Clifford, Sam, Low-Choy, Samantha, Mengersen, Kerrie, & Morawska, Lidia (2012) Semi-parametric modelling of ultrafine particle number concentration. In *ISBA2012*, 25-29 June 2012, Kyoto Terrsa, Kyoto. (Unpublished)

Semi-parametric modelling of ultrafine particle number concentration

Sam Clifford Samantha Low Choy Kerrie Mengersen Lidia Morawska

Science and Engineering Faculty, Queensland University of Technology, Brisbane, Australia

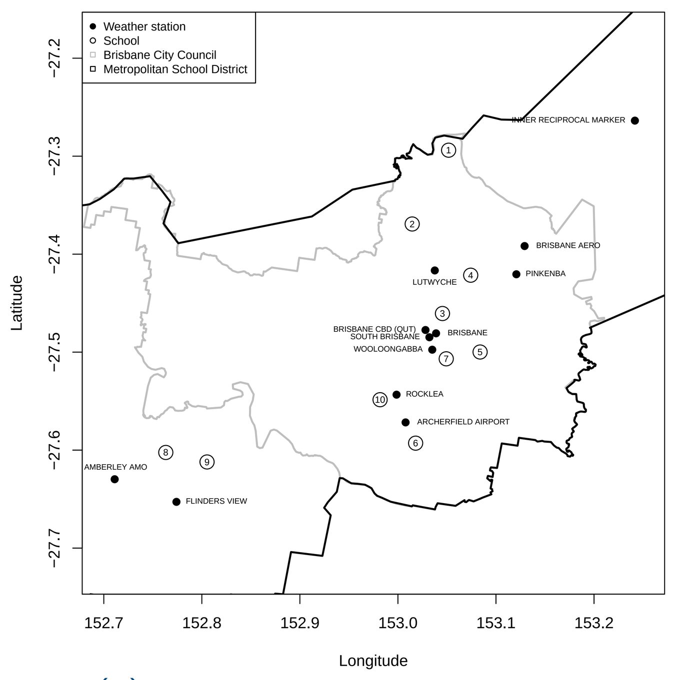
Introduction

Quantifying spatial and/or temporal trends in environmental modelling data requires that measurements be taken at multiple sites. The number of sites and duration of measurement at each site must be balanced against costs of equipment and availability of trained staff. The split panel design comprises short measurement campaigns at multiple locations and continuous monitoring at reference sites [2].

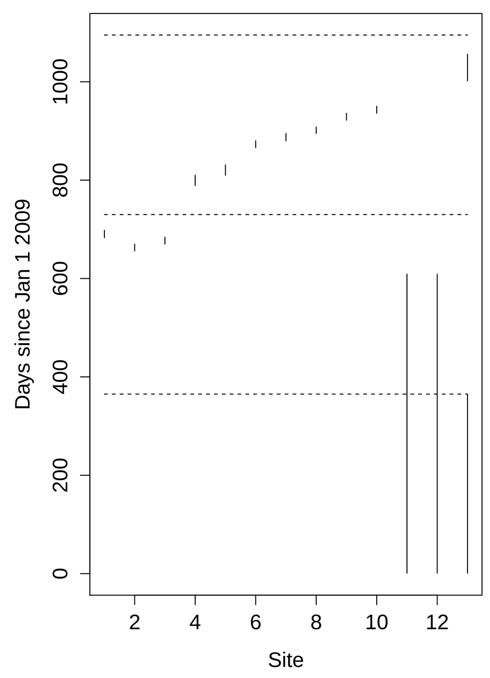
Here we present a modelling approach for a spatio-temporal model of ultrafine particle number concentration (PNC) recorded according to a split panel design. The model describes the temporal trends and background levels at each site.

The data were measured as part of the "Ultrafine Particles from Transport Emissions and Child Health" (UPTECH) project which aims to link air quality measurements, child health outcomes and a questionnaire on the child's history and demographics.

The UPTECH project involves measuring aerosol and particle counts and local meteorology at each of 25 primary schools for two weeks and at three long term monitoring stations, and health outcomes for a cohort of students at each school [3].



