

A Spatially Autocorrelated Weights of Evidence Model

Minfeng Deng^{1,2}

Received 17 September 2009; accepted 19 October 2009
Published online: 7 November 2009

One of the most important features of spatial datasets is that they often exhibit spatial autocorrelation, where locational similarities are observed jointly with similarities in values. Both logistic regression (LR) modelling and weights of evidence (WE) modelling are methods commonly applied in binary pattern recognition. While a spatially autocorrelated variant of the LR model, the so-called autologistic regression (ALR) model, exists in the literature, a spatially autocorrelated variant of the WE model does not exist. In this paper, a spatially autocorrelated weights of evidence (SACWE) model will be proposed. It will be demonstrated that the new model contains the same amount of spatial information as does an ALR model, and it is easy to program and implement. Via a simulation study, it will be shown that, in the presence of spatial autocorrelation, both in terms of in-sample fit and out-of-sample predictions the SACWE model is on par with the ALR model, while significantly outperforming the conventional WE model.

KEY WORDS: Logistic regression, simulation, autologistic regression, spatial prediction.

INTRODUCTION

Weights of evidence (WE) modelling is a Bayesian probability method used in explaining and predicting occurrences of binary events. It relates the presence of a binary response variable to a number of binary maps of geological features, which are used as predictor patterns, and it produces a map of estimated posterior probabilities as the end product. Detailed descriptions of the method can be found in Bonham-Carter, Agterberg, and Wright (1988, 1989), Agterberg, Bonham-Carter, and Wright (1990), and Bonham-Carter (1994). Although in the majority of its empirical applications WE modelling has been used to produce maps of mineral deposit potentials, increasingly it has been adopted in fields beyond mineral exploration.

For instance, Hansen (2000) and Hansen and others (2002) used WE to analyze archeological site distributions, Romero-Calcerrada and Luque (2006) predicted the habitat suitability of *Picoides tridactylus* (three-toed woodpecker), Mathew, Jha, and Rawat (2007) and Dahal and others (2008) studied landslide susceptibility, Emelyanova and others (2008) investigated cattle farm distribution in Australia, amongst others.

Upon close inspection, however, one notices that the binary response variable of interest in these recent studies are likely to be spatially autocorrelated. Human beings are social animals, and human settlements, such as farms, as well as archeological finds, are likely to exhibit strong clustering behavior. Likewise, the presence of animal species in a particular region is likely to be influenced by the presence of the same species in nearby regions. The occurrence of landslides in neighboring areas is also likely to lead to structural instabilities and increase the probability of landslides in the area being surrounded. While significant spatial autocorrelation is likely to be present, it has not been considered in these recent studies.

¹Department of Econometrics and Business Statistics, Monash University, Menzies Building, Clayton Campus, Clayton VIC 3800, Australia.

²To whom correspondence should be addressed; e-mail: minfeng.deng@buseco.monash.edu.au

It has long been recognized that stochastic processes in close spatial proximity often exhibit spatial autocorrelation, where locational similarities are observed in conjunction with similarities in values. The first law of geography, as first described in Tobler (1970), states clearly that “everything is related to everything else, but near things are more related than distant things.” Since then, important works of Cliff and Ord (1973), Ord (1975), Anselin (1980, 1988), Cressie (1993), just to name a few, have made significant contributions in incorporating spatial autocorrelation in spatial modelling. However, spatial autocorrelation has never been formally incorporated into WE modelling. This is perhaps due mainly to the fact that WE modelling was initially developed for mineral potential mapping, where the spatial process under consideration is inanimate and not expected to exhibit significant spatial dependencies. But as WE modelling is increasingly being implemented in fields beyond mineral exploration, the need to incorporate spatial autocorrelation should be acknowledged. It should also be noted that, even if the spatial process of interest is on its own not spatially autocorrelated, it is possible that some of the underlying predictor patterns may be spatially autocorrelated. And if one or more of these spatially autocorrelated predictor patterns are missing from the modelling process, the residuals of the model will exhibit spatial patterns, a situation that has received a lot of attention in the spatial econometrics literature and gave rise to the so-called spatial error model; for more details see Anselin (1988).

Logistic regression (LR) is another loglinear binary events recognition method prominent in the literature; detailed descriptions of its applications in mineral research can be found in Agterberg (1992), Agterberg and others (1993), and Agterberg and Bonham-Carter (1999). Importantly, Besag (1972, 1974, 1975) developed a class of autologistic regression (ALR) models, where response variable values of the spatial neighbors are incorporated into the logistic model, thus explicitly accounting for spatial autocorrelation. Subsequently, Haining (1985) used ALR to investigate spatial price competition, Augustin, Muggleston, and Buckland (1996, 1998) studied spatial distribution of wild life, Wu and Huffer (1997) and Huffer and Wu (1998) studied spatial distribution of plant species, Wintle and Bardos (2006) investigated species-habitat relationships, amongst others. Both simulation studies and empirical studies have shown that, in the presence of significant spatial autocorrelation, the

ALR model outperforms the LR model in terms of fit and predictive ability.

It is well understood that the WE model and the LR model share many similarities, and that in some special cases they are equivalent (see Deng, 2009). Given the importance of spatial autocorrelation in spatial processes and the close links between the WE model and the LR model, development of a WE variant of the ALR model would be a useful addition to the spatial literature. In this paper, I will propose a spatially autocorrelated weights of evidence (SACWE) model, where values of the spatial neighbors are incorporated as additional predictor patterns in the model. It will be demonstrated that the SACWE model contains the same amount of information as the ALR model, and it is easy to program and implement. Via a simulation study, it will be shown that in the presence of spatial autocorrelation the SACWE model significantly outperforms the WE model both in terms of in-sample fit and out-of-sample predictions, and is on par with the ALR model.

