

# Bayesian Spectral Modeling for Multivariate Spatial Distributions of Elemental Concentrations in Soil

Maria A. Terres<sup>\*,†</sup>, Montserrat Fuentes<sup>\*,‡</sup>,  
Dean Hesterberg<sup>§</sup>, and Matthew Polizzotto<sup>¶</sup>

**Abstract.** Recent technological advances have enabled researchers in a variety of fields to collect accurately geocoded data for several variables simultaneously. In many cases it may be most appropriate to jointly model these multivariate spatial processes without constraints on their conditional relationships. When data have been collected on a regular lattice, the multivariate conditionally autoregressive (MCAR) models are a common choice. However, inference from these MCAR models relies heavily on the pre-specified neighborhood structure and often assumes a separable covariance structure. Here, we present a multivariate spatial model using a spectral analysis approach that enables inference on the conditional relationships between the variables that does not rely on a pre-specified neighborhood structure, is non-separable, and is computationally efficient. Covariance and cross-covariance functions are defined in the spectral domain to obtain computational efficiency. The resulting pseudo posterior inference on the correlation matrix allows for quantification of the conditional dependencies. A comparison is made with an MCAR model that is shown to be highly sensitive to the choice of neighborhood. The approaches are illustrated for the toxic element arsenic and four other soil elements whose relative concentrations were measured on a microscale spatial lattice. Understanding conditional relationships between arsenic and other soil elements provides insights for mitigating pervasive arsenic poisoning in drinking water in southern Asia and elsewhere.

**Keywords:** conditional dependence, lattice, non-separable covariance, quasi-matern spectral density, spatial modeling.

## 1 Introduction

Expansive spatial datasets are becoming more common as data-collection efficiency is improved with technological advances. In turn, this has created a need for computationally efficient modeling approaches that can accommodate these large datasets. Examples of such approaches include the predictive process (Banerjee et al., 2008),

---

<sup>\*</sup>Department of Statistics, North Carolina State University, Raleigh, NC 27695-8203

<sup>†</sup>The Climate Corporation, San Francisco, CA 94103, [mterres.stat@gmail.com](mailto:mterres.stat@gmail.com)

<sup>‡</sup>Virginia Commonwealth University, Richmond, VA 23284, [mfuentes@vcu.edu](mailto:mfuentes@vcu.edu)

<sup>§</sup>Soil Science Department, North Carolina State University, Raleigh, NC 27695-7619, [dean\\_hesterberg@ncsu.edu](mailto:dean_hesterberg@ncsu.edu)

<sup>¶</sup>Soil Science Department, North Carolina State University, Raleigh, NC 27695-7619, [matt-polizzotto@ncsu.edu](mailto:matt-polizzotto@ncsu.edu)

nearest-neighbor Gaussian process (Datta et al., 2014), partitioning of the spatial region (Kim et al., 2005), covariance tapering (Sang and Huang, 2012), SPDE approximations to Gaussian random fields (Lindgren et al., 2011) and others. Each of these approaches exhibits unique strengths and weaknesses, as discussed by Stein (2014). In addition, for many modern datasets there may be interest in modeling multiple spatial variables jointly, treating them as dependent spatial random variables, in lieu of a regression approach where several variables are simply being conditioned on. However, few of the aforementioned methods can easily accommodate these multivariate spatial datasets. Development of computationally efficient approaches to handle such data has the potential to greatly improve the extent to which researchers can learn about the relationships between spatial processes.

When data have been collected on a spatial lattice, common in fields such as medical imaging and environmental science, some of the most common modeling approaches come from the family of conditionally autoregressive (CAR) models, laid out by Besag (1974) and extended to the multivariate case (MCAR) by Mardia (1988). At their heart, CAR models have been defined such that the spatial observation at any location will be normally distributed with a mean that is a weighted average of the neighboring observations. Defined through a series of conditional distributions, CAR models assume conditional independence between observations that are not spatially adjacent, and as such are special cases of Markov Random Field (MRF) models. In the literature MCAR models have been criticized for possessing a poorly identified and overly elaborate dependence structure, and multiple reparameterizations have been proposed to improve propriety and model behavior (Gelfand and Vounatsou, 2003; Sain and Cressie, 2007; Zhang et al., 2009; Sain et al., 2011).

Although originally proposed by Mardia (1988), the most prevalent form of an MCAR model is the adaptation by Gelfand and Vounatsou (2003) ensuring distributional propriety. In this form the covariance structure is separable and can be decomposed into a covariance matrix describing the (non-spatial) relationship between the variables and a second covariance matrix describing the spatial dependence shared across all variables. Although this is computationally efficient, it is quite restrictive in its description of the marginal spatial dependencies of the variables. Alternative formulations, such as Jin et al. (2005) and Jin et al. (2007), allow for non-separable formulations but become more computationally expensive. Finally, all of these approaches rely on a pre-defined neighborhood structure that limits the shape and extent of the spatial dependence in a way that is avoided by working with the spectral approach we propose here.

The Markovian structure of the MCAR model constrains spatial dependence to a set of neighbors pre-determined through an adjacency matrix. This is in contrast to geostatistical approaches where spatial dependence is assumed to decay as a smooth function of distance and the covariance parameters. In situations where there is interest in jointly modeling multivariate spatial lattice data while avoiding the separable Markovian structure and propriety issues inherent in MCAR models, spectral analysis procedures provide a natural framework to turn to. Computations are conducted after transforming the data into the “spectral domain,” allowing for greater efficiency

as discussed in Section 3. The literature on spectral analysis techniques is prolific in time series contexts (Priestley, 1981; Koopmans, 1995), and also fairly common in the frequentist spatial literature (Stein, 1999). However with a few exceptions, e.g. Hancock and Stein (1993), Reich and Fuentes (2012), and Stroud et al. (2014), there has been relatively little work in this area when modeling spatial data within a Bayesian framework.

In this paper we present a joint multivariate spatial model using spectral analysis procedures to gain computational efficiency. Each spatial variable has a spectral density controlling the marginal spatial covariance function, while a correlation matrix controls the relative cross-covariances between the variables. The model is fit in a Bayesian framework, allowing for uncertainty quantification in all aspects of the model. As described in Section 3, the multivariate normal likelihood is approximated by the Whittle Likelihood (Whittle, 1954), and all posterior distributions are thus referred to as “pseudo posteriors.” Of particular interest is the pseudo posterior examination of the correlation matrix which provides inference on the nature of the conditional dependencies between the variables.

This methodology is illustrated for an analysis of microscale accumulation of potentially toxic arsenic in soil. Arsenic contamination of drinking water is a widespread human health concern, particularly in Asian countries where over 100 million people routinely consume dangerous levels of arsenic by drinking naturally contaminated well water (Ravenscroft et al., 2009). Several methods of water treatment have been studied with varying success, including the introduction of additional chemical salts or solutions that will react with the arsenic (Jiang et al., 2012; Komárek et al., 2013). Soil chemical processes influence the distribution of arsenic in well water and subsequent threats to human health (Polizzotto et al., 2008). However, because soils comprise multiple elements in multiple mineral and organic components, a more precise understanding of chemical reactions could be aided through a statistical description of the soil elements’ dependencies.

For this analysis synchrotron X-ray fluorescence microprobe ( $\mu$ -XRF) analysis was used to map accumulated arsenic in relation to other chemical elements in thin coatings on a quartz sand grain collected from a soil sample. The technique produces multivariate spatial lattice maps that essentially reflect relative abundances of elements. Interest lies in understanding the conditional dependencies between arsenic and the other soil components, and whether these dependent components serve to mitigate or potentiate the accumulation of arsenic.

In natural systems, the mobility and toxicity of arsenic is largely controlled by various competing abiotic and biotic redox and adsorption processes involving organic matter and iron, aluminum, and manganese oxides (Borch et al., 2009). Many of these reactions have been studied using single or binary mixtures of aqueous species, model minerals and/or organic components. In contrast, less mechanistic detail is known about arsenic behavior in soils, which comprise diverse assemblages of minerals and organic matter. Unlike the common approach of modeling X-ray absorption spectra from soils to identify pure chemical species (Manceau et al., 2014), our approach aims to identify, via element associations, possible interactions between multiple soil components that

cause the reactivity of soil species to differ from what may be expected from model chemical analogues studied in isolation of soil.

Studying the pairwise behavior of soil elements, a common approach for analysis of microscale soil chemical data, provides only a very limited view of their relationships and neglects to account for higher-order interactions between multiple elements. This is in contrast to our hypothesis that interactions between soil components inferred from element pairs may depend on other co-localized elements in complex assemblages of soil components. In order to capture these kinds of behaviors, it is necessary to have a model that is sufficiently flexible in its treatment of conditional dependencies. The model developed here is illustrated for arsenic and several co-localized elements. Such conditional dependency models would enable scientists to better understand how arsenic behaves when multiple components are co-localized within complex soil materials.

The lattice structure of the  $\mu$ -XRF data make spectral procedures a natural choice for model-fitting. This approach was previously explored by Guinness et al. (2014) in a frequentist setup, but the methodology could not accommodate more than three spatial variables and lacked adequate measures of uncertainty. The need for additional model flexibility and improved uncertainty quantification indicated a fresh look at the problem from a Bayesian perspective. Unlike the model developed by Guinness et al. (2014), the modeling framework we outline can easily accommodate any arbitrary number of spatial variables. We illustrate the methodology with five soil elements, including the three that were previously analyzed, providing full descriptions of the uncertainty associated with our estimates based on the pseudo posteriors and producing conditional dependence graphs that clearly illustrate the relationships between the variables.

The remainder of this article is organized as follows. In Section 2 the  $\mu$ -XRF data are presented as a motivating example. Section 3 provides an overview of the spectral analysis approach in the spatial domain, outlining some properties of the commonly used approximations, and specifying additional modeling details. Section 4 outlines details for the MCAR model used for comparison with the proposed spectral model. In Section 5 the model results and potential implications are discussed for both a simulation example as well as the  $\mu$ -XRF data. Finally, Section 6 concludes with a brief discussion of the presented methodology and its potential for future work.

