Bayesian Analysis (2023) **18**, Number 4, pp. 1191–1222

Nearest-Neighbor Mixture Models for Non-Gaussian Spatial Processes*

Xiaotian Zheng†, Athanasios Kottas‡, and Bruno Sansó‡

Abstract. We develop a class of nearest-neighbor mixture models that provide direct, computationally efficient, probabilistic modeling for non-Gaussian geospatial data. The class is defined over a directed acyclic graph, which implies conditional independence in representing a multivariate distribution through factorization into a product of univariate conditionals, and is extended to a full spatial process. We model each conditional as a mixture of spatially varying transition kernels, with locally adaptive weights, for each one of a given number of nearest neighbors. The modeling framework emphasizes direct spatial modeling of non-Gaussian data, in contrast with approaches that introduce a spatial process for transformed data, or for functionals of the data probability distribution. We study model construction and properties analytically through specification of bivariate distributions that define the local transition kernels. This provides a general strategy for modeling different types of non-Gaussian data based on the bivariate distribution family, and offers avenues to incorporate spatial association via different dependence measures. Regarding computation, direct spatial modeling facilitates efficient, full simulation-based inference; moreover, the framework leverages its mixture model structure to avoid computational issues that arise from large matrix operations, and thus has the potential to achieve scalability. We illustrate the methodology using synthetic data examples and an analysis of Mediterranean Sea surface temperature observations.

Keywords: Bayesian hierarchical models, copulas, Markov chain Monte Carlo, spatial statistics, tail dependence.

1 Introduction

Gaussian processes have been widely used as an underlying structure in model-based analysis of irregularly located spatial data in order to capture short range variability. The fruitfulness of these spatial models owes to the simple characterization of the Gaussian process by a mean and a covariance function, and the optimal prediction it provides that justifies kriging. However, the assumption of Gaussianity is restrictive in many fields where the data exhibits non-Gaussian features, for example, vegetation abundance (Eskelson et al., 2011), precipitation data (Sun et al., 2015), contaminated soil (Paul and Cressie, 2011), temperature data (North et al., 2011), and wind speed data (Bevilacqua et al., 2020). This article aims at developing a flexible class of geostatistical models

*This research was supported in part by the National Science Foundation under awards MMS 2050012, DMS 2153277, and SES 1950902.

†School of Mathematics and Applied Statistics, University of Wollongong, Wollongong, New South Wales, Australia, xzheng@uow.edu.au

‡Department of Statistics, University of California, Santa Cruz, California, USA

 https://doi.org/10.1214/23-BA1405

that is customizable to general non-Gaussian distributions, with particular focus on continuous data.

Several approaches have been developed for non-Gaussian geostatistical modeling. A straightforward approach consists of fitting a Gaussian process after transformation of the original data. Possible transformations include Box-Cox (De Oliveira et al., 1997), power (Allcroft and Glasbey, 2003), square-root (Johns et al., 2003), and Tukey g-and-h (Xu and Genton, 2017) transforms, to name a few. A relevant approach that involves estimating the transformation is to use transport maps with Gaussian process priors (Katzfuss and Schäfer, 2023). Alternative to transformations, a non-Gaussian distribution can be represented as a location-scale mixture of Gaussian distributions. This yields Gaussian process extensions that are able to capture skewness and long tails (Kim and Mallick, 2004; Palacios and Steel, 2006; Zhang and El-Shaarawi, 2010; Mahmoudian, 2017; Morris et al., 2017; Zareifard et al., 2018; Tagle et al., 2020; Bevilacqua et al., 2021). Beyond methods based on continuous mixtures of Gaussian distributions, Bayesian nonparametric methods have been explored for geostatistical data modeling, starting with the approach in Gelfand et al. (2005) which extends the Dirichlet process (Ferguson, 1973) to a prior model for random spatial surfaces. We refer to Müller et al. (2018) for a review. From a different perspective, a class of non-Gaussian Matérn fields is formulated with stochastic partial differential equations driven by non-Gaussian noise (Bolin, 2014; Wallin and Bolin, 2015; Bolin and Wallin, 2020).

An alternative popular approach involves a hierarchical model structure that assumes conditionally independent non-Gaussian marginals, combined with a latent spatial process that is associated with some functional or link function of the first-stage marginals. Hereafter, we refer to these models as hierarchical first-stage non-Gaussian models. If the latent process is linked through a function of some parameter(s) of the first-stage marginal which belongs to the exponential dispersion family, the approach is known as the spatial generalized linear mixed model and its extensions (Diggle et al., 1998; Chan and Dong, 2011). Non-Gaussian spatial models that build from copulas (Joe, 2014) can also be classified into this category. Copula models assume pre-specified families of marginals for observations, with a multivariate distribution underlying the copula for a vector of latent variables that are probability integral transformations of the observations (Danaher and Smith, 2011). Spatial copula models replace the multivariate distribution with one that corresponds to a spatial process, thus introducing spatial dependence (Bárdossy, 2006; Ghosh and Mallick, 2011; Krupskii et al., 2018; Beck et al., 2020).

The non-Gaussian modeling framework proposed in this article is distinctly different from the previously mentioned approaches. Our methodology builds on the class of nearest-neighbor processes obtained by extending a joint density for a reference set of locations to the entire spatial domain. The original joint density is factorized into a product of conditionals with respect to a directed acyclic graph (DAG). Deriving each conditional from a Gaussian process results in the nearest-neighbor Gaussian process (NNGP; Datta et al. 2016a). Models defined over DAGs have received substantial attention; see, e.g., Datta et al. (2016b), Finley et al. (2019), Peruzzi et al. (2020), Peruzzi and Dunson (2022b), and Jin et al. (2023). The class of DAG-based models originates

from the approach in Vecchia (1988) that considers nearest-neighbor approximations. Related work that exploits sparsity to approximate an expensive Gaussian likelihood includes Stein et al. (2004), Gramacy and Apley (2015), Sun and Stein (2016), Stroud et al. (2017), Guinness (2018), and Schäfer et al. (2021). Katzfuss and Guinness (2021) provide a further generalization of the Vecchia approximation framework. Considerably less attention, however, has been devoted to defining models over a DAG with non-Gaussian distributions for the conditionals of the joint density. This is in general a difficult problem, as each conditional involves, say, a p-dimensional conditioning set, which requires a coherent model for a $(p+1)$-dimensional non-Gaussian distribution, with p potentially large. In this article, we take on the challenging task of developing a computationally efficient, interpretable framework that provides generality for modeling different types of non-Gaussian data and flexibility for complex spatial dependence.

We overcome the aforementioned challenge, namely, modeling a non-Gaussian conditional density with a high-dimensional conditioning set, by using a structured mixture model. More specifically, each non-Gaussian conditional over a DAG is modeled as a weighted combination of first-order spatially varying transition kernels, each of which depends on an element of the original p-dimensional conditioning set. This approach produces multivariate non-Gaussian distributions by specification of the bivariate distributions that define the local transition kernels. Thus, it provides generality for modeling different non-Gaussian behaviors, since, relative to the multivariate analogue, constructing bivariate distributions is substantially easier, for instance, using bivariate copulas. Moreover, such a model structure offers the convenience of quantifying multivariate dependence through the collection of bivariate distributions. As an illustration, we study tail dependence properties under appropriate families of bivariate distributions, and provide results that guide modeling choices. The modeling framework achieves flexibility by letting both the weights and transition kernels be spatially dependent, inducing sufficient local dependence to describe a wide range of spatial variability. We refer to the resulting geospatial process as the nearest-neighbor mixture process (NNMP).

An important feature of the model structure is that it facilitates the study of conditions for constructing NNMPs with pre-specified families of marginal distributions. Such conditions are easily implemented without parameter constraints, thus resulting in a general modeling tool to describe spatial data distributions that are skewed, heavy-tailed, positive-valued, or have bounded support, as illustrated through several data examples in Section 5. The NNMP framework emphasizes direct modeling by introducing spatial dependence at the data level. It avoids the use of transformations that may distort the Gaussian process properties (Wallin and Bolin, 2015). It is fundamentally different from the class of hierarchical first-stage non-Gaussian models that introduce spatial dependence through functionals of the data probability distribution, such as the transformed mean. Regarding computation, NNMP models do not require estimation of potentially large vectors of spatially correlated latent variables, something unavoidable with hierarchical first-stage non-Gaussian models. Approaches for such models typically resort to approximate inference (Rue et al., 2009; Zilber and Katzfuss, 2021; Sainsbury-Dale et al., 2021), except for certain classes of models that specify particular families of distributions for the data and latent variables to achieve conjugacy (Bradley et al. 2018; Bradley et al. 2020). See also Peruzzi and Dunson (2022a) for a gradient-based Markov

chain Monte Carlo (MCMC) algorithm for sampling the latent variables. Estimation of NNMPs is analogous to that of a finite mixture model, thus avoiding the need to perform costly matrix operations for large data sets, and allowing for computationally efficient, full simulation-based inference. Overall, the NNMP framework offers a flexible class of models that is able to describe complex spatial dependence, coupled with an efficient computational approach, leveraged from the mixture structure of the model.

The rest of the article is organized as follows. We first formulate the general NNMP modeling framework and investigate stationarity conditions in Section 2. Then, in Section 3, we present approaches to construct different components of an NNMP model, and discuss implied model properties. Specific examples of NNMP models illustrate different components of the methodology. Section 4 develops the general approach to Bayesian estimation and prediction under NNMP models. In Section 5, we demonstrate different NNMP models with synthetic data examples and with the analysis of Mediterranean Sea surface temperature data. Finally, Section 6 concludes with a summary and discussion of future work.

2 Nearest-Neighbor Mixture Processes for Spatial Data

2.1 Modeling Framework

Consider a univariate spatial process $\{Z(\boldsymbol{v}) : \boldsymbol{v} \in \mathcal{D}\}$, where $\mathcal{D} \subset \mathbb{R}^p$. We focus on the case with $p = 2$ although our results are general for $p \geq 1$. In what follows, we take $\boldsymbol{v} = (v_1, v_2)$ for any $\boldsymbol{v} \in \mathcal{D}$, where v_1 and v_2 are the x- and y-coordinate of $\boldsymbol{v}$. The spatial domain $\mathcal{D}$ is partitioned into two sets. In particular, the first set, referred to as the reference set and denoted as $\mathcal{S} = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n\}$, consists of a finite collection of locations in $\mathcal{D}$. We use $\boldsymbol{s}$ and $\boldsymbol{u}$ as generic notation for locations in the partitioning sets, that is, $\boldsymbol{v} \equiv \boldsymbol{s}$ for $\boldsymbol{s} \in \mathcal{S}$ and $\boldsymbol{v} \equiv \boldsymbol{u}$ for $\boldsymbol{u} \in \mathcal{D} \setminus \mathcal{S}$.

If we regard the locations in $\mathcal{S}$ as vertices of a DAG, we can factorize the joint density of $\boldsymbol{Z}_{\mathcal{S}}$ into a product of univariate conditionals as

$$p(\boldsymbol{z}_{\mathcal{S}}) = p(z(\boldsymbol{s}_1)) \prod_{i=2}^{n} p(z(\boldsymbol{s}_i) \mid \boldsymbol{z}_{\mathrm{Ne}(\boldsymbol{s}_i)}), \tag{2.1}$$

where $\boldsymbol{z}_{\mathcal{S}} = (z(\boldsymbol{s}_1), \ldots, z(\boldsymbol{s}_n))^\top$, and the set $\mathrm{Ne}(\boldsymbol{s}_i) \subset \mathcal{S}_i = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_{i-1}\}$ consists of parents of $\boldsymbol{s}_i$. The joint density in (2.1) corresponds to a directed graphical model (Jordan 2004; also known as a Bayesian network), with a DAG driving the conditional independence structure between random variables. In particular, conditional on $\boldsymbol{Z}_{\mathrm{Ne}(\boldsymbol{s}_i)}$, $Z(\boldsymbol{s}_i)$ is independent of $\boldsymbol{Z}_{\mathcal{S}_i \setminus \mathrm{Ne}(\boldsymbol{s}_i)}$. The elements of $\mathrm{Ne}(\boldsymbol{s}_i)$ are selected as the nearest neighbors of $\boldsymbol{s}_i$ within $\mathcal{S}_i$ according to geostatistical distance, and $\mathrm{Ne}(\boldsymbol{s}_i)$ is referred to as the nearest-neighbor set of $\boldsymbol{s}_i$, having at most L elements with $L \ll n$. Placing the elements of $\mathrm{Ne}(\boldsymbol{s}_i)$ in ascending order with respect to distance to $\boldsymbol{s}_i$, we have $\mathrm{Ne}(\boldsymbol{s}_i) = (\boldsymbol{s}_{(i1)}, \ldots, \boldsymbol{s}_{(i,i_L)})$, where $i_L = (i-1) \wedge L$. We note that the development of the modeling framework applies for any choice of the neighbor sets.

Constructing an NNMP model involves specification of $p(z(\boldsymbol{s}_i) \mid \boldsymbol{z}_{\mathrm{Ne}(\boldsymbol{s}_i)})$ in (2.1), and extension to an arbitrary finite set in $\mathcal{D} \setminus \mathcal{S}$. Regarding the first stage of the construction,

we define the joint density over the reference set using a structured mixture model for the conditional densities:

$$p(z(\boldsymbol{s}_i) \,|\, \boldsymbol{z}_{\mathrm{Ne}(\boldsymbol{s}_i)}) = \sum_{l=1}^{i_L} w_l(\boldsymbol{s}_i)\, f_{\boldsymbol{s}_i,l}(z(\boldsymbol{s}_i) \,|\, z(\boldsymbol{s}_{(il)})). \tag{2.2}$$

Here, the lth mixture component, $f_{\boldsymbol{s}_i,l}$, corresponds to the conditional density of $z(\boldsymbol{s}_i)$ given the lth nearest neighbor of $\boldsymbol{s}_i$. For every $\boldsymbol{s}_i \in \mathcal{S}$, the weights satisfy $w_l(\boldsymbol{s}_i) \geq 0$, for all l, and $\sum_{l=1}^{i_L} w_l(\boldsymbol{s}_i) = 1$. In a DAG, nearest neighbors in set $\mathrm{Ne}(\boldsymbol{s}_i)$ are vertices that have directed edges pointing to $\boldsymbol{s}_i$. Thus, it is appealing to consider a high-order Markov model in which temporal lags have a similar notion of direction. Our approach to formulating (2.2) is motivated by a class of mixture transition distribution models (Le et al., 1996), which consists of a mixture of first-order transition densities with a vector of common weights. A key feature of the formulation in (2.2) is the decomposition of a non-Gaussian conditional density, with a potentially large conditioning set, into a weighted sum of local conditional densities. This provides flexible, parsimonious modeling of $p(z(\boldsymbol{s}_i) \,|\, \boldsymbol{z}_{\mathrm{Ne}(\boldsymbol{s}_i)})$ through specifying bivariate distributions that define the local conditionals $f_{\boldsymbol{s}_i,l}(z(\boldsymbol{s}_i) \,|\, z(\boldsymbol{s}_{(il)}))$. We provide further discussion on this feature for model construction and relevant properties in the following sections.

Spatial dependence characterized by (2.2) is twofold. First, the weights $w_l(\boldsymbol{s}_i)$ are spatially varying, reflecting the variability in contributions from each neighbor and changes in neighborhood structures for different locations. Secondly, each component $f_{\boldsymbol{s}_i,l}$ is associated with spatially varying parameters indexed at $\boldsymbol{s}_i \in \mathcal{S}$, defined by a probability model or a link function. In general, probability models or link functions for the spatially varying parameters should be considered case by case, given different specifications on the components $f_{\boldsymbol{s}_i,l}$. Details of the construction for the weights and component densities are deferred to Sections 3.1 and 3.2.