THE AUTOLOGISTIC REGRESSION MODEL

Before developing the SACWE model, it is instructive to first look at the ALR model in detail. For spatial observation i , let the response variable be defined by a binary random variable y_i , and $y_i = 1$ when the event is present and $y_i = 0$ when absent. For simplicity and without loss of generality, let there be one predictor pattern, which is defined by a binary random variable $x_{1,i}$, and $x_{1,i} = 1$ when the predictor pattern is present and $x_{1,i} = 0$ when absent. Furthermore, let p_i be the probability of $y_i = 1$. Following the notation of Augustin, Muggleston, and Buckland (1996), the ALR model is specified as:

$$\log\left(\frac{p_i}{1-p_i}\right) = \alpha + \beta_1 x_{1,i} + \beta_{\text{auto}} \text{autocov}_i \quad (1)$$

where α is the usual intercept term, β_1 is the regression coefficient associated with the predictor pattern $x_{1,i}$, and β_{auto} is the regression coefficient associated with the so-called spatial autocovariate, autocov_i , which is calculated as:

$$\text{autocov}_i = \frac{\sum_{j=1}^{k_i} w_{ij} y_j}{\sum_{j=1}^{k_i} w_{ij}} \quad (2)$$

where k_i is the total number of spatial neighbors for i , and w_{ij} is the spatial weight given to its j th

neighbor. It can be seen from Eq. 2 that the spatial autocovariate autocov_i is a weighted average of the spatial neighbors of observation i . The autocovariate autocov_i differs from the normal predictor pattern $x_{1,i}$ in that it is a function of the response variable, and its inclusion explicitly accounts for spatial autocorrelation. It is clear that autocov_i has a lower bound of 0, as in the case of all neighbors of observation i being 0, and an upper bound of 1, as in the case of all neighbors being 1. When β_{auto} is positive, the probability of $y_i = 1$ is positively influenced by the average value of the spatial neighbors of i , and the more neighbors taking on a value of 1 the higher the probability of y_i being equal to 1.

There are a large number of ways of specifying the spatial structure and hence the spatial weights w_{ij} , and a detailed discussion of the topic is beyond the scope of this paper. The most simplistic specification, which is defined by a binary contiguity matrix, can be found in Ord (1975) and Anselin (1988). Essentially, a binary contiguity matrix treats spatial relationships as a binary relationship: one is either a spatial neighbor ($w_{ij} = 1$) or not a spatial neighbor ($w_{ij} = 0$). To demonstrate, consider the regular lattice in Figure 1, where i is the spatial observation of interest. One of the most commonly used binary contiguity matrix, known as the Queen contiguity matrix (named after its resemblance to the movements of the Queen in a chess game), defines all shaded cells in Figure 1 as spatial neighbors of observation i . In this case, $k_i = 8$ and $w_{ij} = 1$ for all of the eight neighbors. From this point onwards, for ease of discussion and without loss of generality, all spatial weights matrices used will be of the Queen contiguity design.

It should be noted that, despite its intuitive appeal and its apparent functional simplicity,

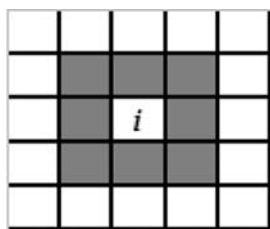


Figure 1. A regular lattice, in which observation i is of interest, and the shaded cells are considered spatial neighbors of i .

estimation of the ALR model as given in Eq. 1 is not straightforward and it still remains an active area of research. In the traditional setting for a logistic regression (LR) model, the response variables are assumed to be independent, and the full likelihood is simply the product of the likelihoods of individual spatial observations. In the ALR model, however, the presence of autocov_i on the right-hand side of Eq. 1 means that the response variables are no longer independent. As shown in Huffer and Wu (1998), the full likelihood for the ALR model is known only to within a normalization constant, which is a function of the regression parameters and is thus intractable except in trivial cases, and the standard maximum likelihood (ML) results do not apply.

A number of estimators have been proposed in the literature, and only the most important ones are outlined below, as a comprehensive literature review is beyond the scope of this paper. Besag (1972, 1974) suggested a simple coding method, in which the spatial sample is divided into spatially independent subsets, for which separate ML estimates are obtained and combined at the end. But the coding method was found to be inefficient and sensitive to choice of coding schemes. Besag (1975) suggested a maximum pseudo-likelihood (MPL) estimator. Essentially, MPL assumes that the spatial units are independent and treats autocov_i as another covariate. Comets (1992) showed that the MPL estimates are consistent and asymptotically normal, although the MPL estimates do not have a valid variance measure. As the MPL estimator is intuitive and can be computed using conventional logistic regression techniques, it is the most widely used estimator in practice. Finally, Wu and Huffer (1997) and Huffer and Wu (1998) developed a Markov Chain Monte Carlo (MCMC) method for approximating the distribution of the ML estimators for spatially autocorrelated binary choice models. Their method is computationally far more intensive, and its convergence depends on the choice of the trial state being sufficiently close to the true MLE value. But they showed that their MCMC MLE produces the best fit in most cases.

Each estimator has its own advantages and disadvantages, and it is clear that the MPL estimator has the closest link with the traditional LR estimator. Therefore, for the purpose of developing a method for a spatially autocorrelated WE model, the MPL method of Besag (1975) appears to be the most appropriate starting point and the benchmark

against which the new model's performance will be compared against.

A NEW SACWE MODEL

Recall that, in Eq. 2, autocov_i is a continuous variable bounded between 0 and 1. In a WE modelling setting, however, all predictor patterns are required to take on binary values, and a predictor pattern with multiple discrete states needs to be redefined into several binary predictor patterns, each representing one of the states. Therefore, to incorporate the spatial term autocov_i in WE modelling one must first redefine autocov_i . Consider a set of nine hypothetical data points in a regular lattice in Figure 2, amongst which observation 5 is the data point of interest. Under the Queen contiguity design, observations (1, 2, 3, 4, 6, 7, 8, 9) are considered spatial neighbors of observation 5. According to Eq. 2, the spatial autocovariate autocov_5 is calculated as:

$$\text{autocov}_5 = \frac{\sum_{j=1}^9 w_{5j} y_j}{\sum_{j=1}^9 w_{5j}} \quad (3)$$

Following the convention in the literature, w_{55} is set to 0 as spatial observation 5 is not considered as its own neighbor, and $w_{5j} = 1$ for $j = 1, 2, \dots, 9$ and $j \neq 5$. Then:

$$\text{autocov}_5 = \frac{y_1 + y_2 + y_3 + y_4 + y_6 + y_7 + y_8 + y_9}{8} \quad (4)$$

And when Eq. 4 is substituted into Eq. 1, one can write:

$$\log\left(\frac{p_5}{1-p_5}\right) = \alpha + \beta_1 x_{1,5} + \frac{\beta_{\text{auto}}}{8} (y_1 + y_2 + y_3 + y_4 + y_6 + y_7 + y_8 + y_9) \quad (5)$$

where all spatial neighbors are treated as equivalent, and that each additional $y_j = 1$ for a spatial neighbor increases the value of the log-linear function by

1	4	7
2	5	8
3	6	9

Figure 2. A regular lattice of nine spatial observations.

exactly $\beta_{\text{auto}}/8$. Let us define a set of eight binary variables:

$$\text{autosum}(k)_5 = 1 \quad \text{iff} \quad \sum_{j=1, j \neq 5}^9 y_j = k, \quad k = 1, 2, \dots, 8 \quad (6)$$

It is clear that the set of $\text{autosum}(k)_5$'s and the continuous variable autocov_5 have a one-to-one correspondence and they contain the same amount of information. For instance, suppose that y_1, y_3 , and y_7 are all equal to 1, while the rest of the spatial neighbors are all 0, then $\text{autocov}_5 = 3/8$, which corresponds uniquely to a set of $\text{autosum}(k)_5$'s, in which $\text{autosum}(3)_5 = 1$ and all other $\text{autosum}(k)_5$'s equal to 0.

