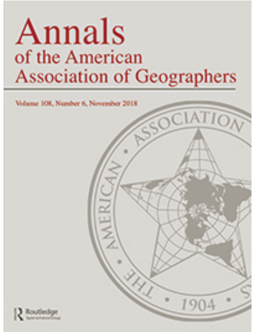

Taylor & Francis
Taylor & Francis Group

Annals of the American Association of Geographers

ISSN: 2469-4452 (Print) 2469-4460 (Online) Journal homepage: https://www.tandfonline.com/loi/raag21

Geographically Weighted Elastic Net: A Variable-Selection and Modeling Method under the Spatially Nonstationary Condition

Kenan Li & Nina S. N. Lam

To cite this article: Kenan Li & Nina S. N. Lam (2018) Geographically Weighted Elastic Net: A Variable-Selection and Modeling Method under the Spatially Nonstationary Condition, Annals of the American Association of Geographers, 108:6, 1582-1600, DOI: 10.1080/24694452.2018.1425129

To link to this article: https://doi.org/10.1080/24694452.2018.1425129

Geographically Weighted Elastic Net: A Variable-Selection and Modeling Method under the Spatially Nonstationary Condition

Kenan Li* and Nina S. N. Lam[†]

*Department of Preventive Medicine, University of Southern California
[†]Department of Environmental Sciences, Louisiana State University

This study develops a linear regression model to select local, low-collinear explanatory variables. This model combines two well-known models: geographically weighted regression (GWR) and elastic net (EN). The GWR model posits that the regression coefficients vary as a function of location and focuses on solving the problem of explaining the relationships under the spatially nonstationary condition, which a global model cannot solve. GWR cannot fulfill the task of variable selection, however, which is problematic when there are many explanatory variables with nonnegligible multicollinearity. On the other hand, the EN model is a member of the regulated regression family. EN can trim the number of explanatory variables and select the most important ones by adding penalty terms in its cost function, and it has been proven to be robust under the high-multicollinearity condition. The EN model is a global model, however, and does not consider the spatial nonstationarity. To overcome these deficiencies, we proposed the geographically weighted elastic net (GWEN) model. GWEN uses the kernel weights derived from GWR and applies EN locally to select variables for each geographical location. The result is a set of locally selected, low-collinear explanatory variables with spatially varying coefficients. We demonstrated the GWEN method on a data set relating population changes to a set of social, economic, and environmental variables in the Lower Mississippi River Basin. The results show that GWEN has the advantages of both the high prediction accuracy of GWR and the low multicollinearity among explanatory variables of EN. *Key Words: elastic net, geographically weighted elastic net, geographically weighted regression, spatial nonstationarity, variable selection.*

本研究发展一个线性迴归模型来选择在地且低共线性的解释变因。此一模型结合了两个知名的模型: 地理加权迴归 (GWR) 与弹性网络 (EN)。GWR 模型假定迴归系数作为区位函数而有所变化, 并聚焦解释空间非静态条件下的关系之问题, 而该问题无法由全球模型解决。但 GWR 无法完成变因选择的任务, 因此当具有诸多无法忽略的多重共线性之解释变因时便会产生问题。此外, EN 模型是调节迴归家族中的一员。EN 能够通过在其成本函数中增加处罚条款, 缩减解释变因的数量, 并选择最重要的变因, 且已被证实在高度多重共线性的条件下是有效的。但 EN 模型是全球模型, 而且不考量空间非静止性。为了克服这些缺陷, 我们提出地理加权弹性网络 (GWEN) 模型。 GWEN 运用 GWR 衍生的核加权, 并将 EN 运用至地方, 以选择各地理区位的变因。该结果是一组地方选择的低度共线解释变因, 并具有空间变异的系数。我们在一组将人口变迁连结至密西西比河流域下游的社会、经济与环境变因组的数据集来展现 GWEN 模型。该结果显示, GWEN 同时具有 GWR 高度预测准确性以及 EN 解释变因的低度多重共线性之优势。 *关键词: 弹性网络, 地理加权弹性网络, 地理加权迴归, 空间非静止性, 变因选择。*

Este estudio desarrolla un modelo de regresión lineal para seleccionar variables explicativas de colineal bajo. Este modelo combina dos modelos bien conocidos: la regresión geográficamente ponderada (GWR) y la red elástica (EN). El modelo GWR plantea que los coeficientes de regresión varían como una función de localización y se enfoca en resolver el problema de explicar las relaciones bajo la condición espacialmente no estacionaria, que un modelo global no puede resolver. Sin embargo, la GWR no puede cumplir la tarea de selección de variables, que resulta problemática cuando hay muchas variables explicativas con multicolinearidad no desdeñable. Por otra parte, el modelo EN hace parte de la familia de regresión regulada. El EN puede recortar el número de variables explicativas y seleccionar las más importantes añadiendo términos de sanción en su función de costo, además de haber resultado robusto bajo la condición de alta multicolinearidad. Sin embargo, el EN es un modelo global que no considera la no estacionalidad espacial. Para remediar estas deficiencias, proponemos el modelo de la red elástica geográficamente ponderada (GWEN). El GWEN usa los pesos kernel derivados de la GWR y aplica localmente el EN para la tarea de seleccionar variables en cada localización geográfica. El resultado es un conjunto de variables explicativas de colineal bajo localmente seleccionadas con coeficientes que varían espacialmente. Hicimos una demostración del método GWEN sobre un

Annals of the American Association of Geographers, 108(6) 2018, pp. 1582–1600 © 2018 by American Association of Geographers
Initial submission, August 2016; revised submissions, February and September 2017; final acceptance, October 2017
Published by Taylor & Francis, LLC.

conjunto de datos que relacionan los cambios de población con un conjunto de variables sociales, económicas y ambientales en la Cuenca del Bajo Río Misisipi. Los resultados muestran que el GWEN reúne las ventajas tanto de la alta exactitud de predicción de la GWR como la baja multicolinearidad propia de las variables explicativas del EN. *Palabras clave: red elástica, red elástica geográficamente ponderada, regresión geográficamente ponderada, no estacionalidad espacial, selección de variables.*

The linear regression model is probably one of the oldest models for identifying relationships among variables. Over the years, researchers have made extensive efforts in modifying the linear regression model into various forms for spatial analysis and modeling (e.g., Casetti 1972; Griffith 1981, 2008; Anselin 1995). In regression analysis, two aspects are very important: the prediction accuracy of the response (dependent) variable and the interpretation of the explanatory (independent) variables. When it comes to analyzing geographical phenomena, spatial nonstationarity and multicollinearity are two of the major obstacles to achieving satisfactory regression results. In the case of spatial nonstationarity, a "simple" global regression model often cannot explain the global relationships between sets of variables accurately because the relationships vary over space. The spatial nonstationarity problem requires a model to allow different relationships to exist within the data set at different locations. On the other hand, multicollinearity is a phenomenon in which two or more explanatory variables in a multiple regression model are highly correlated, meaning that one can be predicted from the others with a substantial degree of accuracy. When multicollinearity exists, the coefficient estimates of the regression model will change erratically in response to small changes in the model or the data, making it difficult to derive valid coefficient estimates for those explanatory variables that exhibit high multicollinearity (Farrar and Glauber 1967). The fact that the explanatory variables are redundant with respect to others also makes the interpretation of the regression difficult. Local linear regression, such as geographically weighted regression (GWR; Brunsdon, Fotheringham and Charlton 1996; Fotheringham, Charlton, and Brunsdon 1996; Fotheringham, Brunsdon, and Charlton 2002), and regularized linear regression, such as elastic net (EN; Zou and Hastie 2005), are products that can be combined to overcome these two obstacles.

The objective of this article is to describe a new method called geographically weighted elastic net (GWEN), which combines GWR and EN to handle spatial nonstationarity and multicollinearity simultaneously. The development of GWEN is motivated by the fact that complex system research is increasingly needed to tackle multifaceted societal problems, and such complex system research requires the analysis of a large number of natural and human factors. Finding a method to better extract the various factors with minimal multicollinearity and maximal explanation of nonstationarity will help improve our understanding and modeling of the system dynamics. We demonstrate the GWEN method using a data set relating population changes to a set of social, economic, and environmental variables in the Mississippi River Delta region (K. Li 2015; Qiang and Lam 2015, 2016). Next we provide a brief background of related research, followed by a description of the methods of GWR, EN, and GWEN and a case study of population changes in the Mississippi Delta. We conclude with a summary of the pros and cons of the new GWEN method.

Background

Brunsdon, Fotheringham, and Charlton (1996) developed GWR in the field of spatial analysis to overcome the issues of spatial nonstationarity (Fotheringham, Charlton, and Brunsdon 1996; Fotheringham, Brunsdon, and Charlton 2002). GWR allows the parameters for each geographical location to be estimated and mapped individually, as opposed to having a single set of globally estimated coefficients fitted to the entire data set. By doing so, the method estimates the spatially varying relationships between explanatory variables and a response variable. Although GWR is not designed to address spatial autocorrelation directly, which is another inherent property of spatial data, the method's ability to handle spatial nonstationarity by capturing local relationships can reduce the effects of spatial autocorrelation on the regression model (Fotheringham, Brunsdon, and Charlton 2002). In other words, low spatial autocorrelation in the residuals is to be expected. If the residuals of an ordinary linear regression are sufficiently autocorrelated, then one of the underlying assumptions of ordinary linear regression is violated and the regression analysis could be unreliable. GWR is preferable in such cases, because GWR accounts for some of the spatial autocorrelation latent in georeferenced variables by transferring them to the spatially varying coefficients. Griffith (2008), however, found that GWR might not be able to overcome the problem of spatial autocorrelation because positive spatial autocorrelation in the residuals from GWR still existed. To illustrate the disagreement among researchers (Fotheringham and Oshan 2016), the spatial autocorrelation of the residuals of several models is tested in the case study that follows.

