

Exploring Spatial and Temporal Heterogeneity of Environmental Noise in Toronto

Camila Casquilho-Resende, Nick Fishbane, Seong-Hwan Jun, and Yunlong Nie Department of Statistics, The University of British Columbia

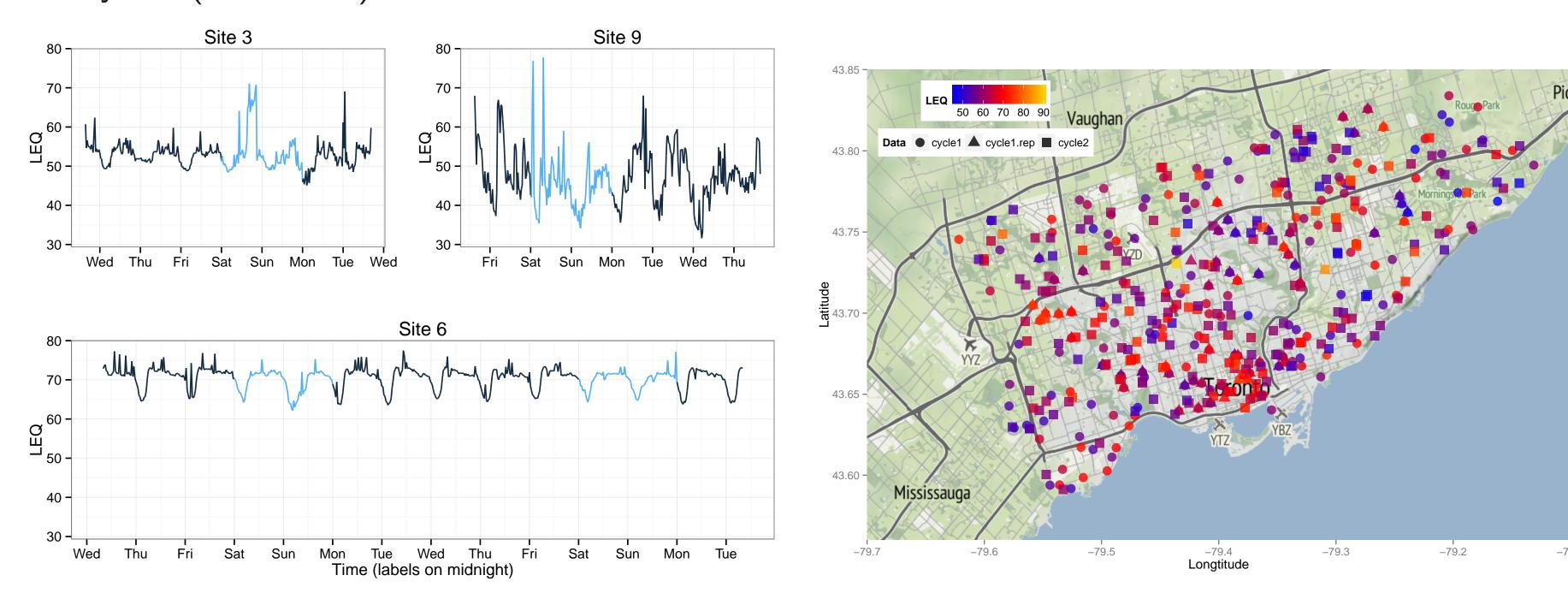
Introduction

Objective

The objective is to analyze the variation of environmental noise across temporal and spatial dimensions. In the process, we identify site characteristics that contribute to the noise level in Toronto.

Data

Noise was measured at 10 sites throughout Toronto over a week (or more) at 30-minute intervals. Time series were recorded at various intervals from mid-June to early September, starting at different times in the week, with various amplitudes and noise levels. This can be seen by the time-series from three of the ten sites. The map shows the site locations of this case study, collected in three phases: Cycle1 (n=241), Cycle1.rep (n=100), and Cycle2 (n=213).



