

Regularised B-splines Projected Gaussian Process Priors to Estimate Time-trends in Age-specific COVID-19 Deaths^{*†}

Mélodie Monod^{‡,†††,†††}, Alexandra Blenkinsop[§], Andrea Brizzi[¶], Yu Chen^{||}, Carlos Cardoso Correia Perello^{**}, Vidoushee Jogarah^{††}, Yuanrong Wang^{‡‡}, Seth Flaxman^{¶¶}, Samir Bhatt^{|||,***,†††} and Oliver Ratmann^{§§,†††,†††}

Abstract. The COVID-19 pandemic has caused severe public health consequences in the United States. In this study, we use a hierarchical Bayesian model to estimate the age-specific COVID-19 attributable deaths over time in the United States. The model is specified by a novel non-parametric spatial approach over time and age, a low-rank Gaussian Process (GP) projected by regularised B-splines. We show that this projection defines a GP with attractive smoothness and computational efficiency properties, derive its kernel function, and discuss the penalty terms induced by the projected GP. Simulation analyses and benchmark results show that the B-splines projected GP may perform better than standard B-splines and Bayesian P-splines, and equivalently well as a standard GP at considerably lower runtimes. We apply the model to weekly, age-stratified COVID-19 attributable deaths reported by the US Centers for Disease Control, which are subject to censoring and reporting biases. Using the B-splines projected GP, we can estimate longitudinal trends in COVID-19 associated deaths across the US by

^{*}This work was supported by the Imperial College Research Computing Service, DOI: [10.14469/hpc/2232](https://doi.org/10.14469/hpc/2232), the Imperial College COVID-19 Response Fund, the EPSRC through the EPSRC Centre for Doctoral Training in Modern Statistics and Statistical Machine Learning and the Imperial President's PhD Scholarships. S. Bhatt acknowledges funding from UK Research and Innovation (MR/V038109/1), the Academy of Medical Sciences Springboard Award (SBF004/1080), The MRC (MR/R015600/1), the Bill & Melinda Gates Foundation (OPP1197730), Imperial College Healthcare NHS Trust-BRC Funding (RDA02), the Novo Nordisk Foundation (NNF20OC0059309) and the NIHR Health Protection Research Unit in Modelling Methodology; and O. Ratmann acknowledges funding from UK Research and Innovation (MR/V038109/1) and the Bill & Melinda Gates Foundation (OPP1175094).

[†]This article was first posted with a typo at the start of Section 2.1, just after equation (2.2): the authors wrote “(daily deaths, d , for state m , in age group b , on week w).” while it should be “(weekly deaths, d , for state m , in age group b , on week w).”. The typo was corrected on 2022-10-17.

[‡]Department of Mathematics, Imperial College London, melodie.monod18@imperial.ac.uk

[§]Department of Mathematics, Imperial College London

[¶]Department of Mathematics, Imperial College London

^{||}Department of Mathematics, Imperial College London

^{**}Department of Mathematics, Imperial College London

^{††}Department of Mathematics, Imperial College London

^{‡‡}Department of Mathematics, Imperial College London

^{§§}Department of Mathematics, Imperial College London, oliver.ratmann@imperial.ac.uk

^{¶¶}Department of Computer Science, University of Oxford

^{|||}MRC Centre for Global Infectious Disease Analysis, Jameel Institute, School of Public Health, Imperial College London

^{***}Section of Epidemiology, Department of Public Health, University of Copenhagen

^{†††}Contributed equally.

^{‡††}Corresponding authors.

1-year age bands. These estimates are instrumental to calculate age-specific mortality rates, describe variation in age-specific deaths across the US, and for fitting epidemic models. Here, we couple the model with age-specific vaccination rates to show that vaccination rates were significantly associated with the magnitude of resurgences in COVID-19 deaths during the summer 2021. With counterfactual analyses, we quantify the avoided COVID-19 deaths under lower vaccination rates and avoidable COVID-19 deaths under higher vaccination rates. The B-splines projected GP priors that we develop are likely an appealing addition to the arsenal of Bayesian regularising priors.

Keywords: Bayesian statistics, non-parametric statistics, Gaussian process approximations, Stan probabilistic computing language, COVID-19, mortality.

MSC2020 subject classifications: Primary 62F15, 60G15, 65D07; secondary 62P10.

1 Introduction

A new pathogen, Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), emerged in the Wuhan region of China in December 2019 and continues to spread worldwide. The resulting disease, COVID-19, is severe, with overall infection fatality rates (IFRs) between 0.1% and 1% (Meyerowitz-Katz and Merone, 2020; Brazeau et al., 2020), which increase exponentially with age (Levin et al., 2020). We note those IFRs are for the original SARS-CoV-2 strain assuming no treatment, vaccination or prior immunity. In the United States (US), vaccines have been initially prioritised to older age groups before being offered to all adults from May 2021. Developed vaccines are highly effective to prevent deaths (Haas et al., 2021; Baden et al., 2021), but uptake has been variable across states (Centers for Disease Control and Prevention, 2021a). Despite increasing vaccination coverage in the US, several US states reported resurgent COVID-19 death waves in the summer of 2021, most notably Florida. These death waves were equal to or greater in magnitude than the death waves observed before vaccination. A key question is whether these death waves are linked to low vaccination coverage in adults.

Here, we provide methods for estimating high resolution age-specific COVID-19 attributable deaths over time. Doing so requires a statistical model because data are partially censored, reported with delays, and reported in age bands that can be inconsistent across locations or do not match those of other data streams. Numerous methods have been developed to model the time evolution of overall COVID-19 attributable deaths (e.g., Lavezzo et al., 2020; Institute for Health Metrics and Evaluation, 2020; Chen et al., 2020; Blangiardo et al., 2020; Zheng et al., 2021; Flaxman et al., 2020), but few are suitable to analyze longitudinal data stratified by age brackets or other discrete strata (Monod et al., 2021), let alone at high resolution such as 1-year age bands. One reason for this paucity of methods is that in SIR-type models, calculating age-specific next generation operators at a time resolution of days becomes computationally slow over observation periods that span years, especially when calculations need to be repeated millions of times within a Bayesian framework (Wikle et al., 2020; Monod et al., 2021). These considerations are prompting us and others to consider non-mechanistic,

flexible estimation approaches (Shah et al., 2020; Pokharel and Deardon, 2021). As our objective is to reconstruct spatio-temporal trends in age-specific COVID-19 attributable deaths, it does not require invoking many of the assumptions or complexities that underlie mechanistic or semi-mechanistic COVID-19 transmission dynamics models, and Bayesian non-parametric models are sufficient. We present a fully Bayesian, computationally scalable approach to estimate a two-dimensional (2D) surface over ages and weeks that describes the time evolution of COVID-19 deaths by 1-year age bands at US state level. To impute missing entries on the surface and estimate global trends over ages and weeks, we borrow information across neighboring entries by using a non-parametric smoothing method.

A natural starting point for modelling a surface is a 2D Gaussian process (GP) (Rasmussen and Williams, 2005). However, their computational complexity makes the use of 2D GPs in a fully Bayesian framework difficult when the surface dimension becomes large, even after using a Kronecker structure on the covariance (Saatçi, 2011; Wilson et al., 2014). Here we adopt a low-rank approximation via a tensor product of B-splines for which the parameters follow a 2D GP. We refer to this method as “regularised B-splines projected 2D GP”. The resulting approach is equivalent to a 2D GP defined by a low-rank covariance matrix projected by B-splines. B-splines are a popular choice for non-parametric modelling, due to their continuity properties ensuring smoothness of the fitted surface, and their easy implementation. But choosing the optimal number and position of knots—the defining grid segments where the surface is expected to change its behavior—on the space of ages and weeks is a complex task. Some approaches have focused on adding a penalty to restrict the flexibility of the fitted surface in a frequentist framework (O’Sullivan, 1986, 1988; Eilers and Marx, 1996; Eilers et al., 2006). Following this idea, we regularise the fitted surface by penalizing the complexity of the low-rank 2D GP’s covariance matrix. We qualitatively compare the penalty induced by this choice to that of related regularisation methods. We benchmark the proposed regularised B-splines projected 2D GP against several other popular smoothing methods, and we demonstrate that our approach results in substantive computational gains over a standard 2D GP for similar estimation accuracy.

This paper aims to reconstruct age-specific COVID-19 attributable deaths on publicly available incomplete and censored age-specific COVID-19 attributable deaths reported by the Centers for Disease Control and Prevention (CDC) (2021c). It is structured as follows. Section 2 introduces the data and their limitations. Section 3 describes the proposed methodology to model trends in the age-composition of COVID-19 deaths, including a theoretical characterization of the penalties introduced by regularised B-splines projected 2D GPs. Section 4 presents a comparison of the proposed method to related smoothing approaches on simulated data, and Section 5 on real world data used for benchmarking. In Section 6 we present the predicted time trends of age-specific COVID-19 deaths, estimated age profile of COVID-19 deaths and predicted age-specific COVID-19 mortality rates. Additionally, we document the marked variation in the summer 2021 resurgence in age-specific COVID-19 deaths across states, and show that strong resurgences in deaths are associated low vaccine coverage. Lastly, Section 7 closes with a discussion.

2 COVID-19 deaths data

2.1 COVID-19 deaths data

The Center for Disease Control (CDC) and National Center for Health Statistics (NCHS) report each week on the total number of deaths involving COVID-19 that have been received and tabulated through the National Vital Statistics database (Centers for Disease Control and Prevention, 2021c) for every US state across the age groups

$$\mathcal{B} = \{0, 1-4, 5-14, 15-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75-84, 85+\}. \quad (2.1)$$

We refer to the latter as cumulative reported COVID-19 attributable deaths, $D_{m,b,w}$, for state m , in age group b and on week w . Historical records are made available by Rearc (2021). To simplify the longitudinal dependence and create a time series from the cumulative counts (King et al., 2015), we obtain the weekly COVID-19 attributable deaths by differencing,

$$d_{m,b,w} = D_{m,b,w+1} - D_{m,b,w}, \quad (2.2)$$

in each location and age group for all but the last available week (weekly deaths, d , for state m , in age group b , on week w). We index the weekly deaths by $w \in \mathcal{W} = \{1, \dots, W\}$.

Reflecting the reporting nature of the age-stratified CDC data, the weekly COVID-19 attributable $d_{m,b,w}$ are subject to reporting delays and do not necessarily correspond to the number of individuals who died of COVID-19 in state m and week w . The CDC does not report the cumulative deaths count if the count is between 1 and 9. Thus, it is not possible to retrieve all weekly deaths, and we refer to the set of weeks that are retrievable in state m for age group b through first order differencing by $\mathcal{W}_{m,b}^{\text{WR}}$, and to the set of weeks that are non-retrievable because of censoring by $\mathcal{W}_{m,b}^{\text{WNR}}$. The censored cumulative deaths are bounded, such that the sum of non-retrievable weekly deaths is also bounded. The exact computations are described in Supplementary Text, Section S2.1 (Monod et al., 2022). In addition, there was no update on July 4, 2020 resulting in missing (i.e., non-reported) cumulative deaths in that week. The weekly deaths of that week and the preceding week are declared missing. Note that the missing weekly deaths are not equivalent to the non-retrievable weekly deaths because we cannot bound their sums. In the main text, we focus on data reported from May 2, 2020 to January 01, 2022. We thus have $W = 88$. Figure 1A shows the COVID-19 attributable weekly deaths data for the four most populous US states, California, Florida, New York and Texas.

