



Geographically weighted elastic net logistic regression

Alexis Comber¹ · Paul Harris²

Received: 9 July 2017 / Accepted: 12 September 2018 / Published online: 26 September 2018
© The Author(s) 2018

Abstract

This paper develops a localized approach to elastic net logistic regression, extending previous research describing a localized elastic net as an extension to a localized ridge regression or a localized lasso. All such models have the objective to capture data relationships that vary across space. Geographically weighted elastic net logistic regression is first evaluated through a simulation experiment and shown to provide a robust approach for local model selection and alleviating local collinearity, before application to two case studies: county-level voting patterns in the 2016 USA presidential election, examining the spatial structure of socio-economic factors associated with voting for Trump, and a species presence–absence data set linked to explanatory environmental and climatic factors at gridded locations covering mainland USA. The approach is compared with other logistic regressions. It improves prediction for the election case study only which exhibits much greater spatial heterogeneity in the binary response than the species case study. Model comparisons show that standard geographically weighted logistic regression over-estimated relationship non-stationarity because it fails to adequately deal with collinearity and model selection. Results are discussed in the context of predictor variable collinearity and selection and the heterogeneities that were observed. Ongoing work is investigating locally derived elastic net parameters.

Keywords GWR · GW-ELNR · Elastic nets · Collinearity · Model selection

JEL Classification C14 · C31

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10109-018-0280-7>) contains supplementary material, which is available to authorized users.

✉ Alexis Comber
a.comber@leeds.ac.uk

¹ School of Geography, University of Leeds, Leeds LS2 9JT, UK

² Rothamsted Research, North Wyke, Okehampton, Devon EX20 2SB, UK

1 Introduction

This paper develops a geographically weighted elastic net logistic regression (GW-ENLR). The context and rationale for this binary response predictor or classifier derive from two areas of interest or concern in regression modelling in geography: (a) non-stationarity in data relationships and (b) collinearity and model selection in the predictor variable data set. For issue (a), this can be dealt with using an explicitly local framework such as the geographically weighed (GW) modelling approach proposed by Fotheringham et al. (2002), where issue (b) is now similarly localized commonly requiring some penalized regression adaptation.

In applications of regression to spatial data, two effects are often of most concern: (1) spatial autocorrelation in the error term and (2) spatial heterogeneity of the regressions coefficients. The former contradicts the underlying assumptions of independence and identically distributed errors, whilst the latter contradicts the assumption of fixed data relationships, when applying a standard regression. The former commonly requires the incorporation of a *stationary* autocorrelation parameter in the regression model, whilst the latter allows the regression coefficients to vary across space—i.e. they are *non-stationary*. In either case, the objective is the same, that is to provide more accurate or realistic coefficients together with associated measures of uncertainty. Often it is difficult to identify or separate one effect from the other, as both effects tend to be strongly inter-linked (e.g. Anselin 1990) and some models aim to capture both effects (e.g. Brunsdon et al. 1998). The focus of this paper lies with effect (1) in accounting for spatial heterogeneity between response and predictor data relationships. In this respect, a GW modelling approach is followed, where localized versions of a generalized linear model (GLM) with a logit link function (Fotheringham et al. 2002) are applied (i.e. a GW logistic regression, GW-LR, as the response is binary) and developed (i.e. via the new geographically weighted elastic net logistic regression (GW-ENLR) model), which are themselves extensions of the classic GW regression (GWR) (Brunsdon et al. 1996) for a Gaussian response.

Model selection is an important component of any regression model construction. It involves identifying which predictor variables to include and/or techniques for transforming and reducing the predictor subset. In the context of this study, model selection is directly linked to the reduction in unwanted collinearity effects among the predictors. Failure to correctly specify a model when collinearity is present can result in a loss of precision and power in the coefficient estimates—leading to poor inferences. This risk commonly increases as more predictors are introduced. Collinearity occurs when pairs of predictor variables have a strong positive or negative relationship with each other and is typically considered a potential problem when these data pairs have correlations of less than -0.8 or greater than $+0.8$, say. Approaches for addressing collinearity include transformations of the predictors such as that found in principal components analysis (PCA) regression, so that the transform negates collinearity altogether, or a related approach, such as that provided by partial least squares regression

(e.g. Frank and Friedman 1993). In this study, a penalized approach is taken via the elastic net, a hybrid of ridge regression and the lasso (i.e. the least absolute shrinkage and selection operator) (Zou and Hastie 2005). Only the lasso and the elastic net additionally provide a model selection function.

Issues of collinearity can be compounded under GW regression models, which use subsets of the data to construct a series of localized regressions. This is because predictor variables in the local subset may be collinear even when collinearity is not observed globally (Wheeler and Tiefelsdorf 2005; Wheeler 2013). For these reasons, a number of extensions have been proposed to address local collinearity including ridge GWR (Wheeler 2007; Brunson et al. 2012; Bárcena et al. 2014; Gollini et al. 2015) and the ‘adjusted/enriched data subset’ GWR models of Brunson et al. (2012) and Bárcena et al. (2014). For addressing collinearity *and* model selection, the GW lasso model has been a natural development (Wheeler 2009; Wang and Li 2017) and more recently, the GW elastic net (Li and Nam 2018), where local collinearity *and* local model selection are addressed. All such models are for Gaussian response variables only.

For other response types, via a GW GLM framework, only a GW logistic lasso exists (Yoneoka et al. 2016). This paper now extends this suite of GW GLM approaches further by proposing an elastic net adaptation to GW-LR (GW-ENLR).¹ It applies this new model to two geographical case studies, one using election data and the another using environmental data, with the aim of improving prediction/classification accuracy and improving understanding of the processes being modelled through the local selection of key predictors that best explain the process, locally. Furthermore, and unlike the studies of Yoneoka et al. (2016) and Li and Nam (2018) with their respective GW logistic lasso and Gaussian-response GW elastic net models, this study evaluates the new GW-ENLR model through a simulation experiment, as similarly advocated in the Gaussian-response GW lasso study of Wheeler (2009).

2 Background: collinearity, ridge regression, the lasso and elastic net

Reliable inference from any regression relies somewhat on the independence of the predictor variables. If these are correlated or collinear (i.e. the columns of the design matrix have an approximate linear dependence), then the regression can become sensitive to random errors in the observed response producing a large variance and reducing model inferential power. Thus, model reliability and precision are affected, resulting in unstable coefficient estimates, inflated standard errors and inferential biases (Dormann et al. 2013). Critically, model extrapolation may be erroneous and there may be problems in separating predictor variable effects (Meloun et al. 2002). For this study, a penalized approach is taken to address adverse effects of predictor collinearity. In particular, an elastic net approach is followed, which is itself an extension or hybridization of the lasso

¹ Such a model has been suggested (e.g. Yoneoka et al. 2016) but not yet been implemented, demonstrated or objectively evaluated, as is the case here.

and ridge regression. Both the lasso and the elastic net also provide predictor variable subset selection.

Ridge regression (Hoerl 1962; Hoerl and Kennard 1970) addresses predictor collinearity by altering the estimator to include a small change to the values of the diagonal of the cross-product matrix, referred to as the ridge. The off-diagonal elements of the cross-product matrix describe the co-variation in the predictor variables for any regression, where the effect of the additional ridge term is to increase the difference between the diagonal and off-diagonal elements of the matrix, thereby reducing the collinearity among the predictors. Unfortunately, there is a cost to implementing these small changes to the diagonal: the estimator becomes biased and the standard errors of the estimates (and t values) are no longer available (which is similarly the case for the lasso and the elastic net). Lasso approaches (Tibshirani 1996) provide model selection and dimensionality reduction in order to overcome model over-fitting and predictor variable collinearity, respectively. The lasso is a shrinkage estimator and generates coefficient estimates that are biased to be small, effectively including a penalty term to constrain them. It resembles ridge regression, but operates by simultaneous continuous variable shrinkage and automatic predictor variable selection. It has some limitations as noted by Zou and Hastie (2005): saturating with short fat data ($p > n$); randomly selecting one of the correlated predictor variables; is potentially dominated by the ridge when the data are long and thin ($n > p$). The elastic net (Zou and Hastie 2005) is a hybrid of ridge regression and lasso regularization. Like the lasso, elastic net can result in model reduction by generating zero-valued coefficients so that the corresponding predictor variable drops. Empirical studies have suggested that the elastic net technique can outperform the lasso (with respect to prediction accuracy) on data with highly correlated predictors (Friedman et al. 2010).

Collinearity can be a more difficult problem to address in localized regressions such as GWR and GW GLMs, as region-specific subsets of the data may exhibit collinearity, even when none is observed globally. In this respect, various GW ridge, GW lasso and GW elastic net models have been proposed (Wheeler 2007, 2009; Brunsdon et al. 2012; Bárcena et al. 2014; Gollini et al. 2015; Yoneoka et al. 2016; Wang and Li 2017; Li and Nam 2018). This study extends this existing suite of models through the development of a GW elastic net logistic regression (i.e. GW-ENLR) for understanding local structure in the data and supporting local variable selection, when the response variable is of a binary form.

