Supplementary Material: Detection of spatiotemporal changepoints – a generalised additive model approach

Michael J. Hollaway & Rebecca Killick

June 2024

1 Simulation Studies Full Details

This section presents the full details of the simulation studies undertaken to evaluate the performance of GAM-PELT. Two main sets of scenarios are run, with the first involving different underlying spatial and temporal distributions but no changepoints (see Table S1); the second introduces changepoints at times 50, 100, and 150 with a change in either the spatial, temporal or spatio-temporal structure between segments (see Table S2). For all scenarios, the number of time points and spatial locations were fixed at 200 and 50 respectively. For each scenario 100 replicates were run and the spatial locations were generated at random at the start of each replicate uniformly from $u \sim$ Unif(-3,3), $v \sim$ Unif(40,60), rounded to 1 decimal place.

In the tables, we use summary notation for the different components of each scenario. In the following we unpack what this means in the practical construction of the simulations. Each component is additive. For all scenarios, the mean and variance are fixed a 0.0 and 1.0 respectively for the first segment and "random draw" means equally weighted.

Time:

Independent Errors are independent and identically distributed $\mathrm{N}(0, 1)$.

AR1:Same Errors are AR1 models with the AR parameter simulated from a Truncated Normal $\mathrm{TN}(0, 1, 0.1, 0.9)$. A single AR parameter is simulated and applied to all spatial locations.

AR1:Random Errors are AR1 models with the AR parameter simulated from a Truncated Normal $\mathrm{TN}(0, 1, 0.1, 0.9)$. A different AR parameter is simulated for each spatial location.

Mean & Variance The mean parameters for each segment are drawn from a normal distribution ($\mathrm{N}(0.0, 5.0)$) and the variance parameters from a lognormal distribution (log-$\mathrm{N}(0.0, \frac{log(10)}{2})$).

Scenario	Time	Space
1	AR1: Same all locations Mean: Same all spatial locations Variance: Same all spatial locations	Constant
2	AR1: Random all locations Mean: Same all spatial locations Variance: Same all spatial locations	Constant
3	Independent	Constant
4	Independent	Independent
5	Independent	Structured correlation
6	Independent	2-D `gamSim` example (`mgcv`)

Table S1: Summary of scenarios where there are no spatial and/or temporal changes in the simulated dataset.

Spatial:

Constant Each spatial location has the same value in space drawn from $\mathrm{N}(0, 5^2)$.

Independent Each spatial location has a different value in space independently drawn from $\mathrm{N}(\mu, \sigma^2, 0.0, 5.0)$.

Structured Each spatial location is spatially correlated using an unconditional Gaussian simulation from the `gstat` R package (E. J. Pebesma 2004; Gräler, E. Pebesma, and Heuvelink 2016). To ensure quantitatively different spatial distributions in each replicate, the range, sill and nmax (number of nearest observations used in Gaussian simulation) were sampled from $range \sim$ Unif(10,1000), $sill \sim$ Unif(5.0,5000.0) and $nmax \sim$ Unif(10,200) respectively.

2-d `gamSim` The underlying spatial distribution was generated using the 2-D `gamSim` example from `mgcv`.

2 AURN dataset details

To demonstrate an application of GAM-PELT on a real-world dataset we utilised data from the United Kingdom (UK)'s Automatic, Urban and Rural Network (AURN) of air quality monitoring stations (Defra 2024a). The AURN is a network of 281 monitoring stations that provide measurements of key air quality pollutants covering the time period of 1^{st} July 1972 through to the present day. Out of the 281 sites, 175 are currently still actively monitoring and are located at various urban, rural, roadside and industrial locations around England, Scotland, Wales and Northern Ireland. Pollutant data is available for ozone (O_3), Nitric oxide (NO), Nitrogen dioxide (NO_2) Nitrogen oxides as nitrogen dioxide (NO_x), Sulphur dioxide (SO_2), Carbon monoxide (CO), particulate matter of

