Modelling spatio-temporal variation in black smoke pollutant concentrations

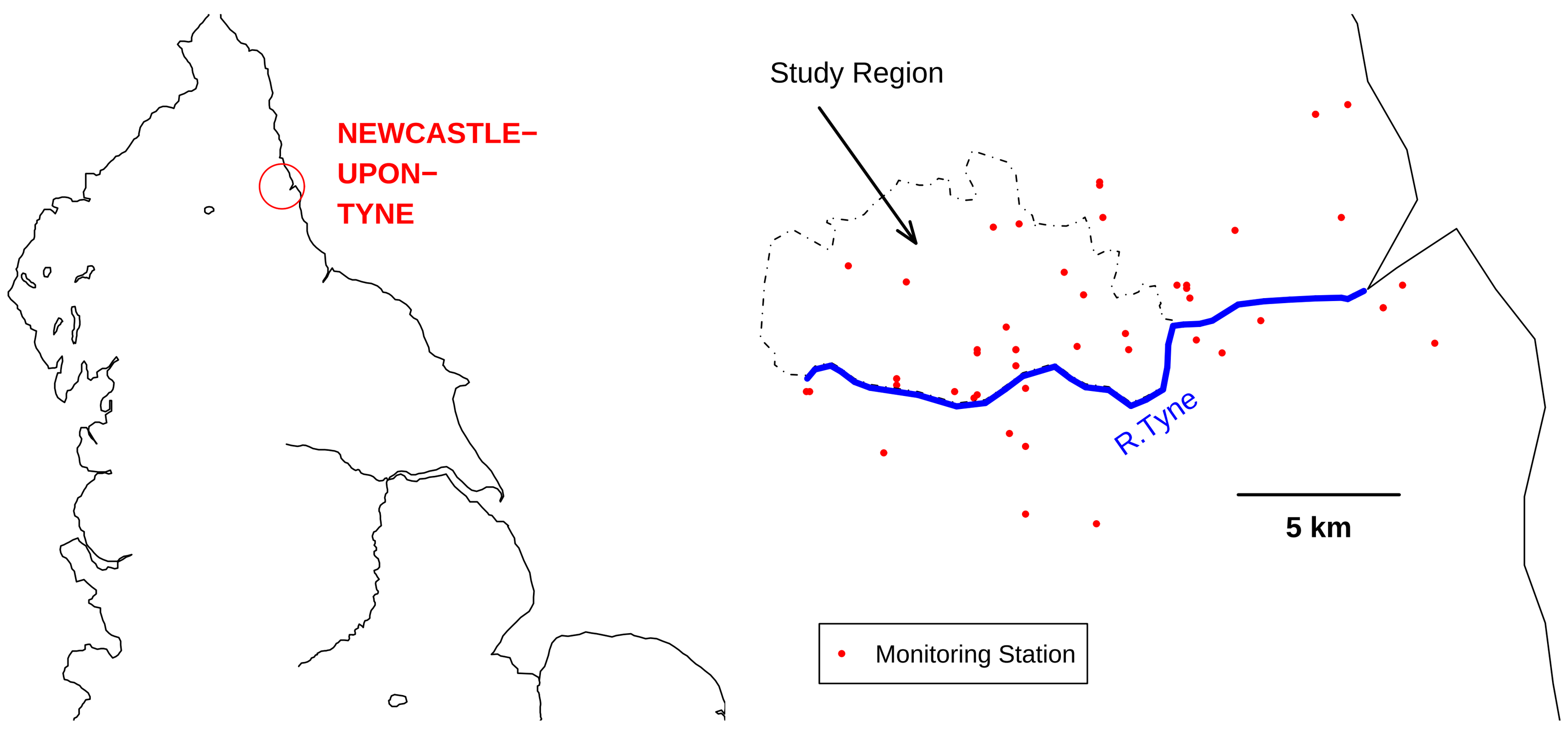
T.R. Fanshawe, P.J. Diggle, Lancaster University, on behalf of the PAMPER team, email: `t.fanshawe@lancaster.ac.uk`

The UK PAMPER Study

The UK PAMPER (Particulate Matter and Perinatal Events Research) study is a historical cohort study to investigate the relationship between pregnancy outcomes and explanatory variables encompassing air pollution, socio-economic and meteorological data.

Location Newcastle-upon-Tyne, 1961-1992

Aim To predict black smoke level for any location in the study region, for any time in the study period.