Generalizing for any spatial observation i , the set of eight binary $\text{autosum}(k)_i$'s, together with the exogenous binary variable $x_{1,i}$, give a total of nine predictor patterns, which are then used in WE modelling. The $\text{autosum}(k)_i$'s are derived from the spatial autocovariate autocov_i and they explicitly account for spatial autocorrelation, hence the new model is termed the SACWE model. Its implementation is relatively straightforward. Once a spatial structure has been specified, one only needs to count the number of neighbors scoring a value of 1 for each spatial unit, and generate a set of binary $\text{autosum}(k)_i$'s. As discussed in the section "The autologistic regression model," under the MPL setting the estimation of an ALR model is the same as that of a typical LR model. Therefore, along the same line of reasoning, the calibration of the SACWE model will be the same as that of a typical WE model.

It should be emphasized that, just as the MPL estimates for an ALR model are only valid in large samples, the results for a SACWE model must also be interpreted with caution and in the context of the available sample size. Currently, exact ML results for ALR models do not exist except for trivial cases. And while MCMC ML procedures have been shown to adequately approximate the distributions of the true ML estimators, for more detailed discussions see Wu and Huffer (1997) and Huffer and Wu (1998); these procedures are computationally highly intensive and do not appear to be transferable to WE modelling. While the inclusion of the $\text{autosum}(k)_i$'s in the SACWE model is a first attempt on accounting for spatial autocorrelation in WE modelling, its finite sample limitations need to be acknowledged.

A SIMULATION STUDY

Simulation Design

A simulation study will now be presented. Let there be six exogenous binary predictor patterns, $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5, x_6)^T$. Moreover, the potential for the binary dependent variable to be spatially autocorrelated needs to be incorporated in the data generating process. It has been well-established in the econometrics literature that underlying every discrete choice model is a so-called latent variable model; for detailed discussion see Johnston and Dinardo (1997). More specifically, suppose there exists a continuous but unobserved latent variable y^* such that:

$$y_i^* = \alpha + \mathbf{X}_i^T \boldsymbol{\beta} + \varepsilon_i \quad (7)$$

where α is the intercept term, \mathbf{X}_i is a (6×1) vector of six exogenous binary predictor patterns for observation i , ε_i is an identically and independently distributed (*i.i.d.*) random disturbance term, and $\boldsymbol{\beta}$ is a (6×1) vector of coefficients associated with the predictor patterns. The binary value of y_i can be defined by the following rule:

$$y_i = 1 \quad \text{if } y_i^* > 0, \text{ and } 0 \text{ otherwise} \quad (8)$$

which implies:

$$p(y_i = 1) = p(y_i^* > 0) \quad (9)$$

Therefore, to simulate spatially autocorrelated binary values of y_i , one can first simulate spatially autocorrelated continuous values of y_i^* . One can write down the well-known spatial autoregressive process (see Anselin, 1988) for the latent variable y^* in matrix notation:

$$\mathbf{y}^* = \rho \mathbf{W} \mathbf{y}^* + \boldsymbol{\alpha} \mathbf{1} + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (10)$$

where \mathbf{y}^* is an $(N \times 1)$ vector of unobserved latent variables, $\mathbf{1}$ is an $(N \times 1)$ vector of 1's, \mathbf{X} is an $(N \times 6)$ matrix of six exogenous binary predictor patterns observed for all spatial units, and $\boldsymbol{\varepsilon}$ is an $(N \times 1)$ vector of *i.i.d.* random disturbances. \mathbf{W} is the $(N \times N)$ spatial weights matrix, where w_{ij} identifies the spatial relationship between the i th and j th spatial unit. In the current study, \mathbf{W} is constructed based on the Queen contiguity structure (Fig. 1), where every spatial unit has eight neighbors. To avoid complications that can arise from the so-called edge effects, where spatial units on the edges of the

map have fewer neighbors and might require special treatments, the spatial units on the edges are excluded from the analysis to ensure that every unit in the current simulation study faces the same spatial structure and has the same number of neighbors. ρ is the so-called spatial autoregressive parameter. For a positively spatially autocorrelated process, ρ is bounded between $(0, 1)$. It is clear that, when $\rho = 0$, the model has no spatial autocorrelation and is reduced to the usual case of LR, while the closer ρ is to 1 the stronger the spatial autocorrelation. Although by definition the vector of latent variables \mathbf{y}^* is latent and unobserved, Eq. 10 has the following reduced form:

$$\mathbf{y}^* = (\mathbf{I} - \rho \mathbf{W})^{-1} (\boldsymbol{\alpha} \mathbf{1} + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon}) \quad (11)$$

where \mathbf{I} is an $(N \times N)$ identity matrix. It is clear that, given a fixed spatial structure specified by \mathbf{W} , and given a set of simulated \mathbf{X} , one can simulate a set of N spatially autocorrelated y_i^* . Then by Eq. 8 one can easily transform the continuous y^* into spatially autocorrelated binary y_i .

It will be assumed that ε_i is *i.i.d.* $N(0, 3^2)$, and x_j follows a Bernoulli distribution with $p(x_j = 1) = 0.5$ for all $j = 1, 2, \dots, 6$. The following parameter values will be used:

Parameter:	α	β_1	β_2	β_3	β_4	β_5	β_6
Value	-5.5	1.5	1.2	1.0	1.5	1.2	1.0

and the effect of spatial autocorrelation will be investigated by changing the value of the spatial autoregressive parameter ρ from 0 (no spatial autocorrelation) to 0.6 (strong spatial autocorrelation), with an increment of 0.1 in each scenario. To demonstrate that the method described above can adequately simulate spatially autocorrelated landscape, Figure 3 contains two typical maps from the simulations. It can be seen that, in the absence of spatial autocorrelation, the presence of the response variable characteristic (i.e., $y_i = 1$) is scattered randomly in the landscape with no specific patterns, while in the presence of strong spatial autocorrelation distinct clusters are formed.