2.2 COVID-19 vaccination data

The CDC reports weekly time series of the proportions of individuals aged 18–64 and 65+ who are fully vaccinated (Centers for Disease Control and Prevention, 2021b), referred to as vaccination rate in the following. Fully vaccinated individuals are defined as having received the second dose of a two-dose vaccine or one dose of a single-dose vaccine. We observe records from March 02, 2021 to January 01, 2022 and denote by $v_{m,c,w}$ the vaccination rate for state m in age group $c \in \mathcal{C}$ at the start of week w ,

$$\mathcal{C} = \{18-64, 65+\}. \quad (2.3)$$

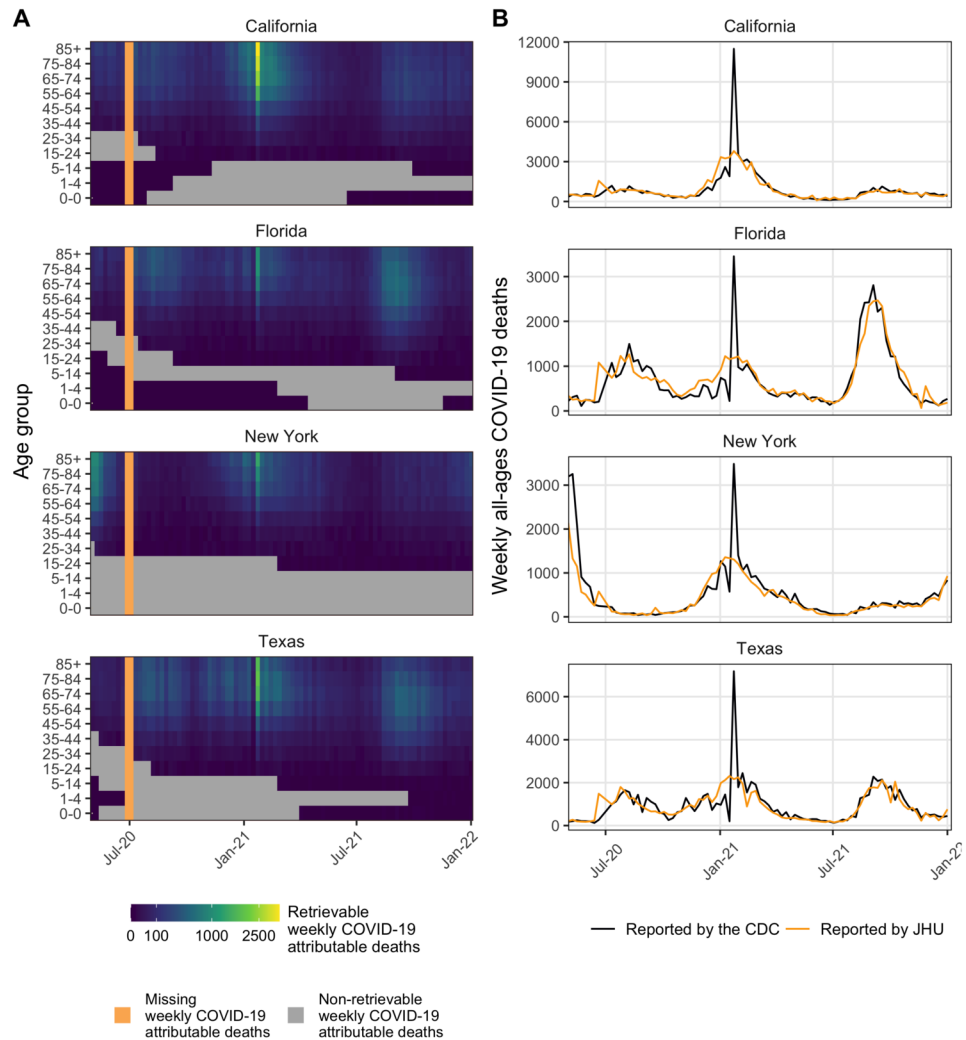


Figure 1: Age-specific weekly COVID-19 attributable deaths in the four most populous US states. (A) The retrievable weekly deaths are computed using the first order difference of reported cumulative deaths by the CDC (2021c). Grey cells indicate non-retrievable weekly deaths due to censored cumulative deaths. Orange bars indicate missing weekly deaths due to non-reported cumulative deaths. **(B)** Weekly all-ages COVID-19 attributable deaths reported by the CDC (black) and by JHU (2020) (yellow). The all-ages weekly COVID-19 attributable deaths is obtained from the CDC data by summing the retrievable weekly COVID-19 attributable deaths across ages. The JHU data are reported regardless of age and are corrected for under/over reporting via now-casting.

2.3 Other data sets

We use two additional data sets. Firstly, we retrieved the weekly COVID-19 attributable deaths regardless of age from John Hopkins University (JHU) for all US states from May 2, 2020 to January 01, 2022 (John Hopkins University of Medicine, 2020), which we denote by $d_{m,w}^{\text{JHU}}$ for state m and week w . In these data, reporting-delayed deaths are back-distributed where possible (this correction is also known as “now-casting”), and for this reason we use the overall deaths as a reference to mitigate the reporting delays in the age-stratified CDC data (2.2). Figure 1B illustrates the corrections applied on the JHU data for under/over reporting via now-casting by comparing the all-ages weekly COVID-19 attributable deaths reported by the JHU and the CDC. Secondly, age-specific weekly COVID-19 attributable deaths are also reported by each state’s Departments of Health (DoH) on their websites or data repositories. For 44 US states data records are available up to April 1, 2021 (Imperial College London COVID-19 Response Team, 2020). There are no records for Montana, Nebraska, New York, Ohio, South Dakota and West Virginia. The DoH do not censor low death counts, and the all-ages sum of the DoH data correspond well to JHU data. This prompts us to use the DoH data as an independent data set to assess the accuracy of our model estimates that are derived from the CDC data.

3 Methods

3.1 Modelling weekly COVID-19 attributable deaths by age

For simplicity we suppress the state index in what follows, with all equations being analogous. Our aim is to estimate the weekly deaths by 1-year age bands, $a \in \mathcal{A} = \{0, 1, \dots, 104, 105\}$ in week $w \in \mathcal{W} = \{1, \dots, W\}$, and we denote their expectation by $\mu_{a,w}$. We first decompose $\mu_{a,w}$ as the product of the weekly deaths for all ages λ_w with the relative contribution $\pi_{a,w}$ of age a to weekly deaths, where $\sum_a \pi_{a,w} = 1$ for all w . Second, we model the longitudinal age composition of deaths $\pi_{a,w}$ as a bivariate random function $(a, w) \rightarrow f(a, w) \in \mathbb{R}$, that is exponentiated, normalised, and evaluated on the 2D grid $\mathcal{A} \times \mathcal{W}$. Our basic latent model structure is thus

$$\begin{aligned} \mu_{a,w} &= \lambda_w \pi_{a,w} \\ \pi_{a,w} &= \text{softmax}\left([f(a, w)]_{a \in \mathcal{A}}\right) \\ &= \frac{\exp f(a, w)}{\sum_{\tilde{a} \in \mathcal{A}} \exp f(\tilde{a}, w)}. \end{aligned} \tag{3.1}$$

To link the expected weekly deaths by 1-year age band, $\mu_{a,w}$, to the data (2.2), we aggregate them over the age groups specified by the CDC, $\mu_{b,w} = \sum_{a \in b} \mu_{a,w}$, for all $b \in \mathcal{B}$. Then, we model the observed, weekly deaths in age b and week w for all ages and weeks through Negative Binomial distributions in the shape-scale parameterisation,

$$d_{b,w} \mid \alpha_{b,w}, \theta \sim \text{NegBin}(\alpha_{b,w}, \theta) \tag{3.2a}$$

$$\mu_{b,w} = \alpha_{b,w} \frac{\theta}{1 - \theta} \tag{3.2b}$$

$$\sigma_{b,w}^2 = \alpha_{b,w} \frac{\theta}{(1-\theta)^2} = \mu_{b,w}(1-\theta)^{-1} = \mu_{b,w}(1+\nu), \quad (3.2c)$$

with mean $\mu_{b,w}$ and variance $\sigma_{b,w}^2$, and where $\nu > 0$ is interpreted as an overdispersion parameter. The purpose of the shape-scale parameterisation with identical θ (3.2a) is that then, the weekly deaths conditional on their total follow a Dirichlet-Multinomial distribution with parameters $\alpha_{b,w}$, resulting in the succinct identity $\alpha_{b,w} = \sum_{a \in \mathcal{B}} \alpha_{a,w}$ (Townes, 2020). For implementation, notice that the shape and scale parameters can be rewritten as $\alpha_{b,w} = \mu_{b,w}/\nu$ and $\theta = \nu/(1+\nu)$.

Our Bayesian model comprises as log likelihood the sum of the log Negative Binomial densities (3.2a) over retrievable weekly deaths parameterised by $\mu_{b,w}$ and ν . Considering data censoring, similar terms involving log cumulative density functions of the Negative Binomial that bound sums of the non-retrievable weekly deaths are added in the log likelihood. Taken together, for the collection of weekly deaths $\mathbf{d} = (d_{b,w}, b \in \mathcal{B}, w \in \{\mathcal{W}_b^{\text{WR}}, \mathcal{W}_b^{\text{WNR}}\})$, our log posterior density is

$$\log p(\boldsymbol{\lambda}, \nu, \mathbf{f}, \boldsymbol{\phi} \mid \mathbf{d}) \propto \sum_{b \in \mathcal{B}} \sum_{w \in \mathcal{W}_b^{\text{WR}}} \log \text{NegBin}(d_{b,w} \mid \lambda_w \pi_{b,w}/\nu, \frac{\nu}{1+\nu}) \quad (3.3a)$$

$$+ \log p(\{d_{b,w}\}_{w \in \mathcal{W}_b^{\text{WNR}}} \mid \dots) \quad (3.3b)$$

$$+ \sum_{w \in \mathcal{W}} \log \text{Exp}(\lambda_w \mid 1/T_w) \quad (3.3c)$$

$$+ \log \text{Exp}(\nu \mid 1) + \log p(\mathbf{f} \mid \boldsymbol{\phi}) + \log p(\boldsymbol{\phi}), \quad (3.3d)$$

where for brevity the term $p(\{d_{b,w}\}_{w \in \mathcal{W}_b^{\text{WNR}}} \mid \dots)$ in (3.3b) is detailed in Supplementary Text, Section S2.2, and the $\pi_{b,w}$ depend on the bivariate random function \mathbf{f} . Throughout we specify weakly informative priors as follows. First, we opt in (3.3c) for a wide Exponential prior on the total number of weekly deaths λ_w . The prior expectation on the total weekly deaths is found by summing the retrievable deaths, such that $T_w = \sum_{b \in \mathcal{B}, w \in \mathcal{W}_b^{\text{WR}}} d_{b,w}$. Second, the overdispersion parameter is given an Exponential prior that peaks at zero and thus favors the least complex model scenario with no overdispersion. The model is completed with a prior on the random function \mathbf{f} and its hyperparameters $\boldsymbol{\phi}$, which we investigate in detail in the next section. Throughout, we checked that priors are only weakly informative, and also considered alternative prior specifications such as Normal prior densities on λ_w or more heavy-tailed priors on the overdispersion coefficient ν , which resulted in posterior inferences that are largely equivalent to the specification in (3.3) (Supplementary Text Section S4).

Considering reporting delays in the data, we explicitly allow for a rescaling of the sum of deaths across age groups in the model according to other curated data sets, which effectively re-distributes reporting delayed deaths in the CDC data set to earlier dates. Specifically, we require that our posterior predictions of the COVID-19 attributable deaths in age group a and week w , $d_{a,w}^*$, sum to $\sum_a d_{a,w}^* = d_w^{\text{JHU}}$. We achieve this by exploiting the probabilistic relationship between Negative Binomial distributions in the

shape-scale parameterisation and the Dirichlet-Multinomial, through

$$p(d_{\cdot,w}^* | \mathbf{d}, d_w^{\text{JHU}}) = \int \text{DirMult}(d_{\cdot,w}^* | d_w^{\text{JHU}}, (\alpha_{a,w})_{a \in \mathcal{A}}) p(\boldsymbol{\lambda}, \nu, \mathbf{f}, \boldsymbol{\phi} | \mathbf{d}) d(\boldsymbol{\lambda}, \nu, \mathbf{f}, \boldsymbol{\phi}). \quad (3.4)$$

We denote by $\mu_{a,w}^*$ the expected predictive COVID-19 attributable deaths in age group a at week w ,

$$\mu_{a,w}^* | \mathbf{d}, d_w^{\text{JHU}} = \mathbb{E}[d_{a,w}^* | \mathbf{d}, d_w^{\text{JHU}}]. \quad (3.5)$$

3.2 Modelling age-specific contributions to COVID-19 weekly deaths

In (3.1), we introduced a 2D function $f(a, w)$, for which we shall find a prior in this section. Let the number of points of the age axis be $n = |\mathcal{A}|$ and on the week axis be $m = |\mathcal{W}|$. The total number of points on the grid is $N = n \times m$. The ensemble of pairs of points is $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_N) = \mathcal{A} \times \mathcal{W}$. We now investigate different modelling approaches for the function $f(a, w)$.