We obtain the NNMP, a legitimate spatial process, by extending (2.2) to an arbitrary set of non-reference locations $\mathcal{U} = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r\}$ where $\mathcal{U} \subset \mathcal{D} \setminus \mathcal{S}$. In particular, we define the conditional density of $\boldsymbol{z}_\mathcal{U}$ given $\boldsymbol{z}_\mathcal{S}$ as

$$p(\boldsymbol{z}_\mathcal{U} \,|\, \boldsymbol{z}_\mathcal{S}) = \prod_{i=1}^{r} p(z(\boldsymbol{u}_i) \,|\, \boldsymbol{z}_{\mathrm{Ne}(\boldsymbol{u}_i)}) = \prod_{i=1}^{r} \sum_{l=1}^{L} w_l(\boldsymbol{u}_i)\, f_{\boldsymbol{u}_i,l}(z(\boldsymbol{u}_i) \mid z(\boldsymbol{u}_{(il)})), \tag{2.3}$$

where the specification on $w_l(\boldsymbol{u}_i)$ and $f_{\boldsymbol{u}_i,l}$, for all i and all l, is analogous to that for (2.2), except that $\mathrm{Ne}(\boldsymbol{u}_i) = \{\boldsymbol{u}_{(i1)}, \ldots, \boldsymbol{u}_{(iL)}\}$ are found in $\mathcal{S}$, where $\boldsymbol{u}_{(il)}$ denotes the lth nearest-neighbor of $\boldsymbol{u}_i$ in terms of geostatistical distance. Building the construction of the neighbor sets $\mathrm{Ne}(\boldsymbol{u}_i)$ on the reference set ensures that $p(\boldsymbol{z}_\mathcal{U} \,|\, \boldsymbol{z}_\mathcal{S})$ is a proper density.

Given (2.2) and (2.3), we can obtain the joint density $p(\boldsymbol{z}_\mathcal{V})$ of a realization $\boldsymbol{z}_\mathcal{V}$ over any finite set of locations $\mathcal{V} \subset \mathcal{D}$. When $\mathcal{V} \subset \mathcal{S}$, the joint density $p(\boldsymbol{z}_\mathcal{V})$ is directly available as the appropriate marginal of $p(\boldsymbol{z}_\mathcal{S})$. Otherwise, we have that $p(\boldsymbol{z}_\mathcal{V}) = \int p(\boldsymbol{z}_\mathcal{U} \,|\, \boldsymbol{z}_\mathcal{S}) p(\boldsymbol{z}_\mathcal{S}) \prod_{\{\boldsymbol{s}_i \in \mathcal{S} \setminus \mathcal{V}\}} dz(\boldsymbol{s}_i)$, where $\mathcal{U} = \mathcal{V} \setminus \mathcal{S}$. If $\mathcal{S} \setminus \mathcal{V}$ is empty, $p(\boldsymbol{z}_\mathcal{V})$ is simply $p(\boldsymbol{z}_\mathcal{U} \,|\, \boldsymbol{z}_\mathcal{S}) p(\boldsymbol{z}_\mathcal{S})$. In general, the joint density $p(\boldsymbol{z}_\mathcal{V})$ of an NNMP is intractable. However,

since both $p(\boldsymbol{z}_{\mathcal{U}} \mid \boldsymbol{z}_{\mathcal{S}})$ and $p(\boldsymbol{z}_{\mathcal{S}})$ are products of mixtures, we can recognize that $p(\boldsymbol{z}_{\mathcal{V}})$ is a finite mixture, which suggests flexibility of the model to capture complex non-Gaussian dependence over the domain $\mathcal{D}$. Moreover, we show in Section 3.2 that for some NNMPs, the joint density $p(\boldsymbol{z}_{\mathcal{V}})$ has a closed-form expression. In the subsequent development of model properties, we will use the conditional density

$$p(z(\boldsymbol{v}) \mid \boldsymbol{z}_{\mathrm{Ne}(\boldsymbol{v})}) = \sum_{l=1}^{L} w_l(\boldsymbol{v})\, f_{\boldsymbol{v},l}(z(\boldsymbol{v}) \mid z(\boldsymbol{v}_{(l)})), \quad \boldsymbol{v} \in \mathcal{D}, \tag{2.4}$$

to characterize an NNMP. If $\boldsymbol{v} \in \mathcal{S}$, (2.4) is equivalent to (2.2). Otherwise, (2.4) corresponds to the conditional in (2.3).

Note that spatial locations are not naturally ordered, and thus a different ordering on the reference set locations results in different neighbor sets. Literature that considers nearest-neighbor models for Gaussian data generally orders locations based on sorting coordinates. Our default choice is a random ordering, which has been shown to improve model performance relative to coordinate-based orderings (Guinness, 2018). For the NNMP models illustrated in this article, we found through simulation experiments that there were no discernible differences between the inferences based on $p(\boldsymbol{z}_{\mathcal{S}})$ given different random orderings. A further remark is that the ordering of the reference set $\mathcal{S}$ is typically reserved for observed data. Thus, the ordering effect lies in the model estimation based on (2.2) with realization $\boldsymbol{z}_{\mathcal{S}}$. Spatial prediction typically rests on locations outside $\mathcal{S}$ using (2.3), where the ordering effect disappears.

Before closing this section, we remark on the connection and a key conceptual difference between the NNMP modeling framework and the nearest-neighbor spatial models discussed in the Introduction. The above derivation of the NNMP follows the general steps to obtain a nearest-neighbor process (Datta et al., 2016a), i.e., specifying a joint distribution on $\mathcal{S}$ according to a DAG, and then extending the joint distribution to $\mathcal{D} \setminus \mathcal{S}$. Thus, the NNMP satisfies the Kolmogorov consistency conditions. On the other hand, in contrast to NNGP models, we do not posit a parent process. For our modeling approach, the right-hand-side of (2.1) does not serve the role of an approximation to the density of a Gaussian process realization. Our modeling perspective is distinctly different in that we directly model the joint density $p(\boldsymbol{z}_{\mathcal{S}})$, utilizing the nearest-neighbor DAG representation with the structured mixture in (2.2), which is key to constructing general non-Gaussian spatial processes.

2.2 NNMPs with Stationary Marginal Distributions

We develop a sufficient condition to construct NNMPs with general stationary marginal distributions. The key feature of this result is that the condition relies on the bivariate distributions that define the first-order transition kernels in (2.4) without the need to impose restrictions on the parameter space. The Supplementary Material includes the proof of Proposition 1, as well as of Propositions 2, 3 and 4, and of Corollary 1, formulated later in Sections 3.2 and 3.3.

Proposition 1. *Consider an NNMP for which the component density $f_{\boldsymbol{v},l}$ in (2.4) is specified by the conditional density of $U_{\boldsymbol{v},l}$ given $V_{\boldsymbol{v},l}$, where the random vector $(U_{\boldsymbol{v},l}, V_{\boldsymbol{v},l})$*

follows a bivariate distribution with marginal densities $f_{U_{\boldsymbol{v},l}}$ and $f_{V_{\boldsymbol{v},l}}$, for $l = 1, \ldots, L$. Consider a pre-specified density f_Z, and assume that: $Z(\boldsymbol{s}_1) \sim f_Z$, $\boldsymbol{s}_1 \in \mathcal{S}$; and, for every $\boldsymbol{v} \in \mathcal{D}$, $f_{U_{\boldsymbol{v},l}}(z) = f_{V_{\boldsymbol{v},l}}(z) = f_Z(z)$, for all z and for all l. Then, the NNMP has stationary marginal density f_Z.

This result builds from the one in Zheng et al. (2022) where mixture transition distribution models with stationary marginal distributions were constructed. It applies regardless of $Z(\boldsymbol{v})$ being a continuous, discrete or mixed random variable, thus allowing for a wide range of non-Gaussian marginal distributions and a general (linear or non-linear) form for the expectation with respect to the conditional density p in (2.4).

As previously discussed, the mixture model formulation for the conditional density in (2.4) induces a finite mixture for the NNMP finite-dimensional distributions. On the other hand, due to the mixture form, an explicit expression for the covariance function is difficult to derive. Consider NNMPs that satisfy the conditions in Proposition 1 with a stationary marginal density f_Z that has finite first and second moment. A recursive equation can be obtained when the conditional expectation with respect to $(U_{\boldsymbol{v},l}, V_{\boldsymbol{v},l})$ is linear, that is, $E(U_{\boldsymbol{v},l} \,|\, V_{\boldsymbol{v},l} = z) = a_l(\boldsymbol{v}) + b_l(\boldsymbol{v})\, z$ for some $a_l(\boldsymbol{v}), b_l(\boldsymbol{v}) \in \mathbb{R}$, $l = 1, \ldots, L$, and for all $\boldsymbol{v} \in \mathcal{D}$. Without loss of generality, we assume the first moment of $Z(\boldsymbol{v})$ with respect to f_Z is zero. Then the covariance over any two locations $\boldsymbol{v}_1, \boldsymbol{v}_2 \in \mathcal{D}$ is

$$\begin{aligned}&\mathrm{Cov}(Z(\boldsymbol{v}_1), Z(\boldsymbol{v}_2)) \\ &\quad = \begin{cases} \sum_{l=1}^{L} w_l(\boldsymbol{s}_i)\, b_l(\boldsymbol{s}_i)\, E(Z(\boldsymbol{s}_j) Z(\boldsymbol{s}_{(il)})), & \boldsymbol{v}_1 \equiv \boldsymbol{s}_i \in \mathcal{S}, \boldsymbol{v}_2 \equiv \boldsymbol{s}_j \in \mathcal{S}, \\ \sum_{l=1}^{L} w_l(\boldsymbol{v}_1)\, b_l(\boldsymbol{v}_1)\, E(Z(\boldsymbol{s}_j) Z(\boldsymbol{v}_{(1l)})), & \boldsymbol{v}_1 \notin \mathcal{S}, \boldsymbol{v}_2 \equiv \boldsymbol{s}_j \in \mathcal{S}, \\ \sum_{l=1}^{L} \sum_{l'=1}^{L} w_{ll'} \{a_{ll'} + b_{ll'} E(Z(\boldsymbol{v}_{(1l)}) Z(\boldsymbol{v}_{(2l')}))\}, & \boldsymbol{v}_1, \boldsymbol{v}_2 \notin \mathcal{S}, \end{cases}\end{aligned} \tag{2.5}$$

where $w_{ll'} \equiv w_l(\boldsymbol{v}_1) w_{l'}(\boldsymbol{v}_2)$, $a_{ll'} \equiv a_l(\boldsymbol{v}_1) a_{l'}(\boldsymbol{v}_2)$, $b_{ll'} \equiv b_l(\boldsymbol{v}_1) b_{l'}(\boldsymbol{v}_2)$, and without loss of generality, we assume $i > j$. The covariance in (2.5) implies that, even though the NNMP has a stationary marginal distribution, it is second-order non-stationary.

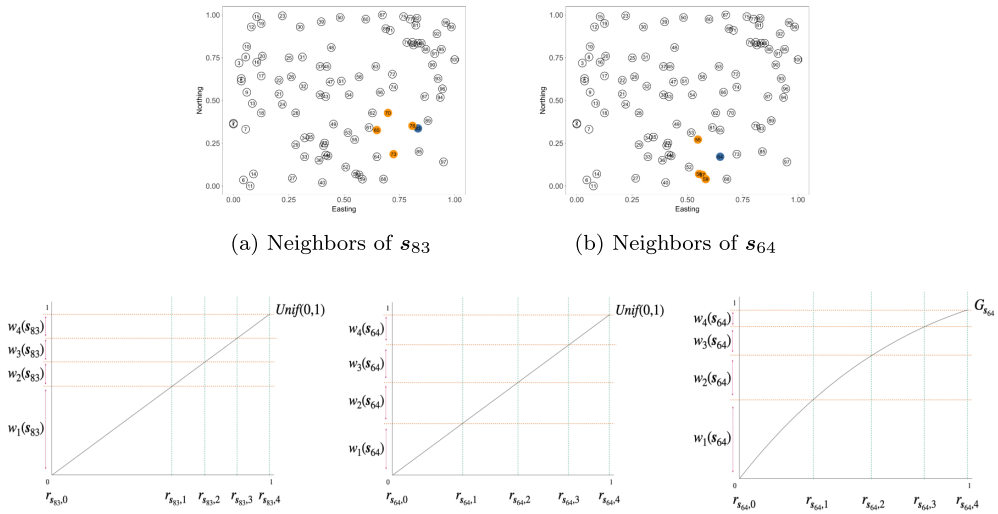
3 Construction of NNMP Model Components

Here, we present our methodology to construct the different components of the general NNMP framework, in particular, the weights (Section 3.1) and the mixture component densities (Section 3.2). In Section 3.3, we show that the NNMP construction facilitates the study of multivariate dependence.

3.1 Spatially Varying Weights

As discussed in Section 2.1, the weights should vary in space, and be able to adjust to different neighborhood structures. Let $\boldsymbol{w}(\boldsymbol{v}) = (w_1(\boldsymbol{v}), \ldots, w_L(\boldsymbol{v}))^\top$ be the vector of weights for $\boldsymbol{v} \in \mathcal{D}$. We seek a model that can accommodate the following scenarios:

(i) For any $\boldsymbol{v} \in \mathcal{D}$ with a set of neighbors, the weights that indicate the contribution of each neighbor can vary as $\boldsymbol{v}$ moves in $\mathcal{D}$;



(a) Neighbors of $\boldsymbol{s}_{83}$ (b) Neighbors of $\boldsymbol{s}_{64}$

(c) Weights $w_l(\boldsymbol{s}_{83})$ (d) Weights $w_l(\boldsymbol{s}_{64})$ (e) Weights $w_l(\boldsymbol{s}_{64})$

Figure 1: Illustration of the NNMP weights. Panels (a) and (b) highlight two locations $\boldsymbol{s}_{83}$ and $\boldsymbol{s}_{64}$ (blue) and their neighbors (orange), among reference locations $\{\boldsymbol{s}_i\}_{i=1}^{100} \subset [0,1] \times [0,1]$. Panels (c) and (d) show weights $\{w_l(\boldsymbol{s}_{83})\}_{l=1}^{4}$ and $\{w_l(\boldsymbol{s}_{64}\}_{l=1}^{4}$ that correspond to two different neighborhood structures, respectively, under a uniform distribution Unif$(0,1)$ with cutoff points $\{r_{\boldsymbol{s}_{83},l}\}_{l=1}^{4}$ and $\{r_{\boldsymbol{s}_{64},l}\}_{l=1}^{4}$ located differently. Panel (e) shows the weights $\{w_l(\boldsymbol{s}_{64})\}_{l=1}^{4}$ for scenario (b) after adjustment by a spatially dependent distribution $G_{\boldsymbol{s}_{64}}$, under the same cutoff points $\{r_{\boldsymbol{s}_{64},l}\}_{l=1}^{4}$.

(ii) For any $\boldsymbol{v}, \boldsymbol{v}'$ in $\mathcal{D}$, the vectors $\boldsymbol{w}(\boldsymbol{v})$ and $\boldsymbol{w}(\boldsymbol{v}')$ can be different if their neighborhood structures are different.

An example of (ii) is illustrated in Figure 1(a)-(b). In the set of reference locations $\mathcal{S} = \{\boldsymbol{s}_i\}_{i=1}^{100} \subset [0,1] \times [0,1]$, location $\boldsymbol{s}_{83}$ has one close and three distant neighbors, while all neighbors of location $\boldsymbol{s}_{64}$ have similar distances to $\boldsymbol{s}_{64}$.

