Regularised B-splines projected Gaussian Process priors to estimate time-trends in age-specific COVID-19 deaths

Mélodie Monod
CDT Modern Statistics and Machine Learning, Imperial College London
Machine Learning & Global Health Network

ISBA World Meeting
29.06.2022

Imperial College London
MACHINE LEARNING
& GLOBAL HEALTH NETWORK

Road map

- **The regularised B-splines projected Gaussian Process prior**

- Theoretical properties: **smoothness** and **computational efficiency**

- **Benchmark** against standard GP, standard B-splines, Bayesian P-splines, …

- **Applications**
    1. estimating time-trends in age-specific COVID-19 deaths
    2. estimating sex and age-specific flows of HIV-1 transmission
    3. estimating sex and age-specific contact patterns

Regularised B-splines projected Gaussian Process priors

Problem set-up

Let our support be a 2D grid $\mathcal{G}$ with points $(x_1, \ldots, x_n) = \mathcal{X} \subset \mathbb{R}$ on the first dimension and points $(y_1, \ldots, y_m) = \mathcal{Y} \subset \mathbb{R}$ on the second dimension. The grid $\mathcal{G} = \mathcal{X} \times \mathcal{Y}$ is the set of $N = n \times m$ coordinates points.

A multivariate random variable $\boldsymbol{z} = \{z(x, y)\}_{(x,y)\in\mathcal{G}}$ is defined on the 2D grid. Its likelihood takes the form,

$$z(x, y) \mid f(x, y), \theta \sim F\big(f(x, y), \theta\big),$$

where $\boldsymbol{F}$ is a distribution parametrised by a 2D random function $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{y})$ and some parameters $\boldsymbol{\theta}$.

Two-dimensional Gaussian Process

A natural starting point for modelling the random function $\boldsymbol{f}$ is a two-dimension Gaussian Process,

$$\boldsymbol{f} \mid \boldsymbol{\phi} \sim \mathcal{GP}(0, \boldsymbol{K}),$$

where $\boldsymbol{K} = \boldsymbol{K}(\boldsymbol{\phi})$ with entries defined by some kernel function k(.,.).

Because our inputs are on a Cartesian product grid, we can decompose the covariance matrix (Saatçi, 2011, Gonen et al., 2011),

$$\boldsymbol{K} = \boldsymbol{K}_2 \otimes \boldsymbol{K}_1.$$

Advantage: Inherits all the properties of Gaussian Processes

Disadvantage: The time complexity of this approach scales to $\mathcal{O}(2N^{3/2})$.

Two-dimensional B-splines surface

B-splines basis functions are constructed from polynomial pieces that are joined at certain values over the input space, called knots. We use cubic B-splines.

$\boldsymbol{f}$ can be modelled with a tensor product of B-splines given by,

$$f(x,y) = \sum_{i=1}^{I} \sum_{j=1}^{J} \beta_{i,j}\, B_i^1(x)\, B_j^2(y).$$

Where $\boldsymbol{B}^1$ of size $\boldsymbol{I} \times \boldsymbol{n}$ and $\boldsymbol{B}^2$ of size $\boldsymbol{J} \times \boldsymbol{m}$ are matrices of B-splines basis functions defined over $\boldsymbol{\mathcal{X}}$ and over $\boldsymbol{\mathcal{Y}}$.

Advantage: Smooth (piecewise infinitely differentiable between the knots, and of continuity $\boldsymbol{C^2}$ on the knots.

Disadvantage: Need to choose the number of knots. To few knots will not capture complex signals and too many will overfit the surface.

Regularised B-splines

Idea: Use many knots and apply a regularisation on the B-splines parameters to impose smoothness.

- Frequentist: smoothing splines (O'Sullivan, 1986,9) and P-splines (Eilers and Marx, 1996; Eilers et al., 2006).
  → Penalty applied on the second derivative of the fitted curve or on finite differences of adjacent B-splines coefficients.

- Bayesian: Bayesian P-splines (Land and and Brezger (2004, 2006).
  → Conditional Autoregressive prior on the B-splines parameters.

B-splines projected two-dimensional Gaussian Process prior

Regularised two-dimensional B-splines view:

$$f(x, y) = \sum_{i=1}^{I} \sum_{j=1}^{J} \beta_{i,j} \, B_i^1(x) \, B_j^2(y).$$

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

where $\boldsymbol{K}_\beta = \boldsymbol{K}_\beta(\boldsymbol{\phi})$ with entries defined by a kernel function $\boldsymbol{k}_\beta$

Low-rank two-dimensional Gaussian Process view:

$$\boldsymbol{f} \mid \boldsymbol{\phi} \sim \mathcal{GP}\Big(0, \big(\boldsymbol{B}^2 \otimes \boldsymbol{B}^1\big)^T \boldsymbol{K}_\beta \big(\boldsymbol{B}^2 \otimes \boldsymbol{B}^1\big)\Big).$$

B-splines projected two-dimensional Gaussian Process prior

Advantages:

1. The 2D random function $\boldsymbol{f}$ inherits smoothness properties from the B-splines. We show that the kernel function obtained by projecting a base kernel function with cubic B-splines is C^2.

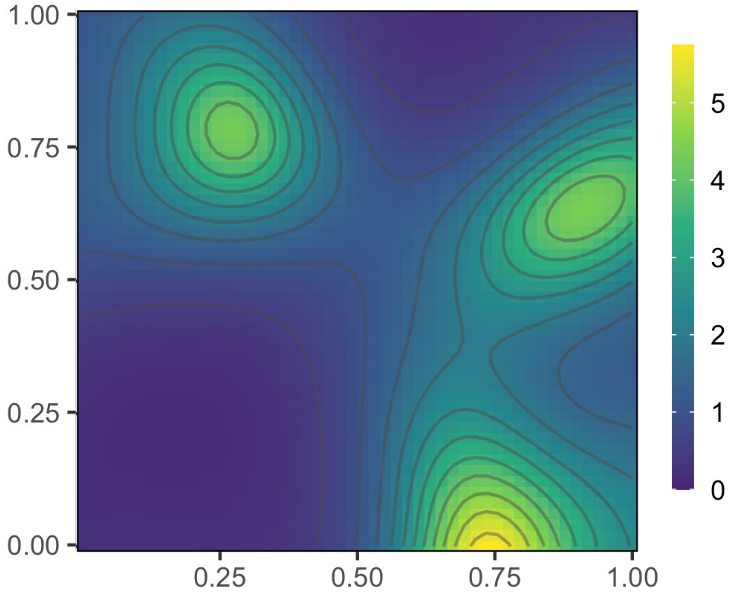
2. The B-splines parameters β are regularised as the negative determinant of the covariance matrix in the log-likelihood plays the same role of a complexity penalty.