(a) Locations of measurement sites



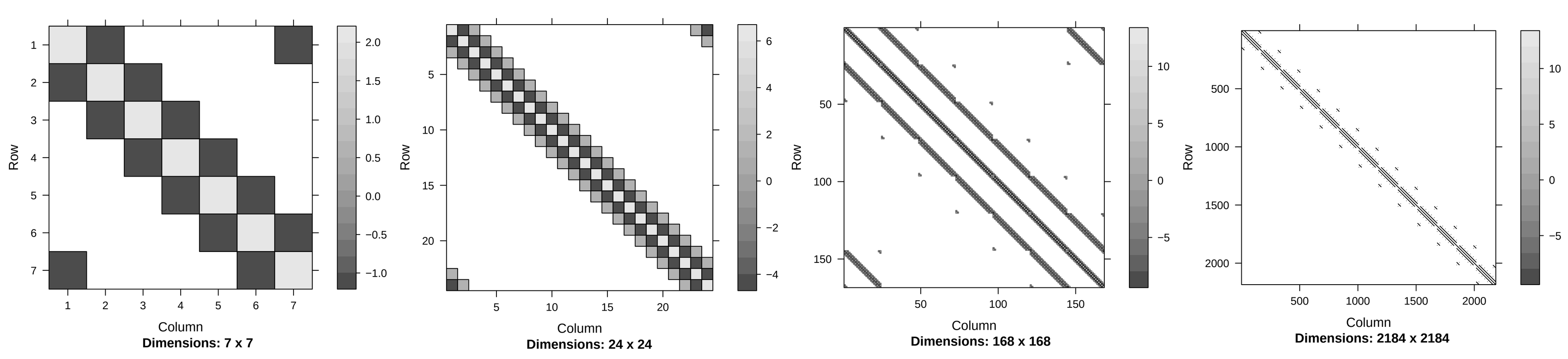
(b) Days of measurement at schools and long term monitoring stations, expressed as number of days since January 1 2009. Dashed lines represent the end of each year.

Figure: Panel designs for UPTECH project. Sites 1 to 10 correspond to schools 1 to 10, site 11 to Woolloongabba, 12 to Rocklea and 13 to QUT.

Three condensation particle counters (CPCs) are deployed at each school and one at each reference site. Placement of CPCs at each school is approximately colinear with the first closest to the upwind road, the second in the middle of the school and the third furthest from the upwind road.

Methodology

Hourly averaged ultrafine PNC is modelled with the Generalised Additive Model in INLA [7]. A daily trend may be included in the model with a second order random walk model for hour of the day. A weekly trend may be included in the model with a first order random walk penalty. The tensor product of these trends (and their precision matrices) correspond to a joint daily and weekly trend, i.e. the daily trend varies over the week. The precision matrices for the multivariate normal distributions of the parameters for these terms are shown below.



(a) Weekly (b) Daily (c) Joint (d) Joint at 13 sites

Figure: Prior precision matrices for modelling temporal trends

We assume that the mean at each site is independent and that the three locations within each school may be treated either by pooling as repeated measures, by assuming independence or explicitly assuming exchangeability.

Four competing regression models are fit, each with a different specification of the spatial and temporal relationships, and choose with the DIC:

1. separable trends, CPCs at each site as repeated measures
2. tensor product of daily and weekly trends, repeated measures
3. separable trends, CPCs exchangeable
4. separable trends, CPCs independent

All models will include a temporal trend common to all sites. The general form for the regression is, thus,

$$\begin{aligned}\log y_{ijk} &= \beta_0 + \beta_{1jk} + f_0(\text{hour}_i, \text{weekday}_i) + f_j\left(\text{hour}_{ij}, \text{weekday}_{ij}\right) + \varepsilon_{ij} \\ \varepsilon_{ij} &\sim \mathcal{N}\left(0, \sigma^2\right)\end{aligned}$$

