Bayesian Geostatistical Modeling for Discrete-Valued Processes

Xiaotian Zheng, Athanasios Kottas, and Bruno Sansó
Department of Statistics, University of California Santa Cruz

March 2, 2022

Abstract

We introduce a flexible and scalable class of Bayesian geostatistical models for discrete data, based on the class of nearest neighbor mixture transition distribution processes (NNMP), referred to as discrete NNMP. The proposed class characterizes spatial variability by a weighted combination of first-order conditional probability mass functions (pmfs) for each one of a given number of neighbors. The approach supports flexible modeling for multivariate dependence through specification of general bivariate discrete distributions that define the conditional pmfs. Moreover, the discrete NNMP allows for construction of models given a pre-specified family of marginal distributions that can vary in space, facilitating covariate inclusion. In particular, we develop a modeling and inferential framework for copula-based NNMPs that can attain flexible dependence structures, motivating the use of bivariate copula families for spatial processes. Compared to the traditional class of spatial generalized linear mixed models, where spatial dependence is introduced through a transformation of response means, our process-based modeling approach provides both computational and inferential advantages. We illustrate the benefits with synthetic data examples and an analysis of North American Breeding Bird Survey data.

Bayesian hierarchical models; Copula functions; Count data; Mixture transition distribution; Nearest neighbors; Spatial classification.

1 Introduction

Discrete geostatistical data arise in many areas, such as biology, ecology, and forestry. Such data sets consist of observations that take discrete values and are indexed in a continuous spatial domain. As an example, consider observations for counts of a species of interest, commonly used to estimate the species distribution over a geographical domain.

The most common approach to modeling such data is through a spatial generalized linear mixed model (SGLMM, Diggle et al. (1998)), under which an exponential family distribution is specified for the response at a given location, assuming independence between locations, conditional on an underlying spatial process. Such process is specified in the second stage of the SGLMM through a link function that associates the response mean to a set of spatial random effects. A Gaussian process is typically used for the spatial random effects. Thus, SGLMMs provide a general modeling tool for geostatistical discrete data applications (Wikle, 2002; Recta et al., 2012; Zhang et al., 2020).

However, SGLMMs have a number of drawbacks. First, they do not correspond to spatial processes for the observed data. Since the spatial random effects are incorporated into the transformed mean, SGLMMs model spatial structure on a function of the response means, not the observations directly. Thus, the model may impose a strong correlation between means over locations that are close, even though the corresponding observations may not be strongly correlated. In addition, the SGLMM specification poses computational challenges. Unlike Gaussian geostatistical models, the spatial random effects cannot be marginalized out. Under simulation-based inference, estimating the spatial random effects generally requires sampling a large number of highly correlated parameters within a Markov Chain Monte Carlo (MCMC) algorithm, which is likely to produce slow convergence, and a large memory footprint. Although efficient computational strategies have been explored in the literature (Christensen and Waagepetersen, 2002; Christensen et al., 2006; Sengupta and Cressie, 2013), the computational challenge is unavoidable, especially for large spatial datasets.

An alternative to SGLMMs involves Gaussian copula models which construct random fields given a pre-specified family of marginal distributions. Here, the joint cumulative

distribution function (cdf) of the spatial responses is characterized by a Gaussian copula corresponding to an underlying Gaussian process; see, e.g., Madsen (2009), Kazianka and Pilz (2010), and Han and De Oliveira (2016). Gaussian copulas provide simplicity in specifying spatial dependence, and flexibility in selecting discrete marginal distributions. However, the evaluation of the resulting likelihood requires efficient approximations of high-dimensional multivariate Gaussian integrals, limiting the applicability of this class of models.

In this paper, we introduce a new class of spatial process models for discrete geostatistical data. This class builds from the nearest-neighbor mixture transition distribution process (NNMP), proposed by Zheng et al. (2021) for modeling large continuous geostatistical data. The NNMP structured mixture formulation is motivated by mixture transition distribution (MTD) models for non-Gaussian time series (Le et al., 1996). In particular, Zheng et al. (2022) discuss construction of stationary MTDs for both continuous and discrete time series, using particular bivariate distributions from the literature.

The contribution of this paper is threefold. First, we develop a discrete analogue of the NNMP, referred to as the discrete NNMP, with particular focus on using bivariate copulas to define the spatially varying conditional probability mass functions (pmfs) for the structured mixture that gives rise to the joint distribution. We show that the joint pmf of the discrete copula NNMP can be further decomposed into a collection of bivariate copulas, providing interpretability for model construction using different families of copulas. In fact, our approach allows for the use of arbitrary bivariate copula families, which enhances model flexibility and enables the description of complex spatial dependencies. We demonstrate with a simulation study the impact of using different copula families, exploring alternatives to the traditional Gaussian copula for spatial modeling. Secondly, we extend the first-order strict stationarity result in Zheng et al. (2021). The extension is key for discrete NNMPs, providing a constructive approach to develop models with spatially varying marginal pmfs. This can be used, for example, to incorporate either continuous or discrete covariates, which is practically important in the context of regression modeling for discrete-valued spatial responses. Finally, utilizing the stationarity extension result, we

develop a Bayesian hierarchical framework that consists of using uniform random variables to transform discrete variables into continuous ones. The proposed approach leverages the properties of copulas for continuous random vectors, thus facilitating the use of different copulas as well as efficient computation. We show through a simulation study that, compared with popular SGLMM methods, this approach yields reliable posterior inference at a much lower computational cost.

The paper is organized as follows. In Section 2, we introduce NNMPs for discrete data, with copula-based discrete NNMPs developed in Section 3. Section 4 presents the Bayesian model formulation for inference, validation and prediction, followed by illustration with synthetic and real datasets in Section 5. Finally, Section 6 concludes with a summary and discussion.

2   NNMPs for discrete data

2.1   Modeling framework

Consider a univariate spatial process $Y(\boldsymbol{v})$ indexed by $\boldsymbol{v} \in \mathcal{D} \subset \mathbb{R}^p$, for $p \geq 1$. Let $\boldsymbol{y}_{\mathcal{S}} = (y(\boldsymbol{s}_1), \ldots, y(\boldsymbol{s}_n))^\top$ be a realization of the process $Y(\boldsymbol{v})$, where $\mathcal{S} = (\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n)$ denotes the reference set. Using a directed acyclic graph (DAG) with vertexes given by $y(\boldsymbol{s}_i)$ for the locations in $\mathcal{S}$, the joint density $p(\boldsymbol{y}_{\mathcal{S}})$ can be expressed as:

$$p(\boldsymbol{y}_{\mathcal{S}}) = p(y(\boldsymbol{s}_1)) \prod_{i=2}^{n} p(y(\boldsymbol{s}_i) \mid y(\boldsymbol{s}_{i-1}), \ldots, y(\boldsymbol{s}_1)), \tag{1}$$

where the conditional distributions depend on the set of parents of each vertex in the DAG.

Reducing the size of the conditioning set to be at most L, we obtain a valid joint density for $\boldsymbol{y}_{\mathcal{S}}$ that approximates (1) as

$$\tilde{p}(\boldsymbol{y}_{\mathcal{S}}) = p(y(\boldsymbol{s}_1)) \prod_{i=2}^{n} p(y(\boldsymbol{s}_i) \mid \boldsymbol{y}_{\mathrm{Ne}(\boldsymbol{s}_i)}), \tag{2}$$

where $\mathrm{Ne}(\boldsymbol{s}_i)$ is a subset of $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_{i-1}\}$, and $\boldsymbol{y}_{\mathrm{Ne}(\boldsymbol{s}_i)}$ is the vector formed by stacking

the process realization over $\text{Ne}(\boldsymbol{s}_i)$. Traditionally, the elements of $\text{Ne}(\boldsymbol{s}_i)$ are selected as the nearest neighbors of $\boldsymbol{s}_i$ within $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_{i-1}\}$, for $i = 2, \ldots, n$, according to a specified distance in $\mathcal{D}$. Ordering the elements of $\text{Ne}(\boldsymbol{s}_i)$ in ascending order with respect to distance to $\boldsymbol{s}_i$, we have $\text{Ne}(\boldsymbol{s}_i) = (\boldsymbol{s}_{(i1)}, \ldots, \boldsymbol{s}_{(i,i_L)})$, where $i_L := (i-1) \wedge L$. The joint density in (2) constructed using nearest neighbors has been explored for fast likelihood approximations (Vecchia, 1988; Katzfuss and Guinness, 2021), and extended to nearest-neighbor Gaussian process models for Gaussian data (Datta et al., 2016), and to NNMPs for continuous, non-Gaussian data (Zheng et al., 2021). We note that the factorization in (1) implicitly requires a topological ordering on the locations as they are not naturally ordered. Effects of the ordering on the approximation have been studied in the literature. Here, we adopt a random ordering, which is shown to give sharper approximation than coordinate-based orderings (Guinness, 2018).

Here, we introduce NNMPs for discrete-valued spatial processes, referred to as discrete NNMPs. Such models are derived by the following two steps. The first step consists of building a valid joint density over $\mathcal{S}$ by modeling the conditional densities in the product of the right hand side of (2) with a weighted combination of conditional pmfs:

$$p(y(\boldsymbol{s}_i) \,|\, \boldsymbol{y}_{\text{Ne}(\boldsymbol{s}_i)}) = \sum_{l=1}^{i_L} w_l(\boldsymbol{s}_i) \, f_{\boldsymbol{s}_i, l}(y(\boldsymbol{s}_i) \,|\, y(\boldsymbol{s}_{(il)})), \tag{3}$$

where $w_l(\boldsymbol{s}_i) \geq 0$ for every $\boldsymbol{s}_i \in \mathcal{S}$ and for all l, and $\sum_{l=1}^{i_L} w_l(\boldsymbol{s}_i) = 1$.

There are two model elements in (3) that describe spatial variability: the mixture component pmfs $f_{\boldsymbol{s}_i, l}$, and the weights $w_l(\boldsymbol{s}_i)$. We defer the specification of the pmfs $f_{\boldsymbol{s}_i, l}$ to the next section. Following Zheng et al. (2021), we define the weights as increments of a logit Gaussian cdf $G_{\boldsymbol{s}_i}$, i.e., $w_l(\boldsymbol{s}_i) = G_{\boldsymbol{s}_i}(r_{\boldsymbol{s}_i, l}) - G_{\boldsymbol{s}_i}(r_{\boldsymbol{s}_i, l-1})$, for $l = 1, \ldots, i_L$. Here, $0 = r_{\boldsymbol{s}_i, 0} < r_{\boldsymbol{s}_i, 1} < \cdots < r_{\boldsymbol{s}_i, i_L - 1} < r_{\boldsymbol{s}_i, i_L} = 1$ are random cutoff points such that $r_{\boldsymbol{s}_i, l} - r_{\boldsymbol{s}_i, l-1} = k'(\boldsymbol{s}_i, \boldsymbol{s}_{(il)}) / \sum_{l=1}^{i_L} k'(\boldsymbol{s}_i, \boldsymbol{s}_{(il)})$, for some bounded kernel $k' : \mathcal{D} \times \mathcal{D} \to [0, 1]$. Convenient choices for k' are kernels that compute the correlation between two points. The underlying Gaussian distribution for $G_{\boldsymbol{s}_i}$ has mean $\mu(\boldsymbol{s}_i) = \gamma_0 + \gamma_1 s_{i1} + \gamma_2 s_{i2}$, and variance κ^2, with $\boldsymbol{s}_i = (s_{i1}, s_{i2})$ where s_{i1} and s_{i2} correspond to the $x-$ and $y-$ coordinates of location $\boldsymbol{s}_i$.

This formulation allows for spatial dependence among the weights through $\mu(\boldsymbol{s}_i)$. Also, the random cutoff points can flexibly reflect the neighbor structure of $\boldsymbol{s}_i$.

The second step completes the construction of a valid stochastic process over $\mathcal{D}$ by extending (3) to an arbitrary finite set of locations outside $\mathcal{S}$, denoted as $\mathcal{U} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r)$, where $\mathcal{U} \subset \mathcal{D} \setminus \mathcal{S}$. In particular, we define the pmf of $\boldsymbol{y}_{\mathcal{U}}$ conditional on $\boldsymbol{y}_{\mathcal{S}}$ as

$$\tilde{p}(\boldsymbol{y}_{\mathcal{U}} \mid \boldsymbol{y}_{\mathcal{S}}) = \prod_{i=1}^{r} p(y(\boldsymbol{u}_i) \mid \boldsymbol{y}_{\mathrm{Ne}(\boldsymbol{u}_i)}) = \prod_{i=1}^{r} \sum_{l=1}^{L} w_l(\boldsymbol{u}_i)\, f_{\boldsymbol{u}_i,l}(y(\boldsymbol{u}_i) \mid y(\boldsymbol{u}_{(il)})), \qquad (4)$$

where the weights and conditional pmfs are defined analogously to Equation (3), and the points $(\boldsymbol{u}_{(i1)}, \ldots, \boldsymbol{u}_{(iL)})$ in $\mathrm{Ne}(\boldsymbol{u}_i)$ are the first L locations in $\mathcal{S}$ that are closest to $\boldsymbol{u}_i$.

In fact, given (3) and (4), a discrete-valued spatial process over $\mathcal{D}$ is well defined, based on the definition of nearest-neighbor processes (Datta et al., 2016). For any finite set $\mathcal{V} \subset \mathcal{D}$ that is not a subset of $\mathcal{S}$, the joint pmf over $\mathcal{V}$ is obtained by marginalizing $\tilde{p}(\boldsymbol{y}_{\mathcal{U}} \mid \boldsymbol{y}_{\mathcal{S}})\tilde{p}(\boldsymbol{y}_{\mathcal{S}})$ over $\boldsymbol{y}_{\mathcal{S} \setminus \mathcal{V}}$, where $\mathcal{U} = \mathcal{V} \setminus \mathcal{S}$.

We note that the model involves the neighborhood size L in both (3) and (4). Our prior model for the spatially varying weights supports the strategy of using an over-specified L that gives a large neighbor set, with important neighbors assigned large weights a posteriori. For specific data examples, a sensitivity analysis for L can be further carried out to find an optimal L according to standard model comparison metrics or scoring rules. This is illustrated with the real data application; see Section 5.3 and the supplementary material.

Practically, Equations (3) and (4) serve different purposes. The reference set $\mathcal{S}$ is often reserved for observed data, so model estimation is based on (3), while spatial prediction at new locations outside the reference set relies on (4). Henceforth, we use

$$p(y(\boldsymbol{v}) \mid \boldsymbol{y}_{\mathrm{Ne}(\boldsymbol{v})}) = \sum_{l=1}^{L} w_l(\boldsymbol{v})\, f_{\boldsymbol{v},l}(y(\boldsymbol{v}) \mid y(\boldsymbol{v}_{(l)})) \qquad (5)$$

to characterize discrete NNMPs, where $\boldsymbol{v}$ is a generic location in $\mathcal{D}$. The neighbor set $\mathrm{Ne}(\boldsymbol{v})$ contains the first L locations in $\mathcal{S}$ that are closest to $\boldsymbol{v}$. We place these locations in ascending order according to distance, denoted as $\mathrm{Ne}(\boldsymbol{v}) = (\boldsymbol{v}_{(1)}, \ldots, \boldsymbol{v}_{(L)})$.

The discrete NNMP formulation implies two distinct features that set it apart from

6

SGLMMs. In a SGLMM, responses $y(\boldsymbol{v})$ are conditionally independent with distribution $f(y(\boldsymbol{v}) \mid z(\boldsymbol{v}), \boldsymbol{\beta}, r) = a(y(\boldsymbol{v}), r) \exp\left(r\{y(\boldsymbol{v})\eta(\boldsymbol{v}) - \psi(\eta(\boldsymbol{v}))\}\right)$, where $z(\boldsymbol{v})$ is a spatial random effect, $\boldsymbol{\beta}$ are regression parameters, r is a dispersion parameter, and $h(\eta(\boldsymbol{v})) = \boldsymbol{x}(\boldsymbol{v})^{\top}\boldsymbol{\beta}+z(\boldsymbol{v})$ for some link function h. The joint distribution of observations $(y(\boldsymbol{s}_1), \ldots, y(\boldsymbol{s}_n))$ involves integrating out the spatial random effects, i.e., $\int \{\prod_{i=1}^{n} f(y(\boldsymbol{s}_i) \mid z(\boldsymbol{s}_i), \boldsymbol{\beta}, r)\}p(\boldsymbol{z}_{\mathcal{S}})d\boldsymbol{z}_{\mathcal{S}}$, where $\boldsymbol{z}_{\mathcal{S}} = (z(\boldsymbol{s}_1), \ldots, z(\boldsymbol{s}_n))^{\top}$. This restricts the choice of $z(\boldsymbol{v})$ to stochastic processes for which the corresponding joint densities are easy to work with, limiting the range of spatial variability the SGLMM can describe over the domain. In practice, $z(\boldsymbol{v})$ is commonly assumed to be a Gaussian process. This limitation, however, does not affect discrete NNMPs, as the spatial dependence is introduced at the data level. The joint pmf of a discrete NNMP is fully specified through (3) and (4), which is a finite mixture of generic spatial components that can flexibly capture spatial variability. In addition, the mixture model structure of discrete NNMPs allows for efficient implementation, using inference approaches for mixtures.

2.2 Model construction with spatially varying marginals

The key ingredient in constructing discrete NNMPs lies in the specification of the mixture component conditional pmfs $f_{\boldsymbol{v},l}$. There are many avenues to specify $f_{\boldsymbol{v},l}$. As each conditional pmf corresponds to a bivariate random vector, say $(U_{\boldsymbol{v},l}, V_{\boldsymbol{v},l})$, our strategy is to model $f_{\boldsymbol{v},l}$ through its bivariate pmf, denoted as $f_{U_{\boldsymbol{v},l}, V_{\boldsymbol{v},l}}$. Let $f_{U_{\boldsymbol{v},l}}$ and $f_{V_{\boldsymbol{v},l}}$ be the marginal pmfs of $(U_{\boldsymbol{v},l}, V_{\boldsymbol{v},l})$, such that $f_{\boldsymbol{v},l} \equiv f_{U_{\boldsymbol{v},l}|V_{\boldsymbol{v},l}} = f_{U_{\boldsymbol{v},l}, V_{\boldsymbol{v},l}}/f_{V_{\boldsymbol{v},l}}$. The benefits of this strategy are twofold. First, it simplifies the multivariate dependence specification by focusing on the bivariate random vectors $(U_{\boldsymbol{v},l}, V_{\boldsymbol{v},l})$. The multivariate dependence will be induced by bivariate distributions through the model's mixture formulation. Second, the strategy allows for the construction of models with a pre-specified family of marginal distributions, facilitating the study of local variability. For example, it is common in discrete geostatistical data modeling to include covariates through the (transformed) mean of the marginal distribution.

The second feature of this strategy relies on an extension of the first-order strict sta-

tionarity result from Zheng et al. (2021). Based on that result, an NNMP has stationary marginal pmf f_Y if $f_{U_{\boldsymbol{v},l}} = f_{V_{\boldsymbol{v},l}} = f_Y$, for all $\boldsymbol{v}$ and all l. Here, we generalize the result such that discrete NNMPs can be built from pre-specified spatially varying marginal pmfs $g_{\boldsymbol{v}}$, where $g_{\boldsymbol{v}}$ is the marginal pmf of $Y(\boldsymbol{v})$. The generalization of the stationarity proposition applies to all NNMPs. For the interest of this paper, we summarize the result in the following proposition for discrete NNMPs.