3 Methods

To situate the new GW-ENLR approach, details on logistic regression (LR), elastic net logistic regression (ENLR) and GW logistic regression (GW-LR) are also provided. These regressions will also be fitted to the case study data sets, for context and to demonstrate the benefits of the new GW-ENLR model.

3.1 Logistic regression (LR)

An LR is a specific GLM with a logit link function of any number Q which is defined as:

$$\text{logit}(Q) = \frac{\exp(Q)}{1 + \exp(Q)} \quad (1)$$

$$\text{pr}(y_{i1} = 1) = \text{logit}\left(\beta_0 + \sum_{k=1}^m \beta_k x_{ik}\right) \quad (2)$$

where y_{i1} is a 0/1 indicator at location i , β_0 is the intercept term, x_{ik} is the value of the k th predictor variable at location i , m is the number of predictor variables and β_k is the regression coefficient for the k th predictor variable. Observe that although the data are associated with locations (or observations) i , spatial effects are not accounted for in this basic regression (i.e. spatial effects are naively assumed as being unimportant).

3.2 Elastic net logistic regression (ENLR)

The elastic net (Zou and Hastie 2005; Hastie and Qian 2014) is a model selection technique, that seeks to identify which predictor variables to include in a regression model. For a parameter α strictly between zero and one, and a nonnegative regularization parameter λ , the elastic net objective function for an LR (ENLR) uses the negative binomial log-likelihood:

$$\min_{\beta_0, \beta} - \left[\frac{1}{N} \sum_{i=1}^N y_{i1} (\beta_0 + x_i^T \beta) - \log \left(1 + \exp(\beta_0 + x_i^T \beta) \right) \right] + \lambda \left[\frac{1}{2} (1 - \alpha) \|\beta\|_2^2 + \alpha \|\beta\|_1 \right] \quad (3)$$

where N is the number of observations (or sample locations in this study's context), the coefficients β_0 and β are scalar and m -vector, respectively, and α and λ denote L_1 - and L_2 -norms, respectively. The problem is solved over a grid of values of λ covering the entire range. The elastic net penalty is controlled by α , and bridges the gap between the lasso and ridge regression: the elastic net is equivalent to the lasso when $\alpha=1$ and as α decreases towards zero, the elastic net approaches a ridge regression. The tuning parameter λ controls the overall strength of the penalty and thus determines which predictors should remain in the regression. The choice of λ can be guided via an optimum value found by a n -fold cross-validation (CV) minimization approach (see Hastie and Qian 2014), whose outcome can be tailored (i.e. set higher or lower) depending on the number of predictors that the analyst wishes to retain. If λ is set to zero, then the elastic net is simply equivalent to its standard regression (in

this case, an LR). The lasso and ridge regression penalties are often referred to as the L_1 -norm and L_2 -norm penalties, respectively (i.e. the elastic net is a combination of L_1 -norm and L_2 -norm penalties). In applications of ENLR, it is also possible to set limits on the coefficients themselves (e.g. for instances where only positive coefficients are sought) and find the associated penalties, accordingly. It is also possible to apply separate penalty factors to each coefficient, which can be useful when a given predictor is so important, it needs to be retained in the model, regardless.

3.3 GW logistic regression (GW-LR)

For GW models, a moving-window kernel is used, where data falling under the kernel are weighted by their distance to the kernel centre using a distance–decay function. These weighted data subsets are then used to calculate location-specific models or statistics, for example, basic local linear regressions with GWR (Brunsdon et al. 1996), local summary statistics (Brunsdon et al. 2002), local PCAs (Harris et al. 2011), and more recently, local contingency matrices (Comber et al. 2017). Outputs from a GW model are then mapped to provide an assessment of process heterogeneity, for example the local coefficients from GWR, the local loadings from a GW PCA.

A GW-LR (Fotheringham et al. 2002) is similar in form to the LR model given in expression (2), except that a GW-LR has locations associated with the model intercept and coefficient terms, as follows:

$$pr(y_{i1} = 1) = \text{logit} \left(\beta_{i0} + \sum_{k=1}^m \beta_{ik} x_{ik} \right) \quad (4)$$

where β_{i0} is the intercept term at location i , and β_{ik} is the local regression coefficient for the k th predictor variable at location i . As a GW-LR provides local intercepts and local coefficients, they can be mapped, along with their standard errors to investigate the nature and importance of relationship non-stationarity in the data (e.g. see Atkinson et al. 2003; Windle et al. 2010; Rodrigues et al. 2014).

The kernel bandwidth in any GW model can be defined as a fixed distance or as an adaptive distance where a fixed number of local data points are used for each local model calculation. The size of the bandwidth is critical as it controls how data nearer to the kernel centre contribute to the local calculation than data farther away. Small bandwidths give greater weighting to nearby data than large bandwidths do. If a bandwidth is very large relative to the sample area (for fixed) or the sample size (for adaptive), then the process under investigation is likely stationary and the usual global model suffices. Unless good reason to, it is usually ill-advised to choose a bandwidth subjectively, and for any GW regression, a bandwidth can objectively be found by minimizing a model fit diagnostic. Options include a onefold CV score (Brunsdon et al. 1996), which optimizes model prediction accuracy, and an Akaike information criterion (AIC) (Fotheringham et al. 2002) approach, which optimizes model parsimony by trading off prediction accuracy with model complexity. In this

study, a CV approach was used to calibrate the GW-LR models, all using a distance–decay bi-square kernel, where weights $w_{i,j}$ are calculated as follows:

$$w_{i,j} = 1 - ((d_{i,j})^2/b^2) \quad (5)$$

where $d_{i,j}$ is the distance from the centre of the kernel at location j , to the observation point at location i , and b is the bandwidth. Once the bandwidth is determined using only the observed data at locations j and i , that bandwidth can be used to find the GW model outputs at all locations (observed and unobserved).

3.4 GW elastic net logistic regression (GW-ENLR)

The GW-ENLR model is a direct analogy to GW-LR in expression (4), but now with a locally defined elastic net form from expression (3). The implementation of GW-ENLR simply requires the fitting of an ENLR to each location-specific GW data subset until all regression locations are visited. Thus, expression (3) is now location-specific, with:

$$\min_{\beta_0, \beta_i} - \left[\frac{1}{N} \sum_{i=1}^N y_{i1} (\beta_0 + x_i^T \beta_i) - \log \left(1 + \exp(\beta_0 + x_i^T \beta_i) \right) \right] + \lambda \left[\frac{1}{2} (1 - \alpha) \|\beta_i\|_2^2 + \alpha \|\beta_i\|_1 \right] \quad (6)$$

A key issue in any GW regression that seeks to address local collinearity and/or local model selection is that the nature of collinearity/model selection will vary from data subset to subset (i.e. it is location-specific) and as such, is inherently dependent on the kernel bandwidth b , which sets the local scale of each individual regression fit. In this respect, the penalty parameters of a ridge, a lasso and an elastic are commonly jointly estimated with the bandwidth by minimizing some CV-based objective function. For such penalties, this optimization can be done in a global fashion (Wheeler 2007; Yoneoka et al. 2016; Wang and Li 2017; Li and Nam 2018) or a local fashion (Wheeler 2009; Brunson et al. 2012; Bárcena et al. 2014; Gollini et al. 2015), where the latter can always be set to default to the global approach. To date, the bandwidth has always been globally defined in a penalized GW regression,² where the use of a fixed and an adaptive distance bandwidth provides different options in this respect.³ The local approach has the benefit in that it tries to ensure a given set of local penalties directly suit the local data structure, whilst the global approach only provides an ‘on average’ penalty solution (i.e. it is only locally appropriate in a broad global sense). Conversely, the global approach can provide results, that when mapped are more easily interpreted or compared, as the penalties are the

² Only the studies of Brunson et al. (2012), and Bárcena et al. (2014) additionally investigate locally defined bandwidths to address local collinearity, but now without any penalty term.

³ Li and Nam (2018) experiment with fixed and adaptive bandwidths, sometimes referring to the latter as local. This is not exactly the case, as the former is globally set by distance, whilst the latter is globally set by the number of neighbours.

same everywhere. Regardless of which approach is taken, the understanding of relationship non-stationarity should only ever be enhanced (not compromised) by the additional complexity of the penalized GW regression adaption. Comparison with the corresponding basic GW regression is always recommended.

Of course, of the three penalized approaches, the elastic net provides the greatest challenge when used in a GW regression, as the bandwidth b , the tuning parameter λ (reflecting the strength of the penalty) and the parameter α (reflecting a lasso or ridge regression) all need to be derived in some manner. In the Gaussian-response case, Li and Nam (2018) provide a CV-based coordinate decent solution, where b and α are globally derived, but instead of globally deriving λ , the number of predictor variables retained for each local regression is globally derived instead. The latter of which indirectly assures local values of λ , whereas if λ were to be globally derived, then the number of retained predictors for each local regression would be location-specific instead.

