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

A. Gilardi[1,*], R. Borgoni[1] and J.Mateu[2]

[1] *Full address of first and second author; andrea.gilardi@unimib.it, riccardo.borgoni@unimib.it*
[2] *Full address of third author; mateu@uji.es*
[*] *Corresponding author*

Abstract. *The estimation of traffic volumes on a street network represents a fundamental step to improve transport planning protocols and develop effective road safety interventions. The traditional ways to derive traffic figures involve manual counts or fine-tuned automatic tools (e.g. cameras or inductive loops). Unfortunately, the manual counts are extremely time consuming, whereas the fixed instruments are typically very expensive and geographically sparse. However, given the increasing availability of mobile sensors (e.g. smartphones and GPS sat-nav), in the last years we observed a surge of methods to infer traffic counts from geo-referenced mobile devices. This paper proposes a spatial statistical calibration technique to combine accurate fixed counts and extensive GPS mobile data for the estimation of traffic flows, re-adapting the statistical methods to the spatial network context. The suggested methodology is exemplified using data collected in the City of Leeds (UK).*

Keywords. *Geographical weighted regression; Spatial networks; Statistical Calibration; Traffic flows*

---

1 Introduction

The estimation of traffic volumes on a street network represents a critical issue to improve transport planning protocols and develop effective road safety interventions. In fact, the traditional ways to produce traffic figures typically involve manual counts with ad-hoc cameras or automatic counts with road-fixed sensors (e.g. inductive loops and spirals). Unfortunately, both techniques have several limitations linked to their limited spatial coverage and high economical costs of installation and maintenance. More recently, traffic information have been collected by geo-referenced mobile sensors (e.g. smartphones and sat-navs) using ad-hoc models. These mobile sensors have several advantages, such as extremely detailed spatial resolution and extensive spatial coverage. However, since not all vehicles are equipped with GPS devices, traffic counts from mobile sensors typically underestimate the real flows.

The precision of the mobile information can be improved by integrating mobile figures using more traditional road-fixed sensors. This allows one to calibrate extensive mobile measurements using more accurate data. We propose a statistical technique to spatially calibrate mobile sensor data using geographical weighted regression (GWR). Being traffic flows a classical example of a phenomenon occurring in a spatial network, the usual GWR was modified to take into account the spatial domain.

2 Spatial calibration by geographical weighted regression

The term *statistical calibration* represent a series of techniques adopted in several research fields to adjust the values of one measurement, say X, using some other measurements, say Y. The need for calibration arises when X is more expensive or more difficult to measure than Y or when the values of X are not recorded and cannot be retrieved [3].

We consider an absolute calibration problem where X is assumed to be measured without error. In particular, X represents the traffic flows measured by the fixed traffic cameras installed on a restricted number of segments of the city network, whereas Y represents the traffic counts derived from mobile sensors installed on cars and available on each street segment. The next two sections will briefly introduce the regression calibration problem and present its spatial re-adaptation in this context.

2.1 Regression Calibration

In a linear regression calibration context, it is assumed that, given a sample of n pairs of observations, the relationship between an imprecise measure Y and a gold-standard or reference value X has the following functional form: $Y_i = f(X_i) = \beta_0 + \beta_1 X_i + \varepsilon_i,\ i = 1,\dots,n$ where ε_i represents the measurement error of the less precise instrument. Two main approaches have been developed when Y is calibrated on X, the so called *classical* and *inverse calibration*.

The classical calibration technique is articulated in two steps. In the first step, Y is regressed on X to obtain the estimated model $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$. Then, the estimated regression model is inverted to obtained predictions of X, i.e. $\hat{X} = (Y - \hat{\beta}_0)\hat{\beta}_1^{-1}$, using known experimental values of Y for each unit where X is not available. On the other hand, in the inverse calibration approach, X is regressed on Y and the values of X are predicted from the estimated model, i.e. $\hat{X} = \hat{\alpha}_0 + \hat{\alpha}_1 Y$, for the units where the gold-standard is not available.

2.2 Spatial calibration via GWR

The analysis of spatial data typically requires ad-hoc adjustments to take into account the nature of the spatial domain. In fact, the relationship between the traffic flows measured by fixed cameras and mobile devices can change according to the spatial location. Hence, a global calibration as described in the previous section is not appropriate and a more local approach would seem preferable.