Proposition 1. *Consider a discrete NNMP model for spatial process $\{Y(\boldsymbol{v}) : \boldsymbol{v} \in \mathcal{D}\}$, and a collection of spatially varying pmfs $\{g_{\boldsymbol{v}} : \boldsymbol{v} \in \mathcal{D}\}$. If, for each $\boldsymbol{v}$, the marginal pmfs of the mixture component bivariate distributions are such that $f_{U_{\boldsymbol{v},l}} = g_{\boldsymbol{v}}$ and $f_{V_{\boldsymbol{v},l}} = g_{\boldsymbol{v}_{(l)}}$, the discrete NNMP has marginal pmf $g_{\boldsymbol{v}}$ for $Y(\boldsymbol{v})$, for every $\boldsymbol{v} \in \mathcal{D}$.*

A natural example for $\{g_{\boldsymbol{v}} : \boldsymbol{v} \in \mathcal{D}\}$ is a family of distributions with (at least) one of its parameters indexed in space, i.e., $g_{\boldsymbol{v}}(\cdot) = g(\cdot \mid \theta(\boldsymbol{v}), \boldsymbol{\xi})$, in particular, through spatially varying covariates. Using a link function for $\theta(\boldsymbol{v})$, we can include such covariates that provide additional spatially referenced information. A more general example involves partitioning the domain into several regions, where in each region, $g_{\boldsymbol{v}}$ is associated with a different family of marginal distributions. A relevant application is estimation of the abundance of a species that shows overdispersion in most areas, but underdispersion in areas where the species is less prevalent (Wu et al., 2015). Overall, Proposition 1 provides flexibility for construction of discrete-valued spatial models with specific marginal pmfs.

We develop next a key component of the methodology, that is, discrete copula NNMP model construction and inference. Given a family of marginal pmfs $g_{\boldsymbol{v}}$, we create spatial copulas for random vectors $(U_{\boldsymbol{v},l}, V_{\boldsymbol{v},l})$. We begin with copulas for a set of base random vectors (U_l, V_l), and extend them to be spatially dependent by modeling the copula parameter that controls the dependence structure as spatially varying. Together with Proposition 1, this strategy allows for construction of discrete NNMPs with marginal pmfs in general families.

8

3   Discrete copula NNMPs

3.1   Copula functions

A bivariate copula function $C : [0,1]^2 \to [0,1]$ is a distribution function whose marginals are uniform distributions on $[0,1]$. Following Sklar (1959), given a random vector (Z_1, Z_2) with joint probability distribution F and marginals F_1 and F_2, there exists a copula function C such that $F(z_1, z_2) = C(F_1(z_1), F_2(z_2))$. If F_1 and F_2 are continuous, C is unique. In this case, the copula density is $c(z_1, z_2) = \partial C(F_1(z_1), F_2(z_2))/(\partial F_1 \partial F_2)$, and the joint density is $f(z_1, z_2) = c(z_1, z_2)f_1(z_1)f_2(z_2)$, where f_1 and f_2 are the densities of F_1 and F_2, respectively.

If both marginals are discrete, the copula C is only unique on the set $\mathrm{Ran}(F_1) \times \mathrm{Ran}(F_2)$, where $\mathrm{Ran}(F_j)$ consists of all possible values of F_j, $j = 1, 2$ (Joe, 2014). Nevertheless, if C is a copula and F_1 and F_2 are discrete distribution functions, then $F(z_1, z_2) = C(F_1(z_1), F_2(z_2))$ is a valid joint distribution; in practice, we select a parametric family for C (Smith and Khaled, 2012). Note that, in contrast with the continuous case, when the marginals are discrete, some popular dependence measures, such as Kendall's τ, will depend on the marginals (Denuit and Lambert, 2005; Genest and Nešlehová, 2007). Consequently, the Kendall's τ of the random vector (Z_1, Z_2) will not be equivalent to the Kendall's τ of the copula. Without loss of generality, hereafter, we assume the bivariate copula carries a single parameter.

3.2   Copula NNMPs for discrete geostatistical data

Here, we introduce copula NNMPs with discrete marginals, with focus on using copulas to specify the bivariate distributions of the mixture components. Dropping the dependence on l for clarity, consider a random vector (U, V) with discrete marginal distributions F_U, F_V, and marginal pmfs f_U, f_V. Let $a_u = F_U(u^-)$ and $b_u = F_U(u)$, where $F_U(u^-)$ denotes the left limit of F_U at u. If U is ordinal, $F_U(u^-) = F_U(u-1)$. Analogous definitions of a_v and b_v apply for V. The joint pmf $f_{U,V}$ of (U, V) is obtained by finite differences,

$$f_{U,V}(u, v) = C(b_u, b_v) - C(b_u, a_v) - C(a_u, b_v) + C(a_u, a_v). \tag{6}$$

Let $c(u, v) = f_{U,V}(u, v)/(f_U(u)f_V(v))$, such that $f_{U,V}(u, v) = c(u, v)f_U(u)f_V(v)$, using a notation that is analogous to that of the joint density when (U, V) is continuous. Therefore, the conditional pmf, $f_{U|V}(u \,|\, v) = c(u, v)f_U(u)$.

To specify the distribution of base random vector (U_l, V_l), we use copula C_l with parameter η_l. For a parsimonious location-dependent model, we create spatially varying copulas $C_{\boldsymbol{v},l}$ on $(U_{\boldsymbol{v},l}, V_{\boldsymbol{v},l})$ by extending η_l to $\eta_l(\boldsymbol{v})$. In practice, we associate $\eta_l(\boldsymbol{v})$ to a spatial kernel that depends on $\boldsymbol{v} \in \mathcal{D}$ through a link function. Using Proposition 1 with a family of marginal pmfs $g_{\boldsymbol{v}}$, the joint pmf on $(U_{\boldsymbol{v},l}, V_{\boldsymbol{v},l})$ is $f_{U_{\boldsymbol{v},l}, V_{\boldsymbol{v},l}}(u, v) = c_{\boldsymbol{v},l}(u, v)f_{U_{\boldsymbol{v},l}}(u)f_{V_{\boldsymbol{v},l}}(v)$, where $f_{U_{\boldsymbol{v},l}} = g_{\boldsymbol{v}}$ and $f_{V_{\boldsymbol{v},l}} = g_{\boldsymbol{v}_{(l)}}$, and the conditional pmf is $f_{\boldsymbol{v},l}(u \,|\, v) = c_{\boldsymbol{v},l}(u, v)g_{\boldsymbol{v}}(u)$. Finally, the conditional pmf of the discrete copula NNMP model is given by

$$p(y(\boldsymbol{v}) \mid \boldsymbol{y}_{\mathrm{Ne}(\boldsymbol{v})}) = \sum_{l=1}^{L} w_l(\boldsymbol{v})\, c_{\boldsymbol{v},l}(y(\boldsymbol{v}), y(\boldsymbol{v}_{(l)}))\, g_{\boldsymbol{v}}(y(\boldsymbol{v})), \tag{7}$$

where the marginal pmf for $Y(\boldsymbol{v})$ is $g_{\boldsymbol{v}}$.

Recall that an NNMP model involves two sets of locations, the reference set $\mathcal{S}$ and nonreference set $\mathcal{U}$. As done in practice, we take the reference set $\mathcal{S}$ to correspond to the observed locations, and consider a generic finite set $\mathcal{U}$ such that $\mathcal{S} \cap \mathcal{U} = \emptyset$. Then, the joint pmf $\tilde{p}(\boldsymbol{y}_{\mathcal{V}})$ over set $\mathcal{V} = \mathcal{S} \cup \mathcal{U}$ describes the NNMP distribution over any finite set of locations that includes the observed locations. In general, for a discrete NNMP, an explicit expression for $\tilde{p}(\boldsymbol{y}_{\mathcal{V}})$ is not available, since working with a bivariate discrete distribution and its conditional pmf is difficult. However, using copulas to specify the bivariate mixture component yields a structured conditional pmf and allows for the study of the joint pmf. The following proposition provides an explicit expression for $\tilde{p}(\boldsymbol{y}_{\mathcal{V}})$ under a discrete copula NNMP. The proof of the proposition can be found in the supplementary material.

Proposition 2. *Consider a discrete copula NNMP model for spatial process $\{Y(\boldsymbol{v}) : \boldsymbol{v} \in \mathcal{D}\}$, with $\mathcal{S} = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n\}$ and $\mathcal{U} = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m\}$, where $n \geq 2$, $m \geq 1$, and $\mathcal{S} \cap \mathcal{U} = \emptyset$. Take $\mathcal{V} = \mathcal{S} \cup \mathcal{U}$, and let $\boldsymbol{y}_{\mathcal{V}} = (y(\boldsymbol{s}_1), \ldots, y(\boldsymbol{s}_n), y(\boldsymbol{u}_1), \ldots, y(\boldsymbol{u}_m))^\top$. Then the joint pmf*

of $\boldsymbol{y}_\mathcal{V}$ is $\tilde{p}(\boldsymbol{y}_\mathcal{V}) = \tilde{p}(\boldsymbol{y}_\mathcal{U} \mid \boldsymbol{y}_\mathcal{S})\tilde{p}(\boldsymbol{y}_\mathcal{S})$, where

$$\tilde{p}(\boldsymbol{y}_\mathcal{S}) = \prod_{i=1}^{n} g_{\boldsymbol{s}_i}(y(\boldsymbol{s}_i)) \sum_{l_n=1}^{n_L} \cdots \sum_{l_2=1}^{2_L} w_{\boldsymbol{s}_n,l_n} \dots w_{\boldsymbol{s}_2,l_2} c_{\boldsymbol{s}_n,l_n} \dots c_{\boldsymbol{s}_2,l_2},$$

$$\tilde{p}(\boldsymbol{y}_\mathcal{U} \mid \boldsymbol{y}_\mathcal{S}) = \prod_{i=1}^{m} g_{\boldsymbol{u}_i}(y(\boldsymbol{u}_i)) \sum_{l_m=1}^{L} \cdots \sum_{l_1=1}^{L} w_{\boldsymbol{u}_m,l_m} \dots w_{\boldsymbol{u}_1,l_1} c_{\boldsymbol{u}_m,l_m} \dots c_{\boldsymbol{u}_1,l_1}.$$

(8)

where $w_{\boldsymbol{s}_i,l_i} \equiv w_{l_i}(\boldsymbol{s}_i)$ and $c_{\boldsymbol{s}_i,l_i} \equiv c_{\boldsymbol{s}_i,l_i}(y(\boldsymbol{s}_i), y(\boldsymbol{s}_{(i,l_i)}))$, for $l_i = 1, \dots, i_L$, $i = 2, \dots, n$, and $w_{\boldsymbol{u}_i,l_i} \equiv w_{l_i}(\boldsymbol{u}_i)$ and $c_{\boldsymbol{u}_i,l_i} \equiv c_{\boldsymbol{u}_i,l_i}(y(\boldsymbol{u}_i), y(\boldsymbol{u}_{(i,l_i)}))$, for $l_i = 1, \dots, L$, $i = 1, \dots, m$.

We note that Proposition 2 also applies when $\boldsymbol{y}_\mathcal{V}$ is continuous. It indicates that, given the sequence of pmfs $g_{\boldsymbol{v}}$, the joint pmf of $\boldsymbol{y}_\mathcal{V}$ is determined by the collection of bivariate copulas, motivating the use of different copula families to construct discrete NNMPs. To balance flexibility and scalability, our strategy is to take all copulas C_l in one family with the same link function for the copula parameters. Table 1 presents three examples with copula parameters modeled via a link function $k : \mathcal{D} \times \mathcal{D} \to [0, 1]$. In particular, the Gumbel and Clayton copulas are asymmetric. They exhibit greater dependence in the positive and negative tails, respectively. In the first simulation example, we demonstrate that when the underlying spatial dependence is non-Gaussian, it may be appropriate to choose asymmetric copulas. We present next an example of a discrete copula NNMP construction.

Example 1. *Gaussian copula NNMP with negative binomial marginals.* For the family of marginal pmfs $g_{\boldsymbol{v}}$, consider the negative binomial distribution with mean $\alpha(\boldsymbol{v})$ and dispersion parameter r, denoted as $\text{NB}(\alpha(\boldsymbol{v}), r)$. Therefore, $g_{\boldsymbol{v}}(y) = \binom{y+r-1}{y}(p(\boldsymbol{v}))^r(1 - p(\boldsymbol{v}))^y$, with $p(\boldsymbol{v}) = r/(\alpha(\boldsymbol{v}) + r)$. To include a vector of covariates $\boldsymbol{x}(\boldsymbol{v})$, we take a log-link function for $\alpha(\boldsymbol{v})$ such that $\log(\alpha(\boldsymbol{v})) = \boldsymbol{x}(\boldsymbol{v})^\top \boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is a vector of regression parameters. We first specify Gaussian copulas C_l with correlation parameters ρ_l for the base random vectors (U_l, V_l). We then modify the correlation parameters ρ_l using a correlation function k for all l such that $\rho_l(\boldsymbol{v}) := k(\boldsymbol{v}, \boldsymbol{v}_{(l)})$, creating a sequence of spatially varying copulas $C_{\boldsymbol{v},l}$. The resulting model is given by (7) with $g_{\boldsymbol{v}} = \text{NB}(\alpha(\boldsymbol{v}), r)$.

Table 1: Examples of spatial copulas $C_{\boldsymbol{v},l}$ and corresponding link functions, $k : \mathcal{D} \times \mathcal{D} \to [0,1]$.

	$C_{\boldsymbol{v},l}(z_1, z_2)$	link function
Gaussian	$\Phi_2(\Phi^{-1}(z_1), \Phi^{-1}(z_2))$	$\rho_l(\boldsymbol{v}) = k(\boldsymbol{v}, \boldsymbol{v}_{(l)})$
Gumbel	$\exp(-((-\log z_1)^{\eta_l(\boldsymbol{v})} + (-\log z_2)^{\eta_l(\boldsymbol{v})})^{1/\eta_l(\boldsymbol{v})})$	$\eta_l(\boldsymbol{v}) = (1 - k(\boldsymbol{v}, \boldsymbol{v}_{(l)}))^{-1}$
Clayton	$(z_1^{-\delta_l(\boldsymbol{v})} + z_2^{-\delta_l(\boldsymbol{v})} - 1)^{-1/\delta_l(\boldsymbol{v})}$	$\delta_l(\boldsymbol{v}) = 2k(\boldsymbol{v}, \boldsymbol{v}_{(l)})/(1 - k(\boldsymbol{v}, \boldsymbol{v}_{(l)}))$

Note: the bivariate cdf Φ_2 corresponds to the standard bivariate Gaussian distribution with correlation $\rho \in (0,1)$, and the cdf Φ corresponds to the standard univariate Gaussian distribution.

3.3 Inference for discrete copula NNMPs

A traditional copula model for an n-variate discrete-valued vector involves evaluating 2^n terms of n-dimensional copulas. Unless n is very small, the computation is infeasible. Notable exceptions are discrete vine copula models (Panagiotelis et al., 2012) that decompose a multivariate pmf into bivariate copulas and marginals under a set of trees. The computations for likelihood evaluations grow quadratically in n. Discrete copula NNMPs compare favorably with discrete vine models, as the structured mixture formulation results in only $4nL$ bivariate copula function evaluations for the likelihood, providing linear growth in n.

Here, we develop a framework for discrete copula NNMP inference, based on transforming the discrete random variables to continuous ones by adding auxiliary variables, using the continuous extension (CE) approach in Denuit and Lambert (2005). Working with continuous marginals improves computational efficiency and stability: the likelihood requires only nL bivariate copula density evaluations; and, computing the conditional pmf using the finite differences in (6) is bypassed, thus avoiding numerical instability especially for copulas that are not analytically available, such as the Gaussian copula. Moreover, this framework makes more efficient the key task of spatial prediction over unobserved sites by avoiding computation that involves inverting the conditional cdf based on (6).

We associate each $Y(\boldsymbol{v})$ with a continuous random variable $Y^*(\boldsymbol{v})$, such that $Y^*(\boldsymbol{v}) = Y(\boldsymbol{v}) - O(\boldsymbol{v})$, where $O(\boldsymbol{v})$ is a continuous uniform random variable on $(0,1)$, independent of $Y(\boldsymbol{v})$ and of $O(\boldsymbol{v}')$, for $\boldsymbol{v}' \neq \boldsymbol{v}$. We refer to $Y^*(\boldsymbol{v})$ as the continued $Y(\boldsymbol{v})$ by $O(\boldsymbol{v})$. Let $Q_{\boldsymbol{v}}$ and $g_{\boldsymbol{v}}$ be the marginal cdf and pmf of $Y(\boldsymbol{v})$, respectively. Then, the marginal cdf

and density of $Y^*(\boldsymbol{v})$ are $Q_{\boldsymbol{v}}^*(y^*(\boldsymbol{v})) = Q_{\boldsymbol{v}}([y^*(\boldsymbol{v})]) + (y^*(\boldsymbol{v}) - [y^*(\boldsymbol{v})])g_{\boldsymbol{v}}([y^*(\boldsymbol{v}) + 1])$, and $g_{\boldsymbol{v}}^*(y^*(\boldsymbol{v})) = g_{\boldsymbol{v}}([y^*(\boldsymbol{v}) + 1])$, respectively, where $[x]$ denotes the integer part of x.

Based on marginal densities $g_{\boldsymbol{v}}^*$, we take spatial copulas $C_{\boldsymbol{v},l}^* = C_{\boldsymbol{v},l}$ for continuous random vectors $(U_{\boldsymbol{v},l}^*, V_{\boldsymbol{v},l}^*)$, with marginals $f_{U_{\boldsymbol{v},l}^*} = g_{\boldsymbol{v}}^*$ and $f_{V_{\boldsymbol{v},l}^*} = g_{\boldsymbol{v}_{(l)}}^*$, using copulas $C_{\boldsymbol{v},l}$ from the original NNMP model. The joint density on $(U_{\boldsymbol{v},l}^*, V_{\boldsymbol{v},l}^*)$ is $f_{U_{\boldsymbol{v},l}^*, V_{\boldsymbol{v},l}^*}(u, v) = c_{\boldsymbol{v},l}^*(u, v)g_{\boldsymbol{v}}^*(u)g_{\boldsymbol{v}_{(l)}}^*(v)$, and the conditional density is $f_{\boldsymbol{v},l}^*(u \,|\, v) = c_{\boldsymbol{v},l}^*(u, v)g_{\boldsymbol{v}}^*(u)$, where $c_{\boldsymbol{v},l}^*$ is the copula density. Denote by $\boldsymbol{y}_{\mathrm{Ne}(\boldsymbol{v})}^*$ the vector that contains the continued elements of $\boldsymbol{y}_{\mathrm{Ne}(\boldsymbol{v})}$, and $\boldsymbol{o}_{\mathrm{Ne}(\boldsymbol{v})}$ the vector of auxiliary variables for elements of $\boldsymbol{y}_{\mathrm{Ne}(\boldsymbol{v})}$. Then, the implied model on $y^*(\boldsymbol{v})$ is

$$p(y^*(\boldsymbol{v}) \,|\, D^*(\boldsymbol{v})) = \sum_{l=1}^{L} w_l(\boldsymbol{v})\, c_{\boldsymbol{v},l}^*(y^*(\boldsymbol{v}), y^*(\boldsymbol{v}_{(l)}))\, g_{\boldsymbol{v}}^*(y^*(\boldsymbol{v})) \tag{9}$$

where $y^*(\boldsymbol{v}) = y(\boldsymbol{v}) - o(\boldsymbol{v})$, and $D^*(\boldsymbol{v}) = \{\boldsymbol{y}_{\mathrm{Ne}(\boldsymbol{v})}^*, o(\boldsymbol{v}), \boldsymbol{o}_{\mathrm{Ne}(\boldsymbol{v})}\}$. Based on Proposition 1, model (9) has marginal density $g_{\boldsymbol{v}}^*$ for $Y^*(\boldsymbol{v})$. To recover $y(\boldsymbol{v})$, we first generate $y^*(\boldsymbol{v})$ from the extended model, and then set $y(\boldsymbol{v}) = [y^*(\boldsymbol{v}) + 1]$.