## 2 Multivariate Spatial $\mu$ -XRF Quartz Sand Grain Data

Data generated for this analysis were obtained by introducing solutions of arsenic to a quartz sand grain collected from a soil sample and subsequently assessing the microscale spatial distributions of arsenic and a number of native soil elements. The sand grain analyzed was separated from a surface soil sample collected from a forest at the Central Crops Research Station in Clayton, NC. The spatial distributions of elements were mapped by  $\mu$ -XRF using Beamline X27A at the National Synchrotron Light Source (NSLS), Brookhaven National Laboratory. Our analysis focuses on element maps collected after sequential treatments of the grain with 100 and 1000  $\mu$ M arsenic (III) treatments. A  $350 \times 450$   $\mu$ m region of interest (ROI) was mapped using an X-ray beam of approximately  $10 \times 10$   $\mu$ m<sup>2</sup> to yield a  $35 \times 45$  pixel array. Resulting elemental maps

reflect the relative abundance of elements within each  $10 \times 10 \times \sim 15 \mu\text{m}^3$  voxel analyzed by the incident X-ray beam. Additional details of data collection were outlined by Guinness et al. (2014).

We focus our analysis on soil concentrations of arsenic (As), iron (Fe), chromium (Cr), nickel (Ni) and zinc (Zn); iron is an abundant soil element found in As-reactive Fe-oxide minerals, and the other four are trace elements. The elements are modeled on the log-scale and, since our interest lies primarily in understanding the dependence structure between these elements, we center each spatial process by subtracting the sample mean and assume zero means in our modeling. Arsenic and iron were most abundant, averaging 7.23 and 9.47 on the log-scale, with the other elements averaging between 5.35 and 5.60. To specify the model, let  $\mathbf{Z}^{(m)}(\mathbf{s})$  be the centered log fluorescence signal at location  $\mathbf{s} \in \mathbb{J}_N$  for soil element  $m$ , for  $m = 1, \dots, M$ , with  $M = 5$ . The  $\mathbf{Z}^{(m)}(\mathbf{s})$  are spatial quantities, with each assumed to be a realization from a Gaussian process and are modeled jointly in the spectral domain, as described in Section 3.

The spatial maps of these five soil elements are provided in Figure 1. Note that some concentrated hotspots of iron, chromium and nickel (e.g. at coordinates (300,200)) are known contaminants from stainless steel deposited on the sample during handling prior to reaction. The correlation between arsenic, iron, and to some extent chromium is readily apparent, with similar spatial features throughout the region. The correlations with nickel and zinc are less apparent. Through our joint modeling procedure we seek to make inference regarding the conditional dependencies exhibited by these five elements.

## 3 Spectral Methods

### 3.1 Computing the Likelihood

Consider a spatial process  $\mathbf{Z} = (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_N))'$  assumed to be a realization from a Gaussian process with stationary covariance function  $c(\mathbf{h}) = \text{cov}(Z(\mathbf{s}), Z(\mathbf{s} + \mathbf{h}))$  that depends on parameters  $\theta$ . Let  $\Sigma_\theta$  denote the corresponding covariance matrix. Then  $\mathbf{Z} \sim N(\mathbf{0}, \Sigma_\theta)$  and the log-likelihood can be computed,

$$\log(p(\mathbf{Z}|\theta)) = -\frac{1}{2}(\log \det \Sigma_\theta + \mathbf{Z}'\Sigma_\theta^{-1}\mathbf{Z}), \quad (1)$$

where proportionality constants have been ignored. Normal likelihoods are generally easy to compute when  $N$  is small and are commonly used in spatial modeling (Stein, 1999; Cressie and Wikle, 2011; Banerjee et al., 2014). However, the matrix inverse  $\Sigma_\theta^{-1}$  requires  $O(N^3)$  floating point operations (flops), rendering computation of this likelihood undesirable when working in large dimensions. In a Bayesian analysis this likelihood may need to be computed multiple times in each iterate of the Gibbs sampler. Such repeated computation of the likelihood may be a critical component in determining the overall computational efficiency of the proposed approach.

When the observations are available on a lattice of size  $N = n_1 \times n_2$ , denoted  $\mathbb{J}_N$ , then it is convenient to work in the spectral domain since the normal log-likelihood can be approximated using the Whittle likelihood (Whittle, 1954; Zimmerman, 1989).

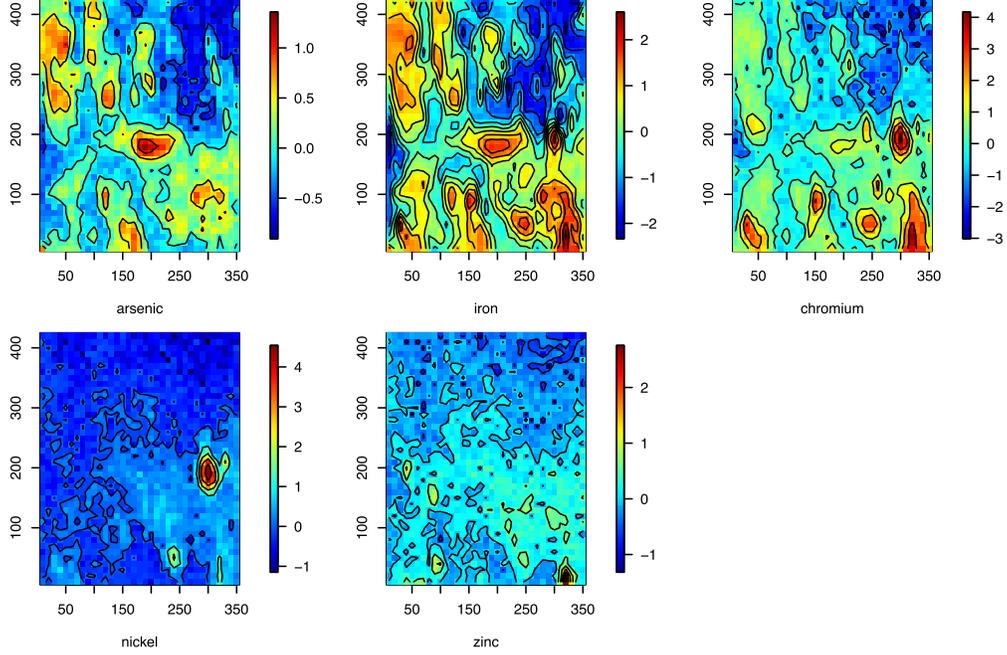


Figure 1: The logged and centered soil components arsenic, iron, chromium, nickel, and zinc.

This approximation takes advantage of the computational efficiency of the Fast Fourier Transform (FFT), dramatically reducing computation time. This approach is outlined below.

Bochner's Theorem states that any stationary covariance function can be represented as an inverse Fourier transform,

$$c(\mathbf{h}) = \int_{\mathbb{R}^2} \exp(i\boldsymbol{\omega}'\mathbf{h})dF(\boldsymbol{\omega}), \quad (2)$$

where  $\mathbf{h}$  is a separation vector for spatial locations  $\mathbf{s}$  and  $\mathbf{s} + \mathbf{h}$ , and  $\boldsymbol{\omega} = (\omega_1, \omega_2)$  is a bivariate spectral frequency. We assume there exists some continuous differentiable  $f(\boldsymbol{\omega})$  such that  $dF(\boldsymbol{\omega}) = f(\boldsymbol{\omega})d\boldsymbol{\omega}$ , commonly referred to as the spectral density. The Spectral Representation Theorem states that the spatial process associated with this covariance function can similarly be represented,

$$Z(\mathbf{s}) = \int_{\mathbb{R}^2} \exp(i\boldsymbol{\omega}'\mathbf{s})d\tilde{Z}(\boldsymbol{\omega}), \quad (3)$$

where  $d\tilde{Z}(\boldsymbol{\omega})$  have uncorrelated increments and  $E|d\tilde{Z}(\boldsymbol{\omega})|^2 = f(\boldsymbol{\omega})$  (Yaglom, 1987). It should be noted that the above theorems are stated in the spatial context, focusing on two dimensions, but similar statements can be made for  $\mathbb{R}^d$  with  $d > 2$ .

When the data have been observed on a grid, there is inadequate information to fully recover the continuous process. In particular, if the data are observed at uniformly spaced locations with intervals of length  $\delta$ , i.e. the process can be written  $Z(\delta\mathbf{x})$  for  $\mathbf{x} \in \mathbb{Z}^2$ , then the associated spectra are constrained to frequencies in the finite interval  $-\pi/\delta \leq \omega \leq \pi/\delta$ . This can be seen by observing that  $\exp(i\boldsymbol{\omega}'\mathbf{h}) = \exp(i(\boldsymbol{\omega} + 2\pi j/\delta)'\mathbf{h})$ , the so called ‘‘aliasing’’ phenomenon.

When working with lattice data the aliasing must be accounted for in the choice of the associated spectral density. One approach is to choose a spectral density defined on the real plane, then accumulate the density over all aliased frequencies. An alternative is to work with a spectral density that has been explicitly defined to have support on  $\boldsymbol{\omega} \in [-\pi/\delta, \pi/\delta]^2$ . We follow the latter approach in our analyses, selecting the quasi-Matérn spectral density introduced by Guinness et al. (2014). To ensure a real-valued process, the spectral densities must additionally be even functions symmetric around zero, which is again satisfied by the quasi-Matérn spectral density.

In practice we cannot compute the integrals in (2) and (3), so we instead approximate them with discrete sums,

$$c(\mathbf{h}) = \sum_{j \in \mathbb{J}_N} e^{i\boldsymbol{\omega}_j'\mathbf{h}} f(\boldsymbol{\omega}_j), \quad (4)$$

$$Z(\mathbf{s}) = \sum_{j \in \mathbb{J}_N} e^{i\boldsymbol{\omega}_j'\mathbf{s}} \tilde{Z}(\boldsymbol{\omega}_j) \quad (5)$$

evaluated at the Fourier frequencies  $\boldsymbol{\omega}_j = (2\pi j_1/n_1, 2\pi j_2/n_2)$ ,  $j = (j_1, j_2) \in \mathbb{J}_N$ . These approximations have a long history of use in spectral modeling of spatial processes, and their properties are well established (Whittle, 1954; Guyon, 1982). In particular, these approximations are known to perform well for stationary processes where data have been observed on a regular lattice, though work exists extending applicability beyond both of these assumptions, such as Fuentes (2007) and Stroud et al. (2014).

