Spatial statistical calibration on linear networks: an application to the analysis of traffic volumes

Presenter: **Andrea Gilardi**[1]

Coauthor: **Riccardo Borgoni**[1]

Coauthor: **Jorge Mateu**[2]

2022-06-02

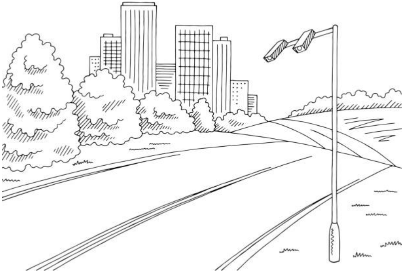
Conference: METMA X

1: University of Milano - Bicocca
2: Universitat Jaume I

Introduction

- The estimation of traffic volumes at the street network level represents a fundamental step in several areas of research (e.g. transport planning and road safety analysis).
- The traditional ways to derive traffic counts involve inductive loops or manual counts from fixed cameras. These figures are usually very accurate.
- However, manual estimates are extremely time consuming, whereas inductive loops are economically expensive and, therefore, geographically sparse.

For example:



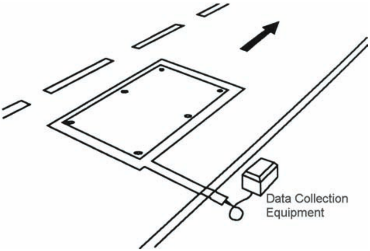


Figure 1: (Fixed Cameras) (Inductive Loops)

Source: Getty Images/iStockphoto by Aluna1 (Left); Traffic Detector Handbook: Third Edition—Volume I by U.S. Department of Transportation (Right).

Introduction (cont)

- In the last years, several agencies developed tools to collect traffic volumes from spatially referenced mobile sensors (e.g. smartphones and sat-navs).
- Mobile data has several advantages such as extremely detailed spatial resolution and extensive coverage.
- However, since not all vehicles are equipped with GPS devices, traffic counts derived from mobile sensors typically underestimate the real flows.

Introduction (cont)

For example:

Source: https://www.tomtom.com/products/road-traffic-data-analytics/

Introduction (cont)

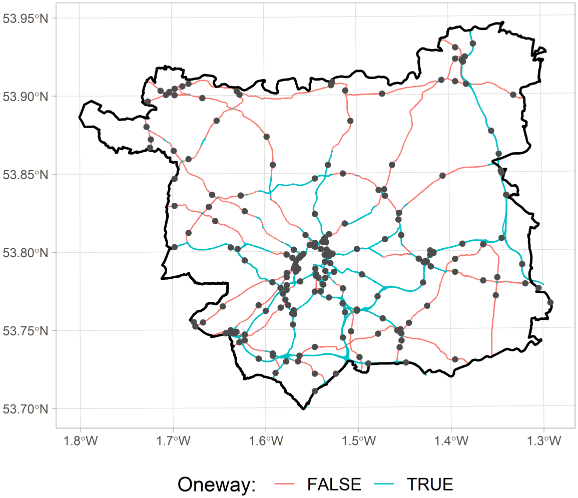
- The objective of this talk is to present a procedure to integrate traffic estimates derived from mobile devices and counts recorded by traditional road-fixed sensors.
- Therefore, we propose a statistical technique to spatially calibrate mobile sensor data using a Geographically Weighted Regression (GWR).
- Moreover, since the traffic flows are a typically example of phenomena occurring on a linear network, the classical GWR was adapted to network data.

Street network and traffic data

- The case study is the metropolitan area of Leeds (UK). Moreover, considering the importance of the recent COVID outbreak, we focused our analysis on the year 2019.
- The street network (18,002 segments covering ∼800km) and the daily traffic estimates at the road segment level from mobile sensors were provided by TomTom.
- On the other hand, we accessed official data shared by the Department for Transport to obtain the Average Annual Daily Flow (AADF) at 364 fixed count-points in Leeds.

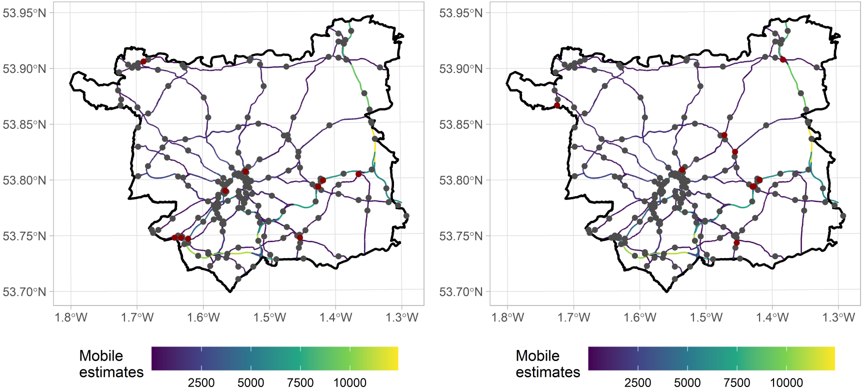
Street network, traffic estimates and fixed count points