Regarding the existing literature, statistical inference for spatial copula models based on the CE approach is typically conducted by maximizing the expected likelihood with respect to the auxiliary variables (Madsen, 2009; Hughes, 2015). We develop inferential methods under the Bayesian framework. Posterior simulation based on (9) takes advantage of copula properties for continuous random variables, thus providing efficient computation for both model estimation and prediction.

4 Bayesian implementation

4.1 Hierarchical model formulation

Assume that $\boldsymbol{y}_{\mathcal{S}} = (y(\boldsymbol{s}_1), \ldots, y(\boldsymbol{s}_n))^\top$ is a realization of a discrete copula NNMP with spatially varying marginal pmfs through spatially dependent covariates, $g_{\boldsymbol{s}_i}(y(\boldsymbol{s}_i)) \equiv g(y(\boldsymbol{s}_i) \,|\, \boldsymbol{\beta}, \boldsymbol{\xi})$. Here, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^\top$, where β_0 is an intercept and $(\beta_1, \ldots, \beta_p)^\top$ is the regression parameter vector for covariates $\boldsymbol{x}(\boldsymbol{s}_i)$, and $\boldsymbol{\xi}$ collects all other parameters of g. The copula

parameter is modeled through a link function k with parameter(s) $\boldsymbol{\phi}$. We use the CE approach associating each $y(\boldsymbol{s}_i)$ with $y^*(\boldsymbol{s}_i)$, such that $y^*(\boldsymbol{s}_i) = y(\boldsymbol{s}_i) - o_i$, where $o_i \equiv o(\boldsymbol{s}_i)$ is uniformly distributed on $(0,1)$, independent of $y(\boldsymbol{s}_i)$ and of o_j, for $j \neq i$. Moreover, denote by $\boldsymbol{\zeta}$ the parameter of the cutoff point kernel for the mixture weights, defined in Section 2.1.

The formulation of the mixture weights allows us to augment the model with a sequence of auxiliary variables, $\{t_i : i = 3, \ldots, n\}$, where t_i is a Gaussian random variable with mean $\mu(\boldsymbol{s}_i)$ and variance κ^2. The augmented model for the data can be expressed as

$$y(\boldsymbol{s}_i) = y^*(\boldsymbol{s}_i) + o_i, \ \ o_i \overset{i.i.d.}{\sim} \text{Unif}(0,1), \ \ i = 1, \ldots, n,$$

$$y^*(\boldsymbol{s}_1) \mid \boldsymbol{\beta}, \boldsymbol{\xi} \sim g_{\boldsymbol{s}_1}^*(y^*(\boldsymbol{s}_1)), \ \ y^*(\boldsymbol{s}_2) \mid y^*(\boldsymbol{s}_1), \boldsymbol{\phi}, \boldsymbol{\beta}, \boldsymbol{\xi} \sim f_{\boldsymbol{s}_2,1}^*(y^*(\boldsymbol{s}_2) \mid y^*(\boldsymbol{s}_1)),$$

$$y^*(\boldsymbol{s}_i) \mid \{y^*(\boldsymbol{s}_{(il)})\}_{l=1}^{i_L}, t_i, \boldsymbol{\phi}, \boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{\zeta} \overset{ind.}{\sim} \sum_{l=1}^{i_L} f_{\boldsymbol{s}_i,l}^*(y^*(\boldsymbol{s}_i) \mid y^*(\boldsymbol{s}_{(il)})) \mathbb{1}_{(r_{\boldsymbol{s}_i,l-1}^*, r_{\boldsymbol{s}_i,l}^*)}(t_i), \ \ i = 3, \ldots, n,$$

$$t_i \mid \boldsymbol{\gamma}, \kappa^2 \overset{ind.}{\sim} N(t_i \mid \gamma_0 + \gamma_1 s_{i1} + \gamma_2 s_{i2}, \kappa^2), \ \ i = 3, \ldots, n,$$

where $f_{\boldsymbol{s}_i,l}^*(y^*(\boldsymbol{s}_i) \mid y^*(\boldsymbol{s}_{(il)})) = c_{\boldsymbol{s}_i,l}^*(y^*(\boldsymbol{s}_i), y^*(\boldsymbol{s}_{(il)})) g_{\boldsymbol{s}_i}^*(y^*(\boldsymbol{s}_i))$, and $r_{\boldsymbol{s}_i,l}^* = \log\{r_{\boldsymbol{s}_i,l}/(1 - r_{\boldsymbol{s}_i,l})\}$, for $l = 1, \ldots, i_L$. The full Bayesian model is completed with prior specification for parameters $\boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{\phi}, \boldsymbol{\zeta}, \boldsymbol{\gamma} = (\gamma_0, \gamma_1, \gamma_2)^\top$ and κ^2. The priors for $\boldsymbol{\xi}$, $\boldsymbol{\phi}$, and $\boldsymbol{\zeta}$ depend on the choices of the pmf $g_{\boldsymbol{s}_i}$, the copula $C_{\boldsymbol{s}_i,l}^*$, and the kernel k', respectively. For parameters $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and κ^2, we consider $N(\boldsymbol{\beta} \mid \mu_{\boldsymbol{\beta}}, \boldsymbol{V_\beta})$, $N(\boldsymbol{\gamma} \mid \mu_{\boldsymbol{\gamma}}, \boldsymbol{V_\gamma})$, and $\text{IG}(\kappa^2 \mid u_{\kappa^2}, v_{\kappa^2})$ priors, where IG denotes the inverse gamma distribution.

4.2 Model estimation, validation and prediction

We outline the MCMC sampler for parameters $(\boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{\phi}, \boldsymbol{\zeta}, \boldsymbol{\gamma}, \kappa^2)$, and latent variables $\{t_i\}_{i=3}^n$ and $\{o_i\}_{i=1}^n$. We note that there is a set of configuration variables $\{\ell_i\}_{i=3}^n$ in one-to-one correspondence with t_i, i.e., $\ell_i = l$ if and only if $t_i \in (r_{\boldsymbol{s}_i,l-1}^*, r_{\boldsymbol{s}_i,l}^*)$, for $l = 1, \ldots, i_L$.

The updates for parameters $\boldsymbol{\beta}$, $\boldsymbol{\xi}$ and $\boldsymbol{\phi}$ require Metropolis steps, since they enter in copula densities $c_{\boldsymbol{s}_i,l}^*$. We use a Metropolis step also for kernel k' parameter $\boldsymbol{\zeta}$, which is involved in the definition of the mixture weights. Let $\boldsymbol{D}$ be the $(n-2) \times 3$ matrix with ith row $(1, s_{i+2,1}, s_{i+2,2})$. The posterior full conditional of $\boldsymbol{\gamma}$ is $N(\boldsymbol{\gamma} \mid \boldsymbol{\mu_\gamma^*}, \boldsymbol{V_\gamma^*})$, where $\boldsymbol{V_\gamma^*} =$

$(\boldsymbol{V}_\gamma^{-1} + \kappa^{-2}\boldsymbol{D}^\top\boldsymbol{D})^{-1}$ and $\boldsymbol{\mu}_\gamma^* = \boldsymbol{V}_\gamma^*(\boldsymbol{V}_\gamma^{-1}\boldsymbol{\mu}_\gamma + \kappa^{-2}\boldsymbol{D}^\top\boldsymbol{t})$, with the vector $\boldsymbol{t} = (t_3, \ldots, t_n)^\top$. The posterior full conditional distribution of κ^2 is $\text{IG}(\kappa^2 \mid u_{\kappa^2} + (n-2)/2, v_{\kappa^2} + \sum_{i=3}^n (t_i - \mu(\boldsymbol{s}_i))^2/2)$.

The posterior full conditional distribution for each latent variable t_i, $i = 3, \ldots, n$, can be expressed as $\sum_{l=1}^{i_L} q_l(\boldsymbol{s}_i)\,\text{TN}(t_i \mid \mu(\boldsymbol{s}_i), \kappa^2; r_{\boldsymbol{s}_i,l-1}^* < t_i \leq r_{\boldsymbol{s}_i,l}^*)$, where TN denotes the truncated normal distribution over the indicated interval, and $q_l(\boldsymbol{s}_i) \propto w_l(\boldsymbol{s}_i)\,c_{\boldsymbol{s}_i,l}^*(y^*(\boldsymbol{s}_i), y^*(\boldsymbol{s}_{(il)}))$, for $l = 1, \ldots, i_L$. Hence, each t_i can be readily updated by sampling from the l-th truncated normal with probability proportional to $q_l(\boldsymbol{s}_i)$. For auxiliary variables o_i, the posterior full conditional of o_1 is proportional to $\prod_{\{j:\boldsymbol{s}_{(j,\ell_j)}=\boldsymbol{s}_1\}} c_{\boldsymbol{s}_j,\ell_j}^*(y(\boldsymbol{s}_j) - o_j, y(\boldsymbol{s}_1) - o_1)$, and that of o_i, $i \geq 2$, is proportional to $c_{\boldsymbol{s}_i,\ell_i}^*(y(\boldsymbol{s}_i) - o_i, y(\boldsymbol{s}_{(i,\ell_i)}) - o_{(i,\ell_i)}) \prod_{\{j:\boldsymbol{s}_{(j,\ell_j)}=\boldsymbol{s}_i\}} c_{\boldsymbol{s}_j,\ell_j}^*(y(\boldsymbol{s}_j) - o_j, y(\boldsymbol{s}_i) - o_i)$, where $\ell_2 = 1$ and $o_{(i,\ell_i)} \equiv o(\boldsymbol{s}_{(i,\ell_i)})$. We update each latent variable o_i with an independent Metropolis step with a $\text{Unif}(0,1)$ proposal distribution.

The likelihood of the continued model admits the form $g_{\boldsymbol{s}_1}(y^*(\boldsymbol{s}_1)) \prod_{i=2}^n p(y^*(\boldsymbol{s}_i) \mid D^*(\boldsymbol{s}_i))$. The product formulation allows for model validation, using a generalization of the randomized quantile residuals proposed by Dunn and Smyth (1996) for independent data. Specifically, we define the marginal quantile residual, $r_1 = \Phi^{-1}(Q_{\boldsymbol{s}_1}^*(y^*(\boldsymbol{s}_1)))$, and the ith conditional quantile residual, $r_i = \Phi^{-1}(F(y^*(\boldsymbol{s}_i) \mid D^*(\boldsymbol{s}_i)))$, $i = 2, \ldots, n$, where F is the conditional cdf of $y^*(\boldsymbol{s}_i)$. If the model is correctly specified, the residuals r_i, $i = 1, \ldots, n$, would be independent and identically distributed as a standard Gaussian distribution.

Finally, we turn to posterior predictive inference at a new location $\boldsymbol{v}_0$. If $\boldsymbol{v}_0 \notin \mathcal{S}$, for each posterior sample, we first compute the cutoff points $r_{\boldsymbol{v}_0,l}$, such that $r_{\boldsymbol{v}_0,l} - r_{\boldsymbol{v}_0,l-1} = k'(\boldsymbol{v}_0, \boldsymbol{v}_{(0l)})/\sum_{l=1}^L k'(\boldsymbol{v}_0, \boldsymbol{v}_{(0l)})$, and the weights $w_l(\boldsymbol{v}_0) = G_{\boldsymbol{v}_0}(r_{\boldsymbol{v}_0,l}) - G_{\boldsymbol{v}_0}(r_{\boldsymbol{v}_0,l-1})$, for $l = 1, \ldots, L$. We then generate $y^*(\boldsymbol{v}_0)$ based on (9), and set $y(\boldsymbol{v}_0) = [y^*(\boldsymbol{v}_0) + 1]$. If $\boldsymbol{v}_0 \equiv \boldsymbol{s}_i \in \mathcal{S}$, we generate $y(\boldsymbol{v}_0)$ similarly, the difference being that we now use the posterior samples for the mixture weights obtained from the MCMC algorithm.

5   Data illustrations

To illustrate the proposed methodology, we present two synthetic data examples and a real data analysis. The goal of the first simulation experiment is to investigate the flexibility

of discrete copula NNMPs, using different copula functions to define the NNMP mixture components. In the second experiment, we demonstrate the inferential and computational advantages of our approach for count data modeling, compared to SGLMMs. Implementation details for the models are provided in the supplementary material. Since our purpose is primarily demonstrative, we took $L = 10$ for the simulation experiments. A comprehensive sensitivity analysis for L was conducted for the real data application of Section 5.3, with details provided in the supplementary material.

In both simulated data examples, we ran the MCMC algorithm for each copula NNMP model for 20000 iterations, discarding the first 4000 iterations, and collecting posterior samples every four iterations. The SGLMM models were implemented using the spBayes package in R (Finley et al., 2007); we ran the algorithm for 40000 iterations and collected posterior samples every five iterations, with the first 20000 as burn-in.

We compare models based on parameter estimates, root mean squared prediction error (RMSPE), 95% credible interval width (95% CI width), 95% credible interval coverage rate (95% CI cover), continuous ranked probability score (CRPS; Gneiting and Raftery 2007), energy score (ES; Gneiting and Raftery 2007), and variogram score of order one (VS; Scheuerer and Hamill 2015). The energy score is a multivariate extension of the CRPS, while the variogram score examines pairwise differences of the components of the multivariate quantity. Both the ES and VS allow for comparison of model predictive performance with respect to dependence structure.

5.1   First simulation experiment

We first generated sites over a regular grid of 120×120 resolution on a unit square domain, and then simulated data from $y(\boldsymbol{v}) = F_Y^{-1}\big(F_Z(z(\boldsymbol{v}))\big)$, where F_Y corresponds to the Poisson distribution with rate parameter $\lambda_0 = 5$, and $z(\boldsymbol{v})$ is the skew-Gaussian random field from Zhang and El-Shaarawi (2010) with stationary marginal distribution F_Z. More specifically, $z(\boldsymbol{v}) = \sigma_1 |\omega_1(\boldsymbol{v})| + \sigma_2 \omega_2(\boldsymbol{v})$, where both $\omega_1(\boldsymbol{v})$ and $\omega_2(\boldsymbol{v})$ are standard Gaussian processes with exponential correlation function based on range parameter 0.1. The density of F_Z is $f_Z(z) = 2 N(z \,|\, 0, \sigma_1^2 + \sigma_2^2) \, \Phi(\sigma_1 z / (\sigma_2 \sqrt{\sigma_1^2 + \sigma_2^2}))$, where $\sigma_1 \in \mathbb{R}$ controls the skewness,

Table 2: Simulation example 1: posterior mean and 95% CI estimates for the rate parameter λ of the Poisson NNMP marginal distribution, and scores for comparison of Gaussian, Gumbel and Clayton copula NNMP models, under each of the three simulation scenarios for σ_1.

| | $\sigma_1 = 1$ | | | $\sigma_1 = 3$ | | | $\sigma_1 = 10$ | | |
	λ			λ			λ		
Gaussian	4.55 (4.16, 4.94)			4.71 (4.37, 5.07)			4.88 (4.55, 5.22)		
Gumbel	4.78 (4.39, 5.21)			4.88 (4.56, 5.24)			4.94 (4.66, 5.23)		
Clayton	5.33 (4.99, 5.68)			5.25 (4.96, 5.56)			5.36 (5.08, 5.65)		
	$\sigma_1 = 1$			$\sigma_1 = 3$			$\sigma_1 = 10$		
	CRPS	ES	VS	CRPS	ES	VS	CRPS	ES	VS
Gaussian	0.69	12.77	94855	0.85	15.54	124893	0.93	16.98	138592
Gumbel	0.69	12.58	92278	0.85	15.32	120932	0.92	16.71	134774
Clayton	0.75	14.34	125800	0.90	17.36	164148	1.00	18.70	174123

and $\sigma_2 > 0$ is a scale parameter. We took $\sigma_2 = 1$, and $\sigma_1 = 1, 3, 10$, which corresponds to positive weak, moderate, and strong skewness.

We considered three discrete copula NNMPs with stationary Poisson marginals, i.e., $g_{\boldsymbol{v}} = f_Y$, for all $\boldsymbol{v}$, where f_Y is the Poisson pmf with rate λ. The three models correspond to the copulas in Table 1, with the link function k given by an exponential correlation function with range parameter denoted by ϕ_1, ϕ_2, and ϕ_3 for the Gaussian, Gumbel, and Clayton copula models, respectively. We specified the cutoff point kernel through an exponential correlation function with range parameter ζ_1, ζ_2, and ζ_3 for the Gaussian, Gumbel, and Clayton copula models, respectively. The Bayesian models are fully specified with an IG(3, 1) prior for the ϕ and ζ parameters, and with $N(\boldsymbol{\gamma} \,|\, (-1.5, 0, 0)^{\top}, 2\mathbf{I}_3)$ and IG($\kappa^2 \,|\, 3, 1$) priors. Finally, the prior for the rate parameter λ was taken as Ga(1, 1), where Ga(a, b) denotes the gamma distribution with mean a/b. We simulated 1000 responses and used 800 of them to fit the three NNMP models. The remaining 200 observations were used for model comparison.

Table 2 provides estimates for the rate parameter λ of the Poisson marginal distribution, and predictive performance metrics. For all three cases for $\sigma_1 = 1, 3, 10$, the Gumbel model yields the more accurate estimates for λ. In particular, the Gumbel model's 95% CIs include the true parameter value, whereas those of the Gaussian and Clayton models failed to cover it when $\sigma_1 = 1$ and $\sigma_1 = 10$, respectively. Regarding predictive performance,

17

the Gumbel model outperforms to a smaller or larger extent the other two models across different scenarios. Predictive random fields under the three models are provided in the supplementary material. We found that prediction by the Clayton model was not able to recover large values. Compared to the Gaussian model, the Gumbel model recovered large values slightly better. Overall, this example demonstrates that, when the underlying spatial dependence is driven by non-Gaussian processes, it is practically useful to consider copulas from asymmetric families, including use of appropriate model comparison tools.

5.2 Second simulation experiment

We generated data over a grid of sites with 120×120 resolution, uniformly on the square $[0,1] \times [0,1]$, using a Poisson SGLMM with $y(\boldsymbol{v}) \,|\, \eta(\boldsymbol{v}) \sim \text{Pois}(\eta(\boldsymbol{v}))$, and $\log(\eta(\boldsymbol{v})) = \beta_0 + v_1\beta_1 + v_2\beta_2 + z(\boldsymbol{v})$, where $\boldsymbol{v} = (v_1, v_2)$, and $z(\boldsymbol{v})$ is a zero-centered Gaussian process (GP) with variance parameter $\sigma^2 = 0.2$ and an exponential correlation function with range parameter $\phi_0 = 1/12$. We set the regression coefficients $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^\top = (1.5, 1, 2)^\top$, resulting in a random field with a trend, as shown in Figure 1(a).