3. The time complexity compared to a full rank two-dimensional Gaussian Process reduces from $\mathcal{O}(2N^{3/2})$ to $\mathcal{O}(2(I \times J)^{3/2})$

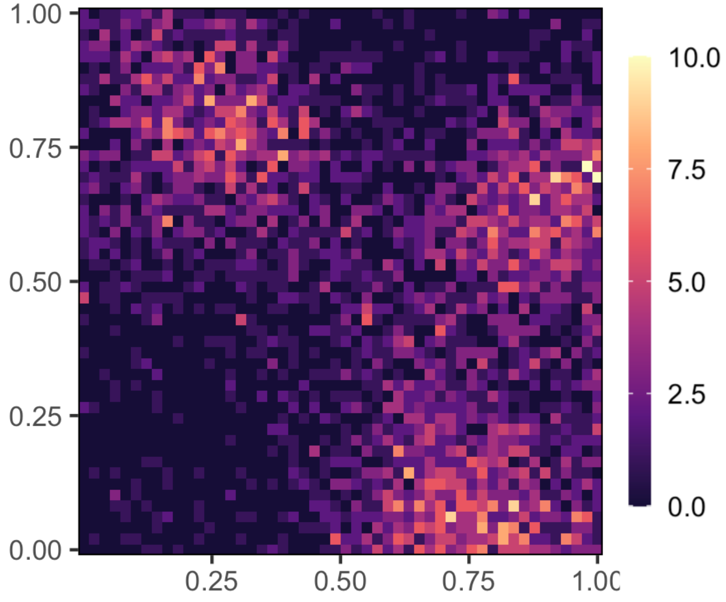
Simulation analysis

Simulated data

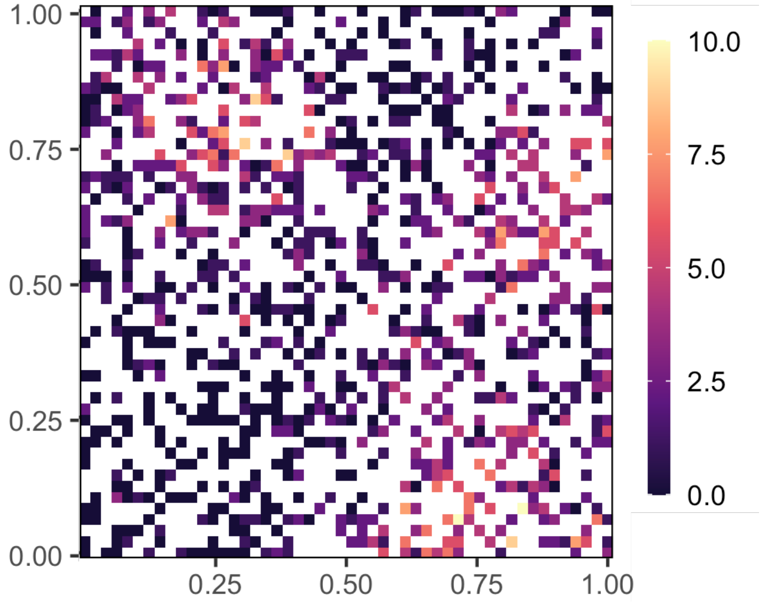
Simulated
mean surface



Simulated
count observations



Simulated
count observations
included in the training set

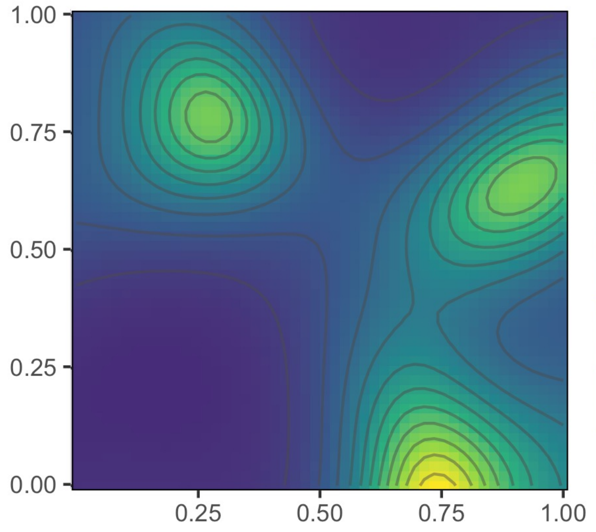


Results

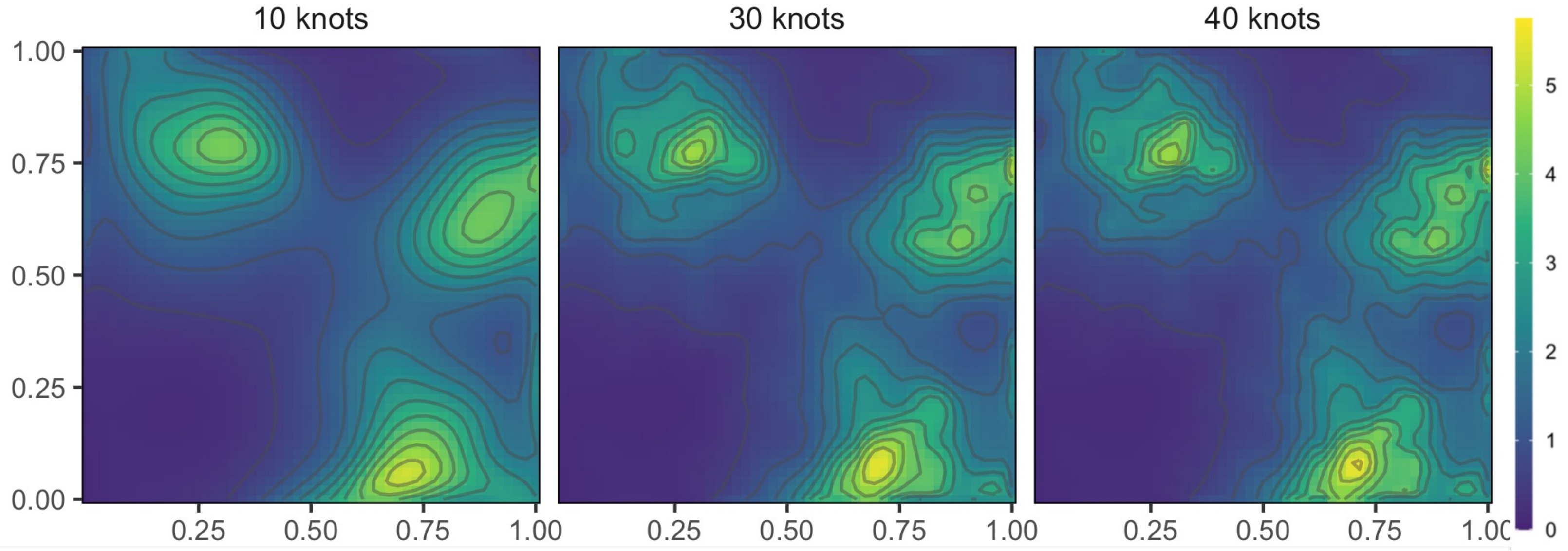
Method	MSE	Runtime in minutes (%reduction runtime)
Standard 2D GP	0.04	47 (0.00%)
Bayesian P-splines		
10 knots	0.09	7 (85.66%)
30 knots	0.13	5 (90.14%)
40 knots	0.13	5 (90.29%)
Regularised B-splines projected 2D GP		
10 knots	0.06	6 (87.50%)
30 knots	0.05	6 (87.07%)
40 knots	0.05	10 (78.28%)

Results

Simulated mean surface

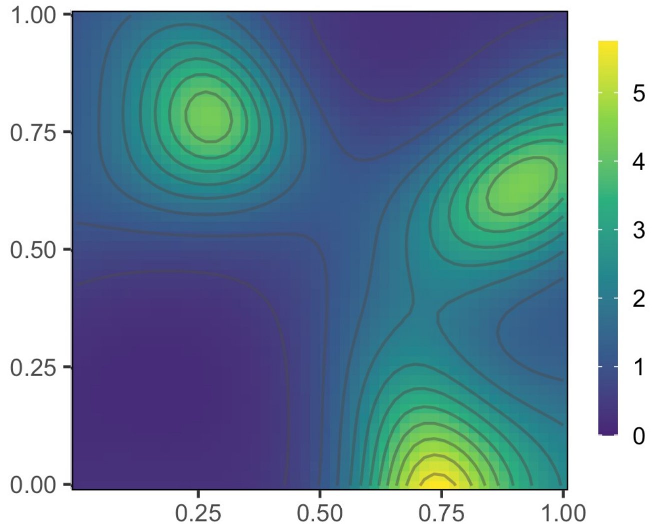


Bayesian P-splines (Lang and Brazer, 2004)