For each iteration, a sample size of 450 (excluding the edge cells) is generated. The simulation is then repeated 1000 times for each scenario and all the results presented below are averaged results, unless specified otherwise.

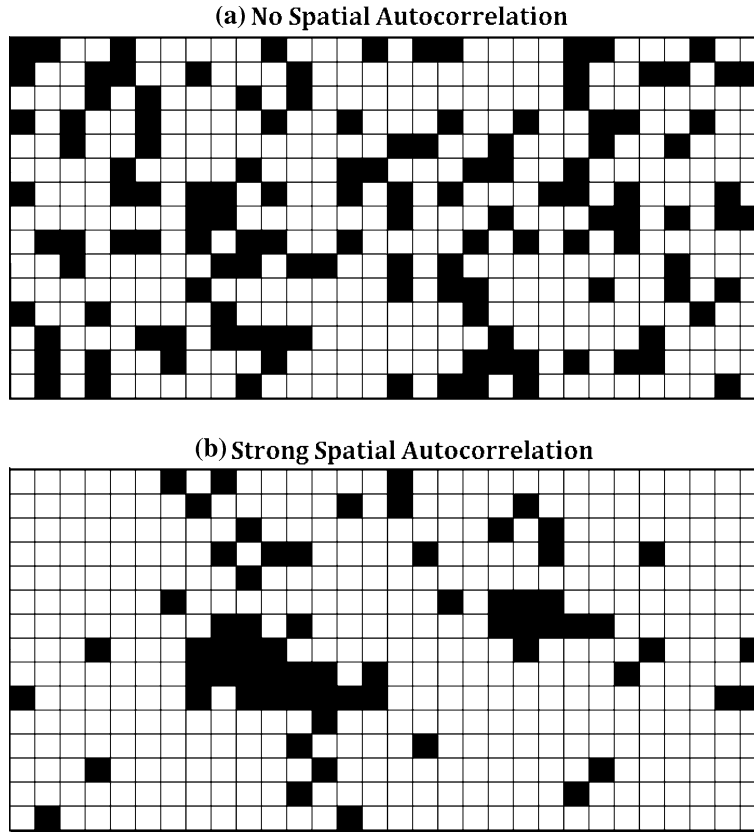


Figure 3. The shaded cells correspond to observations of $y_i = 1$. (a) A typical simulated landscape with no spatial autocorrelation has no distinct patterns; (b) a typical simulated landscape with strong spatial autocorrelation shows distinct clusters.

Simulation Results

The simulation results will be presented in two parts. Firstly, the in-sample fit and estimates of the three models, ALR, WE, and SACWE, will be compared. Secondly, the out-of-sample spatial predictive performances of the three models will be compared.

In-Sample Results

The in-sample fit of the three models will first be compared. Two measure of goodness of fit will be used in this study. The first one is the sum of squared residuals (SSR), as defined in Amemiya (1981):

$$SSR = \sum_{i=1}^{450} [y_i - \hat{p}(y_i = 1)]^2 \quad (12)$$

where $\hat{p}(y_i = 1)$ is the estimated probability of $y_i = 1$ computed by the model. If a model offers a good fit,

$\hat{p}(y_i = 1)$ will be high for cases where $y_i = 1$, and low for cases where $y_i = 0$. Clearly, a low value of SSR indicates a good fit. Table 1 provides a comparison of SSR for all three models (ALR, WE, and SACWE) for increasing values of the spatial autoregressive parameter ρ . From the last two columns of Table 1, it is clear that, for all values of ρ , ALR and SACWE outperform WE. When $\rho = 0$, the difference in SSR between the three models is minimal. This is expected, as in the case of zero spatial autocorrelation the three models are expected to be equivalent. But as ρ increases, i.e., as the strength of spatial autocorrelation increases, while the difference in SSR remains small between ALR and SACWE, both models exhibit increasingly superior fit over WE.

Another validation method often used for validating binary response models is receiver operator characteristic (ROC) curve analysis, which produces a plot of true positive identification rates

Table 1. Comparison of In-Sample SSR

ρ	SSR of Individual Models			% Difference Relative to SSR(WE)	
	ALR	WE	SACWE	ALR	SACWE
0.00	80.483	81.110	80.106	-0.77	-1.24
0.10	77.068	77.751	76.693	-0.88	-1.36
0.20	73.033	73.995	72.670	-1.30	-1.79
0.30	67.324	68.852	67.171	-2.22	-2.44
0.40	59.720	61.880	59.703	-3.49	-3.52
0.50	50.289	53.191	50.480	-5.46	-5.10
0.60	38.399	41.574	38.619	-7.64	-7.11

against false positive identification rates for all possible cutoff values. The distance between the ROC curve and the leading diagonal indicates the accuracy of the model in binary classification, and the larger the area under the curve the better the fit of the model. Detailed discussions of ROC analysis can be found in Vining and Gladish (1992), Zweig and Cambell (1993), Mathew, Jha, and Rawat (2007), amongst others. Figure 4 shows the ROC curves for three scenarios, $\rho = 0$, $\rho = 0.3$, and $\rho = 0.6$, respectively. In the absence of spatial autocorrelation ($\rho = 0$), the three ROC curves are almost identical. But as the value of ρ increases, the gap between WE and the two spatially autocorrelated models ALR and SACWE increases. When ρ is large ($\rho = 0.6$), this gap becomes substantial. The ROC curve analysis further confirms that, in the presence of spatial autocorrelation, the fit of SACWE is comparable to that of ALR, and it is superior to WE.

To see why WE underperforms against both ALR and SACWE, Figure 5 contains maps of estimated probabilities $\hat{p}(y_i = 1)$ for all three models in the case where $\rho = 0.6$ and for the same simulated landscape as that in Figure 3b. The darker the cell is the higher the value of the estimated probability. It can be seen that, while the estimated probabilities from both ALR and SACWE models exhibit clusters of high values, the type of behavior expected in the presence of spatial autocorrelation, the WE model fails to identify any clusters. And when these maps are compared against the true spatial landscape in Figure 3b, it can be seen that both ALR and SACWE mimic the clusters in the true landscape successfully, while WE does not.