We considered three models. The first is the negative binomial NNMP model (NBN-NMP) with a Gaussian copula, as discussed in Example 1. The second model (SGLMM-GP) is a Poisson SGLMM with a GP prior assigned to $z(\boldsymbol{v})$. For the last model (SGLMM-GPP), we considered a Poisson SGLMM with spatial random effects $z(\boldsymbol{v})$ corresponding to a Gaussian predictive process (GPP, Banerjee et al. 2008), with 10×10 knots placed on a grid over the domain. We chose the number of knots such that the computing times for the SGLMM-GPP and NBNNMP models are similar. As in the first simulation example, all models were fit to 800 observations and compared on the basis of 200 additional observations.

The regression coefficients for all models were assigned mean-zero, dispersed normal priors. We worked with an exponential correlation function for all models, used for $\rho_l(\boldsymbol{v})$ of the Gaussian copula in the NBNNMP model, and as the correlation function for the GP and GPP in the SGLMMs. The range parameter was assigned an inverse gamma prior $\text{IG}(3, 1)$ for the NBNNMP model, and a uniform prior $\text{Unif}(1/30, 1/3)$ for the other two models. The cutoff point kernel of the NBNNMP was also specified an exponential correlation function,
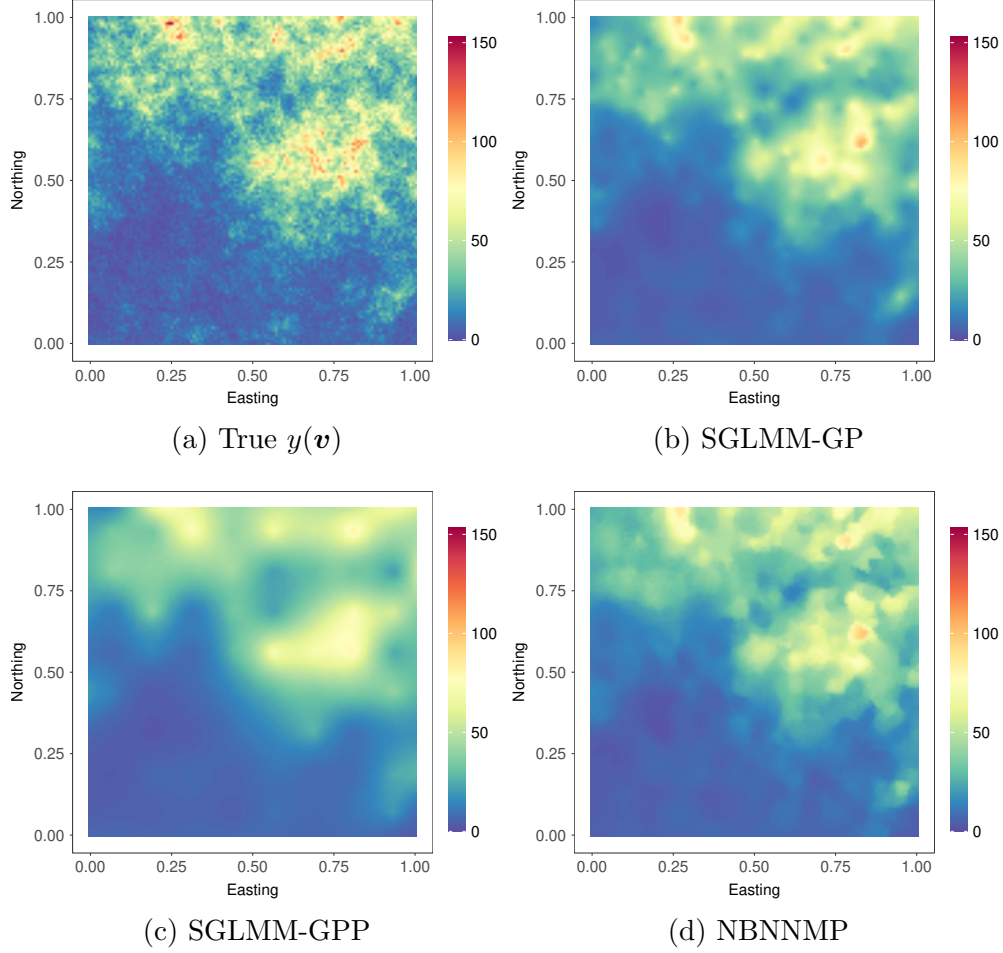
18

(a) True $y(\boldsymbol{v})$ (b) SGLMM-GP

(c) SGLMM-GPP   (d) NBNNMP

Figure 1: Second simulation example. Interpolated surfaces of the true model and posterior median estimates of the SGLMM-GP, SGLMM-GPP and NBNNMP models.

with an IG(3, 1) prior for the range parameter. The variance parameter for the SGLMM models was assigned an inverse gamma prior IG(2, 1). For the logit Gaussian distribution parameters $\boldsymbol{\gamma}$ and κ^2 of the NBNNMP, we used $N((-1.5, 0, 0)^\top, 2\mathbf{I}_3)$ and IG(3, 1) priors, respectively. Finally, we placed a Ga(1, 1) prior on the NBNNMP dispersion parameter r.

Estimates of the regression parameters and performance metrics for out-of-sample prediction are provided in Table 3. We observe that, overall, the NBNNMP model provided the more accurate estimation for $\boldsymbol{\beta}$. Regarding predictive performance, the NBNNMP model outperformed the SGLMM-GPP model by a large margin, and was comparable to the SGLMM-GP model, which corresponds to the data generating process for this simulation experiment. Moreover, the last row of the table highlights the NBNNMP model's

Table 3: Simulation example 2: posterior mean and 95% CI estimates for the regression parameters, performance metrics, and computing time, under the NBNNMP model and the two SGLMM models.

	True	NBNNMP	SGLMM-GP	SGLMM-GPP
β_0	1.5	$1.61\,(1.29, 1.97)$	$1.53\,(1.22, 1.81)$	$1.41\,(1.02, 1.73)$
β_1	1	$0.90\,(0.51, 1.31)$	$0.70\,(0.25, 1.15)$	$0.91\,(0.43, 1.34)$
β_2	2	$1.94\,(1.51, 2.32)$	$2.18\,(1.91, 2.53)$	$2.25\,(1.81, 2.84)$
RMSPE	-	9.06	8.88	10.00
95% CI cover	-	0.98	0.97	0.78
95% CI width	-	37.02	32.24	19.02
CRPS	-	4.58	4.52	5.37
ES	-	92.07	91.41	107.46
VS	-	5175591	5199629	6378263
Time (mins)	-	11.18	935.02	11.68

huge gains in computing time compared to the SGLMM-GP model.

Figure 1(b)-1(d) plots the posterior median estimates of the random field for the three models. The SGLMM-GPP yields an overly smooth estimate, whereas the SGLMM-GP and NBNNMP models provide similar estimates that approximate well the true surface. Overall, this example illustrates the inferential and computational advantages of discrete copula NNMPs for modeling count data.

5.3 North American Breeding Bird Survey data analysis

The primary source of information on population evolution for birds is count data surveys. Since 1966, the North American Breeding Bird Survey (BBS) has been conducted to monitor bird population change. There are over 4000 sampling units in the survey, each with a 24.5-mile roadside route. Along each route, volunteer observers count the number of birds by sight or sound, in a 3-min period at each of 50 stops (Pardieck et al., 2020). The BBS data are often used to determine temporal or geographical patterns of relative abundance. Spatial maps of relative abundance are crucial for ecological studies.

We are interested in the relative abundance of the Northern Cardinal, a bird species that is prevalent in Eastern United States. Figure 2(a) shows the number of birds observed in 2019, with the sizes of the circle radii proportional to the number of birds at each

(a) Observed counts

(b) Predicted counts

(c) 95% CI widths

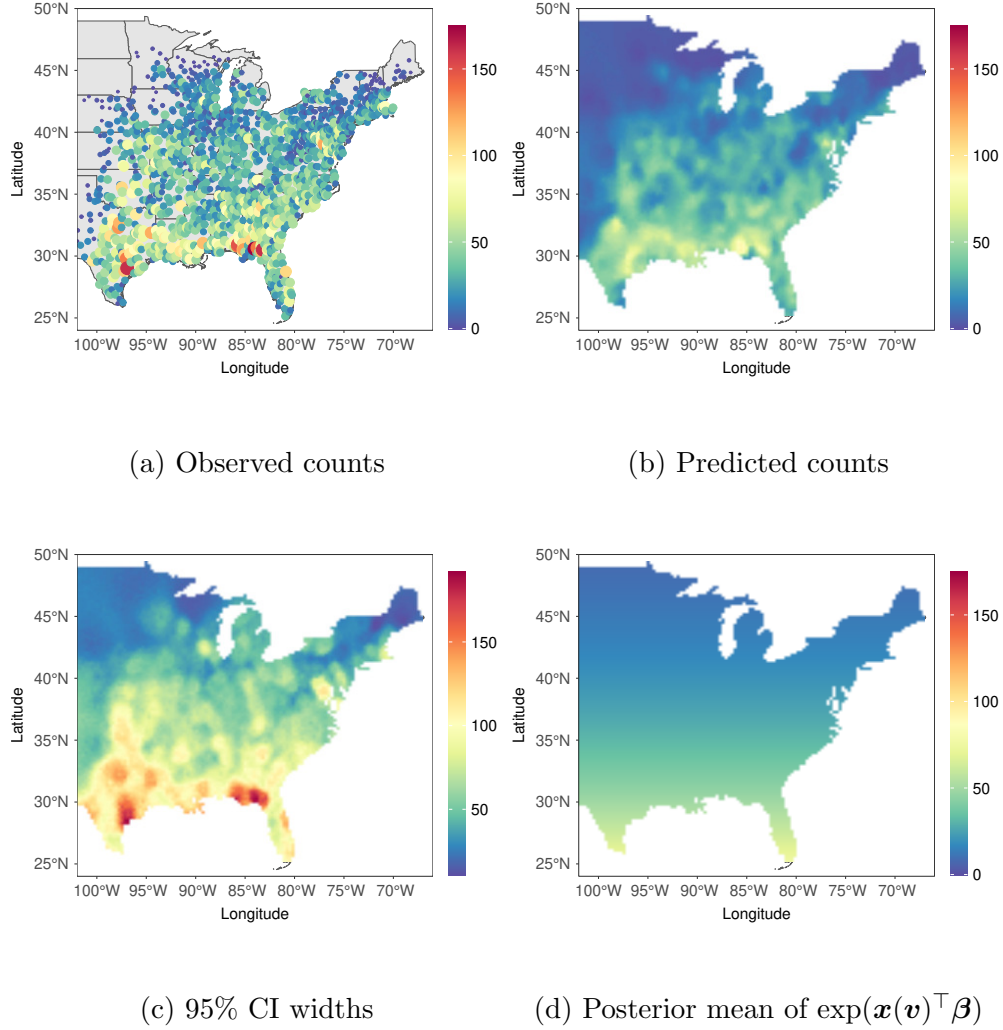(d) Posterior mean of $\exp(\boldsymbol{x}(\boldsymbol{v})^{\top}\boldsymbol{\beta})$

Figure 2: North American Breeding Bird Survey data analysis: (a) observed counts for 2019 BBS of Northern Cardinal, with circle radius proportional to the observed counts; (b) median of the posterior predictive distribution for Northern Cardinal count; (c) widths of the 95% CI of the posterior predictive distribution for Northern Cardinal count; (d) posterior mean of $\exp(\boldsymbol{x}(\boldsymbol{v})^{\top}\boldsymbol{\beta})$.

sampling location. The dataset was extracted with the help of the R package *bbsAssistant* (Burnett et al., 2019); it contains 1515 irregular sampling locations. From Figure 2(a) we observe that the counts tend to increase as latitude decreases, and we thus take latitude as a covariate to account for the long range variability in the population.

We considered the Gaussian copula NBNNMP model defined in Example 1, with spatially varying marginal $\mathrm{NB}(\exp(\boldsymbol{x}(\boldsymbol{v})^{\top}\boldsymbol{\beta}), r)$, where $\boldsymbol{\beta} = (\beta_0, \beta_1)^{\top}$. We used the same link functions and prior specifications as in Section 5.2. We first examined model performance

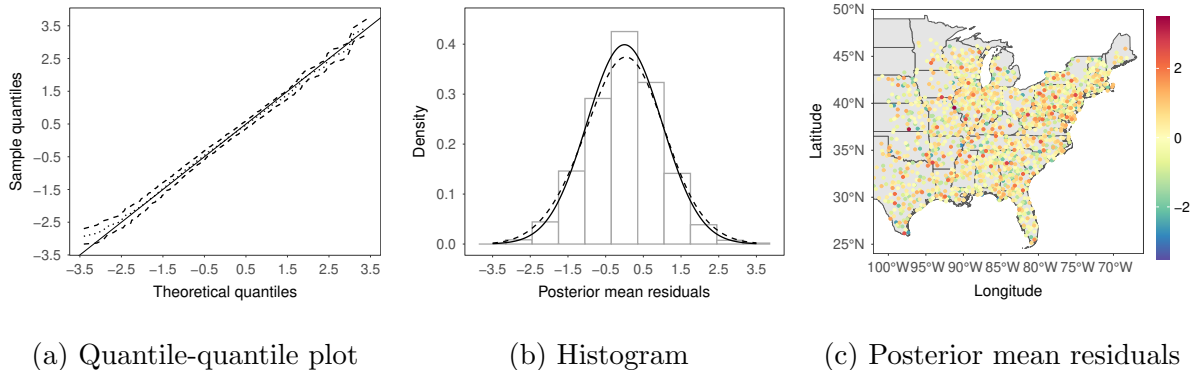(a) Quantile-quantile plot (b) Histogram (c) Posterior mean residuals

Figure 3: North American Breeding Bird Survey data analysis. Randomized quantile residual analysis: (a) dotted and dashed lines correspond to the posterior mean and 95% interval bands, respectively; (b) solid and dashed lines are the standard Gaussian density and the kernel density estimate of the posterior means of the residuals, respectively; (c) spatial plot of the posterior means of the residuals.

under different values of L. Overall, parameter estimates were quite robust. The estimates of mixture weights suggested that the effective number of neighbors for each location was quite consistent for L between 10 and 20. Also, there was no discernible differences for out-of-sample predictive performance. Therefore, we took $L = 20$ as a reasonable upper bound. We also compared NBNNMP models with the three copulas listed in Table 1, using the same link functions for copulas as in Section 5.1. The three models were evaluated based on their predictive performance. Overall, the Gaussian copula outperformed the other two. Details of these analyses are provided in the supplementary material.

We proceeded to analyze the BBS data with the Gaussian copula NBNNMP model with $L = 20$. The posterior mean and 95% CI estimates of the regression parameters β_0 and β_1 are $6.53\,(5.61, 7.38)$ and $-0.09\,(-0.11, -0.06)$, respectively, suggesting an increasing trend in the Northern Cardinal counts as the latitude decreases. The corresponding estimates of the dispersion parameter r are $1.88\,(1.55, 2.22)$, indicating that there is overdispersion over the domain. Figure 2(b) and 2(c) show the posterior predictive median of the counts and the 95% posterior predictive CI width, respectively. Figure 2(b) displays the domain's spatial variability. The estimated uncertainty, as shown in Figure 2(c), is meaningful, as areas with high uncertainty correspond to those where the observed counts are quite heterogeneous. Figure 2(d) provides a spatial map of the mean of the negative binomial marginals, which depicts a North–South trend. Model checking results are shown in Figure

3, including a posterior summary of the Gaussian quantile-quantile plot, and the histogram and spatial plot of the posterior means of the residuals. The results suggest good model fit.

Finally, we compared the NBNNMP with the SGLMM-GP model (details are given in the supplementary material). The parameter estimates of $\boldsymbol{\beta}$ were quite close under the two models. On the other hand, the NBNNMP model resulted in better out-of-sample predictive performance, and, notably, it was substantially more efficient to implement, with computing time 110 times faster than that for the SGLMM-GP model.

6 Discussion

We have introduced a new class of models for discrete geostatistical data, with particular focus on using different families of bivariate copulas to build modeling and inference. Compared to traditional SGLMM methods, the proposed class of models is scalable, and is able to accommodate complex dependence structures.

In general, multivariate discrete distributions are not as tractable as certain families of multivariate continuous distributions, in particular, the Gaussian family. This is the fundamental difficulty of process-based modeling for discrete geostatistical data. Our methodology overcomes this difficulty through a structured mixture model formulation, reducing the specification of a multivariate pmf to that of bivariate copulas that define the mixture components. This formulation yields models for spatial processes that provide flexibility and deliver computational scalability.

In the present work, we explored the strategy of using a single copula family for all bivariate distributions. Exploring the alternative which builds from different copula families for the bivariate distributions remains an interesting question to investigate. We can cast this as a model selection problem and develop algorithms to select models; see examples in Panagiotelis et al. (2017) and Gruber and Czado (2018) in the context of regular vine copula models. Different copula families for bivariate distributions yield more flexibility for the model to capture complex dependence, albeit at the cost of computational scalability. If the main purpose of the application is prediction, rather than model selection, one could

23

explore calibrating the prediction using all candidate copula families. This could be done, for example, with the pseudo Bayesian model averaging approach, where the weight for each model is estimated based on stacking (Yao et al., 2018).

We conduct inference for the discrete copula NNMPs based on the continuous extension approach. Apart from the aforementioned benefits, this approach may allow discrete copula NNMPs to make use of alternative algorithms for faster computation, which are currently being developed for continuous NNMP models. Moreover, with the CE approach, it is possible to develop a class of NNMPs for a multivariate response that consists of both continuous and discrete components, while at the same time retaining computational efficiency.

Supplementary Material

The supplementary material includes proofs and other technical details, sampling algorithm details, and additional results on the data examples of Section 5.1 and 5.3.

References

Banerjee, S., Gelfand, A. E., Finley, A. O., and Sang, H. (2008), "Gaussian predictive process models for large spatial data sets," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70, 825–848.

Burnett, J. L., Wszola, L., and Palomo-Muñoz, G. (2019), "bbsAssistant: An R package for downloading and handling data and information from the North American Breeding Bird Survey," *Journal of Open Source Software*, 4, 1768.

Christensen, O. F., Roberts, G. O., and Sköld, M. (2006), "Robust Markov chain Monte Carlo methods for spatial generalized linear mixed models," *Journal of Computational and Graphical Statistics*, 15, 1–17.

Christensen, O. F. and Waagepetersen, R. (2002), "Bayesian prediction of spatial count data using generalized linear mixed models," *Biometrics*, 58, 280–286.

Datta, A., Banerjee, S., Finley, A. O., and Gelfand, A. E. (2016), "Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets," *Journal of the American Statistical Association*, 111, 800–812.

Denuit, M. and Lambert, P. (2005), "Constraints on concordance measures in bivariate discrete data," *Journal of Multivariate Analysis*, 93, 40–57.

Diggle, P. J., Tawn, J. A., and Moyeed, R. A. (1998), "Model-based geostatistics," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47, 299–350.

Dunn, P. K. and Smyth, G. K. (1996), "Randomized quantile residuals," *Journal of Computational and Graphical Statistics*, 5, 236–244.

Finley, A. O., Banerjee, S., and Carlin, B. P. (2007), "spBayes: An R Package for Univariate and Multivariate Hierarchical Point-Referenced Spatial Models," *Journal of Statistical Software*, 19, 1–24, URL http://www.jstatsoft.org/v19/i04/.

Genest, C. and Nešlehová, J. (2007), "A primer on copulas for count data," *ASTIN Bulletin: The Journal of the IAA*, 37, 475–515.