We thus propose a spatial calibration approach based on geographical weighted regression. GWR is a local form of spatial analysis that allows the estimation of relationships between a dependent variable and a set of predictors that vary over space [2]. More precisely, given a sample of n units in a region S observed at locations $\mathbf{s}_i,\ i = 1,\dots,n$, the GWR model reads as

$$Y(\mathbf{s}_i) = \beta(\mathbf{s}_i)'\mathbf{X}(\mathbf{s}_i) + \varepsilon(\mathbf{s}_i), \quad i = 1,\dots,n, \tag{1}$$

where $Y(\mathbf{s}_i)$ denotes the response variable, $\mathbf{X}(\mathbf{s}_i)$ a column vector of explanatory covariates with a column of constant unitary values that represent the intercept, $\beta(\mathbf{s}_i)$ the corresponding spatially-varying coefficients, and $\varepsilon(\mathbf{s}_i)$ is a zero mean random error.

Parameter estimation at a selected location $\mathbf{s}_j \in S$ is carried out using locally weighted least squares

$$\hat{\beta}(\mathbf{s}_j) = \left[\mathbf{X}'(\mathbf{s})W(\mathbf{s}_j)\mathbf{X}(\mathbf{s})\right]^{-1}\mathbf{X}(\mathbf{s})'W(\mathbf{s}_j)Y(\mathbf{s}_j), \tag{2}$$

where $W(\mathbf{s}_j) = diag(w_{1j}, \ldots, w_{nj})$ is a local weighting square matrix with entry w_{ij} giving the weight associated to unit i when the regression is estimated at location $\mathbf{s}_j$, and $\mathbf{X}(\mathbf{s})$ represents the design matrix. The weights are defined in terms of a kernel function K that decays gradually with d_{ij}, i.e. the distance between the ith observation and the point $\mathbf{s}_j$. In particular, a Gaussian kernel function is adopted in this paper: $K(d_{ij}) = \exp\{-d_{ij}^2/2h\}$, where the bandwidth parameter h determines the spatial range of the kernel. In the case study presented in the next section, the value of h is selected using cross-validation by minimising the mean square error of traffic flows predictions.

Re-adapting the calibration equations described before to the GWR framework is actually straightforward. In fact, since the regression coefficients $\hat{\beta}(\mathbf{s}_j)$ depend upon the spatial locations, the GWR permits one to map the variation in the regression parameters and, more importantly, to calibrate the variable of interest taking into account the spatial pattern of the two measures. More precisely, the inclusion of a GWR into a classical calibration approach can be performed as follows. First, we need to estimate a local model that written as

$$\hat{Y}(\mathbf{s}_i) = \hat{\beta}_0(\mathbf{s}_i) + \hat{\beta}_1(\mathbf{s}_i)X(\mathbf{s}_i), \tag{3}$$

and then we calculate the calibrated value of $X(\mathbf{s})$ by inverting the local equation at any desired location $\mathbf{s}$ i.e. $\hat{X}(\mathbf{s}) = (Y(\mathbf{s}) - \hat{\beta}_0(\mathbf{s}))(\hat{\beta}_1(\mathbf{s}))^{-1}$.

Similarly, the spatial inverse calibration can be performed by estimating a local regression

$$\hat{X}(\mathbf{s}_i) = \hat{\alpha}_0(\mathbf{s}_i) + \hat{\alpha}_1(\mathbf{s}_i)Y(\mathbf{s}_i) \tag{4}$$

which can be used to predict $X(\mathbf{s})$ at any desired location $\mathbf{s}$.