A linear regression model that contains multicollinearity among explanatory variables might yield good fitness and statistical significance, but it will result in parameter estimates that are sensitive to changes in model specification and sample coverage (Farrar and Glauber 1967). Hence, methods to select parsimonious models with the best explanatory variables while minimizing the overlapping among variables would be preferred. Stepwise methods are a widely employed approach to minimizing multicollinearity. Stepwise multiple regression algorithms operate by successive additions (forward selection), successive removals (backward elimination), or a combination of the two (bidirectional elimination), of significant variables according to a specified criterion of variance. Whittingham et al. (2006), however, summarized the principal drawbacks of stepwise multiple regression, which include bias in parameter estimation, inconsistencies among model selection algorithms, an inherent problem of multiple hypothesis testing, and inappropriate focus or reliance on a single best model. The problem of inconsistencies owing to the order of variable entry (or deletion) and the number of candidate variables (Derksen and Keselman 1992) is especially acute when the explanatory variables have multicollinearity (Grafen and Hails 2002).

The multicollinearity problem can be exacerbated in GWR. Wheeler and Tiefelsdorf (2005) suggested that even moderate multicollinearity among locally weighted explanatory variables can lead to strong dependence in the local estimated coefficients. This can be explained by the fact that GWR often involves smaller sample sizes and picks observations with similar values; hence, it is more likely to yield a poorly conditioned design matrix, as demonstrated in some studies (Fotheringham, Kelly, and Charlton 2012; Bãrcena et al. 2014). As an alternative to GWR, Wheeler and Calder (2007) found that the Bayesian regression model with spatially varying coefficient processes generally performs better and degrades less substantially than GWR under strong multicollinearity.

Multicollinearity in GWR can be diagnosed in a number of ways. Fotheringham, Kelly, and Charlton (2012) used Akaike's information criterion (AIC) as a criterion for selecting significant variables from a set of twelve demographic variables in a GWR framework. The calculation of AIC in a model is shown in Equation 1.

$$AIC = 2m + n\ln\left(\frac{RSS}{n}\right), \tag{1}$$

where n is the number of observations, m is the number of explanatory variables, and RSS (residual sum of squares) is the sum of squared errors of prediction in the model. Using AIC as a criterion of fitness is known to be an effective way to create a sparsity model and the method has nice theoretical properties. The optimization of the final regression model that has the least multicollinearity, however, is NP-hard and computationally challenging (Liu and Li 2016). Fotheringham, Kelly, and Charlton (2012) used a stepwise procedure (Step-AIC) instead to select variables to minimize the computational problem. The method starts with no explanatory variables selected, and the variable with the lowest AIC is added at each successive step. The variables selected from this Step-AIC method are global, which means that all of the geolocations share the same set of explanatory variables. Thus, the method also has the same weaknesses as the normal stepwise method mentioned earlier.

Based on the work by Belsley (1991), Wheeler (2007) pioneered the idea of Ridge regression along with GWR (GWR–Ridge) and developed a module for R implementation. Alternatively, Vidaurre, Bielza, and Larrañaga (2012) and Wheeler (2009) proposed the use of least angle regression (LAR) and least absolute shrinkage and selection operator (Lasso) algorithms (Tibshirani 1996) for simultaneous regularization and variable selection in the context of local and geographical regression (GWR–Lasso). Ridge regression and Lasso regression are both regularized regression methods, which place a constraint (i.e., penalty terms) on the regression coefficients to reduce multicollinearity. Integrating regularized regression terms into the GWR framework can improve the model by minimizing multicollinearity, but both Ridge regression and Lasso regression have their own disadvantages (Zou and Hastie 2005). Ridge regression can continuously shrink the values of the parameters (coefficients), but it cannot eliminate any variable through setting its coefficient to zero. As a result, Ridge regression always keeps all of the explanatory variables in the model and is not able to offer a parsimonious model. In contrast, Lasso regression can be used to select variables while minimizing the loss of prediction accuracy, but Lasso regression tends to select only one variable from the group, among which the pairwise correlations are very high. Lasso regression does not care which one is selected, and it has been empirically observed that the prediction performance of Lasso regression is not as good as Ridge regression (Tibshirani 1996). To take advantage of both Ridge regression and Lasso regression, Zou and Hastie (2005) developed a new method called EN. EN is a hybrid of the Ridge and the Lasso (Equation 7), which can serve the purpose of both automatically selecting the variables as with Lasso and keep a high prediction performance when collinearity exists, as does Ridge.

The proposed GWEN method aims to overcome some of the shortcomings discussed in these previous studies by incorporating EN into the GWR framework. The purpose of its development is to model spatially varying relationships, minimize multicollinearity, constrain and stabilize regression coefficients, and lower prediction errors. The penalty terms in the cost functions of the Step-AIC model, the GWR–Lasso model, and the GWR–Ridge model are often referred to as L_0, L_1, and L_2, because they are a function of zero-power, one-power, and two-power, respectively, of the explanatory variables in the models. The details of the cost functions of the Ridge model and the Lasso model are described in the Methods section. Table 1 summarizes and compares the characteristics of the methods just reviewed as well as the proposed GWEN.

Methods

Geographically Weighted Regression

In an ordinary least squares (OLS) regression, the relationship between the response variable and the explanatory variables is modeled as

$$y_i = \sum_{k=1}^{m} X_{ki}\beta_k + \beta_0 + \varepsilon_i, \tag{2}$$

where y_i is the response variable at observation (or location) i, X_{ki} is the value of the kth explanatory variable at location i, β_k is the regression coefficient of the kth explanatory variable, β_0 is the constant to be estimated, and ε_i is the residual at observation i. β_k and β_0 are to be estimated to minimize the following residual term:

$$\frac{1}{2N}\sum_{i=1}^{N} \left. y_i - \beta_0 - \sum_{k=1}^{m} X_{ki}\beta_k \right)^2, \tag{3}$$

where N is the number of observations or data points.

In GWR, a local regression model is fitted at each data point (location) using distance-weighted subsamples defined by a bandwidth. The GWR model for each location i becomes

$$y_i = \sum_{k=1}^{m} X_{ki}\beta_{ki} + \beta_{0i} + \varepsilon_i, \tag{4}$$

where β_{ki} is the local regression coefficient for the kth explanatory variable at location i, and β_{0i} is the intercept at location i. Hence, in GWR, two parameters will need to be determined, the distance decay–based kernel function and the bandwidth. Some commonly used kernel functions are the

Table 1. Variable selection methods used in geographically weighted regression framework

Name	Penalty term	Sources	Variables selected
Step-AIC	L_0	Fotheringham Kelly, and Charlton (2012)	All of the geolocations share the same selection.
GWR–Lasso	L_1	Wheeler (2009)	Each geolocation selects its own explanatory variables.
GWR–Ridge	L_2	Wheeler (2007)	Each geolocation selects its own explanatory variables.
GWEN	Combination of L_1 and L_2	Proposed in this study	Each geolocation selects its own explanatory variables.

Note: AIC = Akaike's information criterion; GWR = geographically weighted regression; Lasso = least absolute shrinkage and selection operator; GWEN = geographically weighted elastic net.

Gaussian function, the *bi*-square nearest neighbor function, and the exponential function (Brunsdon, Fotheringham, and Charlton 1996; Fotheringham, Brunsdon, and Charlton 2002). There are two general schemas to determine the bandwidth: fixed bandwidth or adaptive bandwidth. In a fixed bandwidth schema, the kernel bandwidth is a fixed value and is often estimated by the leave-one-out cross-validation method, which is used to iteratively find the bandwidth with the lowest prediction error of all y_is (i.e., root mean square error [RMSE]). Some researchers also suggest the method of AIC for bandwidth estimation, although it is not clear how the results from these two methods might differ (Wheeler 2007, 2009). In an adaptive bandwidth scheme, the bandwidth is an adaptive value to allow at least a certain number of data points (geolocations) included in the kernel. The predetermined number of data points can be estimated similarly by cross-validation or AIC.

Elastic Net

Because of the need to minimize multicollinearity and achieve parsimony in a single model, Zou and Hastie (2005) developed a new regularization and variable selection method called EN. EN is a hybrid of the Lasso regression and the Ridge regression (Hoerl and Kennard 1988), both of which add a penalty term in the cost function of OLS regression. As mentioned before, these regression methods are often referred to as regularized regressions. Like many other multivariate methods, they require that the variables be normalized before the analysis.

Ridge regression is a technique for analyzing multiple regression data that suffer from multicollinearity. The minimization criterion of Ridge regression is modified from Equation 3:

$$\frac{1}{2N}\sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{k=1}^{m} x_{ki}\beta_k \right)^2 + \lambda_2 \sum_{k=1}^{m} \beta_k^2, \quad (5)$$

where λ_2 is a positive regularization parameter, m is the number of explanatory variables, and β_k denotes the regression coefficient of variable k. λ_2 ranges from 0 to 1; the higher the λ_2 value, the higher the regularization and the larger the shrinkage of the regression coefficient value.

In like manner, Lasso regression is another popular procedure for combating multicollinearity. The difference between Ridge and Lasso is the penalty terms used in their minimization criteria. The minimization criterion of Lasso regression is defined in Equation 6:

$$\frac{1}{2N}\sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{k=1}^{m} x_{ki}\beta_k \right)^2 + \lambda_1 \sum_{k=1}^{m} |\beta_k|. \quad (6)$$

In the same way as Ridge regression, λ_1 ranges from 0 to 1; the higher the λ_1 value, the larger the shrinkage, but Lasso regression can actually achieve the goal of continuously shrinking some of the parameters to zero, thus performing the function of variable selection simultaneously (Zou and Hastie 2005).