The traffic estimates are recorded for each direction of travel, so we will consider a directed linear network:



Street network, traffic estimates and fixed count points (cont)

The two maps show the 2019 AADF estimates in both directions:



Linear Regression Calibration

- Given an imprecise and easy to measure covariate Y and a gold-standard X, the classical linear regression calibration assumes that

  Direct Calibration: $Y = \beta_0 + \beta_1 X + \varepsilon \Longrightarrow \hat{X} = \frac{Y - \hat{\beta}_0}{\hat{\beta}_1}$;
  Inverse Calibration: $X = \alpha_0 + \alpha_1 Y + \varepsilon \Longrightarrow \hat{X} = \hat{\alpha}_0 + \hat{\alpha}_1 Y$;

- The quantities $\alpha_0, \alpha_1, \beta_0, \beta_1$ are parameters, whereas ε is a random error.

- In both cases, the objective is to calibrate the imprecise measurements Y using the gold standards X.

GWR

- The road traffic has an intrinsic spatial nature and the relationship between mobile estimates and fixed-points counts may change according to the spatial location.

- We thus propose a spatial calibration approach based on *Geographically Weighted Regression* (GWR).

- Given a sample of n points, the GWR model reads as

$$Y(\boldsymbol{s}_i) = \beta(\boldsymbol{s}_i)' \boldsymbol{X}(\boldsymbol{s}_i) + \varepsilon_i$$

  where $Y(\boldsymbol{s}_i)$ and $\boldsymbol{X}(\boldsymbol{s}_i)$ denote the response variable and the covariate(s) at location $\boldsymbol{s}_i$, respectively. $\beta(\boldsymbol{s}_i)$ is a vector of spatially varying parameters.

GWR (cont)

- The parameter estimation at location $\boldsymbol{s}_j$ on the network is carried out using locally weighted least squares:

$$\hat{\beta}(\boldsymbol{s}_j) = [X'(\boldsymbol{s}_j)W(\boldsymbol{s}_j)X(\boldsymbol{s}_j)]^{-1}X'(\boldsymbol{s}_j)W(\boldsymbol{s}_j)Y(\boldsymbol{s}_j),$$

  where $\boldsymbol{W}(\boldsymbol{s}_j)$ is a diagonal local weighting square matrix.

- The weights were determined using an exponentially decaying function: $w_{ij} = exp(-d_{ij}^2/2h)$. d_{ij} denotes the shortest-path distance on the directed network between (centroid of the) segment i and location j.

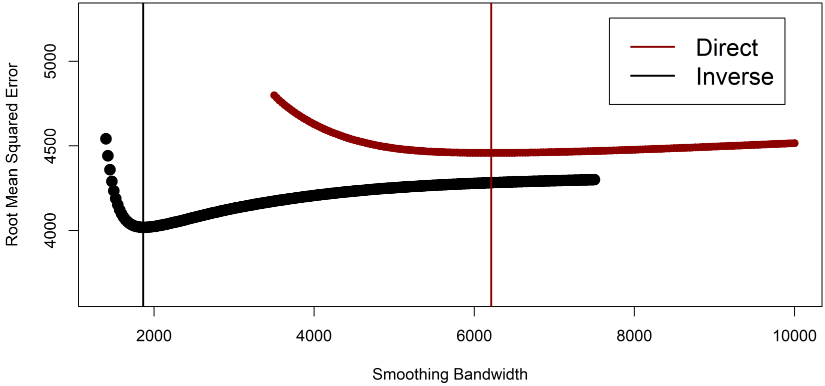
- The bandwidth parameter h was determined by cross-validation minimising the Root Mean Squared Error.

Spatial Regression Calibration

- The inclusion of GWR into a direct calibration approach can be performed as follows:
  1. considering the local model
     $$Y(s_i) = \beta_0(s_i) + \beta_1(s_i)X(s_i) + \varepsilon(s_i),$$
     estimate the spatially varying parameters;
  2. invert the local equation at any location on the network: $\hat{X}(s_i) = (Y(s_i) - \hat{\beta}_0(s_i))/\hat{\beta}_1(s_i)$
- Similarly, the spatial inverse calibration can be performed by estimating a local regression $\hat{X}_i(s_i) = \hat{\alpha}_0(s_i) + \hat{\alpha}_1(s_i)Y(s_i)$ and use it to predict the gold-standard values.