Two-dimensional Gaussian Process

Given observations $(\mathbf{X}, \mathbf{f}) = \{(\mathbf{x}_1, f(\mathbf{x}_1)), \dots, (\mathbf{x}_N, f(\mathbf{x}_N))\}$, we start by considering as model of \mathbf{f} a zero-mean 2D GP,

$$\mathbf{f} | \boldsymbol{\phi} \sim \mathcal{GP}(\mathbf{0}, \mathbf{K}). \quad (3.6)$$

The covariance matrix \mathbf{K} is evaluated at all pairs of points in \mathbf{X} and has entries $\mathbf{K}_{\mathbf{x}, \mathbf{x}'} = \text{Cov}(f(\mathbf{x}), f(\mathbf{x}')) = k(\mathbf{x}, \mathbf{x}')$ with $\mathbf{x}, \mathbf{x}' \in \mathbf{X}$, where $k(\cdot, \cdot)$ is a kernel function. For computational efficiency and because our output is on a multidimensional grid, we decompose the kernel function,

$$k((a, w), (a', w')) = k^1(a, a') k^2(w, w'), \quad (3.7)$$

where $k^1(\cdot, \cdot)$ and $k^2(\cdot, \cdot)$ are kernel functions over ages and weeks, respectively. The corresponding covariance matrix \mathbf{K} is calculated with the Kronecker product $\mathbf{K} = \mathbf{K}^2 \otimes \mathbf{K}^1$, and the number of operations to evaluate the covariance matrix reduces from $\mathcal{O}(N^3)$ to $\mathcal{O}(2N^{3/2})$ (Gönen and Alpaydin, 2011; Saatçi, 2011). We note that the order of the product follows from the fact that the matrix's entries are stacked columnwise. For implementation, we rely on the efficient Kronecker product implementation via matrix-vector products proposed in the Supplementary Material, Section 4 of Wilson et al. (2014). In our applications, we set k^1 and k^2 to be squared exponential kernel functions with variance scale ζ^2 . The scale parameter ζ is attributed a vague prior $\zeta \sim \text{Cauchy}^+(0, 1)$, and even heavier tailed alternatives lead to stable numerical performance in our setting (Supplementary Text Section S4). Specific lengthscales for the rows and the columns, γ_1 and γ_2 , allow for different function's frequency on each dimension.

Considering the length scales, it is commonly recommended to place non-negligible mass on both high-frequency and low-frequency functions, via $\gamma_i \sim \text{Inv-Gamma}(5, 5)$ for $i = 1, 2$ (Stan Development Team, 2020b, chap. 10.3). However, in our setting we found that the corresponding posterior medians are not necessarily within the 95% interquantile range of the prior, and we opted for the heavier tailed $\gamma_i \sim \text{Inv-Gamma}(2, 2)$ for $i = 1, 2$. Under the standard 2D GP prior, $\phi = \{\zeta, \gamma_1, \gamma_2\}$.

B-splines surface

B-splines basis functions – or, more simply, B-splines – are constructed from polynomial pieces that are joined at certain values over the input space, called knots, and defined by a polynomial degree, d , and a non-decreasing sequence of K knots. Given those, the total number of B-splines is $K + d - 1$. We show how a B-spline is constructed in Supplementary Text, Section S3.1. In this paper, we use cubic B-splines by fixing $d = 3$. A fundamental property of B-splines, that constitutes most of their attractiveness, is that they are smooth. More rigorously, a cubic B-spline defined on strictly increasing knots is piecewise infinitely differentiable between the knots, and of continuity C^2 on the knots (Goldman, 2002, chap. 7). \mathbf{f} can be modelled with a tensor product of B-splines given by

$$f(a, w) = \sum_{i=1}^I \sum_{j=1}^J \beta_{i,j} B_i^1(a) B_j^2(w), \quad (3.8)$$

where $B_i^1(\cdot)$ is the i th B-spline defined on a knot vector of length K_1 over the space \mathcal{A} with $i \in \{1, \dots, I\}$ and $I = K_1 + 2$. And, $B_j^2(\cdot)$ is the j th B-spline defined on a knot vector of length K_2 over the space \mathcal{W} with $j \in \{1, \dots, J\}$ and $J = K_2 + 2$. The ensemble of B-splines form a matrix basis denoted by \mathbf{B}^1 and \mathbf{B}^2 of size $I \times n$ and $J \times m$, respectively. The set of pairs of B-splines indices are denoted by $\mathbf{U} = \{1, \dots, I\} \times \{1, \dots, J\}$. $[\beta_{i,j}]_{(i,j) \in \mathbf{U}}$ is a rectangular set of coefficients. We note that to ensure identifiability of the B-spline parameters, the number of evaluated points on each dimension should be greater than the number of B-spline basis functions, i.e., $I < n$ and $J < m$. In our applications, we obtain standard B-splines surfaces by placing independent standard normal priors on all B-splines coefficients. To keep notation streamlined, notice that there are no hyperparameters ϕ under the standard B-splines surface prior.

Regularised B-splines projected two-dimensional Gaussian Process

Given the B-splines indices and corresponding coefficients $\{(i, j), \beta_{i,j}\}_{(i,j) \in \mathbf{U}}$ we place a 2D GP on the coefficients,

$$\beta \mid \phi \sim \mathcal{GP}(0, \mathbf{K}_\beta). \quad (3.9)$$

The covariance matrix has entries $(\mathbf{K}_\beta)_{\mathbf{u}, \mathbf{u}'} = \text{Cov}(\beta_{\mathbf{u}}, \beta_{\mathbf{u}'}) = k_\beta(\mathbf{u}, \mathbf{u}')$ with $\mathbf{u}, \mathbf{u}' \in \mathbf{U}$, where $k_\beta(\cdot, \cdot)$ is a kernel function depending on unknown hyperparameters ϕ . This

approach is equivalent to directly placing a GP on \mathbf{f} with a covariance matrix projected by B-splines. To show this, we rewrite (3.8) as a matrix calculation, $\mathbf{f} = (\mathbf{B}^1)^T \boldsymbol{\beta} \mathbf{B}^2$, and find its vectorized form

$$\text{vec}(\mathbf{f}) = \text{vec}\left((\mathbf{B}^1)^T \boldsymbol{\beta} \mathbf{B}^2\right) = (\mathbf{B}^2 \otimes \mathbf{B}^1)^T \text{vec}(\boldsymbol{\beta}). \quad (3.10)$$

The linear operator produces functions from \mathbb{R}^{IJ} to \mathbb{R}^{nm} . GPs are closed under linear operations (Papoulis and Pillai, 2002, chap.10), and so \mathbf{f} specified by (3.8–3.9) is the GP

$$\mathbf{f} \mid \boldsymbol{\phi} \sim \mathcal{GP}\left(0, (\mathbf{B}^2 \otimes \mathbf{B}^1)^T \mathbf{K}_{\boldsymbol{\beta}} (\mathbf{B}^2 \otimes \mathbf{B}^1)\right). \quad (3.11)$$

In our applications, we decompose the kernel function $k_{\boldsymbol{\beta}}$ into two kernel functions over the rows and columns of the B-spline coefficients matrix as in (3.7). We use squared exponential kernel functions with variance scale ζ^2 and specific lengthscales for the rows and the columns, γ_1 and γ_2 , allowing for different functional frequencies in each dimension. Our priors on the length scales are independent and put non-negligible mass on both high-frequency and low-frequency functions, $\gamma_i \sim \text{Inv-Gamma}(2, 2)$ for $i = 1, 2$, and $\zeta \sim \text{Cauchy}^+(0, 1)$; see also the remarks in the 2D GP section on these choices. Under the regularised B-splines projected 2D GP prior, $\boldsymbol{\phi} = \{\zeta, \gamma_1, \gamma_2\}$.

The kernel function obtained by projecting a base kernel function with cubic B-splines is C^2 as it can be represented as a linear combination of C^2 functions (proof in Supplementary Text, Section S3.2). Therefore the surface obtained when modelling a surface with the B-spline tensor product (3.11), unlike a standard 2D GP (3.6), has inherent smoothness properties not matter the base kernel functions $k_{\boldsymbol{\beta}}$ used.

We can interpret (3.11) as a regularised splines method. Indeed, existing regularised spline methods such as smoothing splines (O’Sullivan, 1986, 1988) and P-splines (Eilers and Marx, 1996; Eilers et al., 2006) aim to minimise in a frequentist setting the loss function

$$\underbrace{\sum_{a \in \mathcal{A}} \sum_{w \in \mathcal{W}} \left(f(a, w) - \sum_{i=1}^I \sum_{j=1}^J \hat{\beta}_{i,j} B_i^1(a) B_j^2(w) \right)^2}_{\text{data fit}} + \underbrace{P}_{\text{penalty}}, \quad (3.12)$$

where $\hat{\beta}$ are the estimated coefficients and P is a penalty applied on the second derivative of the fitted curve for smoothing splines and on finite differences of adjacent B-splines coefficients for P-splines. Now adopting a Bayesian approach, our aim is to maximise the probability of the posterior parameter conditional upon the data, which is proportional to the likelihood multiplied by the prior. By placing a GP on the B-splines coefficients, the prior of \mathbf{f} modelled with the B-spline tensor product (3.11) can be decomposed in

the same way as (3.12),

$$\begin{aligned} \log p(\text{vec}(\mathbf{f}) \mid \boldsymbol{\phi}) &\propto -\frac{1}{2} \left(\text{vec}(\mathbf{f})^T (\mathbf{B}^2 \otimes \mathbf{B}^1)^T \mathbf{K}_\beta (\mathbf{B}^2 \otimes \mathbf{B}^1) \text{vec}(\mathbf{f}) \right. \\ &\quad \left. + \log \left| (\mathbf{B}^2 \otimes \mathbf{B}^1)^T \mathbf{K}_\beta (\mathbf{B}^2 \otimes \mathbf{B}^1) \right| \right) \\ &\propto -\frac{1}{2} \left(\underbrace{\text{vec}(\mathbf{f})^T (\mathbf{B}^2 \otimes \mathbf{B}^1)^T \mathbf{K}_\beta (\mathbf{B}^2 \otimes \mathbf{B}^1) \text{vec}(\mathbf{f})}_{\text{data fit}} + \underbrace{\log |\mathbf{K}_\beta|}_{\text{penalty}} \right), \end{aligned} \quad (3.13)$$

where $|\mathbf{A}|$ denotes the determinant of matrix \mathbf{A} . Thus, the log determinant of the covariance matrix in (3.9) can be interpreted as a complexity penalty, or regulariser, which comes into play if the kernel has a free parameter to control the model complexity (MacKay, 2003, chap. 28).

Bayesian P-splines

Bayesian P-splines have been developed as an extension of the frequentist P-splines, in Lang and Brezger (2004); Brezger and Lang (2006), and impose a spatial dependency on the B-splines coefficients to avoid overfitting. Bayesian P-splines are obtained by placing the following Gaussian Markov Random Field prior on the B-splines coefficients,

$$\beta_{i,j} \mid \beta_{-(i,j)}, \boldsymbol{\phi} \sim \mathcal{N} \left(\sum_{i'=1}^I \sum_{j'=1}^J \frac{\delta_{(i,j),(i',j')} \beta_{i',j'}}{\delta_{(i,j)}^+}, \frac{\tau^2}{\delta_{(i,j)}^+} \right), \quad (3.14)$$

where $\beta_{-(i,j)}$ is the matrix β without entry (i, j) , $\delta_{(i,j),(i',j')}$ evaluates to one if (i, j) and (i', j') are neighboring coordinates vertically or horizontally, and otherwise to zero, and $\delta_{(i,j)}^+ = \sum_{i'=1}^I \sum_{j'=1}^J \delta_{(i,j),(i',j')}$. Here, τ is a spatially varying precision parameter. This model implies that each $\beta_{i,j}$ is normally distributed with a mean equal to the average of its neighbors. We use the Bayesian P-splines prior in simulation analyses and benchmark comparisons, due to their close structural relationship with B-splines projected GP prior. We implemented the Bayesian P-splines prior as in Morris et al. (2019). The scale parameter τ was as before associated with the heavy-tailed prior $\tau \sim \text{Cauchy}^+(0, 1)$, which already resulted in stable numerical performance in our setting and no further customisation was needed. Under the Bayesian P-splines prior, $\boldsymbol{\phi} = \{\tau\}$.

