Urban Environmental Noise Pollution in Paris Using Volunteered Geographic Information (VGI): Spatial Regression Models

By: Juan Nathaniel, Wei Ziyi
Supervisor: Dr. Yan Yingwei

Introduction

Noise pollution has become an increasing concern for public health and the general welfare of communities living in large urban areas, such as Paris. Existing monitoring strategy for noise pollution often relies on taking individual & isolated measurement from key points, which are then interpolated to produce a continuous noise map. This method, although is cost-effective, ignores the possible interaction of noise in one area with noise in its surrounding (e.g. effects of clustering or diffusion). In order to then examine the spatial effects of urban noise, large amount of measurement needs to be taken at a spatial interval meaningful enough to imitate the effect of interpolation. This is where the use of Volunteered Geographic Information (VGI) can be useful to accomplish the goals of this study.

Research Objectives

1. Visualize urban noise level in Paris using VGI data in the year 2017.
2. Examine the presence of any spatial phenomenon from the generated noise map (i.e. spatial autocorrelation) using a suitable neighbourhood structure (i.e. weight matrix).
3. Explore significant variables that can explain the reported noise level & its observable spatial pattern using several spatial regression models: (a) standard regression model (b) spatial lag model (c) spatial error model.
4. Evaluate the resulting models and diagnose them for further improvements.

Materials & Methods

Source of Data

1. VGI noise level: NoiseCapture by noise-planet, an organization which hosts and collates user-captured noise level.
2. Demographic, land-use, and transportation networks: open data from France urban planning authority, APUR.

Spatial Autocorrelation

Spatial autocorrelation is a measure between two values of an attribute that are nearby spatially. Positive spatial autocorrelation suggests that similar values aggregate together while negative spatial autocorrelation suggests otherwise. Moran's I statistic is generally used to measure spatial autocorrelation by measuring the correlation between the dependent variable (y) and the weighted average of y around its neighbors (Wy). This can be interpreted as the slope of the regression between Wy against y (illustrated as Moran scatter plot in **Fig 4**). Neighbors are defined based on either distance or contiguity criteria in a weight matrix (see **Figure 1 & 2**).

Standard Ordinary Least Squares (OLS) Model

A standard linear regression model assumes that each observation of y and the resulting error terms are independently, identically, and normally distributed. However, when spatial autocorrelation exists, the assumption of independent observation is violated. This suggests the presence of spatial patterns among individual noise level. OLS regression has the following general formula:

$$Y = \beta X + \varepsilon$$

Where Y denotes vector of the response variable, X the matrix of explanatory variables, β the matrix of coefficients, and ε the vector of the error terms.

Spatial Lag Model

If the dependent variable, y, is significantly correlated with its neighbor, spatial lag model, rather than linear regression model, can be used by incorporating Wy as an additional variable and using ρ as its coefficient vector in the following manner:

$$Y = \beta X + \rho W y + \varepsilon$$

Spatial Error Model

Alternatively, if our initial linear model generates large residual, then it might be useful to evaluate whether this "unfitness" of the model (i.e. error) exhibits spatial autocorrelation also. If it does, then it might be the case where incorporating other spatially autocorrelated independent variables might improve the overall predictive power of the regression model. Spatial error model is parameterized as follows:

$$Y = \beta X + u,$$
$$\text{where } u = \rho W u + \varepsilon$$

Similar to spatial lag model, Wu is the weighted average of the error term, ε, within a defined neighborhood weight matrix.