We achieve the aforementioned goals with the following model for the weights. Consider a collection of spatially dependent distribution functions $\{G_{\boldsymbol{v}} : \boldsymbol{v} \in \mathcal{D}\}$ supported on $(0,1)$. For each $\boldsymbol{v}$, the weights are defined as the increments of $G_{\boldsymbol{v}}$ with cutoff points $r_{\boldsymbol{v},0}, \ldots, r_{\boldsymbol{v},L}$. More specifically,

$$w_l(\boldsymbol{v}) = \int \mathbb{1}_{(r_{\boldsymbol{v},l-1},\, r_{\boldsymbol{v},l})}(x)\, dG_{\boldsymbol{v}}(x), \quad l = 1, \ldots, L, \tag{3.1}$$

where $\mathbb{1}_A$ denotes the indicator function for set A. The cutoff points $0 = r_{\boldsymbol{v},0} < r_{\boldsymbol{v},1} <$

$\cdots < r_{\boldsymbol{v},L} = 1$ are such that,

$$r_{\boldsymbol{v},l} - r_{\boldsymbol{v},l-1} = \frac{k'(\boldsymbol{v}, \boldsymbol{v}_{(l)} \,|\, \boldsymbol{\zeta})}{\sum_{l=1}^{L} k'(\boldsymbol{v}, \boldsymbol{v}_{(l)} \,|\, \boldsymbol{\zeta})}, \quad l = 1, \ldots, L, \tag{3.2}$$

where $k' : \mathcal{D} \times \mathcal{D} \to [0, 1]$ is a bounded kernel function with parameters $\boldsymbol{\zeta}$.

As an example, consider a simpler version of (3.1) where $G_{\boldsymbol{v}}$ is the uniform distribution, $G_{\boldsymbol{v}} \equiv G = \text{Unif}(0, 1)$ for all $\boldsymbol{v}$, and, in (3.2), $k'(\boldsymbol{v}, \boldsymbol{v}_{(l)} \,|\, \xi) = \exp(-\|\boldsymbol{v} - \boldsymbol{v}_{(l)}\|/\xi)$. Then, the weights become $w_l(\boldsymbol{v}) = \exp(-\|\boldsymbol{v} - \boldsymbol{v}_{(l)}\|/\xi) / \sum_{l=1}^{L} \exp(-\|\boldsymbol{v} - \boldsymbol{v}_{(l)}\|/\xi)$. When ξ is small, the weights will be localized; see, e.g., Figure 1 (c)-(d). When ξ is large, the model will produce similar weights for different locations regardless of their neighborhood structures. To add flexibility we can take G to be spatially dependent, providing a local adjustment to the weights. For instance, the distribution $G_{\boldsymbol{v}}$ in Figure 1(e) allocates more weight to nearer neighbors than the Unif$(0, 1)$ in Figure 1(d).

A convenient choice for $G_{\boldsymbol{v}}$ is the logit Gaussian distribution, denoted by $G_{\boldsymbol{v}}(\cdot \,|\, \mu(\boldsymbol{v}), \kappa^2)$, such that the corresponding Gaussian distribution has mean $\mu(\boldsymbol{v})$ and variance κ^2. The spatial dependence across the weights is introduced through the mean $\mu(\boldsymbol{v}) = \gamma_0 + \gamma_1 v_1 + \gamma_2 v_2$, in order to capture large-scale spatial variability in $\mathcal{D}$. Given the cutoff points and κ^2, a small value of $\mu(\boldsymbol{v})$ favors large weights for the near neighbors. We notice that Cadonna et al. (2019) use a set of fixed, uniform cutoff points on $[0, 1]$, i.e., $r_{\boldsymbol{v},l} - r_{\boldsymbol{v},l-1} = 1/L$, for spectral density estimation, with a collection of logit Gaussian distributions indexed by frequency.

We remark that the modeling approach for the weights seeks to balance flexibility and efficient inference. An alternative is to place a spatial process prior that potentially provides full support for the distributions $G_{\boldsymbol{v}}$, with fixed cutoff points. However, practical specification of such a prior poses the challenge of performing inference given a single realization of $\{Z(\boldsymbol{v})\}$. Thus, we consider instead a parametric family for $G_{\boldsymbol{v}}$, with the combination of (3.1) and (3.2) allowing for a flexible and interpretable model. Moreover, the use of a logit Gaussian distribution facilitates implementation. As detailed in Section 4.2, after conditioning on a vector of auxiliary variables, the full conditional distributions of parameters $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \gamma_2)^\top$ and κ^2 are available in closed-form, facilitating the sampling in an MCMC algorithm. Finally, we point out that our model for $\mu(\boldsymbol{v})$ is primarily illustrative for the NNMP framework. For specific data applications, users can modify it to include spatial variability in different scales, such as using basis function models (Cressie et al., 2022) to construct a random field for $\mu(\boldsymbol{v})$.

An NNMP model requires selection of the neighborhood size L. This can be done using standard model comparison metrics, scoring rules, or information criteria. In general, a larger L increases computational cost. Datta et al. (2016a) conclude that a moderate value ($L \leq 20$) typically suffices for NNGPs. On the other hand, it is possible that information from distant neighbors is also important (Stein et al., 2004). Therefore, one may seek a larger L to include extra neighbor information for large non-Gaussian data sets with complex dependence. Our model for the weights allows taking a relatively large neighbor set without heavily impacting computational demand. In particular, we can

assign small probabilities a priori to distant neighbors. Posterior inference will determine the contribution of each neighbor, with important neighbors being assigned large weights a posteriori.

3.2 Mixture Component Densities

Proposition 1 suggests a general strategy to construct the spatially varying conditional densities $f_{\boldsymbol{v},l}$ in (2.4). That is, the conditional densities $f_{\boldsymbol{v},l} \equiv f_{U_{\boldsymbol{v},l}|V_{\boldsymbol{v},l}}$ correspond to bivariate distributions indexed at $\boldsymbol{v}$, namely, the distributions of $(U_{\boldsymbol{v},l}, V_{\boldsymbol{v},l})$, where $U_{\boldsymbol{v},l}$ and $V_{\boldsymbol{v},l}$ associate with site $\boldsymbol{v}$ and the lth nearest neighbor $\boldsymbol{v}_{(l)}$, respectively.

To balance model flexibility and scalability, we build spatially varying distributions by considering the distribution of random vector (U_l, V_l), for $l = 1, \ldots, L$, and extending some of its parameters to be spatially varying, that is, indexed in $\boldsymbol{v}$. To this end, we use a probability model or a link function. We refer to the random vectors (U_l, V_l) as the set of base random vectors. With a careful choice of the model/function for the spatially varying parameter(s), this construction method reduces significantly the dimension of the parameter space, while preserving the capability of the NNMP model structure to capture spatial dependence.

As an example, for a bivariate Gaussian distribution, take η_l to denote the parameter for the correlation between U_l and V_l. Spatial dependence is then introduced by extending η_l to $\eta_l(\boldsymbol{v})$, that is, a spatial correlation function that computes the correlation between $Z(\boldsymbol{v})$ and $Z(\boldsymbol{v}_{(l)})$. However, the modeling approach is not limited to correlation coefficients. Any measures that quantify the degree of dependence between U_l and V_l can be considered; indeed, Example 3 in this section considers Kendall's τ coefficient. The equation below shows the workflow to introduce spatial dependence:

$$(U_l, V_l \,|\, \eta_l) \xrightarrow{\eta_l(\boldsymbol{v}) \equiv h_l(\boldsymbol{v}, \boldsymbol{v}_{(l)})} (U_{\boldsymbol{v},l}, V_{\boldsymbol{v},l} \,|\, \eta_l(\boldsymbol{v})) \longrightarrow f_{\boldsymbol{v},l} \equiv f_{U_{\boldsymbol{v},l}|V_{\boldsymbol{v},l}}$$

with model/function $h_l : \mathcal{D} \times \mathcal{D} \to H_l$, where H_l is the parameter space of η_l.

We illustrate the method starting with a bivariate Gaussian distribution and its continuous mixtures for real-valued data, followed by a general strategy that uses bivariate copulas for data with general support. In these examples, the bivariate distribution of (U_l, V_l) is specified focusing on the associated marginals. The bivariate distribution specification can also build from compatible conditionals as illustrated in Section 3.3.

Gaussian and Continuous Mixture of Gaussian NNMP Models

Example 1 (Gaussian NNMP). For $l = 1, \ldots, L$, take (U_l, V_l) to be a bivariate Gaussian random vector with mean $\mu_l \mathbf{1}_2$ and covariance matrix $\Sigma_l = \sigma_l^2 \left(\begin{smallmatrix} 1 & \rho_l \\ \rho_l & 1 \end{smallmatrix}\right)$, where $\mathbf{1}_2$ is the two-dimensional column vector of ones, resulting in a Gaussian conditional density $f_{U_l|V_l}(u_l \,|\, v_l) = N(u_l \,|\, (1-\rho_l)\mu_l + \rho_l v_l, \sigma_l^2(1-\rho_l^2))$. If we extend the correlation parameter to be spatially varying, $\rho_l(\boldsymbol{v}) = k_l(\boldsymbol{v}, \boldsymbol{v}_{(l)})$, for a correlation function k_l, we obtain the

spatially varying conditional density,

$$p(z(\boldsymbol{v}) \mid \boldsymbol{z}_{\mathrm{Ne}(\boldsymbol{v})}) = \sum_{l=1}^{L} w_l(\boldsymbol{v})\, N(z(\boldsymbol{v}) \mid (1-\rho_l(\boldsymbol{v}))\mu_l + \rho_l(\boldsymbol{v}) z(\boldsymbol{v}_{(l)}), \sigma_l^2(1-(\rho_l(\boldsymbol{v}))^2)). \quad (3.3)$$

This NNMP is referred to as the Gaussian NNMP (GNNMP). Using Proposition 1, if we take $Z(\boldsymbol{s}_1) \sim N(z \,|\, \mu, \sigma^2)$, and set $\mu_l = \mu$ and $\sigma_l^2 = \sigma^2$, for all l, the model has stationary marginal given by the $N(\mu, \sigma^2)$ distribution. We refer to this model as the stationary GNNMP. The finite-dimensional distribution of the stationary GNNMP is characterized by the following proposition.

Proposition 2. *The finite-dimensional distributions of the stationary GNNMP model are mixtures of multivariate Gaussian distributions.*

Based on the GNNMP, various NNMP models with different families for (U_l, V_l) can be constructed by exploiting location-scale mixtures of Gaussian distributions. We illustrate the approach with the skew-GNNMP model.

Example 2 (Skew-GNNMP). Denote by $\mathrm{TN}(\mu, \sigma^2; a, b)$ the Gaussian distribution with mean μ and variance σ^2, truncated at the interval (a, b). Building from the GNNMP, we start with a conditional bivariate Gaussian distribution for (U_l, V_l), given $z_0 \sim \mathrm{TN}(0, 1; 0, \infty)$, where μ_l is replaced with $\mu_l + \lambda_l z_0$. Marginalizing out z_0 yields the bivariate skew-Gaussian distribution (Azzalini, 2013) for (U_l, V_l). Extending again ρ_l to $\rho_l(\boldsymbol{v})$, for all l, we can express the conditional density $p(z(\boldsymbol{v}) \,|\, \boldsymbol{z}_{\mathrm{Ne}(\boldsymbol{v})})$ for the skew-GNNMP model as

$$\sum_{l=1}^{L} w_l(\boldsymbol{v}) \int_0^{\infty} N(z(\boldsymbol{v}) \,|\, \mu_l(\boldsymbol{v}), \sigma_l^2(\boldsymbol{v}))\, \mathrm{TN}(z_0(\boldsymbol{v}) \,|\, \mu_{0l}(\boldsymbol{v}_{(l)}), \sigma_{0l}^2; 0, \infty)\, dz_0(\boldsymbol{v}), \quad (3.4)$$

where we have: $\mu_l(\boldsymbol{v}) = \{1-\rho_l(\boldsymbol{v})\}\{\mu_l + \lambda_l z_0(\boldsymbol{v})\} + \rho_l(\boldsymbol{v}) z(\boldsymbol{v}_{(l)})$; $\sigma_l^2(\boldsymbol{v}) = \sigma_l^2\{1-(\rho_l(\boldsymbol{v}))^2\}$; $\mu_{0l}(\boldsymbol{v}_{(l)}) = \{z(\boldsymbol{v}_{(l)}) - \mu_l\}\lambda_l/(\sigma_l^2 + \lambda_l^2)$; and $\sigma_{0l}^2 = \sigma_l^2/(\sigma_l^2 + \lambda_l^2)$. Setting $\lambda_l = \lambda$, $\mu_l = \mu$, and $\sigma_l^2 = \sigma^2$, for all l, we obtain the stationary skew-GNNMP model, with skew-Gaussian marginal $f_Z(z) = 2\, N(z \,|\, \mu, \lambda^2 + \sigma^2)\, \Phi((z-\mu)\lambda/(\sigma\sqrt{\lambda^2 + \sigma^2}))$, denoted as $\mathrm{SN}(\mu, \lambda^2 + \sigma^2, \lambda/\sigma)$, where $\lambda \in \mathbb{R}$ controls the skewness, and $\sigma^2 > 0$ is a scale parameter. An extension of the skew-GNNMP that allows λ to vary in space is explored in Section 5.2 for the Mediterranean Sea surface temperature data analysis.

The skew-GNNMP model is an example of a location mixture of Gaussian distributions. Scale mixtures can also be considered to obtain for instance the Student-t model. In that case, we replace the covariance matrix Σ_l with $c\Sigma_l$, taking c as a random variable with an appropriate inverse-gamma distribution. Important families that admit a location and/or scale mixture of Gaussians representation include the skew-t, Laplace, and asymmetric Laplace distributions. Using a similar approach to the one for the skew-GNNMP example, we can construct the corresponding NNMP models.

Copula NNMP Models and Kendall's τ Coefficient

A copula function $C : [0,1]^p \to [0,1]$ is a function such that, for any multivariate distribution function $F(z_1, \ldots, z_p)$, there exists a copula C for which $F(z_1, \ldots, z_p) =$

$C(F_1(z_1), \ldots, F_p(z_p))$, where F_j is the marginal distribution function of Z_j, $j = 1, \ldots, p$ (Sklar, 1959). If Z_j is continuous for all j, C is unique, and the joint density is given by $f(z_1, \ldots, z_p) = c(z_1, \ldots, z_p) \prod_{j=1}^{p} f_j(z_j)$, where $c = \partial^p C / (\partial F_1 \ldots \partial F_p)$ is the copula density, and f_j is the density of Z_j. A copula enables us to separate the modeling of the marginal distributions from the dependence. Thus, the invariant condition in Proposition 1 can be attained by specifying the stationary distribution F_Z as the marginal distribution of (U_l, V_l) for all l. The copula parameter that determines the dependence of (U_l, V_l) can be modeled as spatially varying to create a sequence of spatially dependent bivariate vectors $(U_{\boldsymbol{v},l}, V_{\boldsymbol{v},l})$. Here, we focus on continuous distributions, although this strategy can be applied for any family of distributions for F_Z. We consider bivariate copulas with a single copula parameter, and illustrate next the construction of a copula NNMP given a stationary marginal density f_Z.