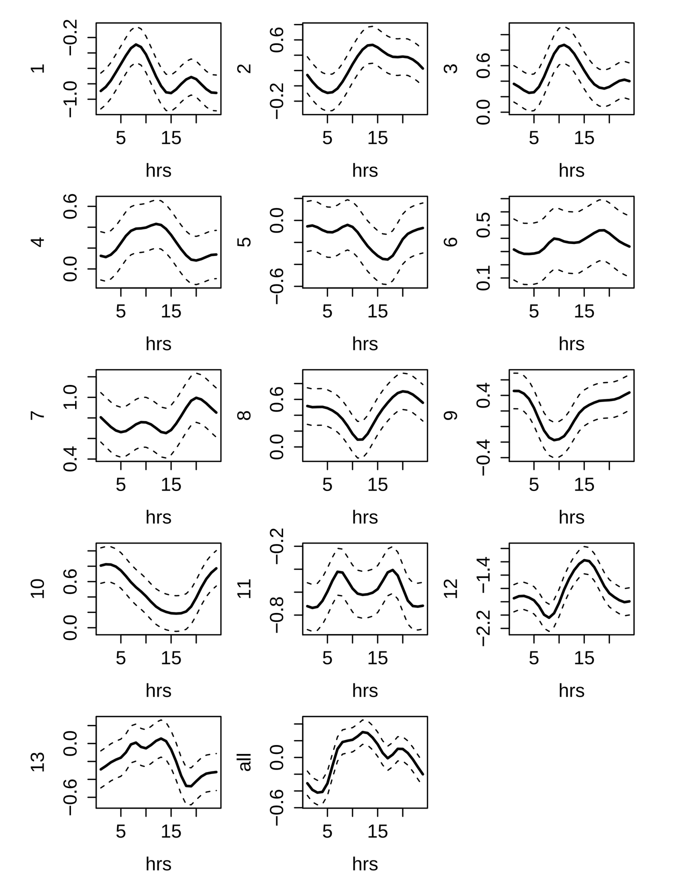
Methodology (cont'd)

where $f_j(\cdot,\cdot)$ represents the temporal trends at site j and $f_0(\cdot,\cdot)$ represents the temporal trends at all sites. These may be additive functions of the hour of the day and day of the week or a joint, non-separable trend. All terms in the regression model are represented as Gaussian Markov Random Fields and are fit using integrated, nested Laplace approximations [8] in R-INLA [7].

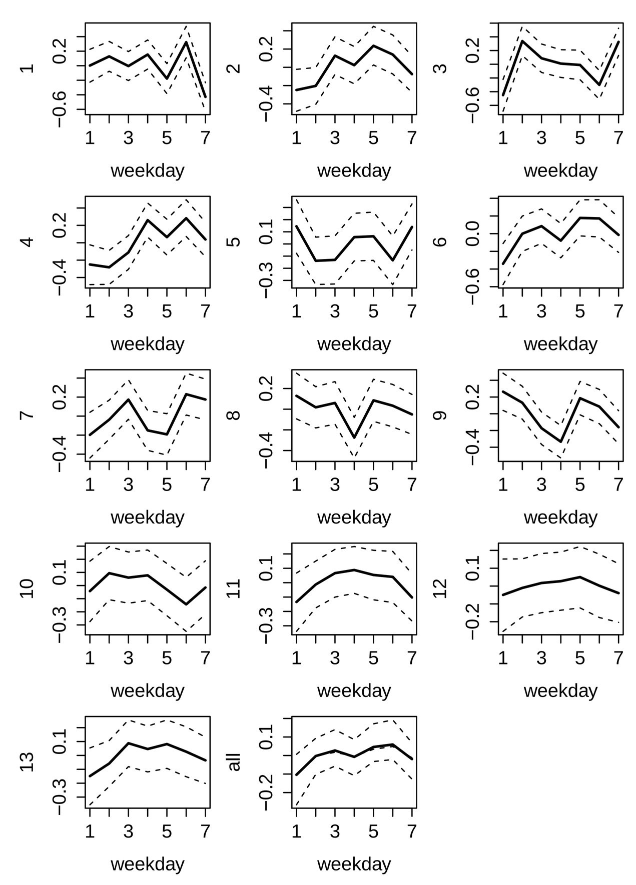
Here β_0 is the overall mean across all sites, with a weakly informative normal prior, and β_{1jk} is the mean of device k at site j and may be modelled as a repeated measure ($\beta_{1jk} \equiv \beta_{1j}$), or as exchangeable or independent. The exchangeable spatial effect prior precision matrix is block diagonal matrix of ones where each block consists of a $K_j \times K_j$ matrix for K_j CPCs deployed at site j.

