Nearest-Neighbor Geostatistical Models for Non-Gaussian Data

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

July 19, 2021

Abstract

We develop a class of nearest neighbor mixture transition distribution process (NNMP) models that provides flexibility and scalability for non-Gaussian geostatistical data. We use a directed acyclic graph to define a proper spatial process with finite-dimensional distributions given by finite mixtures. We develop conditions to construct general NNMP models with pre-specified stationary marginal distributions. We also establish lower bounds for the strength of the tail dependence implied by NNMP models, demonstrating the flexibility of the proposed methodology for modeling multivariate dependence through bivariate distribution specification. To implement inference and prediction, we formulate a Bayesian hierarchical model for the data, using the NNMP prior model for the spatial random effects process. From an inferential point of view, the NNMP model lays out a new computational approach to handling large spatial data sets, leveraging the mixture model structure to avoid computational issues that arise from large matrix operations. We illustrate the benefits of the NNMP modeling framework using synthetic data examples and through analysis of sea surface temperature data from the Mediterranean sea.

Keywords: Bayesian hierarchical models; Copulas; Mixture transition distribution; Stationary marginal distributions; Tail dependence

1 Introduction

Gaussian processes have been widely used as an underlying structure in the 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 unrealistic in many fields where the data exhibits non-Gaussian features such as heavy tails or skewness. Moreover, it is not clear that approaches for scalable models, such as low-rank models and sparsity-inducing models, are extendable when it comes to non-Gaussian data. Accordingly, this article aims at developing a class of geostatistical models that is scalable and customizable to general non-Gaussian nature, with particular focus on continuous data.

Gaussian process-based approaches for modeling continuous, non-Gaussian geostatistical datasets proceed by either representing skewed or long-tailed distributions as location-scale mixtures of Gaussian distributions, or by applying a transformation to a Gaussian process. In the former approach, it is possible to mix over location parameters to capture skewness (Kim and Mallick, 2004; Zhang and El-Shaarawi, 2010; Mahmoudian, 2017) or over scale parameters to capture heavy tails (Palacios and Steel, 2006; Sun et al., 2015). Morris et al. (2017) and Bevilacqua et al. (2020) provide examples of using both types of mixing. Although these models can accommodate skewness and non-standard tail behavior, modeling through mixtures of Gaussian distributions leads to the same computational issues faced by Gaussian process models.

The approach based on transforming a Gaussian process is usually applied to positive continuous data. A popular family of the so-called trans-Gaussian processes is obtained from the Box-Cox family of non-linear transformations (De Oliveira et al., 1997; Allcroft and Glasbey, 2003). Alternatively, Xu and Genton (2017) consider a different family of transformation, the Tukey g-and-h transformation. After transformation, standard Gaussian process statistical analysis can be carried out on the transformed data. However, given a transformation, it is possible that properties of the Gaussian process are not preserved for the transformed process. For example, the square-root transformation induces a non-

stationary covariance function, even though the covariance function associated with the original Gaussian process is stationary (Wallin and Bolin, 2015).

An alternative to Gaussian process-based approaches is to apply a copula for the joint distribution of the underlying spatial process. A copula (Joe, 2014) is used to characterize the dependence structure between random variables, separately from the specification of marginal distributions, so it has been used to describe non-Gaussian spatial variability with general non-Gaussian marginals; see, e.g., Bárdossy (2006), Ghosh and Mallick (2011), Beck et al. (2020). However, copulas need to be used with careful consideration of their properties 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. Moreover, it is difficult to strike a good balance between the flexibility provided by a copula and the computational demand required to fit it; see, e.g., Gräler (2014).

Bayesian nonparametric methods have also been explored for geostatistical modeling, starting with the approach in Gelfand et al. (2005) which extends the Dirichlet process (DP) (Ferguson, 1973) to a prior model for random spatial surfaces. We refer to Müller et al. (2018) for a review. Bayesian nonparametric methods are appealing for their large prior support on the finite-dimensional distributions of the spatial process. However, such models typically require replication at the observed spatial locations for effective inference, and they are computationally prohibitive for large data sets.

The modeling framework proposed in this article is distinctly different from the aforementioned approaches, as it builds on the class of nearest neighbor processes. Nearest neighbor processes are spatial processes obtained by extending a joint density over a reference set to the entire domain, based on a sparse directed acyclic graph (DAG). The joint density is derived from a stochastic process, referred to as the parent process. A Gaussian parent process results in the nearest neighbor Gaussian process (NNGP; Datta et al. 2016a) that has received substantial attention in the recent literature; see, e.g., Datta et al. (2016b), Finley et al. (2019), and Peruzzi et al. (2020), Katzfuss and Guinness (2021). By construction, the NNGP introduces sparsity in the precision matrix, and thus delivers

computational scalability for large spatial data sets.

In this article, we propose a novel family of nearest neighbor processes using a mixture transition distribution (MTD) model (Le et al., 1996; Zheng et al., 2020) for the parent process. The parent MTD process is based on a weighted combination of first-order spatially varying conditional densities, each of which depends on a specific neighbor. Such local dependence, together with location-dependent mixture weights, provide flexible descriptions of the spatial variability. We refer to the resulting process as the nearest neighbor mixture transition distribution process (NNMP). The mixture structure of the parent MTD process gives rise to mixtures for the NNMP finite-dimensional distributions, as well as the convenience of building the multivariate dependence specification through a set of bivariate distributions that define the first-order conditional densities of the parent MTD process. Utilizing this model property, we study the tail dependence and provide results that can guide modeling choices. In addition, extending the temporal MTD framework in Zheng et al. (2020), we develop a sufficient condition to construct NNMPs with general stationary marginal distributions. In essence, the NNMP framework provides a flexible class of models for spatial data that is able to accommodate wide families of marginal distributions, complex spatial dependence, and a variety of tail behaviors, coupled with a scalable computational approach leveraged from the mixture structure of the model.

The rest of the article is organized as follows. In Section 2, we formulate the NNMP using the parent MTD process, present the constructive framework for stationary NNMPs, and study the model's tail dependence properties. Specific examples of NNMP models illustrate different components of the methodology. Section 3 introduces the use of the NNMP as a prior for spatial processes in a Bayesian hierarchical regression setting, and discusses the approach to inference and prediction. In Sections 4 and 5, we illustrate different classes of NNMP-based spatial models with synthetic data examples and with analysis of Mediterranean sea surface temperature data, respectively. Finally, Section 6 concludes with a summary and discussion of future work.

2 Nearest neighbor MTD processes for spatial data

2.1 Modeling framework

Consider a univariate spatial process $\{Z(\boldsymbol{v}) : \boldsymbol{v} \in \mathcal{D}\}$, where $\mathcal{D} \subset \mathbb{R}^p$, for $p \geq 1$. Let $\mathcal{S} = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n\}$ be a finite collection of locations in $\mathcal{D}$, referred to as the reference set. We write the joint density of the random vector $\boldsymbol{z}_{\mathcal{S}} = (Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_n))^\top$ as

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

If we regard the conditioning set of $z(\boldsymbol{s}_i)$ as the set of parents of $z(\boldsymbol{s}_i)$, the joint density $p(\boldsymbol{z}_{\boldsymbol{S}})$ in (1) is a factorization according to a DAG whose vertices are $z(\boldsymbol{s}_i)$. We obtain a sparse DAG by reducing the conditioning set of $z(\boldsymbol{s}_i)$ to a smaller subset, denoted as $\boldsymbol{z}_{\mathrm{Ne}(\boldsymbol{s}_i)}$, with $\mathrm{Ne}(\boldsymbol{s}_i) \subset \mathcal{S}_i = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_{i-1}\}$. We refer to $\mathrm{Ne}(\boldsymbol{s}_i)$ as the neighbor set for $\boldsymbol{s}_i$, having at most L elements with $L \ll n$. The resulting density for the sparse DAG is

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

which has been verified as a proper density (Lauritzen, 1996).

Choosing the neighbor sets $\mathrm{Ne}(\boldsymbol{s}_i)$ creates different sparse DAGs. There are different ways to select members from $\mathcal{S}_i$ for $\mathrm{Ne}(\boldsymbol{s}_i)$; see, for example, Vecchia (1988), Stein et al. (2004), and Gramacy and Apley (2015). Our selection is based on the geostatistical distance between $\boldsymbol{s}_i$ and $\boldsymbol{s}_j \in \mathcal{S}_i$. The selected locations $\boldsymbol{s}_j$ are placed in ascending order according to the distance, denoted as $\boldsymbol{s}_{(i1)}, \ldots, \boldsymbol{s}_{(i,i_L)}$, where $i_L := (i-1) \wedge L$. We note that the development of the proposed framework holds true for any choice of the neighbor sets.

The crucial step of constructing the process model is the specification of a stochastic process over the reference set $\mathcal{S}$. This process characterizes $\tilde{p}(\boldsymbol{z}_{\mathcal{S}})$, so it is appealing to consider processes that have a Markov property, as it naturally connects to the conditional densities in (2). To this end, we consider a MTD process such that the conditional density

for location $\boldsymbol{s}_i \in \mathcal{S}$ in (2) is

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

where $f_{\boldsymbol{s}_i,l}$ is the lth component conditional density of the mixture density p for $\boldsymbol{s}_i \in \mathcal{S}$, and the weights are subject to $\sum_{l=1}^{i_L} w_l(\boldsymbol{s}_i) = 1,\ w_l(\boldsymbol{s}_i) \geq 0$ for every $\boldsymbol{s}_i \in \mathcal{S}$ and for all l.

Spatial dependence characterized by (3) is twofold. First, each component $f_{\boldsymbol{s}_i,l}$ of the density $p(z(\boldsymbol{s}_i) \mid \boldsymbol{z}_{\mathrm{Ne}(\boldsymbol{s}_i)})$ is associated with spatially varying parameters indexed at $\boldsymbol{s}_i \in \mathcal{S}$, defined by a probability model or a link function. Secondly, the weights $w_l(\boldsymbol{s}_i)$ are spatially varying. As each component density $f_{\boldsymbol{s}_i,l}$ depends on a specific neighbor, the weights indicate the contribution of each neighbor of $\boldsymbol{s}_i$. Besides, the weights adapt to the change of locations. For two different $\boldsymbol{s}_i, \boldsymbol{s}_j$ in $\mathcal{S}$, the relative locations of the nearest neighbors $\mathrm{Ne}(\boldsymbol{s}_i)$ to $\boldsymbol{s}_i$ are different from that of $\mathrm{Ne}(\boldsymbol{s}_j)$ to $\boldsymbol{s}_j$. If all elements of $\mathrm{Ne}(\boldsymbol{s}_i)$ are very close to $\boldsymbol{s}_i$, then values of $(w_1(\boldsymbol{s}_i), \ldots, w_{i_L}(\boldsymbol{s}_i))^\top$ should be quite even. On the other hand, if, for $\boldsymbol{s}_j$, only a subset of its neighbors in $\mathrm{Ne}(\boldsymbol{s}_j)$ are close to $\boldsymbol{s}_j$, then the weights corresponding to this subset should receive larger values. We remark that 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 component densities and the weights are deferred to later sections.

To obtain a properly defined spatial process in $\mathcal{D}$, we extend (3) 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

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

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 (3), except that $\mathrm{Ne}(\boldsymbol{u}_i) = \{\boldsymbol{u}_{(i1)}, \ldots, \boldsymbol{u}_{(iL)}\}$ are the first L locations in $\mathcal{S}$ that are closest to $\boldsymbol{u}_i$ in terms of geostatistical distance. Building the construction of the neighbor sets $\mathrm{Ne}(\boldsymbol{u}_i)$ on the reference set ensures that $\tilde{p}(\boldsymbol{z}_\mathcal{U} \,|\, \boldsymbol{z}_\mathcal{S})$ is a proper density.

Given (3) and (4), we can obtain the joint density $\tilde{p}(\boldsymbol{z}_\mathcal{V})$ 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 $\tilde{p}(\boldsymbol{z}_{\mathcal{V}})$ is directly available as the appropriate marginal of $\tilde{p}(\boldsymbol{z}_{\mathcal{S}})$. Otherwise, we have that

$$\tilde{p}(\boldsymbol{z}_{\mathcal{V}}) = \int \tilde{p}(\boldsymbol{z}_{\mathcal{U}} \,|\, \boldsymbol{z}_{\mathcal{S}}) \tilde{p}(\boldsymbol{z}_{\mathcal{S}}) \prod_{\{\boldsymbol{s}_i \in \mathcal{S} \setminus \mathcal{V}\}} dz(\boldsymbol{s}_i), \tag{5}$$

where $\mathcal{U} = \mathcal{V} \setminus \mathcal{S}$. If $\mathcal{S} \setminus \mathcal{V}$ is empty, $\tilde{p}(\boldsymbol{z}_{\mathcal{V}})$ is simply $\tilde{p}(\boldsymbol{z}_{\mathcal{U}} \,|\, \boldsymbol{z}_{\mathcal{S}}) \tilde{p}(\boldsymbol{z}_{\mathcal{S}})$.

The resulting spatial process defined using (3) and (4) is a legitimate process over the entire domain $\mathcal{D}$. This result is a direct application of the nearest neighbor process (Datta et al., 2016a), which is constructed from specifying a parent process that defines $\tilde{p}(\boldsymbol{z}_{\mathcal{S}})$ in (2) over $\mathcal{S}$ based on a DAG, and then extending the parent process to arbitrary finite set $\mathcal{U} \subset \mathcal{D} \backslash \mathcal{S}$. We call the proposed model the nearest neighbor MTD process (NNMP), derived from a parent MTD process. In the subsequent development of the model properties, we will use the associated conditional density

$$p(z(\boldsymbol{v}) \mid \boldsymbol{z}_{\mathrm{Ne}(\boldsymbol{v})}) = \sum_{l=1}^{L} w_l(\boldsymbol{v})\, f_{\boldsymbol{v},l}(z(\boldsymbol{v}) \mid z(\boldsymbol{v}_{(l)})), \quad \boldsymbol{v} \in \mathcal{D}, \tag{6}$$

to characterize an NNMP, where $\mathrm{Ne}(\boldsymbol{v})$ contains the first L locations that are closest to $\boldsymbol{v}$, selected from locations in $\mathcal{S}$. These locations in $\mathrm{Ne}(\boldsymbol{v})$ are placed in ascending order according to distance, denoted as $\boldsymbol{v}_{(1)}, \ldots, \boldsymbol{v}_{(L)}$.

In general, the joint density $\tilde{p}(\boldsymbol{z}_{\mathcal{V}})$ of an NNMP is intractable. However, from (5) where both $\tilde{p}(\boldsymbol{z}_{\mathcal{U}} \,|\, \boldsymbol{z}_{\mathcal{S}})$ and $\tilde{p}(\boldsymbol{z}_{\mathcal{S}})$ are products of mixtures, we can recognize that $\tilde{p}(\boldsymbol{z}_{\mathcal{V}})$ 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 2.3 that for some NNMP models, the joint density $\tilde{p}(\boldsymbol{z}_{\mathcal{V}})$ has a closed-form expression.

Before closing this section, we would like to point out that spatial locations are not naturally ordered. Given a distance function, a different topological ordering on the locations $\boldsymbol{s}_i$ results in different neighbor sets $\mathrm{Ne}(\boldsymbol{s}_i)$. Therefore, a different sparse DAG with density $\tilde{p}(\boldsymbol{z}_{\mathcal{S}})$ is created accordingly for model inference. For the NNMP models illustrated in the data examples, we found by simulation experiments that there were no discernible differences between the inferences based on $\tilde{p}(\boldsymbol{z}_{\mathcal{S}})$, given two different orderings. This ob-

servation is coherent with that from the NNGP models and other literature that considers nearest-neighbor likelihood approximations. Since the approximation of $\tilde{p}(\boldsymbol{z}_{\mathcal{S}})$ to $p(\boldsymbol{z}_{\mathcal{S}})$ depends on the information borrowed from the neighbors, as outlined in Datta et al. (2016a), the effectiveness is determined by the size of $\text{Ne}(\boldsymbol{s}_i)$ rather than the ordering.

2.2 NNMPs with stationary marginal distributions

In this section, we develop a sufficient condition to construct NNMPs with general stationary marginal distributions. The result is given in the following proposition, the proof of which can be found in the Supplementary Material. The key feature of this result is that the condition relies on the distributional assumption for the bivariate distributions that define the MTD component conditional densities in (3) and (4), without the need to impose restrictions on the parameter space.

Proposition 1. *Consider an NNMP characterized by* (6) *for which the component density* $f_{\boldsymbol{v},l}$ *corresponds to a bivariate distribution of a random vector* $(U_{\boldsymbol{v},l}, V_{\boldsymbol{v},l})$ *with marginal densities* $f_{U_{\boldsymbol{v},l}}$ *and* $f_{V_{\boldsymbol{v},l}}$*, for* $l = 1, \ldots, L$*. The NNMP has a stationary marginal density* f_Z *if it satisfies the invariant condition:* $Z(\boldsymbol{s}_1) \sim f_Z$*,* $\boldsymbol{s}_1 \in \mathcal{S}$*, and for every* $\boldsymbol{v} \in \mathcal{D}$*,* $f_Z(z) = f_{U_{\boldsymbol{v},l}}(z) = f_{V_{\boldsymbol{v},l}}(z)$ *for all* z *and for all* l*.*

This result builds from the one in Zheng et al. (2020) where temporal MTD processes 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 marginal distributions and a general functional form, either linear or non-linear, for the expectation with respect to the conditional density p in (3) and (4).

As previously discussed, the mixture formulation of the parent MTD process induces a finite mixture for the finite-dimensional distribution of the NNMP. On the other hand, due to the mixture form, an explicit expression for the covariance is difficult to derive. A recursive equation can be obtained for a class of NNMP models for which the conditional expectation with respect to $(U_{\boldsymbol{v},l}, V_{\boldsymbol{v},l})$ is linear, i.e., $E(U_{\boldsymbol{v},l} \,|\, V_{\boldsymbol{v},l} = z) = a_l(\boldsymbol{v}) + b_l(\boldsymbol{v})\, z$ for some $a_l(\boldsymbol{v}), b_l(\boldsymbol{v}) \in \mathbb{R},\ l = 1, \ldots, L$, and for all $\boldsymbol{v} \in \mathcal{D}$. Suppose the NNMP has a stationary

marginal distribution with finite first and second moments. Without loss of generality, we assume the first moment is zero. Then the covariance over any two locations $\boldsymbol{v}_1, \boldsymbol{v}_2 \in \mathcal{D}$ is

$$\begin{aligned}
&\mathrm{Cov}(Z(\boldsymbol{v}_1), Z(\boldsymbol{v}_2)) \\
&\quad = \begin{cases}
\sum_{l=1}^{L} w_l(\boldsymbol{s}_i)\, b_l(\boldsymbol{s}_i)\, E(Z(\boldsymbol{s}_j) Z(\boldsymbol{s}_{(il)})), & \boldsymbol{v}_1 \equiv \boldsymbol{s}_i \in \mathcal{S}, \boldsymbol{v}_2 \equiv \boldsymbol{s}_j \in \mathcal{S}, \\
\sum_{l=1}^{L} w_l(\boldsymbol{v}_1)\, b_l(\boldsymbol{v}_1)\, E(Z(\boldsymbol{s}_j) Z(\boldsymbol{v}_{(1l)})), & \boldsymbol{v}_1 \notin \mathcal{S}, \boldsymbol{v}_2 \equiv \boldsymbol{s}_j \in \mathcal{S}, \\
\sum_{l=1}^{L} \sum_{l'=1}^{L} w_{ll'} \{a_{ll'} + b_{ll'} E(Z(\boldsymbol{v}_{(1l)}) Z(\boldsymbol{v}_{(2l')}))\}, & \boldsymbol{v}_1, \boldsymbol{v}_2 \notin \mathcal{S},
\end{cases}
\end{aligned} \tag{7}$$

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 (7) implies that, even though the process has a stationary marginal distribution, the NNMP is second-order non-stationary.