For this GW-ENLR study, a basic onefold CV approach is used to find the bandwidth b only (i.e. the same as that used in GW-LR), where the elastic net parameters α and λ are both globally preset through a series of preliminary investigations of the GW-ENLR outputs. This is considered appropriate in the spirit of spatial exploration and sits with this study's goal with respect to the introduction of GW-ENLR. Adaptation of the sophisticated CV-based calibration procedure provided by Li and Nam (2018) in their Gaussian-response model (discussed above) should be straightforward for the purposes of GW-ENLR. However, such an adaptation would only provide global solutions, whereas true local solutions, especially akin to that provided by Brunson et al. (2012); Bárcena et al. (2014); Gollini et al. (2015) would present a challenge. That said, work is ongoing for investigating locally derived α and λ elastic net parameters, through extending the locally compensated ridge GWR model, introduced in Brunson et al. (2012) and fully described in Gollini et al. (2015).

4 Case studies

To demonstrate the value of the new GW-ENLR method, two United States (US) data sets were selected: (a) socio-economic data linked to voting patterns for the 2016 US presidential elections at the US county level, and (b) US environmental data of species presence-absence linked to climatic data. Here LR, ENLR, GW-LR and GW-ENLR are all used to analyse the extent to which the predictor data sets can help classify Trump supporting counties and help predict species presence-absence, respectively. For applying the GW-based models, the expectations are that relationships in these data sets are non-stationary. The respective binary responses were: (1) Trump support, with a value of one indicating that Trump received the most votes in a county, zero otherwise, and (2) species absence, with a value of one indicating species absence, zero otherwise. Each data set has five predictor variables, where a certain degree of collinearity is expected, warranting a penalized component to the study models.

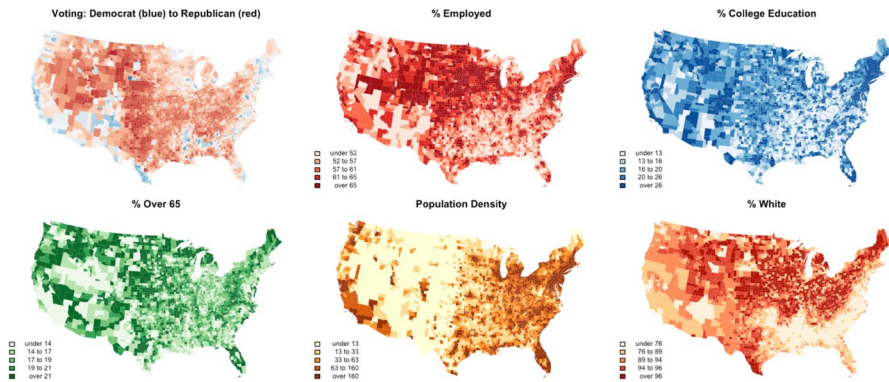


Fig. 1 The election data and socio-economic attributes from the US population census

The predictors used to examine county-level Trump support and species presence–absence were selected simply to illustrate the GW-ENLR method. The focus of the paper is not to show insight into these phenomena or describe them. Rather, the aim is to use case studies from socio-economic and environmental domains for illustration purposes. The predictors for voting were selected because they have been identified by others as factors in Trump support. The species data predictors were selected as the key drivers of plant development (and therefore reproduction and persistence)—thus these should act as valuable explainers of species presence–absence.

4.1 US election data

The US election data were constructed from two sources. First, voting data for each of the 3,108 mainland counties in the US was downloaded from Tony McGovern's Github site (McGovern 2017) and then linked to county outlines in the maps R package (Becker et al. 2016). Next the population census data was downloaded for each county from the US Census Bureau 'QuickFacts' website (US Census 2017) and the following five predictor variables were extracted for each county:

- **% Employed:** In civilian labour force, total, percentage of population age 16 years +, 2011–2015;
- **% College Education:** Bachelor's degree or higher, percentage of persons age 25 years +, 2011–2015;
- **% Over 65:** Persons 65 years and over, percentage, 1 July 2015;
- **Population Density:** Population per square mile, 2010;
- **% White:** White alone, percentage, 1 July 2015.

These socio-economic attributes were then linked to the US counties. A binary response variable was created to indicate whether the county was a Republican, Trump supporting county or not. This indicated the counties where Trump received the most votes. The voting patterns (response) and socio-economic attributes (predictors) are mapped in Fig. 1.

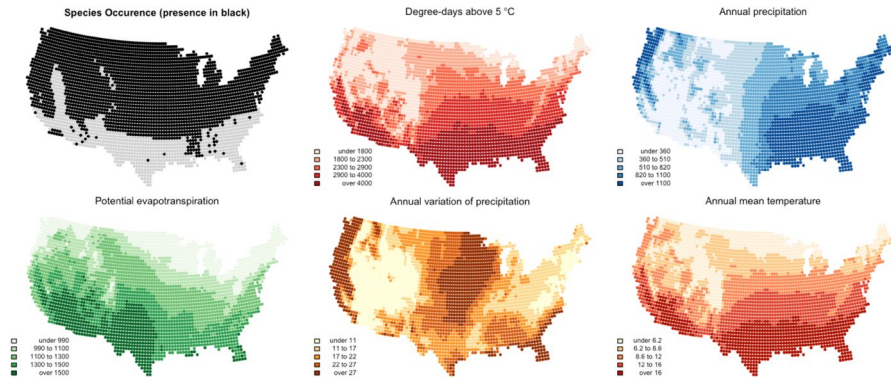


Fig. 2 The presence/absence species data and the environmental predictor variables

4.2 US species data

A second US data set was extracted from the *ecospat* R package (Broennimann et al. 2016). The *ecospat.testNiche.nat* data set covers no specific year and is described as being ‘test data for the niche dynamics analysis in the native range of a hypothetical species’. It includes environmental predictor variables and a binary response variable indicating species presence/absence records for the occurrence of a number of species at 3,259 locations on a grid. The data cover the continent of North America, with an approximate grid spacing of 37.2 km and only records for the mainland US were extracted. In this analysis, the following five environmental attributes were considered as predictors of species presence–absence:

- Degree-days above 5 °C;
- Annual amount of precipitation;
- Potential evapotranspiration;
- Annual variation of precipitation;
- Annual mean temperature.

The spatial distributions of the presence–absence indicator and the environmental predictor variables are shown in Fig. 2.

4.3 Global relationships

An initial data exploration was undertaken to examine the global relationships between the five predictor variables and the binary responses for US counties supporting Trump and for US species presence–absence. The conditional box-plots in Fig. 3 clearly show that most of the predictor variables can be used to discriminate between their respective binary responses, for each case study. The weakest discriminatory powers in this global assessment are found for *Annual*

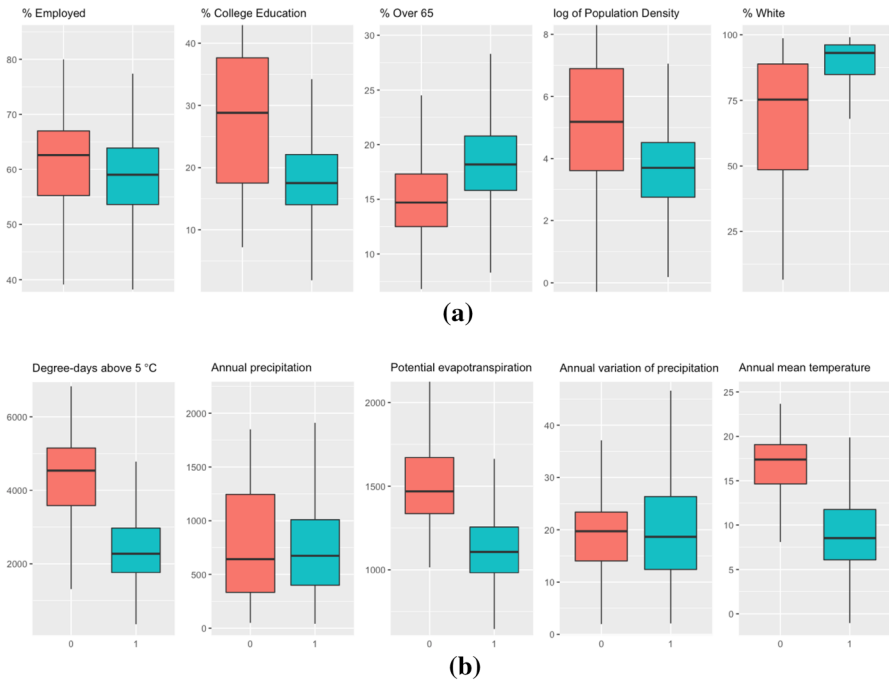


Fig. 3 Conditional boxplots showing the discriminatory powers of each predictor with respect to: **a** counties supporting Trump (blue) versus not supporting Trump (red), and **b** species presence (blue) versus species absence (red) (colour figure online)