Gneiting, T. and Raftery, A. E. (2007), "Strictly proper scoring rules, prediction, and estimation," *Journal of the American Statistical Association*, 102, 359–378.

Gruber, L. F. and Czado, C. (2018), "Bayesian model selection of regular vine copulas," *Bayesian Analysis*, 13, 1111–1135.

Guinness, J. (2018), "Permutation and grouping methods for sharpening Gaussian process approximations," *Technometrics*, 60, 415–429.

Han, Z. and De Oliveira, V. (2016), "On the correlation structure of Gaussian copula models for geostatistical count data," *Australian & New Zealand Journal of Statistics*, 58, 47–69.

Hughes, J. (2015), "copCAR: A flexible regression model for areal data," *Journal of Computational and Graphical Statistics*, 24, 733–755.

Joe, H. (2014), *Dependence modeling with copulas*, Boca Raton, FL: CRC press.

Katzfuss, M. and Guinness, J. (2021), "A general framework for Vecchia approximations of Gaussian processes," *Statistical Science*, 36, 124–141.

Kazianka, H. and Pilz, J. (2010), "Copula-based geostatistical modeling of continuous and discrete data including covariates," *Stochastic Environmental Research and Risk Assessment*, 24, 661–673.

Le, N. D., Martin, R. D., and Raftery, A. E. (1996), "Modeling flat stretches, bursts outliers in time series using mixture transition distribution models," *Journal of the American Statistical Association*, 91, 1504–1515.

Madsen, L. (2009), "Maximum likelihood estimation of regression parameters with spatially dependent discrete data," *Journal of Agricultural, Biological, and Environmental Statistics*, 14, 375–391.

Panagiotelis, A., Czado, C., and Joe, H. (2012), "Pair copula constructions for multivariate discrete data," *Journal of the American Statistical Association*, 107, 1063–1072.

Panagiotelis, A., Czado, C., Joe, H., and Stöber, J. (2017), "Model selection for discrete regular vine copulas," *Computational Statistics & Data Analysis*, 106, 138–152.

Pardieck, K., Ziolkowski Jr, D., Lutmerding, M., Aponte, V., and Hudson, M. (2020), "North American Breeding Bird Survey Dataset 1966–2019: US Geological Survey data release," .

Recta, V., Haran, M., and Rosenberger, J. L. (2012), "A two-stage model for incidence and prevalence in point-level spatial count data," *Environmetrics*, 23, 162–174.

Scheuerer, M. and Hamill, T. M. (2015), "Variogram-based proper scoring rules for probabilistic forecasts of multivariate quantities," *Monthly Weather Review*, 143, 1321–1334.

Sengupta, A. and Cressie, N. (2013), "Hierarchical statistical modeling of big spatial datasets using the exponential family of distributions," *Spatial Statistics*, 4, 14–44.

Sklar, M. (1959), "Fonctions de repartition an dimensions et leurs marges," *Publications de l'Institut de Statistique de L'Université de Paris*, 8, 229–231.

Smith, M. S. and Khaled, M. A. (2012), "Estimation of copula models with discrete margins via Bayesian data augmentation," *Journal of the American Statistical Association*, 107, 290–303.

Vecchia, A. V. (1988), "Estimation and model identification for continuous spatial processes," *Journal of the Royal Statistical Society: Series B (Methodological)*, 50, 297–312.

Wikle, C. K. (2002), "Spatial modeling of count data: A case study in modelling breeding bird survey data on large spatial domains," *Spatial Cluster Modelling*, 199, 209.

Wu, G., Holan, S. H., Nilon, C. H., and Wikle, C. K. (2015), "Bayesian binomial mixture models for estimating abundance in ecological monitoring studies," *The Annals of Applied Statistics*, 9, 1–26.

Yao, Y., Vehtari, A., Simpson, D., and Gelman, A. (2018), "Using stacking to average Bayesian predictive distributions (with discussion)," *Bayesian Analysis*, 13, 917–1007.

Zhang, B., Cressie, N., et al. (2020), "Bayesian Inference of Spatio-Temporal Changes of Arctic Sea Ice," *Bayesian Analysis*, 15, 605–631.

Zhang, H. and El-Shaarawi, A. (2010), "On spatial skew-Gaussian processes and applications," *Environmetrics*, 21, 33–47.

Zheng, X., Kottas, A., and Sansó, B. (2021), "Nearest-Neighbor Geostatistical Models for Non-Gaussian Data," *arXiv:2107.07736*.

Zheng, X., Kottas, A., and Sansó, B. (2022), "On Construction and Estimation of Stationary Mixture Transition Distribution Models," *Journal of Computational and Graphical Statistics*, 31, 283–293.

Supplementary Material for "Bayesian Geostatistical Modeling for Discrete-Value Processes"

A Proof of Proposition 2

Proof. Consider a discrete copula NNMP characterized by

$$p(y(\boldsymbol{v}) \mid \boldsymbol{y}_{\mathrm{Ne}(\boldsymbol{v})}) = \sum_{l=1}^{L} w_l(\boldsymbol{v}) \, c_{\boldsymbol{v},l}(y(\boldsymbol{v}), y(\boldsymbol{v}_{(l)}))g_{\boldsymbol{v}}(y(\boldsymbol{v})),$$

where $g_{\boldsymbol{v}}$ is the marginal pmf of $Y(\boldsymbol{v})$.

Let $\boldsymbol{y}_{\mathcal{V}} = (y(\boldsymbol{s}_1), \ldots, y(\boldsymbol{s}_n), y(\boldsymbol{u}_1), \ldots, y(\boldsymbol{u}_m))^{\top}$ for $n \geq 2$ and $m \geq 1$, where $\mathcal{V} = \mathcal{S} \cup \mathcal{U}$, $\mathcal{S} = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n\}$, $\mathcal{U} = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m\}$, and $\mathcal{S} \cap \mathcal{U} = \emptyset$. The joint pmf of $\boldsymbol{y}_{\mathcal{V}}$ can be written as $\tilde{p}(\boldsymbol{y}_{\mathcal{V}}) = \tilde{p}(\boldsymbol{y}_{\mathcal{U}} \mid \boldsymbol{y}_{\mathcal{S}})\tilde{p}(\boldsymbol{y}_{\mathcal{S}})$. We will first derive the joint pmf $\tilde{p}(\boldsymbol{y}_{\mathcal{S}}) = \tilde{p}(y(\boldsymbol{s}_1), \ldots, y(\boldsymbol{s}_n))$ and then $\tilde{p}(\boldsymbol{y}_{\mathcal{U}} \mid \boldsymbol{y}_{\mathcal{S}})$, where $\boldsymbol{y}_{\mathcal{U}} = (y(\boldsymbol{u}_1), \ldots, y(\boldsymbol{u}_m))^{\top}$.

Let $c_{\boldsymbol{s}_i,l_i} \equiv c_{\boldsymbol{s}_i,l_i}(y(\boldsymbol{s}_i), y(\boldsymbol{s}_{(i,l_i)}))$ and $w_{\boldsymbol{s}_i,l_i} \equiv w_{l_i}(\boldsymbol{s}_i)$ with $l_i = 1, \ldots, i_L$ and $i_L = (i - 1) \wedge L$, for all i. Then

$$\tilde{p}(y(\boldsymbol{s}_1), y(\boldsymbol{s}_2)) = p(y(\boldsymbol{s}_2) \mid y(\boldsymbol{s}_1))g_{\boldsymbol{s}_1}(y(\boldsymbol{s}_1)) = c_{\boldsymbol{s}_2,1}g_{\boldsymbol{s}_2}(y(\boldsymbol{s}_2))g_{\boldsymbol{s}_1}(y(\boldsymbol{s}_1)).$$

Note that by definition of the discrete NNMP, $w_{\boldsymbol{s}_2,1} = 1$. Then

$$\begin{aligned}
\tilde{p}(y(\boldsymbol{s}_1), y(\boldsymbol{s}_2), y(\boldsymbol{s}_3)) &= p(y(\boldsymbol{s}_3) \mid y(\boldsymbol{s}_1), y(\boldsymbol{s}_2))\tilde{p}(y(\boldsymbol{s}_1), y(\boldsymbol{s}_2)) \\
&= \left(\sum_{l_3=1}^{2} w_{\boldsymbol{s}_3,l_3} c_{\boldsymbol{s}_3,l_3} g_{\boldsymbol{s}_3}(y(\boldsymbol{s}_3)) \right) c_{\boldsymbol{s}_2,1}g_{\boldsymbol{s}_2}(y(\boldsymbol{s}_2))g_{\boldsymbol{s}_1}(y(\boldsymbol{s}_1)) \\
&= \prod_{i=1}^{3} g_{\boldsymbol{s}_i}(y(\boldsymbol{s}_i)) \sum_{l_3=1}^{2} w_{\boldsymbol{s}_3,l_3} c_{\boldsymbol{s}_3,l_3} c_{\boldsymbol{s}_2,1} \\
&= \prod_{i=1}^{3} g_{\boldsymbol{s}_i}(y(\boldsymbol{s}_i)) \sum_{l_3=1}^{2} \sum_{l_2=1}^{1} w_{\boldsymbol{s}_3,l_3} w_{\boldsymbol{s}_2,l_2} c_{\boldsymbol{s}_3,l_3} c_{\boldsymbol{s}_2,l_2}.
\end{aligned}$$

Similarly, for $4 \leq n \leq L$, the joint pmf is

$$\tilde{p}(y(\boldsymbol{s}_1), \ldots, y(\boldsymbol{s}_n))$$

$$= p(y(\boldsymbol{s}_n) \,|\, \boldsymbol{y}_{\mathrm{Ne}(\boldsymbol{s}_n)}) \tilde{p}(y(\boldsymbol{s}_1), \ldots, y(\boldsymbol{s}_{n-1}))$$

$$= \left(\sum_{l_n=1}^{n-1} w_{\boldsymbol{s}_n, l_n} c_{\boldsymbol{s}_n, l_n} g_{\boldsymbol{s}_n}(y(\boldsymbol{s}_n)) \right) \left(\prod_{i=1}^{n-1} g_{\boldsymbol{s}_i}(y(\boldsymbol{s}_i)) \sum_{l_{n-1}=1}^{n-2} \cdots \sum_{l_2=1}^{1} w_{\boldsymbol{s}_{n-1}, l_{n-1}} \cdots w_{\boldsymbol{s}_2, l_2} c_{\boldsymbol{s}_{n-1}, l_{n-1}} \cdots c_{\boldsymbol{s}_2, l_2} \right)$$

$$= \prod_{i=1}^{n} g_{\boldsymbol{s}_i}(y(\boldsymbol{s}_i)) \sum_{l_n=1}^{n-1} \cdots \sum_{l_2=1}^{1} w_{\boldsymbol{s}_n, l_n} \cdots w_{\boldsymbol{s}_2, l_2} c_{\boldsymbol{s}_n, l_n} \cdots c_{\boldsymbol{s}_2, l_2}.$$

Finally, for $n > L$, it is easy to show that the joint pmf is

$$\tilde{p}(y(\boldsymbol{s}_1), \ldots, y(\boldsymbol{s}_n))$$

$$= p(y(\boldsymbol{s}_n) \,|\, \boldsymbol{y}_{\mathrm{Ne}(\boldsymbol{s}_n)}) \tilde{p}(y(\boldsymbol{s}_1), \ldots, y(\boldsymbol{s}_{n-1}))$$

$$= \left(\sum_{l_n=1}^{L} w_{\boldsymbol{s}_n, l_n} c_{\boldsymbol{s}_n, l_n} g_{\boldsymbol{s}_n}(y(\boldsymbol{s}_n)) \right)$$

$$\prod_{i=1}^{n-1} g_{\boldsymbol{s}_i}(y(\boldsymbol{s}_i)) \sum_{l_{n-1}=1}^{L} \cdots \sum_{l_{L+1}=1}^{L} \sum_{l_L=1}^{L-1} \cdots \sum_{l_2=1}^{1} w_{\boldsymbol{s}_{n-1}, l_{n-1}} \cdots w_{\boldsymbol{s}_2, l_2} c_{\boldsymbol{s}_{n-1}, l_{n-1}} \cdots c_{\boldsymbol{s}_2, l_2}$$

$$= \prod_{i=1}^{n} g_{\boldsymbol{s}_i}(y(\boldsymbol{s}_i)) \sum_{l_n=1}^{L} \cdots \sum_{l_{L+1}=1}^{L} \sum_{l_L=1}^{L-1} \cdots \sum_{l_2=1}^{1} w_{\boldsymbol{s}_n, l_n} \cdots w_{\boldsymbol{s}_2, l_2} c_{\boldsymbol{s}_n, l_n} \cdots c_{\boldsymbol{s}_2, l_2}.$$

Therefore, we have that, for $n \geq 2$, the joint pmf

$$\tilde{p}(\boldsymbol{y}_{\mathcal{S}}) = \tilde{p}(y(\boldsymbol{s}_1), \ldots, y(\boldsymbol{s}_n)) = \prod_{i=1}^{n} g_{\boldsymbol{s}_i}(y(\boldsymbol{s}_i)) \sum_{l_n=1}^{n_L} \cdots \sum_{l_2=1}^{2_L} w_{\boldsymbol{s}_n, l_n} \cdots w_{\boldsymbol{s}_2, l_2} c_{\boldsymbol{s}_n, l_n} \cdots c_{\boldsymbol{s}_2, l_2}.$$

Turning to the non-reference set $\mathcal{U}$. Let $c_{\boldsymbol{u}_i, l_i} \equiv c_{\boldsymbol{u}_i, l_i}(y(\boldsymbol{u}_i), y(\boldsymbol{u}_{(i, l_i)}))$ and $w_{\boldsymbol{u}_i, l_i} \equiv w_{l_i}(\boldsymbol{u}_i)$ with $l_i = 1, \ldots, L$, for all i. When $m = 1$, $\tilde{p}(\boldsymbol{y}_{\mathcal{U}} \,|\, \boldsymbol{y}_{\mathcal{S}}) = p(y(\boldsymbol{u}_1) \,|\, \boldsymbol{y}_{\mathrm{Ne}(\boldsymbol{u}_1)})$.

When $m \geq 2$, without loss of generality, we consider the case of $m = 2$, i.e., we take

$\mathcal{U} = \{\boldsymbol{u}_1, \boldsymbol{u}_2\}$. Then we have that

$$p(\boldsymbol{y}_\mathcal{U} \mid \boldsymbol{y}_\mathcal{S}) = p(y(\boldsymbol{u}_1) \mid \boldsymbol{y}_{\mathrm{Ne}(\boldsymbol{u}_1)}) p(y(\boldsymbol{u}_2) \mid \boldsymbol{y}_{\mathrm{Ne}(\boldsymbol{u}_2)})$$

$$= \left(\sum_{l_1=1}^{L} w_{\boldsymbol{u}_1, l_1} c_{\boldsymbol{u}_1, l_1} g_{\boldsymbol{u}_1}(y(\boldsymbol{u}_1)) \right) \left(\sum_{l_2=1}^{L} w_{\boldsymbol{u}_2, l_2} c_{\boldsymbol{u}_2, l_2} g_{\boldsymbol{u}_2}(y(\boldsymbol{u}_2)) \right)$$

$$= \prod_{i=1}^{2} g_{\boldsymbol{u}_i}(y(\boldsymbol{u}_i)) \sum_{l_2=1}^{L} \sum_{l_1=1}^{L} w_{\boldsymbol{u}_2, l_2} w_{\boldsymbol{u}_1, l_1} c_{\boldsymbol{u}_2, l_2} c_{\boldsymbol{u}_1, l_1}.$$

Obviously, the result is easily generalized for $\mathcal{U} = \{\boldsymbol{u}_1, \dots, \boldsymbol{u}_m\}$ for any $m > 2$. $\qquad\square$

B   Gaussian, Gumbel, and Clayton copulas

We introduce properties of the Gaussian, Gumbel and Clayton copulas that are useful for the discrete copula NNMP's model estimation and prediction. For more details we refer to Joe (2014). Consider a bivariate vector (X_1, X_2) with marginal cumulative distribution functions (cdfs) such that $F_1(x_1) = t_1$ and $F_2(x_2) = t_2$.

Gaussian copula   A Gaussian copula with correlation $\rho \in (0, 1)$ for (X_1, X_2) is

$$C(F_1(x_1), F_2(x_2) \mid \rho) = C(t_1, t_2 \mid \rho) = \Phi_2(\Phi^{-1}(t_1),\ \Phi^{-1}(t_2) \mid \rho).$$

If both X_1 and X_2 are continuous random variables, the copula has density

$$\frac{1}{\sqrt{1-\rho^2}} \exp\left(\frac{2\rho \Phi^{-1}(t_1) \Phi^{-1}(t_2) - \rho^2 \{(\Phi^{-1}(t_1))^2 + (\Phi^{-1}(t_2))^2)\}}{2(1-\rho^2)} \right).$$

The conditional cdf of T_1 given $T_2 = t_2$, denoted as $C_{1|2}(t_1 \mid t_2)$, is given by

$$C_{1|2}(t_1 \mid t_2) = \frac{\partial C(t_1, t_2)}{\partial t_2} = \Phi\left(\frac{\Phi^{-1}(t_1) - \rho \Phi^{-1}(t_2)}{\sqrt{1-\rho^2}} \right).$$

To simulate X_1 given $X_2 = x_2$, we first compute $t_2 = F_2(x_2)$. We then generate a random number z from a uniform distribution on $[0, 1]$, and compute $t_1 = C_{1|2}^{-1}(z \mid t_2)$ where

$C_{1|2}^{-1}(z \,|\, t_2) = \Phi\left(\sqrt{(1-\rho^2)}\Phi^{-1}(z) + \rho\Phi^{-1}(t_2)\right)$ is the inverse of $C_{1|2}(t_1 \,|\, t_2)$. Finally, we obtain x_1 from the inverse cdf $F_1^{-1}(t_1)$.

Gumbel copula   A Gumbel copula with parameter $\eta \in [1, \infty)$ for (X_1, X_2) is

$$C(F_1(x_1), F_2(x_2) \,|\, \eta) = C(t_1, t_2 \,|\, \eta) = \exp(-((-\log(t_1))^\eta + (-\log(t_2))^\eta)^{1/\eta}).$$

Let $u_1 = -\log(t_1)$ and $u_2 = -\log(t_2)$. If both X_1 and X_2 are continuous random variables, the copula has density

$$\exp(-(u_1^\eta + u_2^\eta)^{1/\eta})((u_1^\eta + u_2^\eta)^{1/\eta} + \eta - 1)(u_1^\eta + u_2^\eta)^{1/\eta-2}(u_1 u_2)^{\eta-1}(t_1 t_2)^{-1}.$$

The conditional cdf of T_1 given $T_2 = t_2$ is

$$C_{1|2}(t_1 \,|\, t_2) = \overline{C}_{1|2}(u_1 \,|\, u_2) = t_2^{-1} \exp(-(u_1^\eta + u_2^\eta)^{1/\eta})(1 + (u_1/u_2)^\eta)^{1/\eta-1},$$

where the conditional cdf $\overline{C}_{1|2}(u_1 \,|\, u_2)$ corresponds to the copula $\overline{C}(u_1, u_2 \,|\, \eta) = \exp(-(u_1^\eta + u_2^\eta)^{1/\eta})$ which is a bivariate exponential survival function, with marginals corresponding to a unit rate exponential distribution. The inverse conditional cdf $C_{1|2}^{-1}(\cdot \,|\, t_2)$ does not have a closed form. To generate X_1 given $X_2 = x_2$, following Joe (2014), we first define $y = (u_1^\eta + u_2^\eta)^{1/\eta}$. Then we have a realization of X_1, say $x_1 = (y_0^\eta - u_2^\eta)^{1/\eta}$, where y_0 is the root of $h(y) = y + (\eta - 1)\log(y) - (u_2 + (\eta - 1)\log(u_2) - \log z) = 0$, where $y \geq u_2$, and z is a random number generated from a uniform distribution on $[0, 1]$.