2.3 Construction of NNMP models

To balance model flexibility and scalability, we build the sequence of bivariate distributions indexed at $\boldsymbol{v}$, namely, $(U_{\boldsymbol{v},l}, V_{\boldsymbol{v},l})$, from a random vector (U_l, V_l), by modifying some of its parameters to be spatially varying, i.e., indexed at $\boldsymbol{v}$, through a probability model or a link function, for $l = 1, \ldots, L$. We refer to the random vectors (U_l, V_l) as the set of base random vectors. With a careful choice of the probability model or link function for the spatially varying parameter, this construction method reduces significantly the dimension of the parameter space, while preserving the capability of the model to capture spatial dependence. This is best illustrated through an example.

Example 1. Gaussian and continuous mixture of Gaussian NNMP models.

Take the set of (U_l, V_l) to be bivariate Gaussian random vectors, with mean $(\mu_l, \mu_l)^\top$ and covariance matrix $\Sigma_l = \sigma_l^2 \left(\begin{smallmatrix} 1 & \rho_l \\ \rho_l & 1 \end{smallmatrix}\right)$, resulting in a Gaussian conditional density $f_{U_l|V_l}(u_l \,|\, v_l) = N(u_l \,|\, (1-\rho_l)\mu_l + \rho_l v_l, \sigma_l^2(1-\rho_l^2))$, for $l = 1, \ldots, L$. Using a correlation function k_l for ρ_l such that $\rho_l(\boldsymbol{v}) = k_l(\boldsymbol{v}, \boldsymbol{v}_{(l)})$, we obtain the spatially varying conditional density,

$$p(z(\boldsymbol{v}) \mid \boldsymbol{z}_{\mathrm{Ne}(\boldsymbol{v})}) = \sum_{l=1}^{L} w_l(\boldsymbol{v})\, N(z(\boldsymbol{v}) \mid (1-\rho_l(\boldsymbol{v}))\mu_l + \rho_l(\boldsymbol{v}) z(\boldsymbol{v}_{(l)}), \sigma_l^2(1-\rho_l(\boldsymbol{v})^2)). \tag{8}$$

This NNMP is referred to as the Gaussian NNMP (GNNMP). If we take $Z(\boldsymbol{s}_1) \sim f_Z = N(z \,|\, \mu, \sigma^2)$, and constrain the parameters to be $\mu = \mu_l$ and $\sigma^2 = \sigma_l^2$ for all l, it can be easily verified that the resulting GNNMP satisfies the invariant condition of Proposition 1, with a stationary marginal $f_Z(z)$. Moreover, when $L = 1$, the model is a Gaussian process. The finite-dimensional distribution of the GNNMP model is characterized by the following proposition, the proof of which is included in the Supplementary Material.

Proposition 2. *Consider the GNNMP characterized by* (8) *with* $\mu = \mu_l$ *and* $\sigma^2 = \sigma_l^2$ *for all* l*. If* $Z(\boldsymbol{s}_1) \sim f_Z = N(z \,|\, \mu, \sigma^2)$*, the GNNMP has stationary marginal* f_Z *and its finite-dimensional distribution is a mixture 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 mixture of Gaussian distributions. For example, replacing μ_l of the bivariate Gaussian distribution by $\lambda_l z_0$ where z_0 follows a Gaussian distribution truncated at $[0, \infty)$, yields a bivariate skew-Gaussian distribution (Azzalini, 2013). We can construct a skew-GNNMP model using the associated conditional density of the bivariate skew-Gaussian distribution. A stationary skew-GNNMP model is demonstrated in the second experiment of the simulation study. Other families that admit a location-scale mixture representation include, for example, Student-t, skew-t, Laplace and asymmetric Laplace. Using the associated conditional density we can construct the corresponding NNMP model. The following proposition characterizes the finite-dimensional distribution of these NNMPs when they satisfy the invariant condition of Proposition 1.

Proposition 3. *Consider a class of NNMPs with stationary marginal distribution such that the family of distributions for the base random vector* (U_l, V_l) *is a location-scale mixture of Gaussian distributions for all* l*. Then, the NNMP finite-dimensional distribution is a finite mixture with mixture components that belong to the same family of the base random vectors.*

Proposition 3 is easily verified. Suppose $\boldsymbol{z}_{\mathcal{V}}$ is a p-dimensional realization from a GNNMP with stationary marginal $f_Z(z) = N(z \,|\, \mu, \sigma^2)$. Then, by Proposition 2, we can write $\tilde{p}(\boldsymbol{z}_{\mathcal{V}}) = \sum_{m=1}^{M} w_m N(\boldsymbol{z}_{\mathcal{V}} \,|\, \mu \mathbf{1}_p, \sigma^2 R_m)$, where $\mathbf{1}_p$ is a column vector of ones and R_m is the correlation matrix of the mth component. Reparameterizing μ and σ^2 with different

distributional assumptions on the transformed parameters, we can obtain different families of distributions for the mixture component of the joint distribution.

2.4 Models based on copulas

As implied by Example 1, a practical choice to construct an NNMP model with a stationary marginal distribution is to consider the same family of bivariate distributions for all base random vectors (U_l, V_l). In order to satisfy the invariant condition of Proposition 1, we require that the corresponding conditional density preserves some spatially varying parameter that is not shared with the stationary marginal distribution. In this regard, a useful strategy is to use a copula for the bivariate distributions of (U_l, V_l).

A copula function $C : [0,1]^p \to [0,1]$ is a function such that, due to Sklar's Theorem (Sklar, 1959), for any multivariate distribution $F(z_1, \ldots, z_p)$, 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, $j = 1, \ldots, p$. If F_j is continuous for all j, C is unique. A copula enables us to separate the modeling of the marginal distributions from the bivariate dependence. Thus, the invariant condition can be easily satisfied 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 can be made spatially varying. We focus on continuous stationary distributions, although this strategy can be applied for any family of distributions for F_Z.

To construct a copula NNMP model, we consider a base copula C_l for (U_l, V_l), for $l = 1, \ldots, L$. We obtain the spatially varying copula $C_{\boldsymbol{v},l}$ for $(U_{\boldsymbol{v},l}, V_{\boldsymbol{v},l})$ by working with spatially varying copula parameter. 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 $f_{U_{\boldsymbol{v},l}}$ and $f_{V_{\boldsymbol{v},l}}$ are the marginal densities of $U_{\boldsymbol{v},l}$ and $V_{\boldsymbol{v},l}$, respectively. Given a pre-specified stationary marginal f_Z from essentially any continuous family of distributions, we replace both $f_{U_{\boldsymbol{v},l}}$ and $f_{V_{\boldsymbol{v},l}}$ with f_Z. We then obtain the conditional density

$$p(z(\boldsymbol{v}) \mid \boldsymbol{z}_{\mathrm{Ne}(\boldsymbol{v})}) = \sum_{l=1}^{L} w_l(\boldsymbol{v})\, c_{\boldsymbol{v},l}(z(\boldsymbol{v}), z(\boldsymbol{v}_{(l)}))f_Z(z(\boldsymbol{v})) \tag{9}$$

that characterizes the stationary copula NNMP. Some of the NNMPs of Section 2.3 can be regarded as special cases of copula NNMPs; e.g., the bivariate Gaussian distribution of (U_l, V_l) in the GNNMP corresponds to a Gaussian copula with Gaussian marginals.

Under the copula framework, one strategy to specify the spatially varying parameters 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, the associated Kendall's τ is given by $\rho_{\tau,l} = 4\int_0^1 \int_0^1 C_l(q_{u_l} q_{v_l}) dq_{u_l} q_{v_l} - 1$, where $q_{u_l} = F_{U_l}(u_l)$ and $q_{v_l} = F_{V_l}(v_l)$. 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, sec. 2.12). We illustrate the strategy with the following example.

Example 2. The bivariate Gumbel copula is an asymmetric copula useful for modeling dependence when the marginal distributions are positive and heavy-tailed. The spatially varying Gumbel copula can be defined as $C_{\boldsymbol{v},l} = \exp(-[(-\log F_{U_{\boldsymbol{v},l}}(z(\boldsymbol{v}))^{\eta_l(\boldsymbol{v})} + (-\log F_{V_{\boldsymbol{v},l}}(z(\boldsymbol{v}_{(l)}))^{\eta_l(\boldsymbol{v})}]^{1/\eta_l(\boldsymbol{v})})$, 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 value 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 $\eta_l(\boldsymbol{v}) \equiv \eta(||\boldsymbol{v} - \boldsymbol{v}_{(l)}||)$ is given by $h_l(||\boldsymbol{v} - \boldsymbol{v}_{(l)}||)$, and $\eta_l(\boldsymbol{v}) \to \infty$ as $||\boldsymbol{v} - \boldsymbol{v}_{(l)}|| \to 0$.

In addition to a convenient strategy to achieve stationarity, copula NNMP models offer avenues to capture complex dependence using general bivariate copulas. In spatial copula modeling, when multivariate copulas are applied to the joint density of a spatial realization, many of them are inappropriate, for example, the Gumbel copula in Example 2 that has exchangeable dependence. Though spatial vine copula models (Gräler, 2014) were proposed to resolve this restriction, their model structure and computation are substantially more complicated than copula NNMP models.

2.5 Mixture component specification and tail dependence

A benefit of building the NNMPs from a set of base random vectors is that specification of the multivariate dependence of $Z(\boldsymbol{v})$ given its neighbors is determined mainly by that of the base random vectors. In this section, we illustrate this attractive property of the model with the establishment of lower bounds for two measures used to assess the strength of the tail dependence of NNMPs.

To establish our results we rely on the assumption 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 $P(U_l > u_l \,|\, V_l = v_l)$ increases as v_l increases. The definition is 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 $P(Z_1 > z_1 \,|\, Z_2 = z_2, \ldots, Z_p = z_p) \leq P(Z_1 > z_1 \,|\, Z_2 = z_2', \ldots, Z_p = z_p')$, for all $(z_2, \ldots, z_p)$ and $(z_2', \ldots, z_p')$ in the support of $(Z_2, \ldots, Z_p)$, where $z_j \leq z_j'$ for $j = 2, \ldots, p$. The conditional density in (6) implies that

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

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 (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 $\lambda_{\mathcal{H},l} = \lim_{q \to 1^-} P(U_l > F_{U_l}^{-1}(q) \mid V_l > F_{V_l}^{-1}(q))$ and $\lambda_{\mathcal{L},l} = \lim_{q \to 0^+} P(U_l \leq F_{U_l}^{-1}(q) \mid V_l \leq F_{V_l}^{-1}(q))$. 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 independent in the lower tail are similarly defined using $\lambda_{\mathcal{L},l}$. Analogously, we can define the upper and lower tail dependence

coefficients for $Z(\boldsymbol{v})$ given its nearest neighbors, namely,

$$\lambda_{\mathcal{H}}(\boldsymbol{v}) = \lim_{q \to 1^-} P(Z(\boldsymbol{v}) > F^{-1}_{Z(\boldsymbol{v})}(q) \mid Z(\boldsymbol{v}_{(1)}) > F^{-1}_{Z(\boldsymbol{v}_{(1)})}(q), \ldots, Z(\boldsymbol{v}_{(L)}) > F^{-1}_{Z(\boldsymbol{v}_{(L)})}(q)),$$

$$\lambda_{\mathcal{L}}(\boldsymbol{v}) = \lim_{q \to 0^+} P(Z(\boldsymbol{v}) \leq F^{-1}_{Z(\boldsymbol{v})}(q) \mid Z(\boldsymbol{v}_{(1)}) \leq F^{-1}_{Z(\boldsymbol{v}_{(1)})}(q), \ldots, Z(\boldsymbol{v}_{(L)}) \leq F^{-1}_{Z(\boldsymbol{v}_{(L)})}(q)).$$

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

Proposition 4. *Consider the NNMP characterized by* (6) *constructed from a set of base random vectors* (U_l, V_l) *supported on* $\Omega \times \Omega$, $\Omega \subset \mathbb{R}$, *with bivariate distribution function* F_{U_l,V_l} *and marginal distribution functions* F_{U_l}, F_{V_l}, *for* $l = 1, \ldots, L$. *Moreover, assume the random variable* U_l *is stochastically increasing in* V_l *for all* l. *Then, for every* $\boldsymbol{v} \in \mathcal{D}$, *the lower bound for the upper tail dependence coefficient* $\lambda_{\mathcal{H}}(\boldsymbol{v})$ *is* $\sum_{l=1}^{L} w_l(\boldsymbol{v}) \lim_{q \to 1^-} P(Z(\boldsymbol{v}) > F^{-1}_{U_{\boldsymbol{v},l}}(q) \,|\, Z(\boldsymbol{v}_{(l)}) = F^{-1}_{V_{\boldsymbol{v},l}}(q))$, *and the lower bound for the lower tail dependence coefficient* $\lambda_{\mathcal{L}}(\boldsymbol{v})$ *is* $\sum_{l=1}^{L} w_l(\boldsymbol{v}) \lim_{q \to 0^+} P(Z(\boldsymbol{v}) \leq F^{-1}_{U_{\boldsymbol{v},l}}(q) \,|\, Z(\boldsymbol{v}_{(l)}) = F^{-1}_{V_{\boldsymbol{v},l}}(q))$.

The proof of the proposition is provided in the Supplementary Material. In a nutshell, Proposition 4 shows that the lower and upper tail dependence coefficients are bounded below by a convex combination of, respectively, the limits of the conditional cumulative distribution functions, and the limits of the conditional survival functions. These are fully determined by the dependence structure of the bivariate distribution for (U_l, V_l). This result is best illustrated with an example.

Example 3. Consider a Lomax NNMP for which the bivariate distributions of the base random vectors correspond to a bivariate Lomax distribution (Arnold et al., 1999), resulting in an associated conditional density of the model as follows

$$p(z(\boldsymbol{v}) \mid \boldsymbol{z}_{\mathrm{Ne}(\boldsymbol{v})}) = \sum_{l=1}^{L} w_l(\boldsymbol{v})\, P(z(\boldsymbol{v}) \mid z(\boldsymbol{v}_{(l)}) + \phi_l, \alpha_l(\boldsymbol{v})),$$

where $P(x \,|\, \phi, \alpha) = \alpha\phi^{-1}(1 + x\phi^{-1})^{-(\alpha+1)}$ denotes the Lomax density, a shifted version of the Pareto Type I density. A small value of α indicates a heavy tail. The component conditional survival function of the Lomax NNMP, expressed in terms of the quantile q, is $\left\{1 + F^{-1}_{U_{\boldsymbol{v},l}}(q)/(F^{-1}_{V_{\boldsymbol{v},l}}(q) + \phi_l)\right\}^{-\alpha_l(\boldsymbol{v})}$ 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.

Proposition 2 holds for the general NNMP 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, the proof of which is given in the Supplementary Material.

Corollary 1. *Suppose that the base random vectors (U_l, V_l) in Proposition 4 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 $(U_{\boldsymbol{v},l}, V_{\boldsymbol{v},l})$ induced from (U_l, V_l).*

Under Corollary 1, if the bivariate distribution of (U_l, V_l) is symmetric, e.g., an elliptically symmetric distribution, the upper and lower tail dependence coefficients coincide, and can simply be denoted as $\lambda(\boldsymbol{v})$. Then we have 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 c.d.f., 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 the conditional c.d.f. $F_{U_l|V_l}(F_{U_l}^{-1}(q) \,|\, F_{V_l}^{-1}(1))$ is positive at $q = 1$. Likewise, a non-zero $F_{U_l|V_l}(F_{U_l}^{-1}(q) \,|\, F_{V_l}^{-1}(0))$ at $q = 0$ indicates some strength of dependence in the lower tails. The functions $F_{U_l|V_l}(\cdot \mid F_{V_l}^{-1}(0))$ and $F_{U_l|V_l}(\cdot \mid F_{V_l}^{-1}(1))$ are referred to as the boundary conditional c.d.f.s.

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

tail dependence is defined likewise using $F_{1|2}(F^{-1}_{Z(\boldsymbol{v})}(q) \,|\, F^{-1}_{Z_{\mathrm{Ne}(\boldsymbol{v})}}(0))$. The proposition below provides lower bounds for the boundary conditional c.d.f.s. The proof can be found in the Supplementary Material.

Proposition 5. *Consider the NNMP characterized by* (6) *constructed from a set of base random vectors* (U_l, V_l) *supported on* $\Omega \times \Omega$, $\Omega \subset \mathbb{R}$, *with bivariate distribution function* F_{U_l,V_l} *and marginal distribution functions* F_{U_l}, F_{V_l}, *for* $l = 1, \ldots, L$. *Moreover, assume the random variable* U_l *is stochastically increasing in* V_l *for all* l. *Let* $\lambda_{\mathcal{H},l}(\boldsymbol{v})$ *and* $\lambda_{\mathcal{L},l}(\boldsymbol{v})$ *be the upper and lower component tail dependence coefficients, respectively, for* $l = 1, \ldots, L$. *If given* $\boldsymbol{v}$, *there exists* $\lambda_{\mathcal{L},l}(\boldsymbol{v}) > 0$ *for some* l, *then the conditional c.d.f.* $F_{1|2}(F^{-1}_{Z(\boldsymbol{v})}(q) \,|\, F^{-1}_{Z_{Ne(\boldsymbol{v})}}(0))$ *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 given* $\boldsymbol{v}$, *there exists* $\lambda_{\mathcal{H},l}(\boldsymbol{v}) > 0$ *for some* l, *then the conditional c.d.f.* $F_{1|2}(F^{-1}_{Z(\boldsymbol{v})}(q) \,|\, F^{-1}_{Z_{Ne(\boldsymbol{v})}}(1))$ *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 5 complements Proposition 4 to assess the strength of the tail dependence. It readily applies for bivariate distributions, especially for copulas which yield explicit expressions for the tail dependence coefficients. For example, the spatially varying Gumbel copula $C_{\boldsymbol{v},l}$ in Example 2 has upper tail dependence coefficient $2 - 2^{1/\eta_l(\boldsymbol{v})} > 0$ for $\eta_l(\boldsymbol{v}) > 1$, so 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, $F_{1|2}(F^{-1}_{Z(\boldsymbol{v})}(q) \,|\, F^{-1}_{Z_{\mathrm{Ne}(\boldsymbol{v})}}(1))$ degenerates at $q = 1$, implying strongest tail dependence.

