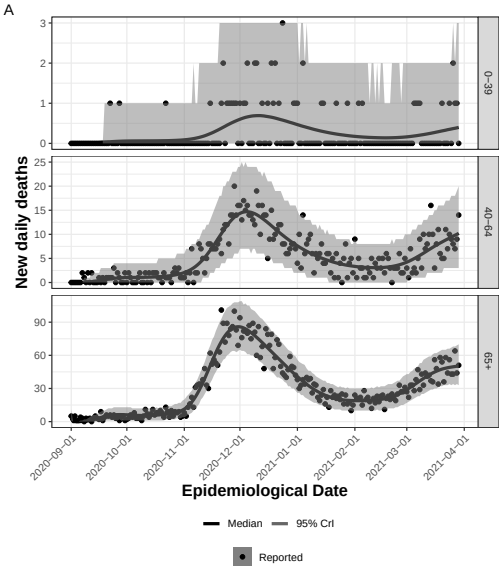
Greece – SBM model

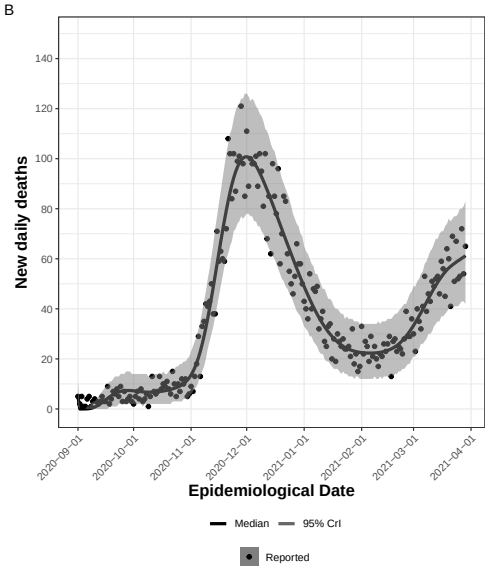


Greece – Model expansion

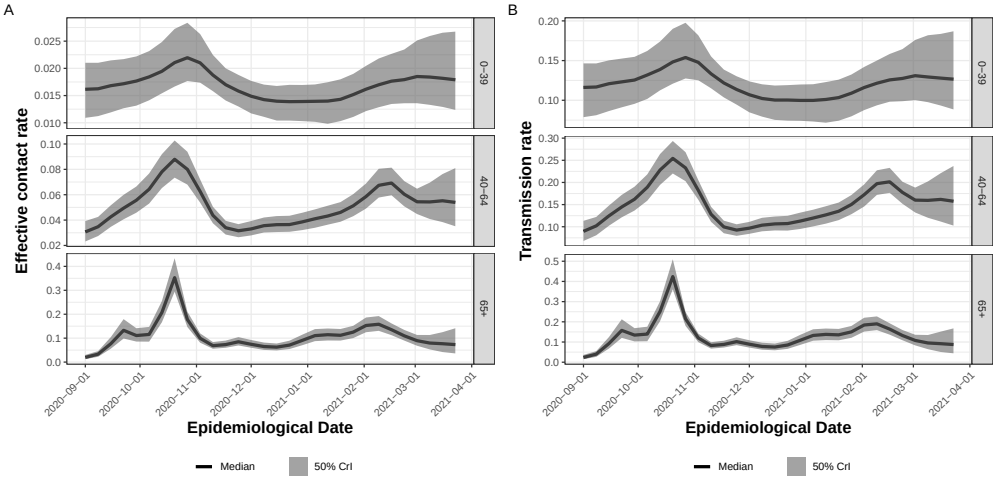
- The SBM model is not flexible enough to accommodate for age-specific trends in SARS-CoV-2 transmission.
- Assessment of the fidelity of the SBM model to the data at a latent level: prior-data conflict at a latent level.
- Expand the SBM model to the MBM model in the spirit of Gustafson (2005) and inspect the effect on the contact survey data.