Clayton copula   A Clayton copula with parameter $\delta \in [0, \infty)$ for (X_1, X_2) is

$$C(F_1(x_1), F_2(x_2) \,|\, \delta) = C(t_1, t_2 \,|\, \delta) = (t_1^{-\delta} + t_2^{-\delta} - 1)^{-1/\delta}.$$

If both X_1 and X_2 are continuous random variables, the copula has density

$$(1 + \delta)(t_1 t_2)^{-\delta-1}(t_1^{-\delta} + t_2^{-\delta} - 1)^{-2-1/\delta}.$$

The conditional cdf of T_1 given $T_2 = t_2$ is

$$C_{1|2}(t_1 \,|\, t_2) = (1 + t_2^{\delta}(t_1^{-\delta} - 1))^{-1-1/\delta}.$$

To simulate X_1 given X_2, we first compute $t_2 = F_2(x_2)$, and generate a uniform random number z on $[0,1]$. Then we compute $t_1 = C_{1|2}^{-1}(z \,|\, t_2)$ where $C_{1|2}^{-1}(z \,|\, t_2) = ((z^{-\delta/(1+\delta)} - 1)t_2^{-\delta} + 1)^{-1/\delta}$. Finally, we obtain x_1 from the inverse cdf $F_1^{-1}(t_1)$.

C   Implementation details

In this section, we introduce necessary posterior simulation steps for the Poisson NNMP (PONNMP) and negative binomial NNMP (NBNNMP) models illustrated in the data examples. For both models, we use an exponential correlation function with range parameter ϕ to create spatial copulas. More specifically, given two different sites $\boldsymbol{v} \neq \boldsymbol{v}'$, the link functions for parameters of the Gaussian, Gumbel and Clayton copulas, respectively, are

$$
\begin{aligned}
\rho(||\boldsymbol{v} - \boldsymbol{v}'||) &= \exp(-||\boldsymbol{v} - \boldsymbol{v}'||/\phi), \\
\eta(||\boldsymbol{v} - \boldsymbol{v}'||) &= \min\{(1 - \exp(-||\boldsymbol{v} - \boldsymbol{v}'||/\phi))^{-1}, 50\}, \\
\delta(||\boldsymbol{v} - \boldsymbol{v}'||) &= \min\{2\exp(-||\boldsymbol{v} - \boldsymbol{v}'||/\phi)/(1 - \exp(-||\boldsymbol{v} - \boldsymbol{v}'||/\phi)), 98\},
\end{aligned}
$$

where the upper bounds 50 and 98 for Gumbel and Clayton copulas are chosen for numerical stability. When $\eta(d_0) = 50$, $\exp(-d_0/\phi) = 0.98$. Similarly, when $\delta(d_0) = 98$, $\exp(-d_0/\phi) = 0.98$. Both link functions imply that given ϕ, the dependence implied by the copulas stays the same for any distance between $\boldsymbol{v}$ and $\boldsymbol{v}'$ smaller than d_0.

We assume that $\boldsymbol{y}_{\mathcal{S}} = (y(\boldsymbol{s}_1), \ldots, y(\boldsymbol{s}_n))^{\top}$ is a vector of observations, where $\mathcal{S} = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n\}$ is the reference set. Each $y(\boldsymbol{s}_i)$ is associated with $y^*(\boldsymbol{s}_i)$ such that $y^*(\boldsymbol{s}_i) = y(\boldsymbol{s}_i) - o_i$, where $o_i \equiv o(\boldsymbol{s}_i)$, $o(\boldsymbol{s}_i) \overset{i.i.d.}{\sim} \text{Unif}(0,1)$, for $i = 1, \ldots, n$. The auxiliary variables o_i is independent of $y(\boldsymbol{s}_i)$ and of o_j for $j \neq i$. Let $\boldsymbol{y}^*_{\text{Ne}(\boldsymbol{s}_i)} = (y^*(\boldsymbol{s}_{(i1)}), \ldots, y^*(\boldsymbol{s}_{(i,i_L)}))^{\top}$ and $\boldsymbol{o}_{\text{Ne}(\boldsymbol{s}_i)} = (o(\boldsymbol{s}_{(i1)}), \ldots, o(\boldsymbol{s}_{(i,i_L)}))^{\top}$, for $i = 2, \ldots, n$.

C.1 Poisson NNMP models and inference

The conditional density of the continued Poisson NNMP (PONNMP) over the reference set is given by

$$p(y^*(\boldsymbol{s}_i) \,|\, \boldsymbol{y}^*_{\mathrm{Ne}(\boldsymbol{s}_i)}, o(\boldsymbol{s}_i), \boldsymbol{o}_{\mathrm{Ne}(\boldsymbol{s}_i)}) = \sum_{l=1}^{i_L} w_l(\boldsymbol{s}_i)\, c^*_{\boldsymbol{s}_i,l}(y^*(\boldsymbol{s}_i), y^*(\boldsymbol{s}_{(il)}))\, f^*_Y(y^*(\boldsymbol{s}_i)),$$

for $i = 2, \ldots, n$, where $f^*_Y(y^*(\boldsymbol{s}_i)) = f_Y([y^*(\boldsymbol{s}_i) + 1])$, and f_Y is a Poisson distribution with rate parameter λ. The component $c^*_{\boldsymbol{s}_i,l}$ is the copula density of a spatial copula. We will illustrate the posterior inference using the Gaussian case as an example. The copula density of the spatial Gaussian copula is given by

$$c^*_{\boldsymbol{s}_i,l}(y^*(\boldsymbol{s}_i), y^*(\boldsymbol{s}_{(il)})) =$$
$$\frac{1}{\sqrt{1 - (\rho_l(\boldsymbol{s}_i))^2}} \exp\left(\frac{2\rho(\boldsymbol{s}_i)\Phi^{-1}(q_i)\Phi^{-1}(q_{il}) - (\rho_l(\boldsymbol{s}_i))^2\{(\Phi^{-1}(q_i))^2 + (\Phi^{-1}(q_{il}))^2\}}{2(1 - (\rho_l(\boldsymbol{s}_i))^2)} \right),$$

where $\rho_l(\boldsymbol{s}_i) \equiv \rho(||\boldsymbol{s}_i - \boldsymbol{s}'_{(il)}||) = \exp(-||\boldsymbol{s}_i - \boldsymbol{s}_{(il)}||/\phi)$, $q_i = F^*_Y(y^*(\boldsymbol{s}_i))$, $q_{il} = F^*_Y(y^*(\boldsymbol{s}_{(il)}))$, and F^*_Y is the cdf of f^*_Y.

The formulation of the mixture weights allows us to augment the model with a sequence of auxiliary variables t_i, $i = 3, \ldots, n$, where t_i is a Gaussian random variable with mean $\mu(\boldsymbol{s}_i) = \gamma_0 + s_{i1}\gamma_1 + s_{i2}\gamma_2$ and variance κ^2. The conditional density of the augmented model on $y^*(\boldsymbol{s}_i)$ is

$$p(y^*(\boldsymbol{s}_i) \,|\, \boldsymbol{y}^*_{\mathrm{Ne}(\boldsymbol{s}_i)}, o(\boldsymbol{s}_i), \boldsymbol{o}_{\mathrm{Ne}(\boldsymbol{s}_i)}) = \sum_{l=1}^{i_L} c^*_{\boldsymbol{s}_i,l}(y^*(\boldsymbol{s}_i), y^*(\boldsymbol{s}_{(il)}))\, f^*_Y(y^*(\boldsymbol{s}_i))\, \mathbb{1}_{(r^*_{\boldsymbol{s}_i,l-1}, r^*_{\boldsymbol{s}_i,l})}(t_i),$$

for $i = 3, \ldots, n$, where $r^*_{\boldsymbol{s}_i,l} = \log(r_{\boldsymbol{s}_i,l}/(1 - r_{\boldsymbol{s}_i,l}))$ for $l = 1, \ldots, i_L$. The random cutoff points $r_{\boldsymbol{s}_i,l}$ is defined such that $r_{\boldsymbol{s}_i,l} - r_{\boldsymbol{s}_i,l-1} = k'(\boldsymbol{s}_i, \boldsymbol{s}_{(il)})/\sum_{l=1}^{i_L} k'(\boldsymbol{s}_i, \boldsymbol{s}_{(il)})$, where $k'(\boldsymbol{s}_i, \boldsymbol{s}_{(il)}) = \exp(||\boldsymbol{s}_i - \boldsymbol{s}_{(il)}||/\zeta)$.

Let $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \gamma_2)$. The Bayesian model is completed with prior specifications for parameters $(\lambda, \phi, \zeta, \boldsymbol{\gamma}, \kappa^2)$. Let $f^*_{\boldsymbol{s}_i,l}(y^*(\boldsymbol{s}_i) \,|\, y^*(\boldsymbol{s}_{(il)})) = c^*_{\boldsymbol{s}_i,l}(y^*(\boldsymbol{s}_i), y^*(\boldsymbol{s}_{(il)}))\, f^*_Y(y^*(\boldsymbol{s}_i))$. With customary prior specifications, the posterior distribution of the parameters and latent vari-

ables can be written as

$$\text{Ga}(\lambda \,|\, u_\lambda, v_\lambda) \times \text{IG}(\phi \,|\, u_\phi, v_\phi) \times \text{IG}(\zeta \,|\, u_\zeta, v_\zeta) \times N(\boldsymbol{\gamma} \,|\, \boldsymbol{\mu}_\gamma, \boldsymbol{V}_\gamma) \times \text{IG}(\kappa^2 \,|\, u_{\kappa^2}, v_{\kappa^2})$$

$$\times\, N(\boldsymbol{t} \,|\, \boldsymbol{D}\boldsymbol{\gamma}, \kappa^2 \mathbf{I}_{n-2})) \times f_Y^*(y(\boldsymbol{s}_1) - o_1 \,|\, \lambda) \times f_{\boldsymbol{s}_2,1}^*(y(\boldsymbol{s}_2) - o_2 \,|\, y(\boldsymbol{s}_1) - o_1, \lambda, \phi)$$

$$\times\, \prod_{i=1}^{n} \text{Unif}(o_i \,|\, 0, 1) \times \prod_{i=3}^{n} \sum_{l=1}^{i_L} f_{\boldsymbol{s}_i,l}^*(y(\boldsymbol{s}_i) - o_i \,|\, y(\boldsymbol{s}_{(il)}) - o_{(il)}, \lambda, \phi) \mathbb{1}_{(r_{\boldsymbol{s}_i,l-1}^*, r_{\boldsymbol{s}_i,l}^*)}(t_i),$$

where $o_{(il)} \equiv o(\boldsymbol{s}_{(il)})$, the vector $\boldsymbol{t} = (t_3, \ldots, t_n)^\top$, and the matrix $\boldsymbol{D}$ is $(n-2) \times 3$ such that the ith row is $(1, s_{2+i,1}, s_{2+i,2})$.

The Monte Carlo Markov chain (MCMC) algorithm to obtain posterior samples consists of updates from the posterior full conditional distribution of each of $(\lambda, \phi, \zeta, \boldsymbol{\gamma}, \kappa^2, \{t_i\}_{i=3}^n, \{o_i\}_{i=1}^n)$. The posterior full conditional distributions of each of $(\boldsymbol{\gamma}, \kappa^2, \{t_i\}_{i=3}^n, \{o_i\}_{i=1}^n)$ are described in the main paper. We focus on the posterior updates for (λ, ϕ, ζ). Note that there is a set of configuration variables $\{\ell_i\}_{i=3}^n$ in one-to-one correspondence with t_i, i.e., $\ell_i = l$ if and only if $t_i \in (r_{\boldsymbol{s}_i,l-1}^*, r_{\boldsymbol{s}_i,l}^*)$, for $l = 1, \ldots, i_L$. We take $\ell_2 = 1$. The posterior full conditional distributions of λ and ϕ are proportional to $\text{Ga}(\lambda \,|\, u_\lambda, v_\lambda) f_Y^*(y(\boldsymbol{s}_1) - o_1) \prod_{i=2}^n f_{\boldsymbol{s}_i,\ell_i}^*(y(\boldsymbol{s}_i) - o_i \,|\, y(\boldsymbol{s}_{(i,\ell_i)}) - o_{(i,\ell_i)})$ and $\text{IG}(\phi \,|\, u_\phi, v_\phi) \prod_{i=2}^n c_{\boldsymbol{s}_i,\ell_i}^*(y(\boldsymbol{s}_i) - o_i, y(\boldsymbol{s}_{(i,\ell_i)}) - o_{(i,\ell_i)})$, respectively. For each parameter, we update it on its log scale with a random walk Metropolis step. To update ζ, we first marginalize out the latent variables t_i from the joint posterior distribution. The posterior full conditional distribution of ζ is proportional to $\text{IG}(\zeta \,|\, u_\zeta, v_\zeta) \prod_{i=3}^n \{G_{\boldsymbol{s}_i}(r_{\boldsymbol{s}_i,\ell_i} \,|\, \mu(\boldsymbol{s}_i), \kappa^2) - G_{\boldsymbol{s}_i}(r_{\boldsymbol{s}_i,\ell_i-1} \,|\, \mu(\boldsymbol{s}_i), \kappa^2)\}$. We update ζ on its log scale with a random walk Metropolis step.

C.2   Negative binomial NNMP models and inference

The conditional density of the continued negative binomial NNMP (NBNNMP) over the reference set is given by

$$p(y^*(\boldsymbol{s}_i) \,|\, \boldsymbol{y}_{\text{Ne}(\boldsymbol{s}_i)}^*, o(\boldsymbol{s}_i), \boldsymbol{o}_{\text{Ne}(\boldsymbol{s}_i)}) = \sum_{l=1}^{i_L} w_l(\boldsymbol{s}_i) c_{\boldsymbol{s}_i,l}^*(y^*(\boldsymbol{s}_i), y^*(\boldsymbol{s}_{(il)})) g_{\boldsymbol{s}_i}(y^*(\boldsymbol{s}_i)),$$

for $i = 2, \ldots, n$, where $g_{\boldsymbol{s}_i}^*(y^*(\boldsymbol{s}_i)) = g_{\boldsymbol{s}_i}([y^*(\boldsymbol{s}_i) + 1])$, and $g_{\boldsymbol{s}_i}$ is a negative binomial distribution with mean $\alpha(\boldsymbol{s}_i) = \exp(\boldsymbol{x}(\boldsymbol{s}_i)^\top \boldsymbol{\beta})$ and dispersion parameter r. Similar to the Poisson case, we illustrate the posterior inference using a spatial Gaussian copula with copula density given by

$$
c_{\boldsymbol{s}_i,l}^*(y^*(\boldsymbol{s}_i), y^*(\boldsymbol{s}_{(il)})) =
$$
$$
\frac{1}{\sqrt{1 - (\rho_l(\boldsymbol{s}_i))^2}} \exp\left(\frac{2\rho(\boldsymbol{s}_i)\Phi^{-1}(q_i)\Phi^{-1}(q_{il}) - (\rho_l(\boldsymbol{s}_i))^2\{(\Phi^{-1}(q_i))^2 + (\Phi^{-1}(q_{(il)}))^2\}}{2(1 - (\rho_l(\boldsymbol{s}_i))^2)}\right),
$$

where $\rho_l(\boldsymbol{s}_i) \equiv \rho(||\boldsymbol{s}_i - \boldsymbol{s}'_{(il)}||) = \exp(-||\boldsymbol{s}_i - \boldsymbol{s}_{(il)}||/\phi)$, $q_i = Q_{\boldsymbol{s}_i}^*(y^*(\boldsymbol{s}_i))$, $q_{il} = Q_{\boldsymbol{s}_{(il)}}^*(y^*(\boldsymbol{s}_{(il)}))$, and $Q_{\boldsymbol{s}_i}^*$ is the cdf of $g_{\boldsymbol{s}_i}^*$ for all $\boldsymbol{s}_i$.

Similarly, we use an exponential correlation function for the cutoff point kernel k', and augment the model with a set of Gaussian random variables t_i with mean $\mu(\boldsymbol{s}_i)$ and κ^2. Let $f_{\boldsymbol{s}_i,l}^*(y^*(\boldsymbol{s}_i) \,|\, y^*(\boldsymbol{s}_{(il)})) = c_{\boldsymbol{s}_i,l}^*(y^*(\boldsymbol{s}_i), y^*(\boldsymbol{s}_{(il)}))g_{\boldsymbol{s}_i}^*(y^*(\boldsymbol{s}_i))$. With customary prior specifications, the joint posterior distribution is given by

$$
N(\boldsymbol{\beta} \,|\, \boldsymbol{\mu}_\beta, \boldsymbol{V}_\beta) \times \mathrm{IG}(r \,|\, u_r, v_r) \times \mathrm{IG}(\phi \,|\, u_\phi, v_\phi) \times \mathrm{IG}(\zeta \,|\, u_\zeta, v_\zeta) \times N(\boldsymbol{\gamma} \,|\, \boldsymbol{\mu}_\gamma, \boldsymbol{V}_\gamma) \times \mathrm{IG}(\kappa^2 \,|\, u_{\kappa^2}, v_{\kappa^2})
$$
$$
\times N(\boldsymbol{t} \,|\, \boldsymbol{D}\boldsymbol{\gamma}, \kappa^2 \mathbf{I}_{n-2})) \times g_{\boldsymbol{s}_1}^*(y(\boldsymbol{s}_1) - o_1 \,|\, \boldsymbol{\beta}, r) \times f_{\boldsymbol{s}_2,1}^*(y(\boldsymbol{s}_2) - o_2 \,|\, y(\boldsymbol{s}_1) - o_1, \boldsymbol{\beta}, r, \phi)
$$
$$
\times \prod_{i=1}^{n} \mathrm{Unif}(o_i \,|\, 0, 1) \times \prod_{i=3}^{n} \sum_{l=1}^{i_L} f_{\boldsymbol{s}_i,l}^*(y(\boldsymbol{s}_i) - o_i \,|\, y(\boldsymbol{s}_{(il)}) - o_{(il)}, \boldsymbol{\beta}, r, \phi) \mathbb{1}_{(r_{\boldsymbol{s}_i,l-1}^*, r_{\boldsymbol{s}_i,l}^*)}(t_i),
$$

where $o_{(il)} \equiv o(\boldsymbol{s}_{(il)})$, the vector $\boldsymbol{t} = (t_3, \ldots, t_n)^\top$, and the matrix $\boldsymbol{D}$ is $(n-2) \times 3$ such that the ith row is $(1, s_{2+i,1}, s_{2+i,2})$.