Due to the lattice structure in the data, the corresponding covariance matrix  $\Sigma_\theta$  will be block circulant and the FFT will effectively diagonalize  $\Sigma_\theta$ . This enables the log-likelihood to be rewritten in the spectral domain, with computation limited by the FFT requiring only  $O(N \log N)$  flops,

$$\log(p(\mathbf{Z}|\theta)) = -\frac{1}{2} \left( \sum_{j \in \mathbb{J}_N} \log(f(\boldsymbol{\omega}_j)) + \sum_{j \in \mathbb{J}_N} F_N(Z_j)^* f(\boldsymbol{\omega}_j)^{-1} F_N(Z_j) \right), \quad (6)$$

where  $F_N(\mathbf{Z})$  is an array denoting the 2-dimensional Fourier transformation of the lattice data  $\mathbf{Z}$  such that the entry  $F_N(Z_j)$  corresponds to the Fourier frequency  $\boldsymbol{\omega}_j$ . The first term in the summation corresponds to the contribution of  $\log \det \Sigma_\theta$  in (1). This equality can be seen in part by noting that the eigenvalues of  $\Sigma_\theta$  correspond to the spectral density  $f(\boldsymbol{\omega})$  evaluated at each of the Fourier frequencies. The second term in the summation is the Fourier representation of the quadratic term in (1), with  $f(\boldsymbol{\omega}_j)^{-1}$  corresponding to the covariance and  $F_N(Z_j)$  and its complex conjugate corresponding to the observed data. These relationships arise by inverting the expressions in (4) and

(5). Further discussion of the Whittle Likelihood is beyond the scope of this paper, and we instead refer interested readers to the extensive literature on that topic (e.g. Whittle, 1954; Guyon, 1982; Dahlhaus and Künsch, 1987; Stein, 1999; Stein et al., 2004).

The disadvantage of this approximation is that the discreteness of (4) and (5) produces the generally undesirable property of periodicity over the range of the data. An adjustment is necessary to mitigate this feature. Some new approaches have recently been proposed in the literature involving an imputed expansion of the lattice such that the periodicity occurs beyond the domain of the observed data (e.g. Guinness and Fuentes, 2016; Stroud et al., 2014). However, we follow the more common approach of data tapering (Dahlhaus, 1983; Dahlhaus and Künsch, 1987). Intuitively, tapering dampens the data along the edges towards zero in a smooth way such that the lattice could be folded into a torus shape without introducing discontinuities.

Specifically, we implement a cosine taper, or Tukey taper (Tukey, 1967; Bloomfield, 2004), defined as,

$$w_d(j) = \begin{cases} \frac{1}{2}(1 + \cos(\frac{2\pi}{r}(j - \frac{r}{2}))), & 0 \leq j < \frac{r}{2}, \\ 1, & \frac{r}{2} \leq j < 1 - \frac{r}{2}, \\ \frac{1}{2}(1 + \cos(\frac{2\pi}{r}(j - 1 + \frac{r}{2}))), & 1 - \frac{r}{2} \leq j \leq 1, \end{cases} \quad (7)$$

for dimensions  $d = 1, 2$  and  $j = 0/n_d, \dots, (n_d - 1)/n_d$ . The parameter  $r$  controls the extent of the tapering, typically chosen to be 5–10% of the observations at each boundary. The original data  $Z_j = Z_{j_1, j_2}$  in the likelihood in (6) is then replaced with the tapered data  $Z_{j_1, j_2} \times w_1(j_1) \times w_2(j_2)$ , and a multiplicative adjustment of  $\prod_{d=1}^2 \sum_{j=1}^{n_d} w_d(j)^2$  is incorporated into the  $F_N(Z_j)$  terms. This cosine taper is a common choice in spectral modeling, tapering most strongly at the margins and smoothly transitioning to untapered data away from the margins. For more information on the properties of the cosine taper readers are referred to Bloomfield (2004).

When the data are multivariate, with lattice observations for  $M$  spatial variables  $\mathbf{Z} = (\mathbf{Z}^{(1)}, \dots, \mathbf{Z}^{(M)})$ , the multivariate process will require a positive definite  $M \times M$  matrix of spectral densities,  $\mathbf{f}(\boldsymbol{\omega})$ . The matrix entries along the diagonal,  $\mathbf{f}_{m,m}(\boldsymbol{\omega})$ , dictate the marginal covariances for each of the variables. Similarly, the matrix entries on the off-diagonal,  $\mathbf{f}_{m,m'}(\boldsymbol{\omega})$ , dictate the cross-covariances between each pair of variables. The likelihood can then be approximated similarly to the univariate case,

$$\log(p(\mathbf{Z}|\theta)) = -\frac{1}{2} \left( \sum_{j \in \mathbb{J}_N} \log \det \mathbf{f}(\boldsymbol{\omega}_j) + \sum_{j \in \mathbb{J}_N} F_N(\mathbf{Z}_j)^* \mathbf{f}(\boldsymbol{\omega}_j)^{-1} F_N(\mathbf{Z}_j) \right), \quad (8)$$

where  $F_N(\mathbf{Z}^{(i)})$  is an array denoting the 2-dimensional Fourier transformation of the lattice data for spatial variable  $\mathbf{Z}^{(i)}$  such that entry  $F_N(\mathbf{Z}_j^{(i)})$  corresponds to the Fourier frequency  $\boldsymbol{\omega}_j$ , the  $M \times 1$  vector  $F_N(\mathbf{Z}_j) = (F_N(\mathbf{Z}_j^{(1)}), \dots, F_N(\mathbf{Z}_j^{(M)}))'$  concatenates the elements of these matrices corresponding to the frequency  $\boldsymbol{\omega}_j$  for each spatial variable, and  $F_N(\mathbf{Z}_j)^*$  is the analogous vector corresponding to the complex conjugate transpose

of the Fourier transformed lattice data. Data tapering can be conducted in the same manner as in the univariate case.

Similar to the univariate case, computation of the multivariate likelihood (8) is limited only by the computation of the FFT on each of the spatial variables and the inversion of the  $M \times M$  matrix  $\mathbf{f}(\boldsymbol{\omega})$ . In addition, the Whittle Likelihood provides an important advantage in the multivariate setting that is not relevant in the univariate case: it avoids specification of a cross-covariance matrix. In a standard setup the cross-covariance matrix would need to be specified in a constrained way in order to ensure positive definiteness. By working in the spectral domain the cross-covariances are specified via spectral densities, allowing for straight forward incorporation of non-separability or dependence on covariates without the risk of issues regarding positive definiteness.

### 3.2 Marginal Spectral Densities

Recall that the covariance function,  $c(\mathbf{h})$ , for each soil element is defined in the spectral domain using a spectral density,  $f(\boldsymbol{\omega})$ . The spectral densities are constrained to be even and to have support on  $[-\pi/\delta, \pi/\delta]^2$  in order to produce real-valued realizations and to avoid the aliasing effect that occurs with lattice observations, as described earlier. While several parametric densities exist in the literature, many are defined on the real line and require care with the aliased frequencies, and others are difficult to interpret.

With the above properties in mind, we focus on the non-separable quasi-Matérn spectral density defined on a lattice,

$$f(\boldsymbol{\omega}) = f(\omega_1, \omega_2) = \frac{\sigma^2}{(1 + (\alpha/\delta)^2(\sin^2(\delta\omega_1/2) + \sin^2(\delta\omega_2/2)))^{\nu+1}}, \quad (9)$$

for  $\alpha, \sigma^2, \nu > 0$  and  $\boldsymbol{\omega} = (\omega_1, \omega_2) \in [-\pi/\delta, \pi/\delta]^2$  as described by Guinness et al. (2014). This spectral density is attractive in part because it approaches the spectral density of the isotropic Matérn covariance function as  $\delta \rightarrow 0$ . In this sense, the quasi-Matérn spectral density is a natural choice for researchers who commonly model point referenced spatial processes and are familiar with the properties of the Matérn class of covariance functions.

The parameter  $\alpha$  functions as a range parameter, indicating the distance at which spatial dependence becomes negligible. The parameter  $\sigma^2$  functions as a variance parameter, controlling the magnitude of the spatial correlation. The parameter  $\nu$  controls the smoothness of the spatial surface. Due to concerns regarding lack of identifiability between the parameters (Zhang, 2004), we fix  $\nu \equiv 1$  in our model fitting.<sup>1</sup> We additionally assume that each  $10 \times 10 \mu\text{m}$  grid cell represents one unit of distance, setting  $\delta \equiv 1$ .

In this proposed framework the covariance structure for the multivariate process is constructed in terms of marginal covariance functions for each of the elements. While

---

<sup>1</sup>Additional values of  $\nu$  were considered to test for sensitivity, and all produced comparable results to those presented here.

some multivariate covariance functions have been proposed in the literature, such as Apanasovich et al. (2012), these approaches are generally quite complex and computationally inefficient. As such, these families of covariance functions were not considered here.

### 3.3 Cross-Covariance Structure

To guarantee a well defined spatial process the  $M \times M$  matrix  $\mathbf{f}(\boldsymbol{\omega})$ , corresponding to the cross-covariance matrix in the spatial domain, must be positive definite. To guarantee this, we follow the approach of Guinness et al. (2014) and write

$$\mathbf{f}(\boldsymbol{\omega}) = \text{diag}(f^{1/2}(\boldsymbol{\omega}))\boldsymbol{\rho}(\boldsymbol{\omega}) \text{diag}(f^{1/2}(\boldsymbol{\omega})), \quad (10)$$

where  $\text{diag}(f^{1/2}(\boldsymbol{\omega}))$  is a diagonal matrix with diagonal entries  $f_m^{1/2}(\boldsymbol{\omega})$ , the square root of the marginal spectral density for the  $m$ th element as defined in (9), for  $m = 1, \dots, M$ . Here,  $\boldsymbol{\rho}(\boldsymbol{\omega})$  is the coherence matrix, a correlation matrix with ones on the diagonal and elements  $\rho_{m,m'}(\boldsymbol{\omega})$  describing the correlation between components  $m$  and  $m'$ . This is analogous to decomposing a covariance matrix into a correlation matrix and vectors of standard deviations, where here the standard deviations are the square roots of the spectral densities instead of the square roots of the variances.

Even in relatively low dimensions, modeling the matrix  $\boldsymbol{\rho}(\boldsymbol{\omega})$  will require the estimation of many correlation functions  $\rho_{m,m'}(\boldsymbol{\omega})$ . For example, in our example with 5 soil elements there would be  $5 \times (5 - 1)/2 = 10$  functions that need to be estimated in the  $\boldsymbol{\rho}(\boldsymbol{\omega})$  matrix. Estimating this many functions may rapidly become prohibitive as the dimension increases. As an alternative, we propose

$$\mathbf{f}(\boldsymbol{\omega}) = \text{diag}(f^{1/2}(\boldsymbol{\omega}))\boldsymbol{\rho} \text{diag}(f^{1/2}(\boldsymbol{\omega})), \quad (11)$$

where the matrix  $\boldsymbol{\rho}$  is constant across frequencies. This assumption implies that for any pair of soil elements, the spatial dependence described by the cross-covariance functions will be dictated solely by the pair of marginal covariance functions and a multiplicative factor  $\rho_{m,m'}$ . While this assumption is primarily motivated by the need for computational efficiency, the resulting model is also better identified and still defines a very flexible framework deemed sufficient to describe the  $\mu$ -XRF data.