Greece – Model expansion

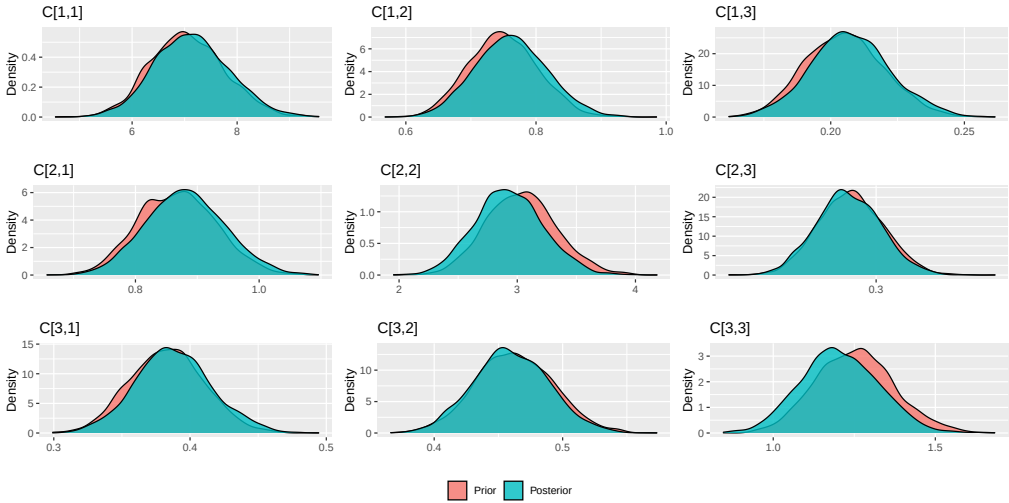




Greece – Model expansion



Greece – Model expansion

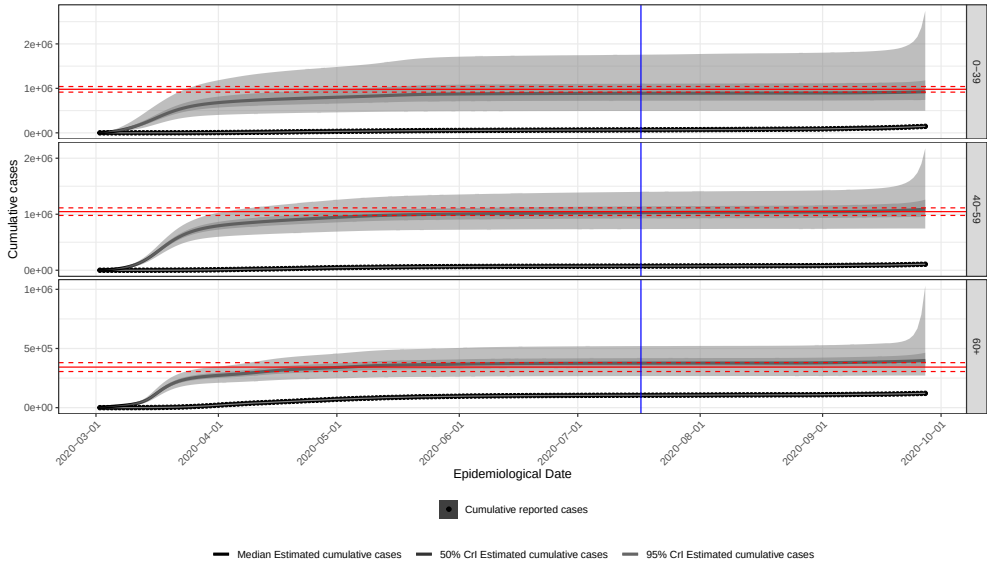


Greece – Model expansion

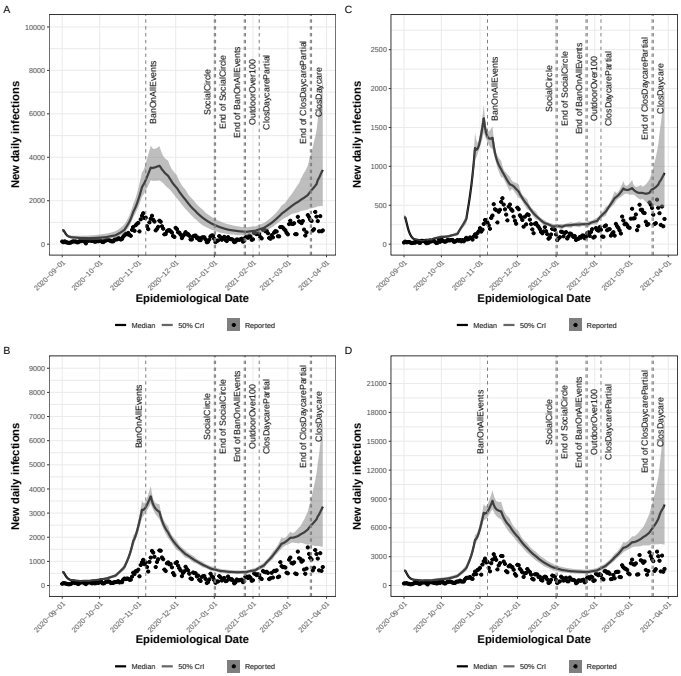
Table 1: Model determination. DIC: Deviance information criterion (Spiegelhalter et al., 2002); $\hat{p}_{DIC}$: estimated effective number of parameters using the DIC; LOOCV: Pareto smoothed importance