Multivariate Spatial Modeling of Conditional Dependence in Microscale Soil Elemental Composition Data

Joseph Guinness[a], Montserrat Fuentes[a], Dean Hesterberg[b], Matthew Polizzotto[b]

[a] *NC State University, Department of Statistics*
[b] *NC State University, Department of Soil Science*

Abstract

The mobility and environmental impacts of toxic trace elements are regulated by their reactions with soils, which are complex heterogeneous mixtures of minerals and organic matter. We describe an experiment that maps the composition of elements on an individual soil sand grain using X-ray fluorescence microprobe analyses, after the grain is treated with arsenic solutions, resulting in multivariate spatial lattice maps of elemental abundance. To understand the behavior of arsenic in soils, it is important to disentangle the complex multivariate relationships among the elements in the sample. The abundance of most elements, including arsenic, correlates strongly with that of iron, but conditional on the amount of iron, some elements mitigate or potentiate the accumulation of arsenic. This problem motivates our work to define conditional correlation in spatial lattice models and give general conditions under which two components are conditionally uncorrelated given the rest. We describe how to enforce that two components are conditionally uncorrelated given a third in parametric models, which provides a basis for likelihood ratio tests for conditional correlation between arsenic and chromium given iron. We show how to apply our results to big datasets using the Whittle likelihood, and we demonstrate through simulation that tapering improves Whittle likelihood parameter estimates governing cross covariance.

Keywords: Multivariate, high-resolution data, elemental composition, conditional dependence

1. Introduction

The discharge of toxic trace elements such as arsenic, mercury, lead, and chromium into the environment from natural and anthropogenic sources may contaminate soils, food, and water. Toxic trace elements are regulated in part by soils, which are complex, heterogeneous mixtures of minerals, organic matter, and living organisms (Brown and Sturchio, 2002; Hesterberg et al., 2011). Analytical techniques based on synchrotron X-ray absorption spectroscopy (XAS) have emerged as powerful tools for determining the chemical forms (speciation) and spatial distribution of chemical elements in soils and other geochemical systems (Hayes et al., 1987; Brown and Sturchio, 2002). New synchrotron X-ray facilities are being designed for faster data collection and aim to achieve spatial resolutions of 1 to 10 nm (Fitts and Thieme, 2012). Although scientists are using these and other highly sophisticated techniques to collect data on soil trace elements, the multicomponent complexity of the geochemical systems hinders translation of the data into a mechanistic understanding of chemical processes that impact the environment. To date, advanced spatial statistical models have not been developed and applied to geochemical systems, limiting the utility and interpretation of the complex data that are being collected. Consequently, new methodologies are needed to keep pace with the rapid advancements in synchrotron X-ray technology that will produce unprecedented amounts of data. The conditional correlation approach developed here is valuable for characterizing spatially heterogeneous composition data from a wide range of soils and other geochemical material and from a variety of microscale analytical techniques.

The objective of this research is to develop statistical models and methods for making inference about multivariate relationships in large spatially correlated lattice data, such as the sand grain data described in Section 2. When the data consist of a number of multivariate observations that can be assumed to be independent, the various multivariate dependence relationships can by studied by computing partial correlation coefficients and by specifying and fitting more complex graphical models (Edwards, 2000). However, the presence of spatial correlation violates the independence assumption. Spectral analysis provides a natural framework for defining and studying the properties of multivariate spatial models (Yaglom, 1987; Gneiting et al., 2010). We define the notion of conditional correlation in spatial lattice models and prove that components j and k are conditionally uncorrelated given the rest of the components if and only if the inverses of the spectral density matrices contain zeroes in the jk'th entries at al-

most every frequency. This is an extension to multiple dimensions of results by Dahlhaus (2000) on conditional correlation in multivariate time series. Dahlhaus (2000) derived asymptotic distributions of statistics based on the inverses of the empirical cross spectral density matrices. In the spatial case, it is not clear how to derive asymptotic distributions for test statistics due to the nontrivial edge effects associated with periodogram estimates of the spectral density when the dimension of the domain is greater than 1 (Guyon, 1982; Dahlhaus and Künsch, 1987). For this reason, we develop a parametric framework for multivariate spatial data with three components and show how to constrain the model to force two of the components to be conditionally uncorrelated given the third. This framework allows for the implementation of a likelihood ratio test for conditional correlation, and it provides a way to model misalignment among the components of the multivariate spatial data.

We demonstrate in a simulation study that the Whittle likelihood (Whittle, 1954) may be used to estimate parameters when the edge effects are controlled with tapering. The Whittle likelihood may be computed efficiently for large datasets and does not suffer from the memory bottleneck associated with storing large covariance matrices. We apply our methods to the multielement spatial data to be described in more detail in Section 2. This dataset contains 4410 observations, so exact maximum likelihood procedures may be applied directly to obtain estimates, which we compare to those found with the Whittle likelihood. We find that arsenic and chromium are significantly conditionally correlated given iron. Further, the correlation is negative, meaning that conditional on the amount of iron present, increased amounts of chromium in this sample tend to mitigate the accumulation of arsenic. This relationship would not be apparent using more traditional methods of analysis–such as elemental correlation and pairwise scatter plots–commonly used by geochemists for similar datasets, indicating that the approach employed here may help uncover new insights into geochemical mechanisms of trace-element binding in soils.

The rest of the paper is organized as follows: Section 2 provides a detailed description of the data collection procedures. In Section 3, we review the spectral representation of stationary multivariate spatial processes, which lends itself to a succinct definition of conditional correlation in terms of the cross spectral density matrices. We also provide parametric forms for spectral densities and coherence functions and show how to force components to be conditionally uncorrelated given the rest. The multivariate spatial Gaussian likelihood is presented, along with the Whittle likelihood approximation. In Section 4, we describe the results of a simulation to study the

behavior of Whittle likelihood estimates of parameters governing the cross covariance functions, and we investigate the role of tapering. Section 5 presents the results of our statistical analyses of the μ-XRF data, and we conclude with a discussion of our findings in Section 6

2. Description of X-ray Fluorescence Data

XAS is an element-specific technique that can identify the spatial location, oxidation state, and local molecular coordination environment of specific elements in a sample. The fluorescence signal arises from excitation of core electrons in all atoms of a given element as selected by the X-ray energy (Kelly et al., 2008). A high-intensity X-ray beam produced by a synchrotron can be focused to a (sub)micron-scale spot size to map the abundance of multiple elements in a multicomponent sample (Sutton and Rivers, 1999). We used synchrotron micro X-ray fluorescence (μ-XRF) to simultaneously collect $420 \times 350\,\mu\text{m}^2$ maps of 11 elements (arsenic, iron, chromium, manganese, calcium, titanium, vanadium, nickel, copper, zinc, gallium, and germanium) from a soil sand grain after chemical treatments with sodium arsenite solutions. The maps were collected at Beamline X27A at the National Synchrotron Light Source at Brookhaven National Laboratory. The soil sand grain was rastered by an approximately $10 \times 10\,\mu\text{m}^2$ focused X-ray beam, producing maps comprised of 42×35 pixels. Resulting fluorescence X-rays were measured at each pixel to simultaneously detect the concentration of elements within the irradiated volume of soil material at that pixel. Based on grain composition, beam flux, and the incident X-ray energy, measurements of the targeted elements were largely constrained to the thin mineral-organic coating on the outer portion of a quartz sand grain. Here, we focus only on the iron, arsenic, and chromium fluorescence data collected after the final arsenic treatment, which showed the strongest correlation. Spatial maps plotted in Figure 1 show that each element exhibited positive spatial correlation at short distances.