It is also instructive to compare the estimate of β_{auto} for the spatial autocovariate autocov_i in ALR

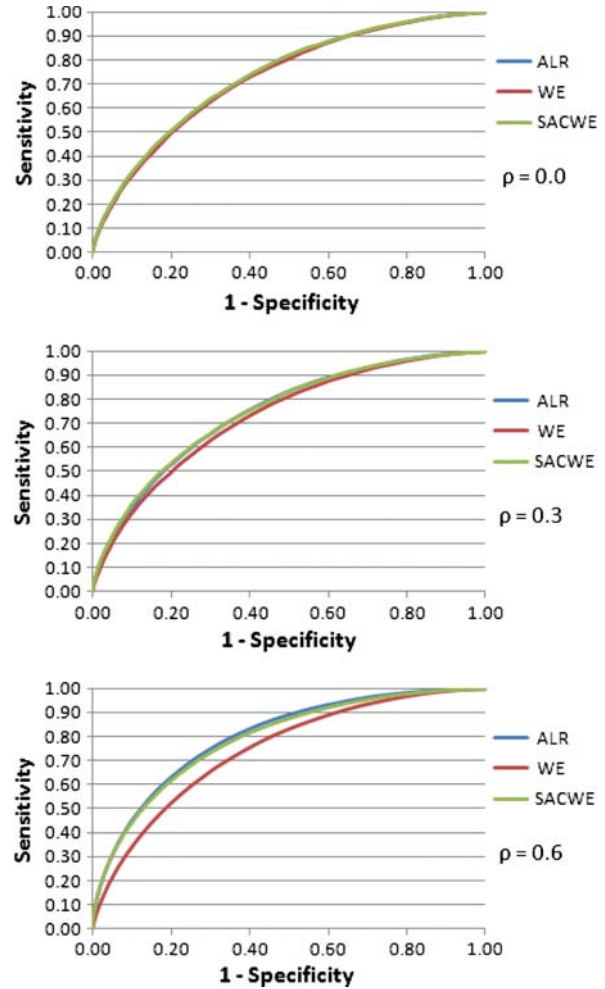


Figure 4. ROC curves of the three models (ALR, WE, and SACWE). As the strength of spatial autocorrelation (determined by ρ) increases, a sizable gap merges between the ROC curves of the spatially autocorrelated ALR and SACWE and that of the spatially uncorrelated WE.

against the contrasts calculated for the set of $\text{autosum}(k)_i$'s in SACWE. Their values are presented in Table 2. Note that, throughout the simulations conducted, $\text{autosum}(8)_i$ was found to be almost always 0, as the case of all eight spatial neighbors being 1 was extremely rare, and the contrast for $\text{autosum}(8)_i$ could not be calculated. But the behavior of the contrasts of the remaining $\text{autosum}(k)_i$'s shows interesting results. First, it is noted that, as ρ increases in value, the estimate of β_{auto} in ALR also increases. Similarly, the values of the contrasts for the $\text{autosum}(k)_i$'s also increase along with ρ in each column of the table, capturing the increasing effect of spatial autocorrelation.

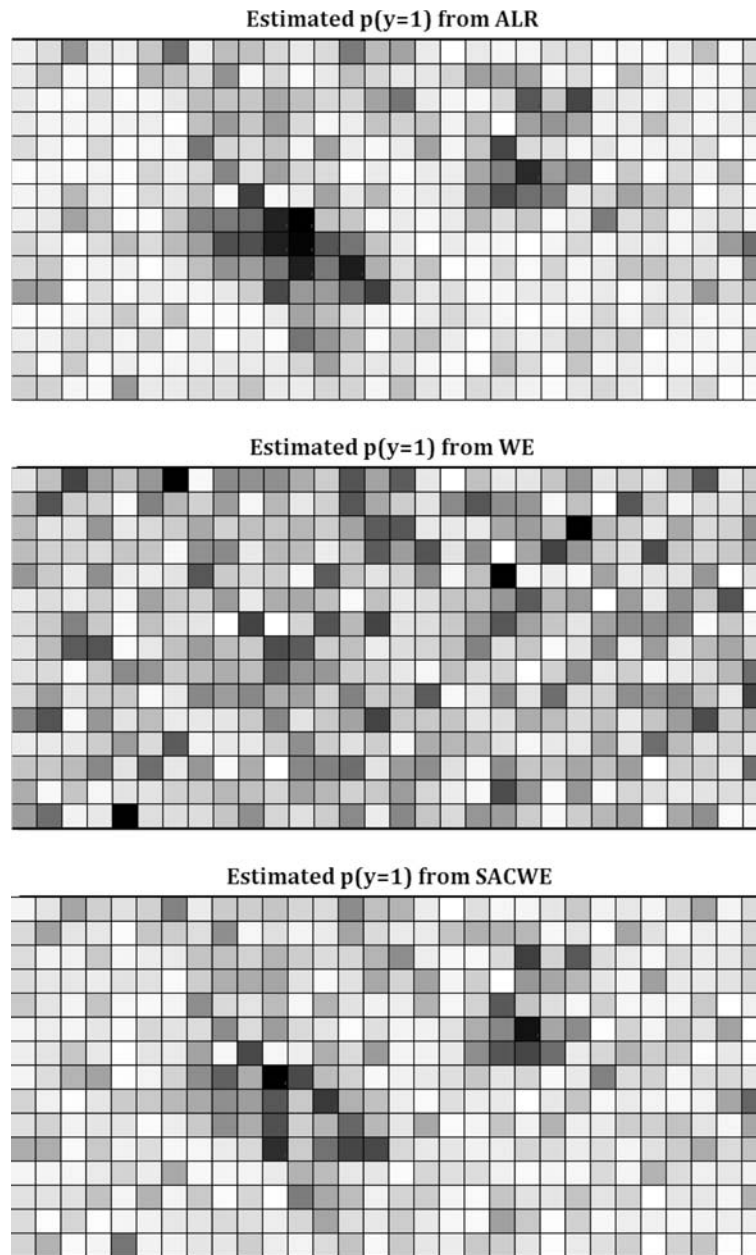


Figure 5. Maps of estimated probabilities from the three models (ALR, WE, and SACWE). The darker the cells are the higher the estimated probabilities are. While both ALR and SACWE show clusters of high probability estimates, WE does not.