Decomposing  $\mathbf{f}(\boldsymbol{\omega})$  as in (11) additionally simplifies the computation of the likelihood. Specifically, it is straight forward to show that  $\det \mathbf{f}(\boldsymbol{\omega}) = (\det \boldsymbol{\rho}) \prod_{m=1}^M f_m(\boldsymbol{\omega})$  and to rewrite the inverted matrix  $\mathbf{f}(\boldsymbol{\omega})^{-1} = \text{diag}(f^{-1/2}(\boldsymbol{\omega}))\boldsymbol{\rho}^{-1}\text{diag}(f^{-1/2}(\boldsymbol{\omega}))$ . Then, substituting into the multivariate likelihood from (8) we have,

$$\begin{aligned} \log(p(\mathbf{Z}|\theta)) = & -\frac{1}{2} \left( N \log \det \boldsymbol{\rho} + \sum_{j \in \mathbb{J}_N} \sum_{m=1}^M \log f_m(\boldsymbol{\omega}_j) \right) \\ & + \sum_{j \in \mathbb{J}_N} F_N(\mathbf{Z}_j)^* \text{diag}(f^{-1/2}(\boldsymbol{\omega}))\boldsymbol{\rho}^{-1} \text{diag}(f^{-1/2}(\boldsymbol{\omega}))F_N(\mathbf{Z}_j). \end{aligned} \quad (12)$$

Note, during model fitting the inversion and determinant of  $\boldsymbol{\rho}$  need only be computed once at each iteration since it does not depend on the frequency.

### 3.4 Conditional Independencies and the Correlation Matrix

As described by Guinness et al. (2014), conditional independencies between the soil components can be explored through examination of  $\mathbf{f}(\boldsymbol{\omega})^{-1}$ , or equivalently through examination of  $\boldsymbol{\rho}^{-1}$  under our proposed parameterization. The Bayesian paradigm allows us to consider uncertainties for each element of the correlation matrix enabling examination of the implied conditional independencies, similar to the approach followed by Hoff (2007). Specifically, consider a vector of normal random variables  $(z_1, z_2, \dots, z_n)'$  with associated correlation matrix  $\mathbf{C}$ . Then in the conditional distribution for  $z_j$  given the other variables,  $p(z_j|z_{-j})$ , the ‘‘coefficients’’ on  $z_{-j}$  will be  $\mathbf{C}_{[j,-j]}\mathbf{C}_{[-j,-j]}^{-1}$ . The sign of these ‘‘coefficients’’ and whether the credible intervals overlap zero will provide inference on the graph of conditional dependence between variables. In this manner, the conditional independencies between the soil elements will be explored in our model through the correlation matrix  $\boldsymbol{\rho}$ .

### 3.5 Prior Specification and Model Fitting

The following prior distributions are assumed for the model parameters,

$$\begin{aligned}\alpha_m | s^2 &\sim TN_{(0,\infty)}(0, s^2), \\ s^2 &\sim IG(2, 2), \\ \sigma_m^2 | \nu_0, \sigma_0^2 &\sim IG(\nu_0/2, \nu_0\sigma_0^2/2), \\ \sigma_0^2 | \nu_0 &\sim IG(2, 2), \\ p(\nu_0) &\propto e^{-\nu_0}, \\ p(\boldsymbol{\rho}) &\propto 1, \quad \boldsymbol{\rho} \in \mathcal{R}^5,\end{aligned}$$

where  $\nu_0 \in \mathbb{Z}^+$ , and  $\mathcal{R}^5$  is the space of all  $5 \times 5$  positive definite correlation matrices. Each spatial process  $Z_m$  has a unique pair of parameters  $(\alpha_m, \sigma_m^2)$ , with hyperpriors for these parameters facilitating sharing of information across soil elements. A Truncated Normal prior is assumed for the range parameter  $\alpha_m$ , ensuring the parameter is positive. The variance parameter for the Truncated Normal,  $s^2$ , has a vague and uninformative Inverse Gamma prior. The prior on  $\sigma_m^2$  is an Inverse Gamma prior, with vague priors on the hyper parameters following the suggestion of Hoff (2010). The prior on  $\boldsymbol{\rho}$  follows the suggestion of Barnard et al. (2000), who similarly decomposed a covariance matrix into standard deviations and a correlation matrix.

We make inference on the model parameters through Markov chain Monte Carlo (MCMC), based on 5000 samples after a thinning of every 25th iterate and a burn-in of 5000 iterations.<sup>2</sup> The  $\alpha_m$  parameters were updated with a random walk Metropolis Hastings step with the proposal variance tuned during a burn-in period to achieve acceptance rates between 0.3 and 0.5. The  $\sigma_m^2$  parameters are also updated with a Metropolis Hastings step, but in this case a more clever proposal distribution can be

---

<sup>2</sup>Using a 2008 MacBook with a 2 GHz Intel Core 2 Duo processor and 4 GB 1067 MHz DDR3 memory, it takes approximately 30 minutes to sample 5000 iterations of the model.

used. If the spatial process  $Z_m$  were being modeled marginally, then  $\sigma_m^2$  would be conjugate,

$$\sigma_m^2 | \theta \sim IG\left(\frac{\nu_0 + J}{2}, \frac{\nu_0 \sigma_0^2 + \sum_{j \in \mathbb{J}_N} F_N(\mathbf{Z}_j^{(m)}) * F_N(\mathbf{Z}_j^{(m)}) (1 + \alpha_m^2 (\sin^2(\omega_{1j}/2) + \sin^2(\omega_{2j}/2)))^2}{2}\right),$$

where  $\theta$  is the collection of all other parameters. This conjugacy breaks down in the multivariate case, but we utilize this marginal conjugate distribution as a proposal in the Metropolis Hastings step with good results. The hyperparameters  $s^2, \nu_0, \sigma_0^2$  are updated via straightforward Gibbs steps.

At each iterate of the Gibbs sampler  $\rho$  is updated by sequentially sampling each entry of the matrix,  $\rho_{i,j}$ , conditional on the others. These conditional distributions restrict  $\rho_{i,j}$  to an interval that will guarantee continued positive definiteness of the full matrix. These distributions are derived by Barnard et al. (2000) and are available up to a proportionality constant. Each  $\rho_{i,j}$  is then sampled using a griddy Gibbs step (Ritter and Tanner, 1992): the proportional likelihood is evaluated for a grid of values in the permitted interval, and an updated draw is sampled according to the resulting weights.

Note that since the multivariate normal likelihood is being approximated with the Whittle Likelihood, we refer to the resulting posteriors as ‘‘pseudo posteriors.’’ It is assumed that the well known theoretical guarantees regarding the MCMC algorithm will similarly apply to the approach using the Whittle Likelihood approximation.

## 4 Multivariate Conditionally Autoregressive Model

Due to its popularity for multivariate lattice data, a separable multivariate conditionally autoregressive (MCAR) approach is compared to the spectral model for both simulated data (Section 5.1) and the  $\mu$ -XRF data (Section 5.3). The MCAR model used for those analyses is described here in the context of the  $\mu$ -XRF data.

The model was defined such that for each soil element  $k$  and each location  $i = 1, \dots, N$  the centered log fluorescence is distributed as a spatial random effect with a separable MCAR prior plus some normally distributed noise. Specifically,

$$\begin{aligned} Z_{k,i} &\sim N(\mu_{k,i}, \tau_k^2), \\ \mu_{k,i} &= \alpha_k + \phi_{k,i}, \\ \phi_i | \phi_{j \neq i} &\sim N\left(\sum_j \frac{w_{i,j}}{w_{i,+}} \phi_j, \frac{1}{w_{i,+}} \Sigma\right), \end{aligned} \tag{13}$$

where  $w_{i,j}$  is an indicator function equal to one if locations  $i$  and  $j$  are neighbors,  $w_{i,+} = \sum_j w_{i,j}$ , and  $\phi_i = (\phi_{1,i}, \dots, \phi_{5,i})'$  is the collection of random effects at location  $i$  for the five soil elements. The MCAR structure assumes the spatial random effect at any location is centered at a weighted average of its neighbors' spatial random effects. For identifiability, these spatial random effects are constrained to sum to zero.

The joint density for  $\phi$  is then,

$$p(\phi|\Sigma) \propto \exp\left(-\frac{1}{2}\phi'(D_w - W) \otimes \Sigma^{-1}\phi\right),$$

where  $D_w - W$  is an  $N \times N$  matrix with  $D_w = \text{diag}(w_{i,+})$ , and  $W$  a proximity matrix such that  $W_{i,j} = w_{i,j}$  and  $W_{i,i} = 0$ . The separability of this model can be easily seen since  $(D_w - W)$  is controlling the spatial dependence, and  $\Sigma^{-1}$  is across the correlations between the soil elements.

The following prior distributions are assumed for the model parameters,

$$\begin{aligned}\alpha_k &\propto 1, \\ \tau_k^2 &\sim IG(2, 2), \\ \Sigma &\sim InvWish(I_5, 5).\end{aligned}$$

The above model was fit in OpenBUGS and R using the package `R2WinBUGS`. All results are based on 5000 sample iterations after a burn-in of 1000 iterations. The above priors were chosen due to constraints imposed by WinBUGS/OpenBUGS – the MCAR prior in WinBUGS/OpenBUGS requires the priors for  $\alpha_k$  and  $\Sigma$  to be flat and Inverse Wishart respectively, with no other choices supported.

## 5 Analysis Results

### 5.1 Simulation Example with Comparison to MCAR

One of the key strengths of the proposed spectral model is its ability to capture conditional relationships between several correlated spatial processes. This feature is first illustrated for a simulation example using a multivariate spatial process that follows the structure described in Section 3. These data are simulated to be the same size as the  $\mu$ -XRF data: five spatial processes on a  $35 \times 45$  unit lattice. The spatial processes are labeled  $z_1, \dots, z_5$ . Conditional independence is induced by introducing zeros into the inverse correlation matrix. Specifically the following pairs are conditionally independent:  $(z_1, z_2)$ ,  $(z_1, z_3)$ , and  $(z_3, z_5)$ .

Three models are fit to these data: the spectral model described in Section 3.5, an MCAR(1), and an MCAR(5). The MCAR models were fitted as described in Section 4 using two neighborhood structures to test for sensitivity in the resulting inference on the conditional relationships.