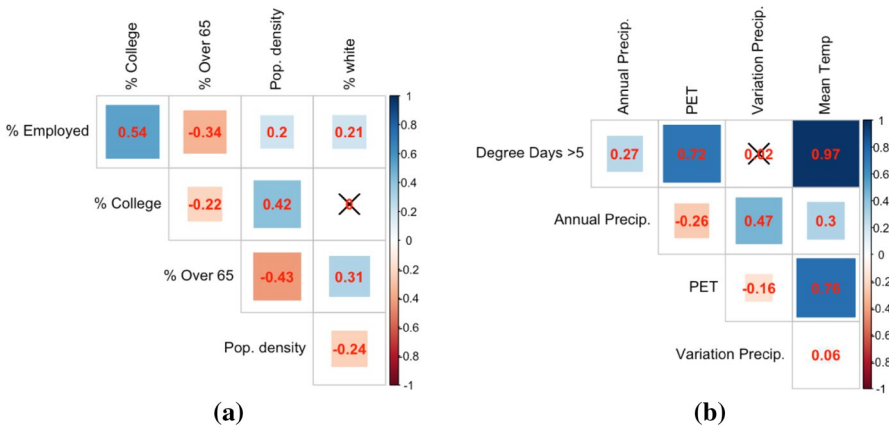


Fig. 4 The correlation matrices of the predictors for **a** the election case study, and **b** the species case study. Note that, insignificant correlations are indicated with cross; the following abbreviations have been used: *Pop.* for population, *precip.* for precipitation, *PET* for potential evapotranspiration

variations in precipitation and *Annual amounts of precipitation* in explaining the species presence–absence response.

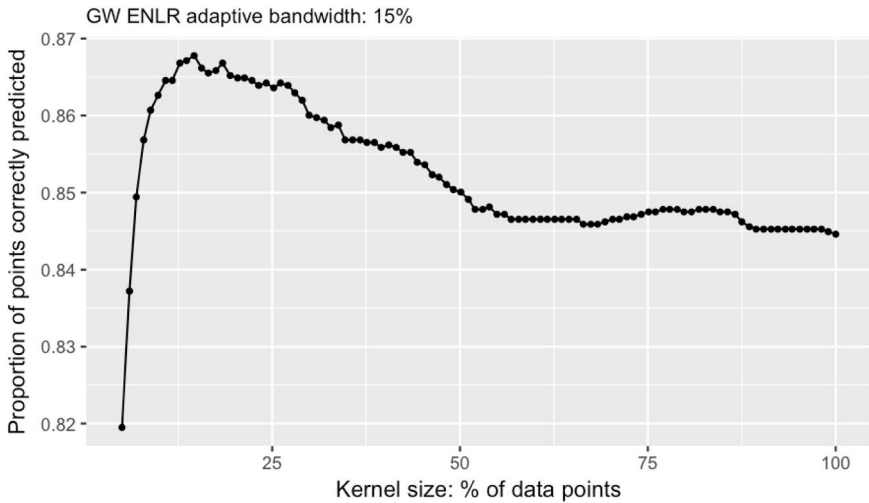
Global collinearity for each case study was measured via the design matrix condition number (CN) where a CN greater than 30 suggests the presence of ‘significant’ collinearity amongst the predictor variables (Belsley et al. 1980; Gollini et al. 2015). These were found to be 18.5 and 31.0 for the election and species case study, respectively. The correlation matrices for the predictors are also given in Fig. 4, for both case studies. Clearly, there is some evidence of global collinearity, supporting the use of penalized regressions to the species data set. This is less evident in the election data set, but as with both studies, collinearity may be stronger locally than that observed globally.

4.4 Study regressions: LR, ENLR, GW-LR and GW-ENLR

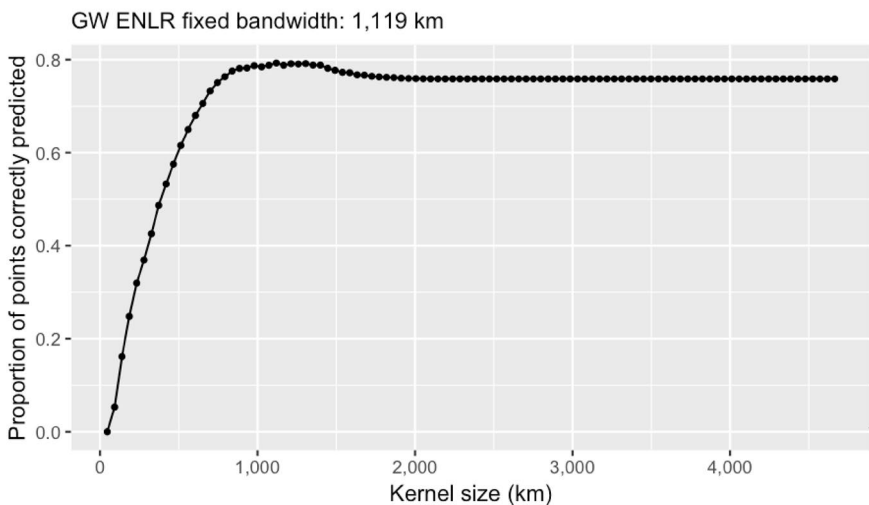
A series of regression analysis were undertaken to compare the LR, ENLR, GW-LR and GW-ENLR models as described in Sect. 3. For both case studies, the predictor variables were linearly rescaled to [0.001, 1], and in the case of the species presence–absence data, absence was modelled. To determine the optimal bandwidth for GW-LR and for GW-ENLR, in each of the case studies, sequences of bandwidths were passed to GW-LR and GW-ENLR and the same leave-one-out CV procedure applied. The US election case study applied an adaptive bandwidth ranging from 1 to 100% in steps of one percentage point and the US species case study applied a fixed bandwidth ranging from 50 to 4,600 km in steps of 50 km. For each case study, the optimal bandwidths were those that resulted in the highest proportion of correct (leave-one-out) classifications/predictions and were passed to the final GW-LR and GW-ENLR calibrations.

The optimal (adaptive) bandwidth for the election GW-LR model was estimated at 11%: 327 counties out of 3,108, whereas the optimal (fixed) bandwidth for the species presence–absence GW-LR model was estimated at 418 km, or approximately 9% of the maximum potential bandwidth of 4,600 km. This suggests that there is a high degree of spatial variation in the association between the predictor variables and the response variable in both the election and species presence–absence case studies. Interestingly, the optimal (adaptive) bandwidth for the election GW-ENLR model increases to 15% of nearest data, whilst the optimal (fixed) bandwidth for the species presence–absence GW-ENLR model has a relatively large increase to 1,119 km. This suggests that failing to adequately deal with collinearity (and model selection), the perception of relationship non-stationarity is stronger than it actually is.

Plots of the bandwidth versus the proportions of correctly predicted data points (i.e. the CV score) are shown in Fig. 5 for both case studies, but for GW-ENLR only. The election case study has a clear maximum around 15% with some noise as the bandwidth increases and declines afterwards. For the species case study, the maximum is not so pronounced. This is not surprising as the species occurrence map indicates a strong north–south divide, entailing a likely difficulty in fitting many local ENLRs of the GW-ENLR model. (As for many of the small data subsets



(a)



(b)

Fig. 5 The proportions of data point correctly predicted for different bandwidths for: **a** the election case study, and **b** the species presence–absence case study

residing in either the north or the south, the binary response has no variance—consisting only of ones or only of zeros.) This is in contrast to the election case study where there is greater spatial heterogeneity in the binary response. Both plots indicate that a GW-ENLR model outperforms its global counterpart in ENLR, as maximums do not indicate a stationary process.

For each case study, Table 1 summarizes the accuracy results for LR, ENLR, GW-LR and GW-ENLR using the proportion of correctly predicted responses by

Table 1 Proportion of correctly predicted responses for both case studies, for all four study regressions: logistic regression (LR), elastic net logistic regression (ENLR), a geographically weighted logistic regression (GW-LR) and geographically weighted elastic net logistic regression (GW-ENLR). The GW-LR and GW-ENLR bandwidths are also given, as are the ENLR and GW-ENLR penalty parameters

Model	Election case study			Species case study		
	Accuracy (%)	B (%)	α and λ	Accuracy (%)	b (km)	α and λ
LR	65.5	–	–	77.3	–	–
ENLR	88.7	–	0.75 and 0.06	87.8	–	0.75 and 0.06
GW-LR	73.3	10.5	–	87.2	417.7	–
GW-ENLR	92.0	15.0	0.75 and 0.02	82.4	1119.0	0.75 and 0.02

each model. The optimal GW-LR and GW-ENLR bandwidths are also given, as are the penalty parameters specified for ENLR and for GW-ENLR. For the ENLR models, the penalty parameters were chosen optimally according to the CV functions found in the *glmnet* R package, v 2.0-10 (Friedman et al. 2010). For the GW-ENLR models, the penalty parameters were chosen according to a set of preliminary analysis for each case study, with a judged re-calibrating of the GW-ENLR model with different penalty parameter pairs and assessing outputs.

