Locally accurate prediction standard errors with spatially varying regression coefficient models
A comparison of techniques

Paul Harris and Stewart Fotheringham
National Centre for Geocomputation
University of Ireland Maynooth
Maynooth, Co. Kildare, Ireland
{paul.harris, stewart.fotheringham}@num.ie

Chris Brunsdon
Department of Geography
University of Leicester
Leicester, LE1 7RH, UK
cb179@le.ac.uk

Abstract—This study assesses the prediction and prediction uncertainty performance of models that cater for both: (i) a nonstationary relationship between the response and a contextual variable and (ii) a nonstationary residual variance (or variogram), at point locations for a single realisation spatial process. Here the crucial aspect of the model specification is allowing the residual variance (or variogram) to vary across space. Without this, the estimated prediction standard errors are only likely to be accurate in a global (or overall) sense and not the desired, local sense. Locally-accurate prediction standard errors, allow locally-relevant prediction confidence intervals and/or locally-relevant estimates of risk (e.g. the risk of exceeding some critical threshold) which is not only valuable to researchers who attempt to model spatial processes, but also to policy makers who need to plan and manage the outcomes of spatial processes at different spatial scales.

Keywords: geographically weighted regression, moving window kriging, heteroskedastic, bayesian prediction models

1. INTRODUCTION

Models involving spatially varying regression coefficients have been used in the analysis of geographical data for some time now. Here our focus is on the geographically weighted regression (GWR) (Brunsdon et al. 1996) model, but other approaches, such as the Bayesian spatially varying coefficient (SVC) model of Gelfand et al. (2003) provide an alternative methodology. In all cases, the aim is to calibrate a model of the form:

\[ z_i = \beta_0(x_i) + \sum_{k=1}^{K} \beta_k(x_i) y_{ik} + r_i \]

where \( z_i \) is the dependent variable with \( i = 1, \ldots, n \); \( y_1, y_2, \ldots, y_k \) are \( k \) independent covariates; \( \beta_k(x_i) \) is a realisation of the continuous function \( \beta_k(x) \) at sample location \( i \); \( x \) is any spatial location (observed or unobserved); and \( r_i \) is a random error (or residual) term. In GWR, emphasis is typically on the estimation of the local regression coefficients \( \beta_k(x) \) for: (a) use as an exploratory tool, or (b) the provision of predictions for the dependent variable. In many applications, it is equally as important to provide information about the uncertainty of such predictions, and it is this aspect of modelling with spatially varying coefficients that we focus on in this study. In particular, we empirically investigate, the multiple linear regression (MLR) model, two GWR models and three universal kriging (UK) models, as set out in Table 1. Our intention is not to show that a GWR model is superior to a UK model, but that it may provide an alternative that in some circumstances is worth consideration for other reasons.

<table>
<thead>
<tr>
<th>TABLE 1. MODELS CONSIDERED IN THIS STUDY</th>
</tr>
</thead>
<tbody>
<tr>
<td>MODEL</td>
</tr>
<tr>
<td>-----------------------------------------</td>
</tr>
<tr>
<td>1. MLR</td>
</tr>
<tr>
<td>2. GWR</td>
</tr>
<tr>
<td>3. HETEROSEDASTIC GWR (H-GWR)</td>
</tr>
<tr>
<td>4. UK IN A GLOBAL NEIGHBOURHOOD (UK-GN)</td>
</tr>
<tr>
<td>5. UK IN A LOCAL NEIGHBOURHOOD (UK-LN)</td>
</tr>
<tr>
<td>6. MOVING WINDOW UK (MW-UK)</td>
</tr>
</tbody>
</table>

Models can also be grouped into pairs, where: (i) MLR and UK-GN are benchmark models, which assume that all \( \beta_k(x) \) functions are constant across space; (ii) GWR and UK-LN let \( \beta_k(x) \) vary across space; and (iii) H-GWR and MW-UK allow both data relationships and residual variation to vary across space. For H-GWR, the residual variation like the regression coefficients, is a function of location; and for MW-UK, residual spatial dependence is a function of location (i.e. local variograms are specified, unlike UK-GN and UK-LN which specify a single global variogram only). This entails that only H-GWR and MW-UK are expected to provide locally-accurate prediction standard errors. Furthermore, for processes where relationships clearly vary across the space, the four models that account for this (GWR, H-GWR, UK-LN and MW-UK) should all predict more accurately than those that do not (MLR and UK-GN).