The X-ray fluorescence signals of most elements analyzed (on the log scale) were strongly correlated with the fluorescence signal of iron, as evidenced by the first two pairwise scatter plots in Figure 2. The strong correlation with iron is likely due to the high abundance of iron relative to trace elements, and to the high binding affinity of many trace elements to iron (and aluminum) oxide minerals (McBride, 1989). The third plot shows that arsenic and chromium are positively correlated with each other as well. For understanding arsenic behavior in soils, it is of interest to know whether this positive correlation between arsenic and chromium is due simply to the

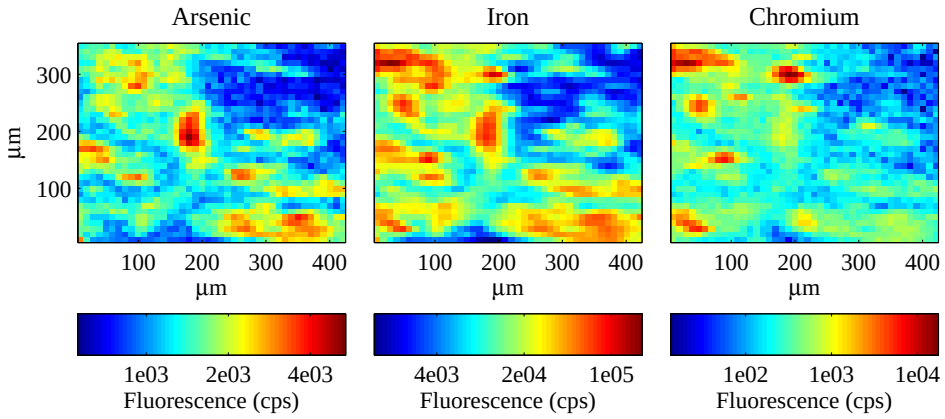

Figure 1: X-ray fluorescence maps measured in counts per second (cps) at each pixel for arsenic, iron, and chromium

fact that these elements are both correlated with iron, or whether there is some dependence between arsenic and chromium that cannot be explained by iron. The first plot in Figure 2 provides insight to this question. It seems clear that for the largest values of iron fluorescence, the fluorescence of arsenic depends nontrivially on that of chromium; greater amounts of chromium tend to decrease the accumulation of arsenic at those locations in the sample. Such a relationship could be indicative of a yet-unknown arsenic oxidation or binding mechanism or reflect differences in iron mineralogy across the sample, so it is important to provide methods for testing the hypothesis of conditional independence under spatial correlation.

3. Multivariate spatial model

In this section we define the statistical models that we apply to the multivariate spatial elemental data discussed in Section 2. Let $\boldsymbol{x} \in \mathbb{R}^d$ and denote by $\boldsymbol{Z}(\boldsymbol{x}) = (Z_1(\boldsymbol{x}), \ldots, Z_p(\boldsymbol{x}))^T$ a multivariate spatial random process on $\mathbb{R}^d$ with $p \in \mathbb{N}$ components. Our μ-XRF data has $p = 3$ components in $d = 2$ dimensions, and $Z_1(\boldsymbol{x})$ refers to the fluorescence of arsenic at location $\boldsymbol{x}$, $Z_2(\boldsymbol{x})$ to that of iron, and $Z_3(\boldsymbol{x})$ to that of chromium. If we assume that $\boldsymbol{Z}$ is a Gaussian process, then its distribution is determined by the mean of each individual observation, $E(Z_j(\boldsymbol{x}))$, and the covariance between any two pairs of observations, $\mathrm{Cov}(Z_j(\boldsymbol{x}), Z_k(\boldsymbol{y}))$. Throughout, we assume that $\mathrm{Var}(Z_j(\boldsymbol{x})) < \infty$. The covariance must satisfy the positive definiteness condition, which states that for any $n \in \mathbb{N}$, constants $a_1, \ldots, a_n$, and

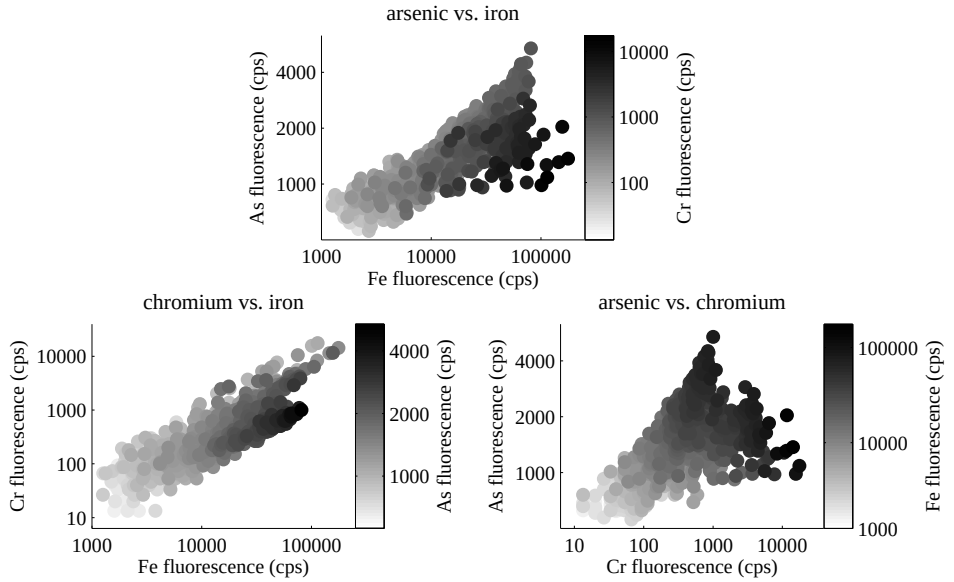

Figure 2: Scatter plots of X-ray fluorescence for all three pairwise relationships among the three elements. Plotting symbol color refers to the fluorescence of the third element, and points are plotted in the order of increasing fluorescence of the third element, so that larger values of the third element cover smaller values.

observations $Z_{j_1}(\boldsymbol{x}_1), \ldots, Z_{j_n}(\boldsymbol{x}_n)$, the variance of the linear combination

$$\mathrm{Var}\left(\sum_{i=1}^{n} a_i Z_{j_i}(\boldsymbol{x}_i)\right) = \sum_{i,k=1}^{n} a_i a_k \mathrm{Cov}(Z_{j_i}(\boldsymbol{x}_i), Z_{j_k}(\boldsymbol{x}_k)) > 0.$$

Often, we assume that the mean and the covariance function are stationary, meaning that $E(Z_j(\boldsymbol{x})) = \mu_j$, and $\mathrm{Cov}(Z_j(\boldsymbol{x}), Z_k(\boldsymbol{y})) = C_{jk}(\boldsymbol{x} - \boldsymbol{y})$, so that the mean function is constant over space for each component, and the covariance function for components j and k depends only on the difference between the two locations.

3.1. Spectral Representation

A stationary multivariate spatial covariance function C_{jk} can be expressed in general as an inverse Fourier transform,

$$C_{jk}(\boldsymbol{x}) = \int_{\mathbb{R}^d} \exp(i\boldsymbol{\omega}'\boldsymbol{x}) dG_{jk}(\boldsymbol{\omega}). \tag{1}$$

We will assume throughout that there exists $g_{jk}(\boldsymbol{\omega})$ for which $dG_{jk}(\boldsymbol{\omega}) = g_{jk}(\boldsymbol{\omega})d\boldsymbol{\omega}$, so (1) becomes

$$C_{jk}(\boldsymbol{x}) = \int_{\mathbb{R}^d} \exp(i\boldsymbol{\omega}'\boldsymbol{x}) g_{jk}(\boldsymbol{\omega}) d\boldsymbol{\omega}, \tag{2}$$

in which case we have the inversion formula

$$g_{jk}(\boldsymbol{\omega}) = \int_{\mathbb{R}^d} \exp(-i\boldsymbol{\omega}'\boldsymbol{x}) C_{jk}(\boldsymbol{x}) d\boldsymbol{x}.$$

In many applications, such as for the μ-XRF data we consider here, we are interested in the process only on a lattice with spacing $\delta \in \mathbb{R}$, that is, $Z_j(\delta\boldsymbol{x})$, $\boldsymbol{x} \in \mathbb{Z}^d$. In this case, the function $\exp(i\boldsymbol{\omega}'(\delta\boldsymbol{x}))$ (as a function of $\boldsymbol{x} \in \mathbb{Z}^d$) is aliased with $\exp(i(\boldsymbol{\omega} + 2\pi\boldsymbol{j}/\delta)'(\delta\boldsymbol{x}))$ for any $\boldsymbol{j} \in \mathbb{Z}^d$. Thus we consider only the aliased spectral density

$$f_{jk}(\boldsymbol{\omega}) = \sum_{\boldsymbol{j} \in \mathbb{Z}^d} g_{jk}(\boldsymbol{\omega} + 2\pi\boldsymbol{j}/\delta),$$

for $\boldsymbol{\omega} \in [-\pi/\delta, \pi/\delta]^d$. The covariance function at lag $\delta\boldsymbol{x}$ can then be expressed as

$$C_{jk}(\delta\boldsymbol{x}) = \int_{[-\pi/\delta, \pi/\delta]^d} \exp(i\boldsymbol{\omega}'(\delta\boldsymbol{x})) f_{jk}(\boldsymbol{\omega}) d\boldsymbol{\omega}, \tag{3}$$

with inversion formula

$$f_{jk}(\boldsymbol{\omega}) = \left(\frac{2\pi}{\delta}\right)^d \sum_{\boldsymbol{x}\in\mathbb{Z}^2} C_{jk}(\delta\boldsymbol{x}) \exp(-i\boldsymbol{\omega}'(\delta\boldsymbol{x})).$$

The conditions for positive definiteness of $C_{jk}(\boldsymbol{x})$ under representation (3) are that the $p \times p$ matrix $\boldsymbol{f}(\boldsymbol{\omega})$ with (j,k)th entry $f_{jk}(\boldsymbol{\omega})$ be positive definite everywhere on $[-\pi/\delta, \pi/\delta]^d$ except possibly on sets of Lebesgue measure zero. The X-ray fluorescence data we consider here are always observed on a lattice, where we define the lattice spacing to be $\delta = 1$, so usually we can ignore δ in the notation.

