The Annals of Statistics
2012, Vol. 40, No. 1, 104–130
DOI: 10.1214/11-AOS948

GOODNESS OF FIT TESTS FOR A CLASS OF MARKOV RANDOM FIELD MODELS

BY MARK S. KAISER, SOUMENDRA N. LAHIRI[1] AND DANIEL J. NORDMAN[2]

Iowa State University, Texas A&M University and Iowa State University

This paper develops goodness of fit statistics that can be used to formally assess Markov random field models for spatial data, when the model distributions are discrete or continuous and potentially parametric. Test statistics are formed from generalized spatial residuals which are collected over groups of nonneighboring spatial observations, called concliques. Under a hypothesized Markov model structure, spatial residuals within each conclique are shown to be independent and identically distributed as uniform variables. The information from a series of concliques can be then pooled into goodness of fit statistics. Under some conditions, large sample distributions of these statistics are explicitly derived for testing both simple and composite hypotheses, where the latter involves additional parametric estimation steps. The distributional results are verified through simulation, and a data example illustrates the method for model assessment.

1. Introduction. Conditionally specified models formulated on the basis of an underlying Markov random field (MRF) are an attractive alternative to continuous random field specification for the analysis of problems that involve spatial dependence structures. By far the most common of such models are those formulated using a conditional Gaussian distribution (e.g., [42]), but models may also be constructed using a number of other conditional distributions such as a beta [23, 31], binary [8], Poisson [4] or Winsorized Poisson [29], and general specifications are available for many exponential families [2, 31].

In an applied spatial setting, we assume that observations are available at a finite set of geo-referenced locations $\{\mathbf{s}_i : i = 1, \dots, N\}$, and to these locations we assign the random variables $\{Y(\mathbf{s}_i) : i = 1, \dots, N\}$. In general, locations are arbitrarily indexed in d-dimensional real space. A MRF is typically constructed by specifying for each location $\mathbf{s}_i$ a *neighborhood*, consisting of other locations on which the full conditional distribution of $Y(\mathbf{s}_i)$ will be functionally dependent. Let the conditional cumulative distribution function (c.d.f.) of $Y(\mathbf{s}_i)$ given $\{Y(\mathbf{s}_j) = y(\mathbf{s}_j) : j \neq i\}$ be denoted as F_i and define $\mathcal{N}_i \equiv \{\mathbf{s}_j \neq \mathbf{s}_i$, and F_i depends functionally on $y(\mathbf{s}_j)\}$. Also define $\mathbf{y}(\mathcal{N}_i) \equiv \{y(\mathbf{s}_j) : \mathbf{s}_j \in \mathcal{N}_i\}$. The Markov

Received April 2010; revised November 2011.
[1]Supported in part by NSF Grants DMS-07-07139 and DMS-10-07703.
[2]Supported in part by NSF Grant DMS-09-06588.
MSC2010 subject classifications. Primary 62F03; secondary 62M30.
Key words and phrases. Increasing domain asymptotics, probability integral transform, spatial processes, spatial residuals.

assumption implies that

$$F_i\big(\cdot|\{y(\mathbf{s}_j):\mathbf{s}_j \neq \mathbf{s}_i\}\big) = F_i\big(\cdot|\{y(\mathbf{s}_j):\mathbf{s}_j \in \mathcal{N}_i\}\big) = F_i(\cdot|\mathbf{y}(\mathcal{N}_i)). \tag{1.1}$$

A model is formulated by specifying, for each $i = 1, \dots, N$, a conditional c.d.f. in (1.1). Conditions necessary for a set of such conditionals to correspond to a joint distribution for $\{Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_N)\}$ are given by Arnold, Castillo and Sarabia [2] and a constructive process with useful conditions sufficient for existence of a joint are laid out in Kaiser and Cressie [30]. Models may be constructed for both discrete and continuous random variables, on regular or irregular lattices, with or without an equal number of neighbors for each location (including $\mathcal{N}_i = \varnothing$ for some locations) and possibly including information from spatial covariates. The construction of models for applications is thus very flexible.

A number of our results and, in particular, Theorem 2.1 to follow, can be generalized to some of the variable situations just described, but it will be beneficial for developing theoretical results to define a setting that is broad but highly structured. We desire a spatial *process* defined on grid nodes of the d-dimensional integer lattice $\mathbb{Z}^d$, where $\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$. We stipulate a number of restrictions for this process that, while not capable of covering all of the finite-dimensional models mentioned previously, is flexible enough to be meaningful in many applied situations. We formally consider specifying an MRF model for a spatial process $\mathbf{Y} \equiv \{Y(\mathbf{s}):\mathbf{s} \in \mathbb{Z}^d\}$, rather than a model (1.1) developed with respect to a finite collection of (possibly nonlattice) data sites $\{Y(\mathbf{s}_i):i = 1, \dots, N\}$. To this end, assume that for any $\mathbf{s} \in \mathbb{Z}^d$ neighborhoods can be constructed using a standard template $\mathcal{M} \subset \mathbb{Z}^d \setminus \{\mathbf{0}\}$ as $\mathcal{N}(\mathbf{s}) = \mathbf{s} + \mathcal{M}$, with $|\mathcal{M}| < \infty$ denoting the size of $\mathcal{M}$. Some examples of $\mathcal{M}$ are given in the next section. We then assume that the process $\mathbf{Y}$ has a stationary distribution function $F(\cdot|\cdot)$ such that, for any $\mathbf{s} \in \mathbb{Z}^d$, the conditional c.d.f. of $Y(\mathbf{s})$ given all remaining variables $\{Y(\mathbf{t}):\mathbf{t} \in \mathbb{Z}^d, \mathbf{t} \neq \mathbf{s}\}$ can be written as

$$F\big(\cdot|\{Y(\mathbf{t}):\mathbf{t} \in \mathbb{Z}^d, \mathbf{t} \neq \mathbf{s}\}\big) = F\big(\cdot|\{Y(\mathbf{t}):\mathbf{t} \in \mathcal{N}(\mathbf{s})\}\big) \tag{1.2}$$

under a Markov assumption.

Given a hypothesized or estimated model, our concern is how one might conduct a goodness of fit (GOF) procedure, either through informal diagnostics or by using formal probability results that lead to a GOF test. The approach we propose here may be viewed within either the context of a pure GOF test to address the question of whether a (possibly fitted) model provides an adequate description of observed data. This is an issue of model assessment and different from model selection, which has been considered, for example, with penalized pseudo-likelihood for parametric MRF models; cf. [11, 21, 26]. Additionally, while other GOF tests may be possible for certain joint model specifications (e.g., a frequency-domain approach for Gaussian processes; cf. [1]), we focus solely on conditional model specifications. The GOF variates introduced in the next section may be used as either diagnostic quantities or as the basis for a formal GOF test as presented in Section 3.

The remainder of this article is organized as follows. In Section 2 we introduce the concept of a conclique and derive GOF variates that form the basis of our approach, using an adaptation of a multivariate probability integral transform (PIT). Section 3 develops a formal methodology for combining these variates over concliques to create GOF tests of Markov models under both simple and composite hypotheses. These tests are omnibus in the sense that they assess the hypothesized model in total, including the neighborhood structure selected, specification of dependence as isotropic or directional, and the form of the modeled conditional distributions. Theoretical results are presented in Section 4 that establish the limiting sampling distributions of GOF tests under the null hypothesis. Section 5 describes a numerical study to support the theoretical findings. Section 6 provides an application of the GOF tests in model assessment for agricultural trials. Section 7 contains concluding remarks and discussions on extensions. Section 8 provides a proof of the foundational conclique result (Theorem 2.1), and all other proofs regarding the asymptotic distribution of GOF test statistics appear in supplementary material [32].

2. Generalized spatial residuals. In this section we derive the basic quantities that form the basis for our GOF procedures. We consider these quantities to be a type of generalized residuals because they fit within the framework suggested by Cox and Snell [9]. In particular, these generalized spatial residuals will be derived using an extended version of Rosenblatt's [41] multivariate PIT combined with a partitioning of spatial locations into sets such that the residuals within each set constitute a random sample from a uniform distribution on the unit interval, under the true model. As discussed by Brockwell [7] and Czado et al. [12], the PIT formulation allows arbitrary model distributions to be considered in assessing GOF, rather than simply continuous ones. Similar transformations, with subsequent formal or informal checks for uniformity, have been important in evaluating the GOF of, and the quality of predictive forecasts from, various models for time series; cf. [13, 15, 16, 19, 24, 27].

2.1. *Concliques.* Before providing the transform that defines our generalized spatial residuals, it is necessary to develop a method for partitioning the total set of spatial locations at which observations are available into subsets with certain properties. We call such sets *concliques* because they are defined as the converse of what are called cliques by Hammersley and Clifford [22]. In the case of regular lattices with neighborhoods defined using either four-nearest or eight-nearest neighbor structures, concliques correspond exactly to the so-called coding sets of Besag [4], which were suggested for use in forming conditional likelihoods for estimation. The key property of concliques, however, allows construction of such sets in more general settings including irregular lattices and hence the new name.

As defined in [22], a *clique* is a set of locations such that each location in the set is a neighbor of every other location in the set. Similar terminology exists in graph theory, where a subset of graph vertices (e.g., locations) form a clique if every two vertices in the subset are connected by an edge [45]. We define a *conclique* as a set of locations such that no location in the set is a neighbor of any other location in the set. Any two members of a conclique may share common neighbors, but they cannot be neighbors themselves. Additionally, every set of a single location can be treated as both a clique or conclique. In the parlance of graphs, the analog of a conclique is a so-called "independent set," defined by a set of vertices in which no two vertices share an edge. This particular graph terminology conflicts with the probabilistic notion of independence, while a "conclique" truly represents a *conditionally* independent set of locations in a MRF model.

While the result of the next subsection holds for any collection of concliques, in practice what is desired is a collection of concliques that suitably partition all observed locations. To achieve this under the process model (1.2), we identify a collection of concliques $\{\mathcal{C}_j : j = 1, \dots, q\}$ that partition the entire grid $\mathbb{Z}^d$. We define a collection of concliques to be a *minimal conclique cover* if it contains the smallest number of concliques needed to partition the set of all locations. In graph theory, this concept is related to determining the smallest (or chromatic) number of colors needed to color a graph (with no two edge-connected vertices sharing the same color) or, equivalently, the smallest number of independent sets needed to partition graph vertices [25]. In practice, identifying a minimal conclique cover is valuable since our procedure produces one test statistic for each conclique in a collection, and those statistics must then be combined into one overall value for a formal GOF test.

EXAMPLE 2.1 (A 4-nearest neighbor model on $\mathbb{Z}^2$). Here, let $\mathbf{s} = (u, v)' \in \mathbb{Z}^2$ for a horizontal coordinate u and a vertical coordinate v. The neighborhood structure of a 4-nearest neighbor model is produced with the template $\mathcal{M} = \{(-1, 0)', (1, 0)', (0, 1)', (0, -1)'\}$, so that $\mathcal{N}(\mathbf{s})$ for a given location $\mathbf{s}$ and neighbors $*$ is as shown in the following figure:

$$\begin{array}{ccc} \cdot & * & \cdot \\ * & \mathbf{s} & * \\ \cdot & * & \cdot \end{array}$$

In this case, the minimal conclique cover contains two members, $\mathcal{C}_1$ and $\mathcal{C}_2$, with elements denoted by 1's and 2's, respectively, as shown below.

Minimal conclique collection for a 4*-nearest neighbor model*:

1	2	1	2	1	2	1	2	1	2
2	1	2	1	2	1	2	1	2	1
1	2	1	2	1	2	1	2	1	2
2	1	2	1	2	1	2	1	2	1
1	2	1	2	1	2	1	2	1	2

EXAMPLE 2.2 (An 8-nearest neighbor model on $\mathbb{Z}^2$). As in the previous example, let $\mathbf{s} = (u, v)'$ but take $\mathcal{M} = \{(u, v)' : \max\{|u|, |v|\} = 1\}$. The neighborhood structure of an 8-nearest neighbor model is then shown in the following figure for a location $\mathbf{s} \in \mathbb{Z}^2$ and neighbors $*$:

$$\begin{matrix} * & * & * \\ * & \mathbf{s} & * \\ * & * & * \end{matrix}$$

For the 8-nearest neighbor model, there are four concliques in the minimal cover, $\mathcal{C}_1, \ldots, \mathcal{C}_4$, with elements denoted by 1's, 2's, 3's and 4's in the following figure, respectively.