Once the performance of our six study models is empirically investigated using an example house price data set, we turn our attention to corresponding models that benefit
from a Bayesian view of uncertainty, via Monte-Carlo Markov Chain (MCMC) methods. Bayesian models are richer in that they provide a full predictive distribution at a target location; unlike the models described which only provide a prediction and an estimate of its standard error (from which a Gaussian predictive distribution needs to be assumed). Here we do not aim to present and evaluate a complete set of Bayesian model formulations, but just provide an introduction to our current research in this area.

II. STUDY MODELS

At some scale, all six study models can be defined using $Z(x) = m(x) + R(x)$, where the random function $Z(x)$ is decomposed into a mean $m(x)$ and residual $R(x)$ component. Both MLR and GWR model $m(x)$ assuming

$R(x)$ is a stationary random function with $E[R(x)] = 0$ and $\text{VAR}[R(x)] = \Sigma$, where the elements of the diagonal $(n \times n)$ matrix $\Sigma = \sigma_r^2 I$ reflect a pure nugget covariogram (i.e. no spatial autocorrelation). Conversely, UK models $m(x)$ with MLR, but where the elements of $\Sigma$ reflect a structured covariogram $C(h)$, where $h$ is the separation distance vector $h = x_i - x_j$ (i.e. spatial autocorrelation exists). As is standard practice, the elements of $\Sigma$ are actually found from the covariogram $y(h)$ and using the relationship $C(h) = \sigma_r^2 - y(h)$. Accordingly, $\Sigma$ is a function of variogram parameters and can be denoted by $\Theta$, where $\Theta$ is a variogram parameter vector typically consisting of a nugget variance, a structural variance and a correlation range. $\Sigma$

A. MLR and GWR

The MLR model can be written as $Z = Y \beta + R$, where $Z$ is the $(n \times 1)$ dependent data vector, $Y$ is the $(n \times k)$ covariate matrix, $\beta$ is a $(k \times 1)$ vector of unknown parameters, and $R$ is a $(n \times 1)$ residual vector. The ordinary least squares parameter estimates $\hat{\beta} = (Y^T Y)^{-1} Y^T Z$ are used to find the MLR prediction $\hat{z}_{MLR}(x) = y(x)^T \hat{\beta}$ at $x$ (where $y(x)$ is a $(k \times 1)$ vector of covariates at $x$). The corresponding GWR model results in $\hat{\beta}(x) = (Y^T W(x) Y)^{-1} Y^T W(x) Z$ where $W(x)$ is a $(n \times n)$ diagonal matrix of spatial weights. The GWR prediction at $x$ is $\hat{z}_{GWR}(x) = y(x)^T \hat{\beta}(x)$. The MLR and GWR prediction variances at $x$, $\sigma^2_{MLR}(x)$ and $\sigma^2_{GWR}(x)$ are estimated using:

$$\text{VAR} \{ \hat{z}(x) - z(x) \} = \hat{\sigma}_r^2 [I + S(x)]$$

Here an unbiased estimate of the residual variance is taken as $\sigma^2 = \text{RSS}/(n - \text{ENP})$, where RSS is the residual sum of squares and ENP is the effective number of parameters of the MLR or GWR model. Furthermore for MLR,

$$S(x) = y(x)^T [Y^T Y]^{-1} y(x)$$

and for GWR,

$$S(x) = y(x)^T [Y^T W(x) Y]^{-1} Y^T W(x) Z [Y^T W(x) Y]^{-1} y(x)$$

respectively. For details, see Leung et al. (2000).