3.3 Numerical inference

All inferences coupled with different priors on the random surface \mathbf{f} were fitted with RStan version 2.21.0, using an adaptive Hamiltonian Monte Carlo (HMC) sampler (Stan Development Team, 2020a). The code is available at <https://github.com/ImperialCollegeLondon/BSplinesProjectedGPs>, and templated Stan model files are in the Supplementary Text Section S5. Using model (3.1–3.2), 8 HMC chains were run in parallel for 1,500 iterations, of which the first 500 iterations were specified as warm-up. Inferences in simulation analysis and benchmarking were performed by running 3 HMC

chains in parallel for 5,000 iterations (resp. 2,000), of which the first 2,500 (resp. 500) were specified as warm-up.

4 Simulation results

Before application to the CDC data, we compare the performance of the spatial priors defined in Section 3.2 at retrieving the mean surface of 2D count data. In the simulations, we consider observations on a 2D grid $\{0, 0.02 \dots, 0.98, 1\} \times \{0, 0.02 \dots, 0.98, 1\}$, such that there are $N = 50 \times 50$ entries. The observations are generated through a Negative Binomial model using as mean the exponential of a zero-mean 2D GP defined by a squared exponential kernel. The variance scale of the kernel is fixed to 1 and the length scale is varied ($\gamma \in \{0.05, 0.25, 1\}$) to generate weakly, mildly and strongly correlated observations. The training set includes 40% uniformly sampled observations (i.e., number of observations in the training set is 1000). Here, we compare the performance of four models with different prior specifications on the mean surface, a standard 2D GP, a standard B-splines surface, Bayesian P-splines and regularised B-splines projected 2D GP. For the methods using B-splines, the same number of equidistant knots are placed along the two axes.

Figure 2A shows the simulated mean surface in the mild correlation scenario obtained with $\gamma = 0.25$, Figure 2B the simulated observations and Figure 2C the simulated observations included in the training set. Figure 3 presents the estimated mean surface when a standard B-spline surface, Bayesian P-splines, and regularised B-splines projected 2D GP priors are used for inference with different numbers of knots. The mean surface estimated by the standard 2D GP model is shown in Supplementary Figure S1.

We find that the standard 2D GP and the regularised B-splines projected 2D GP priors recover well the simulated true mean surface (Figures 3 and Supplementary Fig-

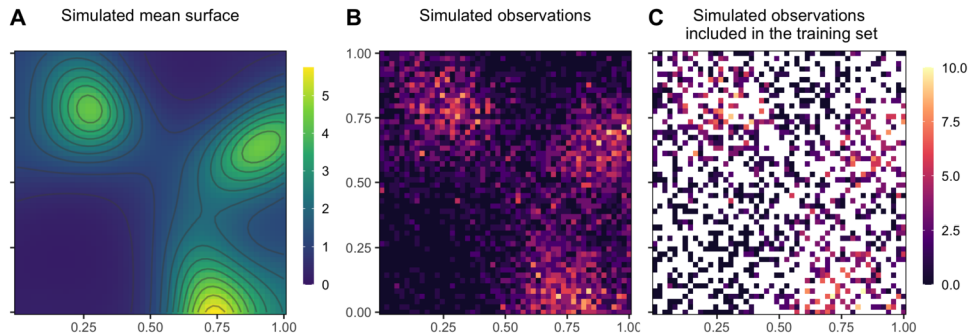


Figure 2: **Mildly correlated 2D simulated observations.** (A) The mean surface is defined by the exponential of a zero-mean 2D GP with a squared exponential kernel implying mild correlations ($\gamma = 0.25$). (B) Count data were simulated from a Negative Binomial model with mean shown in (A) and fixed over-dispersion parameter. (C) 40% of the simulated observations were included in the training set.

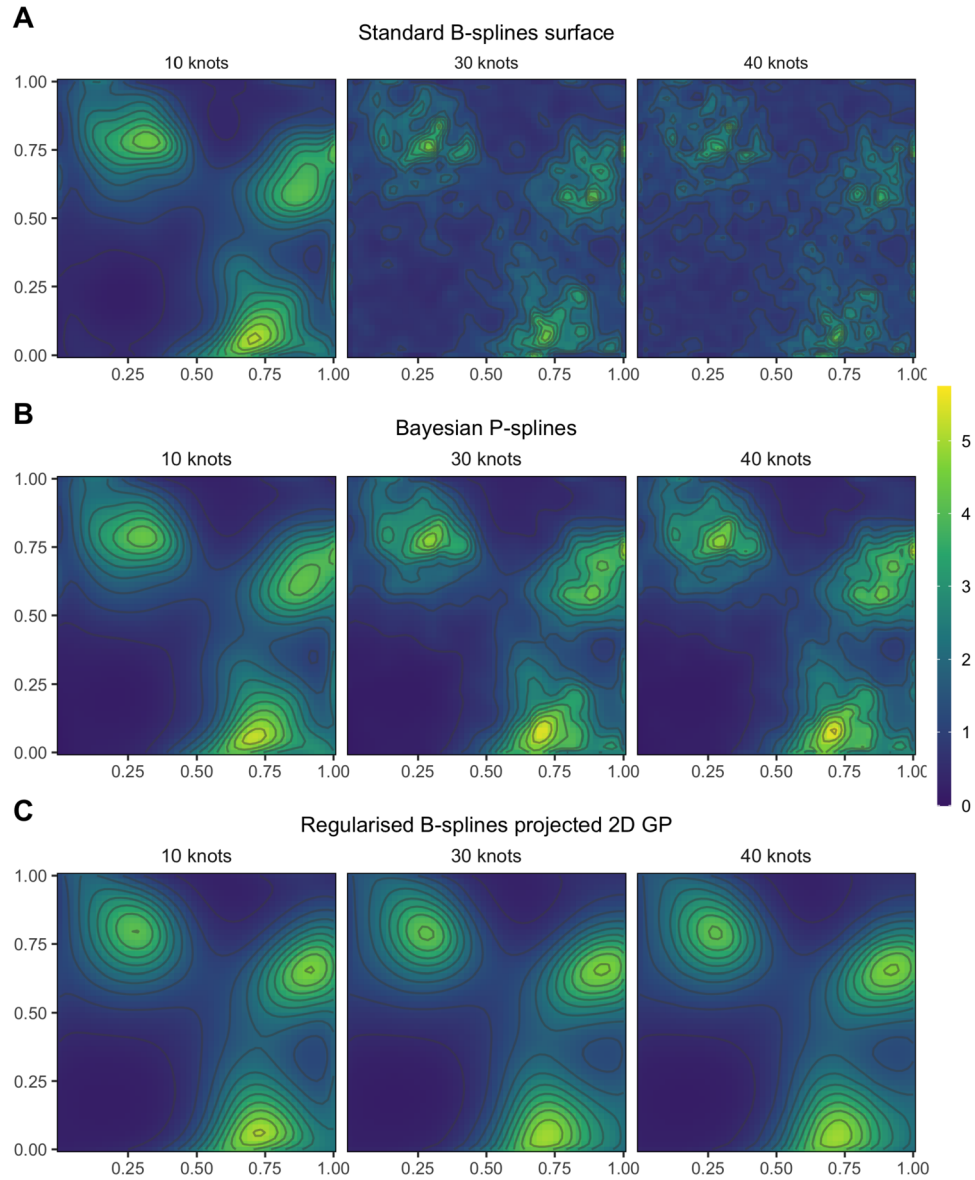


Figure 3: **Estimated mean surfaces using non-regularised and regularised B-splines spatial priors.** Posterior median of the mean surface estimated by (A) standard B-splines prior, (B) Bayesian P-splines prior and (C) regularised B-splines projected 2D GP prior with different number of knots given mildly correlated simulated data ($\gamma = 0.25$). The simulated data are illustrated in Figure 2.

ure S1). Moreover, the regularised B-splines projected 2D GP prior obtains equivalent or better predictive performance as the standard 2D GP for a shorter running time, 70.89% faster on average for 30 knots (Table 1). In contrast, the standard B-splines and the Bayesian P-splines priors overfit the data when the number of knots increase (Figures 3 and Table 1). Similar results were obtained for the other correlation scenarios (Figures S2–S5, Table 1 and Supplementary Table S1).

5 Benchmark results

We next benchmark the performance of the spatial priors defined in Section 3.2 on data unrelated to COVID-19, which has previously been used to benchmark spatial priors. The data set consists of more than 83,000 locations across East Africa with measured deviations in land surface temperatures (Ton et al., 2018). We use the same training data as in Mishra et al. (2020), consisting of 6,000 uniformly sampled locations, and fit them using a Gaussian likelihood with mean surface specified by the various spatial priors, and a free observation variance parameter. We compared in this paper standard 2D B-splines, Bayesian P-splines, and regularised B-splines projected 2D GP, and obtained from Mishra et al. (2020) the performances of a standard 2D GP, low rank Gaussian Markov random field, neural network model and π VAE as a prior for the mean surface. For the spline methods, 125 equidistant knots were placed along both axes. The results are summarised in Supplementary Table S2 and Figure S6. We find that the regularised B-splines projected 2D GP model (testing MSE: 2.96) obtained better predictive performance than the standard B-splines surface model (testing MSE: 4.45), a low rank Gaussian Markov random field (testing MSE: 4.36), and a neural network model (testing MSE: 14.94) and similar predictive performance than Bayesian P-splines (testing MSE: 2.57) and a standard 2D GP (testing MSE: 2.47). It had a worse predictive performance than the π VAE model (testing MSE: 0.38), but training π VAE is not trivial and requires considerable computational runtimes.

6 Age-specific COVID-19 deaths in the United States

Based on the simulation and benchmark results, we use the regularised B-splines projected 2D GP model to reconstruct the time trends in age-specific COVID-19 attributable deaths across the US. The main user input into the model is knot placement. Over the age axis, we placed a knot at the start of each age group reported by the CDC (2.1), and a knot at the largest age considered, amounting to 12 knots at the ages 0, 1, 5, 15, 25, 35, 45, 55, 65, 75, 85, 105. This is the maximum number of relevant knots that one could place because the CDC data are aggregated at the same resolution. Over the time axis, we placed a knot every 6 weeks, from May 2, 2020 to January 01, 2022, amounting to 16 equidistant knots. Sensitivity analyses on knot placement are summarised in Supplementary Table S3.

Overall, fitting the regularised B-splines projected 2D GP model on one state took 1 hour and 30 minutes on a high performance computing environment, with typical effective sample sizes between 1366 to 18836. In comparison, fitting a standard 2D

Method	Simulation scenarios		
	Weakly correlated	Mildly correlated	Strongly correlated
	Mean squared error of the mean surface		
Standard 2D GP	0.08*	0.04*	0.05*
Standard B-splines surface			
Number of knots			
10	0.11	0.15	0.25
30	0.45	0.60	0.90
40	0.67	0.92	1.34
Bayesian P-splines surface			
Number of knots			
10	0.11	0.09	0.10
30	0.14	0.13	0.12
40	0.15	0.13	0.12
Regularised B-splines projected 2D GP			
Number of knots			
10	0.10	0.06	0.06
30	0.09	0.05	0.05*
40	0.08*	0.05	0.05*
	runtime in minutes (−%longest runtime)	runtime in minutes (−%longest runtime)	runtime in minutes (−%longest runtime)
Standard 2D GP	10 (−0.00%)	47 (−0.00%)	21 (−0.00%)
Standard B-splines surface			
Number of knots			
10	1 (−93.85%)	1 (−98.27%)	1 (−96.22%)
30	1 (−88.88%)	1 (−97.53%)	1 (−93.96%)
40	1 (−87.33%)	1 (−96.84%)	1 (−93.17%)
Bayesian P-splines			
Number of knots			
10	9 (−11.59%)	7 (−85.66%)	5 (−76.43%)
30	5 (−52.18%)	5 (−90.14%)	4 (−81.73%)
40	6 (−45.93%)	5 (−90.29%)	5 (−74.60%)
Regularised B-splines projected 2D GP			
Number of knots			
10	5 (−53.96%)	6 (−87.50%)	6 (−69.63%)
30	4 (−61.14%)	6 (−87.07%)	7 (−64.47%)
40	6 (−45.46%)	10 (−78.28%)	14 (−29.92%)

* Best predictive performance.

