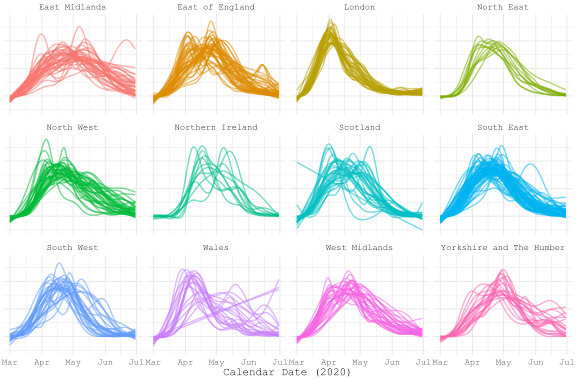
Spatially Aware Temporal Registration of Covid Waves

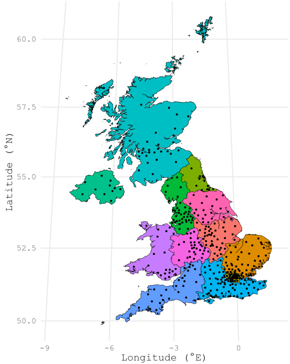
Luke A. Barratt and John A. D. Aston
Statistical Laboratory, DPMMS
University of Cambridge, UK

JSM 2024
Portland, Oregon, USA
5 August 2024

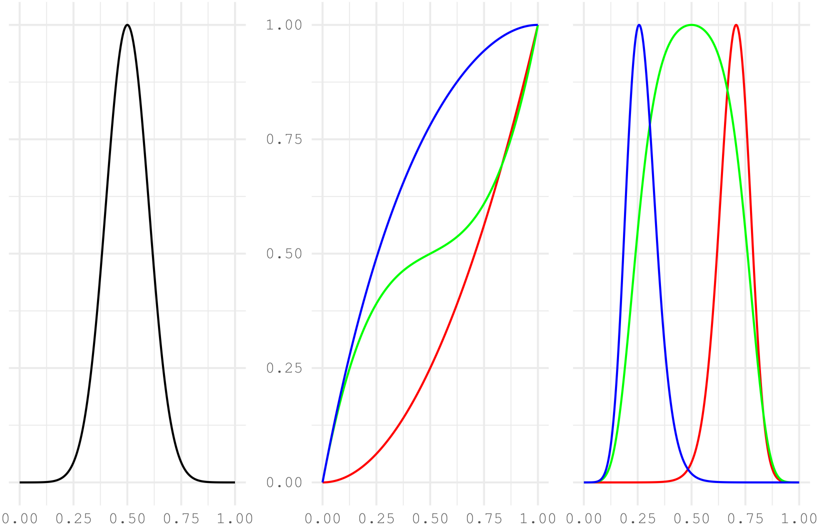
Based on arXiv:2407.17132

The Data


East Midlands
East of England
London
North East
North West
Northern Ireland
Scotland
South East
South West
Wales
West Midlands
Yorkshire and The Humber
Mar
Apr
May
Jun
Jul
Calendar Date (2020)


60.0
57.5
55.0
52.5
50.0
Latitude (°N)
-9
-6
-3
0
Longitude (°E)

Phase Variation



The Statistical Model

$$Y_{ij} = \underbrace{\Xi_i}_{\text{amplitude}} \mu\left(\underbrace{H_i^{-1}}_{\text{phase}}(t_{ij})\right) + \underbrace{\epsilon_{ij}}_{\text{error}}, \quad \underbrace{j \in [m_i]}_{\text{time}}, \quad \underbrace{i \in [n]}_{\text{space}}.$$

Local Variation

$$\Lambda_f(t) := \frac{\int_0^t |f'(\tilde{t})| d\tilde{t}}{\int_0^1 |f'(\tilde{t})| d\tilde{t}}$$

$$\Lambda_{af \circ g} = \Lambda_f \circ g$$

$$H^{-1} = \Lambda_\mu^{-1} \circ \Lambda_{\Xi_\mu \circ H^{-1}}, \qquad \Lambda_\mu^{-1} = \mathbb{E} \Lambda_{\Xi_\mu \circ H^{-1}}^{-1}$$

Optimal Estimation

$$\text{argmin}_{w\in\mathbb{R}^n} \mathbb{E}\left\|\Lambda_\mu^{-1} - \sum_{i=1}^n w_i \Lambda_i^{-1}\right\|^2 \text{ subject to } \mathbb{E}\sum_{i=1}^n w_i \Lambda_i^{-1} = \Lambda_\mu^{-1}$$

$$\mathbb{E}\sum_{i=1}^n w_i \Lambda_i^{-1} = \Lambda_\mu^{-1} \Longleftrightarrow \sum_{i=1}^n w_i = 1$$

$$w^* = \frac{C^{-1}1_n}{1_n^\top C^{-1} 1_n}, \qquad C_{ik} = \mathbb{E}\left\langle \Lambda_i^{-1} - \Lambda_\mu^{-1}, \Lambda_k^{-1} - \Lambda_\mu^{-1} \right\rangle = c(d_{ik})$$

$$\mathbb{E}\left\|\Lambda_i^{-1} - \Lambda_k^{-1}\right\|^2 = 2c(\infty) - 2c(d_{ik})$$

Spatial Registration

Algorithm 1: To estimate the warpings H_i^{-1} from the observations Y_{ij} with pairwise distances d_{ik}.