Results

Spatial Autocorrelation

Figure 1: Moran I's Statistics for Spatially Lagged Noise Level

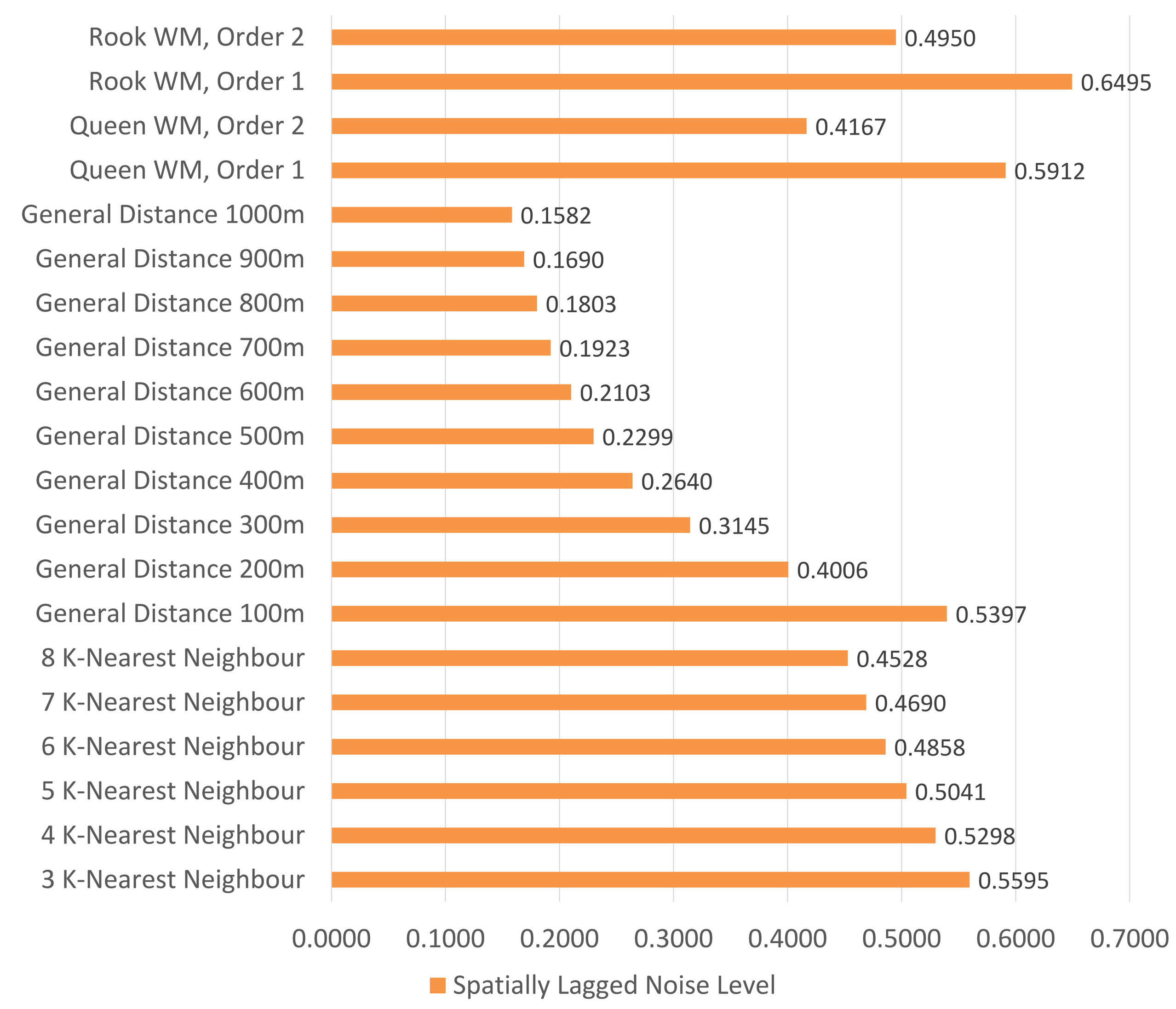
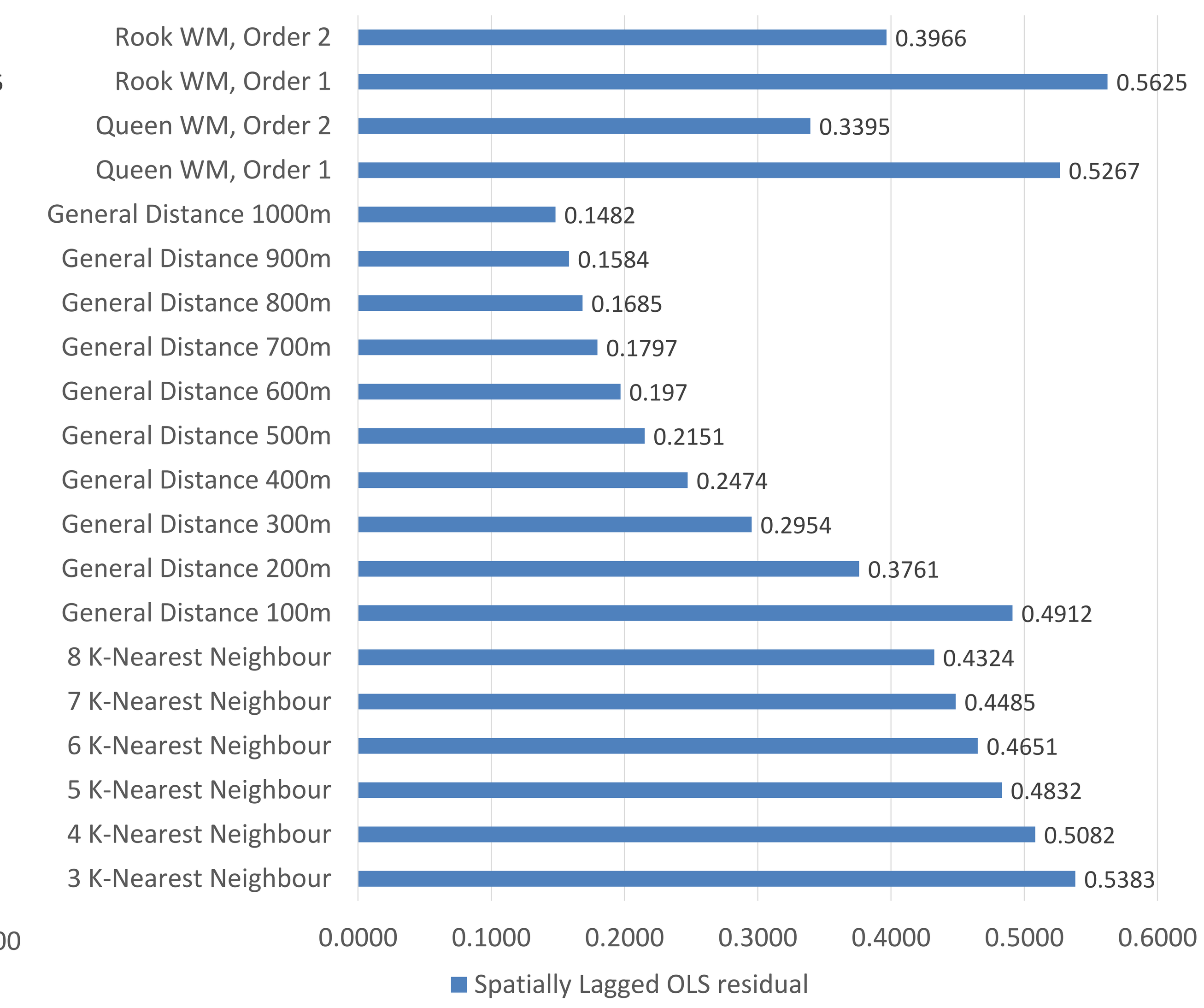