sampling leave-one-out cross-validation information criterion and accompanying standard error (Vehtari et al., 2017) ; $\hat{p}_{loo}$: estimated effective number of parameters using LOOCV and accompanying standard error.

Model	DIC	$\hat{p}_{DIC}$	LOOCV (se)	$\hat{p}_{loo}$ (se)
SBM	2506.8	16.3	2507.4 (56.0)	15.9 (1.2)
MBM	2506.6	26.4	2506.2 (57.3)	24.6 (1.6)

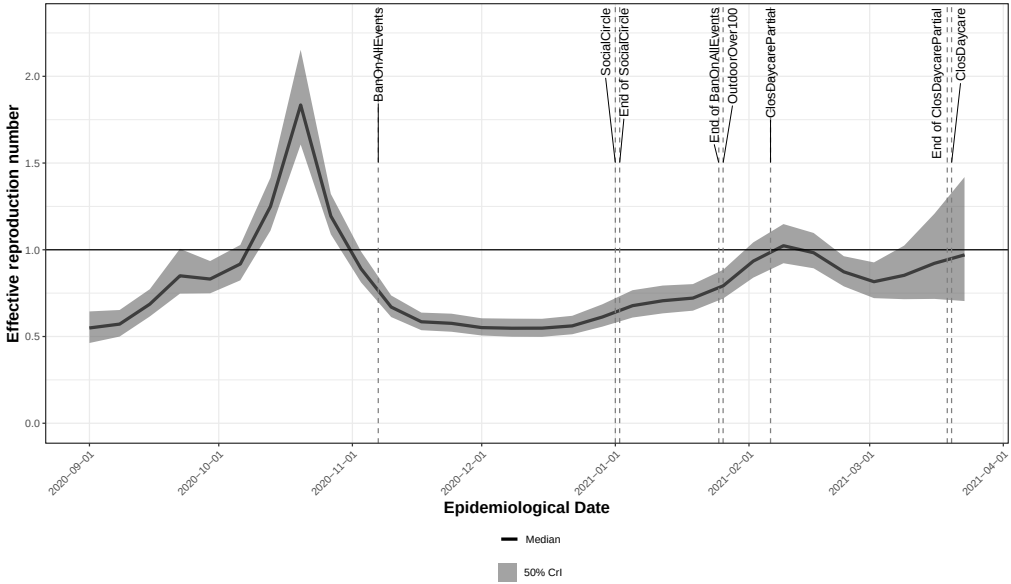
MBM External validation (England)