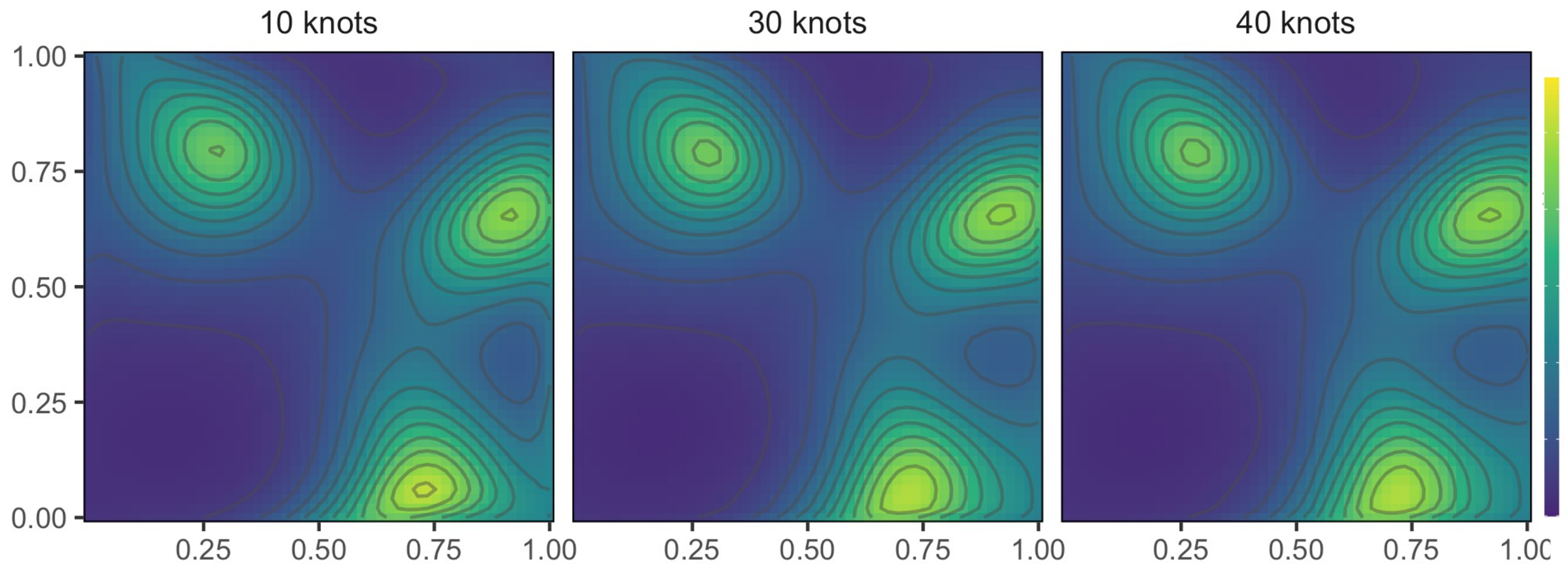


Results

Simulated
mean surface



B-splines projected two-dimensional
Gaussian Process prior



Case study

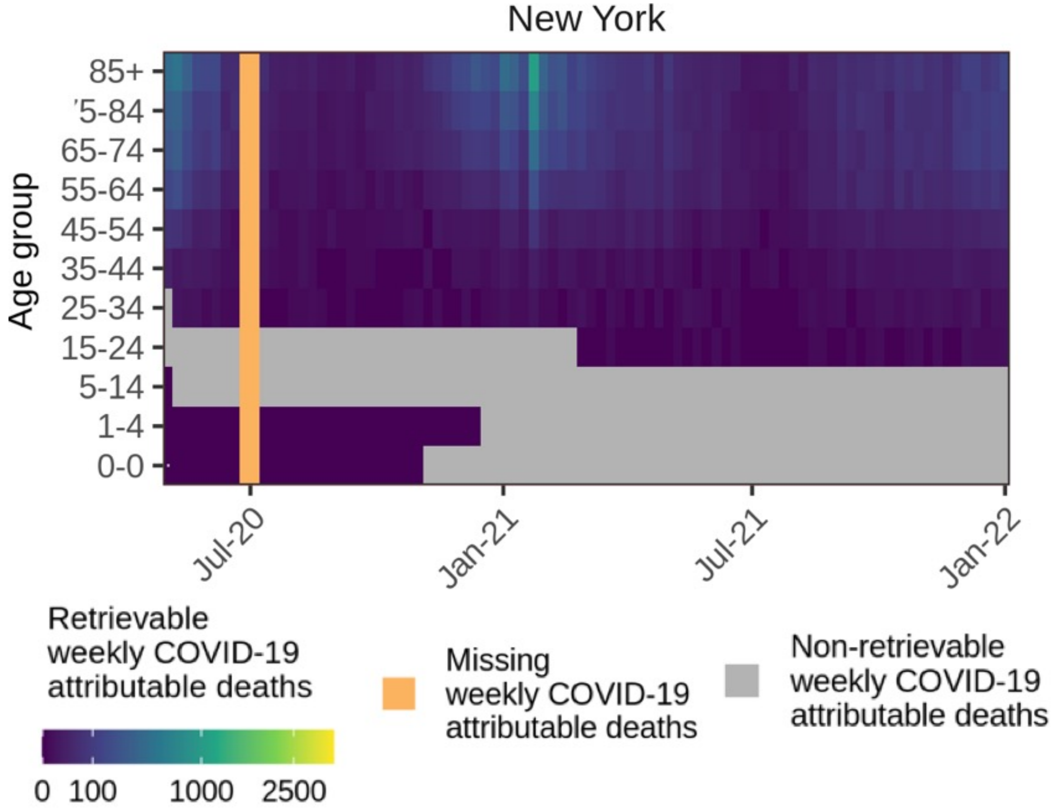
Time-trends in age-specific COVID-19 deaths

Problem and data

- Age and state-specific COVID-19 attributable deaths data over time in the United States are reported by the CDC.

- Those data are partially censored, reported with delays, and reported in age bands $\mathcal{B}$ that may not match those of other data streams.

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



→ we provide methods for estimating high resolution age-specific COVID-19 attributable deaths over time without reporting delays.