As a combination of Ridge and Lasso, EN can serve the purpose of both automatically selecting the variables as does the Lasso and keeping a high prediction performance when collinearity exists as does the Ridge (Zou and Hastie 2005). The penalty term of EN is a trade-off between the Lasso regression penalty (often referred to as L_1 penalty) and the Ridge regression penalty (often referred to as L_2 penalty). The regularization problem of EN is defined in Equation 7.

$$\begin{aligned} &\frac{1}{2N}\sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{k=1}^{m} x_{ki}\beta_k \right)^2 \\ &+\lambda \sum_{k=1}^{m} ((1-\alpha)\beta_k^2 + \alpha|\beta_k|), \end{aligned} \quad (7)$$

where α, ranging from 0 to 1, is the parameter that determines whether the regression model is more like a Lasso or a Ridge regression. The EN is the same as a Lasso regression when α equals 1. On the contrary, as α shrinks toward 0, the EN approaches a Ridge regression. λ is a regularization parameter that determines the number of explanatory variables selected (with nonzero coefficients). λ is 0 when all variables are selected. As λ increases, the number of zero β_k increases, which means that the number of variables selected decreases until it reaches a maximum when no variables are selected. When solving an EN problem, the value of λ is often incrementally tested hundreds of times starting from 0, at which the EN turns into an ordinary regression with no regularization (all nonzero coefficients) to its maximum value at which there is no nonzero β_k. The smallest tested value at which the first zero β_k occurs is defined as the lower boundary, whereas the largest tested value at which the cross-validated RMSE increases dramatically is defined as the upper boundary. The final value of λ is chosen by the user within this safe range according to the desired number of explanatory variables.

Geographically Weighted Elastic Net

This study proposes a method called GWEN to model the relationships between variables locally while selecting the most significant ones to minimize multicollinearity. It can be viewed as a regulated form of GWR. In other words, it generates parsimonious models locally for a geographically based data set.

Definition

Similar to the central idea of GWR, GWEN also uses a search (radius) bandwidth r to identify the data points geographically located near a particular observation (location) P_i. Then β_{ki} is estimated by calibrating a weighted EN based on the data points geographically located within this radius. A defined distance decay function is then used to assign the weight of each data point j to observation i within the radius r_i. Let w_{ij} be the weight of point j to observation i and w_{ij} for the points outside the search radius are set to 0, then the GWEN regularization regression problem for each observation i is defined as follows:

$$\frac{1}{2}\sum_{j=1}^{q}\frac{w_{ij}}{\sum_{k=1}^{N}w_{ij}}\quad y_j-\beta_{0j}-\sum_{k=1}^{m}x_{kj}\beta_{kj}\Bigg)^2$$

$$+\lambda\sum_{k=1}^{m}((1-\alpha)\beta_{ki}^2+\alpha|\beta_{ki}|) \tag{8}$$

$$\begin{cases} w_{ij}=F_{\text{kernel}}(d_{ij},\ r),\ if d_{ij}\le r_i \\ w_{ij}=0,\ if d_{ij}>r_i \end{cases}, \tag{9}$$

where j denotes the data points located within the radius (observation i is included). d_{ij} denotes the distance between observation i and point j, F_{kernel} is a defined distance decay function, r_i is the defined search radius for F_{kernel}, and q is the number of data points within the radius. In a fixed bandwidth schema, r_i is a predetermined constant, whereas in an adaptive bandwidth scheme, r_i is estimated by a predetermined number of neighbor data points.

Solution

Solving a GWEN is basically like solving a weighted EN for each observation iteratively using local weights assigned by the defined kernel function. The solution method is based on the coordinate descent algorithm used by Friedman et al. (2007; Friedman, Hastie, and Tibshirani 2010) for fitting generalized linear models with EN penalties. The coordinate descent algorithm is a nonderivative optimization algorithm for finding the minimum of a given function. The underlying idea of the algorithm is that the minimization of a multivariable function can be achieved by minimizing the function along each variable's direction at a time; that is, by solving the univariate (much simpler) optimization problems in a loop (Wright 2015). For the multivariate minimization problem as in GWEN, all of the β_{ki} are to be estimated. For a specified observation P_i, the coordinate descent algorithm first generates an initial guess for each β_{ki}. Then, the algorithm solves a simplified univariate problem by finding the optimized value for each β_{ki} one by one (totally k steps). After going through each k, the univariate optimization values of the first round are obtained and the initial guess values are substituted with the new values, and these values will be used as initial values for the next round of iteration. The iteration stops until the minimum of the multivariate function converges.

Before using the coordinate descent algorithm to solve a GWEN problem (Equations 8 and 9), three parameters need to be derived: the search bandwidth or the kernel distance r, the penalty coefficient λ, and the coefficient α, which determine whether the regression problem is more like a Lasso or a Ridge regression. We used the K-fold cross-validation method, which uses the criterion of minimizing the RMSE between the predicted and the real values to iteratively derive these three parameters. K is a user-provided or default integer ($K = 10$ was used in this study). The tuning range of r is set by the user, and its values depend on the research problem. The tuning range of α is from 0 to 1, and the tuning range of λ is determined by the solution algorithm (the upper boundary and the lower boundary mentioned before).

There are two options in setting the parameters during the cross-validation process of tuning the parameters in the weighted EN for each individual observation. The parameters can be set either locally, which means that the weighted EN for each individual observation has its own parameter setting, or globally, which means that all of the observations' weighted EN share the same parameter setting. We chose different options for the three parameters, α, r, and λ. First, we set α globally to make sure that every observation has the same trade-off between the L_1 (Lasso) and L_2 (Ridge) penalties. For r, if there is small spatial variation in the density of observations, a global kernel bandwidth can ensure that the weighted EN for each observation is calibrated by a similar number of neighborhood observations. If the density of the observations varies a lot spatially, however, some of the observations might have the problem of lacking neighbors or even no neighbors for a given global bandwidth. In such cases, tuning r locally (adaptive bandwidth scheme) to make sure that each observation has a sufficient and similar number of neighbors is preferable. Thus, in the case of large spatial variation of observation locations, a global number of neighbors instead of a global r is suggested, whereas in the opposite case, a global r can be used to save computational time.

For λ, because it affects the number of selected variables indirectly, a global setting of it will lead to different numbers of explanatory variables in different observations' regression functions (Equation 8). In addition, a global value of λ might exceed the proper tuning range for certain observations, which means that for certain observations, the defined global value might be either outside of the upper tuning boundary, causing a dramatic increase in RMSE, or outside of the lower tuning boundary, causing no variable selected at all. Thus, instead of a global setting of λ, we chose to use a global setting of the number of variables to be selected to make sure that the local models are comparable. Once the number of explanatory variables is set globally, the solution algorithm will set the local value of λ accordingly.

The coordinate descent algorithm and the cross-validation tuning process constitute the GWEN solution method, and the whole procedure is listed in the Appendix. The computational complexity of GWEN is $O(m^3 + m^2n)$, where m is the number of explanatory variables and n is the number of geolocations.

A GWEN Analysis of Population Changes in the Lower Mississippi River Basin

We used a case study of the population changes from 2000 to 2010 in the Lower Mississippi River Basin area to demonstrate the ability of the proposed GWEN in handling the spatial nonstationarity and the multicollinearity problem. The study area contains three major cities in Louisiana (New Orleans, Baton Rouge, and Lafayette) and is very diverse in both the natural and human environments (Figure 1). The region is also very vulnerable to coastal hazards and climate change effects such as sea-level rise and land subsidence. The southern coastal part of the study area, including the City of New Orleans, has been

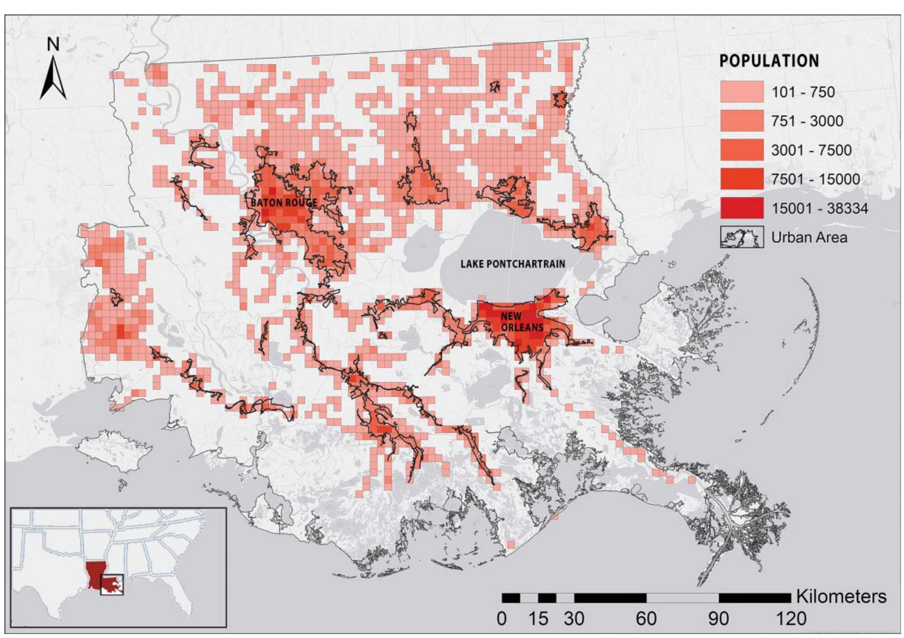

Figure 1. The population count in 2010 defined in 3 km × 3 km grids in the study area. (Color figure available online.)

suffering from steady population decline over the past decade, whereas the northern part of the study area has experienced rapid population growth. The disappearing land in the coastal area has been a major problem facing the region, which prompts the question of whether southern Louisiana is sustainable given the continued land loss and population decline. Answering this critical question requires a system-level analysis of the complex coupling effects between the natural and human factors. Finding out the major factors of population changes is the first step toward more accurate modeling, which is needed for scenario simulation so that the population change patterns can be evaluated under various sustainability options (K. Li 2015; Qiang and Lam 2015; Twilley et al. 2016).