Methods Overview

- ► Temporal variation was modeled using a linear mixed effect (LME) model with harmonic and polynomial components
- ► Importance of site characteristics assessed with random forest and Dirichlet process mixture models
- ► Spatial variation of adjusted noise (LEQ) explored using residual kriging by assuming a Gaussian process with Matérn covariance function
- Aforementioned methods reapplied on test data

Temporal Model

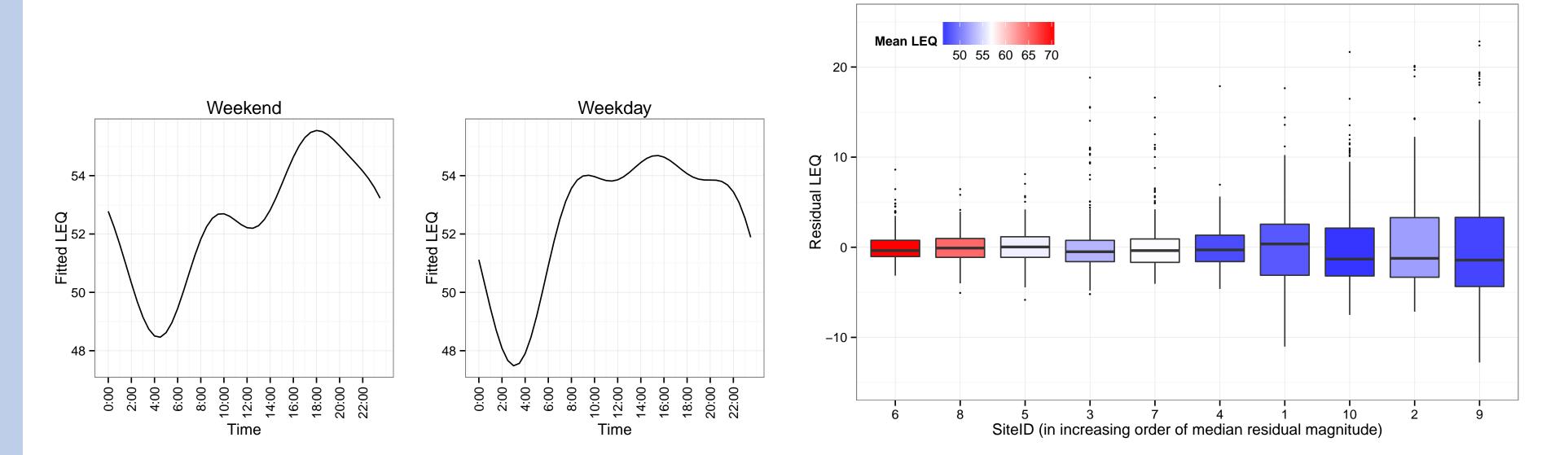
We assume each site to be independent since we observe no relationship between distance and noise correlation between sites. Also, for each site we only consider the "time point" of the observations, which we define as the time of week (of which there are 48*7 = 336). The LME is as follows:

$$egin{aligned} LEQ_{it} &= b_i + d_1(t) + d_2(t)\mathbb{1}_{[t \in W]} + w(t) + \epsilon_{it} \quad b_i \sim \mathcal{N}(0,\sigma^2) \ d_j(t) &= eta_{0j} + eta_{1j}\sin\left(rac{2\pi t}{24}
ight) + eta_{2j}\cos\left(rac{2\pi t}{24}
ight) + eta_{3j}\sin\left(rac{2\pi t}{12}
ight) \ + eta_{4j}\sin\left(rac{2\pi t}{8}
ight) + eta_{5j}\cos\left(rac{2\pi t}{8}
ight) \quad j = 1,2 \ w(t) &= lpha_1 t + lpha_2 t^2 + lpha_3 t^3 \end{aligned}$$