The MCMC algorithm to obtain posterior samples consists of updates from the posterior full conditional distribution of each of $(\boldsymbol{\beta}, r, \phi, \zeta, \boldsymbol{\gamma}, \kappa^2, \{t_i\}_{i=3}^n, \{o_i\}_{i=1}^n)$. The posterior full conditional distributions of each of $(\boldsymbol{\gamma}, \kappa^2, \{t_i\}_{i=3}^n, \{o_i\}_{i=1}^n)$ are described in the main paper. We focus on the posterior updates for $(\boldsymbol{\beta}, r, \phi, \zeta)$. Note that there is a set of configuration variables $\{\ell_i\}_{i=3}^n$ in one-to-one correspondence with t_i, i.e., $\ell_i = l$ if and only if $t_i \in (r_{\boldsymbol{s}_i,l-1}^*, r_{\boldsymbol{s}_i,l}^*)$, for $l = 1, \ldots, i_L$. We take $\ell_2 = 1$. The posterior full conditional distributions of $\boldsymbol{\beta}$ and r are proportional to $N(\boldsymbol{\beta} \,|\, \boldsymbol{\mu}_\beta, \boldsymbol{V}_\beta)g_{\boldsymbol{s}_1}^*(y(\boldsymbol{s}_1) - o_1) \prod_{i=2}^n f_{\boldsymbol{s}_i,\ell_i}^*(y(\boldsymbol{s}_i) - o_i \,|\, y(\boldsymbol{s}_{(i,\ell_i)}) - o_{(i,\ell_i)})$ and $\mathrm{IG}(r \,|\, u_r, v_r)g_{\boldsymbol{s}_1}^*(y(\boldsymbol{s}_1) - o_1) \prod_{i=2}^n f_{\boldsymbol{s}_i,\ell_i}^*(y(\boldsymbol{s}_i) -$

$o_i \,|\, y(\boldsymbol{s}_{(i,\ell_i)}) - o_{(i,\ell_i)})$, respectively. We use a random walk Metropolis step to update $\boldsymbol{\beta}$ and r on its log scale, respectively. The posterior full conditional distribution of ϕ is proportional to $\text{IG}(\phi \,|\, u_\phi, v_\phi) \prod_{i=2}^n c^*_{\boldsymbol{s}_i, \ell_i}(y(\boldsymbol{s}_i) - o_i, y(\boldsymbol{s}_{(i,\ell_i)}) - o_{(i,\ell_i)})$. We update ϕ on its log scale with a random walk Metropolis step. To update ζ, we first marginalize out the latent variables t_i from the joint posterior distribution. The posterior full conditional distribution of ζ is proportional to $\text{IG}(\zeta \,|\, u_\zeta, v_\zeta) \prod_{i=3}^n \{G_{\boldsymbol{s}_i}(r_{\boldsymbol{s}_i, \ell_i} \,|\, \mu(\boldsymbol{s}_i), \kappa^2) - G_{\boldsymbol{s}_i}(r_{\boldsymbol{s}_i, \ell_i - 1} \,|\, \mu(\boldsymbol{s}_i), \kappa^2)\}$. We update ζ on its log scale with a random walk Metropolis step.

D   Additional simulation and model checking results

This section presents additional results of the data examples in the main paper. In particular, Section D.1 corresponds to the first simulation experiment. Section D.2 investigates the mixture weights and neighborhood sizes of the Gaussian copula NBNNMP, compares three discrete copula NBNNMPs, and compares the Gaussian copula NBNNMP with the SGLMM-GP, for the real data example. Section D.3 shows the model checking results.

D.1   First simulation experiment

Figure 1 shows the predicted random fields, given by the three discrete copula NNMP models with Poisson stationary marginals, under different scenarios.

As discussed in the main paper, the Clayton model was not able to recover large values. The Gumbel model seems to recover large values slightly better than the Gaussian model.

D.2   North American Breeding Bird Survey data analysis

D.2.1   Analysis of L

We applied the Gaussian copula NBNNMP model to the whole data set with $L = 5, 10, 15, 20$. For each L, we ran the MCMC algorithm for 30000 iterations, discarding the first 10000 iterations, and collecting posterior samples every 5th iteration.

Table 1 provides the posterior means and 95% CI estimates of the model parameters. They were quite robust across different values of L, except for those of ϕ and ζ, even though

Table 1: BBS data analysis: posterior means and 95% CI estimates for the parameters and computing time, under the Gaussian copula NBNNMP models with different values of L.

	L = 5	L = 10	L = 15	L = 20
β_0	6.52 (5.88, 7.33)	6.56 (5.69, 7.22)	6.48 (5.72, 7.28)	6.48 (5.62, 7.29)
β_1	-0.09 (-0.11, -0.07)	-0.09 (-0.11, -0.06)	-0.09 (-0.11, -0.07)	-0.09 (-0.11, -0.06)
ϕ	1.61 (1.26, 2.04)	2.51 (1.80, 3.47)	2.65 (1.93, 3.59)	2.62 (1.81, 3.68)
ζ	0.82 (0.45, 1.82)	1.10 (0.63, 2.15)	1.37 (0.77, 2.70)	1.71 (0.87, 3.80)
r	1.94 (1.65, 2.22)	1.86 (1.51, 2.19)	1.87 (1.54, 2.21)	1.88 (1.53, 2.22)
γ_0	-1.28 (-3.60, 0.96)	-1.29 (-3.49, 1.01)	-1.51 (-3.77, 0.66)	-1.69 (-3.85, 0.41)
γ_1	0.00 (-0.02, 0.03)	0.00 (-0.02, 0.02)	0.00 (-0.02, 0.02)	0.00 (-0.02, 0.02)
γ_2	0.03 (-0.01, 0.08)	0.02 (-0.02, 0.06)	0.01 (-0.02, 0.06)	0.01 (-0.02, 0.05)
κ^2	2.39 (1.48, 3.65)	2.23 (1.46, 3.31)	1.93 (1.24, 2.95)	1.63 (1.09, 2.30)
Time (mins)	29.17	32.71	38.49	50.91

the different credible intervals have substantial overlap. Note that ϕ and ζ are the range parameters of the exponential correlation functions for the Gaussian copula correlation and for the cutoff point kernel, respectively. Since a model with a large value of L includes more distant neighbors, ϕ and ζ should be larger as they indicate effective ranges.

To examine the model performance on estimating the weights, we randomly selected ten locations $(\boldsymbol{s}_{j_1}, \ldots, \boldsymbol{s}_{j_{10}})$ such that $21 \leq j_k \leq 200$ for $k = 1, \ldots, 5$ and $1312 \leq j_k \leq 1512$ for $k = 6, \ldots, 10$. Since we used random ordering to assign indices to the locations, the neighbors of $\boldsymbol{s}_{j_k}$, $k = 1, \ldots, 5$, may consist of distant locations, whereas the neighbors of $\boldsymbol{s}_{j_k}$, $k = 6, \ldots, 10$, were expected to be all nearby. Figures 2 and 3 illustrate the posterior means and 95% CI estimates of the weights at these ten locations. From the figures, we see that the model provided estimates of the weights that adjust to different neighborhood structures. The effective number of neighbors varied across locations. In addition, the estimates of the weights were quite robust as the value of L increased. We can observe that the model was able to penalize irrelevant neighbors by assigning very small probabilities. While $L = 5$ seems too small to work as an upper bound, we observe that when L ranged from 10 to 20, the effective number of weights for each location was quite consistent.

Finally, a sensitivity analysis was carried out to study the impact of L on the model performance. We randomly split the data into two sets, a training set with 1212 observations and a testing set with 300 observations. We then applied the Gaussian copula NBNNMP

Table 2: BBS data analysis: performance metrics of the Gaussian copula NBNNMP models with different values of L.

	RMSPE	95% CI	95% CI width	CRPS	ES	VS
L = 5	19.90	0.93	66.07	9.79	235.34	39759593
L = 6	19.82	0.94	65.91	9.75	234.50	39446330
L = 7	19.83	0.94	66.04	9.75	234.73	39464801
L = 8	19.80	0.94	66.19	9.75	234.36	39345232
L = 9	19.75	0.94	66.33	9.72	233.42	39073447
L = 10	19.72	0.94	66.27	9.72	233.50	39066501
L = 11	19.74	0.94	66.40	9.73	233.75	39179711
L = 12	19.73	0.95	66.67	9.70	233.10	38919544
L = 13	19.73	0.94	66.50	9.71	233.29	38978258
L = 14	19.70	0.95	66.69	9.71	233.20	38920854
L = 15	19.72	0.95	66.70	9.71	233.26	38865662
L = 16	19.73	0.94	66.70	9.72	233.50	38998533
L = 17	19.72	0.95	66.67	9.72	233.55	38982480
L = 18	19.72	0.94	66.80	9.72	233.63	39013058
L = 19	19.74	0.94	66.67	9.72	233.94	39111633
L = 20	19.79	0.94	66.75	9.74	234.30	39194713

with L from 5 to 20, and evaluated the model performance based on out-of-sample predictive performance as shown in Table 2. There were no discernible differences among the models with L between 9 and 20. The conclusion from the robustness analysis of the choice of L is that $L = 20$ works as a reasonable upper bound for this particular data example.

D.2.2 Comparison of three copula NBNNMP models

We compare three discrete copula NBNNMP models with $L = 20$. Each model used either the spatial Gaussian, Gumbel or Clayton copulas, with negative binomial marginals $\mathrm{NB}(\exp(\boldsymbol{x}(\boldsymbol{v})^\top\boldsymbol{\beta}), r)$. We used the same link functions and prior specifications for copulas as in Section 5.1 of the main paper and the same priors for other parameters as in Section 5.2 of the main paper. We fitted the models to 1215 randomly selected observations and used the remaining 300 for model comparison. For each model, we ran the MCMC algorithm for 30000 iterations, discarding the first 10000 iterations, and collected posterior samples every 5th iteration. Table 3 shows the comparison based on out-of-sample predictive performance. Overall, the Gaussian copula outperformed the other two.

Table 3: BBS data analysis: performance metrics for NBNNMPs based on different copulas.

	RMSPE	95%CI cover	95%CI width	CRPS	ES	VS
Gaussian	19.75	0.94	66.62	9.72	233.91	39136486
Gumbel	19.71	0.96	68.77	9.81	236.18	39665090
Clayton	19.97	0.93	71.51	9.91	237.21	39566563

D.2.3 Comparison with the SGLMM method

We also assessed the model performance by comparison with the SGLMM-GP model. Again, we randomly split the data into a training set with 1212 observations and a testing set with 300 observations. We ran the MCMC algorithm for the Gaussian copula NBNNMP ($L = 20$) for 30000 iterations, discarding the first 10000 iterations, and collecting posterior samples every 5th iteration. Since the MCMC for SGLMM-GP involves sampling the spatial random effects, we ran the algorithm for 50000 iterations and collected posterior samples every five iterations, with the first 30000 as burn-in.

Table 4 shows the parameter estimates and predictive performances by the two models. The regression coefficient estimates were quite similar. Both models indicate an increasing trend in the counts as the latitude decreases. Regarding the out-of-sample predictive performance, the NBNNMP model performed uniformly better than the SGLMM-GP, with a huge gain in computing time.

Table 4: BBS data analysis: parameter estimates and performance metrics of the Gaussian copula NBNNMP and the SGLMM-GP models.

	NBNNMP	SGLMM-GP
β_0	6.57 (5.83, 7.19)	6.67 (6.55, 6.81)
β_1	-0.09 (-0.10, -0.07)	-0.10 (-0.10, -0.09)
RMSPE	19.79	20.41
95% CI	0.94	0.94
95% CI width	66.56	76.56
CRPS	9.74	10.10
ES	234.22	239.02
VS	39204378.76	40185343.15
Time (mins)	37.56	4208.33

D.3 Randomized quantile residual analysis

Model checking results using randomized quantile residuals for simulation examples 1 and 2 are illustrated in Figures 4, 5, 6, and Figure 7, respectively. For simulation example 1, each figure corresponds to a scenario and contains posterior summary of the Gaussian quantile-quantile plot, the histogram and spatial plot of the posterior means of the residuals. We can see that in all cases, the results indicate good model fits.

Additional References

Joe, H. (2014). *Dependence modeling with copulas*. Boca Raton, FL: CRC press.

(a) True $y(\boldsymbol{v})$ ($\sigma_1 = 1$) (b) True $y(\boldsymbol{v})$ ($\sigma_1 = 3$) (c) True $y(\boldsymbol{v})$ ($\sigma_1 = 10$)

(d) PONNMP (Gaussian) (e) PONNMP (Gaussian) (f) PONNMP (Gaussian)

(g) PONNMP (Gumbel) (h) PONNMP (Gumbel) (i) PONNMP (Gumbel)