The resulting conditional dependence “coefficients” from the spectral model and the MCAR models are plotted in Figure 2. The spectral estimates are black dots, the MCAR(1) estimates are open triangles, and the MCAR(5) estimates are crosses. The 95% credible intervals associated with each estimate are shown as vertical lines, and the true values are plotted as red circles. Each subplot demonstrates the conditional relationships between the element given on the y-axis with those on the x-axis. For example, the true values in the first subplot demonstrate that  $z_1$  conditional on the other four elements is independent of  $z_2$  and  $z_3$ .

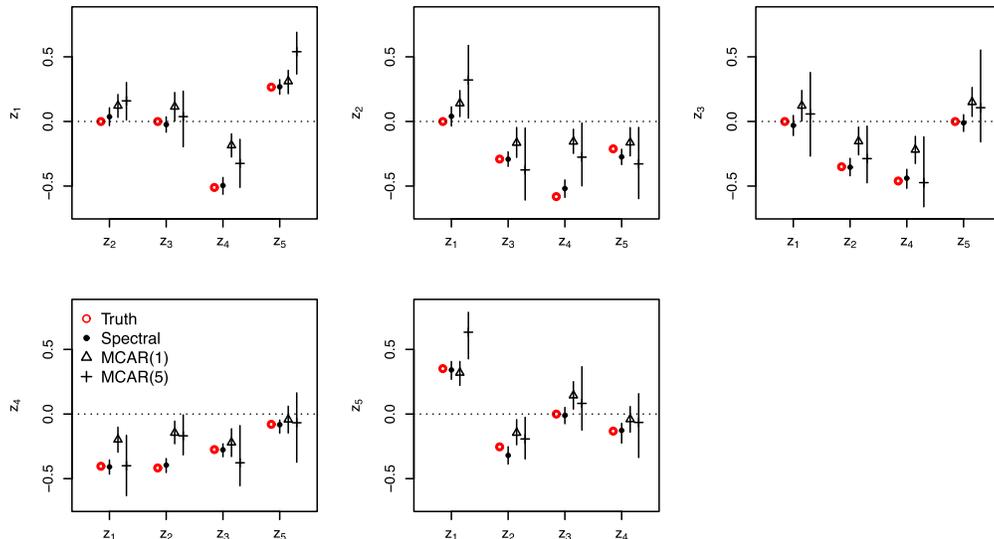


Figure 2: Estimated “coefficients” on  $z_{-j}$  describing conditional dependencies between the simulated soil components, defined to be  $\mathbf{C}_{[j,-j]}\mathbf{C}_{[-j,-j]}^{-1}$ . The red circles indicate the true “coefficients” computed based on the correlation matrix used for simulation, the black dots are estimates using the spectral modeling approach, the open triangles are estimates from an MCAR(1) model, and the crosses are estimates from an MCAR(5) model.

The estimates from the spectral model align well with the true “coefficients”, with all of the conditional independencies correctly identified and with relatively narrow credible intervals. In contrast, the MCAR(1) intervals are wider than the spectral approach, and the MCAR(5) intervals are even wider. MCAR(1) did not correctly identify any true zero, and it erroneously identified  $z_4$  and  $z_5$  as being conditionally independent. MCAR(5) correctly identified two true zeros, but it also erroneously identified  $z_4$  and  $z_5$  as being conditionally independent. There was no clear pattern of superior performance when comparing MCAR(1) and MCAR(5); sometimes the estimate from MCAR(1) was closer to the truth and sometimes MCAR(5) was closer.

In summary, the spectral method outperforms the MCAR approach. It produces estimates of the conditional relationship that are closer to the true values and have narrower credible intervals, it reliably identifies pairs that are conditionally independent, and it does not require manual selection of a neighborhood structure that can sway the resulting inference (as seen for the MCAR(1) and MCAR(5) models here).

## 5.2 $\mu$ -XRF Data – Spectral Approach

The model described in Section 3.5 was fitted to the  $\mu$ -XRF data with a 10% taper yielding covariance and cross-covariance functions for each of the five soil elements and

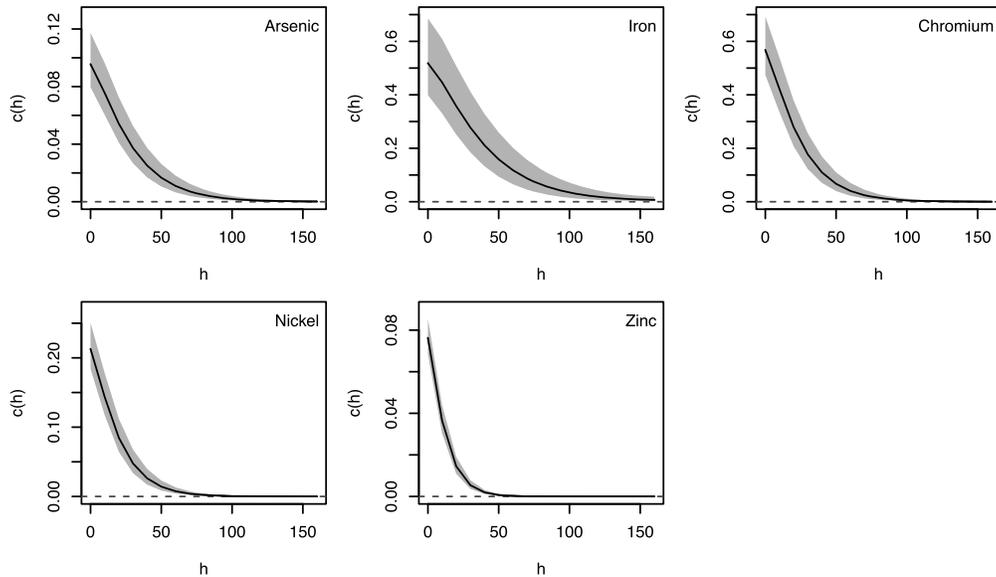


Figure 3: Estimated covariance functions (95% pointwise intervals in grey) for each of the five soil components, plotted for different distances  $h$  ( $\mu\text{m}$ ). A horizontal dashed line indicates a covariance of zero.

the ten soil element pairs. To check for sensitivity, the model was additionally run for a 5% and 15% taper with comparable results (not shown here). The covariance functions are provided in Figure 3 and indicate varying marginal variances and spatial ranges across the soil elements, with iron having the longest spatial dependence and zinc having the shortest. The ability to capture this variability in spatial dependence is a direct outcome of the flexible spectral specification of the model and would be far more difficult to capture in a MRF model with a pre-specified neighborhood structure. In addition, the Bayesian approach provides straightforward descriptions of the uncertainty surrounding these covariance functions, shown here through grey shading representing 95% pointwise credible intervals.

The cross-covariance functions are provided in Figure 4 and similarly illustrate the benefits of this spectral Bayesian modeling approach. In particular, the credible intervals for the cross-covariances between arsenic/nickel, arsenic/zinc and nickel/zinc clearly overlap zero. If one were to plot only the estimated covariance functions, then it would appear that these soil elements are minimally positively correlated with one another. However, the 95% credible intervals enabled by the Bayesian framework make it quite clear that one cannot infer any pairwise relationships between these pairs of elements.

In addition to learning about the pairwise relationships between these soil elements, researchers are particularly interested in learning multi-element effects on the chemical reactivity of an element, i.e., the nature of pairwise relationships when we condition on the presence of other elements in the sample. Using the approach outlined in Sec-

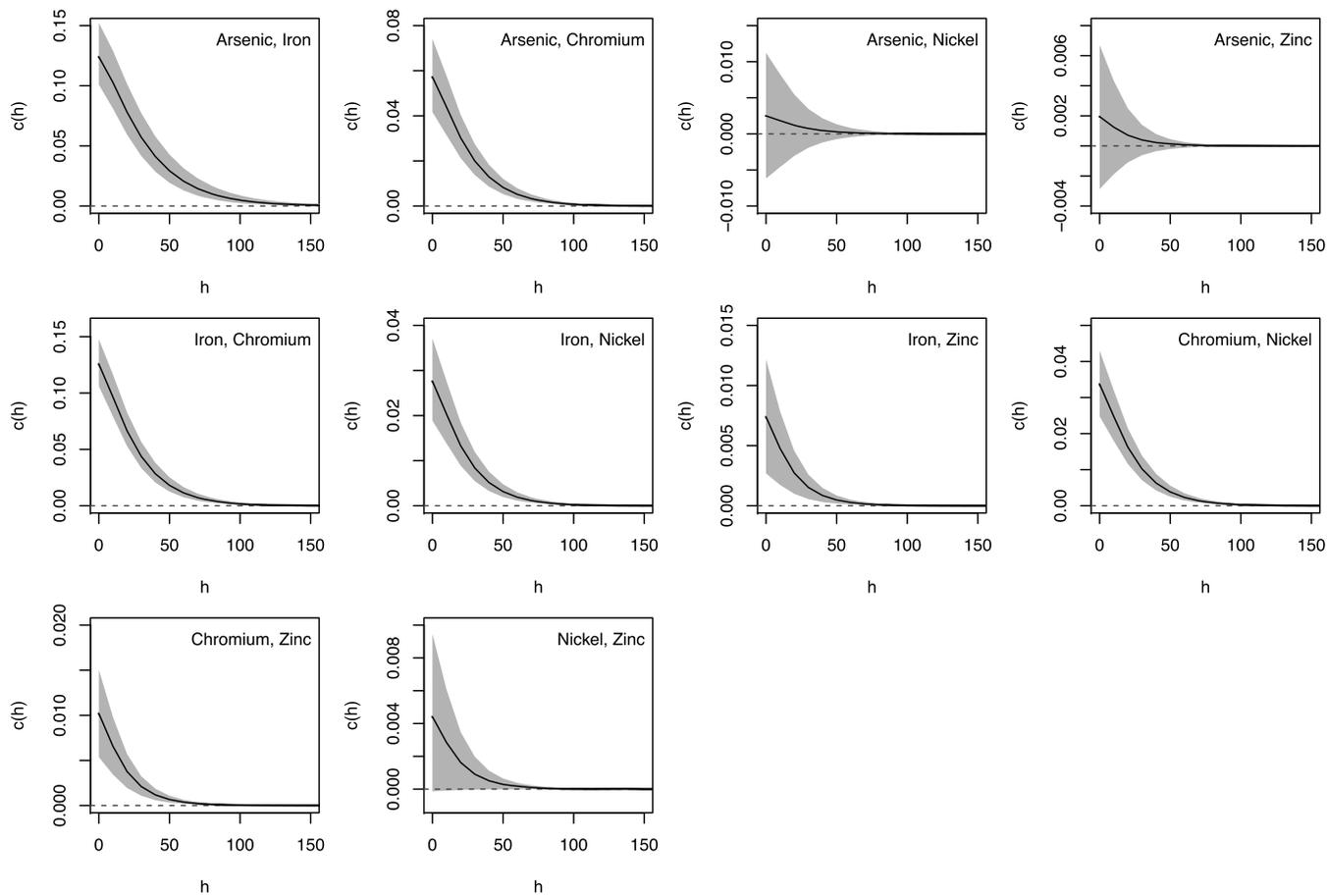


Figure 4: Estimated cross-covariance functions (95% pointwise intervals in grey) for each of the soil component pairs, plotted for different distances  $h$  ( $\mu\text{m}$ ). A horizontal dashed line indicates a cross-covariance of zero.

tion 3.4, for each soil element we consider the “coefficients” that would be placed on the other soil elements in the conditional distribution implied by the fitted joint model. Pseudo posterior estimates and credible intervals for these “coefficients” are provided in Figure 5, with each subplot corresponding to one of the five soil elements. For example, looking at the first subplot one can see that arsenic has a strong relationship with iron and a slight negative relationship with nickel, but does not have a significant relationship with chromium and zinc. In conjunction with Figure 4, we can infer that arsenic and chromium will be positively correlated when examined pairwise, but that this relationship disappears when the other soil elements are accounted for.