The two utilities of the GWEN method are that, first, GWEN has the GWR's advantage of lowering the regression residuals by providing local relationships and, second, GWEN has the EN's advantage of maintaining prediction performance and eliminating the multicollinearity and selecting the explanatory variables. The results generated from the GWEN model can be used to facilitate the modeling of the coupled dynamics between natural and human factors. In geosimulation using the bottom-up approach, a main challenge is defining the behaviors of the "bottom" units. Very few existing models have focused on capturing the local relationships, which are needed to model a complex phenomenon in a diverse region. Appling GWEN in modeling population changes in this study area should help in extracting the local rules and key variables governing the population dynamics, as well as evaluating how they differentiate over space. These local rules will help inform the building of a bottom-up simulation model in the study area.

Data Description

The data set used was created by K. Li (2015), which includes the variable of population change in number from 2000 to 2010 as the response variable and thirty-five social–environmental variables as explanatory variables for the study area (see Table 2). The original data were obtained from various sources in different scales and resolutions. Areal interpolation was used to transform the census data into a 3 km × 3 km grid lattice (Goodchild and Lam 1980; Lam 1983, 2009). The areal interpolation procedure used is an "intelligent" areal interpolation method that has the volume-preserving property (Shu et al. 2010; Cromley et al. 2012). In this study, we used the developed land cover area layer as an ancillary ("intelligent") layer to derive the weights for interpolating each census variable from census tract or block group boundaries into the 3 km × 3 km grid cells. This areal interpolation method with an additional control variable is very similar to the principle of dasymetric mapping, a mapping technique designed to reflect within-zone variations (Lam 1983, 2009). The data observations were the cells with population greater than 100 in the year 2000. There were a total of 1,565 data observations in the study area. Figure 1 shows the study area with the population counts for these data observations.

Model Specification

This section specifies the kernel weighting function, the parameter settings, and the method to indicate the multicollinearity. Because the goal of this article is to introduce the method, we used a fixed bandwidth instead of adaptive bandwidth so that we have a simpler environment to compare with. Hence, we set a global value r for all r_i. The global r was determined by cross-validation described next. A Gaussian function, a commonly used weighting function, was used for all of the observations. Equation 10 shows the Gaussian function used to calculate the weights w_{ij}:

$$\begin{cases} w_{ij} = e^{-\frac{{d_{ij}}^2}{\sigma^2}}, \; if \; d_{ij} \le r, \\ w_{ij} = 0, \; if \; d_{ij} > r \end{cases} \tag{10}$$

where σ determines the rate of the distance decay. The larger the σ, the slower the decay speed. Other notations are the same as in Equation 9. In this study, the value σ was set to make the weights at the edge of the bandwidth (radius) around 1 percent of the weight at the center.

We used the cross-validation method (as described in the Solution section) to derive the values of α (trade-off between Ridge and Lasso) and the bandwidth r. The tuning range of the bandwidth was from twenty-five to forty-five grids (i.e., 75–135 km) with an increment of five grids (15 km). The tuning range of α was from 0.1 to 1, with 0.2 as the incremental step. We set the cross-

Table 2. Acronyms and descriptions of the variables used in this study

Socioeconomic and housing variables			
Categories	Acronym	Description	Year
Housing	Occupied	Percent; occupied housing units	2000
	NonVehicle	Percent; occupied housing units with no vehicles available	2000
	NonFuel	Percent; occupied housing units with no house heating fuel used	2000
	NonPlumb	Percent; occupied housing units lacking complete plumbing facilities	2000
	NonKitchen	Percent; occupied housing units lacking complete kitchen facilities	2000
	NonTele	Percent; occupied housing units with no telephone service	2000
	NonMtg	Percent; specified owner-occupied units without mortgage	2000
	OwnerR	Percent; owner-occupied housing units	2000
	MedValue	Number; median value of specified owner-occupied units (dollars)	2000
	MedRent	Number; median gross rent of specified renter-occupied units (dollars)	2000
Households	OCST20	Percent; owner cost as a percentage of household income less than 15 percent	2000
	OCST35	Percent; owner cost as a percentage of household income more than 35 percent	2000
	Rent15	Percent; gross rent as a percentage of household income less than 15 percent	2000
	Rent35	Percent; gross rent as a percentage of household income more than 35 percent	2000
	OCSTWMTG	Number; median selected monthly owner costs with a mortgage (dollars)	2000
	OCSTNMRG	Number; median selected monthly owner costs without a mortgage (dollars)	2000
	HhSize	Number; average household size	2000
	MeanTime	Number; mean travel time to work (minutes)	2000
	MedIcm	Number; households median income (dollars)	2000
Individuals	Female	Percent; total female population	2000
	Population	Number; total population	2000
	Under5	Percent; total population under 5 years old	2000
	Over65	Percent; total population over 65 years old	2000
	HighSch	Percent; population over 25 years old with high school graduation or higher	2000
	Married	Percent; population over 15 years old and now married (except separated)	2000
	Employed	Percent; population over 16 years old employed	2000
	Poverty	Percent; individuals below poverty level	2000
Structures	Road	Number; road density	2007
	Pipeline	Number; pipeline density	2007
	GasWell	Number; oil and gas injection well density	2007
Environmental	Damages	Number; property damages by natural hazards (coastal, flood, hurricane, thunderstorm, and tornado) in 2010 inflation rate (dollars)	2000– 2010
	Subsidence	Number; subsidence rate interpolated by empirical Bayesian kriging using benchmarks	2004
	Elevation	Number; mean elevation calculated from LiDAR images (meters)	2000–2010
Land use and land cover	Developed	Percent; percentage of developed land use area	2001
	Water	Percent; percentage of open water land use area	2001

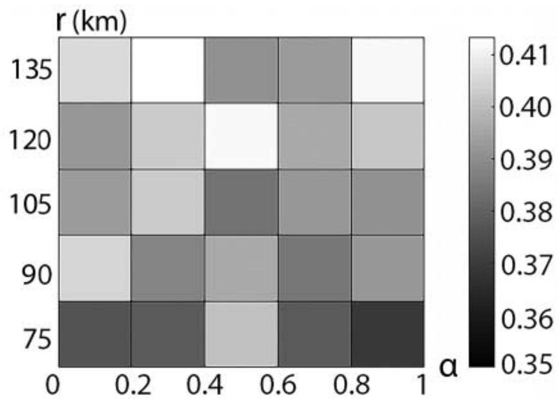

Figure 2. Average mean square errors of the geographically weighted elastic net model using different bandwidth r and α combinations.

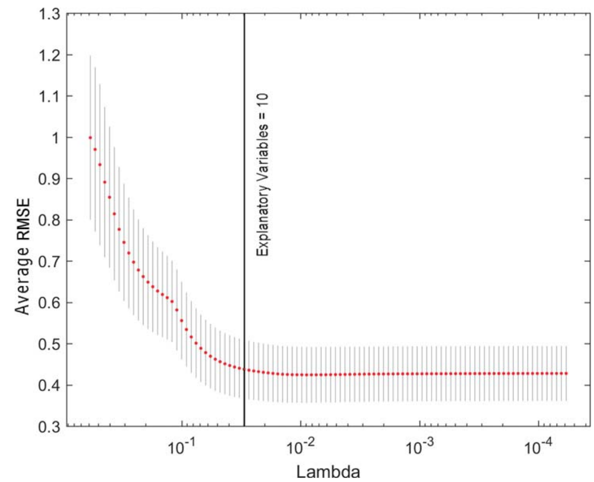

Figure 3. Average root mean square errors of all of the local submodels in geographically weighted elastic net under different values of λ. RMSE = root mean square error. (Color figure available online.)

validation to tenfold, which means that the original data set was randomly partitioned into ten subsets. Of the ten subsets, nine were used for training the model and the remaining one for testing the model. The cross-validation process was repeated ten times (folds), with each of the ten subsets used once as the test data. The RMSE of each combination of α and r within their defined tuning ranges is displayed in Figure 2, which shows no obvious correlations between α, r, and RMSE. We picked the lowest RMSE combination of $\alpha = 0.9$ and $r = 75$ km as the final parameter setting of the GWEN model in this study.

We set the number of variables to be selected to ten according to the inflection point of the curve (Figure 3) showing the GWEN model's average RMSEs over the λ values (which indicates the number of selected variables). We used a variance inflation factor (VIF), which measures how much the variance of an estimated regression coefficient for an explanatory variable is increased because of multicollinearity, to quantify the severity of the multicollinearity. VIF for each explanatory variable k is calculated using Equation 11:

$$VIF(k) = \frac{1}{1 - {R_k}^2}, \tag{11}$$

where ${R_k}^2$ is the coefficient of determination between variable k and the other explanatory variables. Generally, a rule of thumb is that if VIF(k) is greater than 10, then the multicollinearity of variable k is high (Kutner et al. 2008).

Global Moran's I was calculated using Equation 12 to evaluate the spatial autocorrelation of the residuals of different methods (H. Li, Calder, and Cressie 2007). The results are included in Table 3. Equation 12 is

$$I = \frac{n}{W} \frac{\sum_i \sum_j w_{ij}(x_i - \overline{x})(x_j - \overline{x})}{\sum_i (x_i - \overline{x})^2}, \tag{12}$$

where x is the variable of interest indexed by i and j, n is the total number of observations, w_{ij} is the weight between observation x_i and x_j, and W is the sum of all w_{ij}. We used an inverse distance weighting function to determine the weights in this study, Moran's I values range from –1 to +1, from perfect negative to perfect positive spatial autocorrelation. Values close to 0 imply little spatial autocorrelation.