(j) PONNMP (Clayton) (k) PONNMP (Clayton) (l) PONNMP (Clayton)
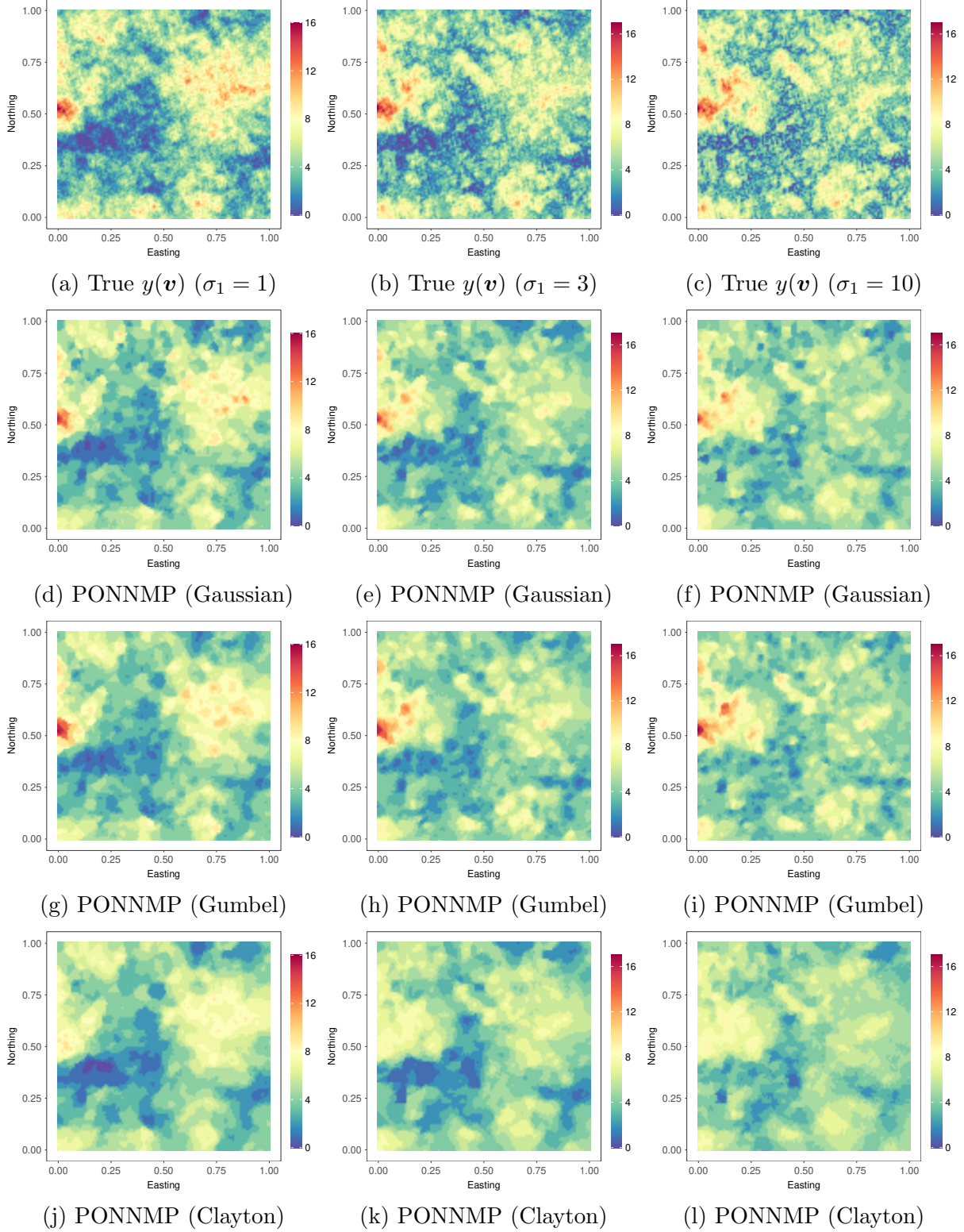
Figure 1: Simulated data example 1. Interpolated surfaces of the true model (first row), and posterior median estimates of the Poisson NNMP (PONNMP) models using Gaussian (second row), Gumbel (third row), and Clayton (fourth row) copulas. Columns from left to right correspond to scenarios with $\sigma_1 = 1, 3, 10$, respectively.
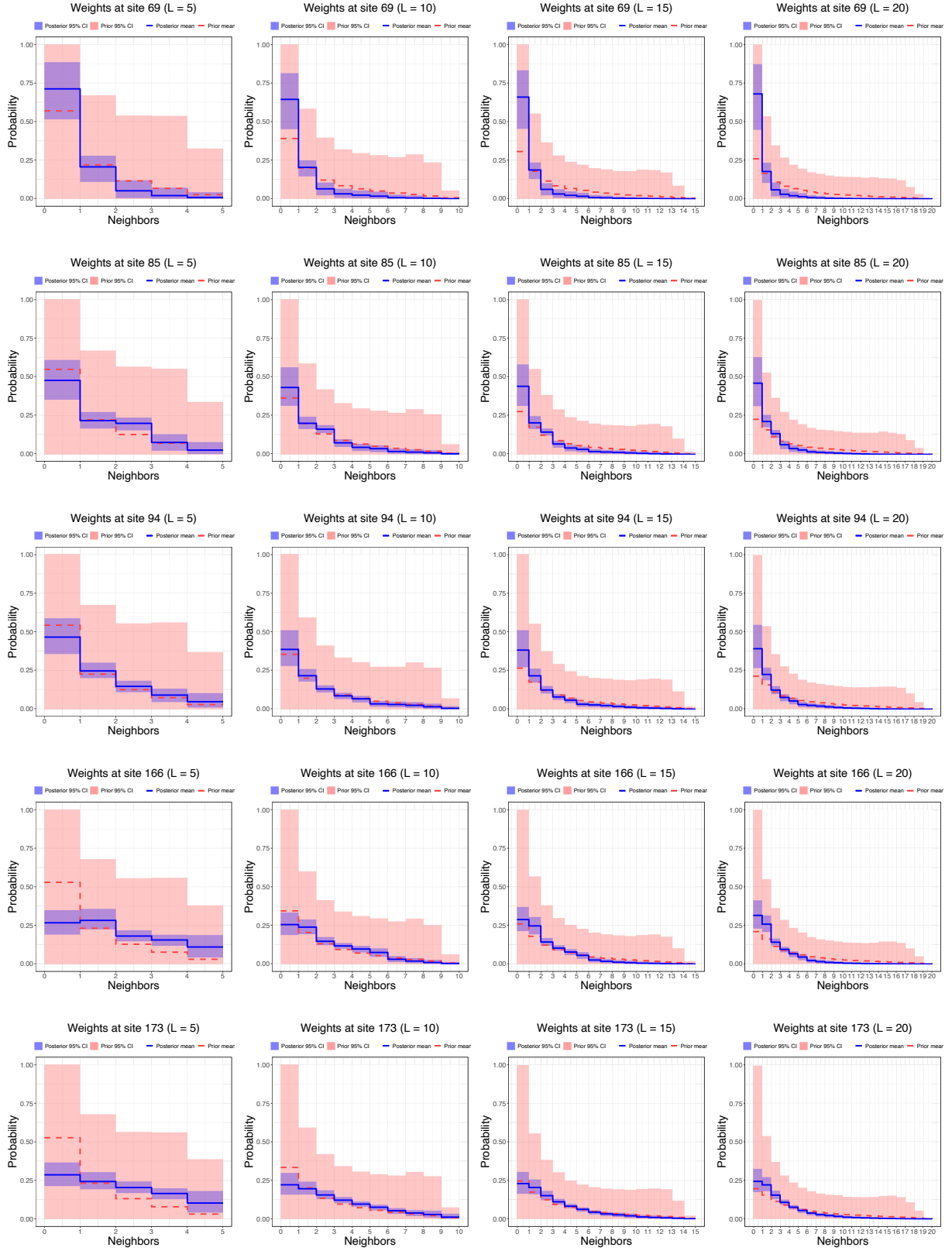
Figure 2: BBS data analysis: Posterior means and 95% CI estimates of the weights of the first five locations.
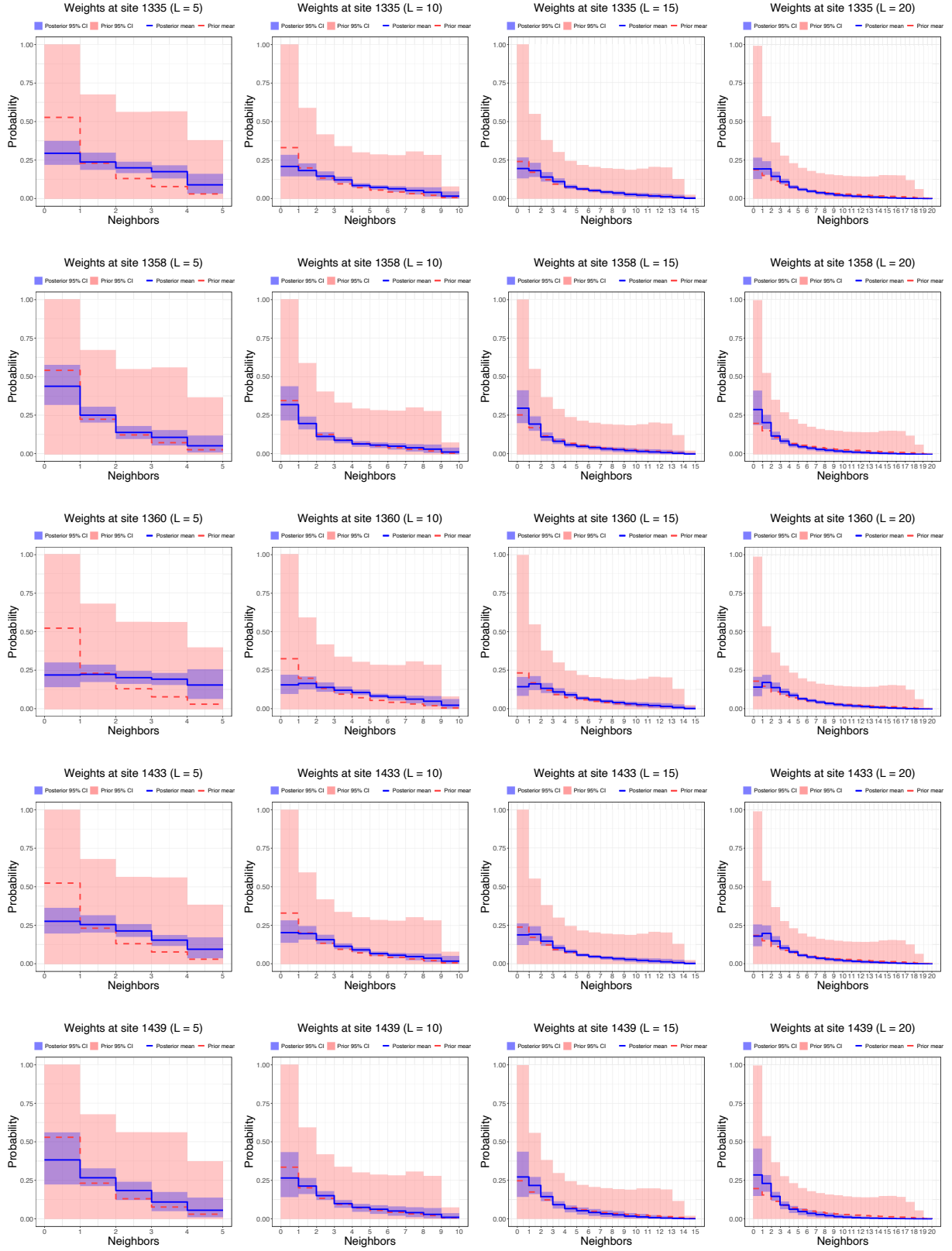
Figure 3: BBS data analysis: Posterior means and 95% CI estimates of the weights of the last five locations.
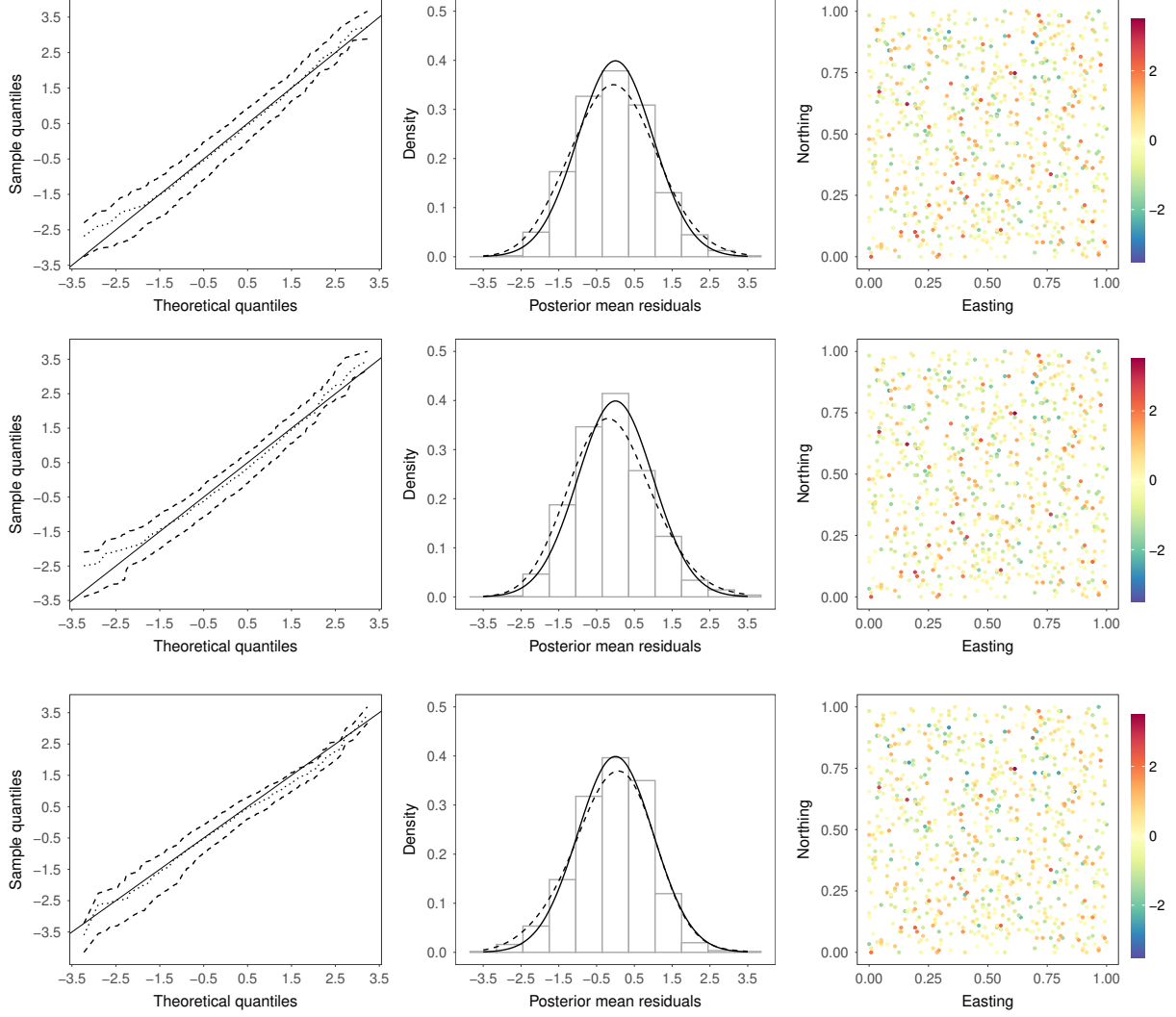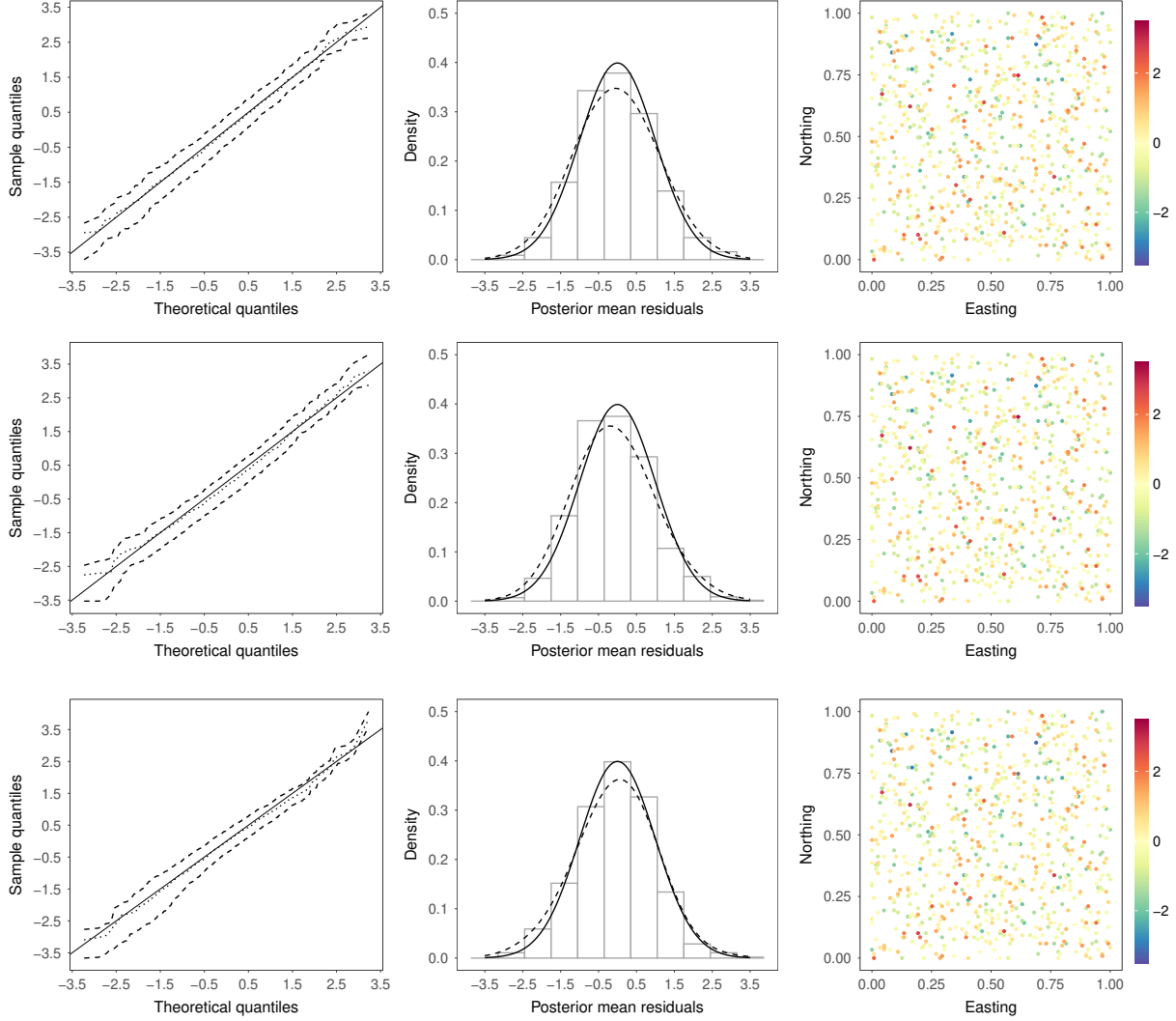
Figure 4: Simulated data example 1 - randomized quantile residual analysis for Scenario 1 ($\sigma_1 = 1$). Left column: Gaussian quantile-quantile plots. Dotted and dashed lines correspond to the posterior mean and 95% interval bands, respectively. Middle column: Histograms of the posterior means of the residuals. Solid and dashed lines are the standard Gaussian density and the kernel density estimate of the posterior means of the residuals, respectively. Right column: spatial plots of the posterior means of the residuals. Rows from top to bottom correspond to the Gaussian, Gumbel, and Clayton models, respectively.

Figure 5: Simulated data example 1 - randomized quantile residual analysis for Scenario 2 ($\sigma_1 = 3$). Left column: Gaussian quantile-quantile plots. Dotted and dashed lines correspond to the posterior mean and 95% interval bands, respectively. Middle column: Histograms of the posterior means of the residuals. Solid and dashed lines are the standard Gaussian density and the kernel density estimate of the posterior means of the residuals, respectively. Right column: spatial plots of the posterior means of the residuals. Rows from top to bottom correspond to the Gaussian, Gumbel, and Clayton models, respectively.
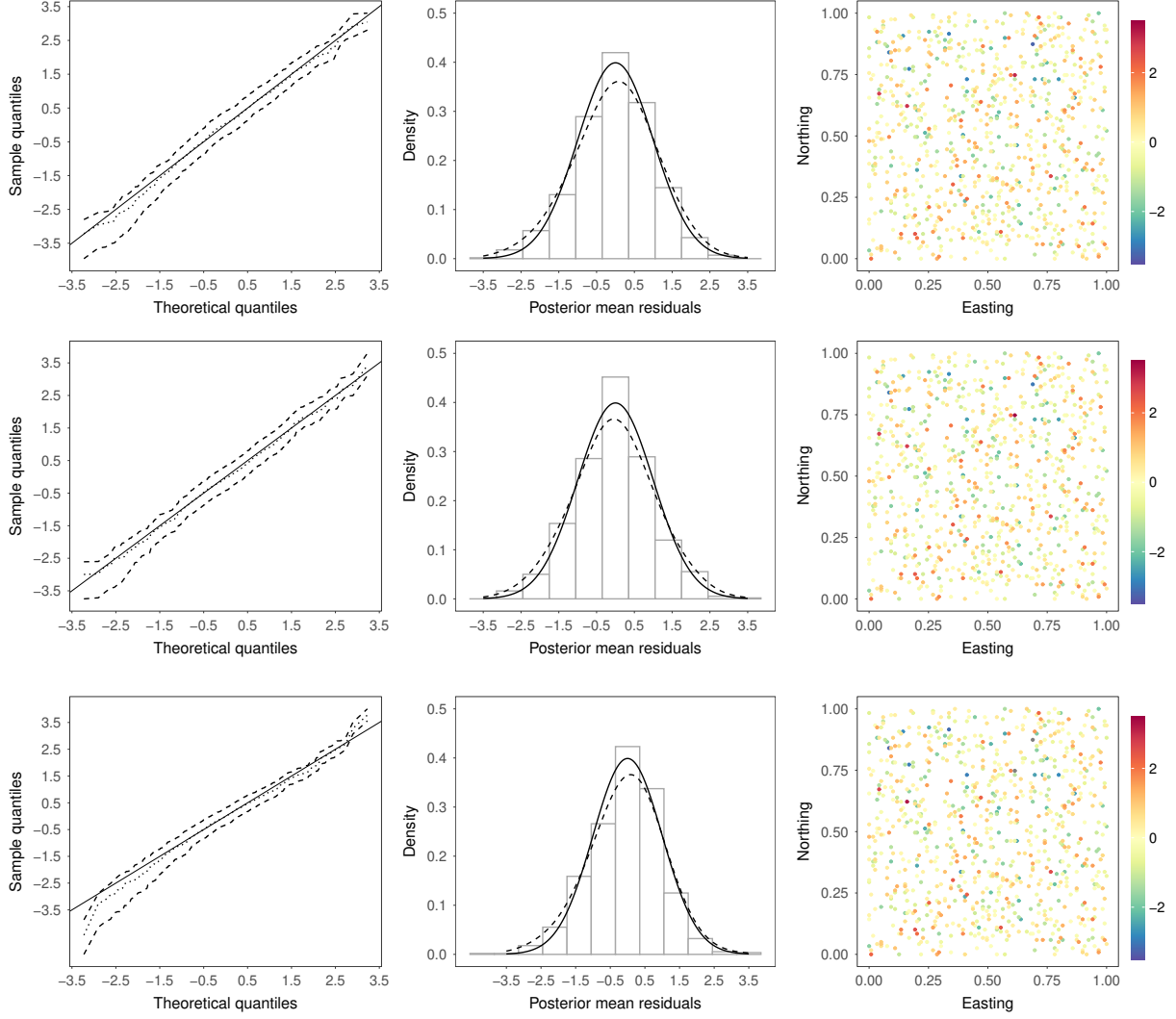
Figure 6: Simulated data example 1 - randomized quantile residual analysis for Scenario 3 ($\sigma_1 = 10$). Left column: Gaussian quantile-quantile plots. Dotted and dashed lines correspond to the posterior mean and 95% interval bands, respectively. Middle column: Histograms of the posterior means of the residuals. Solid and dashed lines are the standard Gaussian density and the kernel density estimate of the posterior means of the residuals, respectively. Right column: spatial plots of the posterior means of the residuals. Rows from top to bottom correspond to the Gaussian, Gumbel, and Clayton models, respectively.
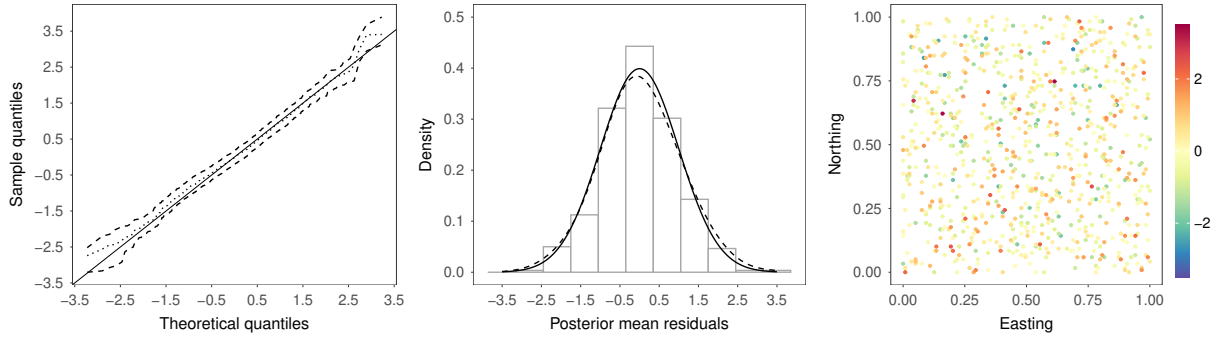
Figure 7: Simulated data example 2 - randomized quantile residual analysis for the NBNNMP model. Left panel: Gaussian quantile-quantile plot. Dotted and dashed lines correspond to the posterior mean and 95% interval bands, respectively. Middle panel: Histogram of the posterior means of the residuals. Solid and dashed lines are the standard Gaussian density and the kernel density estimate of the posterior means of the residuals, respectively. Right panel: spatial plot of the posterior means of the residuals.