To illustrate these conditional relationships more intuitively, a conditional dependence graph is provided in Figure 6. Each of the soil elements is represented by a node, with edges connecting the nodes when there is a positive (indicated by a ‘+’) or negative (indicated by a ‘-’) conditional relationship between the elements. For example, there is no edge connecting arsenic and chromium, suggesting that they are independent conditional on nickel and iron. That is, the observed pairwise correlation between arsenic and chromium can be explained statistically via the relationships each have with nickel and iron.

Finally, as a form of model-checking, we can compare the analysis results for arsenic, iron and chromium to those of Guinness et al. (2014). The covariance functions in Figure 3 are consistent with those found in the previous work, as are the shapes of the cross-covariance functions in Figure 4. This agreement is encouraging and suggests that minimal meaningful flexibility was lost by setting  $\rho(\omega) \equiv \rho$ . Additionally, Guinness et al. (2014) explored the dependence between arsenic and chromium through a hypothesis testing approach and found “strong but not overwhelming evidence for conditional dependence.” In contrast, our analysis suggests that when nickel and zinc are incorporated into the model, the conditional dependence between chromium and arsenic is no longer significant. I.e., not surprisingly, the nature of these conditional dependencies will change depending on the soil elements being accounted for. This highlights the importance of being able to accommodate several variables of interest, previously impossible using the earlier model.

Our model can also be checked with respect to the known chemistry of arsenic. Specifically, we consider the strong positive marginal and conditional relationships detected between arsenic and iron. Scanning electron microscopy – energy-dispersive X-ray (SEM-EDX) analysis indicated that visibly dark spots on the sand-grain sample corresponded with hotspots (Figure 1) that are enriched in iron, chromium, and nickel at a ratio corresponding with stainless steel (SEM-EDX data not shown). We deduced that the contamination was introduced to the sample during mounting with stainless steel forceps. However, the most common form of iron in the sample (Fe(III)-oxides) was natural and has a high capacity to bind arsenic, with the effects of stainless steel adsorption being trivial in comparison. This known binding between the iron and arsenic is consistent with the positive relationship indicated by the model. Moreover, the model was able to separate the negative effect of the minor stainless steel contamination from that of the pervasive, arsenate-adsorbing Fe(III)-oxides

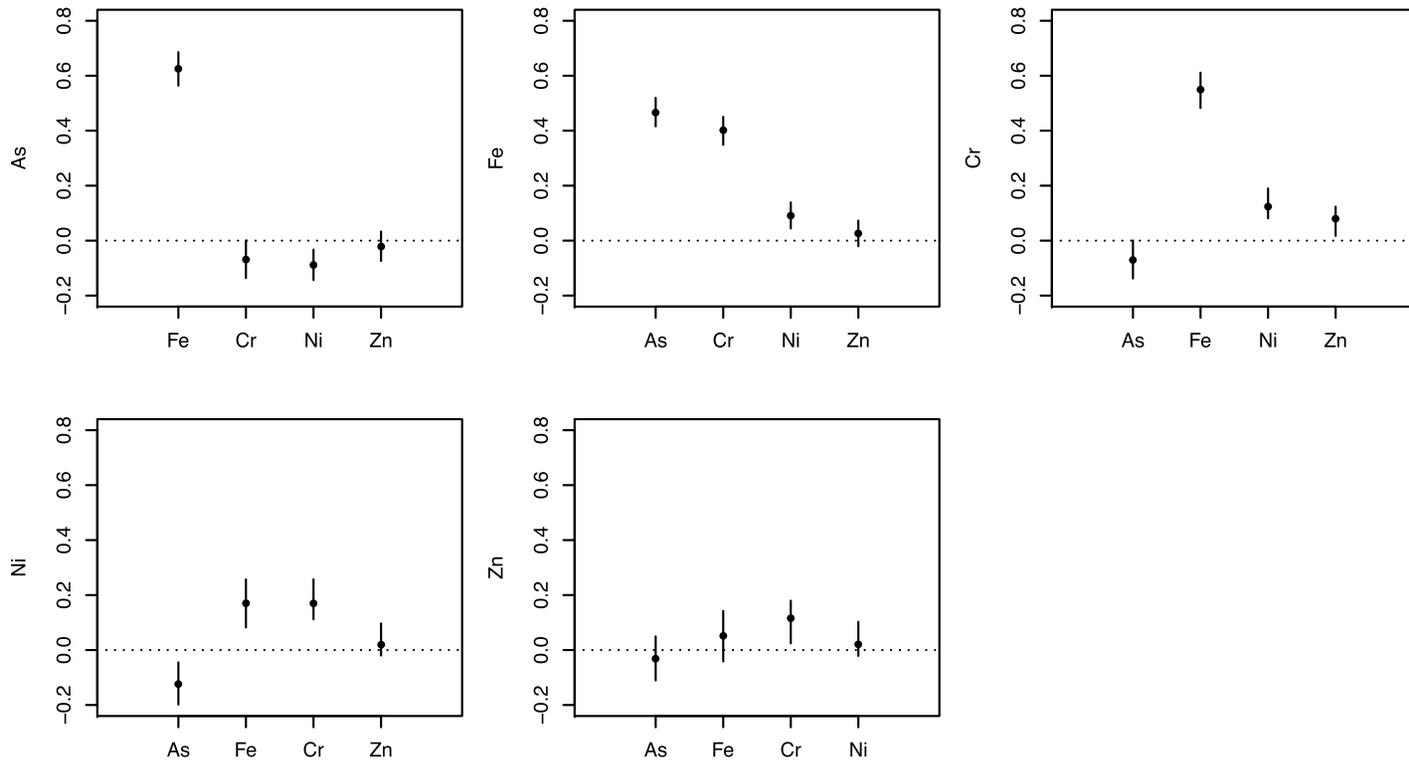


Figure 5: Estimated “coefficients” on  $z_{-j}$  describing conditional dependencies between the soil components, defined to be  $\mathbf{C}_{[j,-j]} \mathbf{C}_{[-j,-j]}^{-1}$ .

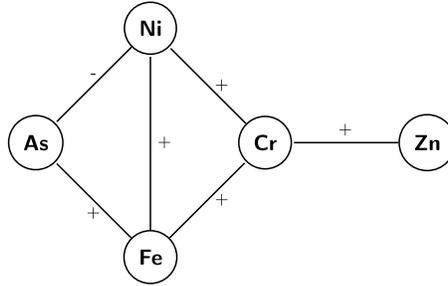


Figure 6: Reduced conditional dependence graph for the five soil components.

### 5.3 $\mu$ -XRF Data – MCAR Approach

For comparative purposes, a separable multivariate conditionally autoregressive break (MCAR) model was additionally fit to these data. As discussed earlier, this is one of the most common Bayesian approaches to modeling areal data due to its structural simplicity and relative computational efficiency. However, this simplicity imposes clear limitations on the model’s flexibility, as we illustrate below.

One of the common selling points for separable MCAR models is their simplicity and fast computation. However, when fitting this model on the soil data some unexpected computational challenges arose. When the model was run without saving the samples of the spatial random effects the computation time was approximately 15 minutes. Of course, the spatial random effects are the main component of this model and should ideally be examined. Unfortunately, when the spatial random effects were included in the list of parameters to return, the code (Terres et al., 2016) ran for over 4 hours and was eventually aborted. This is likely due to the large size of the data, consisting of  $35 \times 42 = 1470$  areal units for each of the five soil elements, equating to 7350 spatial random effects! In contrast, if a researcher were studying a phenomenon across the United States, then a state-level outcome would require only 50 random effects.

The model defined in (13) relies on an explicit choice of neighborhood structure, defined through the  $w_{i,j}$ s, corresponding to an assumed extent of spatial dependence. To test for sensitivity, the MCAR model was fit with two different neighborhood structures. In the first setup any location within  $10 \mu\text{m}$  was identified as a neighbor, setting  $w_{i,j} = 1$ . In the second setup this distance was extended to  $50 \mu\text{m}$ , consistent with some of the spatial dependencies that were seen in the results from the spectral modeling.

The MCAR model differs from the spectral model in several key ways that limit its flexibility as well as our capacity to make comparisons between the two models. The first is the separable nature of the multivariate modeling. There is a matrix  $\Sigma$  that controls the correlation structure between the five soil elements, and this is completely separated from the neighbor-averaging approach to modeling the spatial structure. In contrast, the spectral model discussed earlier is non-separable, allowing for this dependence to differ across different spatial distances. Second, the separable MCAR model requires a pre-defined neighborhood structure that is shared across all soil elements. I.e., not

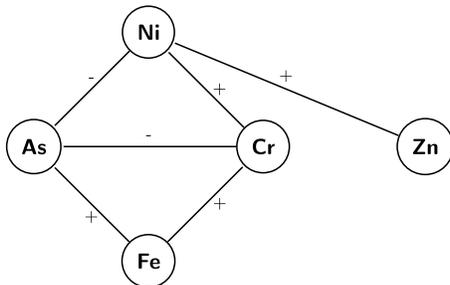


Figure 7: Reduced conditional dependence graph for the five soil components using MCAR with neighbors within distance 1 (10  $\mu\text{m}$ ).