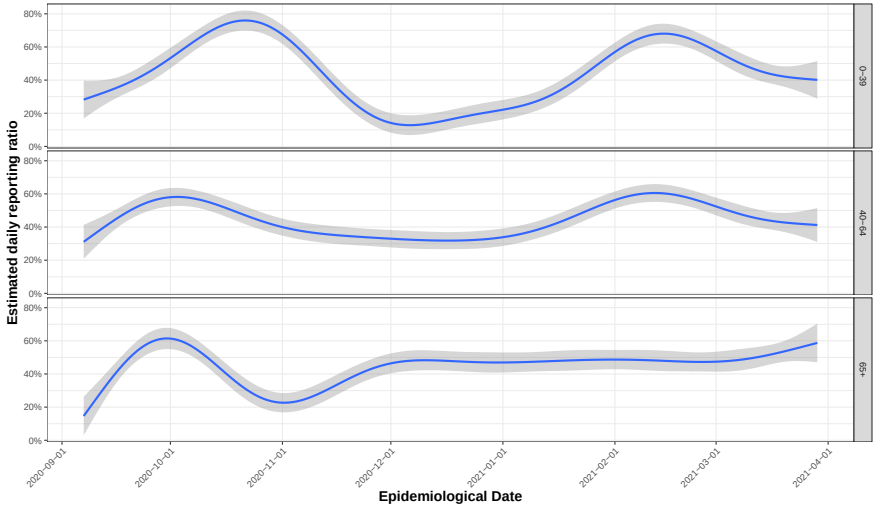
Greece – Age-specific infections (MBM)



Greece – Effective reproduction number (MBM)



Greece – Age-specific reporting ratio (MBM)



Exchangeable SDEs

- A Bayesian multi-task learning framework.
- Exchangeable GBMs - fall within the broad family of multi-output Gaussian Processes (GP).
- Multi-output GPs implemented for a simpler class of Bayesian infectious disease models (Seymour et al., 2022).
- Each Brownian motion path is drawn from an underlying population.
- Can be extended to include alternative Gaussian Processes to the GBM.

Exchangeable SDEs

- *A-priori* belief - shared structure across tasks.
- A Bayesian hierarchical model for:
  - Capturing cross-dependencies between tasks.
  - Introducing a natural interpretation based on the underlying mean stochastic process.
- Exchangeable GBMs improve predictive ability of the model compared to independent GBMs.

Concluding remarks

- Epidemic model driven by the distinct features of COVID-19 data:
  - A data-driven approach to inferring the mechanism governing COVID-19 transmission.
  - Diffusion processes that are a-priori independent.
  - Assess fidelity to the data at a latent level and resolve corresponding prior-data conflict via model expansion.
- Advantages:
  - Synthesis of multiple data sources, publicly available across countries.
  - Remove need for additional information on timing of interventions and hypotheses about their impact on transmission.
  - Model allows policymakers to assess effect of NPIs on each age group.

Concluding remarks

- Limitations:
  - Proposed hierarchical model better suited to the pre-vaccine era.
  - Variations in reporting procedure of deaths and mortality definitions across time and countries.
  - Parametric assumptions for infection-to-death distribution.
  - Account for hospital-acquired infections.
  - Time-invariant age-stratified IFRs.
- Moving forward:
  - Age-stratified IFRs & integration of further healthcare surveillance data.
  - Exchangeable diffusion processes: shared structure between dynamic transmission rates for individuals of different age groups.

References

Betancourt, M. (2018). A conceptual introduction to Hamiltonian Monte Carlo. https://arxiv.org/abs/1701.02434.