Minimal conclique cover for an 8*-nearest neighbor model*:

1	2	1	2	1	2	1	2	1	2
3	4	3	4	3	4	3	4	3	4
1	2	1	2	1	2	1	2	1	2
3	4	3	4	3	4	3	4	3	4
1	2	1	2	1	2	1	2	1	2

2.2. *Defining generalized spatial residuals.* Let $\{A(\mathbf{s}) : \mathbf{s} \in \mathbb{Z}^d\}$ denote a collection of independent and identically distributed (i.i.d.) random variables, which are Uniform (0, 1) and also independent of the spatial process $\mathbf{Y}$. For any $\mathbf{s} \in \mathbb{Z}^d$, we then define a random generalized spatial residual as

$$\begin{aligned} U(\mathbf{s}) = {} & \big(1 - A(\mathbf{s})\big) \cdot F\big(Y(\mathbf{s})|\{Y(\mathbf{t}) : \mathbf{t} \in \mathcal{N}(\mathbf{s})\}\big) \\ & + A(\mathbf{s}) \cdot F^-\big(Y(\mathbf{s})|\{Y(\mathbf{t}) : \mathbf{t} \in \mathcal{N}(\mathbf{s})\}\big), \end{aligned} \tag{2.1}$$

where $F(\cdot|\cdot)$ denotes the (stationary) c.d.f. from (1.2), and $F^-(\cdot|\cdot)$ denotes the left limit of the c.d.f., that is, $F^-(y|\{Y(\mathbf{t}) : \mathbf{t} \in \mathcal{N}(\mathbf{s})\}) = P(Y(\mathbf{s}) < y|\{Y(\mathbf{t}) : \mathbf{t} \in \mathcal{N}(\mathbf{s})\})$, $y \in \mathbb{R}$. This residual applies the notion of a randomized PIT [7], allowing for a noncontinuous c.d.f. $F(\cdot|\cdot)$ to be considered. When $F(\cdot|\cdot)$ is continuous, the spatial residual reduces to a PIT $U(\mathbf{s}) = F(Y(\mathbf{s})|\{Y(\mathbf{t}) : \mathbf{t} \in \mathcal{N}(\mathbf{s})\})$ in Rosenblatt's [41] format. Given that a collection of concliques is available for a particular situation, the fundamental result that serves as the basis for our GOF procedures is as follows.

THEOREM 2.1. *Let the spatial process* $\{Y(\mathbf{s}) : \mathbf{s} \in \mathbb{Z}^d\}$ *have conditional distribution functions as in* (1.2), *and let* $\{\mathcal{C}_j : j = 1, \ldots, q\}$ *be a collection of concliques that partition the integer grid* $\mathbb{Z}^d$. *Then for any* $j = 1, \ldots, q$, *the variables* $\{U(\mathbf{s}) : \mathbf{s} \in \mathcal{C}_j\}$ *given by* (2.1) *are i.i.d. Uniform* (0, 1) *variables.*

Typically, the conditional c.d.f. $F(\cdot|\cdot)$ of expression (1.2) will be a parameterized function, and we now write this as $F_\theta(\cdot|\cdot)$ to emphasize the parametrization.

Let θ_0 denote the true value of the parameter. In an application we have available a set of observations taken to represent realizations of the random variables $\{Y(\mathbf{s}_i): i = 1, \ldots, N\}$. Theorem 2.1 indicates that if we compute generalized spatial residuals as, in the notation of (2.1),

$$U(\mathbf{s}_i) = \big(1 - A(\mathbf{s}_i)\big) \cdot F_{\theta_0}\big(y(\mathbf{s}_i) | \{y(\mathbf{t}) : \mathbf{t} \in \mathcal{N}(\mathbf{s}_i)\}\big) + A(\mathbf{s}_i) \cdot F^{-}_{\theta_0}\big(y(\mathbf{s}_i) | \{y(\mathbf{t}) : \mathbf{t} \in \mathcal{N}(\mathbf{s}_i)\}\big), \qquad \mathbf{s}_i \in \mathcal{C}_j, \tag{2.2}$$

then within any conclique $\mathcal{C}_j$ these variables should behave as a random sample from a uniform distribution on the unit interval. If we use a minimal conclique cover having q members, then we will have q sets of residuals, each of which should behave as a random sample from a uniform distribution. These sets of residuals will not, however, be independent, so we will not have a total collection that behaves as q independent random samples.

In practice we will usually also replace the parameter θ with an estimate $\hat{\theta}$ computed on the basis of the observations so that, technically, the values within any conclique will not actually be independent either. We expect, however, that if the model is appropriate, then these residuals will exhibit approximately the same behavior as independent uniform variates, in the same way that ordinary residuals from a linear regression model with normal errors behave as an approximate random sample of normal variates, despite the fact that they cannot technically represent such a sample.

A basic diagnostic plot can be constructed by plotting the empirical distribution function of each set of residuals $\{u(\mathbf{s}_i): \mathbf{s}_i \in \mathcal{C}_j\}$, $j = 1, \ldots, q$, and examining them for departures from a standard uniform distribution function. See, for instance, Gneiting et al. [19], Section 3.1, for a summary of graphical approaches for exploring uniformity in PIT values. Tests for uniformity may be used for individual sets of residuals to guide the decision about whether a given fitted model is adequate or to choose between two competing (even nonnested) models. Such procedures do not constitute a formal GOF test, however, because there is no guarantee that results will agree across differing sets of residuals in a conclique cover. Formal procedures for combining evidence from the residual sets into one overall GOF test are presented in the next section.

3. Methodology: Goodness of fit tests.

3.1. *General setting.* Suppose that for a set of locations on the d-dimensional integer lattice $\{\mathbf{s}_1, \ldots, \mathbf{s}_N\} \subset \mathbb{Z}^d$, we want to assess the GOF of a conditional model specification, based on a set of observed values $\{Y(\mathbf{s}_i): i = 1, \ldots, N\}$. We assume that the observed values are a partial realization of a class of process models defined on $\mathbb{Z}^d$ for which the conditional c.d.f. of $Y(\mathbf{s})$ given $\{Y(\mathbf{t}): \mathbf{t} \neq \mathbf{s}\}$ belongs to a class of parameterized conditional distribution functions,

$$\mathcal{F}_\theta = \big\{F_\theta\big(\cdot | \{Y(\mathbf{t}) : \mathbf{t} \in \mathcal{N}(\mathbf{s})\}\big) : \theta \in \Theta\big\}, \tag{3.1}$$

where $\Theta \subseteq \mathbb{R}^p$, $1 \leq p < \infty$, is a parameter space, $\mathcal{N}(\mathbf{s}) = \mathbf{s} + \mathcal{M}$ and, analogously to (1.2), $\mathcal{M} \subset \mathbb{Z}^d \setminus \{\mathbf{0}\}$. Two testing problems fit into this framework, where the null hypothesis is simple and where it is composite.

In the next subsections, we describe GOF tests for simple and composite hypotheses based on the observations $\{Y(\mathbf{s}_i) : i = 1, \dots, N\}$, which are assumed to have arisen in the following way. Suppose that $R \subset \mathbb{R}^d$ denotes a sampling region within which N observations are obtained at a set of sampling locations $\mathcal{S}_N \equiv R \cap \mathbb{Z}^d = \{\mathbf{s}_1, \dots, \mathbf{s}_N\}$. Define the interior of the set of sampling locations as $\mathcal{S}_N^{\mathrm{int}} \equiv \{\mathbf{s} \in \mathcal{S}_N : \mathcal{N}(\mathbf{s}) \subset \mathcal{S}_N\}$. Locations in this set are those sampling locations for which all neighbors are also sampling locations, allowing generalized spatial residuals to be computed for all $\mathbf{s} \in \mathcal{S}_N^{\mathrm{int}}$, even if the physical sampling region R is irregular. Finally, let $\mathcal{C}_{1N}, \dots, \mathcal{C}_{qN}$ denote the conclique partition of $\mathcal{S}_N^{\mathrm{int}}$ determined by $\mathcal{C}_{jN} = \mathcal{C}_j \cap \mathcal{S}_N^{\mathrm{int}}$, $j = 1, \dots, q$. In practice we will desire a minimal conclique cover but this is not necessary in what follows.

3.2. *Testing a simple null hypothesis*. First consider the case of the simple (S) null hypothesis in which the testing problem is given by, for some specified $\theta_0 \in \Theta$,

$H_0(S)$: The data $\{Y(\mathbf{s}_i) : i = 1, \dots, N\}$ represent a partial sample of the process model class (3.1) with $\theta = \theta_0$;

$H_1(S)$: Not $H_0(S)$.

To construct test statistics appropriate for these hypotheses, we consider the generalized spatial residuals under $H_0(S)$,

$$U(\mathbf{s}) = \big(1 - A(\mathbf{s})\big) \cdot F_{\theta_0}\big(Y(\mathbf{s}) | \{Y(\mathbf{t}) : \mathbf{t} \in \mathcal{N}(\mathbf{s})\}\big) + A(\mathbf{s}) \cdot F_{\theta_0}^{-}\big(Y(\mathbf{s}) | \{Y(\mathbf{t}) : \mathbf{t} \in \mathcal{N}(\mathbf{s})\}\big), \qquad \mathbf{s} \in \mathcal{S}_N^{\mathrm{int}}. \tag{3.2}$$

Now define, for $j = 1, \dots, q$, the (generalized residual) empirical distribution function over the jth conclique by

$$G_{jN}(u) = \frac{1}{|\mathcal{C}_{jN}|} \sum_{\mathbf{s} \in \mathcal{C}_{jN}} \mathbb{I}\big(U(\mathbf{s}) \leq u\big),$$

$u \in [0, 1]$. Here and in the following, $\mathbb{I}(A)$ denotes the indicator function of a statement A, where $\mathbb{I}(A) = 1$ if A is true and $\mathbb{I}(A) = 0$ otherwise. Note that under $H_0(S)$, $E\{G_{jN}(u)\} = u$, $u \in [0, 1]$, as a result of Theorem 2.1. Hence, to assess the GOF of the model over the jth conclique $\mathcal{C}_j$, we consider the scaled deviations of the empirical distribution function from the Uniform (0, 1) distribution,

$$W_{jN}(u) \equiv N^{1/2}\big(G_{jN}(u) - u\big), \qquad u \in [0, 1]. \tag{3.3}$$

A number of GOF test statistics for testing $H_0(S)$ may be obtained by combining the W_{jN}'s in different ways:

$$T_{1N} = \max_{j=1,\dots,q} \sup_{u\in[0,1]} |W_{jN}(u)|, \tag{3.4}$$

$$T_{2N} = \left(\frac{1}{q}\sum_{j=1}^{q}\Big[\sup_{u\in[0,1]} |W_{jN}(u)|\Big]^2\right)^{1/2}, \tag{3.5}$$

$$T_{3N} = \max_{j=1,\dots,q}\left(\int_0^1 |W_{jN}(u)|^r\,du\right)^{1/r}, \tag{3.6}$$

$$T_{4N} = \frac{1}{q}\sum_{j=1}^{q}\left(\int_0^1 |W_{jN}(u)|^r\,du\right)^{1/r}, \tag{3.7}$$

where $r \in [1, \infty)$ in (3.6) and (3.7). Note that T_{1N} and T_{2N} are obtained by combining conclique-wise Kolmogorov–Smirnov test statistics, while T_{3N} and T_{4N} are obtained by combining conclique-wise Cramér–von Mises test statistics. While our statistics are based exclusively on paired differences (e.g., $G_{jN}(u) - u$, $u \in [0, 1]$), other test statistics may be formulated to assess agreement between the empirical G_{jN} and Uniform(0, 1) distributions, such as GOF tests based on ϕ-divergences studied in [24]. In Section 4, we provide asymptotic distributions for the empirical processes (3.3), which may be an ingredient for determining limit distributions of statistics based on ϕ-divergences; cf. Theorem 3.1 [24].