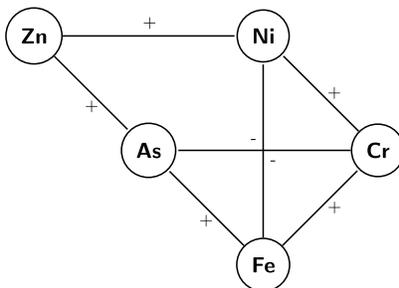


Figure 8: Reduced conditional dependence graph for the five soil components using MCAR with neighbors within distance 5 (50  $\mu\text{m}$ ).

only is the practitioner hand-selecting the distances at which spatial dependence is non-negligible, but this spatial dependence must be the same for all five soil elements. Looking at Figure 1, as well as the results of the spectral model in Figure 3, this is clearly not the case. Some of these assumptions can be relaxed in more complex setups for the MCAR model, but this is at the sacrifice of computational efficiency.

Despite the differences in model structure, the results of the MCAR model enables assessment of conditional dependencies similarly to the spectral approach described in Section 3.4. Here the assessment is made based on the matrix  $\Sigma$ , controlling the covariance between the soil elements, which can be decomposed similarly to the  $\rho$  matrix from the spectral model. This approach allows for construction of graphs similar to 6, shown in Figures 7 and 8 for the MCAR models with first order neighbors and fifth order neighbors respectively. The stark differences between these graphs emphasizes the reliance of this model on the neighborhood structure that, in contrast to the spectral model, is not explicitly learned from the data.

A more formal model comparison can be made between the two MCAR models and the proposed spectral model using the deviance information criterion (DIC). The DIC

is defined as follows,

$$\begin{aligned} DIC &= D(\bar{\theta}) + 2p_D, \\ D(\theta) &= -2 \log p(\mathbf{Z}|\theta), \\ p_D &= \frac{1}{2} \text{Var}_{\theta|\mathbf{Z}}(D(\theta)), \end{aligned}$$

where  $\bar{\theta}$  is the posterior mean of the parameters. The first term in the DIC is the deviance computed using the posterior mean of the parameters and is intended to be a measure of the goodness of fit. The second term is a penalty for the number of parameters in the model. Here the definition of  $p_D$  is based on the posterior variance of the deviance, following the suggestion of Gelman et al. (2014) and Spiegelhalter et al. (2014).

The computed DICs for each model are provided in Table 1. Recalling that lower DIC values indicate a better fit, the MCAR(5) appears to be the worst model for these data and the spectral model appears to be the best model.

Model	DIC	$ DIC - DIC_{best} $
Spectral	-3583	0
MCAR(1)	355636	359219
MCAR(5)	1103108	1106691

Table 1: Comparison of DIC across the three models, ordered from best (lowest DIC) to worst (highest DIC). The third column provides the absolute difference between each model’s DIC and that of the best model.

## 6 Discussion

The framework we have outlined enables joint spatial modeling of multiple spatial processes in a computationally efficient manner through the use of Fourier transformations and spectral analysis theory. A primary advantage of this joint modeling is the ability to make inference on conditional dependencies between the spatial processes while allowing for flexibility in the spatial properties of each process individually. The representation of the covariance structure using spectral densities enables a flexible structure for the covariance functions and cross-covariance functions, with properties such as non-separability and covariate dependence being easily incorporated. The covariance parameters provide inference on the extent of spatial dependence, avoiding specification of a neighborhood matrix required for other methods. By implementing the approach in a Bayesian setup, we can fully quantify the uncertainty associated with our estimates of covariance functions via the pseudo posterior while simultaneously taking advantage of the computationally efficient aspects of spectral modeling. All of this is achieved by utilizing the Whittle likelihood approximation to the multivariate normal likelihood, since the true likelihood may be prohibitively expensive to compute in models with comparable flexibility.

This approach was illustrated for soil elements mapped in thin mineral-organic coatings on the surface of a quartz soil-sand grain using  $\mu$ -XRF analysis, including separating

effects of elements that are known to occur at least partially as non-reactive stainless-steel contamination. Our proposed model provides a framework for assessing the conditional relationships between several soil elements that can have different geochemical effects on arsenic accumulation in soils. Interactive effects between pollutants, multiple minerals and/or organic matter co-localized in soil microsites are difficult to disentangle using current analytical techniques, thereby limiting our predictive capabilities for natural environmental settings.

Understanding how microscale relationships between soil elements affect arsenic binding is one approach for determining the importance of multi-component interactions in soils that are not accounted for by single-component model systems. Specifically, molecular-scale reactions known from model systems control the partitioning of contaminants between soil solids and soil water which, in turn, controls the macroscale impacts of trace-element contaminants. Macroscale reactions may include mobilization into the human food chain or drinking water, which is illustrated by the mass poisoning of humans in southern Asia. The modeling presented here provides an approach to couple microscale spatial data with existing molecular-scale knowledge to infer specific chemical controls on element partitioning. By quantifying these relationships, researchers can then develop better strategies to mitigate human exposure to environmental contaminants. With additional overlapping spatial data sets, the analyses presented here could reasonably be expanded to other soil elements such as aluminum, manganese, and carbon that are also known to impact soil arsenic reactivity.

Unlike our proposed approach, previous work analyzing these data (Guinness et al., 2014) was not able to account for more than three soil elements at a time and also lacked adequate descriptions of uncertainty. Despite using a similar spectral representation for the likelihood, the authors adopted a hypothesis testing framework in order to assess the conditional relationships between soil elements. In their framework each pairwise correlation function was characterized in part through a multiplicative scalar, which indicated conditional independence when set to zero. The model was fit with and without this parameter set to zero, and the resulting log-likelihood ratio was computed and compared to theoretical distribution quantiles. The reliance on a single parameter hypothesis test limited the parameterization of the joint model and prevented extensions to dimensions greater than three. In the case of soil elements, as well as other multivariate problems for which the methodology may be applicable, restricting analysis to three variables is a severe limitation.

In contrast, by approaching the problem in a Bayesian framework one can take advantage of the same computational benefits of the spectral representation and simultaneously make a more thorough analysis of the conditional relationships of interest. Instead of rigidly defining the model around a single parameter of interest, a prior can be put on the correlation matrix allowing for a flexible model structure that can accommodate an arbitrary number of variables. In this setup, inference on the conditional relationships is a natural byproduct of the pseudo posterior inference on the correlation matrix and can be made simultaneously for several variables.

The standard Bayesian approaches to modeling lattice data revolve around MCAR models, specifically separable MCAR models when computational efficiency is a concern.

With these models, as was illustrated in Sections 5.1 and 5.3, the resulting inference on the conditional relationships between variables will be heavily dependent on the neighborhood structure. This neighborhood structure is not explicitly learned from the data; it is instead pre-specified based on an assumed distance of spatial dependence. To select a neighborhood structure from the data, a model comparison approach would be necessary post-model fitting. This is in contrast to methods utilizing a covariance function where a range parameter is fitted to the data along with the other parameters. Since this spatial dependence impacts the resulting inference on the conditional relationships between soil elements, it is convenient for it to be included during the model-fitting.

In a separable MCAR model the spatial behavior and multivariate correlation structures are incorporated into the model somewhat independently – one component of the model controls spatial dependence while another controls the correlation between soil elements. As such, in these models the neighborhood structure is shared across all soil elements. As was seen from the results shown in Figure 3, the assumption of equal spatial ranges does not appear to hold true for the soil data. In addition, the separable MCAR requires a fixed (non-spatial) covariance matrix to describe the dependencies across the soil elements. The assumption of separability simplifies the model structure and reduces computation times, but separable models are often criticized for being unrealistic.

Fundamentally the spectral modeling approach proposed here is very similar to an MCAR model. Both assume a Gaussian distribution to describe a spatial process, relying on parametric forms to describe the spatial dependencies. However, the spectral approach has some advantages over the separable MCAR that is used when there are concerns about computational efficiency. Unlike the MCAR, the spectral approach allows for inference regarding the extent of the spatial dependence across distances, and the nonseparable specification of the covariance structure allows for the spatial dependence to differ across the five soil elements. As was shown here, the MCAR model’s lack of flexibility in these regards can greatly affect posterior inference regarding conditional relationships between the elements.

Several areas for future work can be suggested by the proposed methodology. In its current form the model is appropriate for jointly modeling multiple spatial variables observed on a complete lattice. However, future adaptations can be envisioned to accommodate such spatial processes that are non-stationary (Fuentes, 2002), and/or were collected on an incomplete lattice (Fuentes, 2007; Stroud et al., 2014; Guinness and Fuentes, 2016). Specifically, recent papers have introduced methodology for imputation at unobserved lattice locations for a univariate spatial process. However, no such methodology exists for the multivariate case explored in this paper. In addition to enabling modeling when there is missing data, such imputation methods would enable posterior predictive checks to assess the extent of the spatial correlation that was captured by the multivariate model.

Finally, future work could add even more flexibility to the model structure. Flexibility to model more complex dependence structures could be achieved through the introduction of a nonparametric spectral density. Additionally, in the current paper the correlation matrix  $\rho$  was assumed to be constant across frequencies in the above

analyses, enabling straight forward sampling and interpretation of the conditional relationships. In the future this assumption may be relaxed in order to achieve a more flexible model structure.

## Supplementary Material

Supplementary Material of “Bayesian Spectral Modeling for Multivariate Spatial Distributions of Elemental Concentrations in Soil” (DOI: [10.1214/16-BA1034SUPP](https://doi.org/10.1214/16-BA1034SUPP); .zip).