Birrell, P., Blake, J., van Leeuwen, E., Gent, N., and De Angelis, D. (2021). Real-time nowcasting and forecasting of COVID-19 dynamics in England: the first wave. *Phil Trans R Soc B*, 376:20200279.

Chatzilena, A., Demiris, N., and Kalogeropoulos, K. (2022). A modelling framework for the analysis of the transmission of SARS-CoV2. https://arxiv.org/abs/2203.03773.

Davies, N., Klepac, P., Liu, Y., Prem, K., Jit, M., CMMID COVID-19 working group, and Eggo, R. (2020). Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nat. Med.*, 26:1205–1211.

Dureau, J., Kalogeropoulos, K., and Baguelin, M. (2013). Capturing the time-varying drivers of an epidemic using stochastic dynamical systems. *Biostatistics*, 14(3):541–555.

Flaxman, S., Mishra, S., Gandy, A., Unwin, H., Mellan, T., Coupland, H., Whittaker, C., Zhu, H., Berah, T., and Eaton, J. (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, 584(7820):257–261.

Gosh, S., Birrell, P., and De Angelis, D. (2022). An approximate diffusion process for environmental stochasticity in infectious disease transmission modelling. https://arxiv.org/abs/2208.14363.

Gustafson, P. (2005). On Model Expansion, Model Contraction, Identifiability and Prior Information: Two Illustrative Scenarios Involving Mismeasured Variables. *Stat Sci*, 20(2):111–140.

Hauser, A., Counotte, M., Margossian, C., Konstantinoudis, G., Low, N., Althaus, C., and Riou, J. (2020). Estimation of SARS-CoV-2 mortality during the early stages of an epidemic: A modeling study in Hubei, China, and six regions in Europe. *PLoS Med*, 17(7):1–17.

Hellenic National Public Health Organization (2022). COVID-19 Daily Deaths. https://eody.gov.gr/epidimiologika-statistika-dedomena/imerisies-ektheseis-covid-19/.

References (cont.)

Knock, E., Whittles, L., Lees, J., Perez-Guzman, P., Verity, R., FitzJohn, R., Gaythorpe, K., Imai, N., Hinsley, W., Okell, L., Rosello, A., Kantas, N., Walters, C., Bhatia, S., Watson, O., Whittaker, C., Cattarino, L., Boonyasiri, A., Djaafara, B., Fraser, K., Fu, H., Wang, H., Xi, X., Donnelly, C., Jauneikaite, E., Laydon, D., White, P., Ghani, A., Ferguson, N., Cori, A., and Baguelin, M. (2021). Key epidemiological drivers and impact of interventions in the 2020 SARS-CoV-2 epidemic in England. *Sci. Transl. Med.*, 13(602):eabg4262.

Li, R., Pei, S., Chen, B., Song, Y., Zhang, T., Yang, W., and Shaman., J. (2020). Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (sars-cov-2). *Science*, 368(6490):89–493.

Liu, T., Hu, J.and Kang, M., Lin, L., Zhong, H., Xiao, J., He, G., Song, T., Huang, Q., Rong, Z., Deng, A., Zeng, W., Tan, X., Zeng, S., Zhu, Z., Li, J., Wan, D., Lu, J., Deng, H., and Ma, W. (2020). Transmission dynamics of 2019 novel coronavirus (2019-nCoV). *Lancet*.

Monod, M., Blenkinsop, A., Xi, X., Hebert, D., Bershan, S., Tietze, S., Bradley, V., Chen, Y., Coupland, H., Filippi, S., Ish-Horowicz, J., McManus, M., Mellan, T., Gandy, A., Hutchinson, M., Unwin, J., Vollmer, C., Michaela, A., Weber, S., Zhu, H., Bezancon, A., Ferguson, N., Mishra, S., Flaxman, S., Bhatt, S., and Ratmann, O. (2021). Age groups that sustain resurging COVID-19 epidemics in the United States. *Science*, 371(6536):eabe8372.