Fit quintic splines with REML smoothing to the Y_{ij} and differentiate for $\widehat{\mathrm{D}Y}_i$;

Compute $\hat{\Lambda}_i(t) := \int_0^t \left|\widehat{\mathrm{D}Y}_i(\tilde{t})\right| \mathrm{d}\tilde{t} \Big/ \int_0^1 \left|\widehat{\mathrm{D}Y}_i(\tilde{t})\right| \mathrm{d}\tilde{t}$;

Compute the variogram data $(d_{ik}, \frac{1}{2}\|\hat{\Lambda}_i^{-1} - \hat{\Lambda}_k^{-1}\|^2)_{i \neq k}$;

Fit the Matérn model by iteratively reweighted least squares, producing $\hat{\gamma}$;

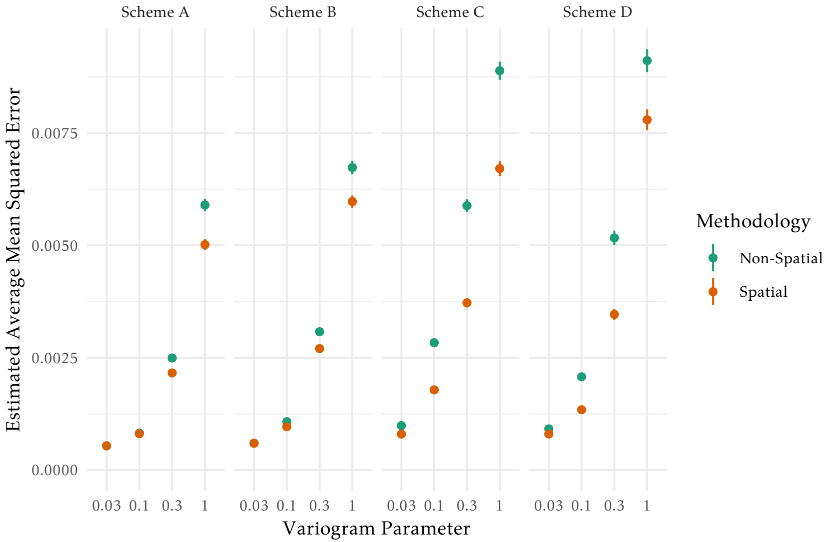
Estimate the covariances $\hat{C}_{ik} := 2\hat{\gamma}(\infty) - 2\hat{\gamma}(d_{ik})$;

Estimate the optimal weights $\hat{w} := \hat{C}^{-1} 1_n \Big/ 1_n^\top \hat{C}^{-1} 1_n$;

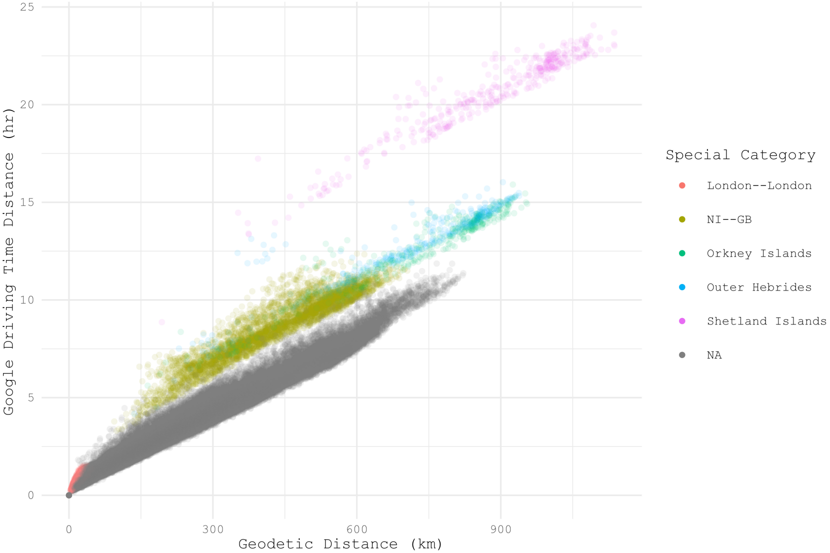
Compute $\hat{\Lambda}_\mu := \left(\sum_{i=1}^n \hat{w}_i \hat{\Lambda}_i^{-1}\right)^{-1}$;

Estimate the warpings $\hat{H}_i^{-1} := \hat{\Lambda}_\mu^{-1} \circ \hat{\Lambda}_i$;