Table 1: **Performance of four spatial priors on simulated data. (Top)** Mean squared errors between the posterior median estimates of the mean surface and the true mean surface evaluated on the test data withheld from fitting. **(Bottom)** Proportional decreases in runtimes compared to the slowest model (for which the proportional decrease is null).

GP required 7 hours and 50 minutes. There were no reported divergences in Stan’s Hamiltonian Monte Carlo algorithm. After model fitting, all prior specifications were checked by comparing their 95% interquartile ranges against those of the corresponding marginal posterior distributions, and sensitivity analyses using alternative priors suggest our inferences are robust under a broad range of prior specifications (Supplementary Text Section S4).

6.1 Predicted number of age-specific COVID-19 attributable deaths

Figure 4 shows the predicted age-specific weekly attributable COVID-19 deaths in the four most populous US states, and Supplementary Figure S7 for the next six most populous US states. The predictions are obtained for each year of age (3.4), and are aggregated into the age bands 0–54, 55–74 and 75+ for direct comparison to the CDC data, shown as black dots. The data and predictions clearly reflect the multiple COVID-19 waves across the four states and high numbers of deaths in the elderly. Importantly, through the Bayesian model, we calibrate the predictions so that the predicted total weekly deaths across all age groups match the curated all-ages JHU data, and uncertainties in the age-specific predictions propagate through the generative model and fluctuations in observed age composition of deaths. The predictions thus resolve key anomalies in the age-specific CDC data relative to the curated JHU data. For example, our model predicts considerable under-reporting in the age-specific deaths in California, Florida, New York (state) and Texas in the CDC data between November 2020 to January 2021. Further, unrealistically high numbers of deaths were reported in the week starting January 23, 2021, and are in our predictions interpreted as delayed data entries and redistributed among previous weeks. Similar observations apply throughout (Supplementary Figure S7).

We validated our predictions of the age profile of COVID-19 attributable deaths using an external data set, i.e. that was not used to inform the model. The external data set includes age-specific COVID-19 deaths reported directly by US states DoH (see Section 2). We computed the empirical age-specific contribution to weekly COVID-19 attributable deaths from the DoH data, and found that overall, 93.05% of the data lied inside the 95% posterior predictive intervals.

We also fitted the model to data with a standard 2D GP smoothing prior, standard 2D B-splines, and Bayesian P-splines on data from one state, Florida. When using a standard 2D B-splines surface and Bayesian P-splines, the estimated 2D surface over ages and weeks was often wiggly with large uncertainty (Supplementary Text, Section S4). Assessing prediction accuracy with the expected log pointwise predictive density, we found that the B-splines projected 2D GP priors outperformed the other smoothing priors (Supplementary Table S4).

6.2 Estimated age profile of COVID-19 attributable deaths

Figure 4 prompts the question if the age profile of COVID-19 attributable deaths differs across states and how it changed over time. Using the Bayesian model, we can

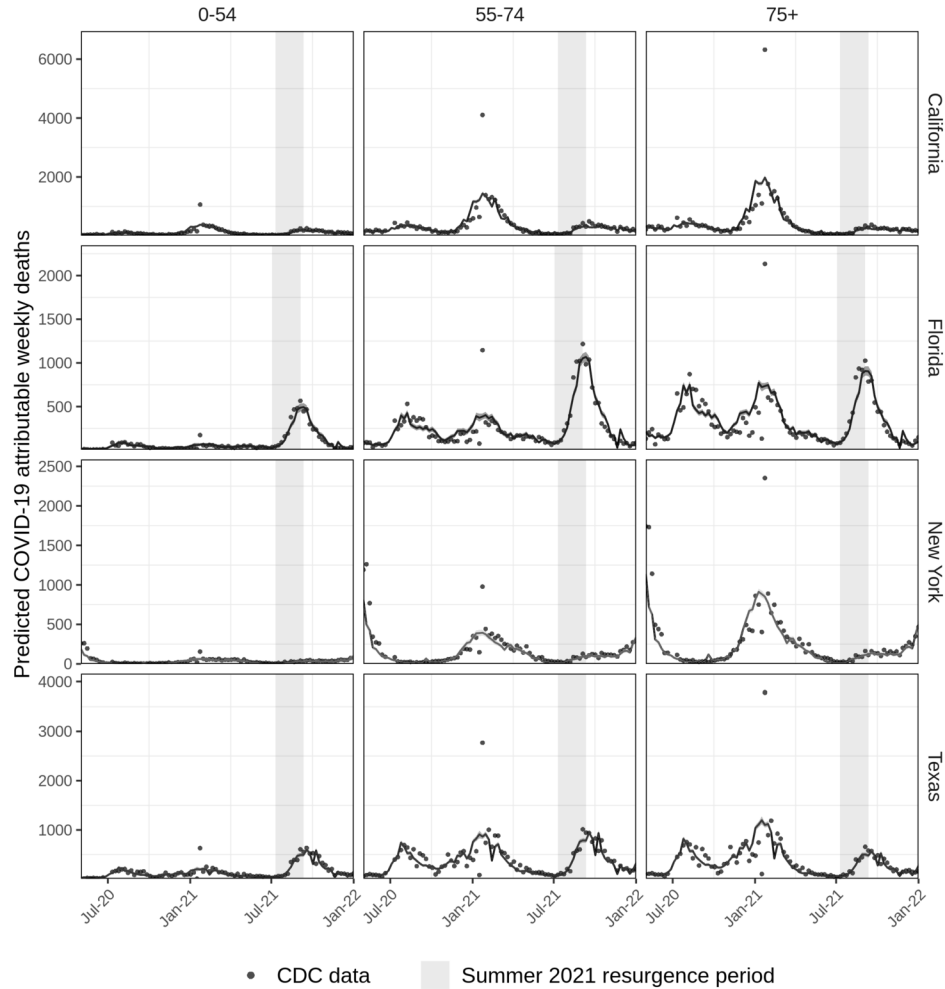


Figure 4: **Predicted age-specific COVID-19 attributable deaths in the four most populous US states.** Shown are the posterior median (line) and 95% credible intervals (ribbon) of the predicted weekly COVID-19 attributable deaths obtained with (3.4). Column facets indicate the age group and row facets indicate the state among which the age-specific COVID-19 attributable deaths were predicted. The CDC data are shown with dots.

provide estimates of the contribution to COVID-19 weekly deaths over time for each year of age and US state through the posterior distributions of $\pi_{a,w}$ in (3.1). We note that such estimates could not have been obtained from the raw CDC data. First, because the data are aggregated by age groups. And second, because the contribution to COVID-19 weekly deaths from age groups for which low death counts were censored is under-estimated and the contribution to COVID-19 attributable deaths from other age

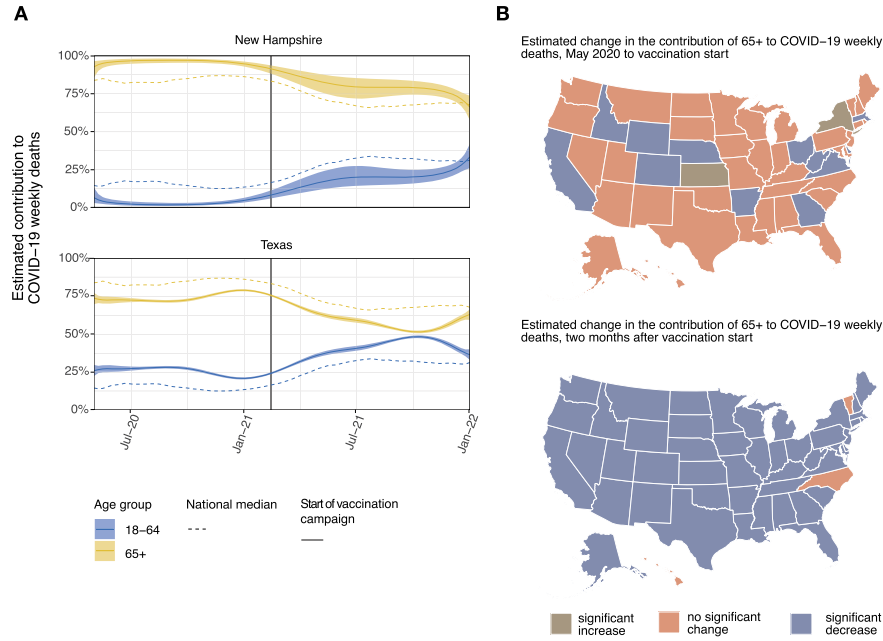


Figure 5: **Estimated age profile of COVID-19 weekly deaths.** (A) Posterior median estimates (line) and 95% credible intervals (ribbon) of the contribution of age groups 18–64 and 65+ to COVID-19 attributable weekly deaths over time (3.1) in New Hampshire and Texas. The dashed line shows the national posterior median estimate. The grey vertical line indicates the start of the COVID-19 vaccination campaign. (B) Estimated change in the contribution of age group 65+ to COVID-19 attributable weekly deaths in the week when COVID-19 vaccination campaigns started relative to the beginning of the COVID-19 pandemic (top), and 8 weeks after the COVID-19 vaccination campaigns started (bottom).

groups is over-estimated. For ease of interpretation, we present the posterior contributions to COVID-19 weekly deaths, $\pi_{a,w}$, aggregated by age groups 18–64 and 65+. We show the inferred age profile of COVID-19 weekly deaths for New Hampshire and Texas in Figure 5A, and for the other states in Supplementary Figure S8. We find substantial variation in the age profile of COVID-19 deaths across states. For example, New Hampshire observed a statistically significantly greater contribution of 65+ to weekly COVID-19 deaths compared to the national median (dotted line), and Texas a statistically significantly lower contribution of 65+ to weekly COVID-19 deaths than the national median. These differences may reflect heterogeneity in the age composition of the populations, the larger burden of COVID-19 in ethnic minority groups (that are also younger), and other factors (Hooper et al., 2020; Wadhera et al., 2020; Millett et al., 2020; Katherine Schaeffer, 2019).

The Bayesian model also allows to investigate how and where the age profile of COVID-19 attributable deaths changed over time, by computing ratios of the form $\pi_{a,w}/\pi_{a,w'}$ for different time points w and w' along with associated posterior uncertainty intervals. Figure 5B shows the states with statistically significant changes in the age profile of COVID-19 attributable deaths before and after the start of COVID-19 vaccination campaigns. We refer to the beginning of the COVID-19 pandemic as the first week of data availability, May 2, 2020, and the start of the COVID-19 vaccination campaigns as the first week when at least 5% of the general population was vaccinated (retrieved from Centers for Disease Control and Prevention, 2021a). Comparing the age profile of deaths at the beginning of the COVID-19 pandemic to the week when COVID-19 vaccination campaigns started, we find that the contribution of 65+ to COVID-19 weekly deaths did not change significantly in the majority of US states (36 states, shown in orange). However 8 weeks after the COVID-19 vaccination campaigns started, the contribution of 65+ to COVID-19 weekly deaths was significantly lower in almost all states when compared to the week in which COVID-19 vaccination campaigns started (shown in blue), illustrating the beneficial population-level effect of prioritising vaccine allocation to higher-risk older age groups.

6.3 Predicted age-specific COVID-19 mortality rates

There is substantial interest in quantifying the total cumulative death toll attributable to COVID-19, and placing the cumulative mortality rates into the wider context of previous global pandemics and other leading causes of death (Cronin and Evans, 2021; Simonsen and Viboud, 2021; Ahmad and Anderson, 2021). Using the Bayesian model, we can predict cumulative COVID-19 mortality rates over time for each year of age and US state by cumulating the weekly $d_{a,w}^*$ in (3.4) over time and dividing by the corresponding population sizes. We used as denominator the 2019 US Census Bureau population size projections (U.S. Census Bureau, 2019). For ease of interpretation, we report cumulative COVID-19 mortality rates aggregated by age groups 25–54, 54–75, 75–84 and 85+ as of January 01, 2022.

The predicted cumulative COVID-19 mortality rates in individuals aged 85+ are shown in Figure 6A, and those for age groups 25–54, 54–75, 75–84 in Supplementary Figure S9. By January 01, 2022, we predict cumulative COVID-19 attributable mortality rates of 32.09‰ [31.86–32.32] in individuals aged 85+, 13.14‰ [13.05–13.23] in individuals aged 75–84, 4.2‰ [4.18–4.22] in individuals aged 55–74, and 0.69‰ [0.69–0.7] in individuals aged 25–54. Figure 6A illustrates the extensive variation in COVID-19 attributable mortality rates across the US. Lamentably, we predict that as of January 01, 2022, the cumulative death rates in individuals aged 85+ exceeded 40‰ in Massachusetts, Rhode Island, Indiana, South Dakota, New Jersey and North Dakota (in decreasing order). Notably, when considering mortality rates in other age groups, other states ranked highest, indicating that factors other than overall disease burden contributed considerably to the high mortality rates in the elderly (Supplementary Figure S9).