3.2. Conditional Correlation

In multivariate modeling, it is often natural to consider the conditional dependence structure of the various components. For example, our μ-XRF data elicit the question of whether arsenic and chromium are conditionally uncorrelated given iron. In general, one may be interested in the conditional correlation between components i and j in the p-variate spatial process. For simplicity here, we assume that $\mu_j = 0$ for every j, although the results generalize easily to non-zero means. To understand how to define and specify models for which components are conditionally uncorrelated, we define $\boldsymbol{Z}_{-ij}(\boldsymbol{x}) = (Z_k(\boldsymbol{x}) : k \neq i, j)$ and the residual processes

$$\varepsilon_{ij}(\boldsymbol{x}) = Z_i(\boldsymbol{x}) - \sum_{\boldsymbol{u}\in\mathbb{Z}^d} a_{ij}(\boldsymbol{x}-\boldsymbol{u})^T \boldsymbol{Z}_{-ij}(\boldsymbol{u}),$$
$$\varepsilon_{ji}(\boldsymbol{x}) = Z_j(\boldsymbol{x}) - \sum_{\boldsymbol{u}\in\mathbb{Z}^d} a_{ji}(\boldsymbol{x}-\boldsymbol{u})^T \boldsymbol{Z}_{-ij}(\boldsymbol{u}),$$

where $a_{ij}(\boldsymbol{u})$ and $a_{ji}(\boldsymbol{u})$ are the optimal $(p-2) \times 1$ linear filters chosen to minimize $\mathrm{Var}(\varepsilon_{ij}(\boldsymbol{x}))$ and $\mathrm{Var}(\varepsilon_{ji}(\boldsymbol{x}))$. In other words, $a_{ij}(\cdot)$ and $a_{ji}(\cdot)$ define the best linear unbiased predictors of $Z_i(\boldsymbol{x})$ and $Z_j(\boldsymbol{x})$ given $\boldsymbol{Z}_{-ij}(\cdot)$. Then we define the conditional covariance between components i and j given the rest as

$$R_{ij}(\boldsymbol{u}) = \mathrm{Cov}(\varepsilon_{ij}(\boldsymbol{x}+\boldsymbol{u}), \varepsilon_{ji}(\boldsymbol{x})),$$

and say that components i and j are conditionally uncorrelated given the rest if $R_{ij}(\boldsymbol{u}) = 0$ for every $\boldsymbol{u} \in \mathbb{Z}^d$. This is an extension to more than one dimension of the definition given in Dahlhaus (2000), and in the case of no spatial correlation, this definition reduces to the usual definition of conditional correlation.

Stationary multivariate process models with this property are most easily specified in the spectral domain, as seen in the following theorem.

Theorem 1. *Components i and j are conditionally uncorrelated given the rest of the components if and only if $\boldsymbol{f}^{-1}(\boldsymbol{\omega})$ has a zero in the (i,j)'th entry for almost every $\boldsymbol{\omega}$.*

A proof of this result is given in the appendix. In the case of $p = 3$ components, which is of interest for our μ-XRF elemental data, it is straightforward to place conditions on the entries of the (3×3) matrix $\boldsymbol{f}(\boldsymbol{\omega})$ in order to guarantee a zero in the inverse. Specifically, $\boldsymbol{f}^{-1}(\boldsymbol{\omega})$ has a zero in entry $(1,3)$ if and only if

$$f_{12}(\boldsymbol{\omega})f_{23}(\boldsymbol{\omega}) = f_{22}(\boldsymbol{\omega})f_{13}(\boldsymbol{\omega}), \tag{4}$$

and we show in Subsection 3.3 how to enforce this constraint when the spectral density matrices are specified parametrically.

3.3. Specific Model for $\boldsymbol{f}$

A general specification for $f_{jk}(\boldsymbol{\omega})$ that guarantees positive definite $\boldsymbol{f}(\boldsymbol{\omega})$ is

$$f_{jk}(\boldsymbol{\omega}) = \sqrt{s_j(\boldsymbol{\omega})s_k(\boldsymbol{\omega})}\rho_{jk}(\boldsymbol{\omega}), \tag{5}$$

where $s_j(\boldsymbol{\omega}), s_k(\boldsymbol{\omega}) > 0$ and integrable, and $\rho_{jk}(\boldsymbol{\omega}) = |\rho_{jk}(\boldsymbol{\omega})| \exp(i\phi_{jk}(\boldsymbol{\omega}))$ is the (j,k)th entry of the coherence matrix $\boldsymbol{\rho}(\boldsymbol{\omega})$, a $p \times p$ a Hermitian positive definite matrix with ones on the diagonal, that is, a correlation matrix for complex-valued random variables. The possibility that $\phi_{jk}(\boldsymbol{\omega}) \neq 0$ allows for phase relationships among the various components. Further, the marginal spectrum for the jth component is $s_j(\boldsymbol{\omega})$.

Under the specification in (5), we can write

$$\boldsymbol{f}(\boldsymbol{\omega}) = D(\boldsymbol{s}(\boldsymbol{\omega}))\boldsymbol{\rho}(\boldsymbol{\omega})D(\boldsymbol{s}(\boldsymbol{\omega})),$$

where $D(\boldsymbol{s}(\boldsymbol{\omega})) = \text{diag}(\sqrt{s_1(\boldsymbol{\omega})}, \ldots, \sqrt{s_p(\boldsymbol{\omega})})$. This implies that $\boldsymbol{f}(\boldsymbol{\omega})$ is positive definite, since pre- and post-multiplying a positive definite matrix by a square full rank matrix and its transpose gives a positive definite matrix, and it implies that

$$\boldsymbol{f}^{-1}(\boldsymbol{\omega}) = D^{-1}(\boldsymbol{s}(\boldsymbol{\omega}))\boldsymbol{\rho}^{-1}(\boldsymbol{\omega})D^{-1}(\boldsymbol{s}(\boldsymbol{\omega})).$$

Since $D(\boldsymbol{s}(\boldsymbol{\omega}))$ is diagonal, a zero in the inverse of $\boldsymbol{f}(\boldsymbol{\omega})$ corresponds to a zero in the inverse of $\boldsymbol{\rho}(\boldsymbol{\omega})$, so the specification in (5) is also useful to specifying conditional dependence relationships through the matrices $\boldsymbol{\rho}(\boldsymbol{\omega})$.

To see how this is useful in the 3-component case that we consider here, we write

$$\boldsymbol{\rho}(\boldsymbol{\omega}) = \begin{bmatrix} 1 & \rho_{12}(\boldsymbol{\omega})^* & \rho_{13}(\boldsymbol{\omega})^* \\ \rho_{12}(\boldsymbol{\omega}) & 1 & \rho_{23}(\boldsymbol{\omega})^* \\ \rho_{13}(\boldsymbol{\omega}) & \rho_{23}(\boldsymbol{\omega}) & 1 \end{bmatrix},$$

where * denotes complex conjugation. The following theorem describes how to constrain $\rho_{13}(\boldsymbol{\omega})$ in order to force components 1 and 3 to be conditionally uncorrelated given component 2 when the phase has a certain form.

Theorem 2. *If $p = 3$, then*

(i) Components 1 and 3 are conditionally uncorrelated given component 2 if and only if $\rho_{13}(\boldsymbol{\omega}) = \rho_{12}(\boldsymbol{\omega})\rho_{23}(\boldsymbol{\omega})$ for almost every $\boldsymbol{\omega} \in \mathbb{T}^d$; in this case, $\boldsymbol{\rho}(\boldsymbol{\omega})$ is positive definite if and only if $|\rho_{12}(\boldsymbol{\omega})|^2 < 1$ and $|\rho_{23}(\boldsymbol{\omega})|^2 < 1$.