For this study, GWR’s weighting matrix is specified using an isotropic exponential kernel. Here weights at $x$ accord to $w(x) = \exp(-d(x)/\beta)$ where the bandwidth parameter is the distance $b$ and $d(x)$ is the distance between $x$ and a sample location $i$. An optimal bandwidth is found in an adaptive form using leave-one-out cross-validation, where the root mean squared prediction error (RMSPE) is calculated for a range of bandwidths and the bandwidth that gives the minimum RMSPE is considered optimal.

B. H-GWR

A heteroskedastic GWR model is calibrated where $\sigma^2$ is replaced with a local estimate $\hat{\sigma}(x)$. Here we specify an iterative model, where at each iteration step, a re-weighted GWR fit is found (Fotheringham et al. 2002). The re-weighting is based on a mean smoothing over the observed squared residuals, which provide our $\sigma^2(x)$ data. In particular, the usual geographic weight at $x$ is multiplied by the inverse of $\sigma^2(x)$, where this second weighting corrects for local heteroskedasticity. The algorithm is applied with updated estimates of $\hat{\beta}(x)$ and $\hat{\sigma}^2(x)$ until an acceptable level of convergence is reached. As the parameter estimates $\hat{\beta}(x)$ are updated, then the H-GWR prediction at $x$ is also updated to give $\hat{z}_{H-GWR}(x) = y(x)^T \hat{\beta}(x)$ and the final estimate $\hat{\sigma}^2(x)$ replaces $\hat{\sigma}^2$ in expression 2 to give this prediction variance at $x$, $\sigma^2_{H-GWR}(x) = \hat{\sigma}^2(x)[I + S(x)]$. Thus H-GWR can improve prediction accuracy, as well as prediction uncertainty accuracy over a basic model. The bandwidth (and kernel function) specified with H-GWR is the same as that found optimally with basic GWR. The same bandwidth and kernel function is also used to smooth the observed squared residuals.

C. UK-GN and UK-LN

For unbiased variogram estimation in UK, it is necessary to find the variogram of the residual process $Z - Y \beta$. However, $\beta$ is unknown and can only be estimated efficiently with generalised least squares which itself needs to be calibrated using unbiased variogram information. This is an analytical impasse, which can be addressed via restricted maximum likelihood (REML), to first identify relatively unbiased estimates of $\Theta$ and then in turn, relatively unbiased estimates of $\beta$. Thus REML is used to parameterise UK, where an isotropic exponential variogram model-type is specified. The parameter estimates for the MLR component are $\hat{\sigma}_{GWS} = [Y^T [\Sigma_0]^{-1} Y]^{-1} Z$ and $\hat{z}_{GWS}(x) = y(x)^T \hat{\beta}_{GWS} + \sigma_0^2 [\Sigma_0]^{-1} (Z - Y \hat{\beta}_{GWS})$ is the UK prediction at $x$; where $\sigma_0$ is a $(n \times 1)$ vector of spatial
covariances between residuals at \( x \) and the sample locations. The UK variance at \( x \) is:

\[
\sigma^2_{UX}(x) = \left( \sigma^2 - \sigma_0^T \left[ \Sigma_0 \right]^{-1} \sigma_0 \right) + \left( \left[ y(x) - Y^T [\Sigma_0]^{-1} \sigma_0 \right] \left[ y(x) - Y^T [\Sigma_0]^{-1} \sigma_0 \right] \right)
\]  

To find \( \hat{z}_{UK}(x) \) and \( \sigma^2_{UX}(x) \) both implicit- and explicit-type solutions are possible, although these are identical if a global neighbourhood (UK-GN) is specified (Hengl et al. 2007). Here an implicit solution is adopted as it allows the MLR component to be calibrated locally when UK is specified with local neighbourhoods (UK-LN). An optimal neighbourhood size is found using the same cross-validation procedure as that used to find the bandwidth in GWR. Technically, local residual variogram parameters should now be estimated that are specific to \( x \) (as in MW-UK), but instead, the parameters of the globally-found residual variogram are retained and this modelling decision is known as quasi-stationarity. For this study, the approximation is specifically used to model nonstationary relationships via the UK model and as such, UK-LN acts as a direct geostatistical alternative to GWR.