Figure 2: Moran I's Statistics for Spatially Lagged OLS Residual



Note: (a) All weight matrices are significant at $p < 0.001$;
(b) Rook, Order 1 has the highest Moran's I statistic for spatially lagged 'Noise Level' variable and its OLS residual. Hence, we will use Rook Order 1 Weight Matrix as our definition of neighbors in the subsequent spatial regression models.

Data Exploration

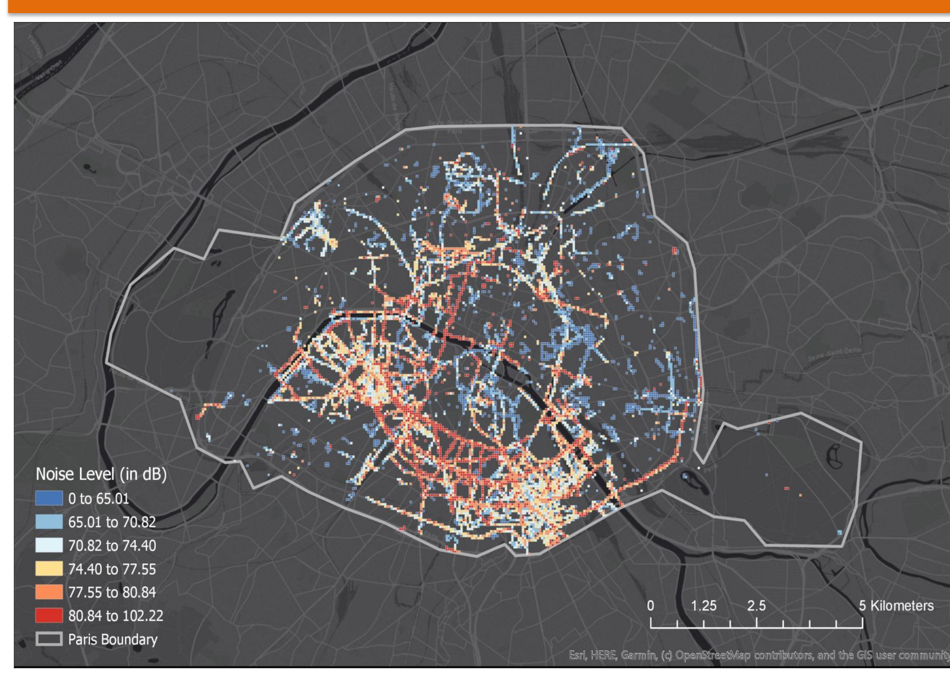


Figure 3: VGI Noise Level Quantile Map (in dB) where blue indicates low noise level and red high noise level

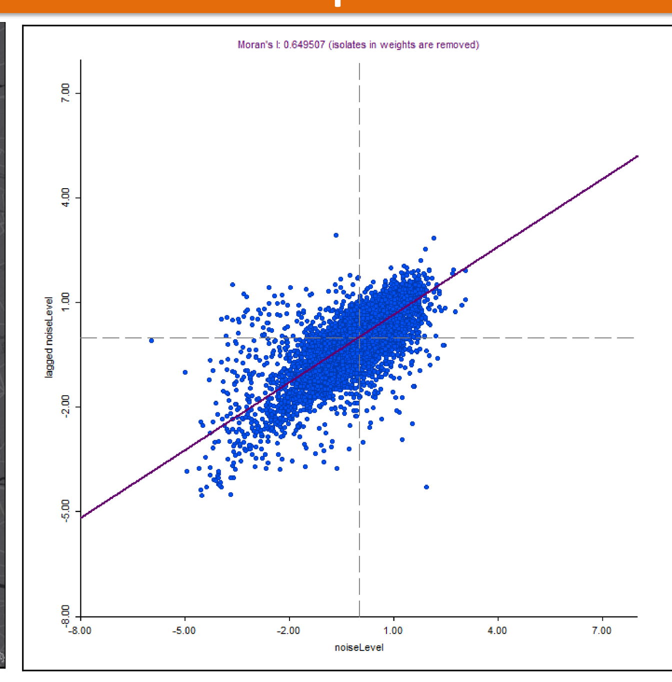


Figure 4: Rook order 1's Moran scatter plot (spatially-lagged noise level against noise level)