To illustrate this point further, Figure 6B shows the predicted cumulative mortality rates in age group 85+ relative to the rate in age group 55–84 in the same state, i.e.

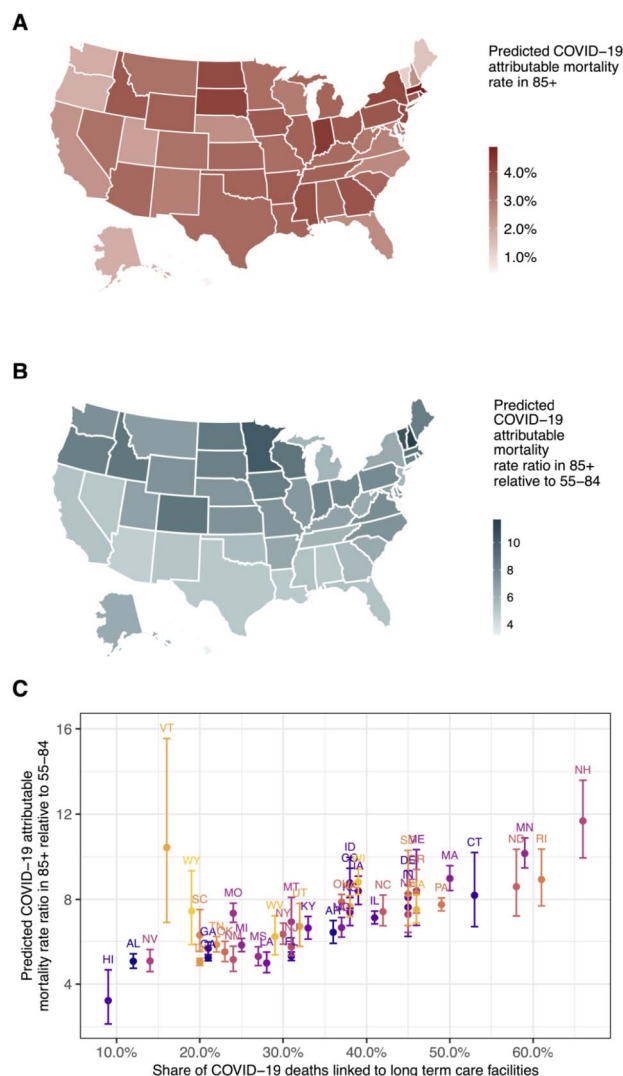


Figure 6: **Predicted COVID-19 attributable cumulative mortality rates.** (A) Posterior median estimates of the predicted COVID-19 attributable mortality rates as of January 01, 2022 in individuals aged 85+. (B) Posterior median estimates of the mortality rate ratio in individuals aged 85+ relative to those aged 55–84 as of June 01, 2021. Note we used a different observation end point here to investigate mortality rate ratios in the context of available data on deaths in nursing homes and other long-term care facilities. (C) Association between the share of COVID-19 attributable deaths linked to nursing homes and other long-term care facilities (The New York Times, 2021) as of June 01, 2021 (x-axis) and posterior median (point) and 95% credible interval (error bar) estimates of the predicted COVID-19 attributable mortality rate ratio in individuals aged 85+ relative to those aged 55–84, again as of June 01, 2021 (y-axis).

the mortality rate ratio of predicted cumulative COVID-19 attributable death. We find the mortality rate ratio ranged from 3.34 [2.27–4.74] in Hawaii to 12.25 [10.47–14.29] in New Hampshire. Intriguingly, we find that the variation in the mortality rate ratio in individuals aged 85+ relative to those aged 55–84 was highly and linearly correlated with the share of COVID-19 deaths linked to nursing homes and other long term care facilities, which we obtained from the New York Times and that were reported up to June 01, 2021 (The New York Times, 2021) (estimated posterior median and 95% credible interval of Pearson correlation coefficient 0.7 [0.49–0.82]). We cannot come to causal conclusions from this univariate analysis, but the estimated strong association effect suggests at least that the observed variation in age-specific mortality rate ratios could be linked to differences in COVID-19 care, interventions, and treatment in nursing homes across the US.

6.4 Variation in the scale of the summer 2021 resurgence

In July 2021, COVID-19 attributable cases and deaths started to rebound (John Hopkins University of Medicine, 2020). As is evident from the empirical data (Figure 1A) and confirmed in our reporting delays-adjusted age-specific predictions (Figure 4), the increase in COVID-19 attributable deaths was highly heterogeneous across US states, with the largest increases seen in Florida, relative to the magnitude of COVID-19 deaths in the previous waves. We hypothesised that the variation in the magnitudes of the summer 2021 resurgence in COVID-19 deaths could be associated with differences in how comprehensively age-specific populations in each state had been vaccinated by that time point. We note that the reporting strata of the weekly deaths (2.1) do not coincide with the reporting strata of the vaccination rates (2.3). Thus, the two data sets cannot be directly compared, and our Bayesian approach is advantageous because we can straightforwardly aggregate predictions by one-year age groups to the stratifications used in both data sets.

To test the hypothesis, we studied the upward trend in COVID-19 attributable deaths from July 01, 2021 over the next 10 weeks. For each state, we define the start of the summer 2021 resurgence as the first week, from July 01 2021, for which a 4-week central moving average on the all-ages weekly COVID-19 deaths started increasing, and which we denote by w_m^{start} . We only evaluated the next 10 weeks during which resurgences peaked to avoid the influence of other factors, and we denote the last week of the summer 2021 resurgence period by w_m^{stop} . The state-specific summer 2021 resurgence period is shown as a shaded area in Figure 4 and Supplementary Figure S7. Next, we considered the relative COVID-19 deaths $r_{m,c,w}$ in state m , age group c , and week w defined as the ratio of the expected posterior predictive weekly deaths in week w relative to its maximum value attained in previous waves before the summer 2021 resurgence period,

$$r_{m,c,w} = \mu_{m,c,w}^* / \mu_{m,c}^{\star, \max}, \quad (6.1)$$

for $w \in \{w_m^{\text{start}}, \dots, w_m^{\text{stop}}\}$ and $c \in \mathcal{C} = \{18-64, 65+\}$, which are the age strata in which the CDC vaccination rates are reported. Here, $\mu_{m,c,w}^*$ are the expected weekly deaths for state m in age group c and week w as in (3.5), and $\mu_{m,c}^{\star, \max} = \max(\{\mu_{m,c,w}^*\}_{w=1, \dots, (w_m^{\text{start}}-1)})$.

By dividing the expected weekly deaths by the maximum value attained on previous waves, the resulting relative COVID-19 deaths in (6.1) are adjusted for state-specific factors that may influence the magnitude of SARS-CoV-2 infections and COVID-19 deaths such as population density and demographic composition. We then formulated a random-effect meta-regression model that uses the CDC vaccine coverage data as predictor for the relative COVID-19 deaths (6.1) in the 50 US states. We defined the pre-resurgence vaccination rates as the proportion of fully vaccinated individuals 14 days before the start of the summer 2021 resurgence period, and denote these for state m and age group c by $v_{m,c}^{\text{pre-resurgence}}$. The meta-regression model is considering data in weeks during the summer resurgence $w \in \{w_m^{\text{start}}, \dots, w_m^{\text{stop}}\}$ and for the age strata $c \in \{18-64, 65+\}$,

$$\begin{aligned}
 r_{m,c,w} &\sim \text{LogNormal}(\xi_{m,c,w}, \kappa_m^2), \\
 \xi_{m,c,w} &= \chi_{m,c} + \psi_{m,c} (w - w_m^{\text{start}}), \\
 \chi_{m,c} &= \chi_c^{\text{base}} + \chi_{m,c}^{\text{state}} + \chi_c^{\text{vacc}} v_{m,c}^{\text{pre-resurgence}} \\
 \psi_{m,c} &= \psi_c^{\text{base}} + \psi_c^{\text{vacc}} v_{m,c}^{\text{pre-resurgence}} \\
 \chi_c^{\text{base}}, \psi_c^{\text{base}} &\sim \mathcal{N}(0, 5^2), \\
 \chi_{c,m}^{\text{state}} &\sim \mathcal{N}(0, \sigma_\chi^2), \\
 \sigma_\chi &\sim \text{Cauchy}^+(0, 1), \\
 \chi_c^{\text{vacc}}, \psi_c^{\text{vacc}} &\sim \mathcal{N}(0, 2.5^2), \\
 \kappa_m &\sim \text{Cauchy}^+(0, 1),
 \end{aligned} \tag{6.2}$$

where $\chi_{m,c}$ correspond to state- and age-specific baseline terms at the start of the summer resurgence, and the $\psi_{m,c}$ capture state- and age-specific growth rates. Here, effects of the pre-resurgence vaccination rates on the baseline and growth rates of the relative deaths are captured respectively by $\chi_{18-64}^{\text{vacc}}$, χ_{65+}^{vacc} and $\psi_{18-64}^{\text{vacc}}$, ψ_{65+}^{vacc} . The random effects on the baseline, $\chi_{m,c}^{\text{state}}$, capture the widely different magnitudes of the epidemic in each state by the start of the summer resurgence. Considering the growth rates, we found that (6.2) already fitted the data very well (Supplementary Figures S10–S11), and so we omitted state-specific random effects on the growth rate in the final model. All prior distributions were specified to be vague, in the sense that the empirical distributions lied well within the prior range.

Figure 7A shows, for illustration, the time evolution of the relative COVID-19 deaths (6.1) among individuals aged 18–64 (left) and 65+ (right) during the summer resurgence for the four most populous states. The corresponding pre-resurgence vaccination rates are shown in colours. Figure 7B shows the relative COVID-19 deaths at the peak of the summer 2021 resurgence wave against the pre-resurgence vaccination rates for the 50 US states. Across the 50 US states, and accounting for uncertainty in the $r_{m,c,w}$ by coupling (3.5) and (6.1–6.2) in the joint Bayesian model, we find that higher pre-resurgence vaccination rates were associated with smaller relative COVID-19 attributable deaths during the summer resurgence (estimated posterior median and 95% credible intervals of $\psi_{18-64}^{\text{vacc}} = -0.44$ $[-0.58, -0.29]$ and $\chi_{65+}^{\text{vacc}} = -2.67$ $[-5.01, -0.33]$; $\chi_{18-64}^{\text{vacc}}$ and ψ_{65+}^{vacc} were negative but not statistically significantly so in term of the 95% posterior credible intervals). A forest plot on effect sizes is shown in Supplementary Figure S12.