National Health Service England (2022). COVID-19 Daily Deaths. `https://www.england.nhs.uk/statistics/statistical-work-areas/covid-19-daily-deaths/`.

Office for National Statistics (2022). Population and migration. `https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates`.

Prem, K., Zandvoort, K., Klepac, P., Eggo, R., Davies, N., Cook, A., and Jit, M. (2021). Projecting contact matrices in 177 geographical regions: An update and comparison with empirical data for the COVID-19 era. *PLoS Comput Biol*, 17(7):1–19.

Seymour, R., Kypraios, T., and O'Neill, P. (2022). Bayesian nonparametric inference for heterogeneously mixing infectious disease models. *Proc. Natl. Acad. Sci.*, 119(10):e2118425119.

References (cont.)

Spiegelhalter, D., Best, N., Carlin, B., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *J. R. Statist. Soc. B*, 64(4):583–639.

UK Health Security Agency (2022). COVID-19 datasets. `https://coronavirus.data.gov.uk/details/download/`.

United Nations: Department of Economic and Social Affairs: Population Division (2020). *World population prospects*. United Nations, New York, NY.

Vehtari, A., Gelman, A., and Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.*, 27:1413–1432.

Ward, H., Atchison, C., Whitaker, M., Ainslie, K., Elliott, J., Okell, L., Redd, R., Ashby, D., Donnelly, C., Barclay, W., Darzi, A., Cooke, G., Riley, S., and Elliott, P. (2021). Sars-cov-2 antibody prevalence in england following the first peak of the pandemic. *Nat. Commun.*, 12:905.

Wood, S., Pya, N., and Säfken, B. (2016). Smoothing parameter and model selection for general smooth models. *J. Am. Statist. Ass.*, 111(516):1548–1563.

Data sources and definitions

Table 2: Data sources and definitions.

Data type	Description	Source	Reference
Mortality	Daily number of new deaths by age group according to the date of death, tested positive for COVID-19. For England, hospital-only.	NHS, England; Hellenic National Public Health Organization	National Health Service England (2022); Hellenic National Public Health Organization (2022)
Infections	Daily number of new cases by age group and specimen date	NHS, England; Hellenic National Public Health Organization	UK Health Security Agency (2022); Hellenic National Public Health Organization (2022)
Age distribution	Age distribution for a given year, broken down by 5-year age bands and gender	United Nations; Office for National Statistics, UK	United Nations: Department of Economic and Social Affairs: Population Division (2020); Office for National Statistics (2022)
Contact matrix	Rate of contacts between age groups		Prem et al. (2021)

Diffusion-driven multi-type transmission model (1)

- Age-stratified deterministic Susceptible-Exposed-Infected-Recovered ($\mathbb{N}_\alpha = S_t^\alpha + E_{1,t}^\alpha + E_{2,t}^\alpha + I_{1,t}^\alpha + I_{2,t}^\alpha + R_t^\alpha$) compartmental model
  - Force of infection:

$$\lambda_\alpha(t) = \beta_t^{\alpha\alpha'} \sum_{\alpha'=1}^{A} \left[C_{\alpha\alpha'} \frac{\left(I_{1,t}^{\alpha'} + I_{2,t}^{\alpha'} \right)}{\mathbb{N}_{\alpha'}} \right], \quad \alpha = 1, \ldots, A.$$

  - Non-linear system of ODEs:

$$\begin{cases} \frac{dS_t^\alpha}{dt} & = -\lambda_\alpha(t) S_t^\alpha, \\ \frac{dE_{1,t}^\alpha}{dt} & = \lambda_\alpha(t) S_t^\alpha - \tau E_{1,t}^\alpha, \quad \frac{dE_{2,t}^\alpha}{dt} = \tau \left(E_{1,t}^\alpha - E_{2,t}^\alpha \right) \\ \frac{dI_{1,t}^\alpha}{dt} & = \tau E_{2,t}^\alpha - \gamma I_{1,t}^\alpha, \quad \frac{dI_{2,t}^\alpha}{dt} = \gamma \left(I_{1,t}^\alpha - I_{2,t}^\alpha \right) \\ \frac{dR_t^\alpha}{dt} & = \gamma I_{2,t}^\alpha \\ \Delta_{t,\alpha}^{\text{infec}} & = \int_{t-1}^{t} \tau E_{2,s}^\alpha ds \end{cases} \tag{1}$$

- Mean Incubation period [$d_E = 2/\tau$]; Mean Infection period [$d_I = 2/\gamma$].