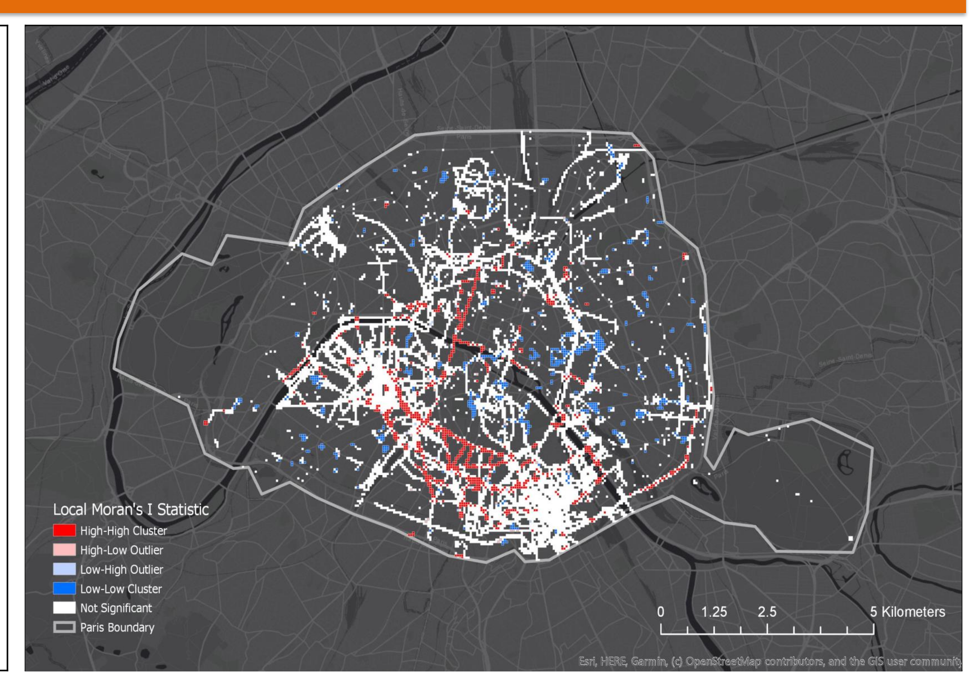


Figure 5: Local hotspot analysis (*Blue*: low noise level neighborhood; *Red*: high noise level neighborhood.