Secondly, the general pattern across the rows of the table is that the contrast for $\text{autosum}(k+1)_i$ exceeds that of $\text{autosum}(k)_i$. Therefore, as the number of spatial neighbors recording 1 increases, the posterior probability estimate also increases in SACWE, a result that is consistent with expectations.

Out-of-Sample Predictions

It has been demonstrated in the previous section that both ALR and SACWE produce superior in-sample fit over WE in the presence of spatial autocorrelation, it is also interesting to compare the

Table 2. Comparison of the Estimated Effects of the Spatial Terms

ALR		SACWE							
$\beta(\text{autocov})$		Contrast of autosum(k)							
ρ	autocov	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$	$k = 8$
0.0	-0.193	0.002	-0.007	-0.018	-0.031	-0.083	0.306	0.921	N/A
0.1	0.408	-0.127	-0.039	0.062	0.116	0.138	0.568	0.981	N/A
0.2	1.043	-0.216	-0.001	0.137	0.245	0.354	0.810	1.117	N/A
0.3	1.750	-0.302	0.034	0.287	0.443	0.655	1.098	1.342	N/A
0.4	2.515	-0.296	0.146	0.447	0.658	1.020	1.376	1.493	N/A
0.5	3.515	-0.250	0.340	0.708	0.999	1.473	1.735	1.400	N/A
0.6	4.697	-0.091	0.599	1.040	1.573	2.041	2.087	1.962	N/A

spatial predictive performance of the three models. If the response variable y_i is continuous, to perform spatial predictions one can utilize the fact that the well-known spatial autoregressive equation:

$$\mathbf{y} = \rho \mathbf{W} \mathbf{y} + \alpha \mathbf{1} + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (13)$$

has a convenient reduced form:

$$\mathbf{y} = (\mathbf{I} - \rho \mathbf{W})^{-1} (\alpha \mathbf{1} + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon}) \quad (14)$$

where the value of the response variable \mathbf{y} is explicitly expressed as a function of \mathbf{X} ; for a more detailed discussion, refer to Anselin (1988). To predict the values of y_i in areas not yet sampled, so long as the exogenous predictor patterns \mathbf{X} are observed for those areas and that the spatial structure (as defined in \mathbf{W}) is known, one can compute their predicted values using Eq. 14.

However, when the response variable takes on discrete values, such as in the case of a binary y_i , analytical solutions for the reduced form do not exist, and simple solutions such as Eq. 14 cannot be used for spatial prediction of binary y_i . Moreover, even if we assume that a spatial model, such as the ALR of Eq. 1, can be perfectly specified and estimated, when the spatial landscape is only partially sampled it is likely that autocov_i cannot be calculated in many cases as some of the spatial neighbors have not yet been observed.

Augustin, Mugglestone, and Buckland (1996, 1998) have suggested incorporating the Gibbs sampler into the ALR model, where the presence/absence values in unsurveyed spatial units are recursively updated given neighboring values until convergence. Their method is computational intensive and does not appear to be directly transferable to the WE method. In the current simulation study, I will propose a simple method, where both model calibration and spatial prediction will utilize

1	5	9
2	6	10
3	7	11
4	8	12

Figure 6. An incompletely sampled spatial dataset of 12 points, where the shaded cells are observed, while the rest are not.

information only from units where observations on y_i are available. To demonstrate, suppose that amongst the 12 data points in Figure 6, units (1, 2, 3, 6, 9, 10) are observed (highlighted in grey), while units (4, 5, 7, 8, 11, 12) are unobserved. For the WE model, the lack of observations on ($y_4, y_5, y_7, y_8, y_{11}, y_{12}$) does not lead to any complications in either model calibration or prediction, as it does not take into account neighboring values. For the ALR and SACWE models, however, both the autocov_i and the set of $\text{autosum}(k)_i$'s need to be modified accordingly. In calibrating the ALR model, taking the 6th observation as an example, the equation becomes:

$$\log\left(\frac{p_6}{1-p_6}\right) = \alpha + \mathbf{X}_6^T \boldsymbol{\beta} + \beta_{\text{auto}} \text{autocov}_6 \quad (15)$$

where

$$\text{autocov}_6 = \frac{y_1 + y_2 + y_3 + y_9 + y_{10}}{5} \quad (16)$$

For unit 6, out of a total of eight possible neighbors only five of which (namely $y_1, y_2, y_3, y_9,$ and y_{10}) are

observed, and the spatial autocovariate autocov_6 is calculated as an averaged value of only these five neighbors. Similarly, in generating the set of $\text{autosum}(k)_6$'s for the SACWE model in using Eq. 6, only values on these five neighbors will be used. Now suppose that one is interested in calculating \hat{p}_7 , the predicted probability of $y_7 = 1$, using the ALR model:

$$\log\left(\frac{\hat{p}_7}{1 - \hat{p}_7}\right) = \alpha + \mathbf{X}_7^T \boldsymbol{\beta} + \beta_{\text{auto}} \text{autocov}_7 \quad (17)$$

where

$$\text{autocov}_7 = \frac{y_2 + y_3 + y_6 + y_{10}}{4} \quad (18)$$

Once again the spatial autocovariate autocov_7 is an averaged value of only the observed neighbors (namely $y_2, y_3, y_6,$ and y_{10}). Similarly, in using the SACWE model, the set of $\text{autosum}(k)_7$'s will also only involve these observed neighbors.

The forecasting experiment will be conducted as follows. Spatial observations will be generated for $\rho = 0.60$, which means significant spatial autocorrelation is present. Out of the simulated spatial sample of $N = 450$, a proportion of the observations will be randomly chosen as areas with ‘‘unobserved’’ y_i and set aside for forecasting performance evaluation, while the remaining observations will be used for model calibration. Three scenarios will be separately investigated: a sparsely sampled landscape of just $\sim 25\%$ of the units having been observed, a half-sampled landscape of $\sim 50\%$ being observed, and a well-covered landscape of $\sim 75\%$ being observed. Figure 7 shows what the three scenarios typically look like. The forecasting performances of the three models will be compared based on the SSR calculated for the unobserved cells, and the results are summarized in Table 3. It is clear that both ALR and SACWE forecast more accurately than WE in all three scenarios. Interestingly, the two spatially autocorrelated models show distinct advantages in all three scenarios, even in the extreme case where the spatial landscape is only sparsely sampled ($\sim 25\%$ observed). As the number of possible neighbors is 8 in a Queen contiguity spatial structure, in the most extreme scenario only one quarter (or 2) of the spatial neighbors are expected to be observed on average for each unit. This suggests that, despite the lack of knowledge on the majority of the neighbors, incorporation of spatial information from those limited number of observed neighbors still help to significantly improve the predictive