3.3. *Testing a composite null hypothesis.* The composite (C) null hypothesis can be stated as

$H_0(C)$: The data $\{Y(\mathbf{s}_i) : i = 1, \dots, N\}$ represent a partial sample of some member of the process model class (3.1) for an unknown θ;

$H_1(C)$: Not $H_0(C)$.

Let $\hat\theta$ denote an estimator of θ based on $\{Y(\mathbf{s}_i) : i = 1, \dots, N\}$. Since θ is unknown, instead of the $U(\mathbf{s})$'s of (3.2), we work with an estimated version of the generalized spatial residuals,

$$\begin{aligned}\hat U(\mathbf{s}) &= \big(1 - A(\mathbf{s})\big)\cdot F_{\hat\theta}\big(Y(\mathbf{s})|\{Y(\mathbf{t}) : \mathbf{t} \in \mathcal{N}(\mathbf{s})\}\big)\\ &\quad + A(\mathbf{s})\cdot F_{\hat\theta}^{-}\big(Y(\mathbf{s})|\{Y(\mathbf{t}) : \mathbf{t} \in \mathcal{N}(\mathbf{s})\}\big), \qquad \mathbf{s} \in \mathcal{S}_N^{\text{int}},\end{aligned} \tag{3.8}$$

where, as before, $\mathcal{N}(\mathbf{s}) = \mathbf{s} + \mathcal{M}$. Note that if $\hat\theta$ is a reasonable estimator of θ and if $F_\theta(\cdot|\cdot)$ is a smooth function of θ, then the $\hat U(\mathbf{s})$'s of (3.8) are approximately distributed as Uniform (0, 1). This suggests that we can base tests of $H_0(C)$ versus $H_1(C)$ on the processes

$$\hat W_{jN}(u) \equiv N^{1/2}\big(\hat G_{jN}(u) - u\big), \qquad u \in [0, 1], \tag{3.9}$$

for $j = 1, \dots, q$, where

$$\hat{G}_{jN}(u) = \frac{1}{|\mathcal{C}_{jN}|} \sum_{\mathbf{s} \in \mathcal{C}_{jN}} \mathbb{I}(\hat{U}(\mathbf{s}) \le u), \qquad u \in [0, 1].$$

The test statistics for testing $H_0(C)$ versus $H_1(C)$ are now given by

$$\hat{T}_{1N}, \dots, \hat{T}_{4N}, \tag{3.10}$$

where $\hat{T}_{jN}$ is obtained by replacing W_{jN} in expressions (3.4)–(3.7) with $\hat{W}_{jN}$. In the next section, we describe the limit distributions of the test statistics under the null hypothesis.

4. Asymptotic distributional results.

4.1. *Basic concliques.* To formulate large sample distributional results for the GOF statistics, we shall assume that the concliques $\mathcal{C}_1, \dots, \mathcal{C}_q$ used for these statistics can be "built up" from unions of structurally more basic concliques, say $\mathcal{C}_1^*, \dots, \mathcal{C}_{q^*}^*$, $q^* \ge q$. For any given template $\mathcal{M} \subset \mathbb{Z}^d \setminus \{\mathbf{0}\}$ defining neighborhoods as $\mathcal{N}(\mathbf{s}) = \mathbf{s} + \mathcal{M}$, $\mathbf{s} \in \mathbb{Z}^d$, we suppose such concliques are constructed as follows.

Let $\mathbf{e}_i \in \mathbb{Z}^d$ denote a vector with 1 in the ith component and 0 elsewhere, and define $m_i \equiv \max\{|\mathbf{e}_i'\mathbf{s}| : \mathbf{s} \in \mathcal{M}\}$ as the maximal absolute value of ith component over integer vectors in the neighborhood template $\mathbf{s} \in \mathcal{M}$, $i = 1, \dots, d$. Define a collection of sublattices as

$$\mathcal{C}_j^* = \{\mathbf{a}_j + \Delta\mathbf{s} : \mathbf{s} \in \mathbb{Z}^d\}, \qquad j = 1, \dots, q^* \equiv \prod_{i=1}^{d} (m_i + 1), \tag{4.1}$$

where $\Delta = \operatorname{diag}(m_1 + 1, \dots, m_d + 1)$ is a positive diagonal matrix and

$$\mathbf{a}_j \in \mathcal{I} \equiv \{(a_1, \dots, a_d)' \in \mathbb{Z}^d : 0 \le a_i \le m_i, i = 1, \dots, d\},$$

where $\mathbf{a}_j \ne \mathbf{a}_k$ if $\mathcal{C}_j^* \ne \mathcal{C}_k^*$.

Proposition 4.1 shows that these sets provide a collection of "basic" concliques (or coding sets) since locations within the same sublattice $\mathcal{C}_j^*$ are separated by directional distances Δ that prohibit neighbors within $\mathcal{C}_j^*$. Additionally, the proposition gives a simple rule for merging basic concliques $\mathcal{C}_j^*$ to create larger concliques $\mathcal{C}_j$. In the following, write $\pm\mathcal{M} = \mathcal{M} \cup -\mathcal{M}$, and define $\|\mathbf{s}\|_\infty = \max_{1 \le i \le d} |s_i|$ for $\mathbf{s} = (s_1, \dots, s_d)' \in \mathbb{Z}^d$.

PROPOSITION 4.1. *Under the process assumptions of Theorem 2.1 and for any neighborhood specified by a finite subset* $\mathcal{M} \subset \mathbb{Z}^d \setminus \{\mathbf{0}\}$:

(a) *sets* $\mathcal{C}_1^*, \dots, \mathcal{C}_{q^*}^*$ *of form* (4.1) *are concliques that partition* $\mathbb{Z}^d$;

(b) *if* $\mathbf{a}_1, \ldots, \mathbf{a}_i, \mathbf{a}_{i+1} \in \mathcal{I}$, $i \geq 1$, *such that* $\mathcal{C} \equiv \bigcup_{j=1}^{i} \mathcal{C}_j^*$ *is a conclique*, *then* $\mathcal{C} \cup \mathcal{C}_{i+1}^*$ *is a conclique if and only if*

$$\mathbf{a}_j - \mathbf{a}_{i+1} + \Delta\mathbf{s} \notin \pm\mathcal{M} \qquad \textit{for all } \mathbf{s} \in \mathbb{Z}^d,\ \|\mathbf{s}\|_\infty \leq 1, \textit{ and any } j = 1, \ldots, i.$$

In addition to providing a systematic approach for building concliques, the purpose of this basic conclique representation is to allow the covariance structure of the limiting Gaussian process of the conclique-wise empirical processes [cf. (3.3)] to be written explicitly and to simplify the distributional results to follow (as basic concliques $\mathcal{C}_j^*$ above have a uniform structure and are translates of one another). With many Markov models on a regular lattice described by the neighborhoods in Besag [4] involving coding sets or "unilateral" structures, there is typically no loss of generality in building a collection of concliques $\mathcal{C}_1, \ldots, \mathcal{C}_q$ from such basic concliques. We illustrate Proposition 4.1 with some examples.

EXAMPLE 2.1 (Continued). Under the four-nearest neighbor structure in $\mathbb{Z}^2$, we have $\mathcal{M} = \{\pm(0,1)', \pm(1,0)'\} = \pm\mathcal{M}$, $m_1 = m_2 = 1$, $\Delta = \text{diag}(2,2)$ and $q^* = 4$, so there are four basic concliques $\{C_j^*\}_{j=1}^4$ determined by the vectors

$$\mathbf{a}_1 = (0,0)', \qquad \mathbf{a}_2 = (1,1)', \qquad \mathbf{a}_3 = (1,0)', \qquad \mathbf{a}_4 = (0,1)'.$$

Because $\mathbf{a}_2 - \mathbf{a}_1 + \Delta \cdot \mathbf{s} = (1,1)' + 2\mathbf{s} \notin \pm\mathcal{M}$ for any $\mathbf{s} \in \mathbb{Z}^2$, $\|\mathbf{s}\|_\infty \leq 1$, then $\mathcal{C}_1 \equiv \mathcal{C}_1^* \cup \mathcal{C}_2^*$ is a conclique, and, similarly, so is $\mathcal{C}_2 \equiv \mathcal{C}_3^* \cup \mathcal{C}_4^*$. Additionally, Proposition 4.1 shows also that $\mathcal{C}_1$, $\mathcal{C}_2$ cannot be further merged so that these represent the previously illustrated minimal conclique cover.

EXAMPLE 2.2 (Continued). Under the eight-nearest neighbor structure in $\mathbb{Z}^2$, we have that $\mathcal{M} = \{\pm(0,1)', \pm(1,0)', \pm(1,1)', \pm(1,-1)'\}$ and the basic concliques $\{\mathcal{C}_j^*\}_{j=1}^4$ are the same as in Example 2.1 and correspond to Besag's [4] coding sets. However, these basic concliques cannot be merged into larger concliques by Proposition 4.1 and hence match the minimal cover of four concliques as illustrated previously (i.e., $\mathcal{C}_j = \mathcal{C}_j^*$).

EXAMPLE 4.1. Under a "simple unilateral" neighbor $\mathcal{M} = \{(0,-1)', (-1,0)'\}$ in $\mathbb{Z}^2$ (cf. [4], Section 6.2), the basic concliques are again the same and Proposition 4.1 gives $\mathcal{C}_1 \equiv \mathcal{C}_1^* \cup \mathcal{C}_2^*$, $\mathcal{C}_2 \equiv \mathcal{C}_3^* \cup \mathcal{C}_4^*$ as a minimal conclique cover.

4.2. *Asymptotic framework*. We now consider a sequence of sampling regions R_n indexed by n. For studying the large sample properties of the proposed GOF statistics, we adopt an "increasing domain spatial asymptotic" structure [10], where the sampling region R_n becomes unbounded as $n \to \infty$. Let R_0 be an open connected subset of $(-1/2, 1/2]^d$ containing the origin. We regard R_0 as a "prototype" of the sampling region R_n. Let $\{\lambda_n\}_{n\geq 1}$ be a sequence of positive numbers such that $\lambda_n \to \infty$ as $n \to \infty$. We assume that the sampling region $R_n = \lambda_n R_0$ is obtained by "inflating" the set R_0 by the scaling factor λ_n (cf. [40]). Since the origin is assumed to lie in R_0, the shape of R_n remains the same for different values of n. To avoid pathological cases, we assume that for any sequence of real numbers $\{a_n\}_{n\geq 1}$ with $a_n \to 0+$ as $n \to \infty$, the number of cubes of the lattice $a_n\mathbb{Z}^d$ that intersect both R_0 and R_0^c is $O((a_n)^{-(d-1)})$ as $n \to \infty$. This implies that, as the sampling region grows, the number of observations near the boundary of R_n is of smaller order $O(N_n^{(d-1)/d})$ than the total number N_n of observations in R_n so that the volume of R_n, N_n and the number of interior locations are equivalent as $n \to \infty$. The boundary condition on R_0 holds for most regions R_n of practical interest, including common convex subsets of $\mathbb{R}^d$, such as rectangles and ellipsoids, as well as for many nonconvex star-shaped sets in $\mathbb{R}^d$. (Recall that a set $A \subset \mathbb{R}^d$ is called star-shaped if for any $x \in A$, the line segment joining x to the origin lies in A.) The latter class of sets may have a fairly irregular shape. See, for example, [38, 43] for more details.