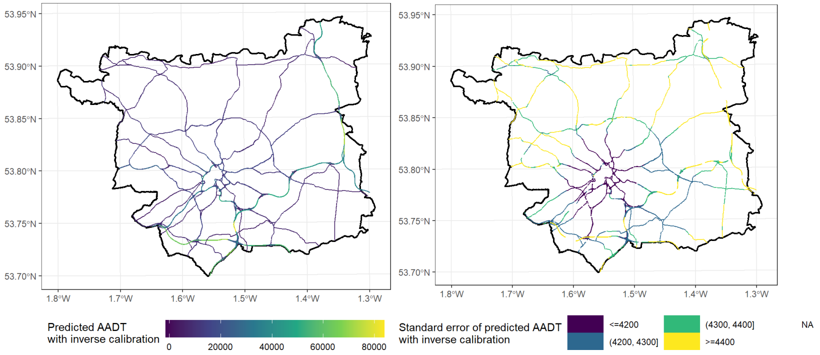
Optimal Bandwidth

Optimal bandwidths for the two types of calibration:



Traffic flows predictions

Predicted gold-standard for each street segment in the network and standard errors of the predictions.

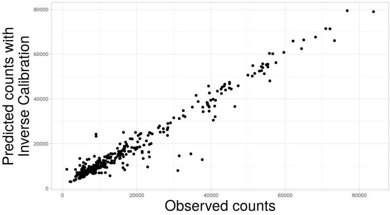


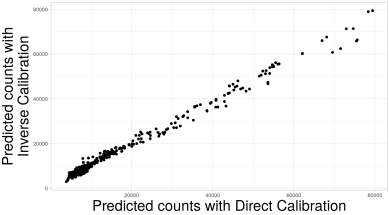
We obtain similar predictions with direct calibration.

Traffic flows predictions

We tested the spatial stationarity of the coefficients using a montecarlo test. We found a strong spatial variability.

Moreover, we compared predicted and observed flows at fixed point cameras locations:





Future works

Now we are working on:

- deriving a procedure to estimate the predictions' standard errors for spatial direct calibration;
- extending the suggested methodology to more sophisticated statistical approaches such as multivariate calibration or truncated calibration;
- exploring the relevance of border effects for the estimation.

Conclusions

- In this talk, we introduced a spatial extension of classical statistical calibration techniques considering direct and inverse calibration.
- In both cases, the idea is to re-align an imprecise measurement (i.e. the tomtom figures) using expensive and precise counts (i.e. the fixed count-points), also taking into account the spatial variation.
- In fact, this is also one of the first statistical application adopting mobile devices to re-calibrate road traffic counts.

References

Marc Barthélemy. “Spatial networks”. In: *Physics reports* 499.1-3 (2011), pp. 1–101.

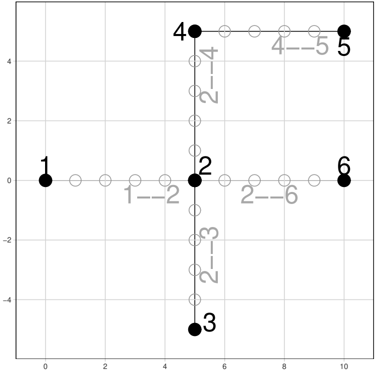
A. Stewart Fotheringham, Chris Brunsdon, and Martin Charlton. *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons, 2003.

Christine Osborne. “Statistical calibration: a review”. In: *International Statistical Review/Revue Internationale de Statistique* (1991), pp. 309–336.

Abbas Sheikh Mohammad Zadeh and Mohammad Ali Rajabi. “Analyzing the effect of the street network configuration on the efficiency of an urban transportation system”. In: *Cities* 31 (2013), pp. 285–297.

Junwei Zeng, Yongsheng Qian, Bingbing Wang, Tingjuan Wang, and Xuting Wei. “The impact of traffic crashes on urban network traffic flow”. In: *Sustainability* 11.14 (2019), p. 3956.

Spatial Linear Networks



Spatial Component

```
Simple feature collection with 5 features
geometry type: LINESTRING
dimension: XY
1 LINESTRING (0 0, 1 0, ...)
2 LINESTRING (5 0, 5 -1, ...)
3 LINESTRING (5 0, 5 1, ...)
4 LINESTRING (5 5, 6 5, ...)
5 LINESTRING (5 0, 6 0, ...)
```

Graph Component

```
6 vertices
[1] 1 2 3 4 5 6
5 edges
[1] 1--2 2--3 2--4 4--5 2--6
```