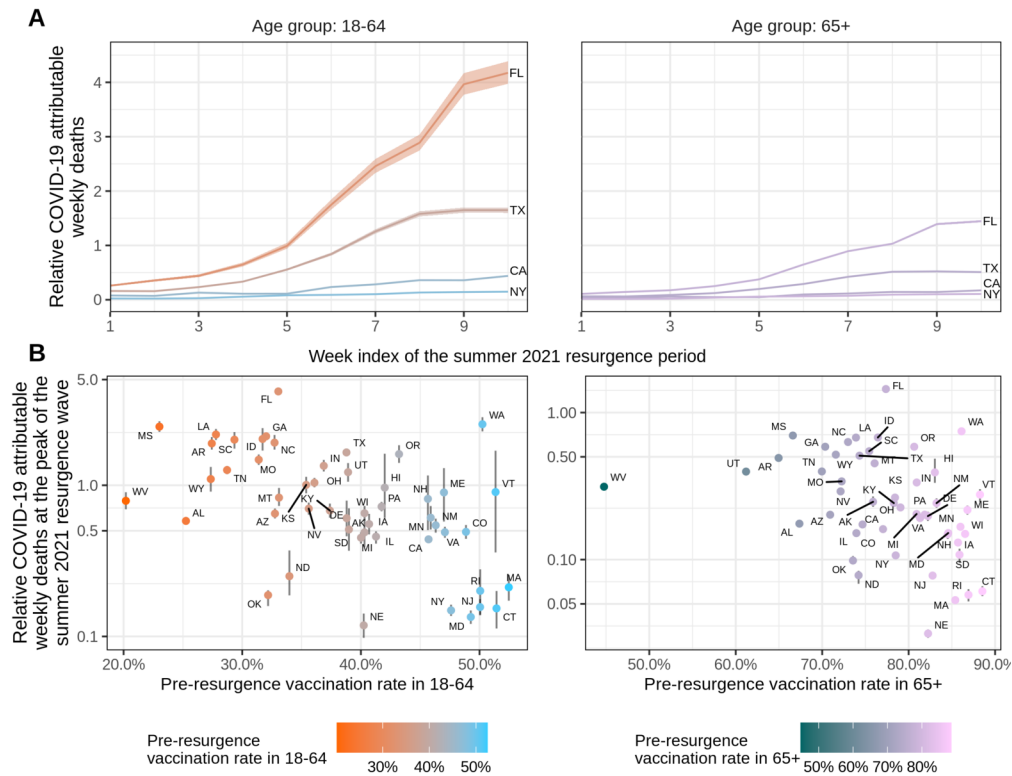


Figure 7: Meta-analysis of resurgent COVID-19 deaths during the summer 2021. (A) Posterior median estimates (line) along with 95% credible intervals (ribbon) of the relative COVID-19 deaths (6.1) over the summer 2021 resurgence period are shown against week indices of the summer 2021 resurgence that correspond to different calendar weeks for each state (see text). Pre-resurgence vaccination rates are shown in colour. Data are shown for the four most populous states for illustration purposes. (B) Posterior median estimates (point) along with 95% credible intervals (error bar) of the relative COVID-19 deaths (6.1) at the peak of the summer 2021 resurgence are shown against the pre-resurgence vaccination rate for the 50 US states. US state abbreviations are shown next to estimates.

6.5 Projected impact of varying vaccine coverage on resurgent COVID-19 deaths

To illustrate the implications of the inferred associations between pre-resurgence vaccine coverage and COVID-19 deaths in summer 2021, we considered four counterfactual analyses. Specifically, we projected the weekly COVID-19 deaths under four counterfactual pre-resurgence vaccine coverage scenarios, assuming that the estimated associations are causal. In the first and second counterfactual projections, the pre-resurgence vaccination rates among 18–64 year olds in all states were fixed respectively to the minimum

and maximum pre-resurgence vaccination rates among 18–64 year olds observed across the 50 US states. In the third and fourth counterfactual projections, the pre-resurgence vaccination rates among 65+ year olds in all states were fixed respectively to the minimum and maximum pre-resurgence vaccination rates among 65+ year olds observed across the 50 US states. The pre-resurgence vaccination rates among 18–64 year olds ranged from 20.19% in West Virginia to 52.46% in Massachusetts. The pre-resurgence vaccination rates among 65+ year olds ranged from 44.76% in West Virginia to 88.53% in Connecticut. We consider these counterfactual scenarios as realistic, as they focus on retrospective evaluations rather than forecasts, and are limited to the observed range of vaccination rates.

Figure 8 summarises the counterfactual projections for all US states. From July 03, 2021 to September 25, 2021 and across the 50 US states, we project that the death toll in 18–64 would have increased by 129.22% [53.51, 271.34] if pre-resurgence vaccination rates among 18–64 in all states had been the same as the minimum pre-resurgence vaccination rate across states (Figure 8A). Furthermore, we project that the death toll in 65+ would have increased by 175.6% [33.3, 499.63] if vaccination rate among 65+ in all states had been the same as the minimum vaccination rate across states (Figure 8B). Conversely, we project that 49.91% [30.81, 64.38] of all deaths in 18–64 were avoidable if pre-resurgence vaccination rates among 18–64 in all states had been the same as the maximum pre-resurgence vaccination rate across states (Figure 8C) and that 31.47% [10.44, 47.7] of deaths in 65+ were avoidable if pre-resurgence vaccination rates among 65+ in all states had been the same as the maximum pre-resurgence vaccination rate across states (Figure 8D). Stratified results by US state are shown in Tables S5–S6.

7 Discussion

We present a Bayesian semi-parametric modelling approach to estimate and report longitudinal trends in COVID-19 attributable deaths by 1-year age bands across the United States. The model is informed by COVID-19 attributable weekly deaths reported by the CDC, the official agency providing such data at the finer spatial resolution in the US. In comparison to the crude CDC data, the predictions are for all weeks, and account for missing and censoring present in the crude data. The predictions of the model also adjust for reporting delays and for over- and under-reporting, by allowing the use of an external dataset for calibration, for which we used the curated JHU all-ages COVID-19 attributable deaths. Importantly, the Bayesian model predicts COVID-19 attributable deaths by 1-year age band. These predictions have broad application, from quantifying the presence and extent of shifts in the age profile of deaths to estimating age-specific mortality rates and uncovering variation in age-specific mortality across states, or characterising the impact of age-prioritized vaccination strategies. Our estimates in Figure 6 and similar should not be misunderstood as a league table, but rather as a tool towards better understanding of the differential impact that COVID-19 had across US states, and as a step to characterize the relative effectiveness of infection control and clinical care measures. The model implementation is portable, freely available and computationally efficient through its implementation in Stan and the use of B-splines projected

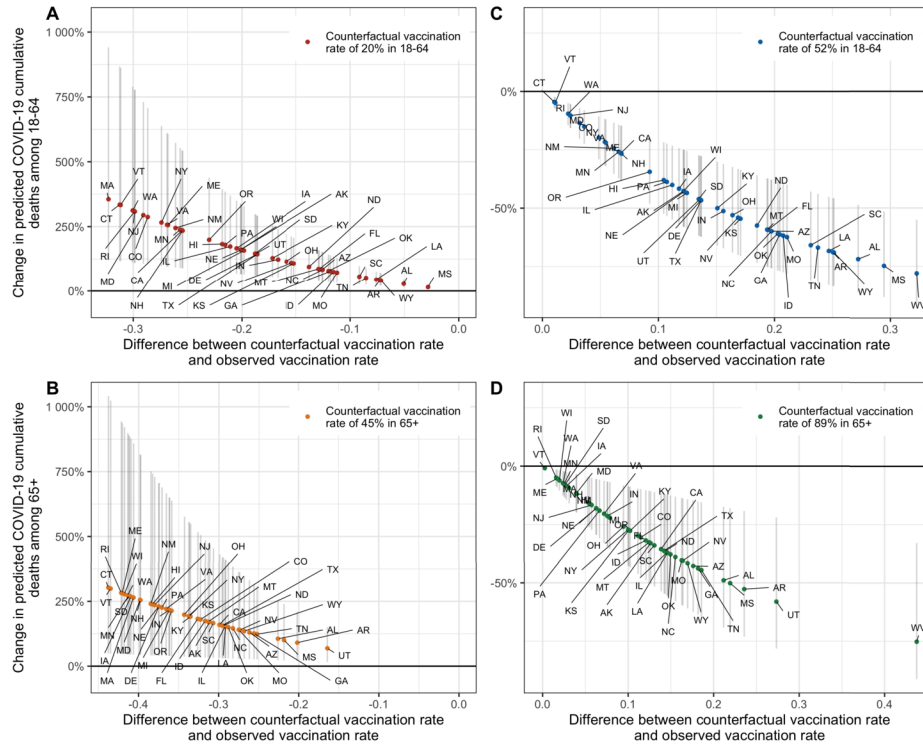


Figure 8: Projected COVID-19 attributable deaths under counterfactual vaccine coverage during the summer 2021 resurgence period. The posterior median estimate (dot) and 95% credible interval (error bars) of the proportional change in the expected posterior predictive cumulative COVID-19 attributable deaths at the end of the summer 2021 resurgence period under the counterfactual scenarios is presented. The x-axis shows the difference from the actual to the counterfactual pre-resurgence vaccination rates assuming that the pre-resurgence vaccination rates among individuals aged 18–64 in all states had been the same as the minimum (A) and maximum (C) pre-resurgence vaccination rate across the 50 US states; that the pre-resurgence vaccination rates among individuals aged 65+ in all states had been the same as the minimum (B) and maximum (D) pre-resurgence vaccination rate across the 50 US states.

Gaussian process priors. For these reasons, we find that the model is a useful tool to improve the reporting of COVID-19 deaths and facilitate demographic and epidemiologic analyses at finer granularity.

Our analyses of the COVID-19 resurgences in the US during the summer 2021 provide further support for the hypothesis that lower vaccine coverages in several states are associated with larger COVID-19 death counts (Sah et al., 2021; Moghadas et al., 2021). However, there are a large number of behavioural, governmental and state-specific differences that modify this effect size. We have here attempted to agnostically account

for these via random effect terms, but it should be noted that we do not explicitly model these factors. Additionally, our study only considered the period of resurgence during the summer 2021. During late 2021 and early 2022, the new SARS-CoV-2 variant of concern Omicron emerged. The epidemiological properties of this variant as well as the change in policy around its control may not allow our model, as is, to generalise well. Applying our model to new variants in different temporal periods will require the relevant context to be taken into account.

We developed B-splines projected Gaussian process priors to describe the time trends in the age profile of COVID-19 deaths through a non-parametric 2D random surface. We placed GP priors on the B-splines coefficients, showed that the model is again a GP, and derived the B-splines projected kernel function of the GP. We showed that the model can be interpreted as regularised B-splines, and find that the negative determinant of the covariance matrix in the log-likelihood, often referred to as Occam's factor (MacKay, 2003, chap. 28), plays the same role as the penalty of smoothing splines and P-splines (O'Sullivan, 1986; Eilers and Marx, 1996). Compared to a standard 2D GP, improvements in runtimes lie both in the fact that fewer parameters on the 2D surface are estimated (from $n \times m$ to $I \times J$), and that the computational complexity in calculating covariance matrices is reduced (from $\mathcal{O}(2(n \times m)^{3/2})$ to $\mathcal{O}(2(I \times J)^{3/2})$), where I and J are the number of B-splines on \mathcal{A} and \mathcal{W} , and n and m are the number of points on \mathcal{A} and \mathcal{W} with $I < n$ and $J < m$. On simulations and on real data, we find the model induces better smoothness properties where these are warranted and avoids over-fitting compared to standard B-splines (d. Boor, 1978) and Bayesian P-splines (Lang and Brezger, 2004) in the settings that we considered. B-splines projected Gaussian process priors are likely an appealing addition to the arsenal of Bayesian regularising priors.

Supplementary Material

Supplementary Material to Regularised B-splines projected Gaussian Process priors to estimate time-trends in age-specific COVID-19 deaths by Monod et al. (DOI: [10.1214/22-BA1334SUPP](https://doi.org/10.1214/22-BA1334SUPP); .pdf). The online Supplementary Material contains the derivation of the likelihood of Section 3.1, the proof of regularised B-splines projected Gaussian Process priors' properties discussed in Section 3.2, the prior sensitivity analyses and the templated Stan model files.