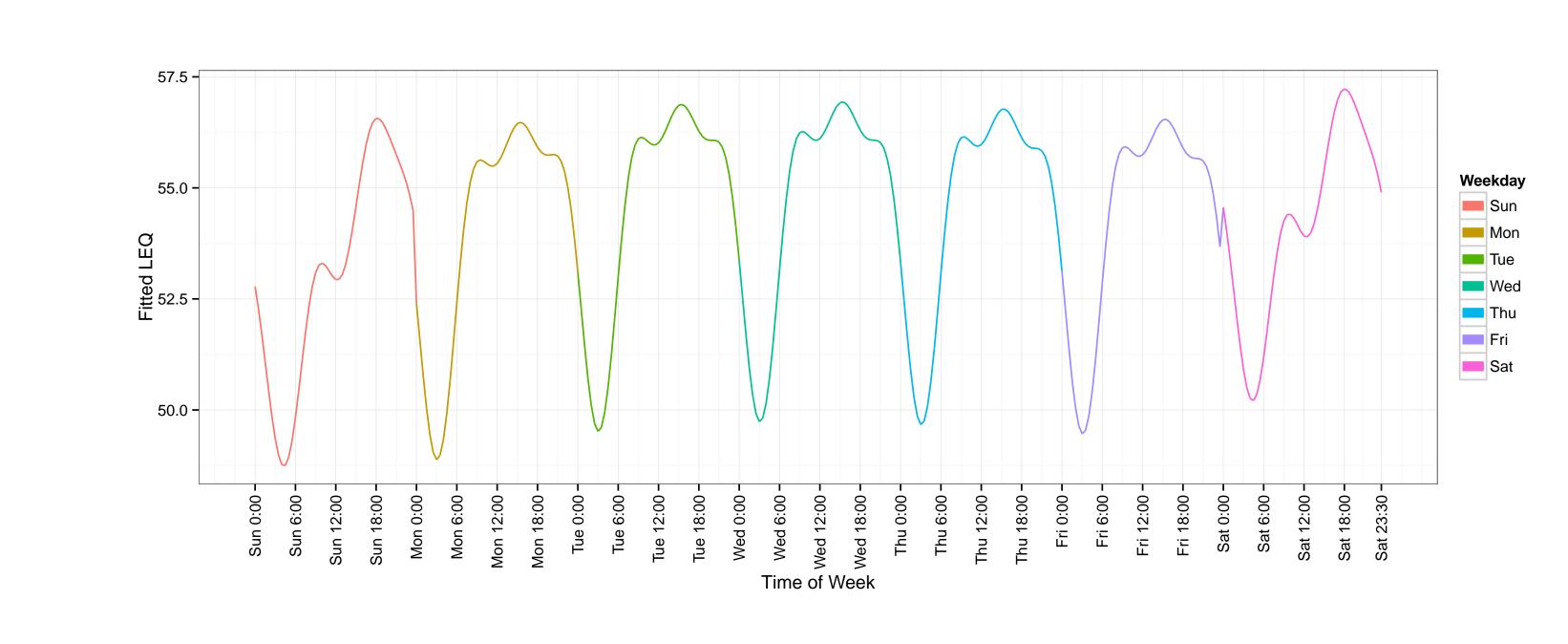
The components (periods) of $d_j(t)$ were chosen based on a spectral analysis on all sites time-series individually. Then, the combination of components and the order of the polynomial are determined by significance and AIC. Finally, the choice of time point baseline reference (t=0 on Sunday midnight) is chosen by AIC. The resulting trend will come from the fixed effects which will look like $d_1(t) + w(t)$ for weekdays and $d_1(t) + d_2(t) + w(t)$ for weekends.

Describing temporal variations

The isolated harmonic functions $d_1(t)$ and $d_1(t) + d_2(t)$ are shown below as well as the overall weekly trend. We see that for weekdays, noise is fairly constant throughout work hours, rising and falling at both rush hours, with a peak at 3:30 PM; on weekends noise levels rise later in the day, and higher at that, peaking at 6:00 PM.