Thus, for both the election and species case studies the ENLR and GW-ENLR models were run with the same α and λ , respectively. The ENLR models were run with $\alpha=0.75$ and $\lambda=0.06$, indicating a tendency to a lasso rather than a ridge regression, but where the penalty is relatively weak. The GW-ENLR models were run with $\alpha=0.75$ and $\lambda=0.02$, again indicating a tendency to a lasso rather than a ridge regression, but where the penalty is weaker than set globally.

Some trends are evident in Table 1:

- For both case studies, LR is always outperformed by the other three models;
- For both case studies, ENLR outperforms both LR and GW-LR;
- ENLR strongly outperforms GW-LR in the election case study, but only marginally in the species presence–absence case study;
- GW-ENLR clearly outperforms the other three models in the election case study only;
- GW-ENLR only outperforms LR in the species presence–absence case study.

As with any set of regression comparisons, it is paramount that the intercept and predictor coefficient estimates are interrogated, as summarized in Table 2 and mapped in Fig. 6 (for three of the predictors, from each case study, but for both GW-based models). The stationary coefficient LR model returns a single, global estimate for each predictor, as does the ENLR model, but where predictors are dropped when their coefficients shrink to zero. The non-stationary coefficient GW-LR and GW-ENLR models each return a series of local coefficient estimates, the interquartile ranges of which give a broad indication of their spatial variation. Values of ‘zero’ in Table 3 indicate that the corresponding predictor was dropped from the global or local regression.

Table 2 The coefficient estimates from the study regression models, where the first, second and third quartiles (Q) of the coefficient distribution for the geographically weighted logistic regression (GW-LR) are given, and where the first, second and third quartiles (Q) of the coefficient distribution for geographically weighted elastic net logistic regression (GW-ENLR) are given together with the minimums and maximums

Case study	Coefficient estimate	LR	ENLR	GW-LR 1st Q	GW-LR Median	GW-LR 3rd Q	GW-ENLR Min	GW-ENLR 1st Q	GW-ENLR median	GW-ENLR 3rd Q	GW-ENLR Max
Election	Intercept	-1.690	0.185	-10.389	-3.766	0.478	0.555	1.493	2.002	2.523	4.394
	% Employed	1.583	Zero	-2.935	1.564	7.005	-2.695	Zero	Zero	Zero	3.023
	% College education	-10.686	-3.464	-14.172	-9.477	-5.287	-22.049	-9.544	-6.761	-3.372	Zero
	% Over 65	1.392	Zero	-1.341	2.722	7.676	-10.984	Zero	2.060	4.729	11.134
	Population Density	-1.994	-0.111	-11.28	-6.482	-2.085	-11.730	-4.443	-0.593	Zero	0.701
	% White	7.501	2.983	9.579	14.263	20.564	-1.641	0.455	2.546	5.894	12.324
Species	Intercept	-8.454	-3.855	-33.906	-11.954	-0.453	-5.466	-1.951	-1.340	-0.989	-0.277
	Degree-days above 5 °C	-11.160	1.715	-118.668	-40.911	4.834	-5.454	Zero	0.239	4.540	11.934
	Annual amount of precipitation	-1.026	Zero	-45.613	-7.262	13.61	-36.173	-10.595	-5.048	-0.091	12.657
	Potential evapotranspiration	3.368	2.239	-10.111	3.67	25.192	-6.589	Zero	Zero	2.163	10.771
	Annual variation of precipitation	-6.341	Zero	-37.138	-7.182	16.992	-41.605	-17.212	-2.155	Zero	47.039
	Annual mean temperature	21.102	2.036	-10.024	28.201	119.371	Zero	Zero	Zero	0.111	3.571

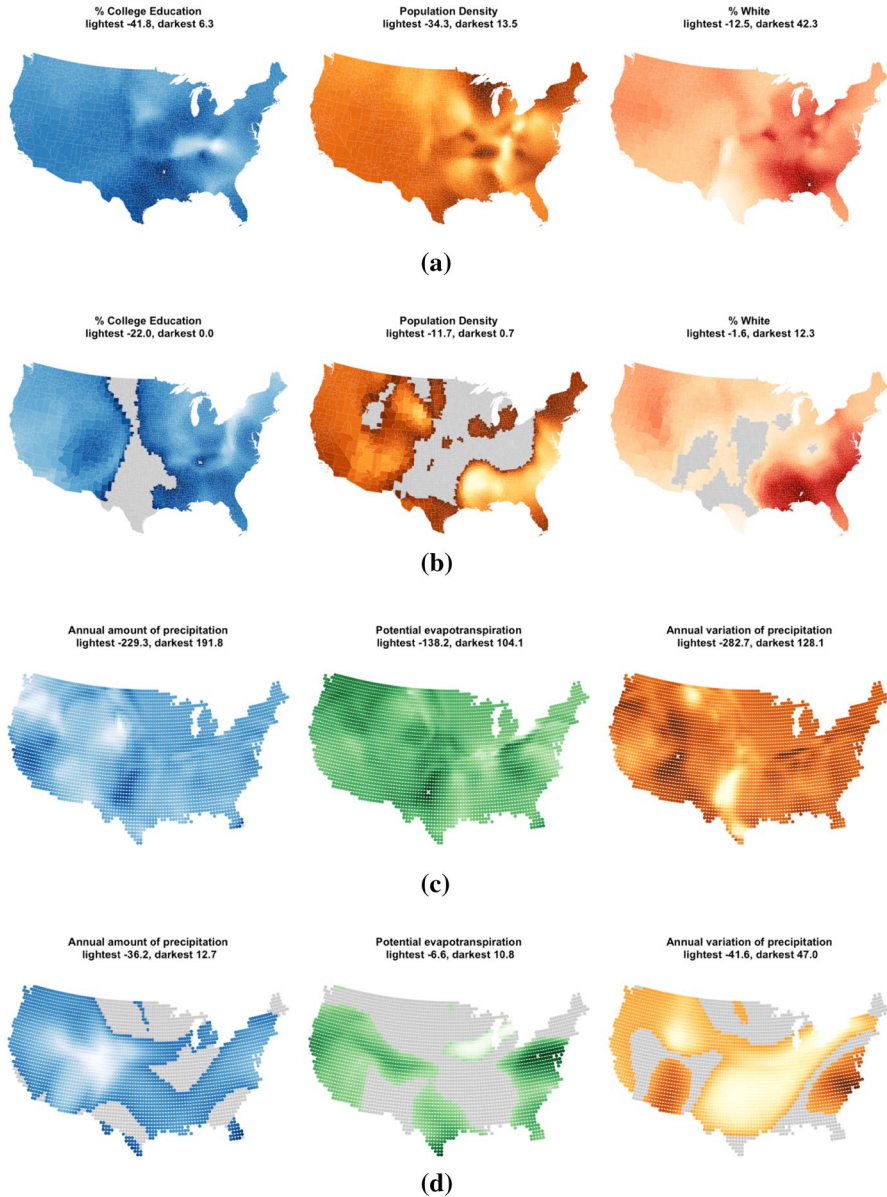


Fig. 6 The spatial distributions of the local coefficient estimates: **a** GW-LR for % *College Education*, *Population Density* and % *White* for the election case study; **b** GW-ENLR for % *College Education*, *Population Density* and % *White* for the election case study; **c** GW-LR for *Degree-days above 5 °C*, *Potential evapotranspiration* and *Annual variation in precipitation* for the species case study; and **d** GW-ENLR for *Degree-days above 5 °C*, *Potential evapotranspiration* and *Annual variation in precipitation* for the species case study. Areas shaded in grey indicate where GW-ENLR did not select the predictor in the local model

Table 3 A summary of the 100 simulations (* indicates over 159 locations)

		Min.	1st Q.	Median	3rd Q.	Max.
α	Global CN	4.5	11.4	14.6	19.7	48.0
0.50	GW-ENLR bandwidth (%)	11.9	15.7	18.2	22.0	36.5
	Counts of local CN > 30*	20	26	29	32	37
	Local collinearity addressed (%)	48.0	62.5	67.7	74.6	85.7
0.75	GW-ENLR bandwidth (%)	11.9	15.7	17.6	20.8	30.2
	Counts of local CN > 30*	23	26	30	33	38
	Local collinearity addressed (%)	62.5	78.8	83.3	87.9	97.1
1.00	GW-ENLR bandwidth (%)	11.9	15.7	18.2	20.8	34.6
	Counts of local CN > 30*	23	26	30	33	38
	Local collinearity addressed (%)	85.2	94.5	96.8	100	100

For the election case study, it is observed:

- The ENLR model drops two predictors in *% Employed* and *% Over 65*, where this is not unexpected given these predictors show the strongest correlations with another predictor (i.e. *% College Education* and *Population Density*, respectively, Fig. 4).
- The coefficients for *% Employed* and *% Over 65*, also change from positive in the LR model to sometimes negative in the other three models, endorsing these coefficients to be those most adversely influenced by collinearity.
- In the GW-ENLR model, all five predictors were selected at some point. Thus, dropping *% Employed* and *% Over 65* globally (in ENLR) appears a poor decision, as these predictors can still provide useful information, locally.
- In the GW-ENLR model, *% Employed*, *% Over 65*, *Population Density* and *% College Education* are most likely to be dropped from the local regression, and in that order.
- The coefficients for *Population Density* are strongly negative in GW-LR, but this weakens in the GW-ENLR fit where more positive coefficients result.
- The GW-LR and GW-ENLR models both suggest non-stationary relationships throughout.