(ii) If $\rho_{13}(\boldsymbol{\omega}) \neq \rho_{12}(\boldsymbol{\omega})\rho_{23}(\boldsymbol{\omega})$ and $\phi_{jk}(\boldsymbol{\omega})$ can be expressed as $\phi_j(\boldsymbol{\omega}) - \phi_k(\boldsymbol{\omega})$ for every j and k, then $\boldsymbol{\rho}(\boldsymbol{\omega})$ is positive definite if and only if $|\rho_{12}(\boldsymbol{\omega})|^2 < 1$, $|\rho_{23}(\boldsymbol{\omega})|^2 < 1$, and

$$|\rho_{12}(\boldsymbol{\omega})||\rho_{23}(\boldsymbol{\omega})| - u(\boldsymbol{\omega}) < |\rho_{13}(\boldsymbol{\omega})| < |\rho_{12}(\boldsymbol{\omega})||\rho_{23}(\boldsymbol{\omega})| + u(\boldsymbol{\omega}),$$

where $u(\boldsymbol{\omega}) = \sqrt{1 - |\rho_{12}(\boldsymbol{\omega})|^2}\sqrt{1 - |\rho_{23}(\boldsymbol{\omega})|^2}$.

A proof is given in the appendix. Theorem 2 makes it possible to perform a likelihood ratio test of the hypothesis that components 1 and 3 are conditionally uncorrelated given component 2 by considering the reduced model in which $\rho_{13}(\boldsymbol{\omega}) = \rho_{12}(\boldsymbol{\omega})\rho_{23}(\boldsymbol{\omega})$ for every $\boldsymbol{\omega} \in \mathbb{T}^d$ versus the full model in which $|\rho_{13}(\boldsymbol{\omega})|$ is allowed to vary across its allowable values for every $\boldsymbol{\omega} \in \mathbb{T}^d$. The fact that the result requires a special form for the phase is a restriction, but the restriction obviously includes models with zero phase, and it also includes translational models, in which the phase can be written as $\phi_{jk}(\boldsymbol{\omega}) = \boldsymbol{\omega}'(\boldsymbol{h}_j - \boldsymbol{h}_k)$, where $\boldsymbol{h}_j = (h_{j1}, h_{j2})$ defines a translation in $\mathbb{Z}^d$.

For the purpose of testing the hypothesis that components 1 and 3 are conditionally uncorrelated given component 2, we propose the following specification for the cross coherence spectra: let $\gamma_{ij}(\boldsymbol{\omega}) : \mathbb{T}^2 \to [-1, 1]$

and $r_{ij} \in (-1, 1)$. Then define

$$
\begin{aligned}
|\rho_{12}(\boldsymbol{\omega})| &= r_{12}\gamma_{12}(\boldsymbol{\omega}) \\
|\rho_{23}(\boldsymbol{\omega})| &= r_{23}\gamma_{23}(\boldsymbol{\omega}) \\
|\rho_{13}(\boldsymbol{\omega})| &= |\rho_{12}(\boldsymbol{\omega})||\rho_{23}(\boldsymbol{\omega})| + r_{13}u(\boldsymbol{\omega})\gamma_{13}(\boldsymbol{\omega}).
\end{aligned}
$$

This choice for the form of $|\rho_{13}(\boldsymbol{\omega})|$ allows for a one-parameter test for conditional correlation since $r_{13} = 0$ implies that components 1 and 3 are conditionally uncorrelated given component 2.

3.4. Parametric forms for spectral densities and coherence

The marginal spectral densities must be positive and integrable to preserve positive definiteness of the covariance function. Finley et al. (2009) suggested models for which the spectral density can be written as $s_j(\boldsymbol{\omega}) = s_{j1}(\omega_1)s_{j2}(\omega_2)$. Thus, we can choose any one-dimensional forms $s_{jk}(\omega)$ available from the time series literature, such as autoregressive or moving average spectral densities (Brockwell and Davis, 2009). So-called separable models are advantageous for exact likelihood computations, but it is also important to propose models that do not possess the separability property.

One particularly flexible parametric family of non-separable spectral density functions is

$$
s_j(\boldsymbol{\omega}) = s_j(\omega_1, \omega_2) = \frac{\sigma_j^2}{\left(1 + \alpha_j^2(\sin^2(\omega_1/2) + \sin^2(\omega_2/2))\right)^{\nu_j+1}}, \qquad (6)
$$

where $\sigma_j, \alpha_j, \nu_j > 0$. This family is of interest because if the lattice spacing is δ, then we may express the spectral density as

$$
s_j(\boldsymbol{\omega}) = s_j(\omega_1, \omega_2) = \frac{\sigma_j^2}{\left(1 + \left(\frac{\alpha_j}{\delta}\right)^2\left(\sin^2\left(\frac{\delta\omega_1}{2}\right) + \sin^2\left(\frac{\delta\omega_2}{2}\right)\right)\right)^{\nu_j+1}},
$$

which tends to the spectral density of the isotropic Matérn covariance function as $\delta \to 0$, so we refer to (6) as the quasi Matérn spectral density. The three parameters $(\sigma_j, \alpha_j, \nu_j)$ may suffer from lack of identifiability in certain asymptotic settings, just as is the case for the Matérn covariance (Zhang, 2004). The form in (6) assumes that the process is invariant to 90 degree rotations of the domain. We discuss in Section 5 how to modify this model to treat the two dimensions differently.

Gaetan and Guyon (2010) describe how to derive spectral density functions for parametric simultaneous auto-regressive (SAR) and conditional

auto-regressive (CAR) models. For example, the isotropic nearest neighbor SAR model in two dimensions is

$$Z_j(x_1, x_2) = \frac{\alpha_j}{4}\left[Z_j(x_1 - 1, x_2) + Z_j(x_1 + 1, x_2) + \right.$$
$$\left. Z_j(x_1, x_2 - 1) + Z_j(x_1, x_2 + 1)\right] + \varepsilon_j(x_1, x_2), \qquad (7)$$

with $|\alpha_j| < 1$, and $\varepsilon_j(x_1, x_2)$ has mean zero and variance σ_j^2. This model has spectral density

$$s_j(\boldsymbol{\omega}) = s_j(\omega_1, \omega_2) = \frac{\sigma_j^2}{\left(1 - \frac{\alpha_j}{2}(\cos(\omega_1) + \cos(\omega_2))\right)^2}. \qquad (8)$$

(Gaetan and Guyon, 2010). If we allow the exponent in the denominator of 8 to vary, we obtain the more flexible model

$$s_j(\boldsymbol{\omega}) = s_j(\omega_1, \omega_2) = \frac{\sigma_j^2}{\left(1 - \frac{\alpha_j}{2}(\cos(\omega_1) + \cos(\omega_2))\right)^{\nu_j+1}}, \qquad (9)$$

where the form of the exponent, ν_j+1, is a convention chosen to correspond to that found in the isotropic Matérn spectral density and the quasi Matérn spectral density in Equation (6).

We use similar parametric representations for the $\gamma_{ij}(\boldsymbol{\omega})$, which are constrained to $[-1, 1]$:

$$\gamma_{ij}(\boldsymbol{\omega}) = \gamma_{ij}(\omega_1, \omega_2) = \frac{1}{\left(1 + a_{ij}^2(\sin^2(\omega_1/2) + \sin^2(\omega_2/2))\right)^{m_{ij}+1}},$$

with $a_{ij}, m_{ij} > 0$, or alternatively we could choose

$$\gamma_{ij}(\boldsymbol{\omega}) = \gamma_{ij}(\omega_1, \omega_2) = \frac{(1 - |a_{ij}|)^{m_{ij}}}{\left(1 - \frac{a_{ij}}{2}(\cos(\omega_1) + \cos(\omega_2))\right)^{m_{ij}+1}},$$

with $|a_{ij}| < 1$ and $m_{ij} > 0$.

3.5. Estimation with maximum likelihood and Whittle likelihood

Let $\boldsymbol{n} = (n_1, \ldots, n_d)$ be the vector containing the dimensions of the lattice, so that we observe $Z_j(\boldsymbol{x})$ for $j = 1, \ldots, p$ at locations $\boldsymbol{x} = (x_1, \ldots, x_p)$ where $\boldsymbol{x} \in \mathbb{J}_{\boldsymbol{n}} = \{\boldsymbol{x} : 1 \le x_j \le n_j, 1 \le j \le p\}$. Then if there are no missing values, we have $n = n_1 \times \ldots \times n_d$ observations. In Section 3, we did not impose any restrictions on the distribution of the process–we described only

their second-order properties. In Section 4, we simulate from mean-zero Gaussian process models with the covariance functions that arise from the spectral densities and coherence functions that we specify.