Model

We estimate the weekly deaths by 1-year age bands $a \in \mathcal{A} = \{0, 1, \ldots, 104, 105\}$ in week $w \in \mathcal{W} = \{1, \ldots, 88\}$ 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 and the relative contribution $\pi_{a,w}$ with $\sum_a \pi_{a,w} = 1$.

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

Model

To link the expected weekly deaths by 1-year age band, $\mu_{a,w}$ to the data, we aggregate them over the age groups specified by the CDC, $\mu_{b,w} = \sum_{a \in b} \mu_{a,w}$ for all $b \in \mathcal{B}$.

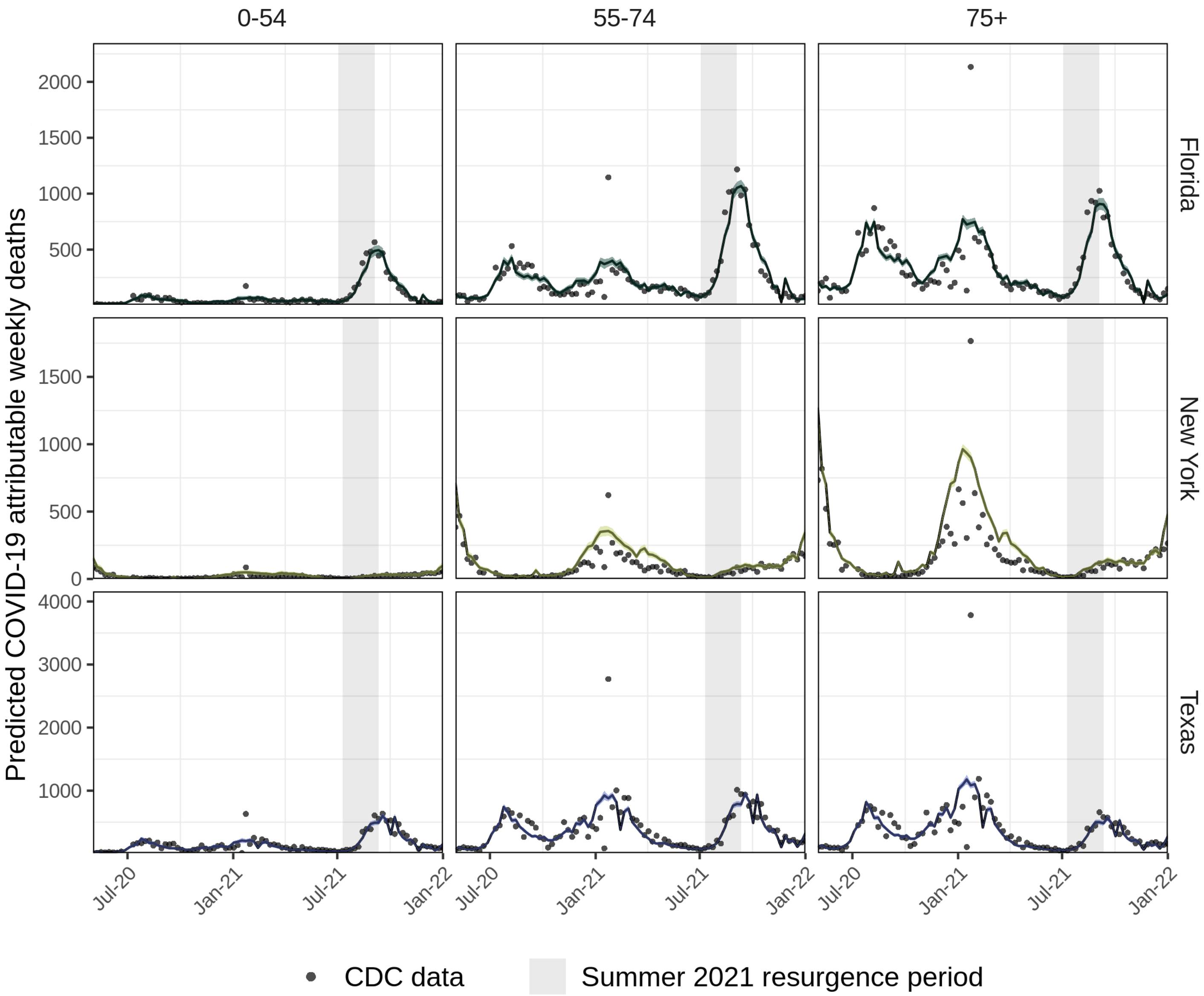
$$d_{b,w} \mid \mu_{b,w}, \theta \sim \mathrm{NegBin}(\mu_{b,w}, \theta)$$