The week-long trend shown below illustrates that Wednesday is the loudest weekday while Saturday is the loudest day of the week. The model was able to fit approximate the observed levels, as seen by the residuals boxplots.

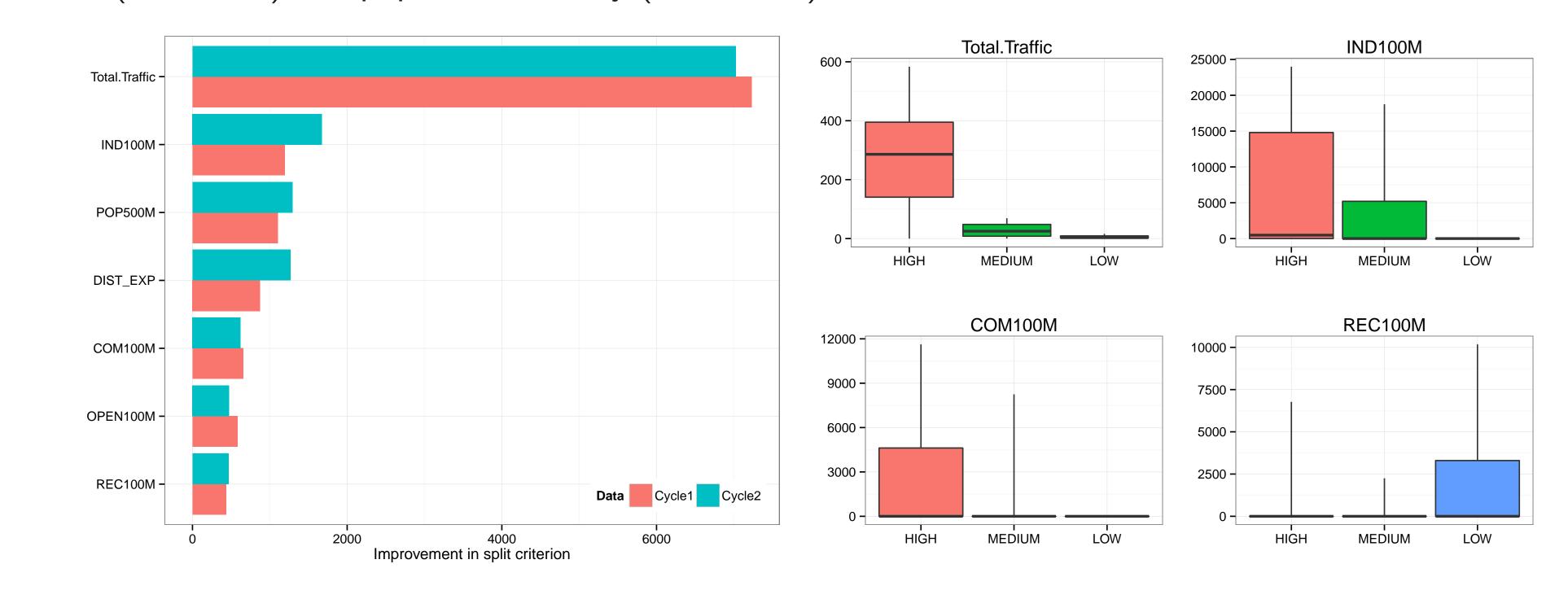


Site Characteristics

The site characteristics can potentially obscure the spatial pattern in the LEQ noise process, thus the need for identifying the important site characteristics.

Random Forest Regression The random forest regression implicitly models the interaction effects and it can identify important variables.

The variable importance plot shows that the traffic count is the most important variable followed by industrial land use (IND100M) and population density (POP500M).



Dirichlet Process Mixture Models

We wish to cluster the sites based on the noise levels and identify the characteristics for each cluster, but the number of clusters K is unknown. An appropriate method in this framework is the Dirichlet Process Mixture Models.

Obtained three clusters, which we label as HIGH, MEDIUM, and LOW noise. High level of traffic count and IND100M for the HIGH cluster and MEDIUM noise sites. Recreational land use (REC100M) is high for LOW noise sites.