3 Bayesian hierarchical model and inference

3.1 Hierarchical model formulation

We consider the following univariate spatially varying regression model,

$$y(\boldsymbol{v}) = \boldsymbol{x}(\boldsymbol{v})^\top \boldsymbol{\beta} + z(\boldsymbol{v}) + \epsilon(\boldsymbol{v}), \quad \boldsymbol{v} \in \mathcal{D}, \tag{10}$$

where $y(\boldsymbol{v}) \in \mathbb{R}$ is a continuous, geostatistical outcome, $\boldsymbol{x}(\boldsymbol{v})$ is a $p \times 1$ vector of spatially referenced predictors, $z(\boldsymbol{v})$ is a spatial process, and $\epsilon(\boldsymbol{v}) \overset{i.i.d.}{\sim} N(0, \tau^2)$ represents the

measurement error. To account for non-Gaussian spatial variability, we assume an NNMP model for $z(\boldsymbol{v})$, with the components $f_{\boldsymbol{v},l}$ built from real-valued bivariate distributions.

A key component of the proposed model formulation is the prior model for the weights. The model allows the weights to vary in space and adjust the differences among the neighbor structures of different reference locations. In particular, we consider a collection of spatially dependent c.d.f.s $\{G_{\boldsymbol{v}} : \boldsymbol{v} \in \mathcal{D}\}$ supported on $(0, 1)$. For each $\boldsymbol{v}$, the weights are defined as the increments of $G_{\boldsymbol{v}}$ with a set of random cutoff points $r_{\boldsymbol{v},0}, \ldots, r_{\boldsymbol{v},L}$. More specifically,

$$w_l(\boldsymbol{v}) = \int 1_{(r_{\boldsymbol{v},l-1}, r_{\boldsymbol{v},l})}(\cdot)\, dG_{\boldsymbol{v}}(\cdot), \quad l = 1, \ldots, L, \tag{11}$$

where 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, for $l = 1, \ldots, L$, $r_{\boldsymbol{v},l} - r_{\boldsymbol{v},l-1} = k'(\boldsymbol{v}, \boldsymbol{v}_{(l)} \,|\, \boldsymbol{\zeta}) / \sum_{l=1}^{L} k'(\boldsymbol{v}, \boldsymbol{v}_{(l)} \,|\, \boldsymbol{\zeta})$, where $k' : \mathcal{D} \times \mathcal{D} \to [0, 1]$ is a bounded kernel function with parameters $\boldsymbol{\zeta}$. With a choice of the kernel k', the random cutoff points reflect the structure in the neighbors $(\boldsymbol{v}_{(1)}, \ldots, \boldsymbol{v}_{(L)})$ of $\boldsymbol{v}$. The choices of the kernel and the associated parameter affect the smoothness of the resulting random field. The distribution function $G_{\boldsymbol{v}}$ is a logit Gaussian distribution such that the corresponding Gaussian distribution has mean $\mu(\boldsymbol{v})$ and variance κ^2. We denote the logit Gaussian distribution as $G_{\boldsymbol{v}}(\cdot \,|\, \mu(\boldsymbol{v}), \kappa^2)$. The spatial dependence across the weights $w_l(\boldsymbol{v})$ is introduced through the distribution functions $G_{\boldsymbol{v}}$ by assuming the mean $\mu(\boldsymbol{v}) = \gamma_0 + \gamma_1 v_1 + \gamma_2 v_2$, where v_1 and v_2 are the first and second coordinates of $\boldsymbol{v}$. Given the cutoff points and κ^2, a smaller value of $\mu(\boldsymbol{v})$ indicates more weights assigned to the nearer neighbors of $\boldsymbol{v}$. As a simpler version of the model in (11), if $G_{\boldsymbol{v}}$ is the uniform distribution on $(0, 1)$, the weights become $k'(\boldsymbol{v}, \boldsymbol{v}_{(l)} \,|\, \boldsymbol{\zeta}) / \sum_{l=1}^{L} k'(\boldsymbol{v}, \boldsymbol{v}_{(l)} \,|\, \boldsymbol{\zeta})$. With a set of fixed, uniform cutoff points on $[0, 1]$, i.e., $r_{\boldsymbol{v},l} - r_{\boldsymbol{v},l-1} = 1/L$, the prior model for the weights was considered in Cadonna et al. (2019) for spectral density estimation, with the collection of logit Gaussian distributions indexed at frequency.

The mixture formulation for $z(\boldsymbol{v})$ includes the choice of the component conditional density, so we employ data augmentation to facilitate posterior simulation. We take the reference set $\mathcal{S} = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n\}$ as a subset of locations where both the outcomes and predictors are observed. For each observed $y(\boldsymbol{s}_i)$, $i = 3, \ldots, n$, we introduce a latent

Gaussian variable t_i with mean $\mu(\boldsymbol{s}_i)$ and variance κ^2. Conditioning on the latent variables, the regression model on the observations $y(\boldsymbol{s}_i)$ can be written as

$$\begin{aligned} y(\boldsymbol{s}_i) \mid z(\boldsymbol{s}_i), \boldsymbol{\beta}, \tau^2 &\overset{ind.}{\sim} N(y(\boldsymbol{s}_i) \mid \boldsymbol{x}(\boldsymbol{s}_i)^\top \boldsymbol{\beta} + z(\boldsymbol{s}_i), \tau^2),\ i = 1, \ldots, n, \\ z(\boldsymbol{s}_1) \mid \boldsymbol{\theta} &\sim p_1(z(\boldsymbol{s}_1) \mid \boldsymbol{\theta}),\ z(\boldsymbol{s}_2) \mid z(\boldsymbol{s}_1), \boldsymbol{\theta} \sim f_{\boldsymbol{s}_2,1}(z(\boldsymbol{s}_2) \mid z(\boldsymbol{s}_1), \boldsymbol{\theta}), \\ z(\boldsymbol{s}_i) \mid \{z(\boldsymbol{s}_{(il)}\}_{l=1}^{i_L}, \boldsymbol{\theta}, \boldsymbol{\zeta} &\overset{ind.}{\sim} \sum_{l=1}^{i_L} f_{\boldsymbol{s}_i,l}(z(\boldsymbol{s}) \mid z(\boldsymbol{s}_{il}), \boldsymbol{\theta}) 1_{(r^*_{\boldsymbol{s}_i,l-1}, r^*_{\boldsymbol{s}_i,l})}(t_i),\ i = 3, \ldots, n, \\ t_i \mid \boldsymbol{\gamma}, \kappa^2 &\overset{ind.}{\sim} N(t_i \mid \gamma_0 + \gamma_1 \boldsymbol{s}_{i1} + \gamma_2 \boldsymbol{s}_{i2}, \kappa^2),\ i = 3, \ldots, n, \end{aligned} \tag{12}$$

where $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \gamma_2)^\top$ and $r^*_{\boldsymbol{s}_i,l} = \log(r_{\boldsymbol{s}_i,l}/(1 - r_{\boldsymbol{s}_i,l})), l = 1, \ldots, i_L, i = 3, \ldots, n$. The vector $\boldsymbol{\theta}$ contains all the parameters of the densities $f_{\boldsymbol{s}_i,l}$ and p_1. The full Bayesian model is completed with prior specification for parameters $\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}, \boldsymbol{\zeta}, \tau^2$ and κ^2. The priors for $\boldsymbol{\theta}$ and $\boldsymbol{\zeta}$ depend on the choices of the densities $f_{\boldsymbol{s}_i,l}$, p_1, and the kernel k', respectively. For parameters $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, τ^2, and κ^2, we specify $N(\boldsymbol{\beta} \,|\, \mu_\beta, \boldsymbol{V}_\beta)$, $N(\boldsymbol{\gamma} \,|\, \mu_\gamma, \boldsymbol{V}_\gamma)$, $\mathrm{IG}(\tau^2 \,|\, u_{\tau^2}, v_{\tau^2})$, and $\mathrm{IG}(\kappa^2 \,|\, u_{\kappa^2}, v_{\kappa^2})$ priors, respectively, where IG denotes the inverse gamma distribution.

3.2 Estimation and prediction

We implement a Gibbs sampler to simulate from the posterior distribution of model parameters $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}, \boldsymbol{\zeta}, \tau^2, \kappa^2)$ and latent variables $\{t_i\}_{i=3}^n$ in (12). Denote by $\boldsymbol{y}_\mathcal{S} = (y(\boldsymbol{s}_1), \ldots, y(\boldsymbol{s}_n))^\top$ and let $\boldsymbol{X}$ be the covariate matrix with the ith row being $\boldsymbol{x}(\boldsymbol{s}_i)^\top$. The posterior full conditional distribution for $\boldsymbol{\beta}$ is $N(\boldsymbol{\beta} \,|\, \boldsymbol{\mu}^*_\beta, \boldsymbol{V}^*_\beta)$ where $\boldsymbol{V}^*_\beta = (\boldsymbol{V}_\beta^{-1} + \tau^{-2}\boldsymbol{X}^\top\boldsymbol{X})^{-1}$ and $\boldsymbol{\mu}^*_\beta = \boldsymbol{V}^*_\beta(\boldsymbol{V}_\beta^{-1}\boldsymbol{\mu}_\beta + \tau^{-2}\boldsymbol{X}^\top(\boldsymbol{y}_\mathcal{S} - \boldsymbol{z}_\mathcal{S}))$. An inverse gamma prior for τ^2 yields an $\mathrm{IG}(\tau^2 \,|\, u_{\tau^2} + n/2, v_{\tau^2} + \sum_{i=1}^n e_i^2/2)$ posterior full conditional, where $e_i = y(\boldsymbol{s}_i) - \boldsymbol{x}(\boldsymbol{s}_i)^\top\boldsymbol{\beta} - z(\boldsymbol{s}_i)$. The posterior full conditional distribution of $\boldsymbol{\theta}$ depends on the choice of $f_{\boldsymbol{s}_i,l}$ and p_1.

We facilitate the update of $z(\boldsymbol{s}_i)$ with a set of configuration variables ℓ_i such that $\ell_2 = 1$ and $\ell_i = l$ if $t_i \in (r^*_{\boldsymbol{s}_i,l-1}, r^*_{\boldsymbol{s}_i,l})$ for $i \geq 3$. The posterior full conditional distribution for $z(\boldsymbol{s}_1)$ is proportional to $N(y(\boldsymbol{s}_1) \,|\, \boldsymbol{x}(\boldsymbol{s}_1)^\top\boldsymbol{\beta} + z(\boldsymbol{s}_1), \tau^2) p_1(z(\boldsymbol{s}_1) \,|\, \boldsymbol{\theta}) \prod_{j:\boldsymbol{s}_{(j,\ell_j)} = \boldsymbol{s}_1} f_{\boldsymbol{s}_j,\ell_j}(z(\boldsymbol{s}_j) \mid z(\boldsymbol{s}_1), \boldsymbol{\theta})$. For $z(\boldsymbol{s}_i)$, $i \geq 2$, the posterior full conditional distribution is proportional to $N(y(\boldsymbol{s}_i) \mid \boldsymbol{x}(\boldsymbol{s}_i)^\top\boldsymbol{\beta} + z(\boldsymbol{s}_i), \tau^2) f_{\boldsymbol{s}_i,\ell_i}(z(\boldsymbol{s}_i) \mid z(\boldsymbol{s}_{(i,\ell_i)}), \boldsymbol{\theta}) \prod_{j:\boldsymbol{s}_{(j,\ell_j)} = \boldsymbol{s}_i} f_{\boldsymbol{s}_j,\ell_j}(z(\boldsymbol{s}_j) \mid z(\boldsymbol{s}_i), \boldsymbol{\theta})$.

We compute the weights $w_l(\boldsymbol{s}_i) = G_{\boldsymbol{s}_i}(r_{\boldsymbol{s}_i,l} \,|\, \mu(\boldsymbol{s}_i), \sigma^2) - G_{\boldsymbol{s}_i}(r_{\boldsymbol{s}_i,l-1} \,|\, \mu(\boldsymbol{s}_i), \sigma^2)$, for $l =$

$1, \ldots, i_L$, and $i = 3, \ldots, n$. The posterior full conditional distribution of $\boldsymbol{\zeta}$ is proportional to $p(\boldsymbol{\zeta}) \prod_{i=3}^{n} (G_{\boldsymbol{s}_i}(r_{\boldsymbol{s}_i,\ell_i}) - G_{\boldsymbol{s}_i}(r_{\boldsymbol{s}_i,\ell_i-1}))$. We update $\boldsymbol{\zeta}$ on its log scale with a random walk Metropolis step. If $\boldsymbol{\zeta}$ is accepted, we update the cutoff points $r_{\boldsymbol{s}_i,l}$ for all i and all l. The posterior full conditional distribution of the latent variable t_i, $i = 3, \ldots, n$, is a piecewise truncated Gaussian distribution $N(t_i \,|\, \mu(\boldsymbol{s}_i), \kappa^2) 1_{(r^*_{\boldsymbol{s}_i,l-1}, r^*_{\boldsymbol{s}_i,l})}(t_i)$ with probability proportional to $w_l(\boldsymbol{s}_i) f_{\boldsymbol{s}_i,l}$. Let $\boldsymbol{D}$ be the matrix such that the ith row is $(1, s_{i1}, s_{i2})$ where s_{i1}, s_{i2} are the first and second coordinates of $\boldsymbol{s}_i$. The posterior full conditional distribution of $\boldsymbol{\gamma}$ is $N(\boldsymbol{\gamma} \,|\, \boldsymbol{\mu}^*_\gamma, \boldsymbol{V}^*_\gamma)$ where $\boldsymbol{V}^*_\gamma = (\boldsymbol{V}^{-1}_\gamma + \kappa^{-2} \boldsymbol{D}^\top \boldsymbol{D})^{-1}$ and $\boldsymbol{\mu}^*_\gamma = \boldsymbol{V}^*_\gamma (\boldsymbol{V}^{-1}_\gamma \boldsymbol{\mu}_\gamma + \kappa^{-2} \boldsymbol{D}^\top \boldsymbol{t})$ with $\boldsymbol{t} = (t_3, \ldots, t_n)^\top$. The posterior full conditional distribution of κ^2 is $\mathrm{IG}(\kappa^2 \,|\, u_{\kappa^2} + (n-2)/2, v_{\kappa^2} + \sum_{i=3}^{n} (t_i - \mu(\boldsymbol{s}_i))^2/2)$.

Let $\boldsymbol{v}_0 \in \mathcal{D}$ where the predictor $x(\boldsymbol{v}_0)$ is observed. We obtain posterior predictive samples of $y(\boldsymbol{v}_0)$ by the following steps. If $\boldsymbol{v}_0 \notin \mathcal{S}$, for each posterior sample of the parameters, we first compute the cutoff points $r_{\boldsymbol{v}_0,l}$ for which $r_{\boldsymbol{v}_0,l} - r_{\boldsymbol{v}_0,l-1} = k'(\boldsymbol{v}_0, \boldsymbol{v}_{(0l)}) / \sum_{l=1}^{L} k'(\boldsymbol{v}_0, \boldsymbol{v}_{(0l)})$, and obtain the weights $w_l(\boldsymbol{v}_0) = G_{\boldsymbol{v}_0}(r_{\boldsymbol{v}_0,l}) - G_{\boldsymbol{v}_0}(r_{\boldsymbol{v}_0,l-1})$ for $l = 1, \ldots, L$. We then predict $z(\boldsymbol{v}_0)$ using (4), and generate $y(\boldsymbol{v}_0)$ using (10). If $\boldsymbol{v}_0 \equiv \boldsymbol{s}_i \in \mathcal{S}$, we generate $y(\boldsymbol{v}_0)$ similar to the earlier case but using posterior samples of the weights obtained from the MCMC, and applying (3) instead of (4) to generate $z(\boldsymbol{v}_0)$.

4 Simulation study

We conduct three simulation experiments to demonstrate the benefits of the proposed modeling framework. First, we demonstrate the effectiveness of the GNNMP in approximating Gaussian random fields, by comparing its performance to the NNGP model. Next, we illustrate the capacity of the skew-GNNMP model to capture different levels of skewness over the domain. Finally, we study inference for tail dependence using copula NNMP models.

In each of the experiments, we created a regular grid of 200×200 resolution on a unit square domain, and generated data on each grid location. We then randomly selected a subset of observations as the reference set for model fitting. For all experiments, we used a random ordering for the reference set. For model assessment and comparison, we used root mean square predictive error (RMSPE), 95% posterior credible interval coverage rate

(95% CI coverage), deviance information criterion (DIC; Spiegelhalter et al. 2002), GG criterion (Gelfand and Ghosh, 1998), continuous ranked probability score (CRPS; Gneiting and Raftery 2007), and log-score (Gneiting and Raftery, 2007).

All posterior analyses are based on posterior samples collected every 10 iterations from a Markov chain of 30000 iterations, with the first 10000 samples being discarded.

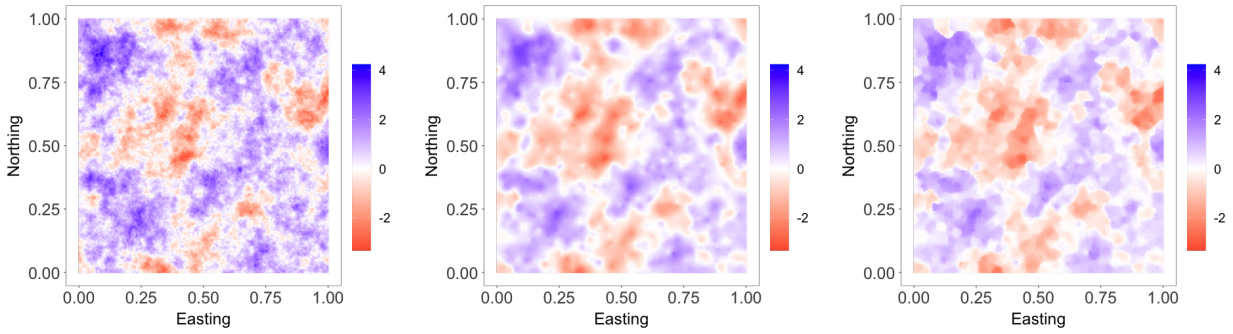
4.1 First experiment

We generated data from the regression in (10), where $z(\boldsymbol{v})$ follows a zero-mean Gaussian process with a unit variance and an exponential correlation function with range parameter 1/12. We included an intercept and a covariate drawn from $N(0,1)$ in the model, and chose $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top = (1,5)^\top$, and $\tau^2 = 0.1$. The setting followed Datta et al. (2016a).

We applied two models. The first one assumes that $z(\boldsymbol{v})$ follows an NNGP model with variance σ_0^2 and exponential correlation function with range paramger ϕ_0. The second one assumes that $z(\boldsymbol{v})$ follows the GNNMP model defined in (8) with $\mu = \mu_l$ and $\sigma^2 = \sigma_l^2$ for all l, such that $z(\boldsymbol{v})$ has a stationary marginal $N(\mu, \sigma^2)$. For the GNNMP, we used exponential correlation functions with range parameter ϕ and ζ, respectively, for the correlation with respect to the component density, and the kernel function that defines the cutoff points for the weights. Details of the MCMC algorithm to implement the GNNMP model are provided in the Supplementary Material. For the NNGP model, we implement the latent NNGP algorithm from the spNNGP package in R (Finley et al., 2020)

For both models, the regression coefficients $\boldsymbol{\beta}$ were assigned flat priors. The variances σ_0^2 and σ^2 received the same inverse gamma prior IG$(2,1)$, and τ^2 was assigned IG$(2, 0.1)$. The range parameter ϕ_0 of the NNGP received a uniform prior Unif$(1/30, 1/3)$, while the range parameters ϕ and ζ of the GNNMP received inverse gamma priors IG$(3, 1/3)$ and IG$(3, 0.2)$, respectively. Regarding the logit Gaussian distribution parameters, $\boldsymbol{\gamma}$ and κ^2, we used $N((-1.5, 0, 0), \text{diag}(2\mathbf{1}_3))$ and IG$(3,1)$ priors, respectively.