D. MW-UK

In essence, our MW-UK model (correctly) applies a UK-LN model with local residual variogram parameters. Here we use REML to fit a pure nugget and three Matérn variogram model-types at each \( x \), and the variogram model with the lowest AIC value is retained. In turn, we then estimate the local MLR parameters at \( x \), optimally. The Matérn function offers much flexibility, where the higher is the value of the its smoothing parameter \( \nu \), the smoother is the process. For this study, \( \nu \) is not estimated but is fixed beforehand at 0.5, 1 or 1.5 (where an exponential model corresponds to a Matérn with \( \nu = 0.5 \)). As MW-UK is a fully-automatic (black-box) method, sensible heuristics are required to choose the starting parameters for each local REML fit. Furthermore, each local MLR component at \( x \) is always specified with the full covariate data set and not some subset of it. We also take the kriging neighbourhood as the same size as that used in the local variography (i.e. no local quasi-stationarity decisions are made). Crucial to the MW-UK model is its window size. This is found in an optimal fashion that reflects both prediction accuracy and prediction uncertainty accuracy (see Harris et al. 2010). Details for specifying a variety of moving window kriging models can be found in Haas (1996); Pardo-Igúzquiza et al. (2005); Lloyd (2010).

The decision of whether or not to apply MW-UK over its stationary counterpart of UK-LN will largely depend on a trade-off between: (a) many ill-fitted local variograms, but with (potentially) more accurate model outputs and (b) a well-fitted (and understood) global variogram with possibly less accurate model outputs. The MW-UK model will be of little worth if local variograms are only marginally different from each other and from the global variogram.

III. STUDY DATA

In order to empirically compare the set of models described, we take as our data a set of house prices in London. Here data were obtained for 11,285 geo-coded properties in London during 1998 through the Nationwide Building Society. This data was used to generate a sample of 2,009 non-overlapping observations, which in turn was split into calibration and (set-aside) validation data sets of 1,405 and 1,404 observations, respectively. Fig. 1 shows the location of the model calibration/validation data.

![Figure 1. Location of model calibration/validation data](image)

For our models, house prices are modelled as a function of twenty-one covariates following the hedonic price model used by Fotheringham et al. (2002). Here house prices in London are regressed on three groups of explanatory variables: (i) house attributes: floor area, property type, construction date, number of bathrooms, provision of garage and central heating; (ii) socioeconomic characteristics: proportion of workforce in professional or managerial occupations, and rate of unemployment at the census output area in which the property is located; and (iii) a location attribute: where postcodes are used to calculate the straight-line distance of each property from the centre of London (this has an accuracy level of 100m).

IV. ANALYSIS: PRELIMINARY RESULTS

Performance results for our six study models at the validation sites are summarised in Table 2. Here the RMSPE and the mean absolute prediction error (MAPE) both measure model prediction accuracy.

The mean squared deviation ratio (MSDR) measures prediction uncertainty accuracy in a global-sense, where values of 1 are sought. In a local-sense, prediction confidence interval (PCI) accuracy is assessed using coverage probabilities. Here the G-statistic measures the accuracy of a model’s PCs (values of 1 are sought) and the MW-statistic gives a rough guide to the precision of a model’s PCs (i.e. this statistic reflects the narrowness of the PCI widths and should be as small as possible).
Results indicate that for all five performance measures, HGWR is the best model.