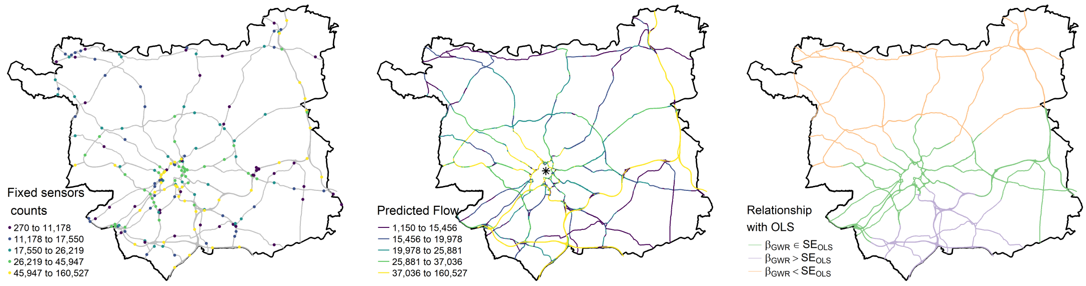
Finally, we note that the usual distance metric adopted in a spatial regression context is the Euclidean distance e.g. $d_{ij} = ||\mathbf{s}_j - \mathbf{s}_i||$. However, the problem that motivated the analysis presented in this paper develops on a one-dimensional linear domain and the sample unit considered below is a road segment represented by its centroid $\mathbf{s}_i$. Therefore, we argue that the distances d_{ij} should be calculated preserving the graph structure of the network. More precisely, indicating by $L = (V, E)$ the one-dimension graph object representing a street network (where V and E denote the sets of vertices and edges, respectively), a path ρ_{ij} connecting any two generic locations $\mathbf{s}_i$ and $\mathbf{s}_j$ on the network is defined as a finite sequence $\{\mathbf{p}_m\}_{m=1}^{M}$ of adjacent vertices in V such that the edges with endpoints $[\mathbf{s}_i, \mathbf{p}_1]$ and $[\mathbf{p_M}, \mathbf{s}_j]$ belong to E. The length of ρ_{ij} can be computed as

$$||\mathbf{s}_i - \mathbf{p}_1|| + \sum_{m=1}^{M-1} ||\mathbf{p}_{m+1} - \mathbf{p}_m|| + ||\mathbf{p}_M - \mathbf{s}_j||,$$

and we define d_{ij} as the minimum length of all paths connecting $\mathbf{s}_i$ and $\mathbf{s}_j$ [1].

3 Results and Conclusions

The statistical methods presented in this paper were exemplified considering fixed and mobile traffic data recorded in the street network of Leeds (UK) from January to December 2019. The road network and the mobile traffic counts were obtained from TomTom Move provider (*https://move.tomtom.com/*). In



(a) (b) (c)

Figure 1: (a) The segments represent the road network of Leeds, whereas the coloured dots represent the location of the fixed cameras and the corresponding traffic figures; (b) Predicted traffic flows for all segments in the network using an inverse spatial calibration. The black star denotes the city centre; (c) Comparison between classical and spatial calibration in terms of estimated slope coefficient.

particular, the network is composed by 8959 geo-referenced segments that are associated to traffic volumes estimated using mobile devices connected to cars and anonymous GPS-equipped smartphones. On the other hand, the fixed camera counts were derived using data shared by the Department for Transport (*https://roadtraffic.dft.gov.uk/downloads*). The linear network, the fixed cameras and the corresponding traffic estimates are displayed in Figure 1(a).

Figure 1(b) displays the predicted traffic flows obtained using the inverse spatial calibration technique. The figure clearly highlights several roads corresponding to a motorway (i.e. the yellow segments connecting the south area with the north/north-east) and the most important arterial thoroughfares. Similar results were obtained in the classical spatial calibration framework. Finally, we explored the stationarity of the relationship between the two available measurements. Figure 1(c) compares the slope estimates given by a non-local inverse calibration obtained using OLS regression with the suggested extension. The results highlight a non-stationarity in the relationship between mobile and fixed counts. The Pearson correlation coefficient between observed and calibrated counts (obtained by a leave-one-out approach) were found equal to 0.956 (classical calibration) and 0.965 (inverse calibration).

We plan to extend the analysis presented in this paper in several directions. In fact, there exists a vast literature on classical and inverse statistical calibration problems (see [3]) that can be extended to the problem at hand. In particular, we will focus on a) deriving estimates of the standard errors of the regression coefficients; b) comparing classical and inverse spatial calibrations using a variety of criteria (e.g. MSE or consistency); c) exploring more sophisticated statistical methods such as multivariate calibration (joining information from different providers), robust GWR and truncated calibration.

References

[1] Baddeley, A., Nair, G., Rakshit, S., McSwiggan, G. and Davies, T.M. (2021). Analysing point patterns on networks - A reivew. *Spatial Statistics* **42**, p. 100435

[2] Fotheringham, A.S., Brunsdon, C. and Charlton, M. (2003) *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons

[3] Osborne, C., 1991. Statistical calibration: a review. *International Statistical Review/Revue Internationale de Statistique*, pp.309-336.