with mean $\mu_{b,w}$ and overdispersion parameter $\theta > 0$.

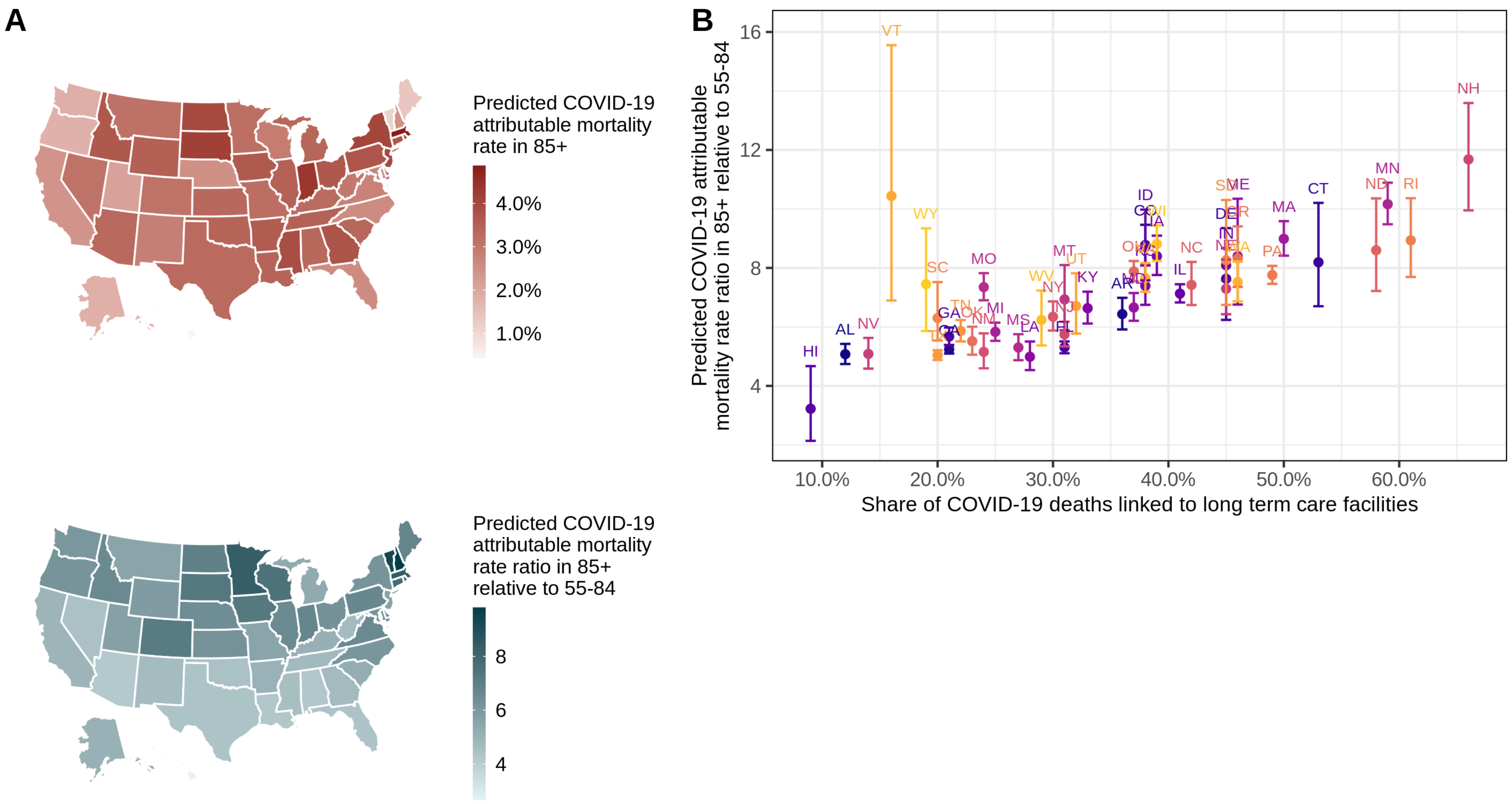
Model

- The JHU data set re-distributes reporting delayed deaths in the CDC data set to earlier dates.
- We require that our posterior predictions of the COVID-19 attributable deaths in age group $\boldsymbol{a}$ and week $\boldsymbol{w}$, $d^{\star}_{a,w}$, sum to $\sum_a d^{\star}_{a,w} = d^{\mathrm{JHU}}_{w}$.