The posterior estimates from the two models for the common parameters, $\boldsymbol{\beta}$ and τ^2, were quite close. The RMSPE, 95% CI coverage, and CRPS from the GNNMP model were 0.03, 0.01, and 0.02 higher than the NNGP model, respectively. The DIC and GG



(a) True GP (b) NNGP ($L = 10$) (c) GNNMP ($L = 10$)

Figure 1: Synthetic data analysis - first experiment. Interpolated surfaces of the true Gaussian process and posterior median estimates from the NNGP and GNNMP models.

criterion are also higher than the NNGP by small margins. The posterior median estimate of the spatial random effects from both models are shown in Figure 1. On the whole, the GNNMP model provides a reasonably good approximation to the Gaussian random field. Moreover, the performance metrics of the GNNMP model are comparable to those of the NNGP model, the model assumptions of which are more well suited to the particular synthetic data example.

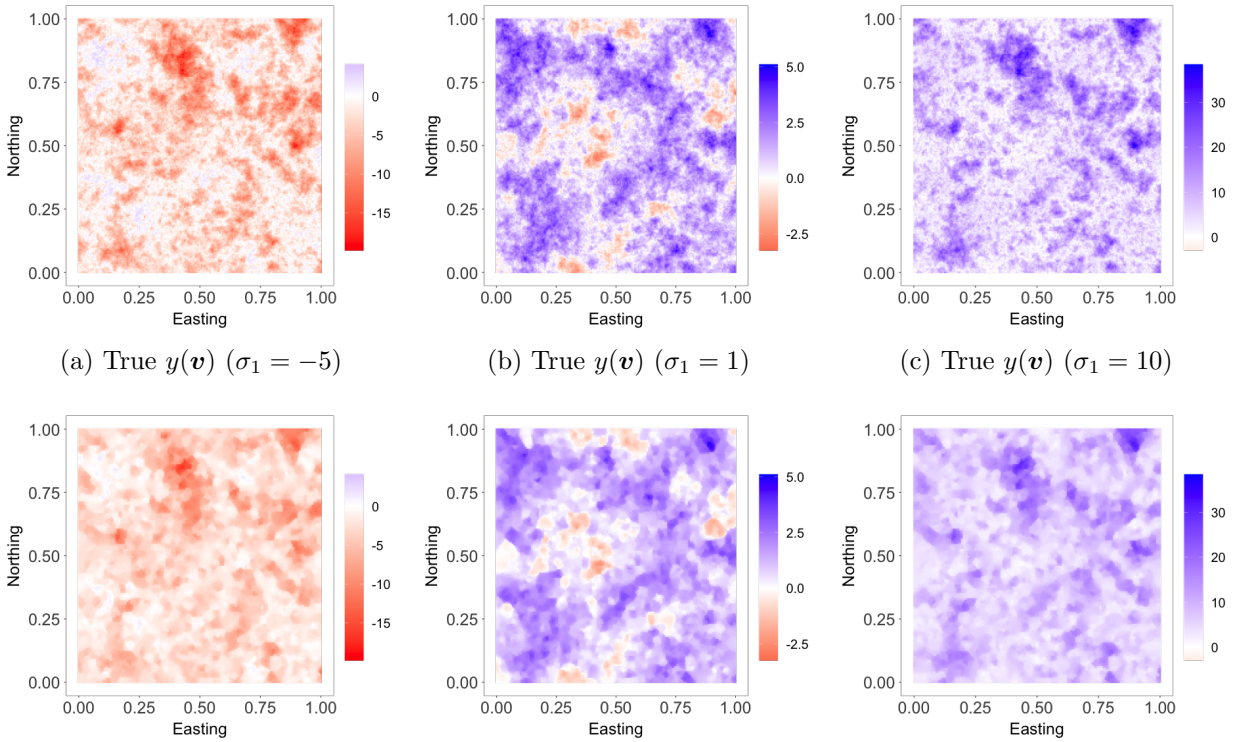
4.2 Second experiment

In this scenario, we generated data from the following skew-Gaussian process (Zhang and El-Shaarawi, 2010),

$$y(\boldsymbol{v}) = \sigma_1 \, |\omega_1(\boldsymbol{v})| + \sigma_2 \, \omega_2(\boldsymbol{v}), \quad \boldsymbol{v} \in \mathcal{D} \tag{13}$$

where $\omega_1(\boldsymbol{v})$ and $\omega_2(\boldsymbol{v})$ are both zero-mean standard Gaussian processes with correlation matrix specified by an exponential correlation function with range parameter $1/12$. The parameter $\sigma_1 \in \mathbb{R}$ controls the skewness, whereas $\sigma_2 > 0$ is a scale parameter. The model has a stationary skew-Gaussian (Azzalini, 2013) marginal density $f_Y(y) = 2N(y \,|\, 0, \sigma_1^2 + \sigma_2^2)\Phi(\sigma_1 y/(\sigma_2\sqrt{\sigma_1^2 + \sigma_2^2}))$. We took $\sigma_2 = 1$, and generated data with $\sigma_1 = -5$, 1 and 10, resulting in three different random fields that are, respectively, moderately negative-skewed, slightly positive-skewed, and strongly positive-skewed, as shown in Figure 2(a)-2(c).

We applied the stationary skew-GNNMP model with $L = 10$ to 2000 observations. The model is obtained by modifying the GNNMP model in (8), taking $\sigma_l^2 = \sigma^2$, and $\mu_l = \lambda z_0$,



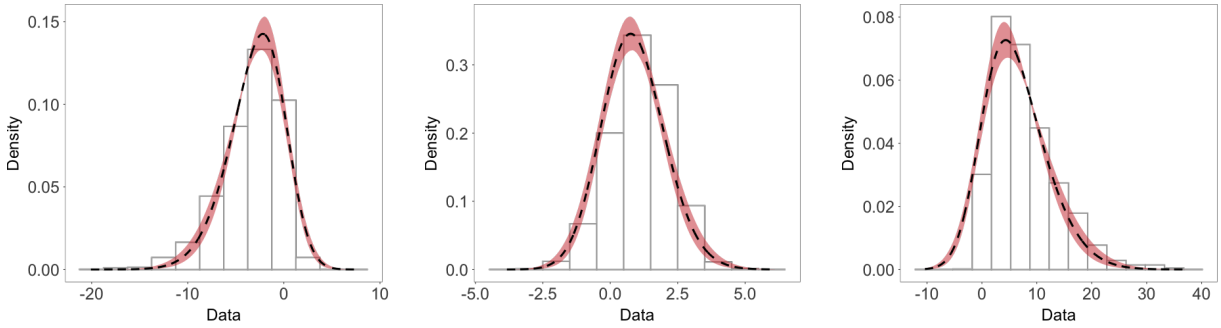
(a) True $y(\boldsymbol{v})$ ($\sigma_1 = -5$) (b) True $y(\boldsymbol{v})$ ($\sigma_1 = 1$) (c) True $y(\boldsymbol{v})$ ($\sigma_1 = 10$)

(d) Skew-GNNMP ($L = 10$) (e) Skew-GNNMP ($L = 10$) (f) Skew-GNNMP ($L = 10$)

Figure 2: Synthetic data analysis - second experiment. Top panels are interpolated surfaces of $y(\boldsymbol{v})$ generated by (13). Bottom panels are the posterior median estimates from the skew-GNNMP model.

with $z_0 \sim N(0,1)I(z_0 \geq 0)$. Here, $\lambda \in \mathbb{R}$ controls the skewness, such that a large positive (negative) value of λ indicates strong positive (negative) skewness. If $\lambda = 0$, the skew-GNNMP model reduces to the GNNMP model. After marginalizing out z_0, we obtain a stationary skew-Gaussian density $f_Z(z) = 2N(z \,|\, 0, \lambda^2+\sigma^2)\Phi(\lambda z/(\sigma\sqrt{\lambda^2+\sigma^2}))$. We completed the full Bayesian specification for the model, by assigning priors $N(\lambda \,|\, 1, 5)$, $\text{IG}(\sigma^2 \,|\, 2, 1)$, $\text{IG}(\phi \,|\, 3, 1/3)$, $\text{IG}(\zeta \,|\, 3, 0.2)$, $N(\boldsymbol{\gamma} \,|\, (-1.5, 0, 0), \text{diag}(2\mathbf{1}_3)))$ and $\text{IG}(\kappa^2 \,|\, 3, 1)$, where ζ is the range parameter of the exponential correlation function for the random cutoff points of the weights. Details of the model and implementation are given in the Supplementary Material.

We focus on the model performance on capturing skewness. The posterior mean and 95% credible interval of λ for the three scenarios were $-3.65\,(-4.10, -3.27)$, $1.09\,(0.91, 1.28)$ and $7.69\,(6.88, 8.68)$, respectively, indicating the model's ability to estimate different levels of skewness. The bottom row of Figure 2 shows that the posterior median estimates of the surfaces capture well features of the true surfaces, even when the level of skewness is



(a) Histogram of $y(\boldsymbol{v})$ ($\sigma_1 = -5$) (b) Histogram of $y(\boldsymbol{v})$ ($\sigma_1 = 1$) (c) Histogram of $y(\boldsymbol{v})$ ($\sigma_1 = 10$)

Figure 3: Synthetic data analysis - second experiment. Posterior means (dashed lines) and 95% credible intervals (shaded regions) for the stationary marginal density.

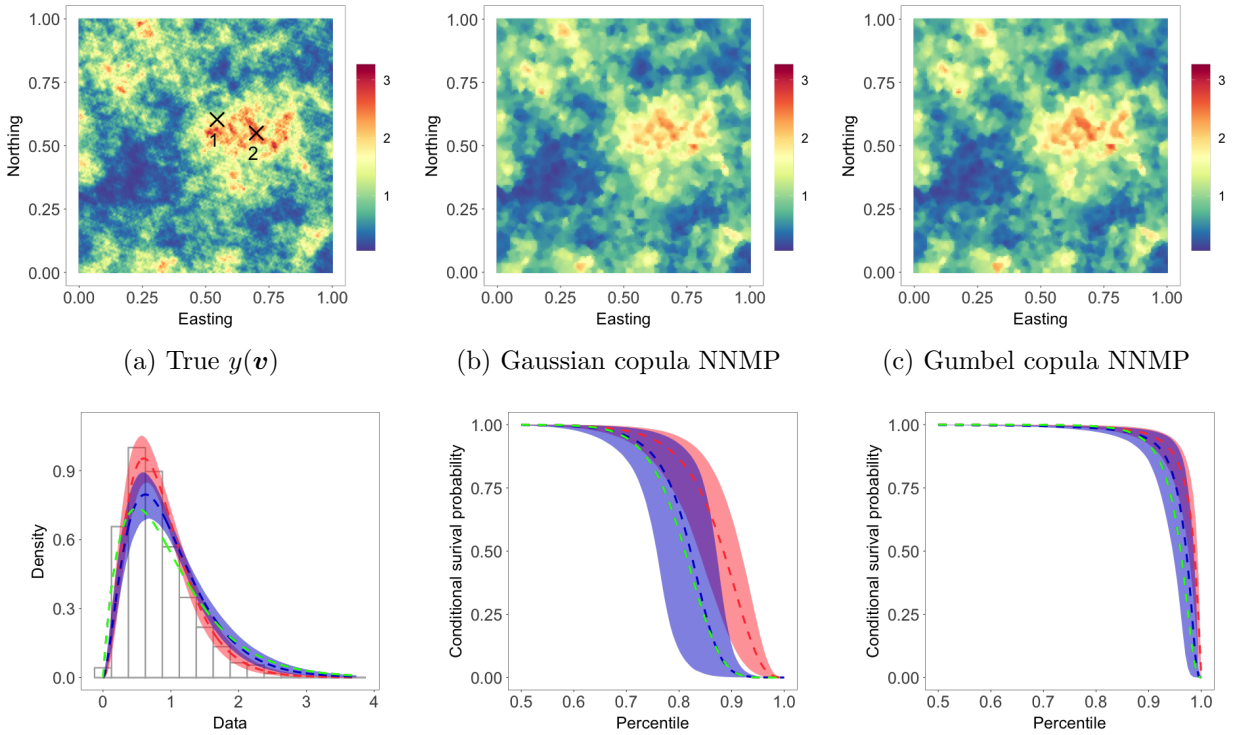
small, thus demonstrating that the model is also able to recover near-Gaussian features. Figure 3 plots the posterior mean and pointwise 95% credible interval for the marginal density, overlaid on the histogram of the simulated data for each of the three cases. These estimates demonstrate the adaptability of the skew-GNNMP model in capturing skewed random fields with different levels of skewness.

4.3 Third experiment

The goal of the last experiment is to demonstrate the use of copulas to construct NNMPs for tail dependence modeling. We note that the focus here is to illustrate the flexibility of the NNMPs with copulas for modeling complex dependence structures, but not for extreme value modeling. To this end, we generated data from the random field

$$y(\boldsymbol{v}) = F^{-1}(T_\nu(\omega(\boldsymbol{v}))), \quad \boldsymbol{v} \in \mathcal{D}, \tag{14}$$

where $\omega(\boldsymbol{v})$ is a zero-mean standard Student-t process with tail parameter ν and scale matrix specified by an exponential correlation function with range parameter ϕ_w. The distribution functions F and T_ν correspond to a gamma distribution $\text{Ga}(2, 2)$ 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 $\chi_\nu = 2T_{\nu+1}(-\sqrt{(\nu+1)(1-\rho_0)/(1+\rho_0)})$. We took $\phi_w = 1/12$, and chose $\nu = 10$ so that the synthetic data exhibits moderate tail dependence at close



(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 4: Synthetic data analysis - third experiment. Top panels are interpolated surfaces of the true field generated by (14) and posterior median estimates from both models with $L = 10$. 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) dash lines and shaded regions are the posterior means and 95% credible intervals from the Gaussian (Gumbel) copula NNMP models.

distance, and the dependence decreases rapidly as the distance d_0 becomes larger. When $\rho_0 = 0.05, 0.5, 0.95$, $\chi_{10} = 0.01, 0.08, 0.61$, respectively.

We applied two copula NNMP models to 2000 observations. The models are of the form in (9) with stationary gamma marginal $\text{Ga}(a, b)$. In the first model, the component copula density $c_{\boldsymbol{v},l}$ corresponds to a bivariate Gaussian copula that is known to be unsuitable for tail dependence modeling. The correlation parameter of the copula was specified by an exponential correlation function with range parameter ϕ_1. In the second model, we consider a spatially varying Gumbel copula as in Example 2. For each component copula density $c_{\boldsymbol{v},l}$, we define the spatially varying parameter through 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

Table 1: Synthetic data analysis - third experiment. Log-scores for subsets that exceed the c-th percentile of the held-out data

c	0	10	30	50	70	90	95
Gaussian copula model	140.009	80.713	30.224	-15.479	-28.806	-20.377	-13.750
Gumbel copula model	118.684	63.962	17.242	-24.300	-28.232	-16.952	-11.422

of the Gumbel copula model stays the same for any distance smaller than d_0 between two locations. For the random cutoff points of the mixture weights, we specified an exponential correlation function with range parameters ζ_1 and ζ_2, respectively, for each model. The Bayesian model is fully specified 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, $N(\boldsymbol{\gamma} \,|\, (-1.5, 0, 0), 2\text{diag}(2\mathbf{1}_3))$ and IG$(\kappa^2 \,|\, 3, 1)$ priors. Model and computation details are provided in the Supplementary Material.

We focus on the performance of the two models with respect to tail dependence inference. Table 1 presents the log-scores for subsets of the held-out data that exceed the c-th percentile of the held-out data. The out-of-sample log-score is the predictive log-likelihood averaging over the model parameters. It reflects the ability of a model to capture dependence structure in the data. We can see that for held-out data that exceed high sample percentiles, the Gumbel copula model gives a higher log-score.

Figure 4 shows the random fields, marginals and conditional survival probabilities estimated by the two models. From Figure 4(a)-4(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, the Gumbel copula NNMP 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 4(a). In particular, Site 1 is surrounded with reference observations with moderate values, while Site 2 is surrounded with large reference observations. We see that the Gumbel copula model provides much closer estimates to the 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.

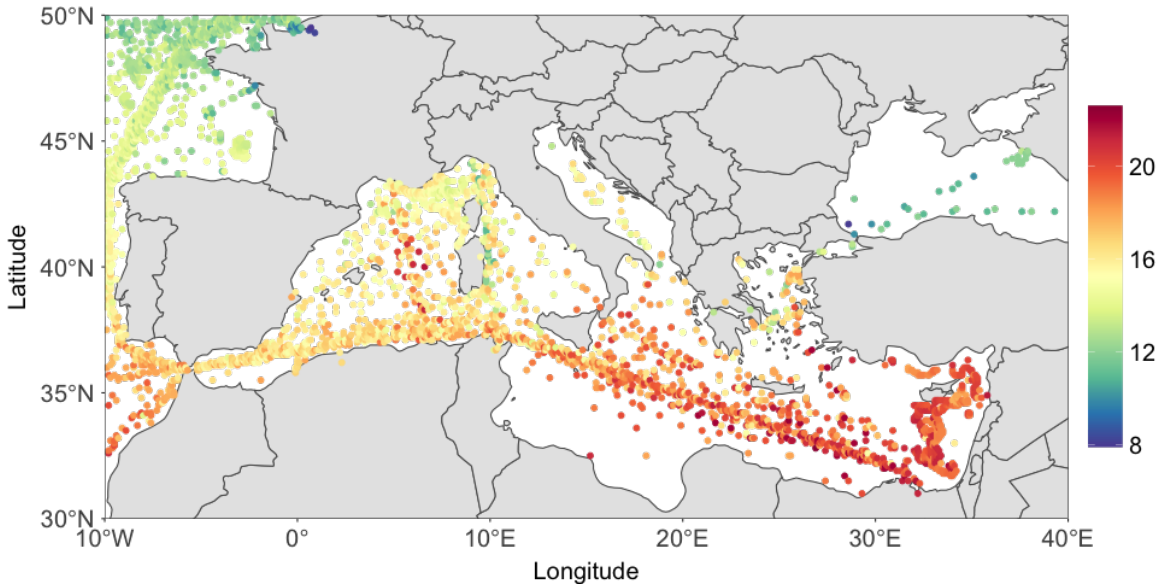


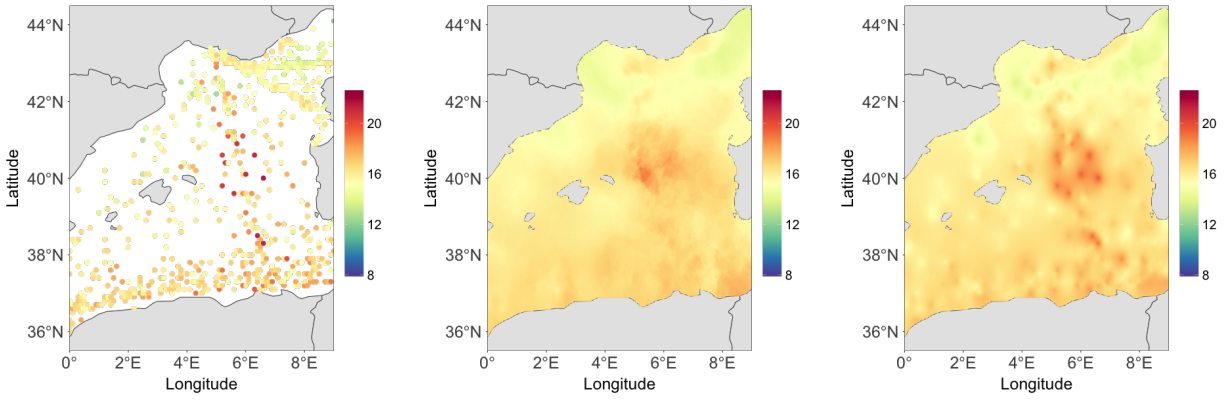
Figure 5: SST data analysis: observed Mediterranean SST.