We want to assess the GOF of the process model specification (1.2), under either the simple or composite hypothesis sets of Section 3. As described in Section 3.1, we suppose that the spatial process is observed at locations on the integer grid $\mathbb{Z}^d$ that fall in the sampling region R_n producing a set of sampling locations $\mathcal{S}_{N_n}$ (indexed by n). To simplify notation, we will use $\mathcal{S}_n$ rather than the more cumbersome $\mathcal{S}_{N_n}$ and $\mathcal{S}_n^{\text{int}}$ rather than $\mathcal{S}_{N_n}^{\text{int}}$. Similarly, we will use W_{jn} to denote the empirical distribution of generalized spatial residuals for the jth conclique under a simple hypothesis as given by (3.3) with $N = N_n$ and $T_{1n}, \dots, T_{4n}$, the corresponding test statistics of (3.4)–(3.7). Also, $\hat{W}_{jn}$, and $\hat{T}_{1n}, \dots, \hat{T}_{4n}$ will denote the quantities in (3.9) and (3.10) with $N = N_n$.

4.3. *Results for the simple testing problem.* For studying the asymptotic distribution of the test statistics $T_{1n}, \dots, T_{4n}$ under the null hypothesis $H_0(S)$, we shall make use of the following condition, which imposes the structure on the concliques described in Section 4.1.

Condition (C.1): Each conclique $\mathcal{C}_1, \dots, \mathcal{C}_q$ is union of basic concliques $\mathcal{C}_1^*, \dots, \mathcal{C}_{q^*}^*$ as in (4.1). Namely, for each $j = 1, \dots, q$, there exists $\mathcal{J}_j \subset \{1, \dots, q^* \equiv \det(\Delta)\}$ where $\mathcal{C}_j = \bigcup_{i \in \mathcal{J}_j} \mathcal{C}_i^*$ and the index sets $\{\mathcal{J}_j\}_{j=1}^q$ are disjoint.

The following result gives the asymptotic null distribution of conclique-wise empirical processes $\mathbf{W}_n = (W_{1n}, \dots, W_{qn})'$ based on the scaled and centered empirical distributions $W_{jn}(u)$, $u \in [0, 1]$, as in (3.3). Note that, while each individual empirical process W_{jn} can be expected to weakly converge to a Brownian bridge under $H_0(S)$ (cf. [3]), the limit law of $\mathbf{W}_n$ will not similarly be distribution-free due to the dependence in observations across concliques. In particular, the null model F_{θ_0} influences the asymptotic covariance structure of $\mathbf{W}_n$.

Let $\mathcal{L}_\infty^q$ denote the collection of bounded vector-valued functions $\mathbf{f} = (f_1, \dots, f_q)' : [0, 1] \to \mathbb{R}^q$ defined on the unit interval. Also, let $|B|$ denote the size of a finite set $B \subset \mathbb{R}$.

THEOREM 4.2. *Suppose that condition* (C.1) *holds*. *Then*, *there exists a zero-mean vector-Gaussian process* $\mathbf{W}(u) = (W_1(u), \dots, W_q(u))'$, $u \in [0, 1]$, *with continuous sample paths on* $[0, 1]$ (*with probability* 1) *such that*

$$\mathbf{W}_n \xrightarrow{d} \mathbf{W} \qquad \text{as } n \to \infty$$

as elements of $\mathcal{L}_\infty^q$. *Further,* $P(\mathbf{W}(u) = \mathbf{0}) = 1$ *for* $u = 0, 1$ *and the* $q \times q$ *covariance matrix function of* $\mathbf{W}$ *is given by*

$$EW_j(u)W_k(v) = \begin{cases} \dfrac{\det(\Delta)}{|\mathcal{J}_j|}(\min\{u, v\} - uv), & \text{if } j = k, \\ \dfrac{\det(\Delta)}{|\mathcal{J}_j| \cdot |\mathcal{J}_k|} \displaystyle\sum_{i \in \mathcal{J}_j, l \in \mathcal{J}_k} \sigma_{i,l}(u, v), & \text{if } j \neq k, \end{cases}$$

for $0 \le u, v \le 1$, $1 \le j, k \le q$ *and*

$$\sigma_{i,l}(u, v) \equiv \sum_{\mathbf{s} \in \mathbb{Z}^d, \|\mathbf{s}\|_\infty \le 1} \{P[U(\mathbf{0}) \le u, U(\mathbf{a}_l - \mathbf{a}_i + \Delta\mathbf{s}) \le v] - uv\} \times \mathbb{I}(\mathbf{a}_l - \mathbf{a}_i + \Delta\mathbf{s} \in \pm\mathcal{M}).$$

The indicator function $\mathbb{I}(\cdot)$ above pinpoints terms in the covariance expression which automatically vanish by the independence of residual variables $U(\mathbf{s})$ within conclique structures (Theorem 2.1). For example, when it is possible to combine two basic concliques $\mathcal{C}_i^*$ and $\mathcal{C}_l^*$, $i \neq l$, into a larger conclique, Proposition 4.1 gives that, for all $\|\mathbf{s}\|_\infty \le 1$, it holds that $\mathbf{a}_i - \mathbf{a}_l + \Delta\mathbf{s} \notin \mathcal{M}$ and so above $\mathbb{I}(\mathbf{a}_i - \mathbf{a}_l + \Delta\mathbf{s} \in \pm\mathcal{M}) = 0$. All sums in the limiting covariance structure then involve only a finite number of terms.

As a direct implication of Theorem 4.2, we get the following result on the asymptotic null distribution of the test statistics $T_{1n}, \dots, T_{4n}$.

COROLLARY 4.3. *Under the conditions of Theorem* 4.2,

$$T_{jn} \xrightarrow{d} \varphi_j(\mathbf{W}) \qquad \text{as } n \to \infty$$

for $j = 1, \dots, 4$, *where the functionals'* φ_j*'s are defined by*

$$\begin{aligned} \varphi_1(\mathbf{f}) &= \max_{1 \le j \le q} \sup_{u \in [0,1]} |f_j(u)|, \\ \varphi_2(\mathbf{f}) &= \left(\frac{1}{q} \sum_{1 \le j \le q} \Big[\sup_{u \in [0,1]} |f_j(u)|\Big]^2\right)^{1/2}, \\ \varphi_3(\mathbf{f}) &= \max_{1 \le j \le q} \left(\int_0^1 |f_j(u)|^r \, du\right)^{1/r}, \\ \varphi_4(\mathbf{f}) &= \frac{1}{q} \sum_{1 \le j \le q} \left(\int_0^1 |f_j(u)|^r \, du\right)^{1/r} \end{aligned} \tag{4.2}$$

for $\mathbf{f} = (f_1, \dots, f_q)' \in \mathcal{L}_\infty^q$, *and for a given* $r \in [1, \infty)$.

4.4. *Results for the composite testing problem.* As for the simple testing problem, here we first derive the asymptotic null distribution of the conclique-wise empirical processes $\hat{\mathbf{W}}_n = (\hat{W}_{1n}, \dots, \hat{W}_{qn})'$ based on scaled and centered empirical distributions $\hat{W}_{jn}(u)$, $u \in [0, 1]$, in (3.9).

Note that the estimator $\hat{\theta}_n$ appears in each summand in $\hat{W}_{jn}$ through the estimated generalized spatial residuals (3.8). In such situations, a common standard approach to deriving asymptotic distributions of empirical processes is based on the concept of *uniform asymptotic linearity* in some local neighborhood of the true parameter value θ_0 (cf. [36, 46]). However, this approach is *not* directly applicable here due to the form of the conditional distribution functions in (3.1) when considered as functions of $\theta \in \Theta$. To establish the limit distribution, we embed the empirical process of the estimated generalized residuals in an enlarged space, namely, the space of locally bounded q-dimensional vector functions on [0, 1], equipped with the metric of uniform convergence on compacts, and then use a version of the continuous mapping theorem; the argument details are provided in [32].

We require some notation and conditions in addition to those introduced in the earlier section. Letting again $|B|$ denote the size of a finite set B, define the strong mixing coefficient of the process $\{Y(\mathbf{s}) : \mathbf{s} \in \mathbb{Z}^d\}$ by

$$\alpha(a; b) = \sup\{|P(A \cap B) - P(A)P(B)| : A \in \mathcal{D}(S_1), B \in \mathcal{D}(S_2), \\ |S_1| \le b, |S_2| \le b, d(S_1, S_2) \ge a, S_1, S_2 \subset \mathbb{Z}^d\},$$

where $\mathcal{D}(S) = \sigma\langle Y(\mathbf{s}) : \mathbf{s} \in S\rangle$ generically denotes the σ-algebra generated by variables $Y(\mathbf{s})$ with locations in $S \subset \mathbb{Z}^d$, $d(S_1, S_2) = \inf\{\|\mathbf{s} - \mathbf{t}\|_1 : \mathbf{s} \in S_1, \mathbf{t} \in S_2\}$, $\|\mathbf{x}\|_1 = \sum_{i=1}^d |x_i|$ for $\mathbf{x} = (x_1, \dots, x_d)' \in \mathbb{R}^d$, and $P(\cdot)$ represents probabilities for the process. Write $F_\theta^{(1)}(\cdot|\cdot)$ and $F_\theta^{(1)-}(\cdot|\cdot)$ to denote $p \times 1$ vectors of first order partial derivatives of $F_\theta(\cdot|\cdot)$ and $F_\theta^-(\cdot|\cdot)$ with respect to θ, when these exist. Let $U_\theta(\mathbf{0}) = (1 - A(\mathbf{0})) \cdot F_\theta(Y(\mathbf{0})|\{Y(\mathbf{t}) : \mathbf{t} \in \mathcal{M}\}) + A(\mathbf{0}) \cdot F_\theta^-(Y(\mathbf{0})|\{Y(\mathbf{t}) : \mathbf{t} \in \mathcal{M}\})$, and denote $\mathbf{U}_\theta^{(1)}(\mathbf{0}) \in \mathbb{R}^p$ as the vector of partial derivatives of $U_\theta(\mathbf{0})$ with respect to θ, when this exists.

Condition (C.2):

(i) There exist constants $\delta_0 \in (0, 1)$, $c_0 \in (0, \infty)$ such that

$$|P(U_\theta(\mathbf{0}) \le u) - P(U_{\theta_0}(\mathbf{0}) \le v)| \le c_0[\|\theta - \theta_0\| + |u - v|]$$

for all $0 \le u, v \le 1$ and $\theta \in \Theta$ satisfying $\max\{\|\theta - \theta_0\|, |u - v|\} \le \delta_0$.

(ii) $\sup\{\|F_\theta^{(1)}(y|\mathbf{x})\| + \|F_\theta^{(1)-}(y|\mathbf{x})\| : \|\theta - \theta_0\| \le \delta_0, y \in \mathbb{R}, \mathbf{x} \in \mathbb{R}^p\} \le c_0$.

(iii) $E\{\sup_{\|\theta-\theta_0\|<\delta} \|\mathbf{U}_\theta^{(1)}(\mathbf{0}) - \mathbf{U}_{\theta_0}^{(1)}(\mathbf{0})\|\} = o(\delta)$ as $\delta \to 0$.

Condition (C.3): Suppose that the joint distribution of $(U_{\theta_0}(\mathbf{0}), \mathbf{U}_{\theta_0}^{(1)}(\mathbf{0}))$ is absolutely continuous with respect to $L \times \mu$ with Radon–Nikodym derivative $\tilde{f}(u, \mathbf{x})$, where L is the Lebesgue measure on $\mathbb{R}$, and μ is a σ-finite measure on $\mathbb{R}^p$. Suppose that

$$\lim_{t\to\infty} \sup_{u\in(0,1)} \int_{\|\mathbf{x}\|>t} \|\mathbf{x}\| \tilde{f}(u,\mathbf{x})\, d\mu(\mathbf{x}) = 0$$

and

$$\int \|\mathbf{x}\| \cdot \sup\{|\tilde{f}(u,\mathbf{x}) - \tilde{f}(v,\mathbf{x})| : |u-v| \le \delta\}\, d\mu(\mathbf{x}) \to 0$$

as $\delta \to 0+$.