Scenario	Time	Space
1a	***All spatial locations change:*** AR1, Mean & Variance	***No change:*** Constant
1b	***All spatial locations change:*** AR1, Mean & Variance	***No change:*** All random
1c	***All spatial locations change:*** AR1, Mean & Variance	***No change:*** Correlated
2a	***1 spatial location at random changes:*** AR1, Mean & Variance	***No change:*** Constant
2b	***1 spatial location at random changes:*** AR1, Mean & Variance	***No change:*** All random
2c	***1 spatial location at random changes:*** AR1, Mean & Variance	***No change:*** Correlated
3a	***No change:*** AR1, Mean & Variance	***Change:*** Constant → Constant
3b	***No change:*** AR1, Mean & Variance	***Change:*** All random → All random
3c	***No change:*** AR1, Mean & Variance	***Change:*** Correlated → Correlated
4a	***AR1, Mean & Variance change:*** Random draw from: 1) All spatial locations change 2) 1 spatial location changes 3) No change	***Spatial structure change:*** Random draw from: 1) Constant 2) All random 3) Correlated 4) No Change
4b	***AR1, Mean & Variance change:*** Random draw from: 1) All spatial locations change 2) 1 spatial location changes	***Spatial structure change:*** Random draw from: 1) Constant 2) All random 3) Structured correlation

Table S2: Summary of scenarios where there are changepoints introduced into the simulated dataset and there is a change in either the spatial, temporal or spatio-temporal structure between segments .

size smaller than 10 micron (PM_{10}), and particulate matter of size smaller than 10 micron ($PM_{2.5}$) at hourly frequency. This provides good coverage of key air quality pollutants across the UK. The AURN data is actively updated daily and is made available under an Open Government Licence by Defra through the UK-Air website (Defra 2024b) or using the API called from the `OpenAir` R package (Carslaw and Ropkins 2012).

For the data application the period 1^{st} February - 31^{st} August 2020 (213 days) was chosen as this covers the timeline of the UK's first nationwide COVID-19 lockdown, whereby impacts on pollutant concentrations would be expected to be seen in some effect at all monitoring locations. We focus on 2 primary (directly emitted) pollutants namely NO_2 (measured at 74 spatial locations) and $PM_{2.5}$ (30 spatial locations), and 1 secondary pollutant (formed in the atmosphere and thus behaves differently), namely O_3 (30 spatial locations). For the purposes of this application, the hourly data was aggregated to daily averages which yielded 74 locations for NO_2, 30 locations for $PM_{2.5}$ and 30 locations for O_3 that had complete daily records for the lockdown period (station locations for each pollutant are shown in Figure S1). Figure S2 shows example daily time series of O_3, NO_2 and $PM_{2.5}$ for sites located in Southend and Wigan Centre where all 3 pollutants are measured concurrently.

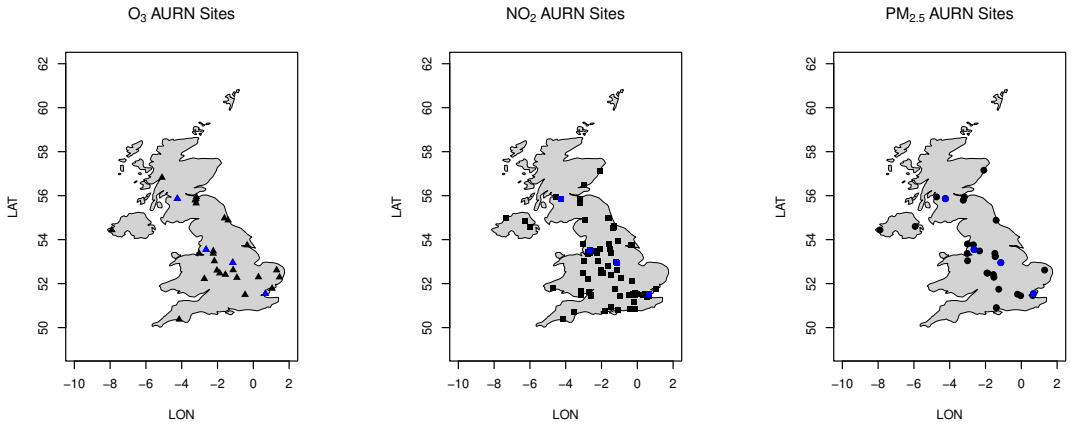

Figure S1: Locations of the AURN measurement locations for O_3 (triangles), NO_2 (squares) and $PM_{2.5}$ (circles). Sites where all 3 pollutants are measured concurrently are highlighted in blue.

References

Pebesma, Edzer J. (2004). "Multivariable geostatistics in S: the gstat package". In: *Computers & Geosciences* 30, pp. 683–691.