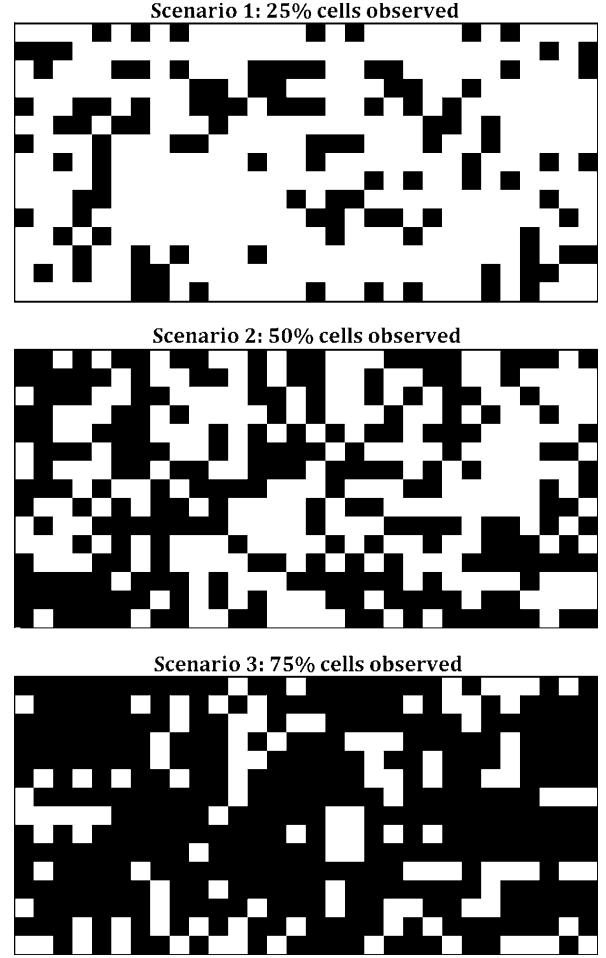


Figure 7. The shaded cells are randomly selected as being observed. Three scenarios are investigated: 25% observed, 50% observed, and 75% observed.

Table 3. Comparison of SSR of Predicted Cells

Scenario	SSR of Individual Models			% Difference Relative to SSR(WE)	
	ALR	WE	SACWE	ALR	SACWE
25% observed	28.812	30.467	28.882	-5.43	-5.20
50% observed	19.414	20.560	19.290	-5.57	-6.17
75% observed	9.531	10.186	9.537	-6.43	-6.37

performance of the model. Another way of explaining this result is that, while many neighbors are not observed and hence not used in forming predictions, the values of those neighbors that are observed take on greater significance. This can be seen from a comparison of the calibrated SACWE contrasts in Table 4. The contrasts for $\text{autosum}(1)$

Table 4. Comparison of the Calibrated SACWE Contrasts

	SACWE Contrasts	
	autosum(1)	autosum(2)
All cells observed	-0.091	0.599
25% observed	0.840	1.893

(having exactly one neighbor recording $y_i = 1$) and autosum(2) (having exactly two neighbor recording $y_i = 1$) are both significantly larger for the scenario where only 25% of the cells are observed. Thus, the positive evidence provided by any $y_i = 1$ neighbor is significantly more important when the landscape is only sparsely sampled.

CONCLUDING REMARKS AND FURTHER ISSUES

The new SACWE model developed in this paper attempts to incorporate spatial autocorrelation into binary pattern recognition. The method is simple and has close links with the well-known ALR method. Additional predictor patterns are generated by counting the number of spatial neighbors recording a value of 1 and are subsequently used in computing posterior probabilities. These additional predictor patterns are shown to contain the same amount of information as the autocovariate in an ALR model. The simulation study conducted show that the SACWE model is on par with ALR both in terms of in-sample fit and out-of-sample predictions, and it significantly outperforms WE when spatial autocorrelation is present. The author hopes that the introduction of the SACWE provides a useful addition to the existing spatial modelling toolbox and can expand the application of GIS-based weights of evidence modelling into research areas previously ignored.

It should be acknowledged that modelling spatially autocorrelated binary data is a complex problem. There remains a number of important issues that have not been resolved in this paper, which will be listed below:

1. In this paper, the spatial landscape is based on a regular lattice, which may not be the case in many empirical research. One way of dealing with this problem could be to

incorporate a distance measure, where only observations within a certain radius can be thought of as spatial neighbors. As a related issue, it is also noted that another popular spatial structure commonly used in empirical research is that of a distance-decay design, where spatial neighbors are considered less important when they are further away. One possible alternative is to divide the spatial neighborhood into zones, where observations within the same zone are treated as equally influential, but zones further away are less important than zones close by. Finally, the choice of cell size in a regular lattice may also be of great importance. An optimal cell size should be chosen judiciously and in the context of the specific spatial process under investigation. The spatial process within each cell should be relatively homogenous, while significant spatial interactions are expected to occur between cells.

2. As emphasized throughout the paper, due to the lack of independence between response variables, statistical properties of the SACWE model must be interpreted cautiously and taking into consideration the available sample size. The endogenous nature of both the spatial autocovariate $auto-cov_i$ in the ALR and the $autosum(k)_i$'s in the SACWE invalidates most of the traditional finite sample results. It remains to be seen whether a more advanced method can be found to circumvent this issue when calibrating the SACWE.
3. The spatial forecasting experiment conducted has assumed that a large spatial landscape is incompletely sampled, and that values within the landscape need to be predicted. But it is also often the case that an entire section of the landscape becomes unobserved. In these situations, information on spatial neighbors is completely missing, and it remains to be seen if it is still possible to produce predictions using SACWE.

ACKNOWLEDGMENT

The author thanks Jerry Jensen and the reviewers for their insightful comments.