Condition (C.4):

(i) There exist zero-mean random variables $\{\mathbf{V}(\mathbf{s}) : \mathbf{s} \in \mathbb{Z}^d\}$ such that

$$N_n^{1/2}(\hat{\theta}_n - \theta_0) = N_n^{-1/2} \sum_{\mathbf{s}\in\mathcal{S}_n} \mathbf{V}(\mathbf{s}) + o_p(1).$$

(ii) For each $\mathbf{s} \in \mathbb{Z}^d$, the variable $\mathbf{V}(\mathbf{s}) = (V_1(\mathbf{s}), \dots, V_p(\mathbf{s}))'$ is $\mathcal{D}(\mathbf{s} + \mathcal{M})$-measurable.

(iii) There exist $a \in (2,\infty)$, $\kappa > 0$ such that $\sup\{E\|\mathbf{V}(\mathbf{s})\|^{2+\kappa} : \mathbf{s} \in \mathbb{Z}^d\} < \infty$ and

$$\sum_{j=1}^{\infty} j^{d-1}\alpha(j;1)^{\kappa/(2+\kappa)} < \infty, \qquad \sum_{j=1}^{\infty} j^{d(2r-1)}\alpha(j;2r-1)^{1/a} < \infty$$

for some integer r satisfying $r > (p+1)/(1-a^{-1})$.

(iv) $\Sigma \equiv \lim_{n\to\infty} \mathrm{Var}(N_n^{-1/2} \sum_{\mathbf{s}\in\mathcal{S}_n} \mathbf{V}(\mathbf{s}))$ exists and is nonsingular.

Conditions (C.2) and (C.3) are exclusively used for handling the effects of the perturbation of the empirical process of the generalized residuals due the estimation of θ. The first displayed condition in (C.3) is an uniform integrability condition, while the second one is a continuity condition on the densities $\tilde{f}(\cdot,\cdot)$ (in u) in a weighted $L^1(\mu)$-norm. Without loss of generality, we shall suppose that $\tilde{f}(u,\mathbf{x}) = 0$ for all $u \notin (0,1)$ except on a set of $\mathbf{x}$-values with μ-measure zero. Condition (C.4) allows us to relate the limit law of the (unperturbed) empirical process part with the variability in estimating θ by $\hat{\theta}_n$. If the conditional model specification is such that the spatial process satisfies Dobrushin's uniqueness condition (cf. [20]), then the MRF is strongly mixing (actually, ϕ-mixing) at an exponential rate and, hence, mixing conditions in (C.4) trivially hold.

THEOREM 4.4. *Suppose that conditions* (C.1)–(C.4) *and the composite null hypothesis* $H_0(C)$ *hold. Then, there exist a zero-mean vector-Gaussian process* $\mathbf{W}(u) = (W_1(u), \dots, W_q(u))'$, $u \in [0,1]$, *with continuous sample paths on* $[0,1]$ (*with probability* 1) *and a random variable* $\mathbf{Z} = (Z_1, \dots, Z_p)' \sim N_p(\mathbf{0}, \Sigma)$, *both defined on a common probability space, such that as* $n \to \infty$,

$$\hat{\mathbf{W}}_n \xrightarrow{d} \mathbf{W} + \mathbf{1} \cdot \mathbf{Z}' \int \mathbf{x}\tilde{f}(\cdot,\mathbf{x})\, d\mu(\mathbf{x})$$

as elements of $\mathcal{L}_\infty^q$, *where* $\mathbf{1} = (1, \ldots, 1)' \in \mathbb{R}^q$. *The* $q \times q$ *covariance matrix function of* $\mathbf{W}$ *is as in Theorem* 4.2 *and for* $j = 1, \ldots, q$, $k = 1, \ldots, p$ *and* $u \in (0, 1)$,

$$EW_j(u)Z_k = \frac{1}{|\mathcal{J}_j|} \sum_{i \in \mathcal{J}_j} \sum_{\mathbf{s} \in \mathbb{Z}^d} E\big(V_k(\mathbf{s} - \mathbf{a}_i) \cdot \mathbb{I}\big(U(\mathbf{0}) \leq u\big)\big).$$

The following result is a direct consequence of Theorem 4.4 and gives the asymptotic distribution of the test statistics under the composite null $H_0(C)$.

COROLLARY 4.5. *Under the conditions of Theorem* 4.4,

$$\hat{T}_{jn} \xrightarrow{d} \varphi_j\left(\mathbf{W} + \mathbf{1} \cdot \mathbf{Z}' \int \mathbf{x} \tilde{f}(\cdot, \mathbf{x})\, d\mu(\mathbf{x})\right) \qquad \text{as } n \to \infty$$

for $j = 1, \ldots, 4$, *where the functionals* φ_j*'s are as defined in* (4.2).

Under the composite null $H_0(C)$, the limiting distributions involved are not distribution-free (i.e., depending on the true model c.d.f. F_{θ_0} in a complex covariance structure). Empirical processes based on PIT residuals with parameter estimates are known to exhibit this behavior in other inference scenarios with time series and independent data (cf. [17]), and often two general approaches are considered for implementing GOF tests [37]: resampling or Khmaladze's [33] martingale transformation. The latter involves a type of continuous de-trending to minimize effects of parameter estimation and has been applied to obtain asymptotically distribution-free tests with other model checks using residual empirical processes based on estimated parameters (cf. [34, 35]). In particular, Bai [3] justified this transformation for tests in parametric, conditionally specified (continuous) distributions for time series, but considered only one empirical process of residuals. If modified to the spatial setting, this result would entail a transformation of $\hat{W}_{jN}$ from one conclique $j = 1, \ldots, 1$ so that its limiting distribution is Brownian motion and distribution-free under $H_0(C)$. The complication here is that with residual empirical processes from multiple concliques, after applying a conclique-wise transformation, the resulting limit distribution of a test statistic under $H_0(C)$ would not be distribution-free due to dependence across concliques (akin to Theorem 4.2 in the case of no parameter estimation). Another option might be to use plug-in estimates of the covariance structure, using, for example, that asymptotic variances of maximum likelihood and pseudolikelihood estimators (i.e., Σ in Theorem 4.4) are known for some Markov field models [21]. But one would also have to estimate other complicated covariances in the limiting distribution of Theorem 4.4, which might be possible with subsampling variance estimation [43].

Spatial resampling methodologies, such as the block bootstrap (cf. [39], Chapter 12), might also be used to approximate sampling distributions of GOF statistics based on spatial residuals and knowledge of the limit distributions in Theorem 4.4 could be applied to toward justifying such bootstrap estimators. Simulations in

Section 5 also suggest that the finite sample versions of the GOF statistics appear to converge fairly quickly to their limits, at least in the case of simple null hypotheses. This implies that, in application, large-sample bootstrap approximations of finite-sample sampling distributions may be reasonable. The theoretical development of a spatial bootstrap for our GOF statistics is outside of the scope of this paper, but in Section 6 we use a parametric spatial bootstrap to calibrate GOF test statistics for a composite null hypothesis.

5. Numerical results. Here we provide a small numerical verification of the large sample distributional results in the simple null hypothesis case, considering observations generated from a conditional Gaussian MRF on $\mathbb{Z}^2$ with a four-nearest neighbor structure specified by $\mathcal{M} = \{\pm(0,1)', \pm(1,0)'\}$ as in Example 2.1. The conditional model family (3.1) of $Y(\mathbf{s})$ given $\{Y(\mathbf{t}) : \mathbf{t} \in \mathcal{N}(\mathbf{s})\}$, $\mathbf{s} \in \mathbb{Z}^2$ ($\mathcal{N}(\mathbf{s}) = \mathbf{s} + \mathcal{M}$), is normal with mean $\mu_{\alpha,\eta}(\mathbf{s}) \equiv \alpha + \eta \sum_{\mathbf{t} \in \mathcal{N}(\mathbf{s})}[Y(\mathbf{t}) - \alpha]$, and variance $\tau^2 > 0$, where $E(Y(\mathbf{s})) = \alpha \in \mathbb{R}$ is the marginal process mean and $|\eta| < 0.25$ denotes a dependence parameter. In total, the model parameters θ are $(\alpha, \tau, \eta)'$.

5.1. *Limit distributions under a simple null hypothesis.* We first examine asymptotic null distributions of GOF test statistics in the simple testing problem $H_0(S) : (\alpha, \tau, \eta)' = (\alpha_0, \tau_0, \eta_0)'$ of Section 3.2 with residuals (3.2) given by $U(\mathbf{s}) = \Phi[\{Y(\mathbf{s}) - \mu_{\alpha_0,\eta_0}(\mathbf{s})\}/\tau_0]$. Here $\Phi(\cdot)$ denotes the standard normal cumulative distribution function, and, for simplicity, we will write hypothesized parameters α_0, τ_0, η_0 as α, τ, η in the following.

As described in Section 3.1, the four-nearest-neighbor structure produces a minimal cover of two concliques $\mathcal{C}_1, \mathcal{C}_2$ (cf. Example 2.1), each of which is a union of two basic concliques $\mathcal{C}_1^*, \ldots, \mathcal{C}_4^*$ provided in Section 4.1. These concliques yield an empirical distribution process $\mathbf{W}_n = (W_{1n}, W_{2n})'$ and GOF test statistics $T_{1n}, \ldots, T_{4n}$ as in (3.4)–(3.7). By Theorem 4.2, $\mathbf{W}_n$ has a mean-zero Gaussian limit $\mathbf{W} = (W_1, W_2)'$ with covariances

$$EW_j(u)W_k(v) = \begin{cases} 2(\min\{u,v\} - uv), & \text{if } j = k, \\ 8\big[P\big(X_1 \le \Phi^{-1}(u), X_2 \le \Phi^{-1}(v)\big) - uv\big], & \text{if } j \ne k, \end{cases} \tag{5.1}$$

$u, v \in [0,1]$, $j, k \in \{1, 2\}$, where vectors (X_1, X_2) in (5.1) are bivariate normal, with marginally standard normal distributions and correlation $-\eta$. Hence, under the simple null hypothesis, the limit process depends on $(\alpha, \tau, \eta)'$ only through the dependence parameter η, which we denote by writing $\mathbf{W} \equiv \mathbf{W}_\eta$.

To understand the distribution of $\varphi_j(\mathbf{W}_\eta)$, $j = 1, 2, 3, 4$, as the asymptotic limit of GOF statistics T_{jn} under Corollary 4.3, we simulated from the theoretical Gaussian process $\mathbf{W}_\eta$ as follows. For each value of $\eta = 0, 0.1, 0.24$, we generated 50,000 sequences of mean-zero bivariate Gaussian variables $(W_1(i/3001), W_2(i/3001))$, $i = 0, \ldots, 3001$, with covariance structure (5.1) over a grid in $[0, 1]$;