For the species presence–absence case study, it is observed:

- The ENLR model drops two predictors in *Annual amount of precipitation* and *Annual variation of precipitation*, but not *Annual mean temperature* or *Degree-days above 5 °C* as might be expected given their strong correlations of Fig. 4.
- However, the coefficients for *Degree-days above 5 °C* change from strongly negative in the LR model, to positive in ENLR, to negative/positive in GW-LR, and finally all positive in GW-ENLR, suggesting this coefficient to be that most adversely influenced by collinearity. Other coefficients change sign across the four models, but not to the extent that *Degree-days above 5 °C* changes.

- In the GW-ENLR model, all five predictors were selected at some point. Thus, dropping *Annual amount of precipitation* and *Annual variation of precipitation* globally (in ENLR) appears a poor decision, as these predictors can still provide useful information, locally.
- Interesting, although *Annual amount of precipitation* and *Annual variation of precipitation* are dropped in ENLR, they are not so commonly dropped in GW-ENLR. Here *Annual mean temperature* and *Potential evapotranspiration* are more likely to be dropped locally. Furthermore, *Annual amount of precipitation* is never dropped locally.
- The GW-LR and GW-ENLR models both suggest non-stationary relationships throughout.

The maps in Fig. 6 support a deeper understanding of the data structure and how best to develop a potential predictive model in two ways. First, the GW-ENLR coefficient maps indicate where local collinearity amongst predictor variables may exist and amongst which predictors. Second, by comparing the GW-LR coefficient maps with those from GW-ENLR, the perceptions of relationship non-stationarity are either confirmed (as local collinearity is not an issue) or thrown into doubt (as local collinearity is an issue). All coefficient maps (both from GW-LR and GW-ENLR) provide information about the spatial structure and spatial interaction of the predictor variables that may provide insight into the nature of the processes being modelled.

For example, the election case study GW-ENLR model (Fig. 6b) suggests that % *College Education* is generally negatively associated with Trump support, but with a swath running North to South where it is increasingly less of a factor, shrinking to zero; whilst for the corresponding GW-LR output (Fig. 6a), some regions within the swath are likely erroneously indicating a positive association. Again, for the election case study, GW-ENLR (Fig. 6b) indicates that *Population density* is mostly negatively associated with Trump support in the South and East indicating more support in rural areas and is more positively associated with Trump support in Texas and Illinois and Missouri and also that % *White* is generally positively associated with Trump support and especially so in the South and East. The described non-stationarities are not so clearly defined in the corresponding GW-LR outputs (Fig. 6a) highlighting likely local instabilities in coefficient estimation due to collinearity.

For the species case study, the spatial patterns in the GW-LR (Fig. 6c) and GW-ENLR (Fig. 6d) coefficients are often quite different, suggesting clear value in the investigation of both model forms. Coefficient maps for the GW-ENLR model (Fig. 6d) suggest that *Annual variation of precipitation* is an important factor negatively associated with species absence to the in the Frontier region of the country, with the most positive associations around in Florida and southern Texas, whilst *Potential evapotranspiration* is a highly localized driver of species absence with negative associations around Lake Michigan and positive associations in Texas, in the West and around the North Eastern metropolitan areas. The selection of *Annual variation of precipitation* in the species case study GW-ENLR model is highly localized, with high negative associations in a swath from Texas to New England and around Montana and positive associations around North Carolina and Arizona.

5 Simulation experiment

To complete the demonstration of the new GW-ENLR model, a simulation experiment was undertaken in order to objectively and robustly determine how it behaved in the context of known local collinearity. Here the aim was to develop a simple exploration of collinearity at the local level with simulated data sets.

In a typical evaluation of spatial varying coefficient (SVC) models, simulated coefficients are generated together with associated predictor and response variables (e.g. Wheeler 2009; Fotheringham and Oshan 2016). As the generated SVCs are known through simulation, they provide a means to objectively assess the accuracy of those estimated by the study model (in this case GW-ENLR). It is also possible to objectively evaluate the prediction accuracy of a SVC model in this manner (e.g. Harris et al. 2010). However, evaluating GW-ENLR through simulation is difficult, as it effectively undertakes a local model selection procedure where coefficients are shrunk to zero making comparisons of the GW-ENLR estimated and the (actual) simulated coefficients awkward. It is still possible to objectively assess the GW-ENLR predictive power through the simulation experiment (i.e. accuracy of the response), but as collinearity primarily affects coefficient estimates (and their interpretation), investigating measures of predictive strength are not entirely relevant. For these reasons, the approach taken here was to examine the performance of GW-ENLR, where the aim was to determine the proportion of occurrences of known local collinearity (through simulation) that were correctly dealt with (through coefficient shrinkage) by the GW-ENLR fit.

To generate the simulated data sets, the same procedure as in Harris et al. (2017) was adopted. Briefly, the experiment generates SVCs $\beta_0, \beta_1, \beta_2, \beta_3$ and then independently, the predictor data, x_1, x_2, x_3 . The coefficient and predictor realizations are then directly used to generate the response variable and the random error data ε_i (given a pre-specified trend to error process ratio). Both the regression coefficients and the predictor variables are (independently) generated using an un-conditional sequential Gaussian co-simulation (Wackernagel 2003). For the predictor variables, the co-simulation parameters are chosen so that relatively strong levels of collinearity are generated between x_2 and x_3 , only. This typically results in correlation coefficients of around $r=0.9$ for this predictor pair. Details are provided in Harris et al. (2017), where for this study, the Gaussian response variable y_i requires conversion to a binomial response variable y_{i1} , following Eq. (1).

In total, 100 simulated data sets are generated and the GW-ENLR model is fitted to each one and its performance assessed. Simulations are generated to the ($n=159$) centroids of the ‘counties of Georgia for the United States’; an educational attainment data set routinely used to demonstrate GWR (Fotheringham et al. 2002) and included in the *spgwr* and *GWmodel* R packages (Bivand et al. 2017; Gollini et al. 2015). The geostatistical-based experiment provides useful stochasticity, enabling nuanced differences to each simulated data set, generated from the same initial specifications.

The result for each of the 100 simulated data sets was a binomial response variable, three predictor variables and four sets of location-specific coefficients, such

that a given simulation typically exhibited moderate coefficient non-stationarity coupled with predictor variable collinearity that was low globally, but moderately strong locally. Collinearity (both globally and locally) was measured via the CN, where a $CN > 30$ is of concern.

For each simulated data set, an optimum GW-ENLR (adaptive) bandwidth was determined first; secondly, having determined the bandwidth, the GW-ENLR was applied to determine the set of local coefficient estimates at each location. The GW-ENLR was specified with three options of $\alpha = 0.50, 0.75$ and 1 , whilst $\lambda = 0.02$, remained as is, each time. For all 100 simulations, the percentage of cases where local collinearity (as measured by local CN) was found that resulted in predictor variable shrinkage under the GW-ENLR was determined for each location.

The results of the simulation are reported in Table 3. For $\alpha = 0.75$, the median number of times that local collinearity was observed in the 100 simulations (through local $CN > 30$) at the 159 locations was 30, and the median GW-ENLR bandwidth was 17.6% of the nearest data points (indicating moderate to strong coefficient heterogeneity, as broadly expected). Instances of local collinearity at the 159 locations were found and correctly dealt with on average 67.7, 83.3 and 96.8% of the time through the GW-ENLR fit (as indicated by a coefficient shrinkage to zero) for $\alpha = 0.50, 0.75$ and 1.00 , respectively. Thus, the new GW-ENLR model performs as expected.