Results and Discussion

We built and trained a GWEN model using the parameters described earlier. We compared the results from applying GWEN to the study area with the results from using seven other related models to provide an evaluation of GWEN's properties and performance. The seven other models were an OLS regression model, a stepwise regression model, a GWR model, an EN model, a Step-AIC model, a GWR–Ridge model, and a GWR–Lasso model. We set the parameters of these seven models according to the GWEN model to make the comparisons among them under the same criteria. The

Table 3. Comparison of the accuracy and multicollinearity of the eight models

Metrics	Ordinary	GWR	GWR–Ridge	Stepwise	Elastic net	Step-AIC	GWR–Lasso	GWEN
Reduce collinearity	No	No	No	Yes	Yes	Yes	Yes	Yes
Local	No	Yes	Yes	No	No	No	Yes	Yes
Root mean square error	0.63	0.42	0.51	0.64	0.65	0.57	0.59	0.55
Mean VIF (mVIF)	3.35	252.64	4.92	1.58	2.15	1.61	2.78	4.04
Maximum VIF (mVIF)	13.10	2,566.20	13.10	3.55	4.06	3.28	2.59	5.20
Minimum VIF (mVIF)	1.10	2.48	1.1	1.04	1.03	1.20	2.05	3.02
Percentage of VIFs (mVIFs) (>10)	0.17	0.91	0.91	0.00	0.00	0.00	0.00	0.00
# (Average) explanatory variables	35.00	35.00	35.00	10.00	10.00	10.00	10.18	10.18
Final AIC	−1,376	−2,645	−2,037	−1,377	−1,328	−1,529	−1,631	−1,851
Global Moran's I	0.069	0.009	0.019	0.075	0.095	0.076	0.046	0.045

Note: GWR = geographically weighted regression; AIC = Akaike's information criterion; Lasso = least absolute shrinkage and selection operator; GWEN = geographically weighted elastic net; VIF = variance inflation factor.

comparison here is not based on the best performance of each method including GWEN, and it is not a performance evaluation of the methods. As shown in Table 3, these methods are not entirely comparable either, because some methods have only one of the two properties (local or reducing multicollinearity), whereas only GWR–Lasso and GWEN have both. Setting some of the parameters constant for all of the methods will help us better understand the properties of GWEN.

For each model, α (if used) was 0.9, bandwidth (if used) was 75 km, and the number of variables to be selected (if applicable) was set to ten. For the variable-selecting models (Stepwise, EN, Step-AIC, GWR–Lasso, and GWEN), the significance level of entering or removing a variable using t values was manually specified to ensure that only ten variables were selected in its final format. In the Step-AIC model, the process of variable selection stops at Step 10. The AIC value for each step is shown in Figure 4, which shows that the decrease of AIC values slows down after around Step 5, implying that the first five variables were the most important. Ten variables were still selected in the Step-AIC model, however, to enable comparison across methods.

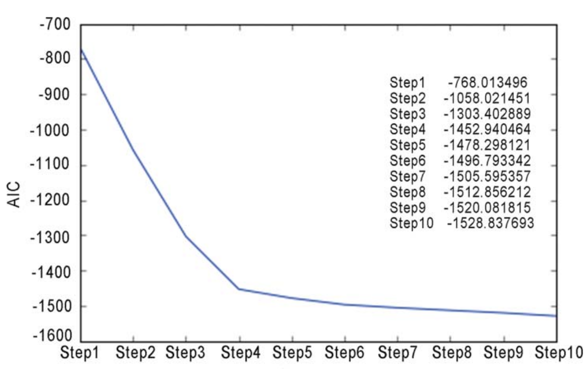

Figure 4. Akaike's information criterion curve for the first ten steps in the Step-AIC model. AIC = Akaike's information criterion. (Color figure available online.)

We evaluated all eight models based on their prediction accuracy (RMSE) and their multicollinearity among the selected explanatory variables (VIF). In addition, we computed the final AIC and spatial autocorrelation for each model. For the local models, including GWR, GWR–Ridge, GWR–Lasso, and GWEN, the mean VIFs (mVIF) calculated from their local submodels were used. Table 3 lists the RMSE; the maximum, minimum, and mean VIFs; the percentage of variables having VIFs ≥ 10; Moran's I; and the AIC. Figure 5 plots the predicted values against the measured values of all eight models.

The subplots in Figure 5 look similar among methods, but a close observation shows that subplots d, e, f, g, and h (GWR, Step-AIC, GWEN, GWR–Lasso, GWR–Ridge) are more concentrated around the $y = x$ diagonal line. This means that the geographically weighted models generally gain better prediction accuracy. All eight models have their predicted values higher than the measured values for the data points near the lower left corner of the subplots. These are the data points that suffered extreme population decreases, and they are mainly located in the New Orleans metropolitan

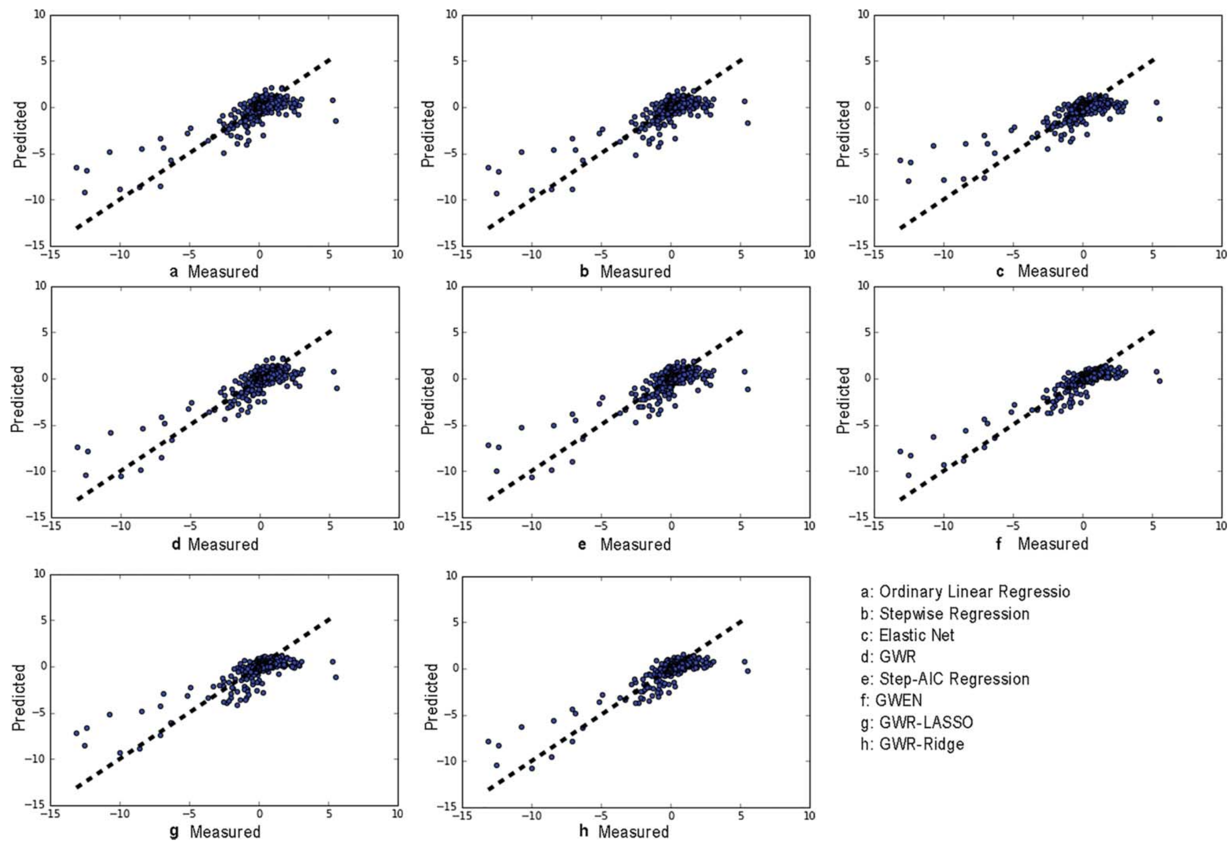

Figure 5. Scatter plots of the measured and the predicted values of the eight models. *Note:* GWR = geographically weighted regression; AIC = Akaike's information criterion; GWEN = geographically weighted elastic net; Lasso = least absolute shrinkage and selection operator. (Color figure available online.)

area (due partly to Hurricane Katrina, which occurred within this study's time span).

Table 3 illustrates several points. First, as expected, the local models such as GWR, GWR–Ridge, GWR–Lasso, and GWEN generally had higher prediction accuracies (smaller RMSE) than the global models such as the OLS, stepwise, and EN models. This implies that with the consideration of spatial nonstationarity, the local models explain the relationships between population changes and the explanatory factors better. Second, also expected, the regularized models (except GWR–Ridge), Step-AIC, and the stepwise models, no matter whether global or local, yielded low multicollinearity (with VIF or mVIF generally smaller than 10). On the contrary, there was a severe multicollinearity problem in the OLS model, GWR, and GWR–Ridge models, and especially in the GWR model, the multicollinearity was very high. Third, when considering only the methods that have the property of reducing multicollinearity (stepwise, EN, Step-AIC, GWR–Lasso, and GWEN), GWEN had the lowest RMSE (0.55), followed by GWR–Lasso (0.59). The slightly higher accuracy of GWEN, however, was compensated by a slightly higher degree of multicollinearity (mVIF = 4.04). Table 3 also shows that in general GWR–Lasso and GWEN are highly comparable. The GWR–Lasso model's prediction accuracy was lower than GWEN, but the mVIF value was lower and the range of VIF values was smaller. Compared with GWR–Lasso, GWEN seems to offer the flexibility of choosing between slightly higher accuracy and enduring slightly more multicollinearity.

For the spatial autocorrelation analysis, all of the Moran's *I* values calculated for the residuals were very close to 0, which means that we did not have seriously clustered residuals using all of the methods in this study (Table 3). In general, the global model with restricted variables (stepwise, EN, and Step-AIC) had higher Moran's *I* values than the local models (GWR, GWR–Ridge, GWR–Lasso, and GWEN). In particular, GWR and GWR–Ridge, with their local property and all thirty-five explanatory variables included, are expected to have better ability to transmit the spatial autocorrelation into their varying coefficients; thus, they had the lowest Moran's I values. On the other hand, GWR–Lasso and GWEN, with approximately ten explanatory variables, do not have the ability to transmit the spatial autocorrelation as much as GWR and GWR–Ridge do. Although the comparison here is only based on one empirical data set, the results suggest that GWEN is comparable to existing methods designed to reduce multicollinearity and handle spatial nonstationarity similar to GWR–Lasso. GWEN has the GWR's advantage of improving the prediction accuracy and the EN's advantage of eliminating multicollinearity and selecting the best explanatory variables.