Example 3 (Copula NNMP). For the bivariate distribution of each (U_l, V_l) with marginals f_{U_l} and f_{V_l}, we consider a copula C_l with parameter η_l, for $l = 1, \ldots, L$. We obtain a spatially varying copula $C_{\boldsymbol{v},l}$ for $(U_{\boldsymbol{v},l}, V_{\boldsymbol{v},l})$ by extending η_l to $\eta_l(\boldsymbol{v})$. The joint density of $(U_{\boldsymbol{v},l}, V_{\boldsymbol{v},l})$ is given by $c_{\boldsymbol{v},l}(z(\boldsymbol{v}), z(\boldsymbol{v}_{(l)})) f_{U_{\boldsymbol{v},l}}(z(\boldsymbol{v})) f_{V_{\boldsymbol{v},l}}(z(\boldsymbol{v}_{(l)}))$, where $c_{\boldsymbol{v},l}$ is the copula density of $C_{\boldsymbol{v},l}$, and $f_{U_{\boldsymbol{v},l}} = f_{U_l}$ and $f_{V_{\boldsymbol{v},l}} = f_{V_l}$ are the marginal densities of $U_{\boldsymbol{v},l}$ and $V_{\boldsymbol{v},l}$, respectively. Given a pre-specified stationary marginal f_Z, we replace both $f_{U_{\boldsymbol{v},l}}$ and $f_{V_{\boldsymbol{v},l}}$ with f_Z, for every $\boldsymbol{v}$ and for all l. We then obtain the conditional density

$$p(z(\boldsymbol{v}) \mid \boldsymbol{z}_{\mathrm{Ne}(\boldsymbol{v})}) = \sum_{l=1}^{L} w_l(\boldsymbol{v})\, c_{\boldsymbol{v},l}(z(\boldsymbol{v}), z(\boldsymbol{v}_{(l)}))\, f_Z(z(\boldsymbol{v})) \tag{3.5}$$

that characterizes the stationary copula NNMP.

Under the copula framework, one strategy to specify the spatially varying parameter is through the Kendall's τ coefficient. The Kendall's τ, taking values in $[-1, 1]$, is a bivariate concordance measure with properties useful for non-Gaussian modeling. In particular, its existence does not require finite second moment and it is invariant under strictly increasing transformations. If (U_l, V_l) is continuous with a copula C_l, its Kendall's τ is $\rho_{\tau,l} = 4 \int_{[0,1]^2} C_l dC_l - 1$. Without loss of generality, we assume C_l carries a single parameter. Taking $A_l \subset [-1, 1]$ as the range of $\rho_{\tau,l}$, we can construct a composition function $h_l := g_l \circ k_l$ for some link function $g_l : A_l \to H_l$ and kernel function $k_l : \mathcal{D} \times \mathcal{D} \to A_l$, where H_l is the parameter space associated with C_l. The kernel k_l should be specified with caution; k_l must satisfy axioms in the definition of a bivariate concordance measure (Joe 2014, Section 2.12). We illustrate the strategy with the following example.

Spatial Gumbel Copula The bivariate Gumbel copula is an asymmetric copula useful for modeling dependence when the marginals are positive and heavy-tailed. The spatial Gumbel copula is defined as

$$C_{\boldsymbol{v},l} = \exp\left(-\left[\left\{-\log F_{U_{\boldsymbol{v},l}}(z(\boldsymbol{v}))\right\}^{\eta_l(\boldsymbol{v})} + \left\{-\log F_{V_{\boldsymbol{v},l}}(z(\boldsymbol{v}_{(l)}))\right\}^{\eta_l(\boldsymbol{v})}\right]^{1/\eta_l(\boldsymbol{v})}\right),$$

where $\eta_l(\boldsymbol{v}) \in [1, \infty)$ and perfect dependence is obtained if $\eta_l(\boldsymbol{v}) \to \infty$. The Kendall's τ is $\rho_{\tau,l}(\boldsymbol{v}) = 1 - \eta_l^{-1}(\boldsymbol{v})$, taking values in $[0, 1]$. We define $\rho_{\tau,l}(\boldsymbol{v}) := k_l(\|\boldsymbol{v} - \boldsymbol{v}_{(l)}\|)$, an

isotropic correlation function. Let $g_l(x) = (1-x)^{-1}$. Then, the function $h_l(\|\boldsymbol{v}-\boldsymbol{v}_{(l)}\|) = g_l \circ k_l(\|\boldsymbol{v}-\boldsymbol{v}_{(l)}\|) = (1-k_l(\|\boldsymbol{v}-\boldsymbol{v}_{(l)}\|))^{-1}$. Thus, the parameter $\eta_l(\boldsymbol{v}) \equiv \eta_l(\|\boldsymbol{v}-\boldsymbol{v}_{(l)}\|)$ is given by $h_l(\|\boldsymbol{v}-\boldsymbol{v}_{(l)}\|)$, and $\eta_l(\boldsymbol{v}) \to \infty$ as $\|\boldsymbol{v}-\boldsymbol{v}_{(l)}\| \to 0$.

After we define a spatially varying copula, we obtain a family of copula NNMPs by choosing a desired family of marginal distributions. Examples of NNMPs with different copulas and marginals are illustrated in Section 5.1.

Copula NNMP models offer avenues to capture complex dependence using general bivariate copulas. Traditional spatial copula models specify the finite dimensional distributions of the underlying spatial process with a multivariate copula. However, the use of multivariate copulas requires careful consideration in a spatial setting. For example, it is common to assume that spatial processes exhibit stronger dependence at smaller distances. Thus, copulas such as the multivariate Archimedean copula, that induce an exchangeable dependence structure, are inappropriate. Though spatial vine copula models (Gräler, 2014) can resolve this restriction, their model structure and computation are substantially more complicated than copula NNMP models.

3.3 Modeling Multivariate Dependence via Mixture Components

Section 3.1 and 3.2 develop approaches to constructing NNMP weights and component densities. The versatility of the framework for modeling non-Gaussian data is demonstrated with the examples of Section 3.2. We have shown that particular non-Gaussian characteristics can be modeled using appropriate sets of base random vectors (U_l, V_l). Spatial dependence between U_l and V_l is then introduced via a chosen dependence measure which is not limited to Pearson's correlation. Combined with the idea of conditional probability modeling via structured mixtures, such a strategy provides a general approach to model non-Gaussian processes with spatially varying dependence.

In this section, we demonstrate that specific NNMP model properties are driven by distributional properties of the base random vectors (U_l, V_l). Note that the NNMP conditional density in (2.4) corresponds to random vector $(Z(\boldsymbol{v}), Z(\boldsymbol{v}_{(1)}), \ldots, Z(\boldsymbol{v}_{(L)}))^\top$. Its multivariate dependence can be quantified through the dependence structure of the bivariate distributions for (U_l, V_l). We illustrate this feature by obtaining lower bounds for two measures used to assess the strength of tail dependence.

The main assumption to establish tail dependence results is that the base random vector (U_l, V_l) has stochastically increasing positive dependence. U_l is said to be stochastically increasing in V_l, if $P\big(U_l > u_l \,|\, V_l = v_l\big)$ increases as v_l increases. The definition can be extended to a multivariate random vector $(Z_1, \ldots, Z_p)$. Z_1 is said to be stochastically increasing in $(Z_2, \ldots, Z_p)$ if $P\big(Z_1 > z_1 \,|\, Z_2 = z_2, \ldots, Z_p = z_p\big) \le P\big(Z_1 > z_1 \,|\, Z_2 = z_2', \ldots, Z_p = z_p'\big)$, for all $(z_2, \ldots, z_p)$ and $(z_2', \ldots, z_p')$ in the support of $(Z_2, \ldots, Z_p)$, where $z_j \le z_j'$, for $j = 2, \ldots, p$. The conditional density in (2.4) implies that

$$P\big(Z(\boldsymbol{v}) > z \,|\, \boldsymbol{Z}_{\mathrm{Ne}(\boldsymbol{v})} = \boldsymbol{z}_{\mathrm{Ne}(\boldsymbol{v})}\big) = \sum_{l=1}^{L} w_l(\boldsymbol{v})\, P\big(Z(\boldsymbol{v}) > z \,|\, Z(\boldsymbol{v}_{(l)}) = z(\boldsymbol{v}_{(l)})\big).$$

Therefore, $Z(\boldsymbol{v})$ is stochastically increasing in $\boldsymbol{Z}_{\mathrm{Ne}(\boldsymbol{v})}$ if $Z(\boldsymbol{v})$ is stochastically increasing in $Z(\boldsymbol{v}_{(l)})$ with respect to $(U_{\boldsymbol{v},l}, V_{\boldsymbol{v},l})$ for all l. If the sequence $(U_{\boldsymbol{v},l}, V_{\boldsymbol{v},l})$ is built from the vector (U_l, V_l), then the set of base random vectors determines the stochastically increasing positive dependence of $Z(\boldsymbol{v})$ given its neighbors.

For a bivariate random vector (U_l, V_l), the upper and lower tail dependence coefficients, denoted as $\lambda_{\mathcal{H},l}$ and $\lambda_{\mathcal{L},l}$, respectively, are $\lambda_{\mathcal{H},l} = \lim_{q\to 1^-} P\big(U_l > F_{U_l}^{-1}(q) \,|\, V_l > F_{V_l}^{-1}(q)\big)$ and $\lambda_{\mathcal{L},l} = \lim_{q\to 0^+} P\big(U_l \le F_{U_l}^{-1}(q) \,|\, V_l \le F_{V_l}^{-1}(q)\big)$. When $\lambda_{\mathcal{H},l} > 0$, we say U_l and V_l have upper tail dependence. When $\lambda_{\mathcal{H},l} = 0$, U_l and V_l are said to be asymptotically independent in the upper tail. Lower tail dependence and asymptotically independence in the lower tail are similarly defined using $\lambda_{\mathcal{L},l}$. Let $F_{Z(\boldsymbol{v})}$ be the marginal distribution function of $Z(\boldsymbol{v})$. Analogously, we can define the upper and lower tail dependence coefficients for $Z(\boldsymbol{v})$ given its nearest neighbors,

$$\lambda_{\mathcal{H}}(\boldsymbol{v}) = \lim_{q\to 1^-} P\big(Z(\boldsymbol{v}) > F_{Z(\boldsymbol{v})}^{-1}(q) \mid Z(\boldsymbol{v}_{(1)}) > F_{Z(\boldsymbol{v}_{(1)})}^{-1}(q), \ldots, Z(\boldsymbol{v}_{(L)}) > F_{Z(\boldsymbol{v}_{(L)})}^{-1}(q)\big),$$
$$\lambda_{\mathcal{L}}(\boldsymbol{v}) = \lim_{q\to 0^+} P\big(Z(\boldsymbol{v}) \le F_{Z(\boldsymbol{v})}^{-1}(q) \mid Z(\boldsymbol{v}_{(1)}) \le F_{Z(\boldsymbol{v}_{(1)})}^{-1}(q), \ldots, Z(\boldsymbol{v}_{(L)}) \le F_{Z(\boldsymbol{v}_{(L)})}^{-1}(q)\big).$$

The following proposition provides lower bounds for the tail dependence coefficients.

Proposition 3. *Consider an NNMP for which the component density $f_{\boldsymbol{v},l}$ is specified by the conditional density of $U_{\boldsymbol{v},l}$ given $V_{\boldsymbol{v},l}$, where the random vector $(U_{\boldsymbol{v},l}, V_{\boldsymbol{v},l})$ follows a bivariate distribution with marginal distribution functions $F_{U_{\boldsymbol{v},l}}$ and $F_{V_{\boldsymbol{v},l}}$, for $l = 1, \ldots, L$. The spatial dependence of $(U_{\boldsymbol{v},l}, V_{\boldsymbol{v},l})$ is built from the base vector (U_l, V_l), which has a bivariate distribution such that U_l is stochastically increasing in V_l, for $l = 1, \ldots, L$. Then, for every $\boldsymbol{v}$, the lower bound for the upper tail dependence coefficient $\lambda_{\mathcal{H}}(\boldsymbol{v})$ is $\sum_{l=1}^{L} w_l(\boldsymbol{v}) \lim_{q\to 1^-} P\big(Z(\boldsymbol{v}) > F_{U_{\boldsymbol{v},l}}^{-1}(q) \,|\, Z(\boldsymbol{v}_{(l)}) = F_{V_{\boldsymbol{v},l}}^{-1}(q)\big)$, and the lower bound for the lower tail dependence coefficient $\lambda_{\mathcal{L}}(\boldsymbol{v})$ is $\sum_{l=1}^{L} w_l(\boldsymbol{v}) \lim_{q\to 0^+} P\big(Z(\boldsymbol{v}) \le F_{U_{\boldsymbol{v},l}}^{-1}(q) \,|\, Z(\boldsymbol{v}_{(l)}) = F_{V_{\boldsymbol{v},l}}^{-1}(q)\big)$.*

Proposition 3 establishes that the lower and upper tail dependence coefficients are bounded below by a convex combination of, respectively, the limits of the conditional distribution functions and the conditional survival functions. These are fully determined by the dependence structure of the bivariate distribution for (U_l, V_l). We illustrate the result of Proposition 3 with the following example.

Lomax NNMP Models Consider bivariate Lomax distributions (Arnold et al., 1999) for (U_l, V_l), $l = 1, \ldots, L$, specified through Lomax conditionals: $f_{U_l|V_l} = \mathrm{Lo}(u \,|\, v + \phi_l, \alpha_l)$ and $f_{V_l|U_l} = \mathrm{Lo}(v \,|\, u + \phi_l, \alpha_l)$. Here, $\mathrm{Lo}(x \,|\, \phi, \alpha) = \alpha\phi^{-1}(1 + x\phi^{-1})^{-(\alpha+1)}$ denotes the Lomax density, i.e., a shifted version of the Pareto Type I density. A small value of α indicates a heavy tail. Extending α_l to $\alpha_l(\boldsymbol{v})$, we obtain a Lomax NNMP conditional density, $p(z(\boldsymbol{v}) \,|\, \boldsymbol{z}_{\mathrm{Ne}(\boldsymbol{v})}) = \sum_{l=1}^{L} w_l(\boldsymbol{v}) \,\mathrm{Lo}(z(\boldsymbol{v}) \,|\, z(\boldsymbol{v}_{(l)}) + \phi_l, \alpha_l(\boldsymbol{v}))$. The component conditional survival function of the Lomax NNMP, expressed in terms of the quantile q, is $\big\{1 + F_{U_{\boldsymbol{v},l}}^{-1}(q)/\big(F_{V_{\boldsymbol{v},l}}^{-1}(q) + \phi_l\big)\big\}^{-\alpha_l(\boldsymbol{v})}$ which converges to $2^{-\alpha_l(\boldsymbol{v})}$ as $q \to 1^-$. Therefore, the lower bound for $\lambda_{\mathcal{H}}(\boldsymbol{v})$ is $\sum_{l=1}^{L} w_l(\boldsymbol{v})\, 2^{-\alpha_l(\boldsymbol{v})}$. As $\alpha_l(\boldsymbol{v}) \to 0$ for all l, the lower bound

for $\lambda_{\mathcal{H}}(\boldsymbol{v})$ tends to one, and hence $\lambda_{\mathcal{H}}(\boldsymbol{v})$ tends to one, since $\lambda_{\mathcal{H}}(\boldsymbol{v}) \leq 1$. As $\alpha_l(\boldsymbol{v}) \to \infty$ for all l, the lower bound tends to zero.

The Lomax example demonstrates an alternate strategy to construct NNMP models: specify the bivariate distribution of (U_l, V_l) through compatible conditionals $f_{U_l|V_l}$ and $f_{V_l|U_l}$, and extend $f_{U_l|V_l}$ to $f_{U_{\boldsymbol{v},l}|V_{\boldsymbol{v},l}}$. Compatibility of conditionals refers to the existence of a bivariate distribution that has the given conditionals (Arnold and Press, 1989). This strategy provides the means to build NNMPs directly according to desired conditional dependence. For example, the Lomax model attains perfect upper tail dependence coefficient, with respect to the conditional density in (2.4), when its component shape parameters $\alpha_l(\boldsymbol{v}) \to 0$ for all l. In general, NNMPs built from compatible conditionals do not have a stationary marginal distribution. This is in contrast to the examples of Section 3.2 where NNMP models are specified focusing on marginals.