<table>
<thead>
<tr>
<th>MODEL</th>
<th>RMSPE x 10^3</th>
<th>MAPE x 10^3</th>
<th>MSDR</th>
<th>G-STAT.</th>
<th>MW-STAT. x 10^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. MLR</td>
<td>38.57</td>
<td>26.96</td>
<td>0.797</td>
<td>0.897</td>
<td>68.50</td>
</tr>
<tr>
<td>2. GWR</td>
<td>32.09</td>
<td>22.36</td>
<td>1.141</td>
<td>0.943</td>
<td>47.08</td>
</tr>
<tr>
<td>3. H-GWR</td>
<td>31.96</td>
<td>21.73</td>
<td>1.124</td>
<td>0.949</td>
<td>45.17</td>
</tr>
<tr>
<td>4. UK-GN</td>
<td>33.93</td>
<td>23.23</td>
<td>0.777</td>
<td>0.883</td>
<td>61.09</td>
</tr>
<tr>
<td>5. UK-LN</td>
<td>34.22</td>
<td>23.40</td>
<td>0.777</td>
<td>0.880</td>
<td>61.89</td>
</tr>
<tr>
<td>6. MW-UK</td>
<td>33.06</td>
<td>23.02</td>
<td>0.801</td>
<td>0.920</td>
<td>51.46</td>
</tr>
</tbody>
</table>

V. BAYESIAN EXTENSIONS

Essentially, all of our study models can be extended to a Bayesian form. In brief, if \( \theta \) is our parameter vector, then:

\[
\pi(\theta) = \frac{1}{(1|\theta, Y)}
\]

where \( \pi(\theta) \) represents prior beliefs about \( \theta \); \( (1|\theta, Y) \) is the likelihood of a set of observed dependent variables \( z \) given covariates \( Y \) and model parameters \( \theta \); and \( (1|\theta, Y) \) represents posterior beliefs about \( \theta \) once observations of \( z \) have been made. Typically, predictions of the dependent variable can be made, given a new set of covariates \( Y' \) using:

\[
\gamma(z'|\theta, Y, z) \propto (1|\theta, Y, z) \pi(\theta|Y, z)
\]

where \( \gamma(z'|\theta, Y, z) \) is the predictive distribution for new values of the dependent variable. From this distribution: predictions, prediction credibility intervals, and the risk of exceeding some threshold can all be computed.

As is common practice, the derivation of the posterior and predictive distributions are not dealt with analytically, but instead a MCMC approach is adopted, where random numbers based on distributions in expressions 4 and 5 are generated (typically around 10,000) and the distributional properties of these random numbers computed. For example, if we drew 10,000 \( z \) values from the distribution set out in expression 5 and computed 2.5 and 97.5 percentiles for each individual \( z \) values, we obtain the prediction credibility intervals.

For a Bayesian version of UK (e.g. Diggle and Ribeiro 2007), the parameters are estimated in the form of expression 4, where in this case \( \theta \) consists of the regression and the variogram model parameters. For a Bayesian version of GWR, some modification of the standard approach is necessary. Here rather than a moving window approach to calibration, we propose a form where each \( \beta_k(x) \) is approximated as a sum of B-spline functions. Each has a finite support and is equal to zero except for a small region in two-dimensional space. Effectively, we write

\[
\beta_k(x) = \sum \alpha_k P_i(x)
\]

and estimate the parameters \( \alpha_k \).

Thus, these parameters are part of \( \theta \). This could result in over-fitting if the value of \( r \) is too high, but \( \pi(\theta) \) is used as a penalty term to prevent such under-smoothing. This Bayesian GWR model is currently in development and represents a more direct alternative to the Bayesian SVC model of Gelfand et al. (2003), than a standard GWR model does.

VI. CONCLUSIONS

Preliminary results using our six core study models indicate that a heteroskedastic GWR model should be the preferred predictor for our study data set. This model can be preferred with respect to both its prediction accuracy and its prediction uncertainty accuracy.

ACKNOWLEDGEMENTS

Research presented in this paper was funded by a Strategic Research Cluster grant (07/SRC/11168) by the Science Foundation Ireland under the National Development Plan. Prof. Brunsdon’s time was made available via the granting of a one year study leave by the University of Leicester.

REFERENCES