5 Application: Mediterranean sea surface temperature data analysis

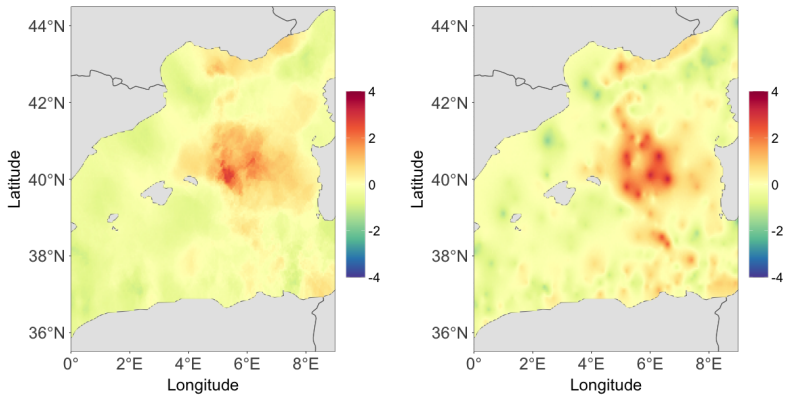
The study of Ocean's dynamics is crucial for understanding of climate variability. One of the most valuable sources of information regarding the evolution of the state of the ocean is provided by the centuries-long record of temperature observations recorded from the surface of the oceans. The record of sea surface temperatures (SST) consists of data collected over time at irregularly scattered locations. This information needs to be processed into spatially continuous fields. In this section, we examine the SST from the Mediterranean sea during December 2003, as shown in Figure 5. The data consist of in situ measurements from different types of devices. Some locations had multiple observations,thus for these sites, we took the median of the observations, resulting in totally 3072 observations.

To deal with the difficulty of producing a spatially continuous field from observations that are irregular-spaced along ship lines and account for the complexity of the surrounding coastlines as well as the circulation system, we considered the spatially varying regression model in (10). From Figure 5, we observed that SST tend to increase as latitude decreases, so we took latitude as a covariate to account for the long range variability in SST.

We first focus on SST along the islands near the shores of Spain, France, Monaco and Italy, between 0 - 9 E. longitude and 33.5 - 44.5 N. latitude. The SST observations in the region, as shown in Figure 6(a), are very heterogeneous, implying that the short range variability is likely to be non-Gaussian. We applied the GNNMP for the spatial random effect in (10), and applied the NNGP for comparison. For both models, we chose the neighbor size $L = 10$ and applied the same prior specifications as in the first experiment. A random topological ordering was used for both models. We took 64 observations, around 10% of the data in the region, as the held-out data for model comparison and used the remaining 580 observations to train the models. For both models, we ran the MCMC with 120000 iterations, discarding the first 20000 samples, and collected samples every 20 iterations.



(a) Regional SST (b) Predicted SST (GNNMP) (c) Predicted SST (NNGP)



(d) Random effect (GNNMP) (e) Random effect (NNGP)

Figure 6: SST data analysis. Panels (a) is observation at the selected region. Panels (b) - (e) are posterior median estimates of the SST and spatial random effects by the GNNMP and NNGP models with $L = 10$.

Table 2: Performance metrics of the GNNMP and NNGP models

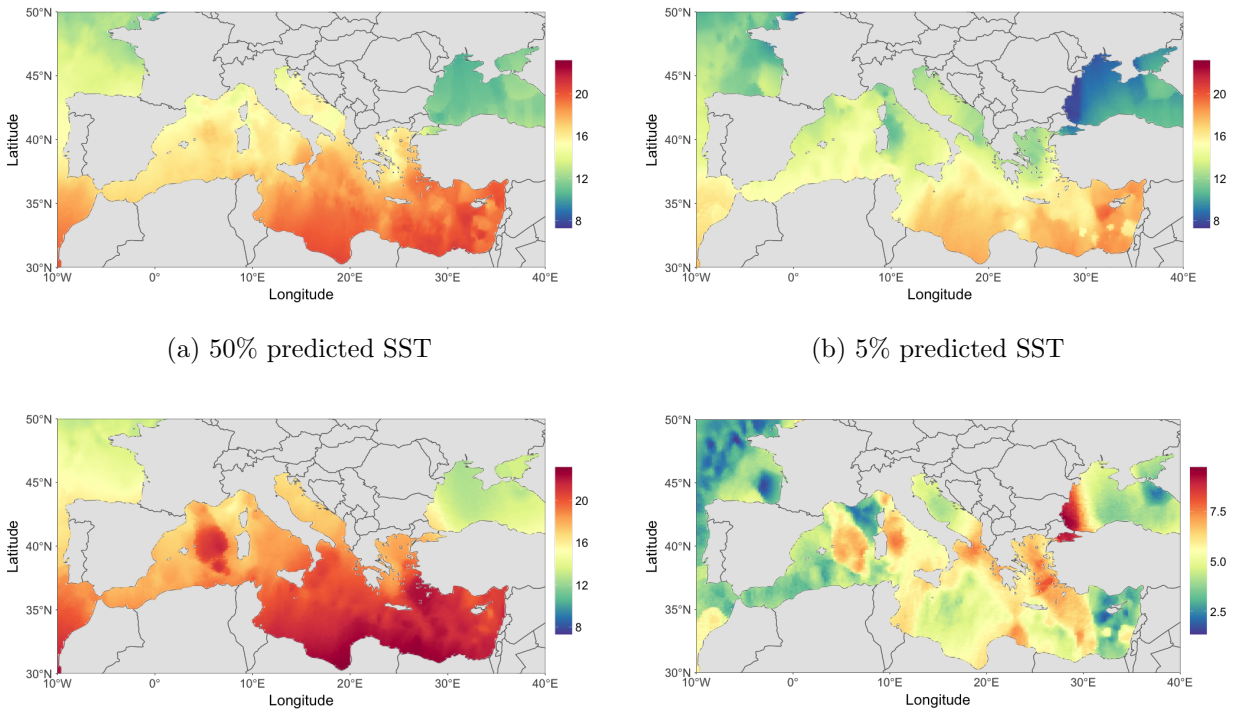
	RMSPE	95% CI	95% CI width	CRPS	G	P	D	DIC
GNNMP	1.03	0.95	4.31	0.56	6.31	103.07	109.37	529.21
NNGP	1.07	0.95	4.35	0.59	176.15	529.06	705.22	1530.67

The posterior mean (95% CI) of the regression intercept from the GNNMP was higher than the NNGP. They were $28.38\,(21.92, 37.60)$ and $25.95\,(22.03, 29.88)$, respectively. The posterior estimates of the coefficient for latitude given by the GNNMP and the NNGP were $-0.31\,(-0.54, -0.15)$ and $-0.25\,(-0.35, -0.15)$, respectively, both indicating that there was a trend of SST decreasing in the latitude at the selected region. For the measurement error τ^2, the GNNMP provided a smaller estimate $0.09\,(0.02, 0.27)$, compared to $0.61\,(0.08, 0.91)$ from the NNGP.

Table 2 shows the performance metrics for both models. Both the DIC and the GG criterion suggest that the GNNMP had a better goodness-of-fit than the NNGP. For out-of-sample prediction, the GNNMP produced smaller RMSPE and CRPS than the NNMP, Both models gave the same accurate 95% CI coverage with similar coverage widths. Figure 6(b)-6(e) show the posterior median estimates of the temperature field and the spatial random effects from both models. Compared to the NNGP, we see that the GNNMP provided much smoother surfaces which closely represents the pattern in the observations.

Given the result for the selected region, we concluded that the GNNMP model was a better candidate to capture the spatial heterogeneity in the Mediterranean SST. We applied the GNNMP model to the whole dataset, with the same prior specification for fitting the regional data. We chose a large $L = 30$ and let the model estimate effective weights for the neighbors. We note that when fitting the model to the whole dataset, the Gibbs sampler experienced slow convergence issues, since the data is strongly heterogeneous and the elements of the spatial random effect $z_{\mathcal{S}}$ was sequentially updated in the sampler. To achieve better mixing, we ran the algorithm for 300000 iterations and collected samples every 30 iterations after the first 150000 as burn-in.

The posterior estimates of the regression parameters β_0 and β_1 were $32.26\,(30.83, 34.25)$ and $-0.42\,(-0.46, -0.38)$, respectively. The estimate suggested that the SST decreased as



(a) 50% predicted SST

(b) 5% predicted SST

(c) 95% predicted SST

(d) 95% CI width

Figure 7: SST data analysis: posterior median (a), 5-th percentile (b), 95-th percentile (c) and 95% credible interval (d) from the GNNMP model with $L = 10$.

the latitude increased for the whole map.

Figure 7(a)-7(c) illustrates the posterior quantile summary of the predicted SST. Compared to Figure 5, we see that the posterior median estimate resembles the observed pattern. The prediction was quite smooth even for areas with few observations. The 95% credible interval width in Figure 7(d) shows that the model describes the uncertainty realistically. The uncertainties were high in areas where there were less observations or the observations were volatile. Overall, we see the model's ability to capture the spatial heterogeneity over the Mediterranean SST.

6 Summary and discussion

We have introduced a class of geostatistical models for large, non-Gaussian data sets, based on nearest neighbor processes. Using an MTD model as the parent process, we have demonstrated the NNMP's flexibility for modeling complex dependence by specification of

a collection of bivariate distributions indexed at space.

The computation of the NNMP not only bypasses all the potential issues from large matrix operations, but also enhances modeling power. Kernel functions, such as wave covariance functions that were impractical for the 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. One limitation of the NNMP's computation, similar to mixture models, is that the Markov chain may experience slow convergence issues. We will need further development on efficient algorithms for fast computation, especially when dealing with large, complex data sets.

An NNMP model requires a selection of the size L of the neighbor sets. In general, a larger L increases computational costs. Datta et al. (2016a) conclude from their experiments that a moderate value L (≤ 20) typically suffices for the NNGP to fit Gaussian data sets. On the other hand, Peruzzi et al. (2020) point out that a smaller L corresponds to a larger Kullback-Leibler divergence of $\tilde{p}(\boldsymbol{z}_{\mathcal{S}})$ from $p(\boldsymbol{z}_{\mathcal{S}})$, regardless of the distributional assumption of the density. Moreover, it is possible that information from the farthest neighbors is also important (Stein et al., 2004). Therefore, for non-Gaussian data sets in which the dependence is complex, one may seek a large L to obtain a better approximation to the full model. Our prior model for the weights enables us to take a relatively large neighbor set with less computational demand. We assign small probabilities a priori to distant neighbors. The contribution of each neighbor is induced by the mixing.

In this article, we have focused on developing the framework for continuous data. The proposed approach can be naturally extended to modeling discrete data, which is being undertaken. Modeling options for geostatistical count data in the existing literature involve either spatial generalized linear mixed models (Diggle et al., 1998) 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. We believe that the extension will provide both inferential and computational benefits to modeling large discrete data sets.

Other research directions include extensions to multivariate and spatio-temporal set-

tings. 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 for both time and space, there is a large scope for the exploration of integrating the time component into the model. An immediate attempt may be to follow the setup in (12) of Datta et al. (2016a), while more interesting avenues will be to modify the weights or the mixture component of the NNMP accordingly.

Supplementary Material

Supplementary Material includes proofs of the propositions and MCMC implementation details of the models applied in the data examples.

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.

Arnold, B. C., Castillo, E., Sarabia, J.-M., and Sarabia, J. M. (1999). *Conditional specification of statistical models*. Springer Science & Business Media.

Azzalini, A. (2013). *The skew-normal and related families*, volume 3. Cambridge University Press.

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

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*, page e2616.

Bevilacqua, M., Caamaño-Carrillo, C., Arellano-Valle, R. B., and Morales-Oñate, V. (2020). Non-Gaussian geostatistical modeling using (skew) t processes. *Scandinavian Journal of Statistics*.

Cadonna, A., Kottas, A., and Prado, R. (2019). Bayesian spectral modeling for multiple time series. *Journal of the American Statistical Association*, 114:1838–1853.

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.

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.

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.

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.

Ferguson, T. S. (1973). A bayesian analysis of some nonparametric problems. *The annals of statistics*, pages 209–230.

Finley, A., Datta, A., and Banerjee, S. (2020). *spNNGP: Spatial Regression Models for Large Datasets using Nearest Neighbor Gaussian Processes*. R package version 0.1.4.

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.

Gelfand, A. E. and Ghosh, S. K. (1998). Model choice: a minimum posterior predictive loss approach. *Biometrika*, 85(1):1–11.

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.

Ghosh, S. and Mallick, B. K. (2011). A hierarchical Bayesian spatio-temporal model for extreme precipitation events. *Environmetrics*, 22(2):192–204.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378.

Gräler, B. (2014). Modelling skewed spatial random fields through the spatial vine copula. *Spatial Statistics*, 10:87–102.

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.

Hua, L. and Joe, H. (2014). Strength of tail dependence based on conditional tail expectation. *Journal of Multivariate Analysis*, 123:143–159.

Joe, H. (2014). *Dependence modeling with copulas*. CRC press.

Katzfuss, M. and Guinness, J. (2021). A general framework for Vecchia approximations of gaussian processes. *Statistical Science*, 36(1):124–141.

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.

Lauritzen, S. L. (1996). *Graphical models*, volume 17. Clarendon Press.

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.

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.

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.

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.

Müller, P., Quintana, F. A., and Page, G. (2018). Nonparametric Bayesian inference in applications. *Statistical Methods & Applications*, 27(2):175–206.

Palacios, M. B. and Steel, M. F. J. (2006). Non-Gaussian Bayesian geostatistical modeling. *Journal of the American Statistical Association*, 101(474):604–618.

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*, 0(0):1–14.

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.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the royal statistical society: Series B (statistical methodology)*, 64(4):583–639.

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.

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.

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.

Wallin, J. and Bolin, D. (2015). Geostatistical modelling using non-Gaussian matérn fields. *Scandinavian Journal of Statistics*, 42(3):872–890.

Xu, G. and Genton, M. G. (2017). Tukey g-and-h random fields. *Journal of the American Statistical Association*, 112(519):1236–1249.

Zhang, H. and El-Shaarawi, A. (2010). On spatial skew-Gaussian processes and applications. *Environmetrics: The official journal of the International Environmetrics Society*, 21(1):33–47.

Zheng, X., Kottas, A., and Sansó, B. (2020). On construction and estimation of stationary mixture transition distribution models. *arXiv:2010.12696*.

Supplementary Material for Nearest-Neighbor Geostatistical Models for Non-Gaussian Data

A Proofs

Proof of Proposition 1. We consider a real-valued univariate spatial process $Z(\boldsymbol{v}), \boldsymbol{v} \in \mathcal{D} \subset \mathbb{R}^p, p \geq 1$. Let $\mathcal{S} \subset \mathcal{D}$ be a reference set. Without loss of generality, we consider the continuous case, i.e., $Z(\boldsymbol{v})$ has a continuous distribution for which its density exists, for all $\boldsymbol{v} \in \mathcal{D}$. To verify the proposition, we partition the domain $\mathcal{D}$ into the reference set $\mathcal{S}$ and the nonreference set $\mathcal{U}$.

Given any $\boldsymbol{v} \in \mathcal{D}$ with a bivariate random vector indexed at $\boldsymbol{v}$, denoted as $(U_{\boldsymbol{v},l}, V_{\boldsymbol{v},l})$, we denote $f_{\boldsymbol{v},l}$ as the conditional density of $U_{\boldsymbol{v},l}$ given $V_{\boldsymbol{v},l}$, and $f_{U_{\boldsymbol{v},l}}, f_{V_{\boldsymbol{v},l}}$ as the marginal densities of $U_{\boldsymbol{v},l}$, $V_{\boldsymbol{v},l}$, respectively. Using the proposition assumption that $f_Z = f_{U_{\boldsymbol{v},l}} = f_{V_{\boldsymbol{v},l}}$, we have that

$$\int_{\mathbb{R}} f_{\boldsymbol{v},l}(u \mid v) f_Z(v) dv = \int_{\mathbb{R}} f_{\boldsymbol{v},l}(u \mid v) f_{V_{\boldsymbol{v},l}}(v) dv = f_{U_{\boldsymbol{v},l}}(u) = f_Z(u), \tag{1}$$

for every $\boldsymbol{v} \in \mathcal{D}$ and for all l.

We first prove the result for the reference set $\mathcal{S}$. By the model assumption, locations in $\mathcal{S}$ are ordered. In this regard, using the proposition assumptions, we can show that $Z(\boldsymbol{s}) \sim f_Z$ for all $\boldsymbol{s} \in \mathcal{S}$ by applying Proposition 1 in Zheng et al. (2020).

Turning to the nonreference set $\mathcal{U}$. Let $g_{\boldsymbol{u}}(z(\boldsymbol{u}))$ be the marginal density of $Z(\boldsymbol{u})$ for every $\boldsymbol{u} \in \mathcal{U}$. Denote by $\tilde{p}(\boldsymbol{z}_{\mathrm{Ne}(\boldsymbol{u})})$ the joint density for the random vector $\boldsymbol{z}_{\mathrm{Ne}(\boldsymbol{u})}$ where $\mathrm{Ne}(\boldsymbol{u}) \subset \mathcal{S}$, so every element of $\boldsymbol{Z}_{\mathrm{Ne}}$ has marginal density f_Z. Then, the marginal density

for $Z(\boldsymbol{u})$ is given by:

$$\begin{aligned} g_{\boldsymbol{u}}(z(\boldsymbol{u})) &= \int_{\mathbb{R}^L} p(z(\boldsymbol{u}) \mid \boldsymbol{z}_{\mathrm{Ne}(\boldsymbol{u})}) \tilde{p}(\boldsymbol{z}_{\mathrm{Ne}(\boldsymbol{u})}) \prod_{\{\boldsymbol{s}_i \in \mathrm{Ne}(\boldsymbol{u})\}} d(z(\boldsymbol{s}_i)) \\ &= \sum_{l=1}^{L} w_l(\boldsymbol{v}) \int_{\mathbb{R}^L} f_{\boldsymbol{v},l}(z(\boldsymbol{u}) \mid z(\boldsymbol{u}_{(l)})) \tilde{p}(\boldsymbol{z}_{\mathrm{Ne}(\boldsymbol{u})}) \prod_{\{\boldsymbol{s}_i \in \mathrm{Ne}(\boldsymbol{u}), \boldsymbol{s}_i \neq \boldsymbol{u}_{(l)}\}} d(z(\boldsymbol{s}_i)) \\ &= \sum_{l=1}^{L} w_l(\boldsymbol{v}) \int_{\mathbb{R}} f_{\boldsymbol{v},l}(z(\boldsymbol{u}) \mid z(\boldsymbol{u}_{(l)})) f_Z(z(\boldsymbol{u}_{(l)})) d(z(\boldsymbol{u}_{(l)})) \\ &= f_Z(z(\boldsymbol{u})), \end{aligned}$$

where the second-to-last equality holds by the result that $Z(\boldsymbol{s}) \sim f_Z$ for all $\boldsymbol{s} \in \mathcal{S}$ and $\mathrm{Ne}(\boldsymbol{u}) \subset \mathcal{S}$ for every $\boldsymbol{u} \in \mathcal{U}$. The last equality follows from (1).