Describing spatial variations

Question: Describe the spatial variations; are there locations in Toronto experiencing higher noise levels?

It is assumed that data $Z(s)=(Z(s_1),\ldots,Z(s_n))$ are a realization of a spatial process $\{Z(s):s\in\mathbb{R}^2\}$. For every site location s_i , the residuals of the following regression model, denoted as $R(s_i)$, retain the spatial variability of the process but some of it has been removed, as a result of the external information used in the modelling.

$$Z(s_i) = \mu(s_i) + arepsilon(s_i), \ arepsilon(s_i) \sim N(0, \psi^2) \ \mu(s_i) = X(s_i)^ op eta$$

These residuals are then kriged by using Ordinary Kriging assuming a Gaussian Process with Matérn covariance model with $\kappa=1.5$ as described below.

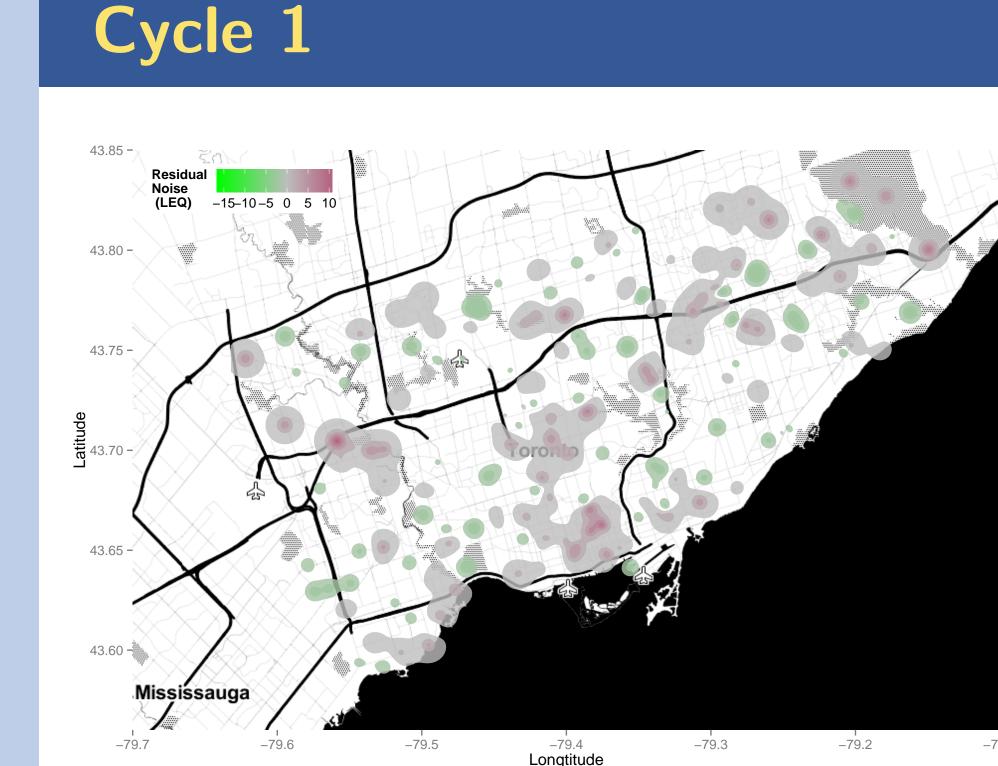
$$R(\mathrm{s}) \sim N(\mu_0, au^2 I + \sigma^2 \Sigma(\phi)).$$

