Detection of temporal changes in spatial dependence

Petrutza C. Caragea[1] and Mark S. Kaiser[1]

[1] Department of Statistics, Iowa State University, Ames, IA 50011-1210, USA

1 Problem Setting

The general problem structure considered in this paper is that of a short to moderate sequence of situations in which a spatial analysis is appropriate at each individual point in time. This structure arises in many situations including environmental monitoring across a region on an annual basis and assessment of the spread of an invasive plant species after introduction to an environment from which it has been absent. In modeling such problems it is often of interest to detect changes in the structure of the underlying scientific processes that lead to change in observable phenomena over time. A key for successful statistical formulation in these situations is that parameter values in a model correspond to components of the underlying scientific process in such a way that the implications of temporal changes in values of the parameters are discernible. For example, temporal changes in the expected values of a model are generally interpreted as corresponding to shifts in the magnitude of the process it represents. But, the well-known lack of a unique decomposition of spatial models into large-scale and small-scale components can complicate meaningful interpretation of parameter value changes over time. Specifically, if the marginal mean structure of a process cannot be isolated in a small set of model parameters, then changes in parameter values may not correspond to easily interpretable shifts in process level or magnitude. In fact, if parameters that appear in the large-scale model component and parameters that appear in the small-scale model component are related to both marginal means and spatial dependence in a highly interacting manner, changes in parameter values over time may not even indicate any meaningful change in the phenomenon under investigation; patterns of changes in interacting parameters that do or do not correspond to substantial changes in scientific process may be exceedingly difficult to identify.
The ideal occurrence in a statistical model is for some parameters to control marginal mean structure, some other parameters to control variance and covariance structure, and for these sets of parameters to be independent. This leads to parameters that are *separable* in the sense of interpretation

and *stable* in the sense that estimators of parameters in one set are not highly influenced by estimators of parameters in the other set, at least in terms of expectation. The Gaussian model perhaps best exemplifies this situation, with some parameters that control only expectation and others that control only covariance, and for which estimation of variance and covariance parameters influence precision of estimators of mean parameters to a much greater degree than bias. These desirable properties of separable parameters and stable estimators are not typically realized in the general class of Markov random field models formulated from exponential family conditional distributions.
This paper will focus on autologistic models of the type introduced by Besag (1974) which are commonly used to model spatial patterns of presence and absence of an event or phenomenon of interest. The traditional parameterization of these models suggests a natural division into parameters that can be used to model large-scale (marginal mean) and small-scale (spatial dependence) structures. But, as will be demonstrated in the sequel, this seemingly natural interpretation does not lead to separable and stable parameters in the sense defined previously and this, in turn, damages our ability to detect meaningful changes in scientific or physical processes on the basis of changes in values of parameter estimates. An alternative parameterization for autologistic models will be proposed that does lead to separable and stable parameters so that changes in individual parameters over time can be interpreted relative to meaningful components of the underlying process.

2 Traditional and Centered Autologistic Models

The setup we consider is that of data collected at n spatial locations $s_i = (u_i, v_i)$, $i = 1, \ldots, n$. and assume that the random variable $Y(s_i)$ has the property that $[Y(s_i) \mid \{Y(s_j) : j \neq i\}] \sim Binary(p_i)$, where $p_i = \text{Prob}[y(s_i) = 1 \mid \text{all } y(s_j), j \neq i]$. For a neighborhood N_i for each random variable, Besag (1974) showed that if p_i are expressed as

$$p_i = \frac{\exp(\theta_i)}{1 + \exp(\theta_i)}, \qquad i = 1, \ldots, n \tag{1}$$

then the natural parameter functions for this formulation must be

$$\theta_i = \alpha_i + \sum_{j \in N_i} \eta_{ij} y(s_j) \qquad i = 1, \ldots, n.$$