Figure 6 maps the λ values of the GWEN model. The λ value is determined by the number of explanatory variables selected. The larger the λ value, the harder the regularization is needed to get the variables trimmed to the desired number. Figure 6 shows that locations with fewer neighbors (more isolated) generally had larger λ values, which also means that local multicollinearities were more severe in those areas.

To evaluate the importance of each explanatory variable in the GWEN model, one should consider both the magnitude of the explanatory variable's coefficient in each geographical location where it is selected and the number of geographical locations for which the variable is selected. Thus, we propose two measures to indicate an explanatory variable's importance. First, the coverage of an explanatory variable k, as defined in Equation 13, is used to show the ratio of the number of geographical locations that use variable k to all geographical locations.

$$\text{coverage}(k) = \frac{N - N_0}{N}, \tag{13}$$

where N_0 is the number of data observations where the coefficient of variable k equals 0, and N is the total number of observations. Second, the loading of an explanatory variable k is the mean squared value of the nonzero coefficients of variable k, as shown in Equation 14.

$$\text{loading}(k) = \frac{1}{N - N_0} \sum_{\beta_{ki}^{std} \neq 0} \left(\beta_{ki}^{std}\right)^2, \tag{14}$$

where $\beta_{ki}{}^{std}$ is the standardized coefficient of variable k at geographical location i. The higher the loading of the variable, the larger the standardized magnitude of its coefficient in the geographical location where it is selected and the more important it is to the response variable at the location. In this case study, because all of the variables were already standardized before the analysis, β_{ki} by itself is the standardized coefficient $\beta_{ki}{}^{std}$.

To indicate the importance of an explanatory variable using a single criterion, we developed another measure called the potency index to combine the coverage and the loading, which is a product of the two

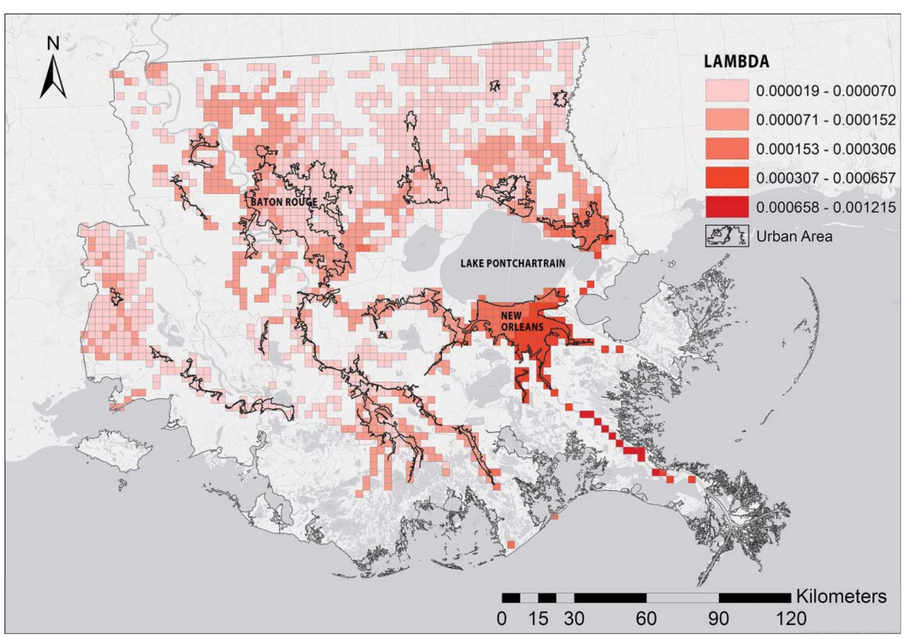

Figure 6. Lambda values of the geographically weighted elastic net model. (Color figure available online.)

Table 4. Coverages, loadings, and potency indexes of the explanatory variables in the geographically weighted elastic net model

Acronym	Coverage	Loading	Potency index	Acronym	Coverage	Loading	Potency index
Population	0.4473	0.2311	0.1034	OCST35	0.3335	0.0055	0.0018
Developed	0.6256	0.0389	0.0243	NonMtg	0.2703	0.0033	0.0009
Damages	0.3757	0.0367	0.0138	NonFuel	0.1265	0.0015	0.0002
Elevation	0.1476	0.0011	0.0002	NonKitchen	0.2486	0.0093	0.0023
Subsidence	0.4958	0.0045	0.0022	NonPlumb	0.2230	5.06E-04	0.0001
Water	0.5974	0.0103	0.0062	NonTele	0.1987	0.0017	0.0003
Employed	0.3661	0.0035	0.0013	NonVehicle	0.3770	0.0703	0.0265
OCSTNMRG	0.2102	0.0029	0.0006	Occupied	0.2716	3.12E-04	0.0001
OCSTWMTG	0.0882	0.0019	0.0002	Over65	0.4492	0.0048	0.0022
Female	0.1725	0.0013	0.0002	OwnerR	0.239	0.0017	0.0004
HighSch	0.1399	0.0017	0.0002	Poverty	0.1125	5.91E-04	0.0001
HhSize	0.122	0.0037	0.0005	Rent15	0.3342	9.28E-04	0.0003
MedValue	0.2524	0.0091	0.0023	Rent35	0.3412	9.78E-04	0.0003
Married	0.2856	0.0017	0.0005	Under5	0.4875	0.0027	0.0013
MedIcm	0.1706	0.0067	0.0011	GasWell	0.2339	0.0309	0.0072
MedRent	0.2914	0.0017	0.0005	Pipeline	0.2479	0.0025	0.0006
MeanTime	0.2658	0.0164	0.0044	Road	0.246	0.0126	0.0031
OCST20	0.3406	0.0063	0.0021				

quantities, as shown in Equation 15.

$$\text{Potency Index}(k) = \text{coverage}(k)\cdot\text{loading}(k) = \frac{1}{N}\sum_{\beta_{ki}\neq 0}(\beta_{ki})^2. \quad (15)$$

The potency index indicates the standardized magnitude of an explanatory variable's coefficient for the whole study area. Table 4 shows the coverage, the loading, and the potency index of all thirty-five explanatory variables in the GWEN model. The higher the coverage, the loading, and the potency index, the more widespread and important the variable is. Table 5 lists the top ten variables derived from EN, stepwise, Step-AIC, GWR–Ridge, GWR–Lasso, and GWEN using the potency index as a criterion.

The top four variables listed in Table 5 are the same for all six methods, with only a slight difference in ranking for two variables (NonVehicle and Developed) between GWEN and GWR–Lasso and the other four methods. The variables Water (percentage of area in open water) and MedValue (median value of owner-occupied units) were picked by four methods but were ranked differently in those four methods. There is a certain degree of overlap in the other selected variables (MeanTime, Under 5, Employed).

Table 5. Comparison of the top ten variables selected by GWEN (using potency index), GWR–Ridge, GWR–Lasso, elastic net, Step-AIC, and stepwise.

Rank	GWR–Ridge	Stepwise	Elastic net	Step-AIC	GWR–Lasso	GWEN
1	Population	Population	Population	Population	Population	Population
2	Developed	Developed	Developed	Developed	NonVehicle	NonVehicle
3	NonVehicle	NonVehicle	NonVehicle	NonVehicle	Developed	Developed
4	Damages	Damages	Damages	Damages	Damages	Damages
5	MedValue	NonPlumbing	Under5	Under5	MeanTime	GasWell
6	OCST20	Under5	Employed	OCST20	GasWell	Water
7	NonMtg	Employed	Water	OCST35	MedValue	MeanTime
8	MeanTime	Pipeline	HighSch	Subsidence	MedInc	Road
9	Married	Water	MedRent	Hhsize	Water	Nonkitchen
10	MedInc	NonFuel	NonFuel	MedValue	Over65	MedValue

Note: GWEN = geographically weighted elastic net; GWR = geographically weighted regression; Lasso = least absolute shrinkage and selection operator; AIC = Akaike's information criterion.

Because GWEN is basically a local version of EN, the coefficient values of the top four variables in GWEN and EN are mapped and compared in Figures 7 through 10.

Figure 7 shows that the total population count in each cell generally had a negative effect on population growth. The EN model implies that this negative effect equally affects all of the geographical units, with a coefficient of –0.7122. In contrast, the GWEN model offers more details about the nonstationarity of this effect. In the GWEN model, this negative impact was mainly concentrated in the New Orleans metropolitan area as well as urban areas north of Lake Pontchartrain. In the metropolitan area of Baton Rouge, the total population count had no effects on the population changes (coefficients close to zero). In some rural areas, some positive effects of population count on population growth are observed.

Figure 8 shows that the percentage of housing units without any vehicle greatly hindered population growth. In the EN model, this kind of impedance was the same throughout the whole study area (with a coefficient of –0.2448), whereas large variance emerged in the GWEN model. The GWEN model shows that low population growth in the area coinciding with low percentage of vehicle possession was obvious in the New Orleans area, and it was slightly weakened in the urban areas north of New Orleans. The effect of the variable hardly existed in rural areas, however.

From Figure 9, the EN results show that the percentage of developed land area was positively associated with population growth for the whole study area with a coefficient value of 0.4442. The GWEN model shows that this positive association varied greatly across space, though. In the New Orleans area, there was no such effect, whereas in the urban areas north of Lake Pontchartrain, the effect was much more obvious.