Proposition 3 holds for the general framework. If the distribution of (U_l, V_l) with $F_{U_l} = F_{V_l}$ has first-order partial derivatives and exchangeable dependence, namely (U_l, V_l) and (V_l, U_l) have the same joint distribution, the lower bounds of the tail dependence coefficients depend on the component tail dependence coefficients. The result is summarized in the following corollary.

Corollary 1. *Suppose that the base random vector (U_l, V_l) in Proposition 3 is exchangeable, and its bivariate distribution with marginals $F_{U_l} = F_{V_l}$ has first-order partial derivatives, for all l. Then the upper and lower tail dependence coefficients $\lambda_{\mathcal{H}}(\boldsymbol{v})$ and $\lambda_{\mathcal{L}}(\boldsymbol{v})$ are bounded below by $\sum_{l=1}^{L} w_l(\boldsymbol{v})\lambda_{\mathcal{H},l}(\boldsymbol{v})/2$ and $\sum_{l=1}^{L} w_l(\boldsymbol{v})\lambda_{\mathcal{L},l}(\boldsymbol{v})/2$, where $\lambda_{\mathcal{H},l}(\boldsymbol{v})$ and $\lambda_{\mathcal{L},l}(\boldsymbol{v})$ are the tail dependence coefficients with respect to $(U_{\boldsymbol{v},l}, V_{\boldsymbol{v},l})$.*

Under Corollary 1, if the bivariate distribution of (U_l, V_l) is symmetric, for instance, an elliptically symmetric distribution, the upper and lower tail dependence coefficients coincide, and can simply be denoted as $\lambda(\boldsymbol{v})$. It follows that $\lambda(\boldsymbol{v}) \geq \sum_{l=1}^{L} w_l(\boldsymbol{v})\lambda_l(\boldsymbol{v})/2$, where $\lambda_l(\boldsymbol{v})$ is the tail dependence coefficient with respect to $(U_{\boldsymbol{v},l}, V_{\boldsymbol{v},l})$.

Tail dependence can also be quantified using the boundary of the conditional distribution function, as proposed in Hua and Joe (2014) for a bivariate random vector. In particular, the upper tail dependence of (U_l, V_l) is said to have some strength if $F_{U_l|V_l}\big(F_{U_l}^{-1}(q) \,|\, F_{V_l}^{-1}(1)\big)$ is positive at $q = 1$. Likewise, a non-zero $F_{U_l|V_l}\big(F_{U_l}^{-1}(q) \,|\, F_{V_l}^{-1}(0)\big)$ at $q = 0$ indicates some strength of dependence in the lower tails. The functions $F_{U_l|V_l}\big(\cdot \mid F_{V_l}^{-1}(0)\big)$ and $F_{U_l|V_l}\big(\cdot \mid F_{V_l}^{-1}(1)\big)$ are referred to as the boundary conditional distribution functions.

We use $F_{1|2}\big(\cdot \,|\, F^{-1}_{\boldsymbol{Z}_{\mathrm{Ne}(\boldsymbol{v})}}(q)\big)$ for simpler notation for the conditional distribution function of $Z(\boldsymbol{v})$, $F\big(\cdot \,|\, Z(\boldsymbol{v}_{(1)}) = F^{-1}_{Z(\boldsymbol{v}_{(1)})}(q), \ldots, Z(\boldsymbol{v}_{(L)}) = F^{-1}_{Z(\boldsymbol{v}_{(L)})}(q)\big)$. Then $F_{1|2}\big(\cdot \,|\, F^{-1}_{\boldsymbol{Z}_{\mathrm{Ne}(\boldsymbol{v})}}(0)\big)$ and $F_{1|2}\big(\cdot \,|\, F^{-1}_{\boldsymbol{Z}_{\mathrm{Ne}(\boldsymbol{v})}}(1)\big)$ are the boundary conditional distribution functions for the NNMP model. The upper tail dependence is said to be i) strongest if $F_{1|2}\big(F^{-1}_{Z(\boldsymbol{v})}(q) \,|\, F^{-1}_{\boldsymbol{Z}_{\mathrm{Ne}(\boldsymbol{v})}}(1)\big)$ equals 0 for $0 \leq q < 1$ and has a mass of 1 at $q = 1$; ii) intermediate if $F_{1|2}\big(F^{-1}_{Z(\boldsymbol{v})}(q) \,|\, F^{-1}_{\boldsymbol{Z}_{\mathrm{Ne}(\boldsymbol{v})}}(1)\big)$ has positive but not unit mass at $q = 1$; iii) weakest if $F_{1|2}\big(F^{-1}_{Z(\boldsymbol{v})}(q) \,|\, F^{-1}_{\boldsymbol{Z}_{\mathrm{Ne}(\boldsymbol{v})}}(1)\big)$ has no mass at $q = 1$. The strength of lower tail dependence

is defined likewise using $F_{1|2}\big(F^{-1}_{Z(\boldsymbol{v})}(q) \,|\, F^{-1}_{\boldsymbol{Z}_{\mathrm{Ne}(\boldsymbol{v})}}(0)\big)$. The following result provides lower bounds for the boundary conditional distribution functions.

Proposition 4. *Consider an NNMP for which the component density $f_{\boldsymbol{v},l}$ is specified by the conditional density of $U_{\boldsymbol{v},l}$ given $V_{\boldsymbol{v},l}$. The spatial dependence of the random vector $(U_{\boldsymbol{v},l}, V_{\boldsymbol{v},l})$ is built from the base vector (U_l, V_l), which has a bivariate distribution such that U_l is stochastically increasing in V_l, for $l = 1, \ldots, L$. Let $\lambda_{\mathcal{L},l}(\boldsymbol{v})$ and $\lambda_{\mathcal{H},l}(\boldsymbol{v})$ be the lower and upper tail dependence coefficients corresponding to $(U_{\boldsymbol{v},l}, V_{\boldsymbol{v},l})$. If for a given $\boldsymbol{v}$, there exists $\lambda_{\mathcal{L},l}(\boldsymbol{v}) > 0$ for some l, then the conditional distribution function $F_{1|2}\big(F^{-1}_{Z(\boldsymbol{v})}(q) \,|\, F^{-1}_{\boldsymbol{Z}_{Ne(\boldsymbol{v})}}(0)\big)$ has strictly positive mass $p_0(\boldsymbol{v})$ at $q = 0$ with $p_0(\boldsymbol{v}) \geq \sum_{l=1}^{L} w_l(\boldsymbol{v})\lambda_{\mathcal{L},l}(\boldsymbol{v})$. Similarly, if for a given $\boldsymbol{v}$, there exists $\lambda_{\mathcal{H},l}(\boldsymbol{v}) > 0$ for some l, then the conditional distribution function $F_{1|2}\big(F^{-1}_{Z(\boldsymbol{v})}(q) \,|\, F^{-1}_{\boldsymbol{Z}_{Ne(\boldsymbol{v})}}(1)\big)$ has strictly positive mass $p_1(\boldsymbol{v})$ at $q = 1$ with $p_1(\boldsymbol{v}) \geq \sum_{l=1}^{L} w_l(\boldsymbol{v})\lambda_{\mathcal{H},l}(\boldsymbol{v})$.*

Proposition 4 complements Proposition 3 to assess strength of tail dependence. When interest lies in tail dependence properties, these propositions can help guide the choice of appropriate families of bivariate distributions for the random vectors (U_l, V_l). The results are particularly useful for bivariate distributions build from copulas, which yield explicit expressions for the tail dependence coefficients. For example, the spatial Gumbel copula $C_{\boldsymbol{v},l}$ in Section 3.2 has upper tail dependence coefficient $2 - 2^{1/\eta_l(\boldsymbol{v})} > 0$ for $\eta_l(\boldsymbol{v}) > 1$. This implies that the tail dependence of a Gumbel copula NNMP model has some strength if $\eta_l(\boldsymbol{v}) > 1$ for some l. In fact, applying the result in Hua and Joe (2014), with a Gumbel copula, $F_{1|2}\big(F^{-1}_{Z(\boldsymbol{v})}(q) \,|\, F^{-1}_{\boldsymbol{Z}_{\mathrm{Ne}(\boldsymbol{v})}}(1)\big)$ degenerates at $q = 1$, implying strongest tail dependence. An example that implements the Gumbel copula NNMP model can be found in Section 5.1.

4 Bayesian Hierarchical Model and Inference

4.1 Hierarchical Model Formulation

We introduce the general approach for NNMP Bayesian implementation, treating the observed spatial responses as an NNMP realization. The inferential framework can be easily extended to incorporate model components that may be needed in practical settings, such as covariates and additional error terms. We illustrate the extensions with the real data analysis in Section 5.2 and in the Supplementary Material, and provide further discussion in Section 6.

Our approach for inference is based on a likelihood conditional on the first L elements of the realization $\boldsymbol{z}_{\mathcal{S}} = (z(\boldsymbol{s}_1), \ldots, z(\boldsymbol{s}_n))^\top$ over the reference set $\mathcal{S} \subset \mathcal{D}$. Following a commonly used approach for mixture models fitting, we use data augmentation to facilitate inference. For $z(\boldsymbol{s}_i)$, $i = L+1, \ldots, n$, we introduce a configuration variable ℓ_i, taking values in $\{1, \ldots, L\}$, such that $P\big(\ell_i \,|\, \boldsymbol{w}(\boldsymbol{s}_i)\big) = \sum_{l=1}^{L} w_l(\boldsymbol{s}_i)\delta_l(\ell_i)$, where $\delta_l(\ell_i) = 1$ if $\ell_i = l$ and 0 otherwise. Conditional on the configuration variables and the vector

$(z(\boldsymbol{s}_1),\ldots,z(\boldsymbol{s}_L))^\top$, the augmented model on $z(\boldsymbol{s}_i)$ is

$$\begin{aligned} z(\boldsymbol{s}_i) \mid z(\boldsymbol{s}_{(i,\ell_i)}), \ell_i, \boldsymbol{\theta} &\overset{ind.}{\sim} f_{\boldsymbol{s}_i,\ell_i}(z(\boldsymbol{s}_i) \mid z(\boldsymbol{s}_{(i,\ell_i)}), \boldsymbol{\theta}), \\ \ell_i \mid \boldsymbol{w}(\boldsymbol{s}_i) &\overset{ind.}{\sim} \sum_{l=1}^{L} w_l(\boldsymbol{s}_i)\delta_l(\ell_i), \end{aligned} \tag{4.1}$$

where $\boldsymbol{\theta}$ collects the parameters of the densities $f_{\boldsymbol{s}_i,l}$ for all l.

The full Bayesian model is completed with prior specification for parameters $\boldsymbol{\theta}, \boldsymbol{\zeta}, \boldsymbol{\gamma}$, and κ^2. The priors for $\boldsymbol{\theta}$ and $\boldsymbol{\zeta}$ depend on the choices of the densities $f_{\boldsymbol{s}_i,l}$ and the cutoff point kernel k', respectively. For parameters $\boldsymbol{\gamma}$ and κ^2, we specify $N(\boldsymbol{\gamma} \mid \boldsymbol{\mu}_\gamma, \boldsymbol{V}_\gamma)$ and $\text{IG}(\kappa^2 \mid u_{\kappa^2}, v_{\kappa^2})$ priors, respectively, where IG denotes the inverse gamma distribution.

4.2 Estimation and Prediction

We implement an MCMC sampler to simulate from the posterior distribution of the model parameters. To allow for efficient simulation of $\boldsymbol{\gamma}$ and κ^2, we associate each $y(\boldsymbol{s}_i)$ with a latent Gaussian variable t_i with mean $\mu(\boldsymbol{s}_i)$ and variance κ^2. There is a one-to-one correspondence between the configuration variables ℓ_i and latent variables t_i: $\ell_i = l$ if and only if $t_i \in (r^*_{\boldsymbol{s}_i,l-1}, r^*_{\boldsymbol{s}_i,l})$ where $r^*_{\boldsymbol{s}_i,l} = \log(r_{\boldsymbol{s}_i,l}/(1 - r_{\boldsymbol{s}_i,l}))$, for $l = 1,\ldots,L$. The posterior distribution of the model parameters, based on the new augmented model, is

$$\begin{aligned} p(\boldsymbol{\theta},\boldsymbol{\zeta},\boldsymbol{\gamma},\kappa^2,\{t_i\}_{i=L+1}^n \mid \boldsymbol{z}_{\mathcal{S}}) \propto\ & \pi_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \times \pi_{\boldsymbol{\zeta}}(\boldsymbol{\zeta}) \times N(\boldsymbol{\gamma} \mid \boldsymbol{\mu}_\gamma, \boldsymbol{V}_\gamma) \times \text{IG}(\kappa^2 \mid u_{\kappa^2}, v_{\kappa^2}) \\ & \times N(\boldsymbol{t} \mid \boldsymbol{D}\boldsymbol{\gamma},\, \kappa^2 \text{I}_{n-L}) \times \prod_{i=L+1}^{n} \sum_{l=1}^{L} f_{\boldsymbol{s}_i,l}(z(\boldsymbol{s}_i) \mid z(\boldsymbol{s}_{(il)}), \boldsymbol{\theta})\, \mathbb{1}_{(r^*_{\boldsymbol{s}_i,l-1}, r^*_{\boldsymbol{s}_i,l})}(t_i), \end{aligned}$$

where $\pi_{\boldsymbol{\theta}}$ and $\pi_{\boldsymbol{\zeta}}$ are the priors for $\boldsymbol{\theta}$ and $\boldsymbol{\zeta}$, respectively, I_{n-L} is an $(n-L)\times(n-L)$ identity matrix, the vector $\boldsymbol{t} = (t_{L+1},\ldots,t_n)^\top$, and the matrix $\boldsymbol{D}$ is $(n-L)\times 3$ such that the ith row is $(1, s_{L+i,1}, s_{L+i,2})$.

The posterior full conditional distribution of $\boldsymbol{\theta}$ depends on the form of $f_{\boldsymbol{s}_i,l}$. Details for the models implemented in Section 5 are given in the Supplementary Material. To update $\boldsymbol{\zeta}$, we first marginalize out the latent variables t_i from the joint posterior distribution. We then update $\boldsymbol{\zeta}$ using a random walk Metropolis step with target density $\pi_{\boldsymbol{\zeta}}(\boldsymbol{\zeta}) \prod_{i=L+1}^n \{G_{\boldsymbol{s}_i}(r_{\boldsymbol{s}_i,\ell_i} \mid \mu(\boldsymbol{s}_i), \kappa^2) - G_{\boldsymbol{s}_i}(r_{\boldsymbol{s}_i,\ell_i-1} \mid \mu(\boldsymbol{s}_i), \kappa^2)\}$. The posterior full conditional distribution of t_i is $\sum_{l=1}^{i_L} q_l(\boldsymbol{s}_i)\text{TN}(t_i \mid \mu(\boldsymbol{s}_i), \kappa^2; r^*_{\boldsymbol{s}_i,l-1}, r^*_{\boldsymbol{s}_i,l})$, where $q_l(\boldsymbol{s}_i) \propto w_l(\boldsymbol{s}_i) f_{\boldsymbol{s}_i,l}(z(\boldsymbol{s}_i) \mid z(\boldsymbol{s}_{(il)}), \boldsymbol{\theta})$ and $w_l(\boldsymbol{s}_i) = G_{\boldsymbol{s}_i}(r_{\boldsymbol{s}_i,l} \mid \mu(\boldsymbol{s}_i), \kappa^2) - G_{\boldsymbol{s}_i}(r_{\boldsymbol{s}_i,l-1} \mid \mu(\boldsymbol{s}_i), \kappa^2)$, for $l = 1,\ldots,L$. Hence, each t_i can be updated by sampling from the l-th truncated Gaussian with probability proportional to $q_l(\boldsymbol{s}_i)$. The posterior full conditional distribution 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 $\boldsymbol{\mu}^*_\gamma = \boldsymbol{V}^*_\gamma(\boldsymbol{V}_\gamma^{-1}\boldsymbol{\mu}_\gamma + \kappa^{-2}\boldsymbol{D}^\top\boldsymbol{t})$. The posterior full conditional distribution of κ^2 is $\text{IG}(\kappa^2 \mid u_{\kappa^2} + (n-L)/2,\ v_{\kappa^2} + \sum_{i=L+1}^n (t_i - \mu(\boldsymbol{s}_i))^2/2)$.