The negative Gaussian loglikelihood is (up to an additive constant)

$$L(\theta) = \frac{1}{2}\log\det\Sigma_\theta + \frac{1}{2}\boldsymbol{Z}^T\Sigma_\theta^{-1}\boldsymbol{Z},$$

where $\boldsymbol{Z}$ is a vector containing the observations, and Σ_θ is the covariance matrix of the observations, respecting the ordering of the observations in $\boldsymbol{Z}$. In general, the computation of the Gaussian loglikelihood requires the Cholesky decomposition of the $pn \times pn$ covariance matrix, which requires $O((np)^3)$ floating point operations, as well as the $O((np)^2)$ memory required for storing it. When the data are observed on a lattice, the Gaussian loglikelihood may be approximated by the Whittle likelihood,

$$L_W(\theta) = \int_{[-\pi,\pi]^d} \left[\log\det \boldsymbol{f}_\theta(\boldsymbol{\omega}) + \boldsymbol{I}(\boldsymbol{\omega})^H \boldsymbol{f}_\theta(\boldsymbol{\omega})^{-1}\boldsymbol{I}(\boldsymbol{\omega})\right] d\boldsymbol{\omega}, \tag{10}$$

where $\boldsymbol{f}_\theta$ is the spectral density with parameter θ, H denotes the Hermitian (conjugate) transpose, and $\boldsymbol{I}(\boldsymbol{\omega}) = (I_j(\boldsymbol{\omega}))$ for $j = 1, \ldots, p$ contains the d-dimensional discrete Fourier transforms (DFT) of the observations,

$$I_j(\boldsymbol{\omega}) = \frac{1}{\sqrt{n}} \sum_{\boldsymbol{x}\in\mathbb{J}_{\boldsymbol{n}}} Z_j(\boldsymbol{x}) \exp(-i\boldsymbol{\omega}'\boldsymbol{x}). \tag{11}$$

Typically, we approximate the integral in (10) by summing over a finite number of Fourier frequencies $\boldsymbol{\omega}$, in which case all necessary values of (11) can be computed efficiently with a fast Fourier transform (FFT) algorithm.

The Whittle likelihood is known to be susceptible to edge effects and produces parameter estimates with slow convergence in models for univariate data when the dimension of the lattice is greater than 1. Indeed, the number of observations near the boundary of the lattice is $O(n^{1-1/d})$, and thus increases nontrivially with the total number of observations when $d > 1$. To alleviate the edge effect issue, Guyon (1982) suggested a modification that arises from the DFT of an unbiased covariance estimate, and Dahlhaus and Künsch (1987) suggested to taper the observations near the boundaries of the lattice and proved that their procedure produces asymtotically efficient parameter estimates when $d \leq 3$. The tapering approach involves multiplying the observations by a window function $h(\boldsymbol{x})$ that approaches zero smoothly when $\boldsymbol{x}$ approaches the boundaries, and the resulting likelihood

approximation becomes

$$L_T(\theta) = \int_{[-\pi,\pi]^d} \left[\log\det \boldsymbol{f}_\theta(\boldsymbol{\omega}) + \boldsymbol{I}_h(\boldsymbol{\omega})^H \boldsymbol{f}_\theta(\boldsymbol{\omega})^{-1} \boldsymbol{I}_h(\boldsymbol{\omega})\right] d\boldsymbol{\omega},$$

where $\boldsymbol{I}_h(\boldsymbol{\omega})$ contains the tapered versions of the DFT

$$I_{h,j}(\boldsymbol{\omega}) = \left(\prod_{k=1}^{d}\sum_{j=1}^{n_k} h_k(x_j)^2\right)^{-1/2} \sum_{\boldsymbol{x}\in\mathbb{J}_d} (Z_j(\boldsymbol{x})h(\boldsymbol{x}))\exp(-i\boldsymbol{\omega}'\boldsymbol{x}), \qquad (12)$$

where $h(\boldsymbol{x}) = h_1(x_1/n_1)\times\ldots\times h_d(x_d/n_d)$ is the product of univariate tapering functions. A common choice for $h_k(u)$ is a cosine taper, also known as a Tukey window. The tapering functions also usually contain a parameter that controls the number of observations near the boundary to taper, and a typical choice is to taper 5-10% of the observations on each boundary.

4. Simulation Study

The simulations in this section study the behavior of Whittle likelihood estimates of parameters governing the coherence of multivariate spatial lattice data when the dimension of the lattice is 2. The purpose of the study is to demonstrate how the strength of spatial dependence affects the parameter estimates in the Whittle likelihood and how tapering may be employed to improve parameter estimation, especially when the sample size is large. We study the performance of tapered and untapered parameter estimates as the size of the lattice increases under two spatial dependence scenarios. In particular, we consider lattices of size $n_1 \times n_2$ for $n_1 = n_2 \in \{10, 30, 50\}$, and we taper either 0% or 10% of the observations on each boundary using a cosine taper. In all simulations, we assume that the spatial covariance functions are invariant to 90-degree rotations and that the spectral density matrices are real, so that the phase relationships are zero. All of the models will assume conditional independence between components 1 and 3 given component 2, and we maximize the Whittle likelihoods without this assumption in place. This allows us to report the sampling distribution of the parameter estimates under the null hypothesis of conditional independence between components 1 and 3 given component 2. The Whittle likelihoods do not use information from the zero frequency.

Due to possible lack of identifiability, we fix $\nu_j = m_{ij} = 1$ in all the simulations and assume they are known. We simulate from the quasi Matérn model with $(\sigma_j, \alpha_j, \nu_j) = (1, \alpha, 1)$ for every j, and $r_{12} = r_{23} = 0.7$,

$n_1 = n_2$	Taper	$r_{12} = 0.7$	$r_{23} = 0.7$	$r_{13} = 0$
10	0%	0.697 (0.014)	0.703 (0.012)	-0.074 (0.053)
	10%	0.697 (0.014)	0.703 (0.013)	-0.061 (0.052)
30	0%	0.699 (0.005)	0.695 (0.004)	-0.014 (0.024)
	10%	0.699 (0.006)	0.696 (0.004)	-0.051 (0.025)
50	0%	0.696 (0.003)	0.696 (0.003)	-0.022 (0.015)
	10%	0.698 (0.003)	0.698 (0.003)	-0.009 (0.010)

Table 1: Average estimates and standard errors (in parentheses) for r_{ij} parameters under Scenario 1 (weaker spatial dependence). True values of the parameters are given at the top of the table.

$n_1 = n_2$	Taper	$a_{12} = 0.3$	$a_{23} = 0.3$
10	0%	0.207 (0.020)	0.194 (0.018)
	10%	0.209 (0.020)	0.196 (0.018)
30	0%	0.279 (0.008)	0.277 (0.007)
	10%	0.287 (0.009)	0.289 (0.007)
50	0%	0.287 (0.004)	0.285 (0.004)
	10%	0.297 (0.004)	0.295 (0.004)

Table 2: Average estimates and standard errors (in parentheses) for a_{ij} parameters under Scenario 1 (weaker spatial dependence). True values of the parameters are given at the top of the table.

$r_{13} = 0$ (conditional independence), $a_{ij} = 0.3$, $m_{ij} = 1$. The values of a_{13} and m_{13} are not relevant for the simulation since $r_{13} = 0$. When fitting the models, we assume that the ν_j and m_{ij} are known, and we estimate σ_j, α_j, r_{ij}, and a_{ij} for i and $j = 1, 2, 3$. Our simulations consider two scenarios: weaker spatial dependence in which $\alpha = 3$ and stronger spatial dependence in which $\alpha = 6$.

In Tables 1-4, we report the average parameter estimates and the standard error of the average over 100 simulations. As expected, parameter estimates generally improve as the sample size increases. All parameters are estimated more accurately by the Whittle likelihoods in the case of weaker spatial dependence versus the case of stronger spatial dependence, which is to be expected because the Whittle likelihood is exact in the case of no spatial dependence (white noise). Tapering does not have much of an effect when the dimensions are 10×10, which is not surprising because 10% tapering means tapering only the observations exactly on the boundary. However, when the dimensions are 50×50, tapering greatly improves estimates of parameters that are not estimated well in the untapered case, with the possible

$n_1 = n_2$	Taper	$r_{12} = 0.7$	$r_{23} = 0.7$	$r_{13} = 0$
10	0%	0.759 (0.014)	0.697 (0.016)	-0.075 (0.055)
	10%	0.758 (0.014)	0.696 (0.016)	-0.054 (0.054)
30	0%	0.669 (0.007)	0.664 (0.006)	-0.021 (0.023)
	10%	0.680 (0.005)	0.675 (0.005)	-0.012 (0.031)
50	0%	0.679 (0.004)	0.686 (0.004)	0.002 (0.018)
	10%	0.692 (0.003)	0.694 (0.003)	0.063 (0.021)

Table 3: Average estimates and standard errors (in parentheses) for r_{ij} parameters under Scenario 2 (stronger spatial dependence). True values of the parameters are given at the top of the table.