Simulation Results



The Google Drive Time Distance

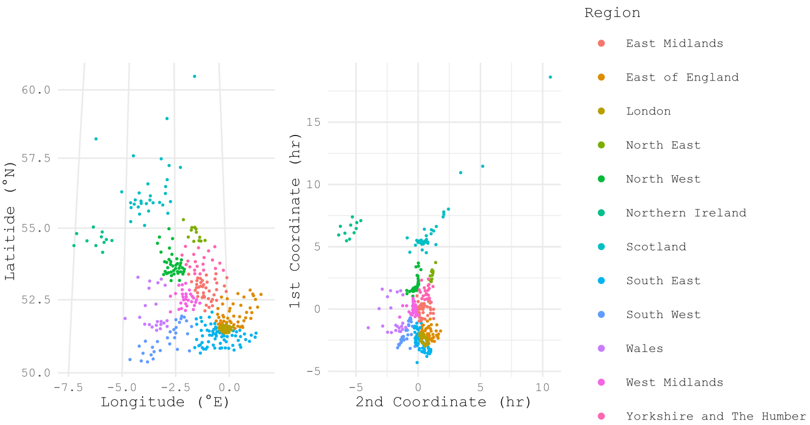


Euclidean Distance Approximation

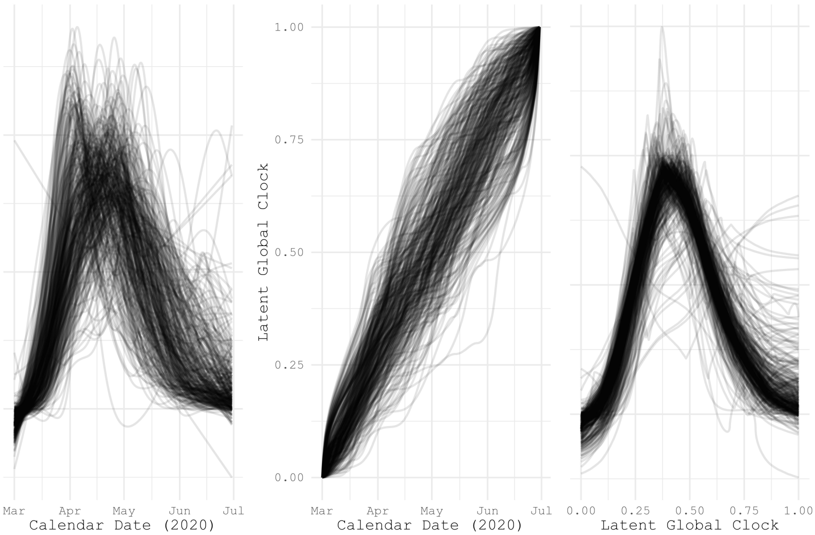
Algorithm 2: To approximate a (non-Euclidean) distance matrix $D \in \mathbb{R}^{n \times n}$ by a Euclidean distance matrix $D^* \in \mathbb{R}^{n \times n}$.

Set $B := -\frac{1}{2} H(D \circ D) H$;
Compute the spectral decomposition $B =: \sum_{i=1}^{n} u_i \lambda_i u_i^\top$, where $\lambda_1 \geq \cdots \geq \lambda_n$;
Choose $p \leq \operatorname{rank} B$;
Set $s_i^* = (\lambda_1 (u_1)_i, \ldots, \lambda_p (u_p)_i)^\top$ for $i \in [n]$;
Set $D_{ik}^* := \|s_i^* - s_k^*\|$ for $i, k \in [n]$;

Approximating Driving Times: 2 Dimensions

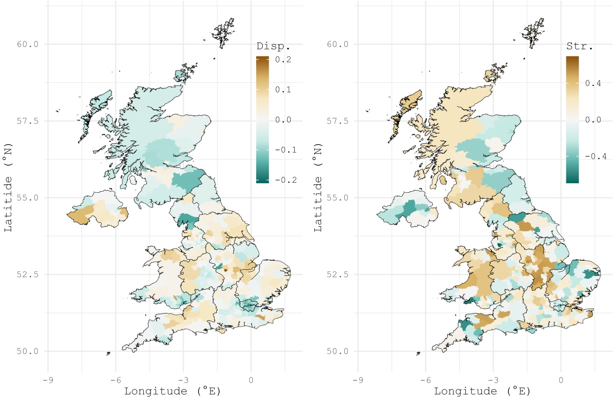


Unaligned to Aligned



Stretch & Displacement I

$$\Delta_i := \int_0^1 t \, \mathrm{d}H_i^{-1}(t) - \frac{1}{2}; \quad \Upsilon_i := \log\left(12 \int_0^1 \left(t - \Delta_i - \frac{1}{2}\right)^2 \mathrm{d}H_i^{-1}(t)\right).$$



Stretch & Displacement II