Turning to prediction, let $\boldsymbol{v}_0 \in \mathcal{D}$. We obtain posterior predictive samples of $z(\boldsymbol{v}_0)$ as follows. If $\boldsymbol{v}_0 \notin \mathcal{S}$, for each posterior sample of the parameters, we first compute

the cutoff points $r_{\boldsymbol{v}_0,l}$, such that $r_{\boldsymbol{v}_0,l} - r_{\boldsymbol{v}_0,l-1} = k'(\boldsymbol{v}_0, \boldsymbol{v}_{(0l)} \,|\, \boldsymbol{\zeta}) / \sum_{l=1}^{L} k'(\boldsymbol{v}_0, \boldsymbol{v}_{(0l)} \,|\, \boldsymbol{\zeta})$, and obtain the weights $w_l(\boldsymbol{v}_0) = G_{\boldsymbol{v}_0}(r_{\boldsymbol{v}_0,l} \,|\, \mu(\boldsymbol{v}_0), \kappa^2) - G_{\boldsymbol{v}_0}(r_{\boldsymbol{v}_0,l-1} \,|\, \mu(\boldsymbol{v}_0), \kappa^2)$, for $l = 1, \ldots, L$. We then predict $z(\boldsymbol{v}_0)$ using (2.3). If $\boldsymbol{v}_0 \equiv \boldsymbol{s}_i \in \mathcal{S}$, we generate $z(\boldsymbol{v}_0)$ similarly, but using samples for the weights collected from the posterior simulation, and applying (2.2) instead of (2.3) to generate $z(\boldsymbol{v}_0)$.

5 Data Illustrations

5.1 Simulation Study

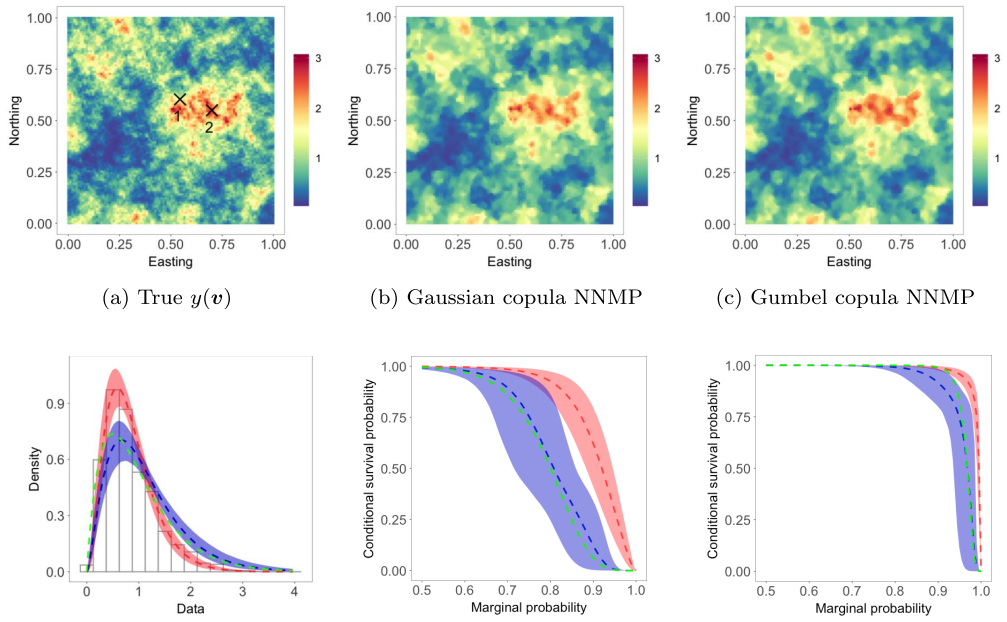
We have conducted several simulation experiments to study the inferential benefits of the NNMP modeling framework. Here, we focus on illustrating the proposed methodology, and present two synthetic data examples corresponding to data that are positive-valued and heavy-tailed, and have bounded support. We also evaluated NNMP model inference and prediction for different non-Gaussian characteristics, such as for data that are proportions or that arise under different levels of skewness, by comparison with alternative spatial models. The models used for comparison and other details of the simulation study can be found in the Supplementary Material. We note that in all simulation experiments, the underlying data generating mechanism is different from the NNMP.

In each experiment, we created a regular grid of 100×100 resolution on a unit square domain, and generated data over the grid. We then randomly selected 2000 locations as the reference set with a random ordering for model fitting. For the purpose of illustration, we chose neighborhood size $L = 10$ for all cases.

Results are based on posterior samples collected every 10 iterations from a Markov chain of 30000 iterations, with the first 10000 samples being discarded. Implementation details for all models are provided in the Supplementary Material. The MCMC algorithms were implemented in the R programming language on a computer with a 2-GHz Intel Core i5 processor and 32-GB RAM. We integrated C++ code for the update of latent variables without particular emphasis on optimizing the code. The computing time for the models in the following experiments was around 20 minutes.

First Simulation Experiment

We demonstrate the use of copulas to construct NNMPs for tail dependence modeling. Our focus is on illustrating the flexibility of copula NNMP models to capture complex dependence structures, and not specifically on extreme value modeling. To this end, we generated data from random field $y(\boldsymbol{v}) = F^{-1}\big(T_\nu(\omega(\boldsymbol{v}))\big), \boldsymbol{v} \in \mathcal{D}$; see Figure 1(a). Here, $\omega(\boldsymbol{v})$ is a standard Student-t process with tail parameter ν and scale matrix specified by an exponential correlation function with range parameter ϕ_w, and the distribution functions F and T_ν correspond to a gamma distribution, $\text{Ga}(2, 2)$ with mean 1, and a standard Student-t distribution with tail parameter ν, respectively. For a given pair of locations in $\mathcal{D}$ with correlation $\rho_0 = \exp(-d_0/\phi_w)$, the corresponding tail dependence coefficient of the random field is $\chi_\nu = 2T_{\nu+1}\big(-\sqrt{(1+\nu)(1-\rho_0)/(1+\rho_0)}\big)$. We took $\phi_w = 1/12$, and chose $\nu = 10$ so that the synthetic data exhibits moderate tail



(a) True $y(\boldsymbol{v})$ (b) Gaussian copula NNMP (c) Gumbel copula NNMP

(d) Estimated marginals (e) Site 1 estimated probabilities (f) Site 2 estimated probabilities

Figure 2: Synthetic data analysis – first simulation experiment. Top panels are the interpolated surface of the true field and posterior median estimates from both models. Bottom panels are estimated marginal densities and conditional survival probabilities from the two models. The green dashed lines correspond to the true model. The red (blue) dashed lines and shaded regions are the posterior mean and 95% credible interval estimates from the Gaussian (Gumbel) copula NNMP models.

dependence at close distance, and the dependence decreases rapidly as the distance d_0 becomes larger. In particular, for $\rho_0 = 0.05, 0.5, 0.95$, we obtain $\chi_{10} = 0.01, 0.08, 0.61$, respectively.

We applied two copula NNMP models. The models are of the form in (3.5) with stationary gamma marginal $\text{Ga}(a, b)$ with mean a/b. In the first model, the component copula density $c_{\boldsymbol{v},l}$ corresponds to a bivariate Gaussian copula, which is known to be unsuitable for tail dependence modeling. The spatially varying correlation parameter of the copula was specified by an exponential correlation function with range parameter ϕ_1. In the second model, we consider a spatial Gumbel copula as in Section 3.2. The copula spatially varying parameter is defined with the link function $\eta_l(\boldsymbol{v}) \equiv \eta_l(\|\boldsymbol{v} - \boldsymbol{v}_{(l)}\|) = \min\{(1 - \exp(-\|\boldsymbol{v} - \boldsymbol{v}_{(l)}\|/\phi_2))^{-1}, 50\}$, where the upper bound 50 ensures numerical stability. When $\eta_l(d_0) = 50$, $\exp(-d_0/\phi_2) = 0.98$. With this link function, we assume that given ϕ_2, the strength of the tail dependence with respect to the lth component of the Gumbel model, stays the same for any distance between two locations smaller than d_0. For the cutoff point kernels, we specified an exponential correlation function with range parameters ζ_1 and ζ_2, respectively, for each model. The Bayesian model is

completed with a IG$(3, 1/3)$ prior for ϕ_1 and ϕ_2, a Ga$(1, 1)$ prior for a and b, a IG$(3, 0.2)$ prior for ζ_1 and ζ_2, and $N(\boldsymbol{\gamma} \,|\, (-1.5, 0, 0)^\top, 2\mathrm{I}_3)$ and IG$(\kappa^2 \,|\, 3, 1)$ priors.

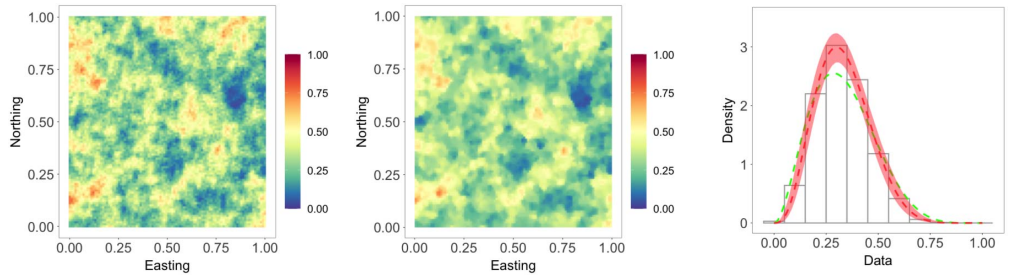
Figure 2 shows the random fields, marginal densities, and conditional survival probabilities estimated by the two models. From Figure 2(a)–2(c), we see that, comparing with the true field, the posterior median estimate by the Gumbel copula model seems to recover the large values better than the Gaussian copula model. Besides, as shown in Figure 2(d), the Gumbel copula model provides a more accurate estimate of the marginal distribution, especially in the tails. We computed the conditional survival probabilities at two different unobserved sites marked in Figure 2(a). In particular, sites 1 and 2 are surrounded by reference set observations with moderate and large values, respectively. The Gumbel copula model provides much more accurate estimates of the conditional survival probabilities, indicating that the model captures better the tail dependence structure in the data. Overall, this example demonstrates that the Gumbel copula NNMP model is a useful option for modeling spatial processes with tail dependence, as well as the flexibility of the NNMP framework for modeling complex dependence using different bivariate copulas.

Second Simulation Experiment

Many spatial processes are measured over a compact interval. As an example, data on proportions are common in ecological applications. In this experiment, we demonstrate the effectiveness of the NNMP model for directly modeling bounded spatial data. In particular, we generated data using the following model, $y(\boldsymbol{v}) = F^{-1}\big(\Phi(\omega(\boldsymbol{v}))\big)$, $\boldsymbol{v} \in \mathcal{D}$, where the distribution function F corresponds to a beta distribution, denoted as Beta(a_0, b_0), and $\omega(\boldsymbol{v})$ is a standard Gaussian process with exponential correlation function with range parameter 0.1. We set $a_0 = 3$, $b_0 = 6$.

We applied a Gaussian copula NNMP model with stationary marginal Beta(a, b), with the same spatial Gaussian copula and prior specification used in the first simulation experiment. Figure 3(b) shows the estimated random field which captures well the main features of the true field in Figure 3(a). The posterior mean and pointwise 95% credible interval of the estimated marginal density in Figure 3(c) overlay on the data histogram. These show that the beta NNMP estimation and prediction provide good approximation to the true field.

Finally, it is worth mentioning that implementing the beta NNMP model is simpler than fitting existing models for data corresponding to proportions. For example, a spatial Gaussian copula model, which corresponds to the data generating process of this experiment, involves computations for large matrices. Alternatively, if a multivariate non-Gaussian copula is used, the resulting likelihood can be intractable and require certain approximations. Another model that is commonly used in this setting is a spatial generalized linear mixed model. The spatial element in the model is introduced through the transformed mean of the observations. A sample-based approach to fit such model requires sampling a large number of highly correlated latent variables. We compared the Beta NNMP model's predictive performance with alternative models for bounded data generated from the two aforementioned settings. The model comparison results



(a) True $y(\boldsymbol{v})$ (b) Beta NNMP (c) Estimated marginals

Figure 3: Synthetic data analysis – second simulation experiment. Panels (a) and (b) are the interpolated surface of the true field and posterior median estimate from the beta NNMP model, respectively. In Panel (c), the green dotted line corresponds to the true marginal. The red dash line and shaded region are the posterior mean and pointwise 95% credible interval for the estimated marginal.

indicate that the beta NNMP model performs well. Interested readers are referred to the Supplementary Material for details.

5.2 Mediterranean Sea Surface Temperature Data Analysis

The study of Ocean's dynamics is crucial for understanding climate variability. One of the most valuable sources of information regarding the evolution of the state of the ocean is provided by the centuries-long record of temperature observations from the surface of the oceans. The record of sea surface temperatures consists of data collected over time at irregularly scattered locations. In this section, we examine the sea surface temperature from the Mediterranean Sea area during December 2003.

It is well known that the Mediterranean Sea area produces very heterogeneous temperature fields. A goal of the spatial analysis of sea surface temperature in the area is to generate a spatially continuous field that accounts for the complexity of the surrounding coastlines as well as the non-linear dynamics of the circulation system. An additional source of complexity comes from the data collection process. Historically, the observations are collected from different types of devices: buckets launched from navigating vessels, readings from the water intake of ships' engine rooms, moored buoys, and drifting buoys (Kirsner and Sansó, 2020). The source of some observations is known, but not all the data are labelled. A thorough case study will be needed to include all this information in order to account for possible heterogeneities due to the different measuring devices. That is beyond the scope of this paper. We will focus on demonstrating the ability of the proposed framework to model non-Gaussian spatial processes that, hopefully, capture the complexities of the physical process and the data collection protocol. We notice that in the original record several sites had multiple observations. In those cases we took the median of the observations, resulting in a total of 1966 observations. The data are shown in Figure 4.

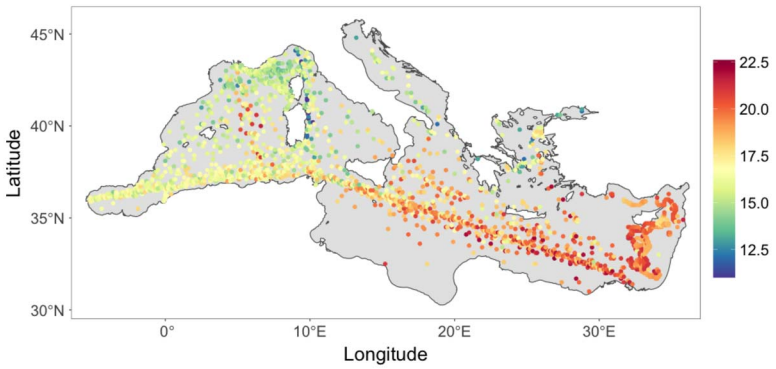


Figure 4: Mediterranean Sea surface temperature data analysis: observed sea surface temperature during December 2003.