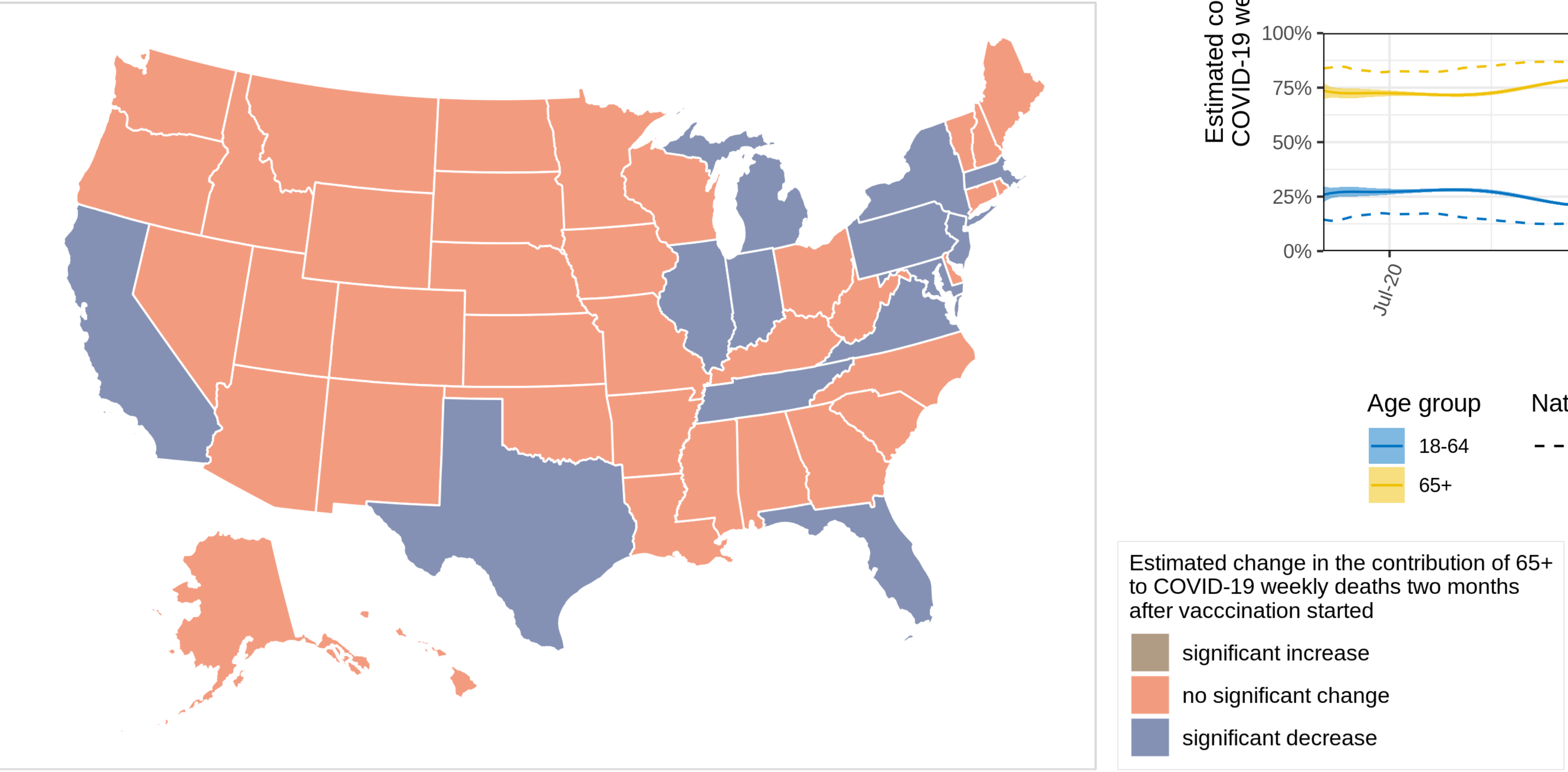
Results: Predicted age-specific COVID-19 attributable deaths over time