$n_1 = n_2$	Taper	$a_{12} = 0.3$	$a_{23} = 0.3$
10	0%	0.250 (0.021)	0.161 (0.020)
	10%	0.249 (0.021)	0.162 (0.020)
30	0%	0.186 (0.012)	0.188 (0.010)
	10%	0.242 (0.010)	0.246 (0.008)
50	0%	0.242 (0.006)	0.245 (0.007)
	10%	0.282 (0.005)	0.282 (0.005)

Table 4: Average estimates and standard errors (in parentheses) for a_{ij} parameters under Scenario 2 (stronger spatial dependence). True values of the parameters are given at the top of the table.

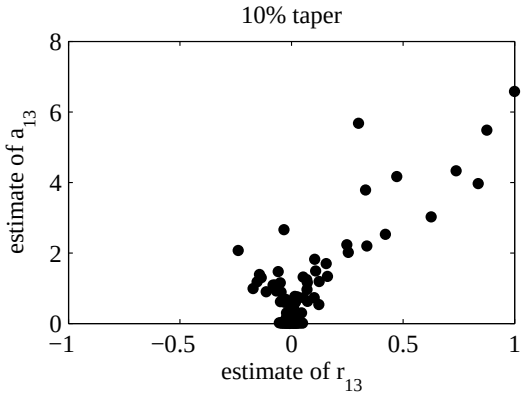

Figure 3: Scatter plots of the estimates of a_{13} versus the estimates of r_{13} with $n_1 = n_2 = 50$ and a 10% taper. When the estimate of r_{13} is not near zero (the truth), the estimate of a_{13} is large, ensuring that $\gamma_{13}(\boldsymbol{\omega})$ is near zero except for at a small range of low frequencies.

exception of r_{13} in the case of strong spatial dependence. We note that all estimates of r_{13} tend to be more variable than are the other parameter estimates. This is likely due to the fact that, in this parameterization, when a_{ij} is large, the coherence at all but the lowest frequencies may remain very small even when $|r_{ij}|$ is not close to zero. In Figure 3, we plot the estimates of a_{13} against the estimates of r_{13} for the case of strong spatial dependence, dimension 50×50, and 10% tapering. Indeed, the estimate of r_{13} is far from zero only in the cases when a_{13} is large.

5. Analysis of X-ray Fluorescence Data

In this section, we analyze the μ-XRF data with the purpose of making inference about the multivariate correlation structure in the presence of spatial dependence. The models discussed are for the centered log fluorescence signals, that is, we construct models for

$$Y_j(\boldsymbol{x}) = Z_j(\boldsymbol{x}) - \overline{Z}_j,$$

where $Z_j(\boldsymbol{x})$ is the log fluorescence at location $\boldsymbol{x}$, and $\overline{Z}_j$ is the sample mean of the log fluorescence of component j. We assume that $Y_j(\boldsymbol{x})$ has mean zero and is Gaussian, so that correlation and dependence are equivalent. We model the spectral densities and coherences with an anisotropic version of the quasi Matérn models discussed in Section 3. The models are

$$s_j(\boldsymbol{\omega}) = \frac{\sigma_j^2}{(1 + \alpha_{1j}^2 \sin^2(\omega_1/2) + \alpha_{2j}^2 \sin^2(\omega_2/2))^{\nu_j+1}},$$
$$\gamma_{ij}(\boldsymbol{\omega}) = \frac{1}{(1 + a_{ij}^2 \sin^2(\omega_1/2) + b_{ij}^2 \sin^2(\omega_2/2))^{m_{ij}+1}}.$$

The parameters α_{1j} and α_{2j} control the spatial range in the vertical and horizontal directions, respectively. Likewise, a_{ij} and b_{ij} control the coherence between components i and j in the vertical and horizontal directions. As in the simulation, we fix the exponent parameters at $\nu_j = 1$ and $m_{ij} = 1$.

The total number of observations in this analysis is 4410, so Gaussian maximum likelihood estimation is time-consuming but not infeasible. Our first task is to maximize the likelihood with and without the constraint $r_{13} = 0$ in place, that is, with and without the constraint that arsenic and chromium are conditionally uncorrelated given iron. For the time being, we assume that the phase is zero, reserving an investigation of the issue of possible misalignment for later in this section. Removing the constraint

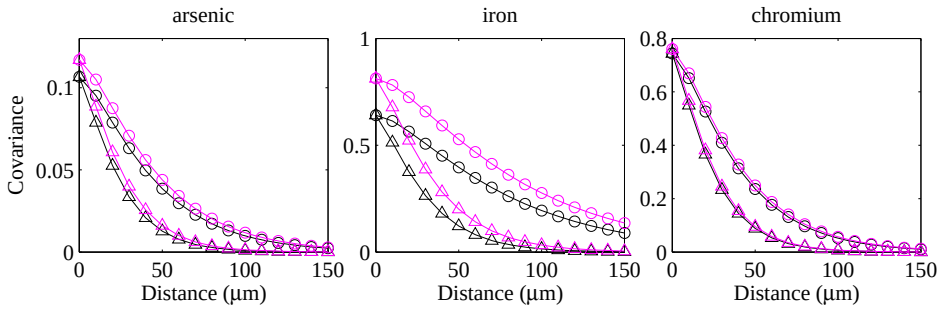

Figure 4: Maximum likelihood (black) and Whittle likelihood (magenta) estimated spatial covariance functions. Circles denote covariance in the horizontal direction, and triangles denote covariance in the vertical direction.

that $r_{13} = 0$ increases the loglikelihood by 9.69 units. The unconstrained model introduces an additional 3 parameters into the model, so under the null hypothesis of conditional independence between arsenic and iron, twice the change in loglikelihood follows approximately a χ^2 distribution with 3 degrees of freedom, leading to a rejection of the null hypothesis. However, for a dataset of this size, a change in loglikelihood of this magnitude offers strong but not overwhelming evidence for conditional dependence–certainly far weaker evidence than would have been obtained had the spatial dependence been ignored.

We also estimate the model using the Whittle likelihood and a 10% cosine taper along each boundary. We plot in Figures 4 and 5 the estimated spatial covariance and cross covariance functions for each element found using exact and Whittle likelihoods, without the constraint that $r_{13} = 0$. We note that in every case the estimated correlation is stronger in the horizontal direction than it is in the vertical direction, a finding that agrees with the behavior apparent in Figure 1. This effect is likely due to the X-ray beam continuously scanning in the horizontal direction and a slightly narrower beam width in the vertical direction. The Whittle likelihood estimates disagree slightly in some cases with the exact likelihood estimates, which is not surprising since tapering essentially discards the observations near the border of the region. Maximum likelihood estimates $r_{13} = -0.39$, so that the coherence between arsenic and chromium is weaker than that expected by a model with conditional independence between arsenic and chromium given iron.

Using the unconstrained fitted model we can construct the cross spectral density matrices for the residual processes (Appendix A) and compute

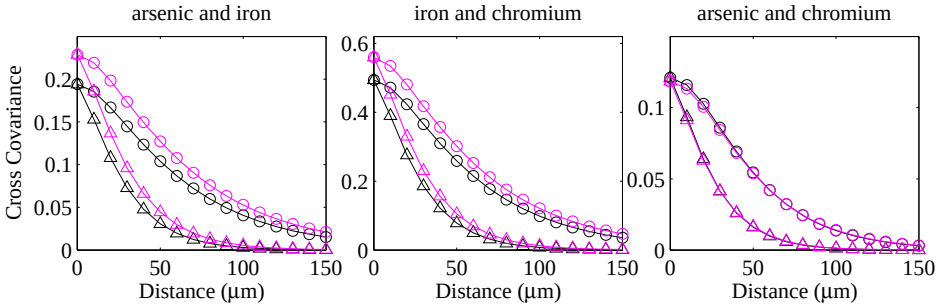

Figure 5: Maximum likelihood (black) and Whittle likelihood (magenta) estimated spatial cross covariance functions. Circles denote covariance in the horizontal direction, and triangles denote covariance in the vertical direction.