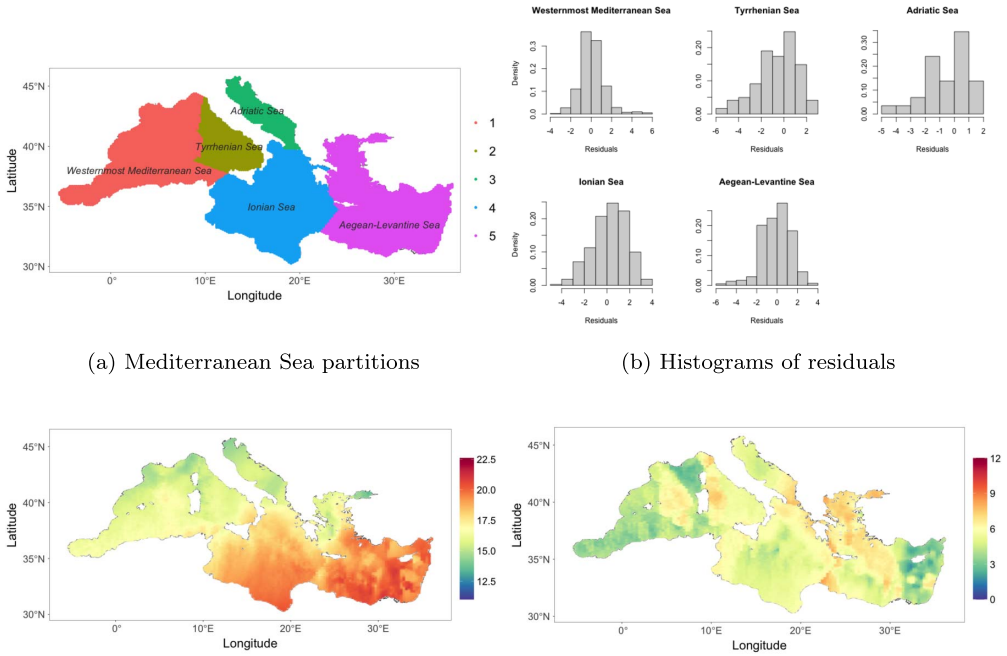
We first examine the Gaussianity assumption for the data. We compare the GNNMP and the NNGP models over a subset of the region where the ocean dynamics are known to be complex and the observations are heterogeneous. The model comparison is detailed in the Supplementary Material and the results support the GNNMP model.

In light of the evidence (Pisano et al., 2020) that sea surface temperature spatial patterns are different over Mediterranean sub-basins, shown in Figure 5(a), which are characterized by different dynamics and high variability of surface currents (Bouzaiene et al., 2020), we further investigate the sea surface temperature over those sub-basins. We fitted a non-spatial linear model to all data, including longitude and latitude as covariates, and obtained residuals from the linear model. Figure 5(b) shows that the histograms of the residuals are asymmetric over the sub-basins, indicating potential skewness in the marginal distribution, with levels of skewness that vary across sub-basins.

An Extended Skew-Gaussian NNMP Model

The exploratory data analysis suggests the need for a spatial model that can capture skewness. We thus analyze the full data set with an extension of the skew-GNNMP model in (3.4).

The new model has two features that extend the skew-GNNMP: (i) it incorporates a fixed effect through the location parameter of the mixture component; (ii) it allows the skewness parameter λ to vary in space. More specifically, the spatially varying conditional density $f_{\boldsymbol{v},l}$ of the model builds from a Gaussian random vector with mean $\left(\boldsymbol{x}(\boldsymbol{v})^\top\boldsymbol{\beta}+\lambda(\boldsymbol{v})z_0(\boldsymbol{v}),\ \boldsymbol{x}(\boldsymbol{v}_{(l)})^\top\boldsymbol{\beta}+\lambda(\boldsymbol{v}_{(l)})z_0(\boldsymbol{v})\right)^\top$ and covariance matrix $\sigma^2\begin{pmatrix}1 & \rho_l(\boldsymbol{v})\\ \rho_l(\boldsymbol{v}) & 1\end{pmatrix}$, where $\boldsymbol{x}(\boldsymbol{v})=(1,v_1,v_2)^\top$ and $z_0(\boldsymbol{v})\sim \mathrm{TN}(0,1;0,\infty)$, for all $\boldsymbol{v}$ and for all l, and (v_1,v_2) are longitude and latitude.



(a) Mediterranean Sea partitions

(b) Histograms of residuals

(c) 50% predicted sea surface temperature

(d) 95% credible interval width

Figure 5: Mediterranean Sea surface temperature data analysis. Panels (a) and (b) are partitions according to Mediterranean sub-basins and histograms of the residuals obtained from a non-spatial linear model. Panels (c) and (d) are posterior median and 95% credible interval estimates of the sea surface temperature from the extended skew-GNNMP model.

The conditional density $p(y(\boldsymbol{v}) \,|\, \boldsymbol{y}_{\mathrm{Ne}(\boldsymbol{v})})$ of the extended model is

$$\sum_{l=1}^{L} w_l(\boldsymbol{v}) \int_0^{\infty} N(y(\boldsymbol{v}) \,|\, \mu_l(\boldsymbol{v}), \sigma_l^2(\boldsymbol{v})) \mathrm{TN}(z_0(\boldsymbol{v}) \,|\, \mu_{0l}(\boldsymbol{v}_{(l)}), \sigma_{0l}^2(\boldsymbol{v}_{(l)}); 0, \infty) dz_0(\boldsymbol{v}), \quad (5.1)$$

with parameters $\mu_l(\boldsymbol{v}) = \boldsymbol{x}(\boldsymbol{v})^\top \boldsymbol{\beta} + \lambda(\boldsymbol{v}) z_0(\boldsymbol{v}) + \rho_l(\boldsymbol{v}) \{y(\boldsymbol{v}_{(l)}) - \boldsymbol{x}(\boldsymbol{v}_{(l)})^\top \boldsymbol{\beta} - \lambda(\boldsymbol{v}_{(l)}) z_0(\boldsymbol{v})\}$, $\sigma_l^2(\boldsymbol{v}) = \sigma^2 \{1 - (\rho_l(\boldsymbol{v}))^2\}$, $\mu_{0l}(\boldsymbol{v}_{(l)}) = \{y(\boldsymbol{v}_{(l)}) - \boldsymbol{x}(\boldsymbol{v}_{(l)})^\top \boldsymbol{\beta}\} \lambda(\boldsymbol{v}_{(l)}) / \{\sigma^2 + (\lambda(\boldsymbol{v}_{(l)}))^2\}$, and $\sigma_{0l}^2(\boldsymbol{v}_{(l)}) = \sigma^2 / \{\sigma^2 + (\lambda(\boldsymbol{v}_{(l)}))^2\}$. Marginalizing out $z_0(\boldsymbol{v})$, the marginal distribution of $Y(\boldsymbol{v})$ is $\mathrm{SN}\big(\boldsymbol{x}(\boldsymbol{v})^\top \boldsymbol{\beta}, (\lambda(\boldsymbol{v}))^2 + \sigma^2, \lambda(\boldsymbol{v})/\sigma\big)$, based on the result of Proposition 1.

We model the spatially varying $\lambda(\boldsymbol{v})$ via a partitioning approach. In particular, we partition the Mediterranean Sea $\mathcal{D}$ according to the sub-basins, that is, $\mathcal{D} = \bigcup_{k=1}^{K} P_k$, $P_i \cap P_j = \emptyset$ for $i \neq j$, where $K = 5$. For all $\boldsymbol{v} \in P_k$, we take $\lambda(\boldsymbol{v}) = \lambda_k$, for $k = 1, \ldots, K$. The partitions $P_1, \ldots, P_K$ correspond to the sub-basins: Westernmost Mediterranean Sea, Tyrrhenian Sea, Adriatic Sea, Ionian Sea, and Aegean-Levantine Sea, respectively.

We applied the extended skew-GNNMP model (5.1) using the whole data set as the reference set, with L chosen to be $10, 15$ or 20. The regression parameters $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^\top$ were assigned mean-zero, dispersed normal priors. For the parameters of the skew-Gaussian marginal, each element of $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_5)$ received a $N(0, 5)$ prior, and σ^2 was assigned an IG$(3, 1)$ prior. We used the same prior specification for other parameters as in the first simulation experiment. Posterior inference was based on thinned samples retaining every 4th iteration, from a total of 30000 samples with a burn-in of 10000 samples.

Results

We focus on the estimation of the regression and marginal distribution parameters $\boldsymbol{\beta}$, $\boldsymbol{\lambda}$ and σ^2. We report the estimates for $L = 15$; they were similar for $L = 10$ or 20. The posterior mean and 95% credible interval estimates of β_0, β_1, and β_2 were $29.07\,(24.80, 31.73)$, $0.12\,(0.09, 0.15)$, and $-0.34\,(-0.41, -0.23)$, indicating that there was an increasing trend in the sea surface temperature as longitude increased and latitude decreased. The corresponding posterior estimates of $(\lambda_1, \ldots, \lambda_5)$ were, respectively, $-0.39\,(-0.97, 0.18)$, $-1.40\,(-2.16, -0.66)$, $-2.49\,(-4.19, -1.16)$, $-1.58\,(-2.55, -0.58)$, and $-2.67\,(-3.90, -1.73)$, and the posterior estimate of σ^2 was $1.50\,(1.17, 1.97)$. These estimates suggest different levels of left skewness over the sub-basins except for the Westernmost Mediterranean Sea.

Figure 5(c) provides the posterior median estimate of the sea surface temperature over a dense grid of locations on the Mediterranean Sea. Compared to Figure 4, the estimate generally resembles the observed pattern. The prediction was quite smooth even for areas with few observations. The 95% credible interval width of the prediction over the gridded locations, as shown in Figure 5(d), demonstrates that the model describes the uncertainty in accordance with the observed data structure; the uncertainty is higher in areas where there are fewer observations or the observations are volatile.

6 Discussion

We have introduced a class of DAG-based mixture models for non-Gaussian processes. Existing approaches to handle non-Gaussian spatial data commonly rely on transforming the data or using latent Gaussian processes. This article adds an alternative that provides direct spatial modeling. On the other hand, it does not preclude the possibility of including a latent process in a hierarchical NNMP model, when the assumption of a latent process is of interest in real data applications. In fact, this idea is illustrated in Section 5.2 where the spatially-dependent skewness parameter is modeled using a partitioning approach. Additionally, the NNMP allows for working with different non-Gaussian characteristics in a unified manner, for example, skewed and heavy-tailed distributions, and distributions that are positive-valued or have bounded support. This is achieved through specifying bivariate distributions for the mixture components with appropriate marginals. We plan to develop a statistical package that implements NNMPs given a pre-specified family of marginals. The package will provide more accessibility to fully Bayesian inference for non-Gaussian spatial processes.

To incorporate covariates, the NNMP can be embedded in an additive or multiplicative model. The former is illustrated in the Supplementary Material with a spatially varying regression model. Under an additive model, the posterior simulation algorithm requires extra care as it involves sequential updating of the elements in $\boldsymbol{z}_{\mathcal{S}}$. This may induce slow convergence behavior. An alternative strategy for covariate inclusion is to model the weights or some parameter(s) of the spatially varying conditional density as a function of covariates. For example, in Section 5.2, we modeled the location parameter of the skew-Gaussian marginal as a linear function of the covariates. Posterior simulation under this approach is easily developed by modifying the update of the relevant parameters discussed in Section 4.2 to that of the regression coefficients.

The NNMP model structure not only bypasses all the potential issues from large matrix operations, but it also enhances modeling power. Kernel functions, such as wave covariance functions, that are impractical for Gaussian process-based models due to numerical instability from matrix inversion, can be used as link functions for the spatially varying parameter of the NNMP. On the other hand, using mixtures to represent a joint distribution may complicate study of general model properties under marginalization; possible exceptions include models that build from the GNNMP which has closed-form finite-dimensional distributions. The same challenge is also found in latent process modeling methods for non-Gaussian spatial data, such as in the class of spatial generalized linear mixed models. Another potential limitation due to the mixture model structure is that the posterior simulation algorithm may experience slow convergence issues. Further development is needed on efficient algorithms for fast computation, especially when dealing with massive, complex data sets.

We have focused in this article on a modeling framework for continuous data. The approach can be naturally extended to handle discrete spatial data. Combining the idea of building NNMPs based on copulas with discrete distributions offers a new modeling avenue to geostatistical models for discrete data, in which a parametric family is typically assumed for the data marginals. However, copulas linking discrete distributions hold different properties from continuous ones because of the unique structures in discrete random variables; see, e.g., the review in Genest and Nešlehová (2007). Consequently, copula modeling for discrete data typically requires careful treatment relative to its continuous analogues, as in, e.g., the pair copula construction (Panagiotelis et al., 2012). Regarding inference, it is well known that discrete copula models present computational challenges (Smith and Khaled, 2012). Extra development of algorithms for fast computation is needed when using the copula NNMP for discrete data. Therefore, we introduce the extension to modeling discrete data in a separate article (Zheng et al., 2023). Modeling options for geostatistical discrete data in the existing literature involve either spatial generalized linear mixed models or spatial copula models (Madsen, 2009). However, owing to their structures, both models have limitations with respect to the distributional assumption for the spatial random effects, as well as in computational efficiency. The extension to discrete NNMP models has the potential to provide both inferential and computational benefits in modeling large discrete data sets.

It is also interesting to explore analysis of spatial extremes using the NNMP framework. We developed guidelines in Section 3.3 to specify NNMPs focusing on strength

of tail dependence. The results highlight the ability of the NNMP model structure to capture local tail dependence at different levels, controlled by the mixture component bivariate distributions, e.g., with a class of bivariate extreme-value copulas. Moreover, using NNMPs for spatial extreme value modeling allows for efficient implementation of inference which is typically a challenge with existing approaches (Davison et al., 2012).

Other research directions include extensions to multivariate and spatio-temporal settings. The former extension requires families of high-dimensional multivariate distributions to construct an NNMP. Effective strategies will be needed to define the spatially varying multivariate distributions that balance flexibility and scalability. When it comes to a joint model over time and space, there is large scope for exploring the integration of the time component into the model, including extending the NNMP weights or the NNMP mixture components.

Supplementary Material

Supplementary Material for "Nearest-Neighbor Mixture Models for Non-Gaussian Spatial Processes" (DOI: 10.1214/23-BA1405SUPP; .pdf). The Supplementary Material includes proofs of the propositions and corollary, additional data examples, and implementation details. The source code and data are available at github.com/xzheng42.