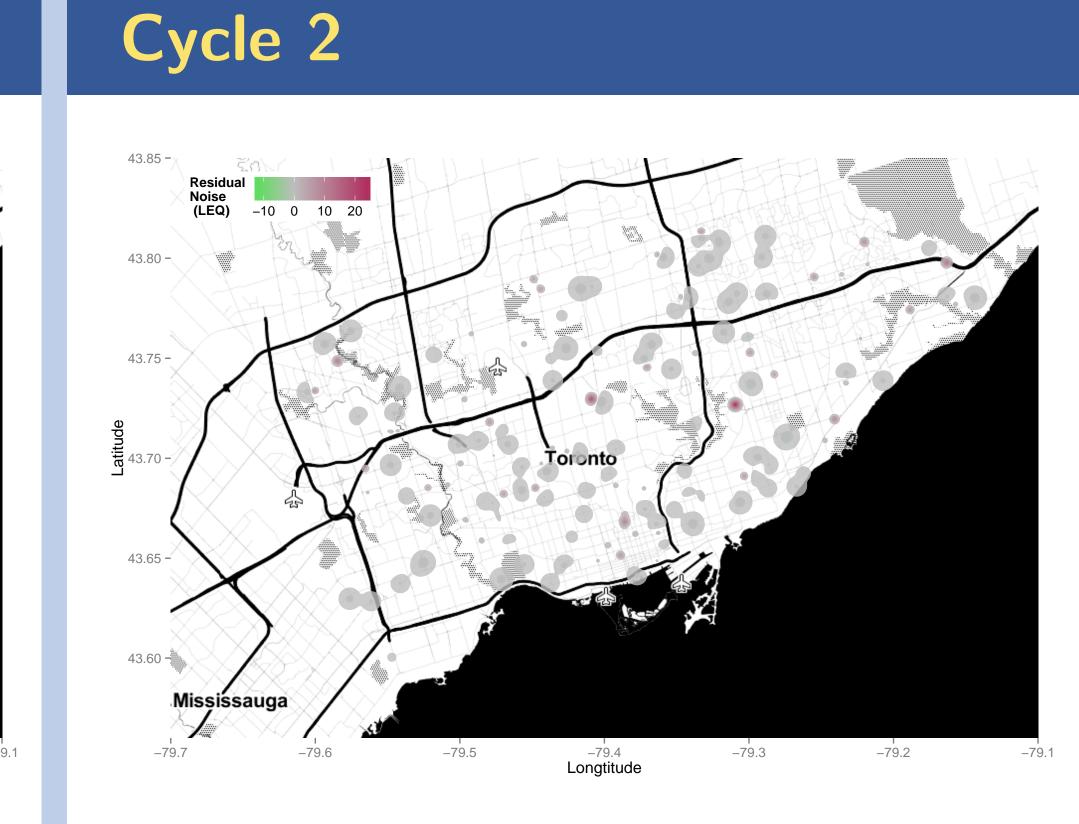
The objective of the kriging procedure is to obtain predictions at unsampled locations using information available elsewhere in the study domain.

Describing spatial variations

Question: Are the spatial variations of Cycle1.rep similar to Cycle1?

- ▶ If Cycle1 and Cycle1.rep follow the same spatial pattern in terms of "unexplained noise", it is likely the kriging prediction of Cycle1.rep obtained from the residual kriging using Cycle1 is somewhat a fair prediction of the noise after adjusting the predictors.
- ➤ We observe a good agreement for the two sets of residuals, so we augment Cycle1 with Cycle1.rep to provide calibrated predictions on Toronto. These are mapped in the first contour plot to show areas of high residual noise, such as downtown and near the Pearson airport, as expected.





Describing spatial variations

Question: Cycles 1 and 2 similarity

- ► The same methods as described above for Cycle1–linear model with chosen covariates and residual kriging—is implemented on Cycle2 to obtain a contour plot for residual noise in Toronto.
- ➤ A large porportion of LEQ variation between sites in Cycle 2 can be explained by site characteristics similarly to Cycle 1. Random forest shows nearly identical results for the importance of site characteristics, so the cause of spatial noise variation is comparable.
- ➤ When employing residual kriging, we obtain the contour map shown above which is dissimilar to that obtained with Cycle 1. This is likely due to the complementarity of the site placements coupled with the low long-distance prediction capability of this particular spatial model.