□

Proof of Proposition 2. We verify the proposition by partitioning the domain $\mathcal{D}$ into the reference set $\mathcal{S}$ and the nonreference set $\mathcal{U}$. We first prove by induction the result for the joint distribution $\tilde{p}(\boldsymbol{z}_{\mathcal{S}})$ over $\mathcal{S}$. Then to complete the proof, it suffices to show that for every location $\boldsymbol{u} \in \mathcal{U}$, the joint density $\tilde{p}(\boldsymbol{z}_{\mathcal{U}_1})$ is a mixture of multivariate Gaussian distributions, where $\mathcal{U}_1 = \mathcal{S} \cup \{\boldsymbol{u}\}$.

Without loss of generality, we assume $\mu = 0$ for the stationary GNNMP with invariant marginal $f_Z(z) = N(z \,|\, 0, \sigma^2)$. The associated spatially varying conditional density is

$$p(z(\boldsymbol{v}) | \boldsymbol{z}_{\mathrm{Ne}(\boldsymbol{v})}) = \sum_{l=1}^{i_L} w_l(\boldsymbol{v}) N(z(\boldsymbol{v}) | \rho_l(\boldsymbol{v}) z(\boldsymbol{v}_{(l)}), \sigma^2(1 - (\rho_l(\boldsymbol{v}))^2)),$$

where for, $i = 3, \ldots, L$, $i_L = i - 1$, and for $i > L$, $i_L = L$. For each i, we denote as $\{w_{i,l_{i-2}}\}_{l_{i-2}=1}^{i_L}$ the vector of mixture weights, as $\{\rho_{i,l_{i-2}}\}_{l_{i-2}=1}^{i_L}$ the vector of the correlation coefficients, and as $\{z_{i,l_{i-2}}\}_{l_{i-2}=1}^{i_L}$ the vector of the nearest neighbors of z_i, for $i \geq 2$, where $w_{i,l_{i-2}} \equiv w_{l_{i-2}}(\boldsymbol{s}_i)$, $\rho_{i,l_{i-2}} \equiv \rho_{l_{i-2}}(\boldsymbol{s}_i)$, $z_{i,l_{i-2}} \equiv z(\boldsymbol{s}_{(i,l_{i-2})})$. We denote by $\boldsymbol{z}_{1:k}$ the realization of $Z(\boldsymbol{s})$ over locations $(\boldsymbol{s}_1, \ldots, \boldsymbol{s}_k)^\top$ for $k \geq 1$, and use $\boldsymbol{z}_{1:k}^{-z_j}$ to denote the random vector $\boldsymbol{z}_{1:k}$ with element z_j removed, $1 \leq j \leq k$. In the following, for a vector $\boldsymbol{a} = (a_1, \ldots, a_m)^\top$, we have that $\boldsymbol{a}c = (a_1 c, \ldots, a_m c)^\top$, where c is a scalar.

Take $Z_1 \sim N(z_1 \,|\, 0, \sigma^2)$. The joint density of $\boldsymbol{z}_{1:2}$ is

$$\tilde{p}(\boldsymbol{z}_{1:2}) = N(z_2|\rho_{2,1}z_1), \sigma^2(1-\rho_{2,1}^2))N(z_1|0,\sigma^2) = N(\boldsymbol{z}_{1:2}|\boldsymbol{0}, \sigma^2\boldsymbol{\Omega}_{2,l_0})$$

where $\boldsymbol{\Omega}_{2,l_0} = \begin{pmatrix} 1 & \rho_{2,1} \\ \rho_{2,1} & 1 \end{pmatrix}$ with $\rho_{2,1} \equiv \rho_{2,l_0}$, i.e. $l_0 = 1$. It follows that $w_{2,l_0} = 1$.

The joint density of $\boldsymbol{z}_{1:3}$ is

$$\begin{aligned}
\tilde{p}(\boldsymbol{z}_{1:3}) &= p_3(z_3|\boldsymbol{z}_{1:2})\tilde{p}(\boldsymbol{z}_{1:2}) \\
&= \sum_{l_1=1}^{2} w_{3,l_1} N(z_3|\rho_{3,l_1}z_{3,l_1}, \sigma^2(1-\rho_{3,l_1}^2))N(\boldsymbol{z}_{1:2}|\boldsymbol{0}, \sigma^2\boldsymbol{\Omega}_{2,l_0}) \\
&= \sum_{l_1=1}^{2} w_{3,l_1} N(z_3|\rho_{3,l_1}z_{3,l_1}, \sigma^2(1-\rho_{3,l_1}^2))N(\boldsymbol{z}_{1:2}^{-z_{3,l_1}}|\rho_{2,l_0}z_{3,l_1}, \sigma^2(1-\rho_{2,l_0}^2))N(z_{3,l,1}|0,\sigma^2) \\
&= \sum_{l_1=1}^{2} w_{3,l_1} N((z_3, \boldsymbol{z}_{1:2}^{-z_{3,l1}})|\boldsymbol{m}_{3,l_1}z_{3,l_1}, \boldsymbol{V}_{3,l_1})N(z_{3,l_1}|0,\sigma^2)
\end{aligned}$$

where $\boldsymbol{m}_{3,l_1} = (\rho_{3,l_1}, \rho_{2,l_0})^\top$, and $\boldsymbol{V}_{3,l_1} = \begin{pmatrix} \sigma^2(1-\rho_{3,l_1}^2) & 0 \\ 0 & \sigma^2(1-\rho_{2,l_0}^2) \end{pmatrix}$. The last equality follows from the fact that a product of conditionally independent Gaussian densities is a Gaussian density.

Making use of the properties of the Gaussian distribution and the model property that has a stationary marginal $N(0, \sigma^2)$, for each l_1, we have that

$$N(\tilde{\boldsymbol{z}}_{1:3,l_1}|\boldsymbol{0}, \sigma^2\boldsymbol{R}_{3,l_1}) = N((z_3, \boldsymbol{z}_{1:2}^{-z_{3,l1}})|\boldsymbol{m}_{3,l_1}z_{3,l_1}, \boldsymbol{V}_{3,l_1})N(z_{3,l_1}|0,\sigma^2),$$

where $\tilde{\boldsymbol{z}}_{1:3,l_1} = (z_3, \boldsymbol{z}_{1:2}^{-z_{3,l_1}}, z_{3,l_1})^\top$, with the following partition relevant to the vector $\tilde{\boldsymbol{z}}_{1:3,l_1}$,

$$\tilde{\boldsymbol{z}}_{1:3,l_1} = \begin{pmatrix} (z_3, \boldsymbol{z}_{1:2}^{-z_{3,l_1}})^\top \\ z_{3,l_1} \end{pmatrix}, \; E(\tilde{\boldsymbol{z}}_{1:3,l_1}) = \begin{pmatrix} \boldsymbol{0} \\ 0 \end{pmatrix}, \; \boldsymbol{R}_{3,l_1} = \begin{pmatrix} \boldsymbol{R}_{3,l_1}^{(11)} & \boldsymbol{R}_{3,l_1}^{(12)} \\ \boldsymbol{R}_{3,l_1}^{(21)} & \boldsymbol{R}_{3,l_1}^{(22)} \end{pmatrix},$$

where $\boldsymbol{R}_{3,l_1}^{(22)} = 1$. It follows that

$$\boldsymbol{m}_{3,l_1}z_{3,l_1} = E((Z_3, \tilde{\boldsymbol{Z}}_{1:2}^{-Z_{3,l_1}}) \,|\, Z_{3,l_1} = z_{3,l_1}) = \boldsymbol{R}_{3,l_1}^{(12)}z_{3,l_1}, \;\; \boldsymbol{V}_{3,l_1} = \sigma^2(\boldsymbol{R}_{3,l_1}^{(11)} - \boldsymbol{R}_{3,l_1}^{(12)}\boldsymbol{R}_{3,l_1}^{(21)}). \tag{2}$$

From Equation (2), we obtain $\boldsymbol{m}_{3,l_1} = \boldsymbol{R}_{3,l_1}^{(12)}$ and $\boldsymbol{R}_{3,l_1} = \begin{pmatrix} 1 & \rho_{2,l_0}\rho_{3,l_1} & \rho_{3,l_1} \\ \rho_{2,l_0}\rho_{3,l_1} & 1 & \rho_{2,l_0} \\ \rho_{3,l_1} & \rho_{2,l_0} & 1 \end{pmatrix}$ for $l_1 =$

1, 2. Then we reorder $\tilde{\boldsymbol{z}}_{1:3,l_1}$ with a matrix $\boldsymbol{B}_{3,l_1}$ such that $\boldsymbol{z}_{1:3} = \boldsymbol{B}_{3,l_1}\tilde{\boldsymbol{z}}_{1:3,l_1}$. It follows that $\boldsymbol{\Omega}_{3,l_1} = \boldsymbol{B}_{3,l_1}\boldsymbol{R}_{3,l_1}\boldsymbol{B}_{3,l_1}^T$, and the joint density is

$$p(\boldsymbol{z}_{1:3}) = \sum_{l_1=1}^{2} w_{3,l_1} N(\boldsymbol{z}_{1:3}|\boldsymbol{0}, \sigma^2\boldsymbol{\Omega}_{3,l_1}) = \sum_{l_1=1}^{2} w_{3,l_1} w_{2,l_0} N(\boldsymbol{z}_{1:3}|\boldsymbol{0}, \sigma^2\boldsymbol{\Omega}_{3,l_1l_0}),$$

since $w_{2,l_0} = 1$ and $l_0 = 1$.

Similarly, the joint density of $\boldsymbol{z}_{1:4}$ is given by

$$\begin{aligned}
\tilde{p}(\boldsymbol{z}_{1:4}) &= p_4(z_4|\boldsymbol{z}_{1:3})\tilde{p}(\boldsymbol{z}_{1:3}) \\
&= \sum_{l_2=1}^{3} w_{4,l_2} N(z_4|\rho_{4,l_2} z_{4,l_2}, \sigma^2(1-\rho_{4,l_2}^2)) \sum_{l_1=1}^{2} w_{3,l_1} N(\boldsymbol{z}_{1:3}|\boldsymbol{0}, \sigma^2\boldsymbol{\Omega}_{3,l_1l_0}) \\
&= \sum_{l_2=1}^{3}\sum_{l_1=1}^{2} w_{4,l_2} w_{3,l_1} N(z_4|\rho_{4,l_2} z_{4,l_2}, \sigma^2(1-\rho_{4,l_2}^2)) \\
&\qquad\qquad N(\boldsymbol{z}_{1:3}^{-z_{4,l_2}} \mid \tilde{\boldsymbol{\Omega}}_{3,l_1l_0}^{(12)} z_{4,l_2}, \sigma^2(\tilde{\boldsymbol{\Omega}}_{3,l_1l_0}^{(11)} - \tilde{\boldsymbol{\Omega}}_{3,l_1l_0}^{(12)}\tilde{\boldsymbol{\Omega}}_{3,l_1l_0}^{(21)})) N(z_{4,l_2}|0,\sigma^2) \\
&= \sum_{l_2=1}^{3}\sum_{l_1=1}^{2} w_{4,l_2} w_{3,l_1} N((z_4, \boldsymbol{z}_{1:3}^{-z_{4,l_2}})|\boldsymbol{m}_{4,l_2l_1} z_{4,l_2}, \boldsymbol{V}_{4,l_2l_1}) N(z_{4-l_2}|0,\sigma^2).
\end{aligned}$$

Similarly, we have that for $l_1 = 1, 2,\ l_2 = 1, 2, 3$,

$$N(\tilde{\boldsymbol{z}}_{1:4,l_2}|\boldsymbol{0}, \sigma^2\boldsymbol{R}_{4,l_2l_1}) = N((z_4, \boldsymbol{z}_{1:3}^{-z_{4,l_2}})|\boldsymbol{m}_{4,l_2l_1} z_{4,l_2}, \boldsymbol{V}_{4,l_2l_1}) N(z_{4-l_2}|0,\sigma^2),$$

and

$$\begin{aligned}
&\tilde{\boldsymbol{\Omega}}_{3,l_1l_0} = \tilde{\boldsymbol{B}}_{4,l_2}\boldsymbol{\Omega}_{3,l_1l_0}\tilde{\boldsymbol{B}}_{4,l_2}^\top,\ \ \tilde{\boldsymbol{z}}_{1:4,l_2} = (z_4, \boldsymbol{z}_{1:3}^{-z_{4,l_2}}, z_{4,l_2})^\top,\ \ \boldsymbol{m}_{4,l_2l_1} = (\rho_{4,l_2}, (\tilde{\boldsymbol{\Omega}}_{3,l_1l_0}^{(12)})^\top)^\top, \\
&\boldsymbol{V}_{4,l_2l_1} = \begin{pmatrix} \sigma^2(1-\rho_{4,l_2}^2) & \boldsymbol{0}^T \\ \boldsymbol{0} & \sigma^2(\tilde{\boldsymbol{\Omega}}_{3,l_1l_0}^{(11)} - \tilde{\boldsymbol{\Omega}}_{3,l_1l_0}^{(12)}\tilde{\boldsymbol{\Omega}}_{3,l_1l_0}^{(21)}) \end{pmatrix}, \\
&\boldsymbol{R}_{4,l_2l_1}^{(12)} = (\boldsymbol{R}_{4,l_2l_1}^{(21)})^\top = \boldsymbol{m}_{4,l_2l_1},\ \ \boldsymbol{R}_{4,l_2l_1}^{(11)} = \boldsymbol{V}_{4,l_2l_1}/\sigma^2 + \boldsymbol{m}_{4,l_2l_1}\boldsymbol{m}_{4,l_2l_1}^T,
\end{aligned}$$

where $\tilde{\boldsymbol{B}}_{4,l_2}$ is a rotation matrix such that $((\boldsymbol{z}_{1:3}^{-z_{4,l_2}})^\top, z_{4,l_2})^\top = \tilde{\boldsymbol{B}}_{4,l_2}\boldsymbol{z}_{1:3}$, so that $\tilde{\boldsymbol{\Omega}}_{3,l_1l_0}^{(22)}$ corresponds to z_{4,l_2}. Then we reorder $\tilde{\boldsymbol{z}}_{1:4,l_2}$ with a matrix $\boldsymbol{B}_{4,l_2}$ such that $\boldsymbol{z}_{1:4} = \boldsymbol{B}_{4,l_2}\tilde{\boldsymbol{z}}_{1:4,l_2}$.

It follows that $\boldsymbol{\Omega}_{4,l_2l_1l_0} = \boldsymbol{B}_{4,l_2}\boldsymbol{R}_{4,l_2l_1}\boldsymbol{B}_{4,l_1}^T$. Then we can obtain the joint density for $\boldsymbol{z}_{1:4}$,

$$\tilde{p}(\boldsymbol{z}_{1:4}) = \sum_{l_2=1}^{3}\sum_{l_1=1}^{2} w_{4,l_2}w_{3,l_1}w_{2,l_0}N(\boldsymbol{z}_{1:4}|\boldsymbol{0}, \sigma^2\boldsymbol{\Omega}_{4,l_2l_1l_0}).$$

Applying the above technique iteratively for $\tilde{p}(\boldsymbol{z}_{1:j})$ for $5 \le j \le k$, we obtain the joint density $\tilde{p}(\boldsymbol{z}_{1:k}) \equiv \tilde{p}(\boldsymbol{z}_{\mathcal{S}})$, for $k \ge 2$, namely,

$$\tilde{p}(\boldsymbol{z}_{1:k}) = \sum_{l_{k-2}=1}^{k_L} \cdots \sum_{l_1=1}^{2} w_{k,l_{k-2}} \dots w_{3,l_1}w_{2,l_0}N(\boldsymbol{z}_{1:k}|\boldsymbol{0}, \sigma^2\boldsymbol{\Omega}_{k,l_{k-2}\dots l_1l_0})$$

where $k_L := (k-1) \wedge L$, $w_{2,l_0} = 1$, $\Omega_{2,l_0} = \begin{pmatrix} 1 & \rho_{2,l_0} \\ \rho_{2,l_0} & 1 \end{pmatrix}$, and for $k \ge 3$,

$$\begin{aligned}
\tilde{\boldsymbol{\Omega}}_{k-1,l_{k-3}\dots l_1l_0} &= \tilde{\boldsymbol{B}}_{k,l_{k-2}}\boldsymbol{\Omega}_{k-1,l_{k-3}\dots l_1l_0}\tilde{\boldsymbol{B}}_{k,l_{k-2}}, \quad \boldsymbol{m}_{k,l_{k-2}\dots l_1} = (\rho_{k,l_{k-2}}, (\tilde{\boldsymbol{\Omega}}^{(12)}_{k-1,l_{k-3}\dots l_1l_0})^\top)^\top, \\
\boldsymbol{V}_{k,l_{k-2}\dots l_1} &= \begin{pmatrix} \sigma^2(1-\rho^2_{k,l_{k-2}}) & \boldsymbol{0} \\ \boldsymbol{0}^T & \sigma^2(\tilde{\boldsymbol{\Omega}}^{(11)}_{k-1,l_{k-3}\dots l_1l_0} - \tilde{\boldsymbol{\Omega}}^{(12)}_{k-1,l_{k-3}\dots l_1l_0}\tilde{\boldsymbol{\Omega}}^{(21)}_{k-1,l_{k-3}\dots l_1l_0}) \end{pmatrix}, \\
\boldsymbol{R}^{(12)}_{k,l_{k-2}\dots l_1} &= (\boldsymbol{R}^{(21)}_{k,l_{k-2}\dots l_1})^\top = \boldsymbol{m}_{k,l_{k-2}\dots l_1}, \quad \boldsymbol{R}^{(11)}_{k,l_{k-2}\dots l_1} = \boldsymbol{V}_{k,l_{k-2}\dots l_1}/\sigma^2 + \boldsymbol{m}_{k,l_{k-2}\dots l_1}\boldsymbol{m}^\top_{k,l_{k-2}\dots l_1}, \\
\boldsymbol{\Omega}_{k,l_{k-2}\dots l_1l_0} &= \boldsymbol{B}_{k,l_{k-2}}\boldsymbol{R}_{k,l_{k-2}\dots l_1}\boldsymbol{B}^T_{k,l_{k-2}},
\end{aligned}$$

where $\tilde{\boldsymbol{B}}_{k,l_{k-2}}$ is the rotation matrix such that $((\boldsymbol{z}^{-z_{k,l_k}}_{1:(k-1)})^\top, z_{k,l_k})^\top = \tilde{\boldsymbol{B}}_{k,l_{k-2}}\boldsymbol{z}_{1:(k-1)}$, and $\boldsymbol{B}_{k,l_{k-2}}$ is the rotation matrix such that the vector $\boldsymbol{z}_{1:k} = \boldsymbol{B}_{k,l_{k-2}}\tilde{\boldsymbol{z}}_{1:k,l_{k-2}}$, where $\tilde{\boldsymbol{z}}_{1:k,l_{k-2}} = (z_k, (\boldsymbol{z}^{-z_{k,l_k}}_{1:(k-1)})^\top, z_{k,l_k})^\top$.