References

Allcroft, D. J. and Glasbey, C. A. (2003). "A latent Gaussian Markov random-field model for spatiotemporal rainfall disaggregation." *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 52(4): 487–498. MR2012972. doi: https://doi.org/10.1111/1467-9876.00419. 1192

Arnold, B. C., Castillo, E., and Sarabia, J. M. (1999). *Conditional Specification of Statistical Models*. New York: Springer. MR1716531. 1204

Arnold, B. C. and Press, S. J. (1989). "Compatible conditional distributions." *Journal of the American Statistical Association*, 84(405): 152–156. MR0999673. 1205

Azzalini, A. (2013). *The Skew-Normal and Related Families*. Cambridge, UK: Cambridge University Press. MR3468021. 1201

Bárdossy, A. (2006). "Copula-based geostatistical models for groundwater quality parameters." *Water Resources Research*, 42(11). 1192

Beck, N., Genest, C., Jalbert, J., and Mailhot, M. (2020). "Predicting extreme surges from sparse data using a copula-based hierarchical Bayesian spatial model." *Environmetrics*, 31(5): e2616. MR4125212. doi: https://doi.org/10.1002/env.2616. 1192

Bevilacqua, M., Caamaño-Carrillo, C., Arellano-Valle, R. B., and Morales-Oñate, V. (2021). "Non-Gaussian geostatistical modeling using (skew) t processes." *Scandinavian Journal of Statistics*, 48(1): 212–245. MR4233171. doi: https://doi.org/10.1111/sjos.12447. 1192

Bevilacqua, M., Caamaño-Carrillo, C., and Gaetan, C. (2020). "On modeling positive continuous data with spatiotemporal dependence." *Environmetrics*, 31(7): e2632. MR4166854. doi: https://doi.org/10.1002/env.2632. 1191

Bolin, D. (2014). "Spatial Matérn fields driven by non-Gaussian noise." *Scandinavian Journal of Statistics*, 41(3): 557–579. MR3249417. doi: https://doi.org/10.1111/sjos.12046. 1192

Bolin, D. and Wallin, J. (2020). "Multivariate type G Matérn stochastic partial differential equation random fields." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(1): 215–239. MR4060983. 1192

Bouzaiene, M., Menna, M., Poulain, P.-M., Bussani, A., and Elhmaidi, D. (2020). "Analysis of the surface dispersion in the Mediterranean sub-basins." *Frontiers in Marine Science*, 7: 486. 1212

Bradley, J. R., Holan, S. H., and Wikle, C. K. (2018). "Computationally efficient multivariate spatio-temporal models for high-dimensional count-valued data (with discussion)." *Bayesian Analysis*, 13(1): 253–310. MR3773410. doi: https://doi.org/10.1214/17-BA1069. 1193

Bradley, J. R., Holan, S. H., and Wikle, C. K. (2020). "Bayesian hierarchical models with conjugate full-conditional distributions for dependent data from the natural exponential family." *Journal of the American Statistical Association*, 115(532): 2037–2052. MR4189775. doi: https://doi.org/10.1080/01621459.2019.1677471. 1193

Cadonna, A., Kottas, A., and Prado, R. (2019). "Bayesian spectral modeling for multiple time series." *Journal of the American Statistical Association*, 114: 1838–1853. MR4047304. doi: https://doi.org/10.1080/01621459.2018.1520114. 1199

Chan, A. B. and Dong, D. (2011). "Generalized Gaussian process models." In *CVPR*, 2681–2688. 1192

Cressie, N., Sainsbury-Dale, M., and Zammit-Mangion, A. (2022). "Basis-function models in spatial statistics." *Annual Review of Statistics and Its Application*, 9(1): 373–400. MR4394913. doi: https://doi.org/10.1146/annurev-statistics-040120-020733. 1199

Danaher, P. J. and Smith, M. S. (2011). "Modeling multivariate distributions using copulas: Applications in marketing." *Marketing Science*, 30(1): 4–21. 1192

Datta, A., Banerjee, S., Finley, A. O., and Gelfand, A. E. (2016a). "Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets." *Journal of the American Statistical Association*, 111(514): 800–812. MR3538706. doi: https://doi.org/10.1080/01621459.2015.1044091. 1192, 1196, 1199

Datta, A., Banerjee, S., Finley, A. O., Hamm, N. A., and Schaap, M. (2016b). "Nonseparable dynamic nearest neighbor Gaussian process models for large spatio-temporal data with an application to particulate matter analysis." *The Annals of Applied Statistics*, 10(3): 1286. MR3553225. doi: https://doi.org/10.1214/16-AOAS931. 1192

Davison, A. C., Padoan, S. A., and Ribatet, M. (2012). "Statistical modeling of spatial

extremes." *Statistical Science*, 27(2): 161–186. MR2963980. doi: https://doi.org/10.1214/11-STS376. 1216

De Oliveira, V., Kedem, B., and Short, D. A. (1997). "Bayesian prediction of transformed Gaussian random fields." *Journal of the American Statistical Association*, 92(440): 1422–1433. MR1615252. doi: https://doi.org/10.2307/2965412. 1192

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(3): 299–350. MR1626544. doi: https://doi.org/10.1111/1467-9876.00113. 1192

Eskelson, B. N., Madsen, L., Hagar, J. C., and Temesgen, H. (2011). "Estimating riparian understory vegetation cover with beta regression and copula models." *Forest Science*, 57(3): 212–221. 1191

Ferguson, T. S. (1973). "A Bayesian analysis of some nonparametric problems." *The Annals of Statistics*, 209–230. MR0350949. 1192

Finley, A. O., Datta, A., Cook, B. D., Morton, D. C., Andersen, H. E., and Banerjee, S. (2019). "Efficient algorithms for Bayesian nearest neighbor Gaussian processes." *Journal of Computational and Graphical Statistics*, 28(2): 401–414. MR3974889. doi: https://doi.org/10.1080/10618600.2018.1537924. 1192

Gelfand, A. E., Kottas, A., and MacEachern, S. N. (2005). "Bayesian nonparametric spatial modeling with Dirichlet process mixing." *Journal of the American Statistical Association*, 100(471): 1021–1035. MR2201028. doi: https://doi.org/10.1198/016214504000002078. 1192

Genest, C. and Nešlehová, J. (2007). "A primer on copulas for count data." *ASTIN Bulletin: The Journal of the IAA*, 37(2): 475–515. MR2422797. doi: https://doi.org/10.2143/AST.37.2.2024077. 1215

Ghosh, S. and Mallick, B. K. (2011). "A hierarchical Bayesian spatio-temporal model for extreme precipitation events." *Environmetrics*, 22(2): 192–204. MR2843351. doi: https://doi.org/10.1002/env.1043. 1192

Gräler, B. (2014). "Modelling skewed spatial random fields through the spatial vine copula." *Spatial Statistics*, 10: 87–102. MR3280092. doi: https://doi.org/10.1016/j.spasta.2014.01.001. 1203

Gramacy, R. B. and Apley, D. W. (2015). "Local Gaussian process approximation for large computer experiments." *Journal of Computational and Graphical Statistics*, 24(2): 561–578. MR3357395. doi: https://doi.org/10.1080/10618600.2014.914442. 1193

Guinness, J. (2018). "Permutation and grouping methods for sharpening Gaussian process approximations." *Technometrics*, 60(4): 415–429. MR3878098. doi: https://doi.org/10.1080/00401706.2018.1437476. 1193, 1196

Hua, L. and Joe, H. (2014). "Strength of tail dependence based on conditional tail expectation." *Journal of Multivariate Analysis*, 123: 143–159. MR3130426. doi: https://doi.org/10.1016/j.jmva.2013.09.001. 1205, 1206

Jin, B., Peruzzi, M., and Dunson, D. B. (2023). "Bag of DAGs: Inferring directional dependence in spatiotemporal processes." *arXiv preprint arXiv:2112.11870*. 1192

Joe, H. (2014). *Dependence Modeling with Copulas*. Boca Raton, FL: CRC Press. MR3328438. 1192, 1202

Johns, C. J., Nychka, D., Kittel, T. G. F., and Daly, C. (2003). "Infilling sparse records of spatial fields." *Journal of the American Statistical Association*, 98(464): 796–806. MR2055488. doi: https://doi.org/10.1198/016214503000000729. 1192

Jordan, M. I. (2004). "Graphical models." *Statistical Science*, 19(1): 140–155. MR2082153. doi: https://doi.org/10.1214/088342304000000026. 1194

Katzfuss, M. and Guinness, J. (2021). "A general framework for Vecchia approximations of Gaussian processes." *Statistical Science*, 36(1): 124–141. MR4194207. doi: https://doi.org/10.1214/19-STS755. 1193

Katzfuss, M. and Schäfer, F. (2023). "Scalable Bayesian transport maps for high-dimensional non-Gaussian spatial fields." *Journal of the American Statistical Association*, 1–15. 1192

Kim, H.-M. and Mallick, B. K. (2004). "A Bayesian prediction using the skew Gaussian distribution." *Journal of Statistical Planning and Inference*, 120(1-2): 85–101. MR2026484. doi: https://doi.org/10.1016/S0378-3758(02)00501-3. 1192

Kirsner, D. and Sansó, B. (2020). "Multi-scale shotgun stochastic search for large spatial datasets." *Computational Statistics & Data Analysis*, 146: 106931. MR4068969. doi: https://doi.org/10.1016/j.csda.2020.106931. 1211

Krupskii, P., Huser, R., and Genton, M. G. (2018). "Factor copula models for replicated spatial data." *Journal of the American Statistical Association*, 113(521): 467–479. MR3803479. doi: https://doi.org/10.1080/01621459.2016.1261712. 1192

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(436): 1504–1515. MR1439090. doi: https://doi.org/10.2307/2291576. 1195

Madsen, L. (2009). "Maximum likelihood estimation of regression parameters with spatially dependent discrete data." *Journal of Agricultural, Biological, and Environmental Statistics*, 14(4): 375–391. MR2750373. doi: https://doi.org/10.1198/jabes.2009.07116. 1215

Mahmoudian, B. (2017). "A skewed and heavy-tailed latent random field model for spatial extremes." *Journal of Computational and Graphical Statistics*, 26(3): 658–670. MR3698675. doi: https://doi.org/10.1080/10618600.2017.1302341. 1192

Morris, S. A., Reich, B. J., Thibaud, E., and Cooley, D. (2017). "A space-time skew-t model for threshold exceedances." *Biometrics*, 73(3): 749–758. MR3713109. doi: https://doi.org/10.1111/biom.12644. 1192

Müller, P., Quintana, F. A., and Page, G. (2018). "Nonparametric Bayesian inference

in applications." *Statistical Methods & Applications*, 27(2): 175–206. MR3807363. doi: https://doi.org/10.1007/s10260-017-0405-z. 1192

North, G. R., Wang, J., and Genton, M. G. (2011). "Correlation models for temperature fields." *Journal of Climate*, 24(22): 5850–5862. 1191

Palacios, M. B. and Steel, M. F. J. (2006). "Non-Gaussian Bayesian geostatistical modeling." *Journal of the American Statistical Association*, 101(474): 604–618. MR2281244. doi: https://doi.org/10.1198/016214505000001195. 1192

Panagiotelis, A., Czado, C., and Joe, H. (2012). "Pair copula constructions for multivariate discrete data." *Journal of the American Statistical Association*, 107(499): 1063–1072. MR3010894. doi: https://doi.org/10.1080/01621459.2012.682850. 1215

Paul, R. and Cressie, N. (2011). "Lognormal block kriging for contaminated soil." *European Journal of Soil Science*, 62(3): 337–345. 1191

Peruzzi, M., Banerjee, S., and Finley, A. O. (2020). "Highly scalable Bayesian geostatistical modeling via meshed Gaussian processes on partitioned domains." *Journal of the American Statistical Association*, 117(538): 969–982. MR4436326. doi: https://doi.org/10.1080/01621459.2020.1833889. 1192

Peruzzi, M. and Dunson, D. B. (2022a). "Spatial meshing for general Bayesian multivariate models." *arXiv preprint arXiv:2201.10080*. 1193

Peruzzi, M. and Dunson, D. B. (2022b). "Spatial multivariate trees for big data Bayesian regression." *Journal of Machine Learning Research*, 23: 17–1. MR4420742. 1192

Pisano, A., Marullo, S., Artale, V., Falcini, F., Yang, C., Leonelli, F. E., Santoleri, R., and Buongiorno Nardelli, B. (2020). "New evidence of Mediterranean climate change and variability from sea surface temperature observations." *Remote Sensing*, 12(1): 132. 1212

Rue, H., Martino, S., and Chopin, N. (2009). "Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2): 319–392. MR2649602. doi: https://doi.org/10.1111/j.1467-9868.2008.00700.x. 1193

Sainsbury-Dale, M., Zammit-Mangion, A., and Cressie, N. (2021). "Modelling, fitting, and prediction with non-Gaussian spatial and spatio-temporal data using FRK." *arXiv preprint arXiv:2110.02507*. 1193

Schäfer, F., Katzfuss, M., and Owhadi, H. (2021). "Sparse Cholesky factorization by Kullback–Leibler minimization." *SIAM Journal on Scientific Computing*, 43(3): A2019–A2046. MR4267493. doi: https://doi.org/10.1137/20M1336254. 1193

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. MR0125600. 1202

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

ciation, 107(497): 290–303. MR2949360. doi: https://doi.org/10.1080/01621459.2011.644501. 1215

Stein, M. L., Chi, Z., and Welty, L. J. (2004). "Approximating likelihoods for large spatial data sets." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(2): 275–296. MR2062376. doi: https://doi.org/10.1046/j.1369-7412.2003.05512.x. 1193, 1199

Stroud, J. R., Stein, M. L., and Lysen, S. (2017). "Bayesian and maximum likelihood estimation for Gaussian processes on an incomplete lattice." *Journal of Computational and Graphical Statistics*, 26(1): 108–120. MR3610412. doi: https://doi.org/10.1080/10618600.2016.1152970. 1193

Sun, Y. and Stein, M. L. (2016). "Statistically and computationally efficient estimating equations for large spatial datasets." *Journal of Computational and Graphical Statistics*, 25(1): 187–208. MR3474043. doi: https://doi.org/10.1080/10618600.2014.975230. 1193

Sun, Y., Stein, M. L., et al. (2015). "A stochastic space-time model for intermittent precipitation occurrences." *The Annals of Applied Statistics*, 9(4): 2110–2132. MR3456368. doi: https://doi.org/10.1214/15-AOAS875. 1191

Tagle, F., Castruccio, S., and Genton, M. G. (2020). "A hierarchical bi-resolution spatial skew-t model." *Spatial Statistics*, 35: 100398. MR4052623. doi: https://doi.org/10.1016/j.spasta.2019.100398. 1192

Vecchia, A. V. (1988). "Estimation and model identification for continuous spatial processes." *Journal of the Royal Statistical Society: Series B (Methodological)*, 50(2): 297–312. MR0964183. 1193

Wallin, J. and Bolin, D. (2015). "Geostatistical modelling using non-Gaussian Matérn fields." *Scandinavian Journal of Statistics*, 42(3): 872–890. MR3391697. doi: https://doi.org/10.1111/sjos.12141. 1192, 1193

Xu, G. and Genton, M. G. (2017). "Tukey g-and-h random fields." *Journal of the American Statistical Association*, 112(519): 1236–1249. MR3735373. doi: https://doi.org/10.1080/01621459.2016.1205501. 1192

Zareifard, H., Khaledi, M. J., Rivaz, F., Vahidi-Asl, M. Q., et al. (2018). "Modeling skewed spatial data using a convolution of Gaussian and log-Gaussian processes." *Bayesian Analysis*, 13(2): 531–557. MR3780434. doi: https://doi.org/10.1214/17-BA1064. 1192

Zhang, H. and El-Shaarawi, A. (2010). "On spatial skew-Gaussian processes and applications." *Environmetrics*, 21(1): 33–47. MR2842222. doi: https://doi.org/10.1002/env.982. 1192

Zheng, X., Kottas, A., and Sansó, B. (2023). "Bayesian geostatistical modeling for discrete-valued processes." *Environmetrics*, 34(7), e2805. MR4659925. doi: https://doi.org/10.1002/env.2805. 1215

Zheng, X., Kottas, A., and Sansó, B. (2022). "On construction and estimation of station-

ary mixture transition distribution models." *Journal of Computational and Graphical Statistics*, 31(1): 283–293. MR4387227. doi: https://doi.org/10.1080/10618600.2021.1981342. 1197

Zilber, D. and Katzfuss, M. (2021). "Vecchia–Laplace approximations of generalized Gaussian processes for big non-Gaussian spatial data." *Computational Statistics & Data Analysis*, 153: 107081. MR4146817. doi: https://doi.org/10.1016/j.csda.2020.107081. 1193

Acknowledgments

This work is part of the Ph.D. dissertation of X. Zheng, completed at University of California, Santa Cruz. The authors wish to thank the reviewers, an Associate Editor, and an Editor for valuable comments.