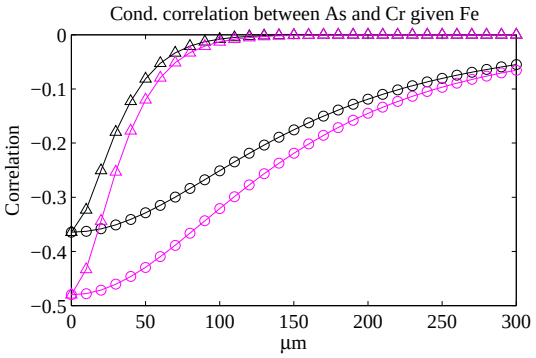

Figure 6: Maximum likelihood (black) and Whittle likelihood (magenta) estimated conditional correlation between arsenic and chromium given iron. Circles denote conditional correlation in the horizontal direction, and triangles denote conditional correlation in the vertical direction.

the conditional spatial correlation function, which is $R_{13}(\boldsymbol{u})$ divided by the square roots of the conditional variances of arsenic and chromium given iron. The resulting correlation function is plotted in Figure 6. We note that the conditional correlation functions are always negative, which agrees with the exploratory analysis seen in Figure 2.

The model we fit here assumes that the cross spectral density matrices are real and thus there are no phase relationships among the elements, but that need not be the case for the data, where misalignment among components may exist. To investigate this issue, we define $\phi_j(\boldsymbol{\omega}) = \omega_1 h_{j1} + \omega_2 h_{j2}$, to be the phase function for component j, which corresponds to a misalignment of h_{j1} units in the vertical direction and h_{j2} units in the horizontal

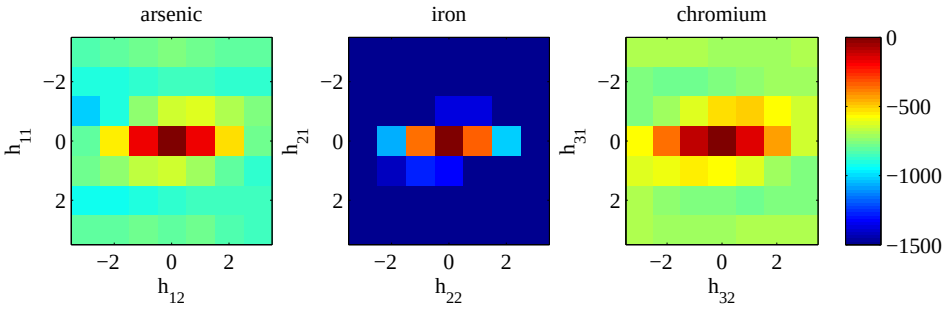

Figure 7: Investigation of possible misalignment. We plot maps of difference in loglikelihood from the maximum likelihood model with zero phases to one in which h_{j1} (vertical axis) or h_{j2} (horizontal axis) are not zero for only one j. All three plots show that the zero phase model maximizes the likelihood.

direction. For each j, we fix $h_{k1} = h_{k2} = 0$ for $k \neq j$ and plot the loglikelihood function evaluated at the maximum likelihood covariance and coherence parameters (under zero phase) for various values of h_{j1} and h_{j2}. We plot in Figure 7 the resulting loglikelihood maps, which shows that the likelihood is maximized at zero phase for each component. Therefore, we do not detect any phase relationships among the elements. It is interesting to note that the loglikelihood decreases more rapidly in the vertical direction in every case, which is the result of stronger spatial dependence in the horizontal direction.

6. Discussion

In synchrotron XAS analysis of geochemical matrices, scatter plots between pairs of elements are often used to determine correlations and to isolate distinct populations of elemental associations for more detailed microscale, chemical-speciation analyses (Manceau et al., 2002). However, significance testing on these simple correlations does not account for spatial correlations of elements. The rapid development of XAS and other new analytical technologies that generate massive quantities of multivariate spatio-chemical data, as well as the scientific need to disentangle the complex multivariate dependence structure in the data, call for new statistical methodology that is able to specify and estimate flexible multivariate dependence structures in a computationally efficient framework. This article provides some of the theoretical groundwork for forcing components to be conditionally uncorrelated in such models, both in a general case in a special

case of three components, where we give a more specific parametric formulation that allows for likelihood ratio testing. We present a simulation study in which we demonstrate that the computationally efficient Whittle likelihood can be used to estimate parameters when the sample size is sufficiently large so as to allow for tapering of observations on the boundary. We also present results of a detailed analysis of our μ-XRF data and conclude that arsenic and chromium are significantly conditionally negatively correlated given iron.

Statistical approaches that account for spatial correlation in microscale analytical data from complex geochemical systems will provide a more robust assessment of chemical relationships that could suggest controlling mechanisms of trace-element mobility and environmental impacts. This article provides an important way forward, but certainly there is much work to be done in developing specific models for conditional correlation in multivariate spatial data. For example, the linear model of co-regionalization (LMC) is a popular choice among practitioners for modeling multivariate spatial data (Banerjee et al., 2003). The LMC has cross spectral density function

$$f_{jk}(\boldsymbol{\omega}) = \sum_{m=1}^{M} T_m(j,k)\rho_m(\boldsymbol{\omega}),$$

where each $\rho_m(\boldsymbol{\omega})$ is individually a valid spectral density function, and each $\boldsymbol{T}_m$ is a positive definite $p \times p$ matrix with entries $T_m(j,k)$. In this model, forcing $\boldsymbol{T}_m^{-1}$ to have a zero in entry (j,k) for every m is *not* sufficient to guarantee that components j and k are conditionally uncorrelated given the rest. In the simplest nontrivial case of $m = 2$ and $p = 3$, we also require $(\boldsymbol{T}_1 + \boldsymbol{T}_2)^{-1}$ to have a zero in entry (j,k). The development of more general necessary and sufficient conditions for components to be conditionally uncorrelated in the LMC is still an open question as far as we know.

It is also important to develop models and methodology for irregularly-spaced multivariate spatial data for which two components are be conditionally uncorrelated given the rest. Fuentes (2007) proposed spectral methods for irregularly-spaced spatial data. Additionally, the work of Stein (2005), may provide some guidance on how to define and enforce conditional independence in multivariate Gaussian process models. In that paper, necessary and sufficient conditions on univariate space-time covariance functions were given to ensure that past and future observations are conditionally independent given the entire current spatial field. It may be possible to adapt those results to obtain conditions for conditional independence in multivariate Gaussian process models.

Theorem 2 provides a means for placing constraints on $\boldsymbol{f}(\boldsymbol{\omega})$ to ensure that its inverse contains a zero in a specified location when there are three components to the process. When there are more than three components, the corresponding constraints may be much more complicated. In these cases, there may be some value to moving to a Bayesian framework in which inference can be made with Markov chain Monte Carlo (MCMC). In this framework, the multivariate models can be specified without any constraints in place, and then MCMC samples can be used to infer the posterior distribution of the entries of the inverses of the spectral density matrices.

Acknowledgments

The authors thank Dr. Ryan Tappero for assistance in collecting data at Beamline X27A. Use of the National Synchrotron Light Source, Brookhaven National Laboratory, was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-AC02-98CH10886. Guinness and Fuentes are supported by the National Science Foundation's Research Network for Statistical Methods for Atmospheric and Oceanic Sciences (STATMOS), award number 1107046.

Appendix A. Proof for Theorem 1

Without loss of generality, let $i = 1$, and $j = 2$, and define

$$\boldsymbol{Y}(\boldsymbol{x}) = \begin{bmatrix} Z_1(\boldsymbol{x}) \\ Z_2(\boldsymbol{x}) \end{bmatrix} \quad \text{and} \quad \boldsymbol{V}(\boldsymbol{x}) = \begin{bmatrix} Z_3(\boldsymbol{x}) \\ \vdots \\ Z_p(\boldsymbol{x}) \end{bmatrix}.$$