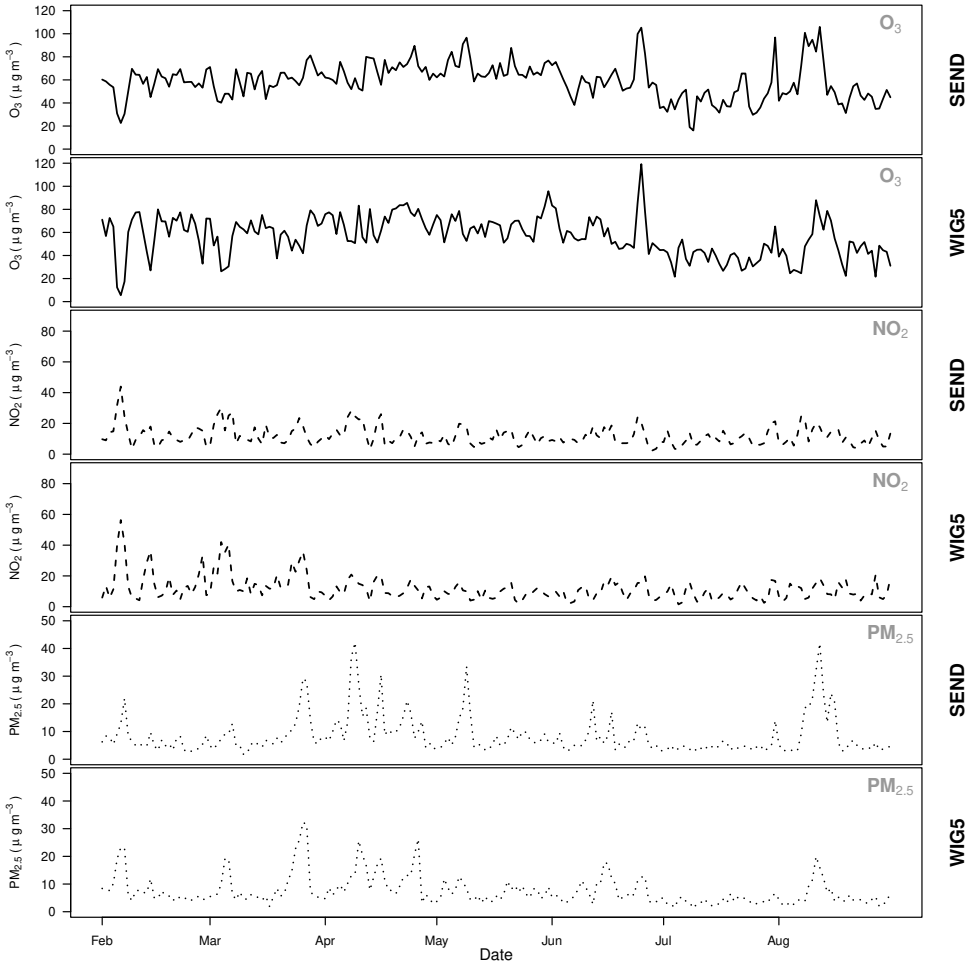

Figure S2: Daily concentration measurements of O_3, NO_2 and $PM_{2.5}$ at Southend-on-Sea (SEND) and Wigan Centre (WIG5) for the period 1^{st} February - 31^{st} August 2020. The top 2 plots show O_3 concentrations, the middle 2 plots show NO_2 concentrations and the bottom 2 plots show $PM_{2.5}$ concentrations.

Carslaw, David C. and Karl Ropkins (2012). "openair — An R package for air quality data analysis". In: *Environmental Modelling & Software* 27–28.0, pp. 52–61. ISSN: 1364-8152. DOI: 10.1016/j.envsoft.2011.09.008.

Gräler, Benedikt, Edzer Pebesma, and Gerard Heuvelink (2016). "Spatio-Temporal Interpolation using gstat". In: *The R Journal* 8 (1), pp. 204–218. URL:

https://journal.r-project.org/archive/2016/RJ-2016-014/index.html.

Defra (2024a). *Automatic Urban and Rural Network (AURN)*. Accessed: 2024-06-06. URL: https://uk-air.defra.gov.uk/networks/network-info?view=aurn.

— (2024b). *UK AIR: Air Information Resource Data Selector*. Accessed: 2024-06-06. URL: https://uk-air.defra.gov.uk/data/data_selector.