References

- Ahmad, F. B. and Anderson, R. N. (2021). "The leading causes of death in the US for 2020." *JAMA*, 325(18): 1829. [975](#)
- Baden, L. R., Sahly, H. M. E., Essink, B., et al. (2021). "Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine." *New England Journal of Medicine*, 384(5): 403–416. [958](#)
- Blangiardo, M., Cameletti, M., Pirani, M., Corsetti, G., Battaglini, M., and Baio, G. (2020). "Estimating weekly excess mortality at sub-national level in Italy during the COVID-19 pandemic." *PLOS ONE*, 15(10): e0240286. [958](#)

- Brazeau, N. F., Verity, R., Jenks, S., Fu, H., Whittaker, C., Winskill, P., Dorigatti, I., Walker, P., Riley, S., Schnekenberg, R. P., Hoeltgebaum, H., Mellan, T. A., Mishra, S., Whittles, L., Bhatt, S., Ghani, A. C., Ferguson, N. M., and Okell, L. C. (2020). “Report 34 – COVID-19 Infection Fatality Ratio Estimates from Seroprevalence.” *Imperial College London*. 958
- Brezger, A. and Lang, S. (2006). “Generalized structured additive regression based on Bayesian P-splines.” *Computational Statistics & Data Analysis*, 50(4): 967–991. MR2210741. doi: <https://doi.org/10.1016/j.csda.2004.10.011>. 967
- Centers for Disease Control and Prevention (2021a). “COVID-19 Vaccination Trends in the United States, National and Jurisdictional.” <https://data.cdc.gov/Vaccinations/COVID-19-Vaccination-Trends-in-the-United-States-N/rh2h-3yt2>. 958, 975
- Centers for Disease Control and Prevention (2021b). “COVID-19 Vaccinations in the United States, Jurisdiction.” <https://data.cdc.gov/Vaccinations/COVID-19-Vaccinations-in-the-United-States-Jurisdi/unsk-b7fc>. 960
- Centers for Disease Control and Prevention (2021c). “Weekly Updates by Select Demographic and Geographic Characteristics.” https://www.cdc.gov/nchs/nvss/vsrr/covid_weekly/index.htm. 959, 960, 961
- Chen, Y.-C., Lu, P.-E., Chang, C.-S., and Liu, T.-H. (2020). “A time-dependent SIR model for COVID-19 with undetectable infected persons.” *IEEE Transactions on Network Science and Engineering*, 7(4): 3279–3294. MR4205631. 958
- Cronin, C. J. and Evans, W. N. (2021). “Excess mortality from COVID and non-COVID causes in minority populations.” *Proceedings of the National Academy of Sciences*, 118(39): e2101386118. 975
- d. Boor, C. (1978). *A Practical Guide to Splines*. New York: Springer Verlag. MR0507062. 982
- Eilers, P. H., Currie, I. D., and Durbán, M. (2006). “Fast and compact smoothing on large multidimensional grids.” *Computational Statistics & Data Analysis*, 50(1): 61–76. MR2196222. doi: <https://doi.org/10.1016/j.csda.2004.07.008>. 959, 966
- Eilers, P. H. C. and Marx, B. D. (1996). “Flexible smoothing with B-splines and penalties.” *Statistical Science*, 11(2). MR1435485. doi: <https://doi.org/10.1214/ss/1038425655>. 959, 966, 982
- Flaxman, S., Mishra, S., Gandy, A., Unwin, H. J. T., Mellan, T. A., Coupland, H., Whittaker, C., Zhu, H., Berah, T., Eaton, J. W., Monod, M., Ferguson, N. M., Okell, L. C., Bhatt, S., et al. (2020). “Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe.” *Nature*, 584(7820): 257–261. 958
- Goldman, R. (2002). *Pyramid Algorithms*. Morgan Kaufmann. 965
- Gönen, M. and Alpaydin, E. (2011). “Multiple kernel learning algorithms.” *Journal of Machine Learning Research*, 12(64): 2211–2268. MR2825425. doi: <https://doi.org/10.1016/j.patcog.2012.09.002>. 964

- Haas, E. J., Angulo, F. J., McLaughlin, J. M., Anis, E., Singer, S. R., Khan, F., Brooks, N., Smaja, M., Mircus, G., Pan, K., Southern, J., Swerdlow, D. L., Jodar, L., Levy, Y., and Alroy-Preis, S. (2021). “Impact and effectiveness of mRNA BNT162b2 vaccine against SARS-CoV-2 infections and COVID-19 cases, hospitalisations, and deaths following a nationwide vaccination campaign in Israel: an observational study using national surveillance data.” *The Lancet*, 397(10287): 1819–1829. 958
- Hooper, M. W., Nápoles, A. M., and Pérez-Stable, E. J. (2020). “COVID-19 and racial/ethnic disparities.” *JAMA*, 323(24): 2466. 974
- Imperial College London COVID-19 Response Team (2020). “COVID-19 Age specific Mortality Data Repository.” Available at <https://github.com/ImperialCollegeLondon/US-covid19-agespecific-mortality-data>. 962
- Institute for Health Metrics and Evaluation (2020). “Modeling COVID-19 scenarios for the United States.” *Nature Medicine*, 27(1): 94–105. 958
- John Hopkins University of Medicine (2020). “Coronavirus Resource Center.” Available at <https://coronavirus.jhu.edu/region>. 961, 962, 977
- Katherine Schaeffer (2019). “The most common age among whites in U.S. is 58 – more than double that of racial and ethnic minorities.” <https://www.pewresearch.org/fact-tank/2019/07/30/most-common-age-among-us-racial-ethnic-groups/>. 974
- King, A. A., de Cellès, M. D., Magpantay, F. M. G., and Rohani, P. (2015). “Avoidable errors in the modelling of outbreaks of emerging pathogens, with special reference to Ebola.” *Proceedings of the Royal Society B: Biological Sciences*, 282(1806): 20150347. 960
- Lang, S. and Brezger, A. (2004). “Bayesian P-splines.” *Journal of Computational and Graphical Statistics*, 13(1): 183–212. MR2044877. doi: <https://doi.org/10.1198/1061860043010>. 967, 982
- Lavezzo, E., Franchin, E., Ciavarella, C., Cuomo-Dannenburg, G., Barzon, L., Vecchio, C. D., Rossi, L., Manganelli, R., Ferguson, N. M., et al. (2020). “Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo’.” *Nature*, 584(7821): 425–429. 958
- Levin, A. T., Hanage, W. P., Owusu-Boaitey, N., Cochran, K. B., Walsh, S. P., and Meyerowitz-Katz, G. (2020). “Assessing the age specificity of infection fatality rates for COVID-19: systematic review, meta-analysis, and public policy implications.” *European Journal of Epidemiology*, 35(12): 1123–1138. 958
- MacKay, D. J. C. (2003). *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, 1st edition. MR2012999. 967, 982
- Meyerowitz-Katz, G. and Merone, L. (2020). “A systematic review and meta-analysis of published research data on COVID-19 infection fatality rates.” *International Journal of Infectious Diseases*, 101: 138–148. 958
- Millett, G. A., Jones, A. T., Benkeser, D., Baral, S., Mercer, L., Beyrer, C., Honer-

- mann, B., Lankiewicz, E., Mena, L., Crowley, J. S., Sherwood, J., and Sullivan, P. S. (2020). “Assessing differential impacts of COVID-19 on black communities.” *Annals of Epidemiology*, 47: 37–44. 974
- Mishra, S., Flaxman, S., Zhu, H., and Bhatt, S. (2020). “ π VAE: Encoding stochastic process priors with variational autoencoders.” *arXiv:2002.06873*. 970
- Moghadas, S. M., Sah, P., Fitzpatrick, M. C., Shoukat, A., Pandey, A., Vilches, T. N., Singer, B. H., Schneider, E. C., and Galvani, A. P. (2021). “COVID-19 deaths and hospitalizations averted by rapid vaccination rollout in the United States.” *MedRxiv*. 981
- Monod, M., Blenkinsop, A., Xi, X., Hebert, D., Bershan, S., Tietze, S., Baguelin, M., Bradley, V. C., Chen, Y., Coupland, H., Filippi, S., Ish-Horowicz, J., McManus, M., Mellan, T., Gandy, A., Hutchinson, M., Unwin, H. J. T., van Elsland, S. L., Vollmer, M. A. C., Weber, S., Zhu, H., Bezancon, A., Ferguson, N. M., 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. 958
- Monod, M., Blenkinsop, A., Brizzi, A., Chen, Y., Cardoso Correia Perello, C., Jogarah, V., Wang, Y., Flaxman, S., Bhatt, S., and Ratmann, O. (2022). “Supplementary Material for “Regularised B-splines projected Gaussian process priors to estimate time-trends in age-specific COVID-19 deaths”.” *Bayesian Analysis*. doi: <https://doi.org/10.1214/22-BA1334SUPP>. 960
- Morris, M., Wheeler-Martin, K., Simpson, D., Mooney, S. J., Gelman, A., and DiMaggio, C. (2019). “Bayesian hierarchical spatial models: Implementing the Besag York Mollié model in stan.” *Spatial and Spatio-temporal Epidemiology*, 31: 100301. 967
- O’Sullivan, F. (1986). “A statistical perspective on ill-posed inverse problems.” *Statistical Science*, 1(4): 502–518. MR0874480. 959, 966, 982
- O’Sullivan, F. (1988). “Fast computation of fully automated log-density and log-hazard estimators.” *SIAM Journal on Scientific and Statistical Computing*, 9(2): 363–379. MR0930052. doi: <https://doi.org/10.1137/0909024>. 959, 966
- Papoulis, A. and Pillai, S. U. (2002). *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill Higher Education. MR0176501. 966
- Pokharel, G. and Deardon, R. (2021). “Emulation-based inference for spatial infectious disease transmission models incorporating event time uncertainty.” *Scandinavian Journal of Statistics*. MR4391061. doi: <https://doi.org/10.1111/sjos.12523>. 959
- Rasmussen, C. E. and Williams, C. K. I. (2005). *Gaussian Processes for Machine Learning*. The MIT Press. MR2514435. 959
- Rearc (2021). “Provisional COVID-19 Death Counts by Sex, Age, and State | CDC / NCHS.” <https://github.com/rearc-data/covid-19-death-counts-sex-age-state>. 960

- Saatçi, Y. (2011). “Scalable Inference for Structured Gaussian Process Models.” *PhD thesis, University of Cambridge*. 959, 964
- Sah, P., Moghadas, S. M., Vilches, T. N., Shoukat, A., Singer, B. H., Hotez, P. J., Schneider, E. C., and Galvani, A. P. (2021). “Implications of suboptimal COVID-19 vaccination coverage in Florida and Texas.” *The Lancet Infectious Diseases*, 21(11): 1493–1494. 981
- Shah, C., Dehmamy, N., Perra, N., Chinazzi, M., Barab’asi, A. L., Vespignani, A., and Yu, R. (2020). “Finding patient zero: learning contagion source with graph neural networks.” *arXiv:2006.11913*. 959
- Simonsen, L. and Viboud, C. (2021). “A comprehensive look at the COVID-19 pandemic death toll.” *eLife*, 10. 975
- Stan Development Team (2020a). “RStan: the R interface to Stan, R package version 2.21.2.” Available at <http://mc-stan.org/>. 967
- Stan Development Team (2020b). “Stan User’s Guide.” Available at https://mc-stan.org/docs/2_29/stan-users-guide/index.html, last accessed 23rd February 2022. 965
- The New York Times (2021). “Nearly One-Third of U.S. Coronavirus Deaths Are Linked to Nursing Homes.” <https://www.nytimes.com/interactive/2020/us/coronavirus-nursing-homes.html>. 976, 977
- Ton, J.-F., Flaxman, S., Sejdinovic, D., and Bhatt, S. (2018). “Spatial mapping with Gaussian processes and nonstationary Fourier features.” *Spatial Statistics*, 28: 59–78. MR3887156. doi: <https://doi.org/10.1016/j.spasta.2018.02.002>. 970
- Townes, F. W. (2020). “Review of probability distributions for modeling count data.” *arXiv:2001.04343*. 963
- U.S. Census Bureau (2019). “Single-race Population Estimates, United States, 2010–2019. July 1st resident population by state, age, sex, single-race, and Hispanic origin, on CDC WONDER Online Database.” Available at <http://wonder.cdc.gov/single-race-single-year-v2019.html>. 975
- Wadhwa, R. K., Wadhwa, P., Gaba, P., Figueroa, J. F., Maddox, K. E. J., Yeh, R. W., and Shen, C. (2020). “Variation in COVID-19 hospitalizations and deaths across New York City boroughs.” *JAMA*, 323(21): 2192. 974
- Wikle, N. B., Tran, T. N.-A., Gentile, B., Leighow, S. M., Albert, E., Strong, E. R., Brinda, K., Inam, H., Yang, F., Hossain, S., Chan, P., Hanage, W. P., Messick, M., Pritchard, J. R., Hanks, E., and Boni, M. F. (2020). “SARS-CoV-2 epidemic after social and economic reopening in three US states reveals shifts in age structure and clinical characteristics.” *MedRxiv*. 958
- Wilson, A. G., Gilboa, E., Nehorai, A., and Cunningham, J. P. (2014). “Fast kernel learning for multidimensional pattern extrapolation.” *Proceedings of the 27th International Conference on Neural Information Processing Systems – Volume 2*, 3626–3634. 959, 964

Zheng, J., Wu, X., Fang, F., Li, J., Wang, Z., Xiao, H., Zhu, J., Pain, C., Linden, P., and Xiang, B. (2021). “Numerical study of COVID-19 spatial-temporal spreading in London.” *Physics of Fluids*, 33(4): 046605. [958](#)

Acknowledgments

We thank Sam Abbott, the editor and two anonymous reviewers for their helpful comments.