The best linear filters a_{12} and a_{21} satisfy

$$E\left[\left(\boldsymbol{Y}(\boldsymbol{x}) - \sum_{\boldsymbol{u}\in\mathbb{Z}^d} \begin{bmatrix} a_{12}(\boldsymbol{x}-\boldsymbol{u})^T \\ a_{21}(\boldsymbol{x}-\boldsymbol{u})^T \end{bmatrix} \boldsymbol{V}(\boldsymbol{u})\right) \boldsymbol{V}(\boldsymbol{y})^T\right] = \boldsymbol{0}_{2\times p-2} \qquad \text{(A.1)}$$

for every $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{Z}^d$. Since we may define a Hilbert space isomorphism between closed linear manifolds of $(Z_1(\boldsymbol{x}), \ldots, Z_p(\boldsymbol{x}))$ and closed linear manifolds of $(\exp(i\boldsymbol{\omega}_1'\boldsymbol{x}), \ldots, \exp(i\boldsymbol{\omega}_p'\boldsymbol{x}))$, studying (A.1) is equivalent to studying

$$f_{YV}(\boldsymbol{\omega}) - A(\boldsymbol{\omega}) f_{VV}(\boldsymbol{\omega}) = 0,$$

where f_{YV} is the cross spectral density matrix between $\boldsymbol{Y}$ and $\boldsymbol{V}$, A contains the discrete Fourier transforms of the components of a_{12} and a_{21},

and f_{VV} is the cross spectral density matrix of $\boldsymbol{V}$ with itself. Therefore $A(\boldsymbol{\omega}) = f_{VY}(\boldsymbol{\omega})f_{VV}(\boldsymbol{\omega})^{-1}$, and the cross spectral density matrix of the residual processes is

$$f_{\varepsilon\varepsilon}(\boldsymbol{\omega}) = f_{YY}(\boldsymbol{\omega}) - f_{YV}(\boldsymbol{\omega})f_{VV}(\boldsymbol{\omega})^{-1}f_{VY}(\boldsymbol{\omega}).$$

Writing

$$\boldsymbol{f}(\boldsymbol{\omega}) = \begin{bmatrix} f_{YY}(\boldsymbol{\omega}) & f_{YV}(\boldsymbol{\omega}) \\ f_{VY}(\boldsymbol{\omega}) & f_{VV}(\boldsymbol{\omega}) \end{bmatrix},$$

and using a well-known result on inverses of block matrices, it is easy to see that the upper left 2×2 submatrix of $\boldsymbol{f}(\boldsymbol{\omega})^{-1}$ is $f_{\varepsilon\varepsilon}(\boldsymbol{\omega})^{-1}$. Writing

$$f_{\varepsilon\varepsilon}(\boldsymbol{\omega}) = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix},$$

the (1,2) entry of $f_{\varepsilon\varepsilon}(\boldsymbol{\omega})^{-1}$ is $-a_{12}/\det f_{\varepsilon\varepsilon}(\boldsymbol{\omega})$. This implies that a zero in $\boldsymbol{f}(\boldsymbol{\omega})^{-1}$ in entry $(1,2)$ for almost every $\boldsymbol{\omega}$ is equivalent to the same being true of the $(1,2)$ entry of $f_{\varepsilon\varepsilon}(\boldsymbol{\omega})$, which is equivalent to the residual processes being uncorrelated at every spatial lag, and thus components 1 and 2 being conditionally uncorrelated given the rest. □

Appendix B. Proof for Theorem 2

The first part of (i) follows directly from (4). To simplify the notation, we write $\rho_{12}(\boldsymbol{\omega}) = a$, $\rho_{23}(\boldsymbol{\omega}) = b$, $\rho_{13}(\boldsymbol{\omega}) = c$. Then $c = ab$, and the Cholesky decomposition of $\boldsymbol{\rho}(\boldsymbol{\omega})$ is

$$\begin{bmatrix} 1 & 0 & 0 \\ a & \sqrt{1-|a|^2} & 0 \\ ab & b\sqrt{1-|a|^2} & \sqrt{1-|b|^2} \end{bmatrix},$$

and the second part of (i) follows from equivalence of Hermitian positive definite matrices and lower triangular decompositions that have real and positive entries on the diagonal.

To prove (ii) we write the Cholesky decomposition of $\boldsymbol{\rho}(\boldsymbol{\omega})$ as

$$\begin{bmatrix} 1 & 0 & 0 \\ a & \sqrt{1-|a|^2} & 0 \\ c & \frac{b-a^*c}{\sqrt{1-|a|^2}} & \sqrt{1-|c|^2-\frac{|b-a^*c|^2}{1-|a|^2}} \end{bmatrix}.$$

If $|a|^2, |b|^2 < 1$, then $\boldsymbol{\rho}(\boldsymbol{\omega})$ is positive definite if and only if

$$(1 - |c|^2)(1 - |a|^2) - |b - a^* c|^2 > 0.$$

This inequality can be simplified to

$$|c|^2 - 2\mathrm{Re}(abc^*) + (|b|^2 + |a|^2 - 1) < 0, \tag{B.1}$$

and we note that

$$\begin{aligned} abc^* &= |a||b||c| \exp(i(\phi_1(\boldsymbol{\omega}) - \phi_2(\boldsymbol{\omega}) + \phi_2(\boldsymbol{\omega}) - \phi_3(\boldsymbol{\omega}) + \phi_3(\boldsymbol{\omega}) - \phi_1(\boldsymbol{\omega}))) \\ &= |a||b||c|, \end{aligned}$$

which means that (B.1) is a real quadratic function of $|c|$, is concave up, and has roots

$$|a||b| \pm \sqrt{1 - |a|^2}\sqrt{1 - |b|^2},$$

completing the proof. □

References

Banerjee, S., Gelfand, A. E., Carlin, B. P., 2003. Hierarchical modeling and analysis for spatial data. CRC Press.

Brockwell, P. J., Davis, R. A., 2009. Time series: theory and methods. Springer.

Brown, G. E., Sturchio, N. C., 2002. An overview of synchrotron radiation applications to low temperature geochemistry and environmental science. Reviews in Mineralogy and Geochemistry 49 (1), 1–115.

Dahlhaus, R., 2000. Graphical interaction models for multivariate time series. Metrika 51 (2), 157–172.

Dahlhaus, R., Künsch, H., 1987. Edge effects and efficient parameter estimation for stationary random fields. Biometrika 74 (4), 877–882.

Edwards, D., 2000. Introduction to graphical modelling. Springer.

Finley, A. O., Banerjee, S., Waldmann, P., Ericsson, T., 2009. Hierarchical spatial modeling of additive and dominance genetic variance for large spatial trial datasets. Biometrics 65 (2), 441–451.

Fitts, J. P., Thieme, J., 2012. Looking into the nanoworld using x-rays. Chemical Geology 329, 1–2.

Fuentes, M., 2007. Approximate likelihood for large irregularly spaced spatial data. Journal of the American Statistical Association 102 (477), 321–331.

Gaetan, C., Guyon, X., 2010. Spatial Statistics and Modeling. Springer.

Gneiting, T., Kleiber, W., Schlather, M., 2010. Matérn cross-covariance functions for multivariate random fields. Journal of the American Statistical Association 105 (491), 1167–1177.

Guyon, X., 1982. Parameter estimation for a stationary process on a d-dimensional lattice. Biometrika 69 (1), 95–105.

Hayes, K. F., Roe, A. L., Brown Jr., G. E., Hodgson, K. O., Leckie, J. O., Parks, G. A., 1987. In situ x-ray absorption study of surface complexes: Selenium oxyanions on α-FeOOH. Science 238 (4828), 783–786.

Hesterberg, D., Duff, M. C., Dixon, J. B., Vepraskas, M. J., 2011. X-ray microspectroscopy and chemical reactions in soil microsites. Journal of Environmental Quality 40 (3), 667–678.

Kelly, S., Hesterberg, D., Ravel, B., 2008. Analysis of soils and minerals using x-ray absorption spectroscopy. Methods of soil analysis. Part 5. Mineralogical methods 5, 387–464.

Manceau, A., Marcus, M. A., Tamura, N., 2002. Quantitative speciation of heavy metals in soils and sediments by synchrotron x-ray techniques. In: Fenter, P. A., Rivers, M. L., Sturchio, N. C., Sutton, S. R. (Eds.), Applications of Synchrotron Radiation in Low-Temperature Geochemistry and Environmental Sciences. Vol. 49. Geochemical Society - Mineralogical Society of America, Washington, D.C., pp. 341–428.

McBride, M., 1989. Reactions controlling heavy metal solubility in soils. In: Advances in soil science. Springer, pp. 1–56.

Stein, M. L., 2005. Space–time covariance functions. Journal of the American Statistical Association 100 (469), 310–321.

Sutton, S. R., Rivers, M. L., 1999. Hard x-ray microprobe techniques and applications. In: Schulze, D., Stucki, J., Bertsch, P. (Eds.), CMS Workshop Lectures, Vol. 9. Synchrotron Methods in Clay Science. The Clay Minerals Society, Boulder, CO, pp. 146–163.

Whittle, P., 1954. On stationary processes in the plane. Biometrika, 434–449.

Yaglom, A. M., 1987. Correlation Theory of Stationary and Related Random Functions I. Springer-Verlag.

Zhang, H., 2004. Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. Journal of the American Statistical Association 99 (465), 250–261.