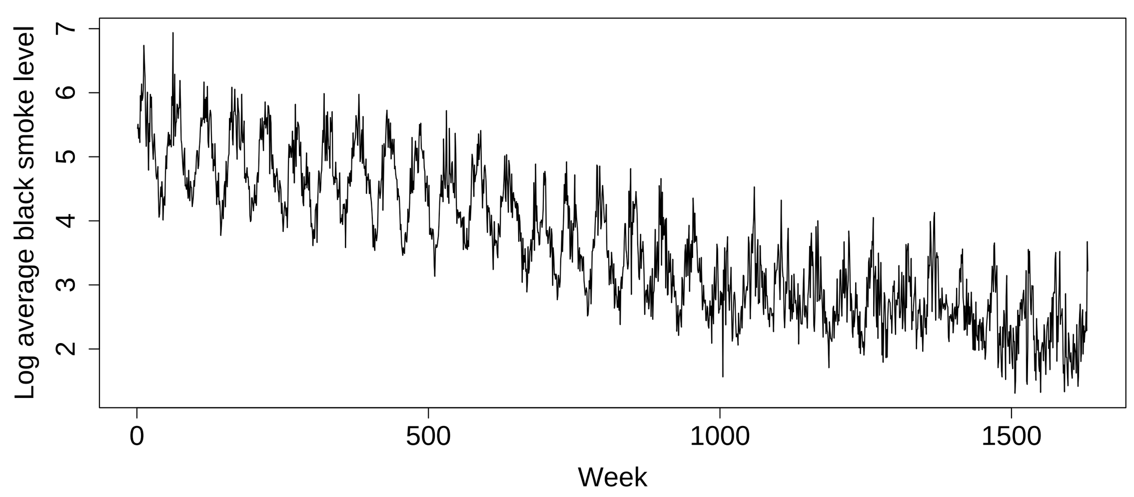
Weekly levels of BS are available from 20 monitoring stations in the study region (above). Whole-region average weekly minimum temperature readings are also available.

Covariates Spatio-temporal variables available include:

- Domestic chimney density within 500m (w_1)
- Distance to nearest industrial area (w_2)
- Binary indicator: residential/non-residential area (w_3)
- Binary indicator: 1956 Clean Air Act implemented? (w_4)

Step 1: Modelling whole-region average BS levels

Log-BS levels ($\log Y$) for the whole study period, averaged across all monitoring stations, are shown below.



We start by modelling seasonal and long-term temporal variations, which are larger than the spatial variation.

Typically the range of log-BS levels of different monitors in a single week are around 1 unit.

There is a strong negative relationship with weekly minimum temperature d_t, and a clear seasonal pattern, levels being higher during the winter.

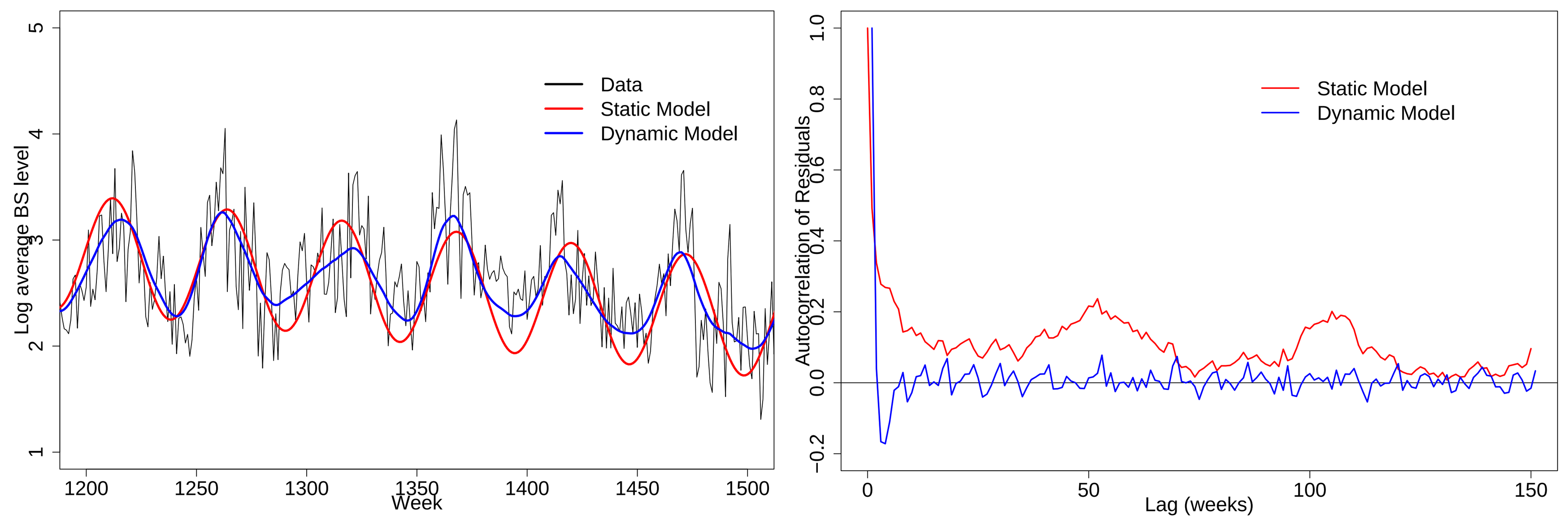
- A dynamic harmonic linear regression model:

$$\log(Y_t) = \alpha + \beta t + \gamma d_t + A_t \cos\left(\frac{2\pi t}{52}\right) + B_t \sin\left(\frac{2\pi t}{52}\right) + \nu_t$$
$$\nu_t \sim \text{IID } N(0, \sigma_V^2)$$

- Residuals from the simple model with static coefficients $A_t \equiv A$, $B_t \equiv B$ are autocorrelated (below).
- An improvement: allow A and B to change slowly over time according to a simple random walk:

$$A_t \sim N(A_{t-1}, \sigma_A^2) \;\; ; \;\; B_t \sim N(B_{t-1}, \sigma_B^2)$$

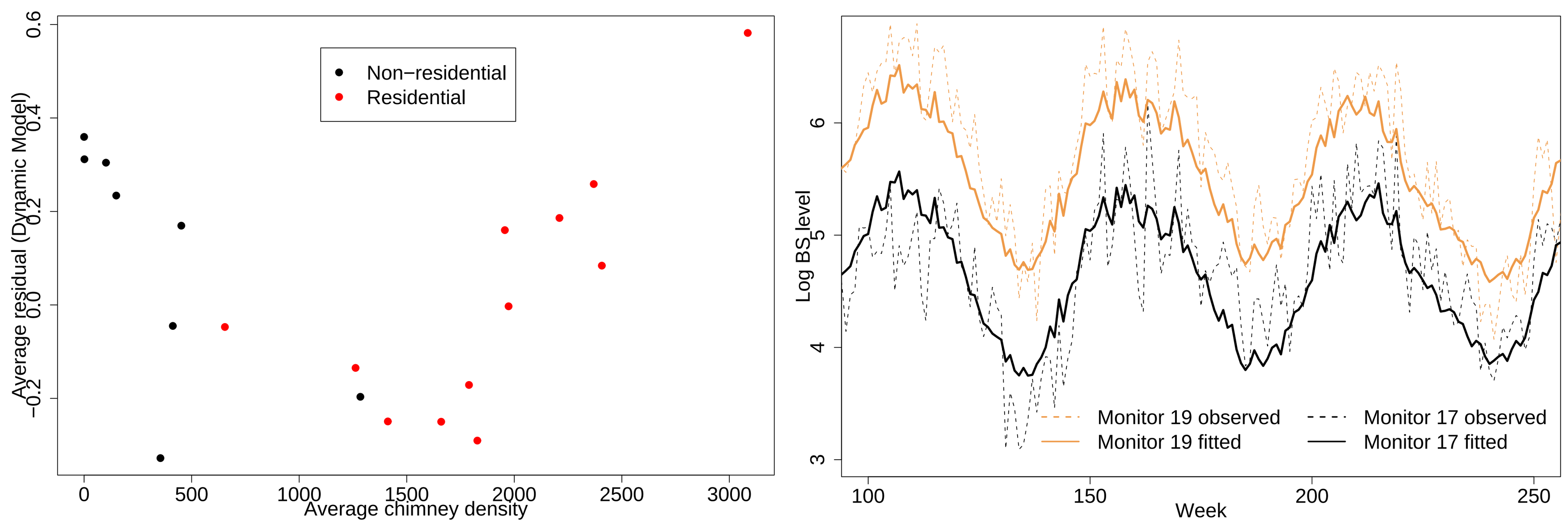
This model removes the residual autocorrelation (below).



Step 2: Modelling spatial variation

- Use BS data from the 20 individual monitoring stations
- Offset: Step 1 fitted values
- Covariates: $w_1, \ldots, w_4$

e.g. Strong evidence that the effect of chimney density differs between residential and non-residential areas.



Model checking There is no evidence from variograms that residuals are temporally or spatially correlated.

- Therefore we consider $Z_{k,t} \sim \text{IID } N(0, \tau^2)$ for all k, t.

Step 3: Prediction at birth locations

- The assumption about the $Z_{k,t}$ enables predictions $Q_{x,t}$ of BS to be made separately for each birth location x and time t in the study period, with standard errors.
- $Q_{x,t}$, integrated over relevant pregnancy periods, will be used as an explanatory variable in the subsequent primary PAMPER analysis.

Hypothesis Higher values of $Q_{x,t}$ are associated with higher probabilities of an adverse pregnancy outcome.