To complete the proof, what remains to show is that the density $\tilde{p}(\boldsymbol{z}_{\mathcal{U}_1})$ is a mixture of multivariate Gaussian distributions, where $\mathcal{U}_1 = \mathcal{S} \cup \{\boldsymbol{u}\}$. We have that

$$\tilde{p}(\boldsymbol{z}_{\mathcal{U}_1}) = \sum_{l=1}^{L} w_l N(z(\boldsymbol{u}) \mid \rho_{\boldsymbol{u},l}z(\boldsymbol{u}_{(l)}), \sigma^2(1-\rho^2_{\boldsymbol{u},l}))\tilde{p}(\boldsymbol{z}_{1:k}),$$

where $z(\boldsymbol{u}_{(l)})$ is an element of $\boldsymbol{z}_{1:k}$ for $l = 1, \dots, L$. We can express each component density $N(\boldsymbol{z}_{1:k}|\boldsymbol{0}, \sigma^2\Omega_{k,l_{k-2}\dots l_1l_0})$ of the joint density $\tilde{p}(\boldsymbol{z}_{1:k})$ as the product of a Gaussian density of $\boldsymbol{Z}^{-Z(\boldsymbol{u}_{(l)})}_{1:k}$ conditional on $Z(\boldsymbol{u}_{(l)})$ and a Gaussian density of $Z(\boldsymbol{u}_{(l)})$. Using the technique in deriving the joint density over $\mathcal{S}$, we can show that $\tilde{p}(\boldsymbol{z}_{\mathcal{U}_1})$ is a mixture of multivariate Gaussian distribution. □

Proof of Proposition 4. For an NNMP $Z(\boldsymbol{v})$, the conditional probability that $Z(\boldsymbol{v})$ is greater than z given its neighbors $\boldsymbol{Z}_{\mathrm{Ne}(\boldsymbol{v})} = \boldsymbol{z}_{\mathrm{Ne}(\boldsymbol{v})}$, where $\boldsymbol{z}_{\mathrm{Ne}(\boldsymbol{v})} = (z_{\boldsymbol{v}_{(1)}}, \ldots, z_{\boldsymbol{v}_{(L)}})$, is

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

where the conditional probability $P(Z(\boldsymbol{v}) > z \,|\, Z(\boldsymbol{v}_{(l)}) = z(\boldsymbol{v}_{(l)}))$ corresponds to the bivariate random vector (U_l, V_l). If U_l is stochastically increasing in V_l for all l, by the assumption that the NNMP model is built from base random vectors (U_l, V_l), we have that $Z(\boldsymbol{v})$ is stochastically increasing in $\boldsymbol{Z}_{\mathrm{Ne}(\boldsymbol{v})}$ for every $\boldsymbol{v} \in \mathcal{D}$, i.e.

$$p(Z(\boldsymbol{v}) > z \,|\, \boldsymbol{Z}_{\mathrm{Ne}(\boldsymbol{v})} = \boldsymbol{z}_{\mathrm{Ne}(\boldsymbol{v})}) \;\leq\; p(Z(\boldsymbol{v}) > z \,|\, \boldsymbol{Z}_{\mathrm{Ne}(\boldsymbol{v})} = \boldsymbol{z}'_{\mathrm{Ne}(\boldsymbol{v})})$$

for all $\boldsymbol{z}_{\mathrm{Ne}(\boldsymbol{v})}$ and $\boldsymbol{z}'_{\mathrm{Ne}(\boldsymbol{v})}$ in $\mathbb{R}^L$, such that $z_{\boldsymbol{v}_{(l)}} \leq z'_{\boldsymbol{v}_{(l)}}$ for all l.

Let $F_{Z(\boldsymbol{v})}$ and $F_{Z(\boldsymbol{v}_{(1)}),\ldots,Z(\boldsymbol{v}_{(L)})}$ be the distribution functions of $Z(\boldsymbol{v})$ and $\boldsymbol{Z}_{\mathrm{Ne}(\boldsymbol{v})}$, respectively. Denote by $S_{Z(\boldsymbol{v}_{(1)}),\ldots,Z(\boldsymbol{v}_{(L)})}(z_1, \ldots, z_L) = P(Z(\boldsymbol{v}_{(1)}) > z_1, \ldots, Z(\boldsymbol{v}_{(L)}) > z_L)$ the joint survival probability. Then for every $\boldsymbol{v} \in \mathcal{D}$ and $q \in (0, 1)$,

$$\begin{aligned}
&P(Z(\boldsymbol{v}) > F^{-1}_{Z(\boldsymbol{v})}(q) \mid Z(\boldsymbol{v}_{(1)}) > F^{-1}_{Z(\boldsymbol{v}_{(1)})}(q), \ldots, Z(\boldsymbol{v}_{(L)}) > F^{-1}_{Z(\boldsymbol{v}_{(L)})}(q)) \\
&= \Bigg\{ \int_{F^{-1}_{Z(\boldsymbol{v}_{(1)})}(q)}^{\infty} \cdots \int_{F^{-1}_{Z(\boldsymbol{v}_{(L)})}(q)}^{\infty} P(Z(\boldsymbol{v}) > F^{-1}_{Z(\boldsymbol{v})}(q) \mid Z(\boldsymbol{v}_{(1)}) = z_1, \ldots, Z(\boldsymbol{v}_{(L)}) = z_L) \\
&\quad dF_{Z(\boldsymbol{v}_{(1)}),\ldots,Z(\boldsymbol{v}_{(L)})}(z_1, \ldots, z_L) \Bigg\} \Big/ S_{Z(\boldsymbol{v}_{(1)}),\ldots,Z(\boldsymbol{v}_{(L)})}(F^{-1}_{Z(\boldsymbol{v}_{(1)})}(q), \ldots, F^{-1}_{Z(\boldsymbol{v}_{(L)})}(q)) \\
&\geq \Bigg\{ \int_{F^{-1}_{Z(\boldsymbol{v}_{(1)})}(q)}^{\infty} \cdots \int_{F^{-1}_{Z(\boldsymbol{v}_{(L)})}(q)}^{\infty} P(Z(\boldsymbol{v}) > F^{-1}_{Z(\boldsymbol{v})}(q) \mid Z(\boldsymbol{v}_{(1)}) = F^{-1}_{Z(\boldsymbol{v}_{(1)})}(q), \ldots, Z(\boldsymbol{v}_{(L)}) = F^{-1}_{Z(\boldsymbol{v}_{(L)})}(q)) \\
&\quad dF_{Z(\boldsymbol{v}_{(1)}),\ldots,Z(\boldsymbol{v}_{(L)})}(z_1, \ldots, z_L) \Bigg\} \Big/ S_{Z(\boldsymbol{v}_{(1)}),\ldots,Z(\boldsymbol{v}_{(L)})}(F^{-1}_{Z(\boldsymbol{v}_{(1)})}(q), \ldots, F^{-1}_{Z(\boldsymbol{v}_{(L)})}(q)) \\
&= P(Z(\boldsymbol{v}) > F^{-1}_{Z(\boldsymbol{v})}(q) \mid Z(\boldsymbol{v}_{(1)}) = F^{-1}_{Z(\boldsymbol{v}_{(1)})}(q), \ldots, Z(\boldsymbol{v}_{(L)}) = F^{-1}_{Z(\boldsymbol{v}_{(L)})}(q)) \\
&= \sum_{l=1}^{L} w_l(\boldsymbol{v}) P(Z(\boldsymbol{v}) > F^{-1}_{U_{\boldsymbol{v},l}}(q) \mid Z(\boldsymbol{v}_{(l)}) = F^{-1}_{V_{\boldsymbol{v},l}}(q)),
\end{aligned} \tag{3}$$

where the first inequality follows from the stochastically increasing positive dependence of $Z(\boldsymbol{v})$ given $\boldsymbol{Z}_{\mathrm{Ne}(\boldsymbol{v})}$.

Taking $q \to 1^-$ on both sides of (3), we obtain

$$\lambda_{\mathcal{H}}(\boldsymbol{v}) \geq \sum_{l=1}^{L} w_l(\boldsymbol{v}) \lim_{q \to 1^-} P(Z(\boldsymbol{v}) > F^{-1}_{U_{\boldsymbol{v},l}}(q) \mid Z(\boldsymbol{v}_{(l)}) = F^{-1}_{V_{\boldsymbol{v},l}}(q)).$$

Similarly, we can obtain the lower bound for $\lambda_{\mathcal{L}}(\boldsymbol{v})$. □

Proof of Corollary 1. We prove the result for $\lambda_{\mathcal{L}}(\boldsymbol{v})$. The result for $\lambda_{\mathcal{H}}(\boldsymbol{v})$ is obtained in a similar way.

Consider a bivariate distribution F_{U_l,V_l} for random vector (U_l, V_l), with marginal distributions $F_{U_l} = F_{V_l} = F_l$ and marginal densities $f_{U_l} = f_{V_l} = f_l$, for all l. The lower tail dependence coefficient is expressed as $\lambda_{\mathcal{L},l} = \lim_{q \to 0^+} \frac{F_{U_l,V_l}(F_l^{-1}(q), F_l^{-1}(q))}{F_{U_l,V_l}(F_l^{-1}(q))}$ with $q \in [0,1]$. If F_{U_l,V_l} has first order partial derivatives, applying the L'Hopital's rule, we obtain

$$\begin{aligned}
\lambda_{\mathcal{L},l} &= \lim_{q \to 0^+} \frac{\partial F_{U_l,V_l}/\partial V_l(F_l^{-1}(q), F_l^{-1}(q)) + \partial F_{U_l,V_l}/\partial U_l(F_l^{-1}(q), F_l^{-1}(q))}{f_l(F_l^{-1}(q))} \\
&= \lim_{q \to 0^+} P(U_l \leq F_l^{-1}(q) \mid V_l = F_l^{-1}(q)) + \lim_{q \to 0^+} P(V_l \leq F_l^{-1}(q) \mid U_l = F_l^{-1}(q)).
\end{aligned}$$

The above is a reproduced result from Theorem 8.57 of Joe (2014). If (U_l, V_l) is exchangeable, we have

$$\lambda_{\mathcal{L},l} = 2 \lim_{q \to 0^+} P(U_l \leq F_l^{-1}(q) \,|\, V_l = F_l^{-1}(q)).$$

If the sequences $(U_{\boldsymbol{v},l}, V_{\boldsymbol{v},l})$ of an NNMP model are built from the base random vectors (U_l, V_l). By our assumption on (U_l, V_l), we have the marginal distributions of $(U_{\boldsymbol{v},l}, V_{\boldsymbol{v},l})$ such that $F_{\boldsymbol{v},l} = F_{U_{\boldsymbol{v},l}} = F_{V_{\boldsymbol{v},l}}$. Then we have

$$\lambda_{\mathcal{L},l}(\boldsymbol{v}) = 2 \lim_{q \to 0^+} P(U_{\boldsymbol{v},l} \leq F^{-1}_{\boldsymbol{v},l}(q) \,|\, V_{\boldsymbol{v},l} = F^{-1}_{\boldsymbol{v},l}(q)).$$

Using the result of Proposition 4, we obtain

$$\lambda_{\mathcal{L}}(\boldsymbol{v}) \geq \sum_{l=1}^{L} w_l(\boldsymbol{v}) \lim_{q \to 0^+} P(U_{\boldsymbol{v},l} \leq F^{-1}_{\boldsymbol{v},l}(q) \mid V_{\boldsymbol{v},l} = F^{-1}_{\boldsymbol{v},l}(q)) = \sum_{l=1}^{L} w_l(\boldsymbol{v}) \lambda_{\mathcal{L},l}(\boldsymbol{v})/2.$$

□

Proof of Proposition 5. By the assumption that U_l is stochastically increasing in V_l

and that $(U_{\boldsymbol{v},l}, V_{\boldsymbol{v},l})$ is constructed based on (U_l, V_l), $U_{\boldsymbol{v},l}$ is stochastically increasing in $V_{\boldsymbol{v},l}$ for all $\boldsymbol{v} \in \mathcal{D}$ and for all l. Then for $Z(\boldsymbol{v})$ with respect to the bivariate distribution of $(U_{\boldsymbol{v},l}, V_{\boldsymbol{v},l})$ with marginal distributions $F_{U_{\boldsymbol{v},l}}$ and $F_{V_{\boldsymbol{v},l}}$, we have that

$$
\begin{aligned}
&P(Z(\boldsymbol{v}) \le F^{-1}_{U_{\boldsymbol{v},l}}(q) \mid Z(\boldsymbol{v}_{(l)}) \le F^{-1}_{V_{\boldsymbol{v},l}}(q)) \\
&= \int_{F^{-1}_{V_{\boldsymbol{v},l}}(0)}^{F^{-1}_{V_{\boldsymbol{v},l}}(q)} P(Z(\boldsymbol{v}) \le F^{-1}_{U_{\boldsymbol{v},l}}(q) \mid Z(\boldsymbol{v}_{(l)}) = z_l) dF_{V_{\boldsymbol{v},l}}(z_l) \Big/ \int_{F^{-1}_{V_{\boldsymbol{v},l}}(0)}^{F^{-1}_{V_{\boldsymbol{v},l}}(q)} dF_{V_{\boldsymbol{v},l}} \\
&\le \int_{F^{-1}_{V_{\boldsymbol{v},l}}(0)}^{F^{-1}_{V_{\boldsymbol{v},l}}(q)} P(Z(\boldsymbol{v}) \le F^{-1}_{U_{\boldsymbol{v},l}}(q) \mid Z(\boldsymbol{v}_{(l)}) = F^{-1}_{V_{\boldsymbol{v},l}}(0)) dF_{V_{\boldsymbol{v},l}}(z_l) \Big/ \int_{F^{-1}_{V_{\boldsymbol{v},l}}(0)}^{F^{-1}_{V_{\boldsymbol{v},l}}(q)} dF_{V_{\boldsymbol{v},l}} \\
&= P(Z(\boldsymbol{v}) \le F^{-1}_{U_{\boldsymbol{v},l}}(q) \mid Z(\boldsymbol{v}_{(l)}) = F^{-1}_{V_{\boldsymbol{v},l}}(0)).
\end{aligned}
$$

It follows that the boundary cdf of the NNMP model

$$
\begin{aligned}
&F_{1|2}(F^{-1}_{Z(\boldsymbol{v})}(q) \mid F^{-1}_{\boldsymbol{Z}_{\mathrm{Ne}(\boldsymbol{v})}}(0)) \\
&= P(Z(\boldsymbol{v}) \le F^{-1}_{Z(\boldsymbol{v})}(q) \mid Z(\boldsymbol{v}_{(1)}) = F^{-1}_{Z(\boldsymbol{v}_{(1)})}(0), \ldots, Z(\boldsymbol{v}_{(L)}) = F^{-1}_{Z(\boldsymbol{v}_{(L)})}(0)) \\
&= \sum_{l=1}^{L} w_l(\boldsymbol{v}) P(Z(\boldsymbol{v}) \le F^{-1}_{U_{\boldsymbol{v},l}}(q) \mid Z(\boldsymbol{v}_{(l)}) = F^{-1}_{V_{\boldsymbol{v},l}}(0)) \qquad (4) \\
&\ge \sum_{l=1}^{L} w_l(\boldsymbol{v}) P(Z(\boldsymbol{v}) \le F^{-1}_{U_{\boldsymbol{v},l}}(q) \mid Z(\boldsymbol{v}_{(l)}) \le F^{-1}_{V_{\boldsymbol{v},l}}(q)),
\end{aligned}
$$

Taking $q \to 0^+$ on both sides of (4), we obtain

$$
F_{1|2}(F^{-1}_{Z(\boldsymbol{v})}(0) \,|\, F^{-1}_{\boldsymbol{Z}_{\mathrm{Ne}(\boldsymbol{v})}}(0)) \ge \sum_{l=1}^{L} w_l(\boldsymbol{v}) \lambda_{\mathcal{L},l}(\boldsymbol{v}).
$$

Hence, if there exists some l such that $\lambda_{\mathcal{L},l}(\boldsymbol{v}) > 0$, the conditional cdf $F_{1|2}(F^{-1}_{Z(\boldsymbol{v})}(q) \,|\, F^{-1}_{\boldsymbol{Z}_{\mathrm{Ne}(\boldsymbol{v})}}(0))$ has strictly positive mass at $q = 0$. We can prove the result for $F_{1|2}(F^{-1}_{Z(\boldsymbol{v})}(q) \mid F^{-1}_{\boldsymbol{Z}_{\mathrm{Ne}(\boldsymbol{v})}}(1))$ in a similar way. □

B Model details and MCMC implementation

In this section, we provide details of the models implemented in the data examples of the paper. Based on the Gibbs sampler described in the paper, we focus on the posterior

updates of the parameters associated with the component densities of the mixture.

B.1 GNNMP models

We consider the spatial regression model on observations

$$y(\boldsymbol{s}_i) = \boldsymbol{x}(\boldsymbol{s}_i)^\top \boldsymbol{\beta} + z(\boldsymbol{s}_i) + \epsilon(\boldsymbol{s}_i), \;\; i = 1, \ldots, n,$$

where $\epsilon(\boldsymbol{s}_i) \overset{i.i.d.}{\sim} N(0, \tau^2)$, and the spatial random effect $z(\boldsymbol{s}_i)$ follows a GNNMP model, namely, $p_1(z(\boldsymbol{s}_1) = N(z(\boldsymbol{s}_1) \,|\, 0, \sigma^2)$, and for $i = 2, \ldots, n$,

$$p(z(\boldsymbol{s}_i) \,|\, \boldsymbol{z}_{\mathrm{Ne}(\boldsymbol{s}_i)}) = \sum_{l=1}^{i_L} w_l(\boldsymbol{s}_i) N(z(\boldsymbol{s}_i) \,|\, \rho_l(\boldsymbol{s}_i) z(\boldsymbol{s}_{(il)}), \sigma^2(1 - \rho_l(\boldsymbol{s}_i)^2)),$$

where $i_L = L \wedge (i-1)$, $\rho_l(\boldsymbol{s}_i) \equiv \rho_l(||\boldsymbol{s}_i - \boldsymbol{s}_{(il)}||) = \exp(-||\boldsymbol{s}_i - \boldsymbol{s}_{(il)}||/\phi)$.

Hierarchical representation of the model can written in the form of (12) of the paper. For the weights, we consider an exponential correlation function with range parameter ζ for the kernel function that defines the random cutoff points. The posterior updates of the weight-relevant parameters can be found in Section 3.2 of the paper. We provide the posterior updates for the component density parameters σ^2, ϕ, and the spatial random effects $z(\boldsymbol{s}_i)$, $i = 1, \ldots, n$. We assign inverse gamma priors $\mathrm{IG}(\sigma^2 \,|\, u_{\sigma^2}, v_{\sigma^2})$ and $\mathrm{IG}(\phi \,|\, u_\phi, v_\phi)$ to σ^2 and ϕ, respectively.