Figure 10 shows that the property damages from natural hazards generally had a negative effect on population growth (with an EN coefficient of –0.2131), especially in the New Orleans area and some other coastal areas, where hurricanes and storm surges frequently strike (including Katrina). The variable had an unexpectedly high negative effect on population changes in the northwest corner of the study area, which could be due to higher vulnerability in the area. In contrast, the damage effect did not seem to affect other inland areas, noting that the population data used in this study is only up to 2010 and does not include the 2016 Louisiana floods. In the rural areas north of Lake Pontchartrain, the damage variable seemed to have a positive association with population growth. Other factors such as economic opportunities might have played a bigger role in the population growth

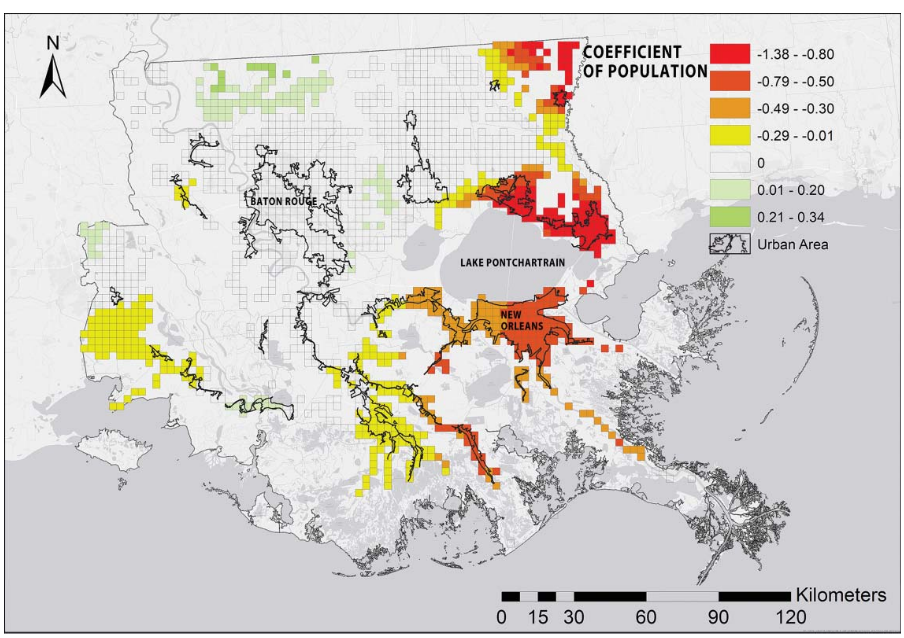

Figure 7. Coefficients of population count in the geographically weighted elastic net model (the elastic net coefficient is –0.7122). (Color figure available online.)

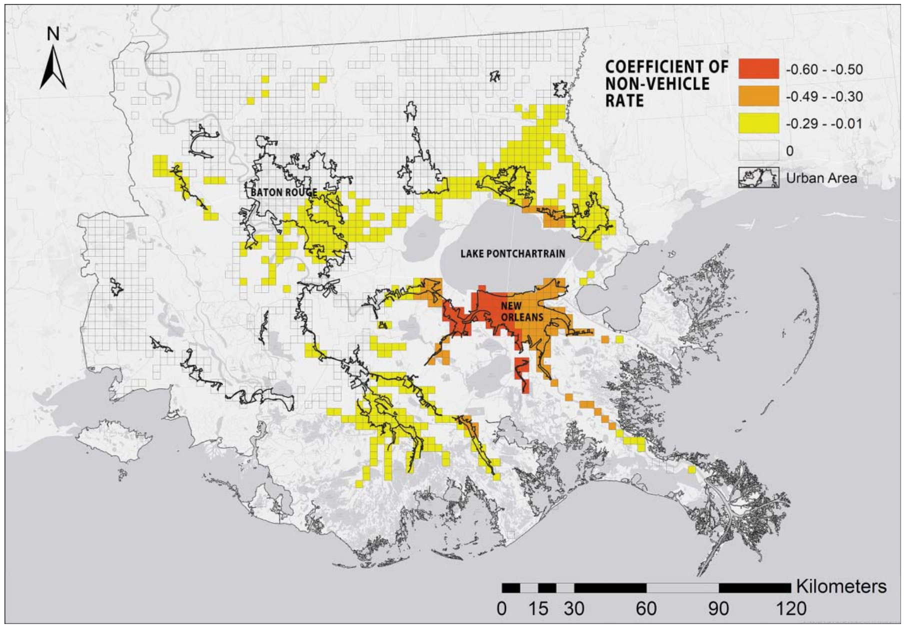

Figure 8. Coefficients of NonVehicle in the geographically weighted elastic net model (the elastic net coefficient is –0.2448). (Color figure available online.)

of the area than the damage resulting from natural hazards.

The comparison between GWEN and EN in Figures 7 through 10 shows that the GWEN model not only performs better in prediction accuracy and eliminating multicollinearity but also reveals the nonstationarity of the explanatory variables. As already mentioned earlier, though, the results from comparing GWEN with all the methods used in this study will need to be interpreted cautiously because the comparison was not based on the best performance of each method but rather a set of fixed parameters

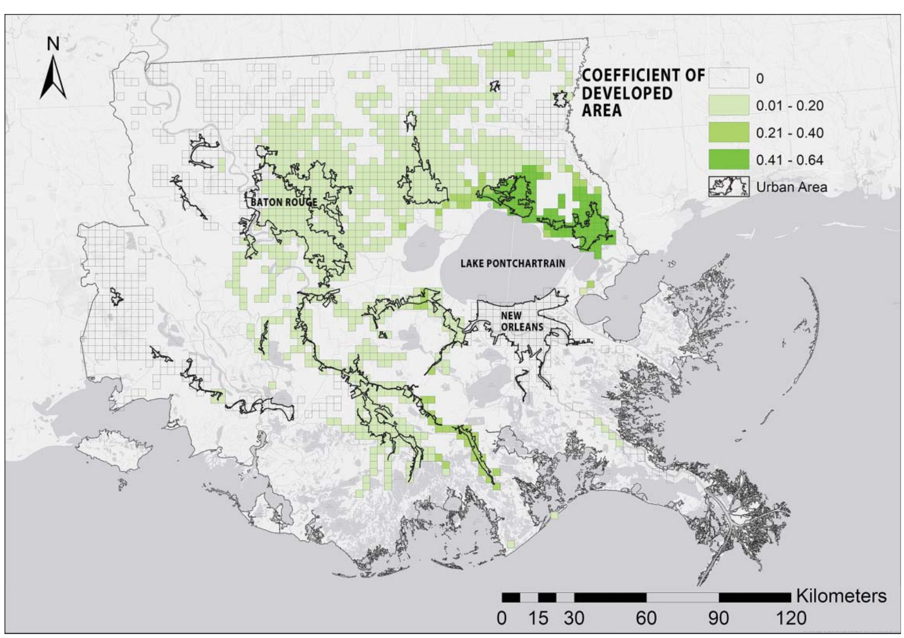

Figure 9. Coefficients of developed land areas in the geographically weighted elastic net model (the elastic net coefficient is 0.4442). (Color figure available online.)

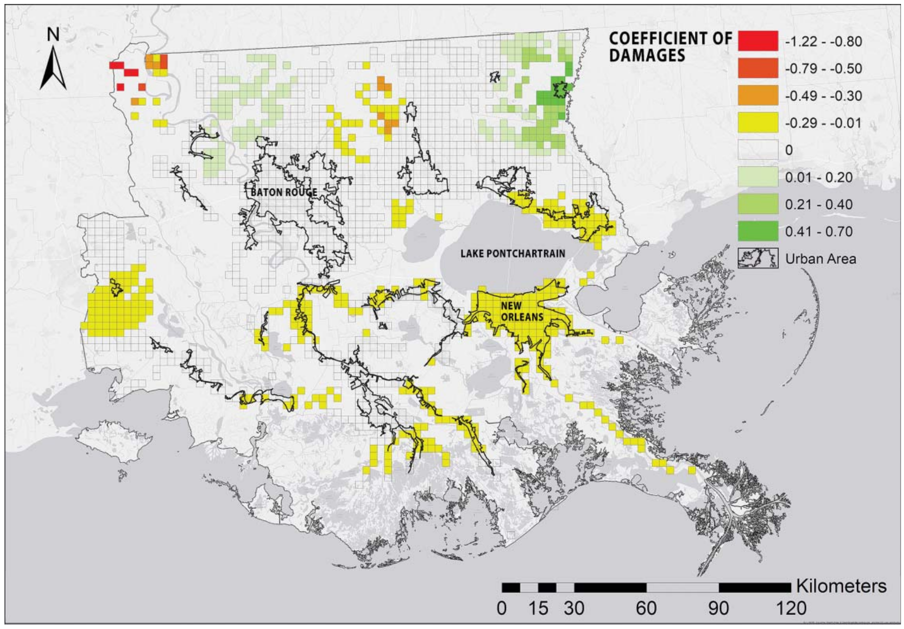

Figure 10. Coefficients of damages in the geographically weighted elastic net model (the elastic net coefficient is –0.2131). (Color figure available online.)

derived semioptimally from GWEN. Comparing these methods would pose a challenge, as some of the methods do not have the same properties as GWEN (local and low collinearity). Moreover, a more reliable comparison would require comparison using more than one real data set and possibly simulated data sets. These are important issues that future studies should address.

Conclusions

This study developed a new method, GWEN, which can be used to select variables to minimize multicollinearity while revealing local spatially varying relationships. We can view GWEN as a two-step method: first generating local submodels for each data observation using a kernel function like GWR and then solving an EN problem for each submodel using the weights from the first step. The study also compared GWEN with seven other regression methods using an empirical data set. The empirical study of population changes in the Lower Mississippi River Basin demonstrates the properties of GWEN and its advantages over a number of regression methods. The empirical study also reveals four variables identified by all eight regression models as the top predictors affecting population changes in the study area: population count, percentage of households with no vehicle, percentage of developed land, and property damage by natural hazards.