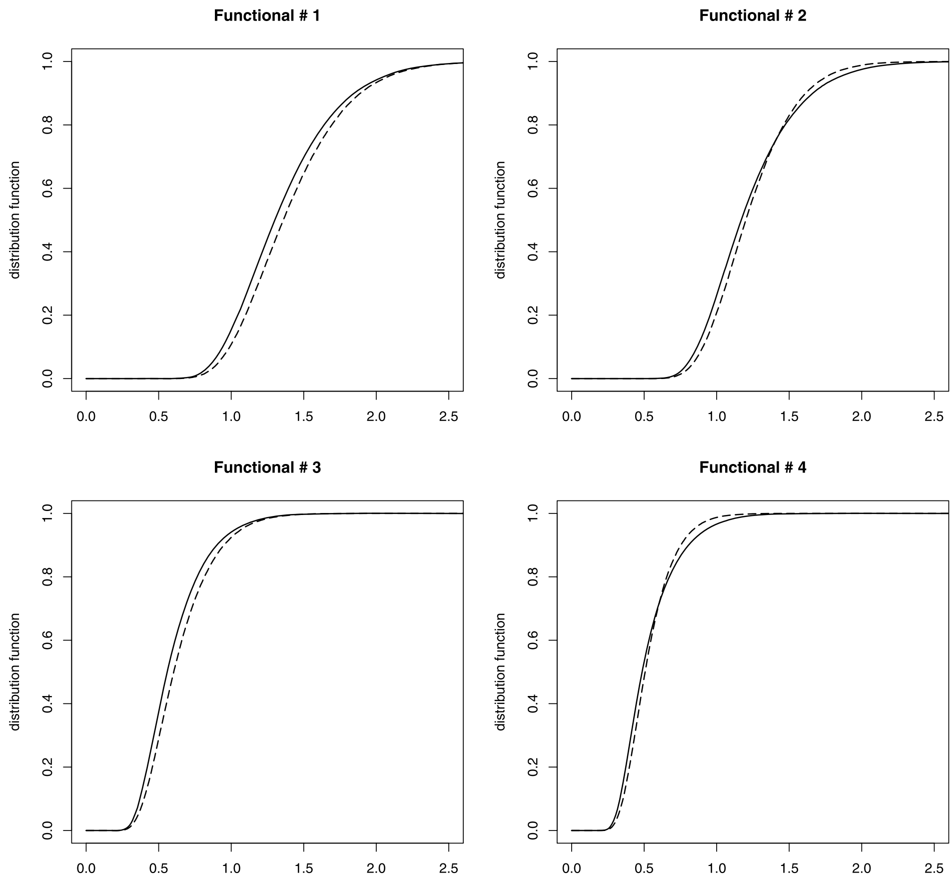

FIG. 1. *Cumulative distribution functions* $F_{\varphi_j(\mathbf{W}_\eta)}(w) = P[\varphi_j(\mathbf{W}_\eta) \le w]$, $w \in \mathbb{R}$, *for limit functionals* $\varphi_1(\mathbf{W}_\eta), \ldots, \varphi_4(\mathbf{W}_\eta)$ *for* $\eta = 0.1$ (*dashed*) *and* $\eta = 0.24$ (*solid*).

the sequence length of 3002 was dictated by computational stability. These provide approximate observations of $\mathbf{W}_\eta$, with η values chosen to reflect no, weak and strong forms of positive spatial dependence. Cumulative distribution functions of each functional $\varphi_1(\mathbf{W}_\eta), \ldots, \varphi_4(\mathbf{W}_\eta)$ were then approximated from $\mathbf{W}_\eta$-realizations. The resulting distribution curves appear in Figure 1 for $\eta = 0.1$ and $\eta = 0.24$, with $\varphi_3(\mathbf{W}_\eta)$ and $\varphi_4(\mathbf{W}_\eta)$ computed using $r = 2$ in (4.2).

5.2. *Comparisons to finite sample distributions.* To compare the agreement of finite sample distributions of T_{jN} under the simple null hypothesis with their limit distributions $\varphi_j(\mathbf{W}_\eta)$, $j = 1, \ldots, 4$, we simulated samples on two grid sizes, a 10×10 grid having $N = 100$ locations and a 30×30 grid having $N = 900$. Here, we simulated 50,000 realizations of conditional Gaussian samples (setting $\alpha = 0$ and $\tau = 1$ with no loss of generality) and evaluated functionals $T_{1N}, \ldots, T_{4N}$ to ap-

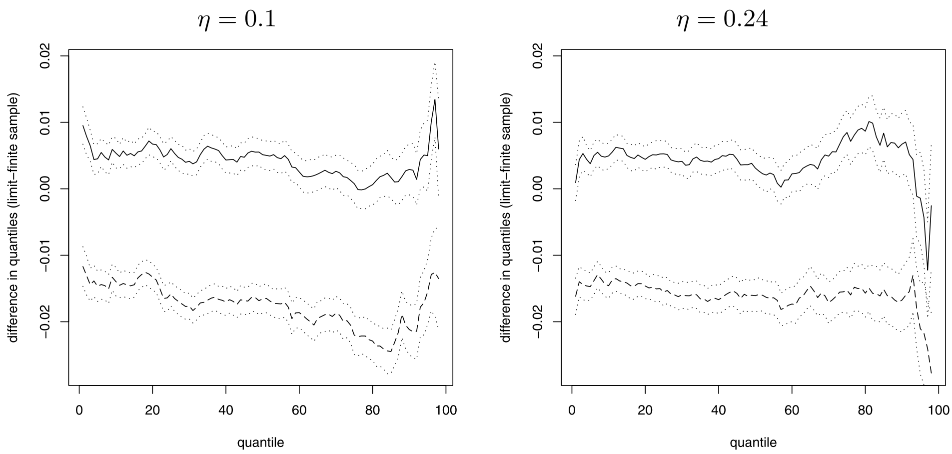

FIG. 2. *Difference in quantiles for* $\varphi_2(\mathbf{W}_\eta)$ *and* T_{2N} *when* $N = 100$ (*dashed line*) *and* 900 (*solid line*) *for* $\eta = 0.1, 0.24$. *Pointwise* 95% *confidence bands* (*dotted*) *indicate the Monte Carlo error in each difference*.

proximate the finite-sample distributions of these GOF statistics. Figure 2 shows the difference between the quantiles of the limit $\varphi_2(\mathbf{W}_\eta)$ and those of T_{2N} for $\eta = 0.1$ and $\eta = 0.24$; the agreement among quantiles for functional 2 is quite good even though this plot was one exhibiting the largest quantile-mismatches among the four GOF functionals. Table 1 shows the proportion of GOF statistics T_{jN} falling above the 95th and 99th quantiles of the corresponding limit $\varphi_j(\mathbf{W}_\eta)$ distribution, $j = 1, 2, 3, 4$. The agreement between the finite-sample and theoretical limit distributions is again close in Table 1.

For various sample sizes and dependence parameters, Table 2 compares the finite-sample distributions of the four GOF statistics $\{T_{jN}\}_{j=1}^4$ against their limit-

TABLE 1

Proportion of GOF statistics T_{jN} *from a conditional Gaussian model falling above the* 95*th and* 99*th quantiles* (*denoted* $q_{95,\eta}$ *and* $q_{99,\eta}$) *of the their limit* $\varphi_j(\mathbf{W}_\eta)$ *distribution*, $j = 1, 2, 3, 4$, *for sample sizes* $N = 100$ *and* $N = 900$ *and with dependence parameters* $\eta = 0, 0.1, 0.24$

		% of $T_{jN} > q_{95,\eta}$				**% of $T_{jN} > q_{99,\eta}$**			
η	**N**	**$j = 1$**	**2**	**3**	**4**	**$j = 1$**	**2**	**3**	**4**
0	100	4.67	4.38	5.09	4.97	0.90	0.90	0.98	0.91
0	900	5.11	4.91	4.95	4.86	1.03	1.09	1.03	1.03
0.1	100	4.60	4.60	4.88	4.92	0.94	0.95	0.95	1.08
0.1	900	5.11	5.13	5.05	5.08	1.08	1.13	1.07	1.15
0.24	100	4.52	4.57	5.02	5.06	0.80	0.76	0.86	0.92
0.24	900	4.92	4.97	4.86	5.03	0.97	0.96	0.93	0.95

TABLE 2

Computed values ($\times 1000$) *from distance metrics comparing finite-sample distributions of statistics* $T_{1N}, \dots, T_{4N}$ *to their limiting distributions* $\varphi_1(\mathbf{W}_\eta), \dots, \varphi_4(\mathbf{W}_\eta)$

		$D_{\mathrm{KS}}(\varphi_j(\mathbf{W}_\eta), T_{jN})$				$D_{\mathrm{CM}}(\varphi_j(\mathbf{W}_\eta), T_{jN})$			
η	N	$j=1$	2	3	4	$j=1$	2	3	4
0	100	19.6	23.0	8.9	9.1	12.9	14.8	3.7	3.2
0	900	6.7	9.7	3.4	4.8	3.8	4.7	1.0	1.4
0.1	100	24.0	27.7	5.3	5.4	16.5	18.2	2.2	1.7
0.1	900	9.9	10.0	6.5	5.7	5.3	4.7	2.5	1.8
0.24	100	21.6	25.2	7.2	7.0	14.7	15.6	2.7	2.2
0.24	900	9.1	8.2	4.2	3.8	4.8	4.8	1.4	1.4
		$D_{\mathrm{KS}}(\varphi_j(\mathbf{W}_{\eta_1}), \varphi_j(\mathbf{W}_{\eta_2}))$				$D_{\mathrm{CM}}(\varphi_j(\mathbf{W}_{\eta_1}), \varphi_j(\mathbf{W}_{\eta_2}))$			
η_1	η_2	$j=1$	2	3	4	$j=1$	2	3	4
0	0.1	14.4	11.9	14.9	13.4	7.0	5.7	7.7	6.1
0.1	0.24	70.0	59.6	90.3	72.9	49.9	35.0	50.4	33.1
0	0.24	81.8	69.1	102.3	84.3	56.6	40.5	58.0	38.9

ing distributions $\varphi_j(\mathbf{W})$ in terms of a Kolmogorov–Smirnov D_{KS} and a Cramér–von Mises-like D_{CM} distance metric, defined by

$$D_{\mathrm{KS}}(X, Z) \equiv \sup_{t \in \mathbb{R}} |F_X(t) - F_Z(t)|,$$

$$D_{\mathrm{CM}}(X, Z) \equiv \left[\int |F_X(t) - F_Z(t)|^2 \, dt\right]^{1/2},$$

relative to the cumulative distributions F_X, F_Z of arbitrary random variables X, Z. To interpret the relative values of these metrics in assessing the distributional distance between T_{jN} and $\varphi_j(\mathbf{W}_\eta)$, it is helpful to reference $D_{\mathrm{KS}}, D_{\mathrm{CM}}$ values for comparing the distributions of $\varphi_j(\mathbf{W}_{\eta_1})$ and $\varphi_j(\mathbf{W}_{\eta_2})$ over parameters $\eta_1 \neq \eta_2$, which Table 2 also provides. Generally, the convergence of the finite-sample distributions T_{jN} to their limits $\varphi_j(\mathbf{W}_\eta)$ appears to occur fairly uniformly over different dependence parameters η and, relative to the distributional differences among different limits [e.g., $\varphi_j(\mathbf{W}_{\eta_1})$ and $\varphi_j(\mathbf{W}_{\eta_2})$], the agreement in distributions of T_{jN} and $\varphi_j(\mathbf{W}_\eta)$ is quite close even for samples of size 100.

5.3. *Power of GOF statistics under simple null hypothesis.* Under the simple null $H_0(S): (\alpha, \tau, \eta)' = (0, 1, 0)'$, we next consider the power of GOF tests based on statistics $T_{1N}, \dots, T_{4N}$ computed from conditional Gaussian data generated with $\eta = 0.1$ and $\eta = 0.24$ and $\alpha = 0$, $\tau = 1$. This gives an idea of the power in testing a hypothesis of no spatial dependence, when the data exhibit forms of positive dependence, both fairly weak ($\eta = 0.1$) and strong ($\eta = 0.24$). For a given

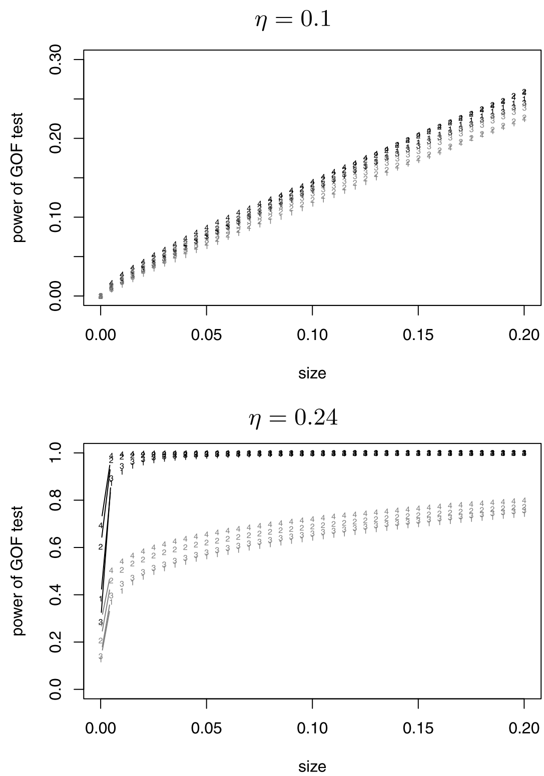

FIG. 3. *Plots of power versus size γ for GOF tests of $H_0 : \eta = 0$ in conditional Gaussian models* (*fixed* $\alpha = 0, \tau = 1$) *based on functionals* $T_{1N}, \dots, T_{4N}$, *determined by data generated under* $\eta = 0.1, 0.24$. *In these power versus size curves*, *each functional is numbered* 1–4 *under sample sizes* $N = 100$ (*grey*) *and* $N = 900$ (*black*).