The posterior full conditional distribution of σ^2 is $\mathrm{IG}(\sigma^2 \,|\, u_{\sigma^2} + n/2, v_{\sigma^2} + \sum_{i=1}^n (z(\boldsymbol{s}_i) - \rho_{\ell_i}(\boldsymbol{s}_i) z(\boldsymbol{s}_{(i,\ell_i)}))^2/(2(1 - (\rho_{\ell_i}(\boldsymbol{s}_i))^2)))$. The posterior full conditional distribution of ϕ is proportional to $\mathrm{IG}(\phi \,|\, u_\phi, v_\phi) \prod_{i=2}^n N(z(\boldsymbol{s}_i) \mid \rho_{\ell_i}(\boldsymbol{s}_i) z(\boldsymbol{s}_{(i,\ell_i)}), \sigma_l^2(1 - (\rho_{\ell_i}(\boldsymbol{s}_i))^2))$. We update ϕ on its log scale with a random walk Metropolis step. Denote by $\boldsymbol{A}_j^{(i)} = \{j : z(\boldsymbol{s}_{(j,\ell_j)}) = z(\boldsymbol{s}_i)\}$, and assume $\rho_l(\boldsymbol{s}_1) = 0, z(\boldsymbol{s}_{(1,l)}) = 0$ for every l. The posterior full conditional of the latent spatial random effects $z(\boldsymbol{s}_i)$ is $N(z(\boldsymbol{s}_i) \,|\, \tilde{\sigma}_i^2 \tilde{\mu}_i, \tilde{\sigma}_i^2)$ where $\tilde{\sigma}_i^2 = (\tau^{-2} + \sigma^{-2}(1 - (\rho_{\ell_i}(\boldsymbol{s}_i))^2)^{-1} + \sum_{j: j \in \boldsymbol{A}_j^{(i)}} \tilde{s}_{ij}^{-2})^{-1}$ and $\tilde{\mu}_i = \tau^{-2}(y(\boldsymbol{s}_i) - x(\boldsymbol{s}_i)^\top \boldsymbol{\beta}) + \sigma^{-2}(1 - (\rho_{\ell_i}(\boldsymbol{s}_i))^2)^{-1} \rho_{\ell_i}(\boldsymbol{s}_i) z(\boldsymbol{s}_{(i,\ell_i)}) + \sum_{j: j \in \boldsymbol{A}_j^{(i)}} z(\boldsymbol{s}_j)(\rho_{\ell_j}(\boldsymbol{s}_j))^{-1} \tilde{s}_{ij}^{-2}$ with $\tilde{s}_{ij}^2 = \sigma^2(1 - (\rho_{\ell_j}(\boldsymbol{s}_j))^2)/(\rho_{\ell_j}(\boldsymbol{s}_j))^2$, for $i = 1, \ldots, n$.

B.2 Skew-GNNMP models

Consider a location mixture of bivariate Gaussian distribution for $\boldsymbol{Z} = (U, V)$, namely,

$$N(\boldsymbol{z} \mid \lambda z_0 \mathbf{1}_2, \sigma^2 \boldsymbol{R}), \;\; z_0 \sim N(z_0 \mid 0, 1) I(z_0 \geq 0), \tag{5}$$

where $\boldsymbol{R} = \left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)$. Marginalizing out z_0, we obtain the joint density of $\boldsymbol{Z}$,

$$f(\boldsymbol{z}) = 2N(\boldsymbol{z} \mid \mathbf{0}, \boldsymbol{\Sigma}) \Phi(\lambda (1 - \lambda^2 \mathbf{1}_2^\top \boldsymbol{\Sigma}^{-1} \mathbf{1}_2)^{-1/2} \mathbf{1}_2^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{z}),$$

where $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{R} + \lambda^2 \mathbf{1}_2 \mathbf{1}_2^\top$. The marginal distribution of $\boldsymbol{Z}$ is a skew-Gaussian distribution, denoted as $\mathrm{SN}(x \,|\, 0, \omega^2, \alpha)$ with $\omega^2 = \lambda^2 + \sigma^2$ and $\alpha = \lambda/\sigma$, with density $f(x) = 2N(x \,|\, 0, \omega^2)\Phi(\alpha x/\omega)$, and $E(X) = \sqrt{2/\pi}\lambda$, where Φ is the cdf of a standard Gaussian distribution. The conditional density of U given $V = v$ is then given by

$$f_{U|V}(u \,|\, v) = N(u \,|\, \tilde{\rho} v, \omega^2(1 - \tilde{\rho}^2)) \Phi(\alpha'(u + v)/\omega') / \Phi(\alpha v/\omega), \tag{6}$$

where $\tilde{\rho} = (\rho\sigma^2 + \lambda^2)/(\sigma^2 + \lambda^2)$, $\alpha' = \lambda/s$, $\omega'^2 = s^2 + 2\lambda^2$ and $s^2 = (1 + \rho)\sigma^2$.

The skew-GNNMP is obtained by using (6) as the component $f_{\boldsymbol{v},l}$ of the mixture, and making the correlation ρ spatially varying with an exponential correlation function such that $\rho_l(\boldsymbol{v}) \equiv \rho_l(||\boldsymbol{v} - \boldsymbol{v}_{(l)}||) = \exp(-||\boldsymbol{v} - \boldsymbol{v}_{(l)}||/\phi)$, for $l = 1, \ldots, L$.

To facilitate computation, we use conditional likelihood inference. That is, based on observations $y(\boldsymbol{s}_i), i = 1, \ldots, L$, we have that

$$\begin{aligned} y(\boldsymbol{s}_i) \mid \boldsymbol{y}_{\mathrm{Ne}(\boldsymbol{s}_i)} &\overset{ind.}{\sim} \sum_{l=1}^{L} b_{\boldsymbol{s}_i,l} N(y(\boldsymbol{s}_i) \mid \tilde{\rho}_l(\boldsymbol{s}_i) y(\boldsymbol{s}_{(il)}), \omega^2(1 - (\tilde{\rho}_l(\boldsymbol{s}_i))^2)) 1_{(r^*_{\boldsymbol{s}_i,l-1}, r^*_{\boldsymbol{s}_i,l})}(t_i), \\ t_i \mid \boldsymbol{\gamma}, \kappa^2 &\overset{ind.}{\sim} N(t_i \mid \gamma_0 + \gamma_1 s_{i1} + \gamma_2 s_{i2}, \kappa^2), \end{aligned}$$

for $i = L + 1, \ldots, n$, where $b_{\boldsymbol{s}_i,l} = \Phi(\alpha'_{\boldsymbol{s}_i,l}(y(\boldsymbol{s}_i) + y(\boldsymbol{s}_{(il)}))/\omega'_{\boldsymbol{s}_i,l}) / \Phi(\lambda y(\boldsymbol{s}_{(il)})/(\sigma\sqrt{\sigma^2 + \lambda^2}))$, $\tilde{\rho}_l(\boldsymbol{s}_i) = (\rho_l(\boldsymbol{s}_i)\sigma^2 + \lambda^2)/(\sigma^2 + \lambda^2)$, $\alpha'_{\boldsymbol{s}_i,l} = \lambda/\sqrt{(1 + \rho_l(\boldsymbol{s}_i))\sigma^2}$, $\omega'^2_{\boldsymbol{s}_i,l} = 1 + (\rho_l(\boldsymbol{s}_i))^2\sigma^2 + 2\lambda^2$, and s_{i1}, s_{i2} are the first and the second coordinates of $\boldsymbol{s}_i$. The transformed cutoff points $r^*_{\boldsymbol{s}_i,l} = \log(r_{\boldsymbol{s}_i,l}/(1 - r_{\boldsymbol{s}_i,l}))$ where $r_{\boldsymbol{s}_i,l}$ is specified using an exponential correlation function $\exp(-||\boldsymbol{s}_i - \boldsymbol{s}_{(il)}||/\zeta)$. We complete the Bayesian model with prior specifications for $\lambda, \sigma^2, \phi, \zeta, \boldsymbol{\gamma}, \kappa^2$. In particular, we consider priors $N(\lambda \,|\, \mu_\lambda, \sigma^2_\lambda)$, $\mathrm{IG}(\sigma^2 \,|\, u_{\sigma^2}, v_{\sigma^2})$,

$\mathrm{IG}(\phi \,|\, u_\phi, v_\phi)$, $\mathrm{IG}(\zeta \,|\, u_\zeta, v_\zeta)$, $N(\boldsymbol{\gamma} \,|\, \boldsymbol{\mu}_\gamma, \boldsymbol{V}_\gamma)$ and $\mathrm{IG}(\kappa^2 \,|\, u_{\kappa^2}, v_{\kappa^2})$.

The posterior updates of the weight-relevant parameters are provided in Section 3.2 of the paper. We present the posterior updates of λ, σ^2 and ϕ. These updates are facilitated with a set of configuration variables ℓ_i such that $\ell_i = l$ if $t_i \in (r^*_{\boldsymbol{s}_i,l-1}, r^*_{\boldsymbol{s}_i,l})$ for $i \geq L+1$. Denote by $f_{\boldsymbol{s}_i,l} = b_{\boldsymbol{s}_i,l} N(y(\boldsymbol{s}_i)\tilde{\rho}_l(\boldsymbol{s}_i) y(\boldsymbol{s}_{(il)}), \omega^2(1-(\tilde{\rho}_l(\boldsymbol{s}_i))^2))$. We use a random walk Metropolis step to update λ with target density $N(\lambda \,|\, \mu_\lambda, \sigma^2_\lambda) \prod_{i=L+1}^n f_{\boldsymbol{s}_i,\ell_i}$. The posterior full conditional distributions of σ^2 and ϕ are proportional to $\mathrm{IG}(\sigma^2 \,|\, u_{\sigma^2}, v_{\sigma^2}) \prod_{i=L+1}^n f_{\boldsymbol{s}_i,\ell_i}$, and $\mathrm{IG}(\phi \,|\, u_\phi, v_\phi) \prod_{i=L+1}^n f_{\boldsymbol{s}_i,\ell_i}$, respectively. For each parameter, we update it on its log scale with a random walk Metropolis step.

Turning to the prediction. We generate predictions based on the skew-Gaussian's location mixture representation in (5). For a location $\boldsymbol{v}_0 \in \mathcal{D}$ outside the reference set, the predictive distribution $p(y(\boldsymbol{v}_0) \,|\, \boldsymbol{y}_{\mathrm{Ne}(\boldsymbol{v}_0)})$ is

$$\sum_{l=1}^{L} w_l(\boldsymbol{v}_0) \int N(y(\boldsymbol{v}_0) \,|\, (1-\rho_l(\boldsymbol{v}_0))\lambda z_0 + \rho_l(\boldsymbol{v}_0) y(\boldsymbol{v}_{(0l)}), \sigma^2(1-(\rho_l(\boldsymbol{v}_0))^2)) \mathrm{TN}(z_0 \,|\, \mu_{0l}, \tau_0^2) dz_0,$$

with $\mu_{0l} = \lambda y(\boldsymbol{v}_{(0l)})/(\lambda^2+\sigma^2)$ and $\tau_0^2 = \sigma^2/(\lambda^2+\sigma^2)$. We denote $\mathrm{TN}(\cdot \,|\, \mu_0, \tau_0^2)$ as the Gaussian distribution truncated at $[0, \infty)$, with mean μ_0 and variance τ_0^2. If $\boldsymbol{v}_0 \equiv \boldsymbol{s}_i$ is in the reference set, we replace L with i_L in the above predictive distribution.

B.3 Copula NNMP models

B.3.1 The Gaussian and Gumbel copula

We consider a bivariate vector (X_1, X_2) with marginal distributions F_1 and F_2 such that $F_1(x_1) = t_1$ and $F_2(x_2) = t_2$. We introduce basic properties of the Gaussian and Gumbel copulas. For more details we refer to Joe (2014).

Gaussian copula A Gaussian copula with correlation $\rho \in (0,1)$ for (X_1, X_2) is $C(t_1, t_2 \,|\, \rho) = \Phi(\Phi^{-1}(t_1) + \Phi^{-1}(t_2))$. The copula is asymptotically independent in both the lower and the upper tails. The corresponding copula density is given by:

$$\frac{1}{\sqrt{1-\rho^2}} \exp\left(\frac{2\rho\Phi^{-1}(t_1)\Phi^{-1}(t_2) - \rho^2(\Phi^{-1}(t_1)^2 + \Phi^{-1}(t_2)^2)}{2(1-\rho^2)} \right). \tag{7}$$

We obtain the spatial varying Gaussian copula $C_{\boldsymbol{v},l}$ by replacing the correlation parameter ρ in (7) with $\rho_l(\boldsymbol{v}) \equiv \rho_l(||\boldsymbol{v} - \boldsymbol{v}_{(l)}||) = \exp(-||\boldsymbol{v} - \boldsymbol{v}_{(l)}||/\phi)$.

Denote by $C_{1|2}(t_1 \,|\, t_2)$ the conditional cdf of T_1 given $T_2 = t_2$. Then we have

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

We sample X_1, given $X_2 = x_2$, with the following steps. Given a realization x_2 of X_2, we compute $t_2 = F_2(x_2)$. We then generate a random number z from a uniform distribution on $[0, 1]$, and compute $t_1 = C^{-1}_{1|2}(z \,|\, t_2)$ where $C^{-1}_{1|2}(z \,|\, t_2) = \Phi\left(\sqrt{(1-\rho^2)}\Phi^{-1}(z) + \rho\Phi^{-1}(t_2)\right)$ is the inverse of $C_{1|2}(t_1 \,|\, t_2)$. Finally, we obtain x_1 from the inverse cdf $F_1^{-1}(t_1)$.

Gumbel copula Let $u_1 = -\log(t_1)$ and $u_2 = -\log(t_2)$. A Gumbel copula with parameter $\eta \in [1, \infty)$ is $C(t_1, t_2 \,|\, \eta) = \exp(-((-\log(t_1))^\eta + (-\log(t_2))^\eta)^{1/\eta})$. It is asymptotically independent in the lower tail and asymptotically dependent in the upper tail with tail dependence coefficient $2 - 2^{1/\eta}$. The corresponding copula density is

$$\exp(-(u_1^\eta + u_2^\eta)^{1/\eta})((u_1^\eta + u_2^\eta)^{1/\eta} + \eta - 1)(u_1^\eta + u_2^\eta)^{1/\eta - 2}(u_1 u_2)^{\eta-1}(t_1 t_2)^{-1}. \tag{8}$$

We obtain the spatially varying Gumbel copula by taking the link function for η in (8) such that $\eta_l(\boldsymbol{v}) \equiv \eta_l(||\boldsymbol{v} - \boldsymbol{v}_{(l)}||) = \min\{(1 - \exp(-||\boldsymbol{v} - \boldsymbol{v}_{(l)}||/\phi))^{-1}, 50\}$, where the upper bound 50 ensures numerical stability.

The Gumbel copula can be written as $\overline{C}(u_1, u_2 \,|\, \eta) = \exp(-(u_1^\eta + u_2^\eta)^{1/\eta})$, which is a bivariate exponential survival function, with marginals corresponding to a unit rate exponential distribution. Then the conditional cdf of T_1 given $T_2 = t_2$ is

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

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

B.3.2 The model and inference

Given observations $y(\boldsymbol{s}_i), i = 1, \ldots, n$, we conduct conditional likelihood inference based on $(y(\boldsymbol{s}_1), \ldots, y(\boldsymbol{s}_L))$. The copula NNMP model with stationary marginal $f_Y = \text{Ga}(a, b)$, and with a set of latent variables t_i, is given by

$$
\begin{aligned}
y(\boldsymbol{s}_i) \mid \boldsymbol{y}_{\text{Ne}(\boldsymbol{s}_i)} &\overset{ind.}{\sim} \sum_{l=1}^{L} c_{\boldsymbol{s}_i,l}(y(\boldsymbol{s}_i), y(\boldsymbol{s}_{(il)})) f_Y(y(\boldsymbol{s}_i)) 1_{(r^*_{\boldsymbol{s}_i,l-1}, r^*_{\boldsymbol{s}_i,l})}(t_i), \\
t_i \mid \boldsymbol{\gamma}, \kappa^2 &\overset{ind.}{\sim} N(t_i \mid \gamma_0 + \gamma_1 s_{i1} + \gamma_2 s_{i2}, \kappa^2),
\end{aligned}
$$

for $i = L+1, \ldots, n$. We denote $c_{\boldsymbol{s}_i,l}$ as the copula density associated with the copula $C_{\boldsymbol{s}_i,l}$ for the component bivariate vector $(U_{\boldsymbol{s}_i,l}, V_{\boldsymbol{s}_i,l})$. In either case of $C_{\boldsymbol{s}_i,l}$ (a spatially varying Gaussian copula or a spatially varying Gumbel copula), there is an associated copula parameter ϕ. The transformed cutoff points $r^*_{\boldsymbol{s}_i,l} = \log(r_{\boldsymbol{s}_i,l}/(1 - r_{\boldsymbol{s}_i,l}))$ where $r_{\boldsymbol{s}_i,l}$ is specified using an exponential correlation function $\exp(-||\boldsymbol{s}_i - \boldsymbol{s}_{(il)}||/\zeta)$. We complete the Bayesian specification by assuming $a \sim \text{Ga}(u_a, v_a)$, $b \sim \text{Ga}(u_b, v_b)$, $\phi \sim \text{IG}(u_\phi, v_\phi)$, $\text{IG}(\zeta \mid u_\zeta, v_\zeta)$, $N(\boldsymbol{\gamma} \mid \boldsymbol{\mu}_\gamma, \boldsymbol{V}_\gamma)$ and $\text{IG}(\kappa^2 \mid u_{\kappa^2}, v_{\kappa^2})$.

The posterior updates of the weight-relevant parameters are provided in Section 3.2 of the paper. We provide the updates for parameters (a, b, ϕ). Let ℓ_i be a configuration variable such that $\ell_i = l$ if $t_i \in (r^*_{\boldsymbol{s}_i,l-1}, r^*_{\boldsymbol{s}_i,l})$ for $i \geq L+1$. Denote by $f_{\boldsymbol{s}_i,l} = c_{\boldsymbol{s}_i,l}(y(\boldsymbol{s}_i), y(\boldsymbol{s}_{(il)})) f_Y(y(\boldsymbol{s}_i))$. The posterior full conditional distributions for parameters a, b and ϕ are proportional to $\text{IG}(a \mid u_a, v_a) \prod_{i=L+1}^n f_{\boldsymbol{s}_i,\ell_i}$, $\text{IG}(b \mid u_b, v_b) \prod_{i=L+1}^n f_{\boldsymbol{s}_i,\ell_i}$, and $\text{IG}(\phi \mid u_\phi, v_\phi) \prod_{i=L+1}^n c_{\boldsymbol{s}_i,l}(y(\boldsymbol{s}_i), y(\boldsymbol{s}_{(il)}))$, respectively. Each parameter is updated on its log scale with a random walk Metropolis step.

Additional References

Joe, H. (2014). *Dependence modeling with copulas*. CRC press.

Zheng, X., Kottas, A., and Sansó, B. (2020). On construction and estimation of stationary mixture transition distribution models. *arXiv:2010.12696.*