Prior distributions

- Initial conditions: t_0 = simulation start; most individuals assumed to be susceptible and are distributed across A according to $\mathbb{N}_\alpha$. The number of people in the exposed compartments at t_0 is controlled by a parameter π, and also distributed according to $\mathbb{N}_\alpha$. Other compartments are set to 0.

Table 3: Model parameters with assumed prior distributions or fixed values.

Symbol	Description	Prior source
Transmission model		
π	Initial proportion of exposed (at t_0)	Beta with $\mathbb{E}[\pi] = 0.1$, $\mathbb{V}[\pi] = 0.05 \cdot \mathbb{E}[\pi]$. Based on Hauser et al. (2020).
$F_{i,j}$	Entries of the contact matrix F	Half-Normal with $\mathbb{E}[F_{i,j}] = \mathcal{F}_{i,j}$, $\mathbb{V}[\pi] = (0.05 \cdot \mathcal{F}_{i,j})^2$ $i,j = 1,\ldots,A$.
x_0	Brownian motion at t_0	$\text{Normal}(0, 5^2)$
x_1	Brownian motion at t_1	$\text{Normal}(0, 4^2)$
x_t	Brownian motion	$x_t \mid x_{t-1}, \sigma_x^2 \sim \text{Normal}(x_{t-1}, \sigma_x^2)$, $t = 2,\ldots,T$.
σ_x	Volatility of the Brownian motion	$\text{Half-Normal}(0, 4)$.
Observational model		
ϕ	Negative Binomial over-dispersion	Exp(0.2). Based on Birrell et al. (2021).
Parameters assumed known		
$\widehat{\text{IFR}}_\alpha$	Infection-fatality rate	Based on Ward et al. (2021).
d_E	Mean Incubation period	3 days. Based on Liu et al. (2020).
d_I	Mean Infection period	4 days. Based on Liu et al. (2020).

Effective Reproduction number

- $R_t^{eff} =$ avg # of secondary infections a case infected at time t would generate, accounting for the finite population size and potential immunity in the population (Davies et al., 2020; Knock et al., 2021).
- The next-generation matrix was calculated as

$$\text{NGM}_{\alpha,\alpha'}(t) = m_{\alpha,\alpha'}(t) d_I \frac{S_\alpha(t)}{\mathbb{N}_{\alpha'}}.$$

- $R_t^{eff} =$ absolute value of the dominant eigenvalue of NGM(t).

Age-specific reporting ratio

- Combine daily laboratory-confirmed COVID-19 infections with estimated daily number of infections in the population.
- Age-specific posterior median of the number of infected individuals in the population at time $t - L$.
- Allows for a reporting delay of L days between infection and report.
- Laboratory-confirmed infections which are reported at time t denoted by $\Delta_{t,\alpha}^{\text{infec,rep}}$.
- Estimated age-specific daily reporting ratio for $\alpha \in \{1, \ldots, A\}$ expressed by

$$\hat{r}_{t,\alpha} = \frac{\Delta_{t,\alpha}^{\text{infec,rep}}}{\hat{\Delta}_{t-L,\alpha}^{\text{infec,pop}}}.$$

We considered a time delay between infection and report (L) equal to 6 days; a time-varying spline-based smoother was applied to $\hat{r}_{t,\alpha}$ via generalized additive model smoothing (Wood et al., 2016).