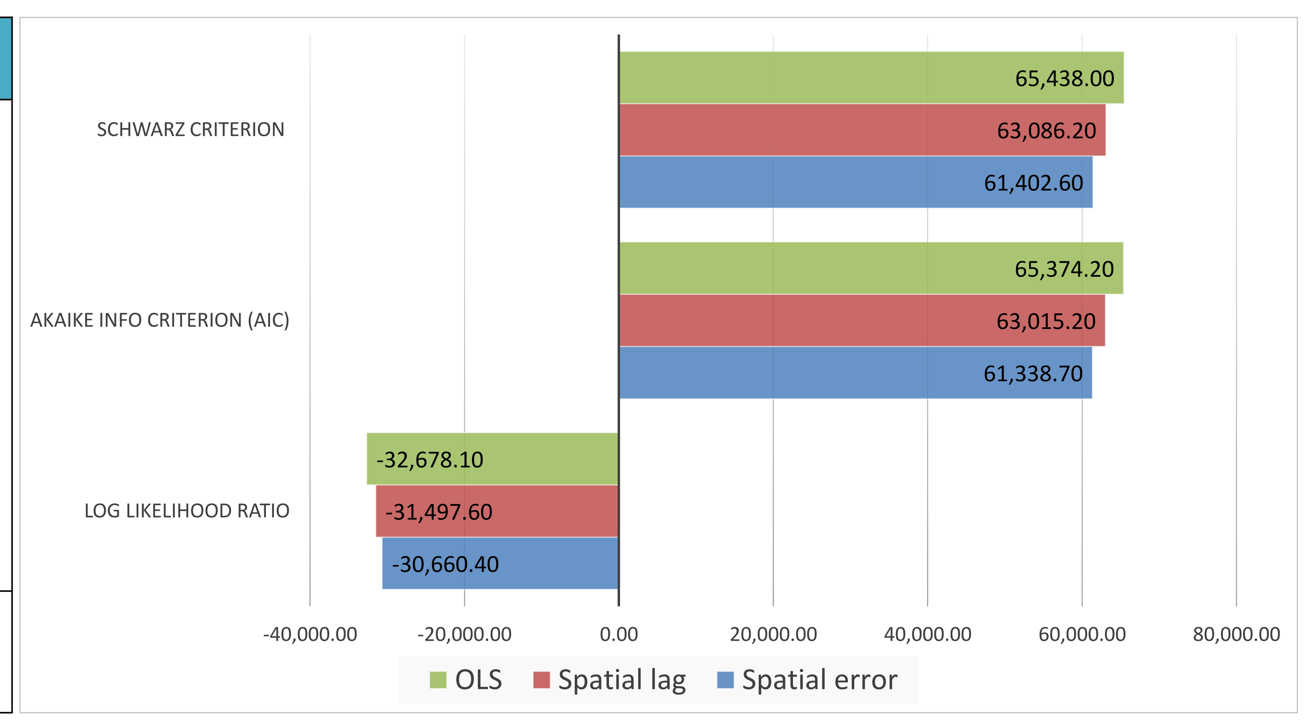
Spatial Regression Models

Table 1: OLS, Spatial Lag, and Spatial Error Regression Models

Variables	OLS	Spatial lag	Spatial error
Residential Use⁺	1.61174	0.9562	-0.2379
Dense Mixed Use⁺	1.46983***	1.0052***	0.7379***
Natural Space Use⁺	-6.5037***	3.1763	-7.5414***
Population Count	-0.0036***	-0.0022***	-0.0028***
Attraction Count	0.5038	0.1072	-0.0643
Public Trans Length	0.0329***	0.0221***	0.0184***
Rail Length	-0.0048***	-0.0019	-0.0014
Primary Road⁺	2.1591***	1.4618***	1.3929***
Spatial lag effect	-	0.2923***	-
Spatial error effect	-	-	0.6353***

Note: ⁺ Categorical Boolean variable (i.e. True or false)
***significant at p < 0.001 for a two-tail test

Figure 6: Spatial Models Diagnostics



Spatial error model is the better model since it has the highest log-likelihood ratio, lowest AIC, and lowest Schwarz Criterion.

Discussion & Conclusions

Evaluation of Spatial Regression Models

1. Rook Order 1 neighborhood structure has the highest Moran I's statistics (scatterplot's gradient) for both the spatial autocorrelation of y (Wy against Y) and OLS error (Wu against ε).
2. Hotspot analysis using Local Indicators of Spatial Association (LISA) shows significant clusters of high-high (i.e. high noise level cells surrounded by high noise level cells) in Southwestern side of Paris (industrial zone) while clusters of low-low in the central & western areas (residential zone).
3. (a) Land-use types are significant predictors for noise level. Dense-mixed use is positively correlated while natural space is negatively correlated with noise level. (b) Transportation networks significantly predict for noise level, with both the length of public rails and the presence of primary road exhibiting significant positive correlations.
4. Spatial error regression model fits the data better given the diagnostic tests. **More spatially-explicit independent variables should be considered to improve the model.**

Implication of Study & Conclusion

1. Noise level exhibits spatial autocorrelation: observation in one area influences noise level in another.
2. Errors are still high (AIC, Schwarz test) even when data are fitted with both spatial regression models. More explanatory variables should be considered in the future.
3. In addition, future research can attempt to combine both spatial lag and error models to evaluate whether greater accuracy can be achieved.
4. We can evaluate whether the use of different cell size (our current study uses 50m x 50m fishnet resolution) will produce different result.

Powered By

ArcGIS® Pro