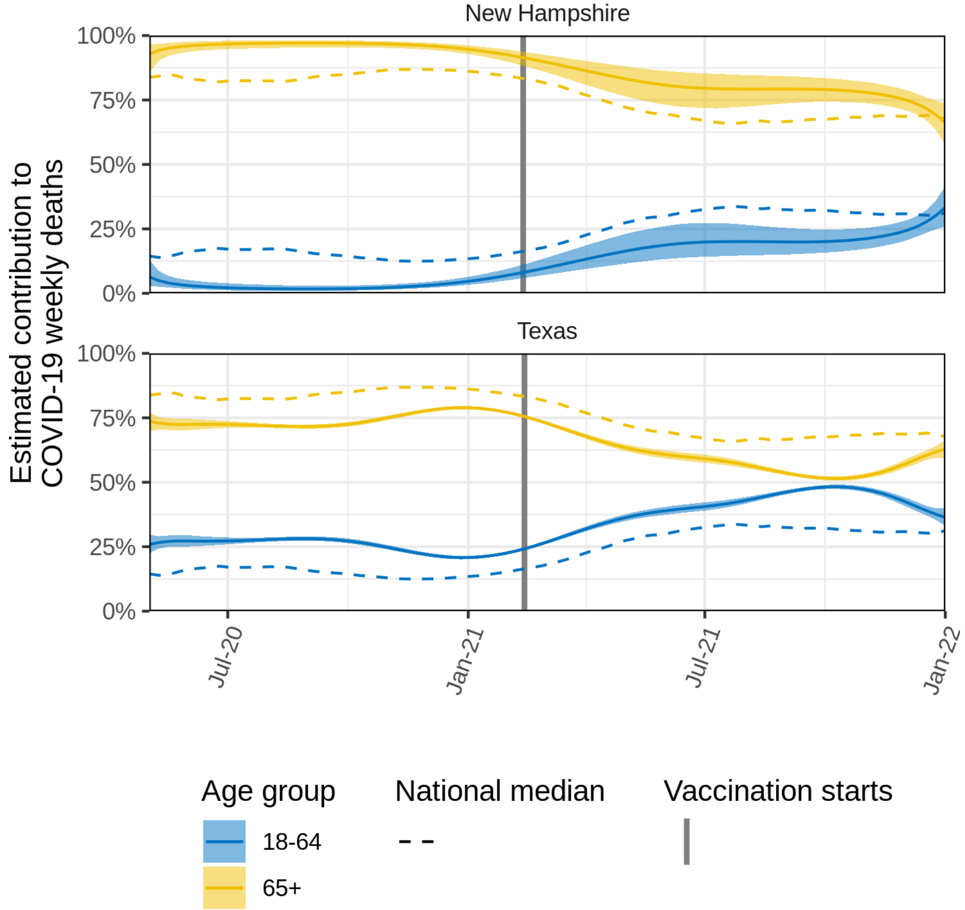


Results: Posterior predictions of COVID-19 attributable mortality



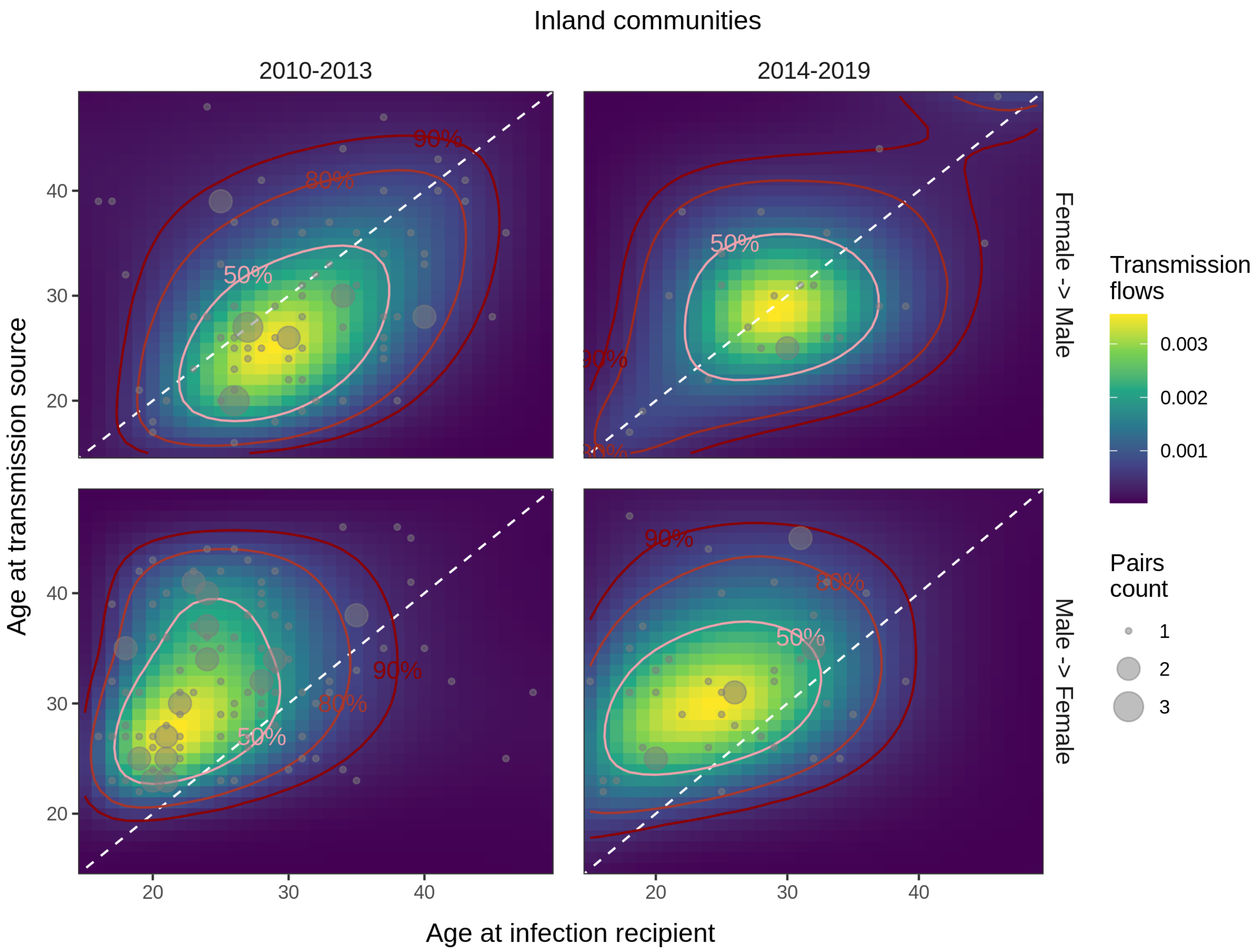
Results: Estimated contribution to COVID-19 deaths over time