## References

- Apanasovich, T. V., Genton, M. G., and Sun, Y. (2012). “A valid Matérn class of cross-covariance functions for multivariate random fields with any number of components.” *Journal of the American Statistical Association*, 107(497): 180–193. [MR2949350](https://doi.org/10.1080/01621459.2011.643197). doi: <https://doi.org/10.1080/01621459.2011.643197>. 10
- Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014). *Hierarchical Modeling and Analysis for Spatial Data*. CRC Press. [MR3362184](https://doi.org/10.1201/b16888). 5
- Banerjee, S., Gelfand, A. E., Finley, A. O., and Sang, H. (2008). “Gaussian predictive process models for large spatial data sets.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(4): 825–848. [MR2523906](https://doi.org/10.1111/j.1467-9868.2008.00663.x). doi: <https://doi.org/10.1111/j.1467-9868.2008.00663.x>. 1
- Barnard, J., McCulloch, R., and Meng, X.-L. (2000). “Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage.” *Statistica Sinica*, 10(4): 1281–1312. [MR1804544](https://doi.org/10.1214/00-SS104). 11, 12
- Besag, J. (1974). “Spatial interaction and the statistical analysis of lattice systems.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 192–236. [MR0373208](https://doi.org/10.1111/j.1467-9868.1974.tb003208.x). 2
- Bloomfield, P. (2004). *Fourier Analysis of Time Series: An Introduction*. John Wiley & Sons. [MR1884963](https://doi.org/10.1002/0471722235). doi: <https://doi.org/10.1002/0471722235>. 8
- Borch, T., Kretzschmar, R., Kappler, A., Cappellen, P. V., Ginder-Vogel, M., Voegelin, A., and Campbell, K. (2009). “Biogeochemical redox processes and their impact on contaminant dynamics.” *Environmental Science & Technology*, 44(1): 15–23. [MR2523906](https://doi.org/10.1021/es80203a001). 3
- Cressie, N. and Wikle, C. K. (2011). *Statistics for Spatio-Temporal Data*. John Wiley & Sons. [MR2848400](https://doi.org/10.1002/9781118134463). 5
- Dahlhaus, R. (1983). “Spectral analysis with tapered data.” *Journal of Time Series Analysis*, 4(3): 163–175. [MR0732895](https://doi.org/10.1111/j.1467-9892.1983.tb00366.x). doi: <https://doi.org/10.1111/j.1467-9892.1983.tb00366.x>. 8
- Dahlhaus, R. and Künsch, H. (1987). “Edge effects and efficient parameter estimation for stationary random fields.” *Biometrika*, 74(4): 877–882. [MR0919857](https://doi.org/10.1093/biomet/74.4.877). doi: <https://doi.org/10.1093/biomet/74.4.877>. 8

- Datta, A., Banerjee, S., Finley, A. O., and Gelfand, A. E. (2014). “Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets.” [arXiv:1406.7343](https://arxiv.org/abs/1406.7343). MR3538706. doi: <https://doi.org/10.1080/01621459.2015.1044091>. 2
- Fuentes, M. (2002). “Spectral methods for nonstationary spatial processes.” *Biometrika*, 89(1): 197–210. MR1888368. doi: <https://doi.org/10.1093/biomet/89.1.197>. 23
- Fuentes, M. (2007). “Approximate likelihood for large irregularly spaced spatial data.” *Journal of the American Statistical Association*, 102(477): 321–331. MR2345545. doi: <https://doi.org/10.1198/016214506000000852>. 7, 23
- Gelfand, A. E. and Vounatsou, P. (2003). “Proper multivariate conditional autoregressive models for spatial data analysis.” *Biostatistics*, 4(1): 11–15. 2
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2014). *Bayesian Data Analysis*, volume 2. Taylor & Francis. 21
- Guinness, J. and Fuentes, M. (2016). “Circulant embedding of approximate covariances for inference from Gaussian data on large lattices.” *Journal of Computational and Graphical Statistics*. doi: <https://doi.org/10.1080/10618600.2016.1164534>. 8, 23
- Guinness, J., Fuentes, M., Hesterberg, D., and Polizzotto, M. (2014). “Multivariate spatial modeling of conditional dependence in microscale soil elemental composition data.” *Spatial Statistics*, 9: 93–108. MR3326833. 4, 5, 7, 9, 10, 11, 17, 22
- Guyon, X. (1982). “Parameter estimation for a stationary process on a d-dimensional lattice.” *Biometrika*, 69(1): 95–105. MR0655674. doi: <https://doi.org/10.1093/biomet/69.1.95>. 7, 8
- Handcock, M. S. and Stein, M. L. (1993). “A Bayesian analysis of kriging.” *Technometrics*, 35(4): 403–410. 3
- Hoff, P. D. (2007). “Extending the rank likelihood for semiparametric copula estimation.” *The Annals of Applied Statistics*, 1(1): 265–283. MR2393851. doi: <https://doi.org/10.1214/07-AOAS107>. 11
- Hoff, P. D. (2010). *A First Course in Bayesian Statistical Methods*. New York: Springer Dordrecht Heidelberg London New York. MR2648134. doi: <https://doi.org/10.1007/978-0-387-92407-6>. 11
- Jiang, J.-Q., Ashekuzzaman, S., Jiang, A., Sharifuzzaman, S., and Chowdhury, S. R. (2012). “Arsenic contaminated groundwater and its treatment options in Bangladesh.” *International Journal of Environmental Research and Public Health*, 10(1): 18–46. 3
- Jin, X., Banerjee, S., and Carlin, B. P. (2007). “Order-free co-regionalized areal data models with application to multiple-disease mapping.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(5): 817–838. MR2368572. doi: <https://doi.org/10.1111/j.1467-9868.2007.00612.x>. 2

- Jin, X., Carlin, B. P., and Banerjee, S. (2005). “Generalized hierarchical multivariate CAR models for areal data.” *Biometrics*, 61(4): 950–961. MR2216188. doi: <https://doi.org/10.1111/j.1541-0420.2005.00359.x>. 2
- Kim, H.-M., Mallick, B. K., and Holmes, C. (2005). “Analyzing nonstationary spatial data using piecewise Gaussian processes.” *Journal of the American Statistical Association*, 100(470): 653–668. MR2160567. doi: <https://doi.org/10.1198/016214504000002014>. 2
- Komárek, M., Vaněk, A., and Ettler, V. (2013). “Chemical stabilization of metals and arsenic in contaminated soils using oxides – a review.” *Environmental Pollution*, 172: 9–22. 3
- Koopmans, L. H. (1995). *The Spectral Analysis of Time Series*. Academic Press. MR1425881. 3
- Lindgren, F., Rue, H., and Lindström, J. (2011). “An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4): 423–498. MR2853727. doi: <https://doi.org/10.1111/j.1467-9868.2011.00777.x>. 2
- Manceau, A., Marcus, M., and Lenoir, T. (2014). “Estimating the number of pure chemical components in a mixture by X-ray absorption spectroscopy.” *Journal of Synchrotron Radiation*, 21(5): 1140–1147. 3
- Mardia, K. (1988). “Multi-dimensional multivariate Gaussian Markov random fields with application to image processing.” *Journal of Multivariate Analysis*, 24(2): 265–284. MR0926357. doi: [https://doi.org/10.1016/0047-259X\(88\)90040-1](https://doi.org/10.1016/0047-259X(88)90040-1). 2
- Polizzotto, M. L., Kocar, B. D., Benner, S. G., Sampson, M., and Fendorf, S. (2008). “Near-surface wetland sediments as a source of arsenic release to ground water in Asia.” *Nature*, 454(7203): 505–508. 3
- Priestley, M. B. (1981). *Spectral Analysis and Time Series*, Vol. 2, Academic Press. 3
- Ravenscroft, P., Brammer, H., and Richards, K. (2009). *Arsenic Pollution: A Global Synthesis*, volume 28. John Wiley & Sons. 3
- Reich, B. J. and Fuentes, M. (2012). “Nonparametric Bayesian models for a spatial covariance.” *Statistical Methodology*, 9(1): 265–274. MR2863613. doi: <https://doi.org/10.1016/j.stamet.2011.01.007>. 3
- Ritter, C. and Tanner, M. A. (1992). “Facilitating the Gibbs sampler: the Gibbs stopper and the griddy-Gibbs sampler.” *Journal of the American Statistical Association*, 87(419): 861–868. 12
- Sain, S. R. and Cressie, N. (2007). “A spatial model for multivariate lattice data.” *Journal of Econometrics*, 140(1): 226–259. MR2395923. doi: <https://doi.org/10.1016/j.jeconom.2006.09.010>. 2

- Sain, S. R., Furrer, R., Cressie, N., et al. (2011). “A spatial analysis of multivariate output from regional climate models.” *The Annals of Applied Statistics*, 5(1): 150–175. MR2810393. doi: <https://doi.org/10.1214/10-AOAS369>. 2
- Sang, H. and Huang, J. Z. (2012). “A full scale approximation of covariance functions for large spatial data sets.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(1): 111–132. MR2885842. doi: <https://doi.org/10.1111/j.1467-9868.2011.01007.x>. 2
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Linde, A. (2014). “The deviance information criterion: 12 years on.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(3): 485–493. MR3210727. doi: <https://doi.org/10.1111/rssb.12062>. 21
- Stein, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer Science & Business Media. MR1697409. doi: <https://doi.org/10.1007/978-1-4612-1494-6>. 3, 5, 8
- Stein, M. L. (2014). “Limitations on low rank approximations for covariance matrices of spatial data.” *Spatial Statistics*, 8: 1–19. MR3326818. doi: <https://doi.org/10.1016/j.spasta.2013.06.003>. 2
- Stein, M. L., Chi, Z., and Welty, L. J. (2004). “Approximating likelihoods for large spatial data sets.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(2): 275–296. MR2062376. doi: <https://doi.org/10.1046/j.1369-7412.2003.05512.x>. 8
- Stroud, J. R., Stein, M. L., and Lysen, S. (2014). “Bayesian and Maximum Likelihood Estimation for Gaussian Processes on an Incomplete Lattice.” [arXiv:1402.4281](https://arxiv.org/abs/1402.4281). 3, 7, 8, 23
- Terres, M. A., Fuentes, M., Hesterberg, D., and Polizzotto, M. (2016). “Supplementary material of “Bayesian spectral modeling for multivariate spatial distributions of elemental concentrations in soil”.” *Bayesian Analysis*. doi: <https://doi.org/10.1214/16-BA1034SUPP>. 19
- Tukey, J. W. (1967). “An introduction to the calculations of numerical spectrum analysis.” *Spectral Analysis of Time Series*, 25. MR0212977. 8
- Whittle, P. (1954). “On stationary processes in the plane.” *Biometrika*, 434–449. MR0067450. 3, 5, 7, 8
- Yaglom, A. (1987). *Correlation Theory of Stationary and Related Random Functions: Vol. 1: Basic Results*. Springer-Verlag. MR0893393. 6
- Zhang, H. (2004). “Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics.” *Journal of the American Statistical Association*, 99(465): 250–261. MR2054303. doi: <https://doi.org/10.1198/016214504000000241>. 9
- Zhang, Y., Hodges, J. S., and Banerjee, S. (2009). “Smoothed ANOVA with spatial effects as a competitor to MCAR in multivariate spatial smoothing.” *The*

*Annals of Applied Statistics*, 3(4): 1805. MR2752159. doi: <https://doi.org/10.1214/09-AOAS267>. 2

Zimmerman, D. L. (1989). “Computationally exploitable structure of covariance matrices and generalized covariance matrices in spatial models.” *Journal of Statistical Computation and Simulation*, 32(1–2): 1–15. MR1006159. doi: <https://doi.org/10.1080/00949658908811149>. 5

#### **Acknowledgments**

The authors thank Joseph Guinness for helpful discussions regarding the spectral analysis theory. Terres and Fuentes were supported by the National Science Foundation’s Research Network for Statistical Methods for Atmospheric and Oceanic Sciences, award number 1107046. Hesterberg and Polizzotto were supported by National Science Foundation Award number 1349374.