Taking the case of $\alpha = 0.75$, the spatial structure of the simulated local collinearity together with the performance of GW-ENLR is shown in Figs. 7 and 8. Figure 7 shows the rate of coefficient shrinkage at each location, first overall, and then for the coefficients associated with each of the three predictor variables, x_1, x_2, x_3 . Clearly, coefficient shrinkage is performing as it should do since it primarily involves the coefficients corresponding to predictor variables x_2 or x_3 , reflecting the high correlation specified between this variable pair only. Sometimes, but to a much lesser extent, the coefficient for x_1 is shrunk also. This reflects the existence of complex controlling influences that each of the three predictor variables have on each other (including the intercept), regardless of 'known' collinearities. Figure 8 shows the frequency of local collinearity observed at each location through the simulated data sets (i.e. local $CN > 30$) and the proportion of times of when it is observed that it is appropriately dealt with by GW-ENLR coefficient shrinkage. Clearly GW-ENLR again performs as it should do in this more general sense.

6 Discussion

The objective of this paper was to develop and apply a localized approach to elastic net logistic regression. Elastic nets and other penalized regression approaches have been proposed as suitable methods for dealing with collinearity and model selection in regression models (Zou and Hastie, 2005; Friedman et al. 2010). The geographically weighted (GW) elastic net logistic regression described in this study has been proposed in work describing GW lasso approaches (Yoneoka et al. 2016), and although an R package has been developed for GW-ENLR (Yoneoka and Saito, 2015), the functions have never worked over the last 3 years. This paper addresses

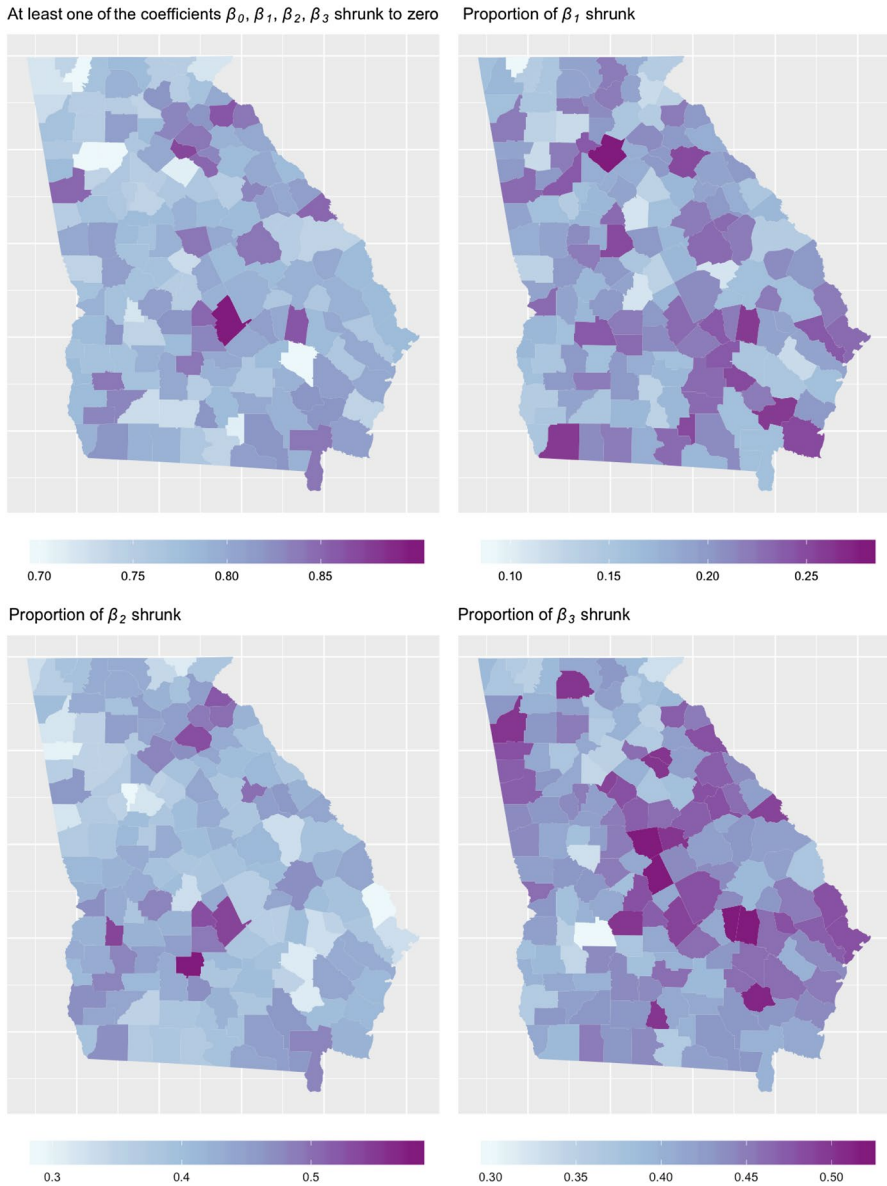


Fig. 7 The proportions of coefficients shrunk by GW-ENLR at each location in the simulated data sets

this gap and demonstrates the application of a framework for GW-ENLR, including tools for optimal adaptive and fixed bandwidth selection, but where the concurrent optimization of the elastic net penalty parameters α and λ is left open for future work, especially with respect to a *local* optimization in the spirit of the locally compensated ridge GWR model of Brunson et al. (2012). Transference of the *global*

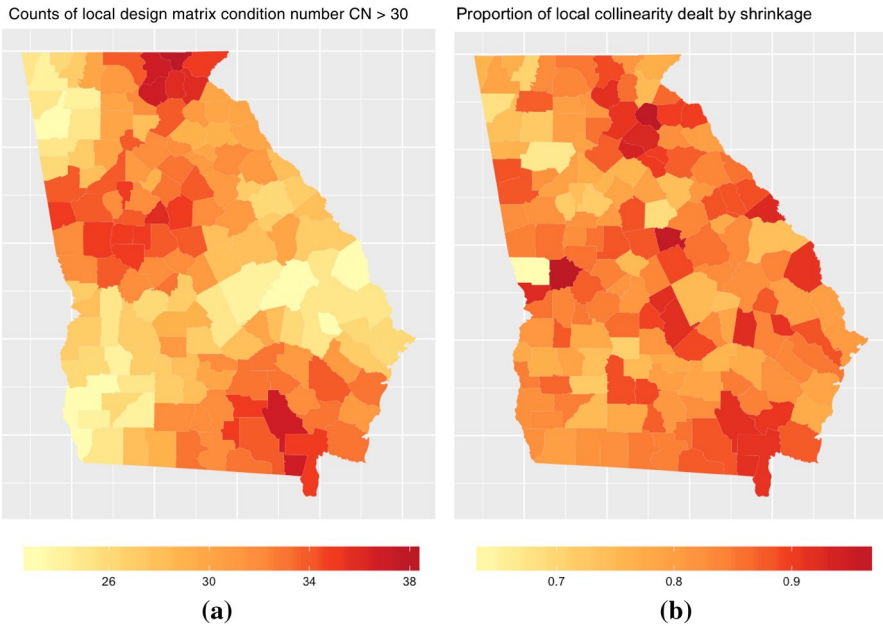


Fig. 8 **a** The frequency of local collinearity observed in each location through the simulated data sets and **b** the proportion of times when it is observed in **(a)** that it is appropriately dealt with by GW-ENLR coefficient shrinkage

optimization procedure for the bandwidth and the penalty parameters α and λ from the Gaussian-response GW elastic net model of Li and Nam (2018) is also possible.

A series of regression models were developed using logistic regression, ENLR, GW logistic regression (GW-LR) and the new GW-ENLR model and applied to two case studies: (1) understanding the socio-economic factors associated with Trump support in the 2016 presidential elections in the USA Trump in the 2016 presidential elections and (2) modelling species presence–absence from environmental data in the USA. For both case studies, the a-spatial logistic regression models were refined by the a-spatial ENLR model and the spatial GW-LR model, which both showed improved prediction rates over logistic regression. Further refinement was demonstrated when ENLR and GW-LR was combined to form GW-ENLR, which further improved performance in the election case study.

The proposed GW-ENLR method adds a new tool for improving predictive performance and refining model selection in local regression. Examining the variables that are locally selected by the GW-ENLR models can provide a deeper geographical understanding of the processes and factors associated with the phenomenon under investigation (e.g. voting behaviours or species persistence) and how they interact. For example, in the election case study, local models of support for Trump consistently included the percentage of white people recorded in the 2015 update of the population census, with coefficient estimates varying spatially, whilst population density and the percentage of people with college educations were selected only in specific locations. Conversely, in the species presence–absence case study, none

of the predictor variables were consistently selected. Instead, there is an East–West split in the factors associated with species presence–absence, driven by potential evapotranspiration in the West and precipitation in the East. Identifying these geographical trends provides the basis for further exploration of the data or the phenomenon under investigation by domain experts. Such spatial detective work is at the core of spatial analysis and GI Science: it is not necessarily the role of those working in spatial analysis, GeoComputation and GI Science to have detailed domain knowledge, but it is important that the discipline continues to develop robust methods for spatial explorations of data and processes.