If these conditional probability mass functions are modeled in exponential family form with natural parameter θ_i and covariate information $x_i; i = 1, \ldots, n$ is incorporated, then the traditional model parameterization is

$$\theta_i = x_i^T \beta + \sum_{j \in N_i} \eta_{ij} y(s_j). \tag{2}$$

One problem with this parametrization is that, in the presence of spatial dependence, the marginal expectation of the process, $E[Y(s_i)]$, does not equal p_i as defined by (1), thus the lack of *separability.* As an alternative to the traditional formulation we propose the following *centered* version of the model,

$$\theta_i = x_i^T\beta + \sum_{j\in N_i} \eta_{ij} \left(y(s_j) - \frac{\exp(x_j^T\beta)}{1+\exp(x_j^T\beta)} \right). \qquad (3)$$

3 Simulation Results

The two desired properties of *separability* and *stability* conferred by the centered parametrization are illustrated by a simulation study. We consider here the case of a set of binary variables on a regular lattice with one covariate constructed to exhibit a SW-NE trend and models with omnidirectional spatial dependence, i.e. $\eta = \eta_{ij},$ for $i, j = 1, \ldots, n$. Since we are interested in exploring the separability and stability of the estimators in the two models for varying values of spatial dependence, we simulated data sets for η varying between 0 and 2.5: when $\eta = 0$, the model reduces to the usual logistic regression model for independent observations, while for $\eta = 2.5$ the spatial dependence is fairly strong, and it tends to dominate the process.
In an effort to assess separability in terms of parameters that may be identified with marginal mean structure, we performed a series of simulations using models (2) and (3). For each pair of simulated data sets, we computed an average value for the data sets by averaging over all locations under each model. The means of these averages across simulated data sets then represent the marginal means of the location averaged values for the models. The resulting values are presented in Figure 1, along with a line that represents the marginal mean of the location averaged values under an independence model, namely,

$$\mu(\beta) = \text{avg}_i \left\{ \frac{\exp(x_i^T\beta)}{1-\exp(x_i^T\beta)} \right\}. \qquad (4)$$

As can be seen in Figure 1, as the dependence parameter η increases, the marginal mean values for model (3) remain more similar to the values in (4) than do the marginal mean values for model (2), indicating that the parameters β allow a constant interpretation in model (3) across levels of dependence, but that this is not true for the corresponding β of model (2). Thus, the large-scale parameter β is more separable from the dependence parameter η in the centered model than the traditional model.
Another simulation was conducted to examine stability in estimation of the dependence parameter η. Simulated data sets were produced from model (3) and estimates were found by applying Besag's pseudo-likelihood procedure

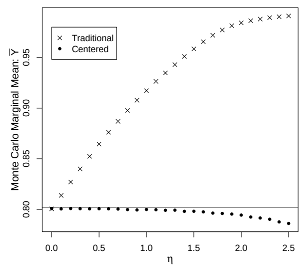

FIGURE 1. Monte Carlo averages across locations for the traditional and centered model. The solid line represents the marginal mean for $\eta = 0$.

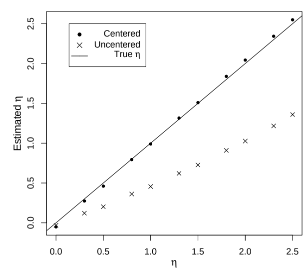

FIGURE 2. Estimated values for the dependence parameter η by the traditional and centered model for various levels of spatial dependence.

under both models (2)and (3). Figure 2 presents Monte Carlo estimates of expected values for these estimators over a range of values of η from 0 to 2.5. As is clearly indicated in this figure, the centered model (3) results in estimated values of spatial dependence that are more accurate.

The upshot of the simulations conducted is that marginal mean structure is embodied in the parameter β and spatial dependence is embodied in the parameter η under the centered model (3) but not the traditional model (2). Interestingly, estimation under either (2) or (3) tends to produce estimated values that fit the observed data sets to roughly the same degree, based on a prediction mean squared error assessment. Thus, while either model may be used to fit a given set of data, individual parameter values for the traditional model (2) do not maintain a constant meaning across different situations, while this is more nearly true for the centered model (3). In the types of monitoring situations described in the Introduction, it would be difficult to interpret shifts in estimated parameters under the traditional model but this would be possible under our proposed centered version.

A long term ecological study on succession on abandoned agricultural fields, known as the Buell-Small Succession Study is used to illustrate the applicability of the centered model for monitoring changes over time in distribution of an invasive plant species.