This study offers a useful tool for researchers who want to explore spatially varying relationships but are hindered by the local multicollinearity of using GWR or those who want to reduce the multicollinearity by selecting the best set of explanatory variables for ease of interpretation but are not satisfied with the accuracy of a global model such as EN. This method is especially useful when the study area is relatively large and the spatial nonstationarity is significant. Future studies should examine in greater detail the performance of GWEN and how it compares with the other methods using both simulated and more real data sets.

Funding

This article is based on work supported by two research grants from the U.S. National Science Foundation: one under the Dynamics of Coupled Natural Human Systems (CNH) Program (Award No. 121211) and the other under the Coastal Science, Engineering and Education for Sustainability (Coastal SEES) Program (Award No. 1427389). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies.

ORCID

Nina S. N. Lam http://orcid.org/0000-0002-5344-9368

References

Anselin, L. 1995. Local indicators of spatial associations—LISA. *Geographical Analysis* 27:93–115.

Bárcena, M., P. Menéndez, M. Palacios, and F. Tusell. 2014. Alleviating the effect of collinearity in geographically weighted regression. *Journal of Geographical Systems* 16 (4):441–66.

Belsey, D. A. 1991. *Conditioning diagnostics: Collinearity and weak data in regression*. New York: Wiley.

Brunsdon, C., A. S. Fotheringham, and M. E. Charlton. 1996. Geographically weighted regression: A method for exploring spatial nonstationarity. *Geographical Analysis* 28 (4):281–98.

Casetti, E. 1972. Generating models by the expansion method: Applications to geographical research. *Geographical Analysis* 4 (1):81–91.

Cromley, R. G., D. M. Hanink, and G. C. Bentley. 2012. A quantile regression approach to areal interpolation. *Annals of the Association of American Geographers* 102 (4):763–77.

Derksen, S., and H. J. Keselman. 1992. Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. *British Journal of Mathematical and Statistical Psychology* 45 (2):265–82.

Farrar, D. E., and R. R. Glauber. 1967. Multicollinearity in regression analysis: The problem revisited. *The Review of Economics and Statistics* 49 (1):92.

Fotheringham, A. S., C. Brunsdon, and M. Charlton. 2002. *Geographically weighted regression: The analysis of spatially varying relationships*. Chichester, UK: Wiley.

Fotheringham, A. S., M. E. Charlton, and C. Brunsdon. 1996. The geography of parameter space: An investigation into spatial non-stationarity. *International Journal of Geographic Information Systems* 10:605–27.

Fotheringham, A. S., M. H. Kelly, and M. Charlton. 2012. The demographic impacts of the Irish famine: Towards a greater geographical understanding. *Transactions of the Institute of British Geographers* 38 (2):221–37.

Fotheringham, A. S., and T. Oshan. 2016. Geographically weighted regression and multicollinearity: Dispelling the myth. *Journal of Geographical Systems* 18 (4):303–29.

Friedman, J., T. Hastie, H. Höfling, and R. Tibshirani. 2007. Pathwise coordinate optimization. *The Annals of Applied Statistics* 1 (2):302–32.

Friedman, J., T. Hastie, and R. Tibshirani. 2010. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* 33 (1):1–22.

Goodchild, M. F., and N. S. Lam. 1980. Areal interpolation: A variant of the traditional spatial problem. *GeoProcessing* 1:297–312.

Grafen, A., and R. Hails. 2002. *Modern statistics for the life sciences*. Oxford, UK: Oxford University Press.

Griffith, D. 1981. Towards a theory of spatial statistics: A rejoinder. *Geographical Analysis* 13:91–93.

———. 2008. Spatial-filtering-based contributions to a critique of geographically weighted regression (GWR). *Environment and Planning A* 40:2751–69.

Hoerl, A., and R. Kennard. 1988. Ridge regression. *Encyclopedia of Statistical Sciences* 8:129–36.

Kutner, M. H., C. Nachtsheim, and J. Neter. 2008. *Applied linear regression models*. Boston: McGraw-Hill.

Lam, N. S. 1983. Spatial interpolation methods: A review. *Cartography and Geographic Information Science* 10 (2):129–50.

———. 2009. Spatial interpolation. *International Encyclopedia of Human Geography* 10:369–76.

Li, H., C. A. Calder, and N. Cressie. 2007. Beyond Moran's I: Testing for spatial dependence based on the spatial autoregressive model. *Geographical Analysis* 39 (4):357–75.

Li, K. 2015. A spatial dynamic model of population changes in a vulnerable coastal environment. PhD dissertation, Louisiana State University and Agricultural and Mechanical College.

Liu, Z., and G. Li. 2016. Efficient regularized regression with L_0 penalty for variable selection and network construction. *Computational and Mathematical Methods in Medicine* 2016:1–11.

Qiang, Y., and N. S. Lam. 2015. Modeling land use and land cover changes in a vulnerable coastal region using artificial neural networks and cellular automata. *Environmental Monitoring and Assessment* 187 (3):57.

———. 2016. The impact of Hurricane Katrina on urban growth in Louisiana: An analysis using data mining and simulation approaches. *International Journal of Geographical Information Science* 30 (9):1832–52.

Shu, Y., N. S. Lam, and M. Reams. 2010. A new method for estimating carbon dioxide emissions from transportation at fine spatial scales. *Environmental Research Letters* 5 (4):044008.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 58:267–88.

Twilley, R. R., S. J. Bentley, Q. Chen, D. A. Edmonds, S. C. Hagen, N. S.-N. Lam, and C. S. Willson, et al. 2016. Co-evolution of wetland landscapes, flooding, and human settlement in the Mississippi River Delta Plain. *Sustainability Science* 11 (4):711–31.

Vidaurre, D., C. Bielza, and P. Larrañaga. 2012. Lazy lasso for local regression. *Computational Statistics* 27 (3):531–50.

Wheeler, D. 2007. Diagnostic tools and a remedial method for collinearity in geographically weighted regression. *Environment and Planning A* 39 (10):2464–81.

———. 2009. Simultaneous coefficient penalization and model selection in geographically weighted regression: The geographically weighted lasso. *Environment and Planning A* 41 (3):722–42.

Wheeler, D., and C. Calder. 2007. An assessment of coefficient accuracy in linear regression models with spatially varying coefficients. *Journal of Geographical Systems* 9 (2):145–66.

Wheeler, D., and M. Tiefelsdorf. 2005. Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *Journal of Geographical Systems* 7 (2):161–87.
Whittingham, M. J., P. A. Stephens, R. B. Bradbury, and R. P. Freckleton. 2006. Why do we still use stepwise modelling in ecology and behavior? *Journal of Animal Ecology* 75 (5):1182–89.
Wright, S. J. 2015. Coordinate descent algorithms. *Mathematical Programming* 151 (1):3–34.
Zou, H., and T. Hastie. 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67 (2):301–20.

KENAN LI is a Postdoctoral Research Associate in the Department of Preventive Medicine (Bio-stats Division) at University of Southern California, Los Angeles, CA 90033. E-mail: kenanl@usc.edu. His research interests include agent-based modeling of the sociosystems, spatial and temporal analysis of the environmental health issues, and developing integrated sensor monitoring systems for epidemiological studies.

NINA S. N. LAM is Professor and E. L. Abraham Distinguished Professor of Louisiana Environmental Studies in the Department of Environmental Sciences at Louisiana State University, Baton Rouge, LA 70803. E-mail: nlam@lsu.edu. Her research interests include GIScience, remote sensing, environmental health, resilience, and sustainability.

Appendix: The GWEN Algorithm

i. Determine the spatial variation of the observation's locations and define either a global kernel bandwidth or a global number of neighborhood observations. In the case of using a global bandwidth, start with the user-defined tuning range and the increment and create a set of kernel bandwidths from the minimum bandwidth to the maximum bandwidth: r_1, r_2, ..., r_n. In the case of using a global number of neighborhood observations, start with the user-defined tuning range and the increment and create a set of number of neighbors from the minimum to the maximum: $n_1, n_2, \ldots, n_n$.

ii. Create a set of αs using the user-defined or default increment from 0.1 to 1: $\alpha_1, \alpha_2, \ldots, \alpha_m$.

iii. Create matrix *RMSE* (*i*, *j*) to document the cross-validated RMSE of the GWEN under α_i, r_j.

iv. For each $i = 1, \ldots, n$, do the following outer loop:

For each $j = 1, \ldots, m$, do the following middle loop:

1. Use the user-provided or default positive integer *K* to divide the observations into *K* groups for *K*-fold cross-validation. The cross-validation process (the inner loop or Step 2) will be repeated *K* times (folds), with each of the *K* subgroups used exactly once as the test (validation) data. The *K* results are then averaged to produce a single accuracy estimation.

2. Using the cross-validation partitions from Step 1 (*K*-folds), for each observation in the training data set (the subgroups other than the test subgroup), do the inner loop:

a. Use r_i or n_i to identify the neighbors of the observation and calculate the weights for these observations using the user-defined kernel function. Use α_j and the calculated weights to build the GWEN problem described by Equations 8 and 9.

b. Identify the maximum value of λ, at which the GWEN problem only gives the null solution (all of the variables are not selected). Using the user-defined or the default λ increment values, create a series of λ from 0 to the maximum.

c. For each λ, use the coordinate descent algorithm (Wright 2015) to solve the GWEN problem and document the solution with the λ value.

d. According to the user-defined number of variables to be selected, find the solution that gives the closest match and the related λ. Document the coefficients, the explanatory variables' VIF values, and the prediction residuals of this solution.

e. Use the test group and calculate the RMSE using the coefficients derived from Step d for cross-validation.

End the inner loop.

3. After iterating Step 2 for all of the *K* cross-validation partitions, average the *K*-fold cross-validated RMSEs calculated from each of the partitions and set it to $RMSE_{ij}$.

End the middle loop.

End the outer loop.

v. Provide the user with the matrix RMSE as a decision support to select the bandwidth and α that will offer the lowest cross-validated RMSE. The coefficients, the explanatory variables' VIF values (Equation 11), and the prediction residuals of the solution under the selected bandwidth and α will be provided as the final results of running the GWEN.