Spatial Statistics 9 (2014) 93–108

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta

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, United States

[b] NC State University, Department of Soil Science, United States

ARTICLE INFO

Article history:
Received 2 October 2013
Accepted 31 March 2014
Available online 18 April 2014

Keywords:
Multivariate
High-resolution data
Elemental composition
Conditional dependence

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.
© 2014 Elsevier B.V. All rights reserved.

* Corresponding author. Tel.: +1 9195130191.
E-mail address: jsguinne@ncsu.edu (J. Guinness).

http://dx.doi.org/10.1016/j.spasta.2014.03.009
2211-6753/© 2014 Elsevier B.V. All rights reserved.

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 nm–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 materials and from a variety of microscale analytical techniques.

The objective of this research is to develop statistical models and methods for making inferences 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 zeros in the jkth entries at almost 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 tended to result in diminished 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.

Download English Version:

https://daneshyari.com/en/article/1064601

Download Persian Version:

https://daneshyari.com/article/1064601

Daneshyari.com