Results and Discussion

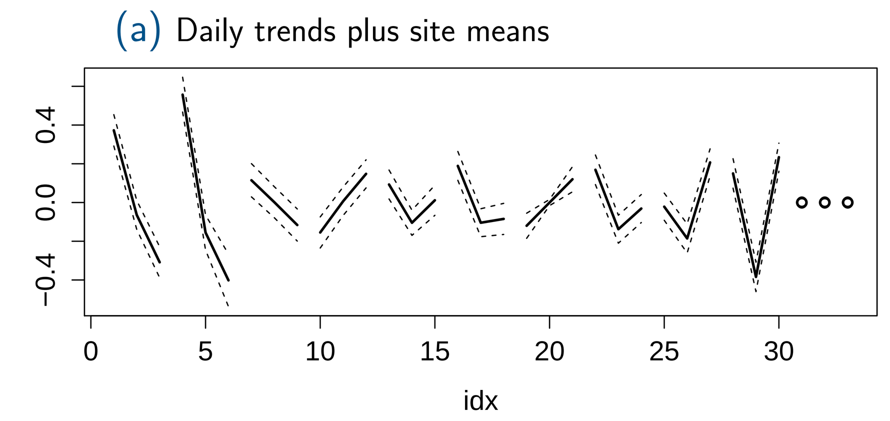
The models with exchangeable and independent priors for the spatial random effect provide qualitatively similar inference and have the lowest DICs of the fitted models (83349 and 88348, respectively, compared to 83817 for separable trends with repeated measures and 84074 for joint trends).



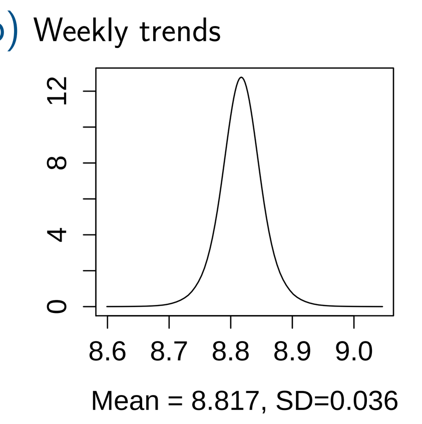
(a) Daily trends plus site means



(b) Weekly trends



(c) Exchangeable spatial random effect



(d) Intercept

Figure: Summary of fitted terms from exchangeable spatial random effect model.

The all-site daily trend contains a morning to midday peak and a small evening peak, corresponding to common traffic patterns and nucleation [6, 5]. The site-specific daily trends differ from each other and represent the deviation from the common daily trend, and contains the effect of local sources such as traffic and land use [4]. The weekly trend has its minimum values on the weekend (days 1 and 7) in the all-site term and this is reinforced at most schools.

The independent or exchangeable within-school spatial random effects provide the best fit according to the DIC. This indicates that including spatial variability within the school provides for a better fit. It is hoped that once the project is finished, a continuous spatial random effect may be fitted to the data from all 25 schools.

The work presented here is in preparation for publication as [1]. This work is funded by an Australian Research Council *Linkage* grant LP0882544, *Quantification of Traffic Generated Nano and Ultrafine Particle Dynamics and Toxicity in Transit Hubs and Transport Corridors*. The industry partners in this project are the Queensland Government agencies Education Queensland and Department of Transport and Main Roads. The authors wish to thank all members of the UPTECH project.

[1] Sam Clifford, Sama Low Choy, Kerrie Mengersen, and Lidia Morawska. Semi-parametric hierarchical modelling of multi-site observational data. *in prep*, 2012.

[2] Melissa J. Dobbie and Brent L. Henderson. Sparse sampling: spatial design for monitoring stream networks. *Statistics Surveys*, 2:113–153, 2008.

[3] Jaime F. Mejía, Samantha Low Choy, Kerrie Mengersen, and Lidia Morawska. Methodology for assessing exposure and impacts of air pollutants in school children: Data collection, analysis and health effects - a literature review. *Atmospheric Environment*, 45(4):813 – 823, 2011.

[4] Jaime F. Mejía, Lidia Morawska, and Kerrie Mengersen. Spatial variation in particle number size distributions in a large metropolitan area. *Atmospheric Chemistry and Physics*, 8:1127–1138, 2008.

[5] Lidia Morawska, E. R. Jayaratne, Kerrie Mengersen, M. Jamriska, and S. Thomas. Differences in airborne particle and gaseous concentrations in urban air between weekdays and weekends. *Atmospheric Environment*, 36:4375–4383, 2002.

[6] Lidia Morawska, S. Thomas, Neville Bofinger, D. Wainwright, and D. Neale. Comprehensive characterization of aerosols in a subtropical urban atmosphere: particle size distribution and correlation with gaseous pollutants. *Atmospheric Environment*, 32:2467–2478, 1998.

[7] Håvard Rue and Sara Martino. *INLA – Functions which allow to perform a full Bayesian analysis of structured additive models using Integrated Nested Laplace Approximaxion*, 0.0 edition, December 2010.

[8] Håvard Rue, Sara Martino, and Nicolas Chopin. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society B*, 71:319–392, 2009.