7 Closing remarks

In conclusion, this paper develops a method for applying ENLR, locally. The rationale for this is simple. First, techniques such as elastic net have been shown to be efficient at model selection and information reduction and can overcome predictor variable collinearity. Second, global, whole map models, such as ENLR, may not be sensitive to local trends in the data and may even obfuscate them. This is because the assumption of spatial stationarity in process and relationships inherent in many statistical models may be inappropriate (Fotheringham and Brunson 1999). Third, a common criticism of GWR is that it is likely to be subject to local collinearity even when none is observed globally (Wheeler and Tiefelsdorf 2005), as the data under the kernel for each local model may include predictors that are highly collinear. This study's GW-ENLR approach combines the local (non-stationary) advantages conferred by GW approaches with the model selection (collinearity) advantages of the elastic net. The results of applying this approach to two case studies—one socio-economic and one environmental—indicate that the resultant maps of coefficients from GW-ENLR can be considered more reliable than that found with GW-LR, and thus, perceptions of relationship non-stationary should also be more assured with GW-ENLR. A short simulation experiment also endorsed this value of the new GW-ENLR model.

Acknowledgements This research was supported by the Natural Environment Research Council Newton Fund Grant (NE/N007433/1) and UK Biotechnology and Biological Sciences Research Council Grants (BBS/E/C/000J0100, BBS/E/C/000I03320 and BBS/E/C/000I0330). The authors thank the anonymous reviewers whose suggestions resulted in a much improved paper. All of the analyses and mapping were undertaken in R 3.4.3 the open source statistical software. The GW-LR analyses used the GWmodel package, v2.0-5 (Gollini et al. 2015), and the ENLR analyses were undertaken using the glmnet package, v 2.0-13 (Friedman et al. 2010). The GW-ENLR code and data used in this paper can be downloaded from <https://github.com/lexcomber/GW-ENLR>.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Anselin L (1990) Spatial dependence and spatial structure instability in applied regression analysis. *J Reg Sci* 30(2):185–207
- Atkinson PM, German SE, Sear DA, Clark MJ (2003) Exploring the relations between river bank erosion and geomorphological controls using geographically weighted logistic regression. *Geogr Anal* 35(1):58–82
- Bárcena MJ, Menéndez P, Palacios MB, Tusell F (2014) Alleviating the effect of collinearity in geographically weighted regression. *J Geogr Syst* 16(4):441–466
- Becker RA, Wilks AR, Brownrigg R, Minka TP, Deckmyn A (2016) maps: Draw geographical maps. R package version 3.1.1. <https://CRAN.R-project.org/package=maps>. Accessed 8 June 2017
- Belsley DA, Kuh E, Welsch RE (1980) Regression diagnostics: identifying influential data and sources of collinearity. Wiley, New York
- Bivand R, Yu D, Nakaya T, Garcia-Lopez MA (2017) spgwr: Geographically weighted regression. R Package Version 0.6–32. <https://cran.r-project.org/web/packages/spgwr/spgwr.pdf>. Accessed on 20 March 2018
- Broennimann O, Di Cola V, Guisan A (2016) ecospat: Spatial ecology miscellaneous methods. R package version 2.1.1. <https://CRAN.R-project.org/package=ecospat>. Accessed 8 June 2017
- Brunsdon C, Fotheringham AS, Charlton M (1996) Geographically weighted regression: a method for exploring spatial nonstationarity. *Geogr Anal* 28(4):281–298
- Brunsdon C, Fotheringham AS, Charlton ME (1998) Spatial nonstationarity and autoregressive models. *Environ Plan A* 30(6):957–993
- Brunsdon C, Fotheringham AS, Charlton M (2002) Geographically weighted summary statistics—a framework for localised exploratory data analysis. *Comput Environ Urban Syst* 26(6):501–524
- Brunsdon C, Charlton M, Harris P (2012) Living with collinearity in local regression models. In: Proceedings of the 10th international symposium on spatial accuracy assessment in natural resources and environmental sciences, Brazil
- Comber A, Brunsdon CF, Charlton M, Harris P (2017) Geographically weighted correspondence matrices for local change analyses and error reporting: mapping the spatial distribution of errors and change. *Remote Sens Lett* 8(3):234–243
- Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, Marquéz JRG, Gruber B, Lafourcade B, Leitão PJ, Münkemüller T, McClean C, Osborne PE, Reineking B, Schröder B, Skidmore AK, Zurell D, Lautenbach S (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36(1):27–46
- Fotheringham AS, Brunsdon C (1999) Local forms of spatial analysis. *Geogr Anal* 31(4):340–358
- Fotheringham AS, Oshan TM (2016) Geographically weighted regression and multicollinearity: dispelling the myth. *J Geogr Syst* 18(4):303–329
- Fotheringham AS, Brunsdon C, Charlton M (2002) Geographically weighted regression: the analysis of spatially varying relationships. Wiley, New York
- Frank IE, Friedman JH (1993) A statistical view of some chemometrics regression tools. *Technometrics* 35(2):109–135
- Friedman J, Hastie T, Tibshirani R (2010) Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 33(1):1–22
- Gollini I, Lu B, Charlton M, Brunsdon C, Harris P (2015) GWmodel: an R package for exploring spatial heterogeneity using geographically weighted models. *J Stat Softw* 63(17):1–50
- Harris P, Fotheringham AS, Crespo R, Charlton M (2010) The use of geographically weighted regression for spatial prediction: an evaluation of models using simulated data sets. *Math Geosci* 42(6):657–680
- Harris P, Brunsdon C, Charlton M (2011) Geographically weighted principal components analysis. *Int J Geogr Inf Sci* 25(10):1717–1736
- Harris P, Brunsdon C, Lu B, Nakaya T, Charlton M (2017) Introducing bootstrap methods to investigate coefficient non-stationarity in spatial regression models. *Spat Stat* 21(A):241–261
- Hastie T, Qian J (2014) Glmnet vignette. http://web.stanford.edu/~hastie/glmnet/glmnet_alpha.html. Accessed 8 June 2017
- Hoerl AE (1962) Application of ridge analysis to regression problems. *Chem Eng Prog* 58(3):54–59
- Hoerl AE, Kennard RW (1970) Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* 12(1):55–67

- Li K, Nam NSN (2018) Geographically weighted elastic net: a variable-selection and modeling method under the spatially nonstationary condition. *Ann Am Assoc Geogr*. <https://doi.org/10.1080/24694452.2018.1425129>
- McGovern T (2017) US president county-level election results for 2012 and 2016. https://github.com/tonmcg/County_Level_Election_Results_12-16 Accessed 8 June 2017
- Meloun M, Militký J, Hill M, Brereton RG (2002) Crucial problems in regression modelling and their solutions. *Analyst* 127(4):433–450
- Rodrigues M, de la Riva J, Fotheringham AS (2014) Modeling the spatial variation of the explanatory factors of human-caused wildfires in Spain using geographically weighted logistic regression. *Appl Geogr* 48:52–63
- Tibshirani R (1996) Regression shrinkage and selection via the lasso. *J R Stat Soc B* 58:267–288
- US Census (2017) QuickFacts. <https://www.census.gov/quickfacts/>. Accessed 8 June 2017
- Wackernagel H (2003) *Multivariate geostatistics: an introduction with applications*. Springer, Berlin
- Wang W, Li D (2017) Structure identification and variable selection in geographically weighted regression models. *J Stat Comput Simul* 87(10):2050–2068
- Wheeler D (2007) Diagnostic tools and a remedial method for collinearity in geographically weighted regression. *Environ Plan A* 39(10):2464–2481
- Wheeler D (2009) Simultaneous coefficient penalization and model selection in geographically weighted regression: the geographically weighted lasso. *Environ Plan A* 41(3):722–742
- Wheeler D (2013) Geographically weighted regression. In: Fischer M, Nijkamp P (eds) *Handbook of regional science*. Springer, Berlin, pp 1435–1460
- Wheeler D, Tiefelsdorf M (2005) Multicollinearity and correlation among regression coefficients in geographically weighted regression. *J Geogr Syst* 7(2):161–187
- Windle MJS, Rose GA, Devillers R, Fortin M-J (2010) Exploring spatial non-stationarity of fisheries survey data using geographically weighted regression (GWR): an example from the Northwest Atlantic. *ICES J Mar Sci* 67(1):145–154
- Yoneoka D, Saito E (2015) GWLelast: geographically weighted logistic elastic net regression. <https://CRAN.R-project.org/package=GWLelast>. Accessed 8 June 2017
- Yoneoka D, Saito E, Nakaoka S (2016) New algorithm for constructing area-based index with geographical heterogeneities and variable selection: an application to gastric cancer screening. *Sci Rep* 6:26582
- Zou H, Hastie T (2005) Regularization and variable selection via the elastic net. *J R Stat Soc B* 67(2):301–320