GOF statistic T_{jN} from a sample of size $N = 100$ or $N = 900$, a size γ test is conducted by rejecting H_0 if T_{jN} exceeds the $1 - \gamma$ quantile of the limit distribution $\varphi(\mathbf{W}_{\eta=0})$ under the null hypothesis. Figure 3 plots power versus size γ for these tests when $\eta = 0.1$ and $\eta = 0.24$, based on 50,000 simulated data sets. Power is low under the alternative $\eta = 0.1$, as might be expected, but considerably higher when $\eta = 0.24$. Tests with functionals T_{2N}, T_{4N} (based conclique-wise averages of GOF statistics) tend to perform similarly and exhibit slightly more power than tests with T_{1N}, T_{3N} (based conclique-wise maxima of GOF statistics).

6. An application to agricultural field trials.

6.1. *The problem.* Besag and Higdon [5] present an analysis of six agricultural field trials of corn varieties conducted in North Carolina using a hierarchical model that included an intrinsic Gaussian MRF as an improper prior for spatial structure. An intrinsic Gaussian MRF results from fixing dependence parameters at the boundary of the parameter space. In discussion of this paper, Smith [44] raised the question of what diagnostics were available to examine potential evidence for spatial structure based on the available data, and presented variograms of three of the trials. Kaiser and Caragea [28] used data from these same three trials to illustrate a model-based diagnostic they called the S-value. Questions about the spatial structure suggested by the data included the possibilities of nonstationarity and directional dependencies. Here, we use data from all six trials to examine the question of whether a simple model with constant mean and unidirectional dependence can be rejected as a plausible representation of spatial structure. Our question is simply one of whether a basic Gaussian MRF with constant mean and a single dependence parameter could be rejected as a possible data generating mechanism for the data, not whether it might be be most preferred model available.

Each field trial consisted of observations of yield from 64 corn varieties with each variety replicated 3 times in each trial. The spatial layout of each trial was essentially that of a 11×18 regular lattice, although the last column of that lattice contained only 5 locations. After subtracting variety by trial means in the same manner as [28, 44], we deleted the last column to obtain a rectangular 11×17 lattice containing 187 observations for each trial. We assumed a four-nearest-neighborhood structure but without use of a border strip, so that locations had a variable number of neighboring observations, 4 for each of the 135 interior locations, 3 for each of the 48 edge locations, and 2 for each of the four corner locations.

6.2. *The model.* Although each trial should nominally have marginal mean zero, to examine a full composite setting we fit a model with conditional Gaussian distributions having expected values $\{\mu(\mathbf{s}_i) : i = 1, \dots, n\}$ and constant conditional variance τ^2 where, with N_i denoting the neighborhood of location $\mathbf{s}_i$; $i = 1, \dots, n$,

$$\mu(\mathbf{s}_i) = \alpha + \eta \sum_{\mathbf{s}_j \in N_i} \{y(\mathbf{s}_j) - \alpha\}. \tag{6.1}$$

The joint distribution of this model is then Gaussian with marginal means $\boldsymbol{\alpha}$ an n-vector with each element equal to α and covariance matrix $(I - C)^{-1}M$ where I is the $n \times n$ identity matrix, M is an $n \times n$ diagonal matrix with all nonzero entries equal to τ^2 and $C = \eta H$ with H an $n \times n$ matrix having element (i, j) equal to 1 if locations $\mathbf{s}_i$ and $\mathbf{s}_j$ are neighbors and 0 otherwise. With this structure, the parameter space of η can be determined to be $(-0.2563, 0.2563)$ based on eigenvalues of H (cf. [10]); this differs slightly from the parameter space for a lattice with four-nearest-neighborhood structure wrapped on a torus due to the size of the lattice and the use of varying numbers of neighbors for edge locations.

6.3. *The GOF procedure.* The model of expression (6.1) was fit to (centered) data from each of the six trials using maximum likelihood estimation. Generalized spatial residuals were computed for each of the two concliques, one having 93

TABLE 3
Estimates for conditional Gaussian models fit to data from six agricultural field trials; *the point estimates for α for all trials differ from zero by at most* 10^{-15}

	Point		Interval		
Trial	τ^2	η	α	τ^2	η
1	95.56	0.2526	(−10.21, 10.40)	(79.43, 119.54)	(0.2107, 0.2544)
2	156.90	0.1855	(−3.19, 3.42)	(125.96, 190.08)	(0.0922, 0.2257)
3	128.94	0.2476	(−7.66, 7.76)	(105.63, 159.54)	(0.1976, 0.2533)
4	129.92	0.2095	(−3.57, 3.74)	(104.54, 159.76)	(0.1264, 0.2380)
5	69.33	0.2522	(−8.29, 8.23)	(57.20, 86.44)	(0.2091, 0.2543)
6	210.75	0.2542	(−20.57, 19.68)	(175.39, 268.45)	(0.2136, 0.2549)

and the other 94 locations. Using the fitted models, a parametric bootstrap procedure was used to arrive at p-values for each of the four test statistics introduced as $\hat{T}_{jN}$; $j = 1, \dots, 4$, in Section 3.3. For each fitted model (i.e., trial) 5000 bootstrap data sets were simulated using a Gibbs algorithm with a burn-in of 500 and spacing of 10, which appeared adequate to result in convergence of the chain based on scale reduction factors [18] and eliminate dependence between successive data sets based on autocorrelations. Model (6.1) was fit, generalized spatial residuals produced and the four test statistics computed for each bootstrap data set, from which p-values were taken as the proportion of simulated test statistic values greater than those from the actual data sets. Bootstrap data sets were also used to produce percentile bootstrap intervals for parameters (cf. [14]). Percentile intervals were chosen because basic bootstrap intervals extended beyond the parameter space for η for each of the six trials.

6.4. *Results*. Results of estimation are presented in Table 3. Intervals were computed at the 95% level and values for η are reported to four decimal places because estimates tended to be close to the upper boundary of the parameter space (0.2563). Overall, estimation was fairly similar for these six trials, which were conducted in different counties of North Carolina, including an indication of high variability in estimating these parameters, particularly α and τ^2. Estimates of η indicate moderate to strong spatial structure in each of the six trials, and estimates of τ^2 indicate substantial local variability despite this structure.

GOF p-values resulting from the parametric bootstrap procedure of Section 6.3 are presented in Table 4 for each of the four test statistics of Section 3.3. Overall these values provide no indication that we are able to dismiss model (6.1) as a plausible representation of the spatial structure present in these data.

7. Conclusions. In this article we have introduced a practical method to assess the aptness of Markov random field models for representing spatial processes.

TABLE 4
Parametric bootstrap p-values for the six agricultural field trials

Trial	T_1	T_2	T_3	T_4
1	0.8348	0.7976	0.7086	0.7530
2	0.3844	0.4182	0.2132	0.3262
3	0.0852	0.1168	0.1506	0.1478
4	0.1656	0.1084	0.1426	0.0972
5	0.2162	0.1828	0.1754	0.2024
6	0.3502	0.2382	0.4642	0.2984

This method is based on special sets of locations we have called concliques that partition the total set of observed locations such that generalized spatial residuals within each conclique approximate realizations of independent random variables on the unit interval. These generalized spatial residuals can be combined across nonindependent concliques in natural ways to produce GOF statistics that correspond to Gaussian empirical processes that have identifiable limit distributions. While those limit distributions can involve complex covariance structures, we have demonstrated that finite sample versions of the GOF statistics appear to converge rather quickly to their limits, at least in the case of a simple null hypothesis. This implies that, in an application, approximation of their limit distributions under a suitable null hypothesis will provide a useful reference distribution against which to compare the value of an observed GOF statistic. The composite hypothesis setting introduces a considerably more complicated situation than does the simple hypothesis setting, because limit laws involve covariances that cannot be easily determined either explicitly or numerically. In an application, resampling methods would seem to hold the greatest promise for approximating distributions of GOF statistics based on generalized spatial residuals. While developing spatial subsampling or block bootstraps (cf. [39], Chapter 12) for this purpose requires further investigation, the use of such resampling was illustrated in this article in the application to agricultural field trials.

We wish to comment on a number of issues that involve the distinction between application of the GOF methodology developed and the production of theoretical results for that methodology. First is the issue of stationarity. There is nothing in the definition or construction of generalized spatial residuals, or GOF statistics constructed from them, that requires a stationary model. All that is needed is identification of a full conditional distribution for each location (1.1) that may then be used in (2.1), and assurance that a joint distribution having these conditionals exists. Assumptions of stationarity made in this article facilitate the production of theoretical results needed to justify use of the methodology. Another issue is application to discrete cases. While the data examples given have considered continuous conditional models, we have applied random generalized spatial residuals to models formed from Winsorized Poisson conditional distributions [29] with promising empirical results. Similar to questions of stationarity and discrete cases, there is nothing in the constructive methodology that requires a regular lattice or that each location have the same number of neighbors. Use of a regular lattice in this article again facilitates the demonstration of theoretical properties, but this is not needed to implement the procedures suggested. The application of Section 6 involved a regular lattice, but no border strip or other boundary conditions were imposed to render neighborhoods of equal size. It should certainly be anticipated that there may be edge effects on GOF statistics as developed here, just as there are edge effects on properties of estimators. How severe these effects might be in various settings, and whether the use of modified boundary conditions (e.g., [6]) could mitigate such effects is an issue in need of additional investigation. Essentially the same thoughts can be offered relative to potential sparseness that might occur in an application. Locations lacking neighbors entirely could be considered members of any conclique one chooses, and construction of GOF statistics would proceed unhindered. What the effects of varying degrees of sparseness are remains an open question. Of course, if no locations have any neighbors, then the methodology presented here reduces to statistics constructed on the basis of the ordinary probability integral transform for independent random variables.

As with all GOF tests, the procedure based on generalized spatial residuals developed in this article serves as a vehicle for assessing a selected model for overall adequacy, not as a vehicle for selection of the most attractive model in the first place. This is important in consideration of fitted models under the composite setting, in which we can think of estimation as having "optimized" a given model structure for description of a set of observed data. There may be two or more such structures that could be, with the best choice of parameter values possible, viewed as plausible data generating mechanisms for a set of observations. This does not necessarily mean, however, that those different structures are equally pleasing as models for the problem under consideration.

Finally, we mention a connection with the assessment of k-step ahead forecasts in a time series setting. Let $\{X_t; t \in \mathbb{Z}\}$ denote a series of random variables observed at discrete, equally spaced, points in time. The probability integral transform with distributions conditioned on the present and past has been used to construct k-step ahead residuals $U_{t+k} = F_{t+k|t}(X_{t+k})$, where the conditioning in F is on $\{X_t, X_{t-1}, \ldots\}$ (e.g., [15, 16, 19]). While our use of the probability integral transform is similar to what is done in this time series setting, the conditioning requirements are quite distinct. In the spatial setting, two spatial residuals are independent only if neither is in the conditioning set of the other (i.e., are both in the same conclique). In time series k-step ahead forecasts, two values, U_i and U_j, will be independent only if either X_i *is* in the conditioning set of X_j, or vice versa. The difference stems from the use of full conditionals in spatial Markov random field models, rather than the sequential conditionals in the time series context which need not invoke a Markov property at all. The approach taken to the development of theoretical results in this article could potentially be used in the time series setting, but the modifications require further investigation.