REFERENCES

- Agterberg, F. P., 1992, Combining indicator patterns in weights of evidence modelling for resource evaluation: *Nonrenew. Res.*, v. 1, no. 1, p. 35–50.
- Agterberg, F. P., and Bonham-Carter, G. F., 1999, Logistic regression and weights of evidence modelling in mineral exploration, *in* Computer Applications in the Mineral Industries: Golden, CO, p. 483–490.
- Agterberg, F. P., Bonham-Carter, G. F., and Wright, D. F., 1990, Statistical pattern integration for mineral exploration, *in* Gaál, G., and Merriam, D. F., eds., Computer Applications in Resource Exploration Prediction and Assessment for Metals and Petroleum: Oxford, Pergamon, p. 1–21.
- Agterberg, F. P., Bonham-Carter, G. F., Wright, D. F., and Cheng, Q., 1993, Weights of evidence modelling and weighted logistic regression for mineral potential mapping, *in* Davis, J. C., and Herzfeld, U. C., eds., Computers in Geology, 25 Years of Progress: Oxford University Press, New York, p. 13–32.
- Amemiya, T., 1981, Qualitative response models: a survey: *J. Econ. Lit.*, v. 19, p. 1483–1536.
- Anselin, L., 1980, Estimation methods for spatial autoregressive structures: a study in spatial econometrics: Regional Science Dissertation and Monograph Series 8, Program in Urban and Regional Studies, Cornell University, Ithaca, New York, 273 p.
- Anselin, L. 1988, Spatial econometrics: methods and models: Kluwer Academic Publishers, Dordrecht, 284 p.
- Augustin, N. H., Muggleston, M. A., and Buckland, S. T., 1996, An autologistic model for the spatial distribution of wildlife: *J. Appl. Ecol.*, v. 33, p. 339–347.
- Augustin, N. H., Muggleston, M. A., and Buckland, S. T., 1998, The role of simulation in modelling spatially correlated data: *Environmetrics*, v. 9, p. 175–196.
- Besag, J., 1972, Nearest-neighbour systems and the auto-logistic model for binary data: *J. R. Stat. Soc. B*, v. 34, p. 75–83.
- Besag, J., 1974, Spatial interaction and the statistical analysis of lattice systems (with discussion): *J. R. Stat. Soc. B*, v. 36, p. 192–236.
- Besag, J., 1975, Statistical analysis of non-lattice data: *Statistician*, v. 24, p. 179–195.
- Bonham-Carter, G. F., 1994, Geographic Information Systems for geoscientists: Oxford, Pergamon, 398 p.
- Bonham-Carter, G. F., Agterberg, F. P., and Wright, D. F., 1988, Integration of geological datasets for gold exploration in Nova Scotia: *Photogram. Remote Sens.*, v. 54, no. 11, p. 1585–1592.
- Bonham-Carter, G. F., Agterberg, F. P., and Wright, D. F., 1989, Weights of evidence modelling: a new approach to mapping mineral potential, *in* Agterberg, F. P. and Bonham-Carter, G. F., eds., Statistical Applications in the Earth Sciences: Geological Survey Canada Paper 9-9, p. 171–183.
- Cliff, A. D., and Ord, J. K., 1973, Spatial autocorrelation: Pion, London, 178 p.
- Comets, F., 1992, On consistency of a class of estimators for exponential families of Markov random fields on the lattice: *Ann. Stat.*, v. 20, no. 1, p. 445–568.
- Cressie, N. A. C., 1993, Statistics for spatial data, revised edition: Wiley, New York, 900 p.
- Dahal, R. K., Hasegawa, S., Nonomura, A., Yamanaka, M., Masuda, T., and Nishino, K., 2008, GIS-based weights-of-evidence modelling of rainfall-induced landslides in small catchments for landslide susceptibility mapping: *Environ. Geol.*, v. 54, p. 311–324.
- Deng, M., 2009, A conditional dependence adjusted weights of evidence model: *Nat. Resour. Res.* doi:10.1007/s11053-009-9101-5.
- Emelyanova, I. V., Donald, G. E., Miron, D. J., Henry, D. A., and Garner, M. G., 2008, Probabilistic modelling of cattle farm distribution in Australia: *Environ. Model. Assess.*, v. 14, p. 449–465.
- Haining, R., 1985, The spatial structure of competition and equilibrium price dispersion: *Geograph. Anal.*, v. 17, p. 231–242.
- Hansen, D. T., 2000, Describing GIS applications: spatial statistics and weights of evidence extension to ArcView in the analysis of the distribution of archeology sites in the landscape: Proceedings of the 20th Annual ESRI International User Conference, San Diego, CA.
- Hansen, D. T., West, J., Simpson, B., and Welch, P., 2002, Modeling spatial uncertainty in analysis of archeological site distribution: Proceedings of the 22nd Annual ESRI International User Conference, San Diego, CA.
- Huffer, F. W., and Wu, H., 1998, Markov Chain Monte Carlo for autologistic regression models with application to the distribution of plant species: *Biometrics*, v. 54, p. 509–524.
- Johnston, J., and Dinardo, J., 1997, Econometric methods, 4th ed.: McGraw-Hill, New York, 531 p.
- Mathew, J., Jha, V. K., and Rawat, G. S., 2007, Weights of evidence modelling for landslide hazard zonation mapping in part of Bhagirathi valley, Uttarakhand: *Curr. Sci.*, v. 92, no. 5, p. 628–638.
- Ord, J. K., 1975, Estimation methods for models of spatial interaction: *J. Am. Stat. Assoc.*, v. 70, p. 120–126.
- Romero-Calcerrada, R., and Luque, S., 2006, Habitat quality assessment using Weights-of-Evidence based GIS modelling: the case of *Picoides tridactylus* as species indicator of the biodiversity value of the Finnish forest: *Ecol. Modell.*, v. 196, p. 62–76.
- Tobler, W. R., 1970, A computer movie simulating urban growth in the Detroit Region: *Econ. Geogr.*, v. 46, p. 234–240.
- Vining, D. J., and Gladish, G. W., 1992, Receiver operating characteristic curves: a basic understanding: *RadioGraphics*, v. 12, p. 1147–1154.
- Wintle, B. A., and Bardos, D. C., 2006, Modeling species-habitat relationships with spatially autocorrelated observation data: *Ecol. Appl.*, v. 16, no. 5, p. 1945–1958.
- Wu, H., and Huffer, F. W., 1997, Modelling the distribution of plant species using the autologistic regression model: *Environ. Ecol. Stat.*, v. 4, p. 49–64.
- Zweig, M. H., and Cambell, G., 1993, Receiver operator characteristic plots: a fundamental evaluation tool in clinical medicine: *Clin. Chem.*, v. 39, p. 561–577.