Other applications

Estimating shifting HIV transmission dynamics in Rakai, Uganda



Imperial College London

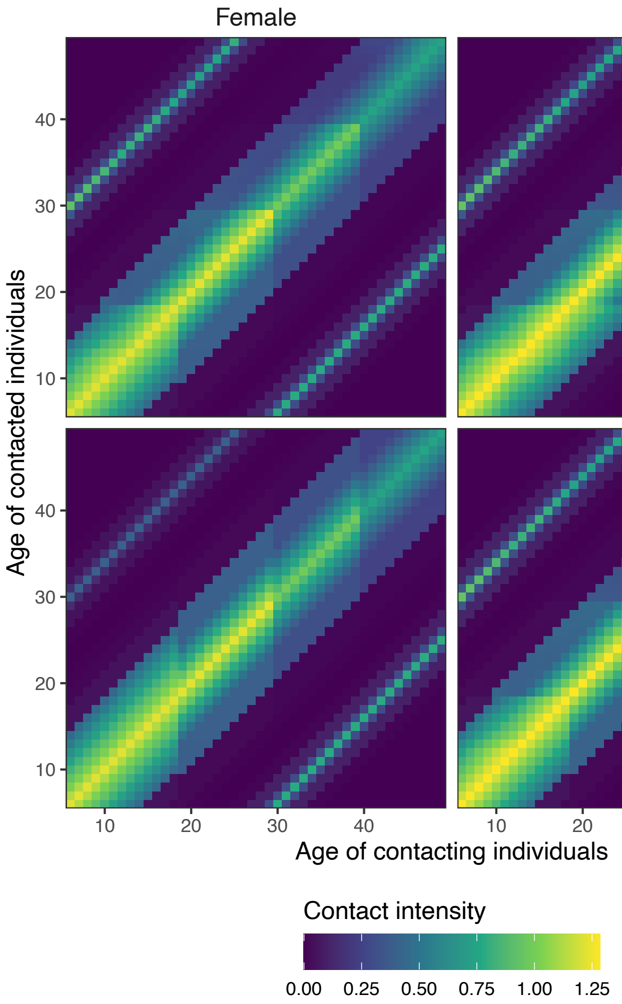
PANGEA HIV

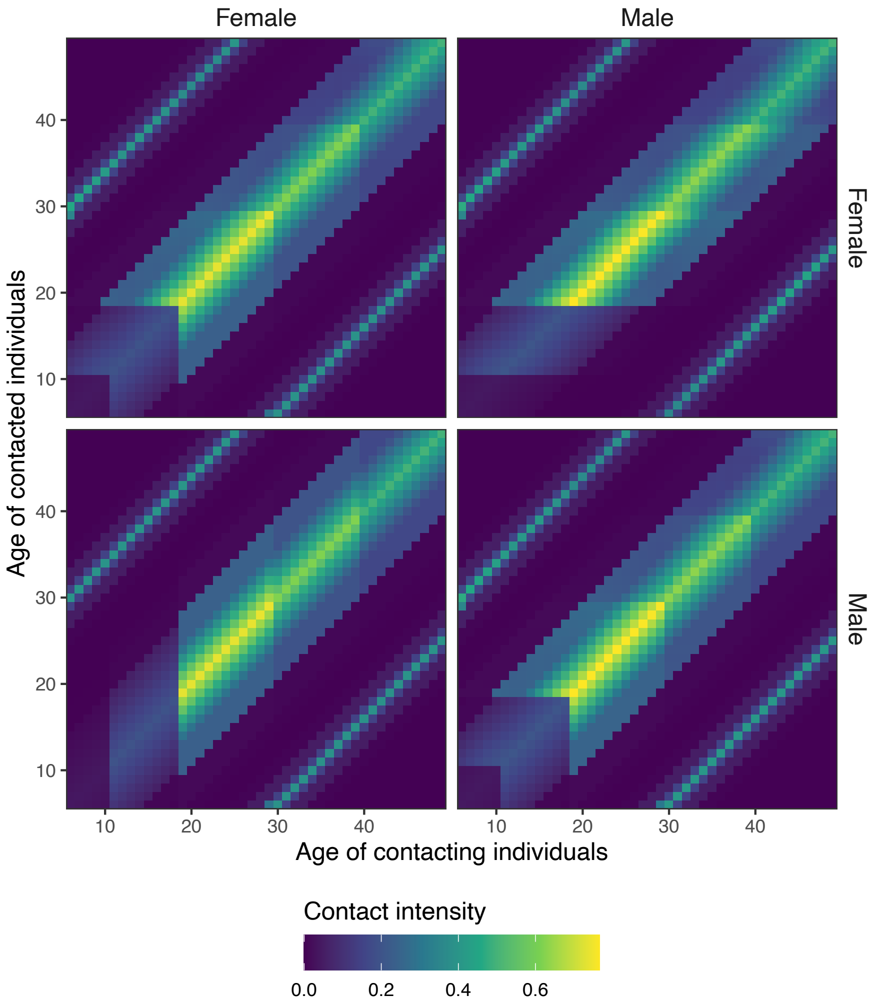
Rakai Health Sciences Program
Improved Health Through Research

BILL & MELINDA GATES foundation

Inferring human contact patterns and contact dynamics in Germany

COVIMOD social contact surveys data collected in Germany during the COVID-19 outbreak.





Discussion

- We develop a novel a low-rank Gaussian Process (GP) projected by regularised B-splines. This projection defines a **new GP with attractive smoothness and computational efficiency properties**.

- Simulation analyses and benchmark results show that the B-splines projected GP may perform better than un-regularised B-splines and Bayesian P-splines, and equivalently well as a standard GP at considerably lower runtimes.

- This approach is **versatile**, applicable to many problems and easy to implement. We provide **templated stan files**.

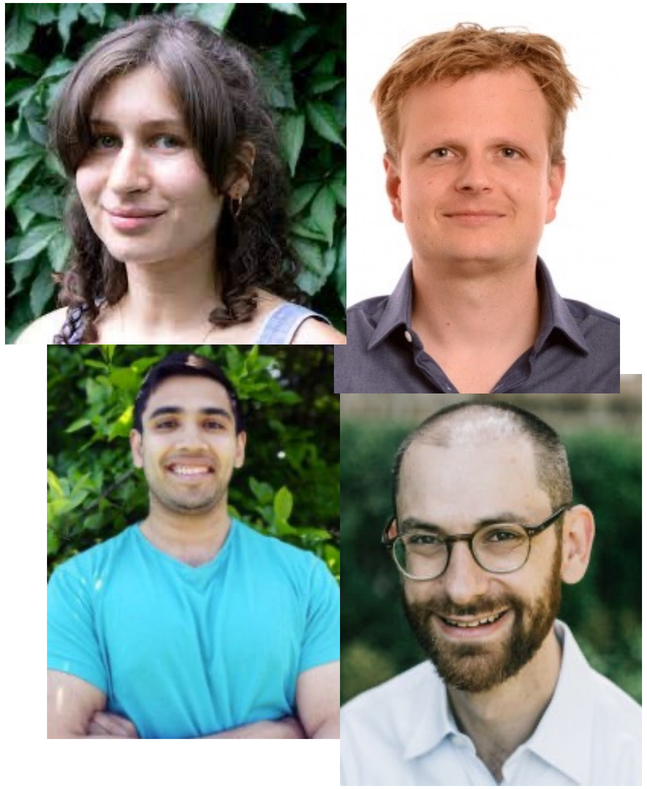
- The B-splines projected GP priors are likely an appealing addition to the arsenal of Bayesian regularising priors.

Thank you

Authors
Mélodie Monod, Alexandra Blenkinsop, Andrea Brizzi, Yu Chen, Carlos Cardoso Correia Perello, Vidoushee Jogarah, Yuanrong Wang, Seth Flaxman, Samir Bhatt and Oliver Ratmann

Funding
Imperial College London COVID-19 Response Fund
EPSRC CDT in Modern Statistics and Statistical Machine Learning
Imperial College London presidential scholarship
Bill & Melinda Gates Foundation
UK Medical Research Council
NIHR Health Protection Research Unit in Modelling Methodology
Community Jameel

Imperial College London
MACHINE LEARNING & GLOBAL HEALTH NETWORK



- Pre-print
  https://arxiv.org/abs/2106.12360

- Data and code including templated stan files CC-BY-4.0
  https://github.com/ImperialCollegeLondon/BSplinesProjectedGPs