8. Proof of generalized spatial residual properties. As Theorem 2.1 provides the main distributional result for generalized spatial residuals (2.1) from concliques, which are fundamental to the GOF test statistics of Sections 3 and 4, we establish Theorem 2.1 here. The proofs of other results from the manuscript are provided in supplementary material [32].

Let $\mathcal{Q}$ denote a finite subset of conclique $\mathcal{C} \subset \mathbb{Z}^d$ with $|\mathcal{Q}| = l \geq 2$, and let $\mathcal{I}_\mathcal{Q} = \bigcup_{\mathbf{s}\in\mathcal{Q}}\{\mathbf{i} : \mathbf{i} \in \mathcal{N}(\mathbf{s})\} = \{\mathbf{s}_1, \dots, \mathbf{s}_L\}$, $L \geq 1$, be the finite index set of all neighbors of sites in $\mathcal{Q}$; additionally, enumerate the l elements of $\mathcal{Q}$ as $\mathcal{Q} = \{\mathbf{s}_{1+L}, \dots, \mathbf{s}_{l+L}\}$, say. With respect to the enumeration of $\mathcal{I}_\mathcal{Q}$ and $\mathcal{Q}$, let $F_1(\cdot)$ denote the marginal c.d.f. of $Y(\mathbf{s}_1)$, and let $F_j(\cdot)$, $2 \leq j \leq L + l$, denote the conditional c.d.f. of $Y(\mathbf{s}_j)$ given $Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_{j-1})$; define the function $F_j^-(\cdot)$ by the left limits of $F_j(\cdot)$. By the randomized PIT [7], $\{(1 - A(\mathbf{s}_j)) \cdot F_j[Y(\mathbf{s}_j)] + A(\mathbf{s}_j) \cdot F_j^-[Y(\mathbf{s}_j)] : j = 1, \dots, L + l\}$ are i.i.d. Uniform (0, 1) random variables.

For any $i, k \in \{L + 1, \dots, L + l\}$, variables $Y(\mathbf{s}_i)$ and $Y(\mathbf{s}_k)$ belong to the conclique $\mathcal{Q}$ so that all neighbors of $Y(\mathbf{s}_i)$ and $Y(\mathbf{s}_k)$ are among $\{Y(\mathbf{s}_j)\}_{j=1}^L$. By the Markov property (1.2), $F_j[Y(\mathbf{s}_j)] = F[Y(\mathbf{s}_j)|\{Y(\mathbf{s}) : \mathbf{s} \in \mathcal{N}(\mathbf{s}_j)\})]$ holds and we may equivalently write (2.1) as $U(\mathbf{s}_j) = (1 - A(\mathbf{s}_j)) \cdot F_j[Y(\mathbf{s}_j)] + A(\mathbf{s}_j) \cdot F_j^-[Y(\mathbf{s}_j)]$ for any $j \in \{L + 1, \dots, L + l\}$, though these relationships may not necessarily hold for $j = 1, \dots, L$. Hence, $\{U(\mathbf{s}) : \mathbf{s} \in \mathcal{Q}\}$ are i.i.d. Uniform (0, 1) variables for any arbitrary finite subset $\mathcal{Q}$ of $\mathcal{C}$.

Acknowledgments. The authors are very grateful to an Associate Editor and three referees for thoughtful comments and suggestions which significantly clarified and improved the manuscript.

SUPPLEMENTARY MATERIAL

Proofs of main results for spatial GOF test statistics (DOI: 10.1214/11-AOS948SUPP; .pdf). A supplement [32] provides proofs of all asymptotic distributional results from Section 4, regarding the conclique-based spatial GOF test statistics in simple and composite null hypothesis settings (Proposition 4.1, Theorem 4.2, Corollary 4.3, Theorem 4.4, Corollary 4.5). The proof in the composite hypothesis case is particularly nonstandard; see Section 4.4.

REFERENCES

[1] ANDERSON, T. W. (1993). Goodness of fit tests for spectral distributions. *Ann. Statist.* **21** 830–847. MR1232521

[2] ARNOLD, B. C., CASTILLO, E. and SARABIA, J. M. (1992). *Conditionally Specified Distributions. Lecture Notes in Statistics* **73**. Springer, Berlin. MR1177398

[3] BAI, J. (2003). Testing parametric conditional distributions of dynamic models. *Rev. Econom. Statist.* **85** 531–549.

[4] BESAG, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. Roy. Statist. Soc. Ser. B* **36** 192–236. MR0373208

[5] BESAG, J. and HIGDON, D. (1999). Bayesian analysis of agricultural field experiments. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **61** 691–746. MR1722238

[6] BESAG, J. and KOOPERBERG, C. (1995). On conditional and intrinsic autoregressions. *Biometrika* **82** 733–746. MR1380811

[7] BROCKWELL, A. E. (2007). Universal residuals: A multivariate transformation. *Statist. Probab. Lett.* **77** 1473–1478. MR2395595

[8] CARAGEA, P. C. and KAISER, M. S. (2009). Autologistic models with interpretable parameters. *J. Agric. Biol. Environ. Stat.* **14** 281–300. MR2750841

[9] COX, D. R. and SNELL, E. J. (1971). On test statistics calculated from residuals. *Biometrika* **58** 589–594. MR0312618

[10] CRESSIE, N. A. C. (1993). *Statistics for Spatial Data*, 2nd ed. Wiley, New York.

[11] CSISZÁR, I. and TALATA, Z. (2006). Consistent estimation of the basic neighborhood of Markov random fields. *Ann. Statist.* **34** 123–145. MR2275237

[12] CZADO, C., GNEITING, T. and HELD, L. (2009). Predictive model assessment for count data. *Biometrics* **65** 1254–1261. MR2756513

[13] DARLING, D. A. (1957). The Kolmogorov–Smirnov, Cramér–von Mises tests. *Ann. Math. Statist.* **28** 823–838. MR0093870

[14] DAVISON, A. C. and HINKLEY, D. V. (1997). *Bootstrap Methods and Their Application. Cambridge Series in Statistical and Probabilistic Mathematics* **1**. Cambridge Univ. Press, Cambridge. MR1478673

[15] DAWID, A. P. (1984). Statistical theory. The prequential approach. *J. Roy. Statist. Soc. Ser. A* **147** 278–292. MR0763811

[16] DIEBOLD, F. X., GUNTHER, T. A. and TAY, A. S. (1998). Evaluating density forecasts with applications to financial risk management. *Internat. Econom. Rev.* **39** 863–883.

[17] DURBIN, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. *Ann. Statist.* **1** 279–290. MR0359131

[18] GELMAN, A. and RUBIN, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statist. Sci.* **7** 457–511.

[19] GNEITING, T., BALABDAOUI, F. and RAFTERY, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **69** 243–268. MR2325275

[20] GUYON, X. (1995). *Random Fields on a Network*. Springer, New York. MR1344683

[21] GUYON, X. and YAO, J.-F. (1999). On the underfitting and overfitting sets of models chosen by order selection criteria. *J. Multivariate Anal.* **70** 221–249. MR1711522

[22] HAMMERSLEY, J. M. and CLIFFORD, P. (1971). Markov fields on finite graphs and lattices. Unpublished manuscript.

[23] HARDOUIN, C. and YAO, J.-F. (2008). Multi-parameter auto-models with applications to cooperative systems and analysis of mixed state data. *Biometrika* **95** 335–349.

[24] JAGER, L. and WELLNER, J. A. (2007). Goodness-of-fit tests via phi-divergences. *Ann. Statist.* **35** 2018–2053. MR2363962

[25] JENSEN, T. R. and TOFT, B. (1995). *Graph Coloring Problems*. Wiley, New York. MR1304254

[26] JI, C. and SEYMOUR, L. (1996). A consistent model selection procedure for Markov random fields based on penalized pseudolikelihood. *Ann. Appl. Probab.* **6** 423–443. MR1398052

[27] JUSTEL, A., PEÑA, D. and ZAMAR, R. (1997). A multivariate Kolmogorov–Smirnov test of goodness of fit. *Statist. Probab. Lett.* **35** 251–259. MR1484961

[28] KAISER, M. S. and CARAGEA, P. C. (2009). Exploring dependence with data on spatial lattices. *Biometrics* **65** 857–865. MR2649858

[29] KAISER, M. S. and CRESSIE, N. (1997). Modeling Poisson variables with positive spatial dependence. *Statist. Probab. Lett.* **35** 423–432. MR1483030

[30] KAISER, M. S. and CRESSIE, N. (2000). The construction of multivariate distributions from Markov random fields. *J. Multivariate Anal.* **73** 199–220. MR1763324

[31] KAISER, M. S., CRESSIE, N. and LEE, J. (2002). Spatial mixture models based on exponential family conditional distributions. *Statist. Sinica* **12** 449–474. MR1902719
[32] KAISER, M. S., LAHIRI, S. N. and NORDMAN, D. J. (2011). Supplement to "Goodness of fit tests for a class of Markov random field models." DOI:10.1214/11-AOS948SUPP.
[33] KHMALADZE, È. V. (1981). A martingale approach in the theory of goodness-of-fit tests. *Theory Probab. Appl.* **26** 240–257.
[34] KHMALADZE, È. V. (1993). Goodness of fit problem and scanning innovation martingales. *Ann. Statist.* **21** 798–829. MR1232520
[35] KHMALADZE, E. V. and KOUL, H. L. (2004). Martingale transforms goodness-of-fit tests in regression models. *Ann. Statist.* **32** 995–1034. MR2065196
[36] KOUL, H. L. (1970). A class of ADF tests for subhypothesis in the multiple linear regression. *Ann. Math. Statist.* **41** 1273–1281. MR0270507
[37] KOUL, H. L. and SAKHANENKO, L. (2005). Goodness-of-fit testing in regression: A finite sample comparison of bootstrap methodology and Khmaladze transformation. *Statist. Probab. Lett.* **74** 290–302. MR2189468
[38] LAHIRI, S. N. (1999). Asymptotic distribution of the empirical spatial cumulative distribution function predictor and prediction bands based on a subsampling method. *Probab. Theory Related Fields* **114** 55–84. MR1697139
[39] LAHIRI, S. N. (2003). Central limit theorems for weighted sums of a spatial process under a class of stochastic and fixed designs. *Sankhyā Ser. A* **65** 356–388. MR2028905
[40] LAHIRI, S. N., KAISER, M. S., CRESSIE, N. and HSU, N.-J. (1999). Prediction of spatial cumulative distribution functions using subsampling. *J. Amer. Statist. Assoc.* **94** 86–110. With comments and a rejoinder by the authors. MR1689216
[41] ROSENBLATT, M. (1952). Remarks on a multivariate transformation. *Ann. Math. Statist.* **23** 470–472. MR0049525
[42] RUE, H. and HELD, L. (2005). *Gaussian Markov Random Fields: Theory and Applications. Monographs on Statistics and Applied Probability* **104**. Chapman and Hall/CRC, Boca Raton, FL. MR2130347
[43] SHERMAN, M. and CARLSTEIN, E. (1994). Nonparametric estimation of the moments of a general statistic computed from spatial data. *J. Amer. Statist. Assoc.* **89** 496–500. MR1294075
[44] SMITH, R. L. (1999). Discussion of "Bayesian analysis of agricultural field experiments," by J. Besag and D. Higdon. *J. Roy. Statist. Soc. Ser. B* **61** 725–727.
[45] SPEED, T. P. (1978). Relations between models for spatial data, contingency tables and Markov fields on graphs. *Suppl. Adv. Appl. Probab.* **10** 111–122.
[46] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York. MR1385671

M. S. KAISER
D. J. NORDMAN
DEPARTMENT OF STATISTICS
IOWA STATE UNIVERSITY
AMES, IOWA 5001
USA
E-MAIL: mskaiser@iastate.edu
dnordman@iastate.edu

S. N. LAHIRI
DEPARTMENT OF STATISTICS
TEXAS A&M UNIVERSITY
COLLEGE STATION, TEXAS 77843
USA
E-MAIL: snlahiri@stat.tamu.edu