Spatial Statistics 6 (2013) 57–77

Contents lists available at ScienceDirect

Spatial Statistics

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

Continental-scale kriging of gold-bearing commodities

Christien Thiart[a,*], Alfred Stein[b]

[a] Department of Statistical Sciences, University of Cape Town, Cape Town, South Africa

[b] Department of Earth Observation Science, Faculty of Geo-Information Science and Earth Observation (ITC), Twente University, Enschede, The Netherlands

ARTICLE INFO

Article history:
Received 11 December 2012
Accepted 19 July 2013
Available online 26 July 2013

Keywords:
Continental kriging
Weighted variogram
Stratification
Gondwana
Gold potential

ABSTRACT

This paper focuses on continental-scale kriging on the African continent using the gold-bearing commodities of the Gondwana Geoscience Indexing Database. The mineral layer contains over 20 000 commodities, each containing information on its ordinal interval size value. Boundaries between class intervals across the database are, however, not uniform. We perform spatial interpolation on a continental scale using the commodity gold as the binary variable. First, we select an appropriate distance metric in order to krige on an essentially spherical surface. We use this metric to implement a valid covariance function. Second, the ordinal size classes of the commodities are combined into a unique size classification. In addition, the commodity size classification is used as a proxy for data reliability and is incorporated by using a weighted variogram. The geology is used to stratify Africa into geologically homogeneous strata, leading to stratified kriging. The best model in each stratum is used to produce a map of gold commodities of Africa including the spatial uncertainties. By integrating advanced techniques with high-quality data, a state-of-the-art map of gold commodities was obtained for Africa, including the spatial uncertainties.

1. Introduction

A major aim of geostatistics is to predict (or map) values of an attribute of interest at unobserved locations using the observed data at known locations. The use of kriging for that purpose assumes

* Corresponding author. Tel.: +27 21 650 3219; fax: +27 216504773.
E-mail addresses: christien.thiart@uct.ac.za (C. Thiart), a.stein@utwente.nl (A. Stein).

2211-6753/$ – see front matter
http://dx.doi.org/10.1016/j.spasta.2013.07.004

that the data originate from a spatial model for a random field. This random field is decomposed into a mean structure and a spatial random process (Diggle and Ribeiro, 2007). The mean structure depends upon a parameter vector $\boldsymbol{\beta}$, whereas the spatial random process depends on the variance–covariance matrix $\boldsymbol{\Sigma}(\mathbf{h}, \boldsymbol{\theta})$, being a function of the Euclidean distance $\mathbf{h}$ between two points and the vector of parameters $\boldsymbol{\theta}$. When recording and analyzing spatial data, distance is generally taken to be Euclidean. The Euclidean distance however is not always appropriate to use, and it could well be distorted on large areas or on spheres. More and more data are measured on the continental scale, and the challenge is to implement spatial interpolation on that scale, comparable to kriging on a sphere (Schaffrin, 1992).

Problems relating to distance are twofold. First, an appropriate distance, or metric, must be defined for the area. Non-Euclidean distances should be considered if the near-spherical shape of the global earth form cannot be ignored and planar distances do not reflect the true distance (Schaffrin, 1992; Jun and Stein, 2007; Banerjee, 2005). Then distances based on great circles should more correctly be invoked (Gneiting, 1999; Cressie et al., 1990). In the past, non-Euclidean distances were considered during the recording of the water temperature and depth in a river, where the distance along the river was preferred to the distance based on the GPS coordinates (Cressie and Majure, 1997; Kern and Higdon, 1999; Curriero, 2006; Ver Hoef et al., 2006). In a different context, Dominici et al. (2000) used a binary distance, i.e., a value of 1 for points in the same area and 0 elsewhere, to link 20 US cities. Banerjee (2005, page 624) showed that the choice of metric influences not only the estimation of the parameters but also the predictions.

Second, the validity of the resultant covariance matrix must be checked. Most textbooks define valid covariance functions based on the Euclidean distance, but for other metrics such covariance functions are not always valid (Banerjee, 2005; Curriero, 2006; Diggle and Ribeiro, 2007; Huang et al., 2011). A broad-based definition of a valid covariance function is that the covariance matrix should be positive definite. Huang et al. (2011) expressed the covariance function on a sphere as a function of a coefficient b_n and n Legendre polynomials. They showed that the covariance function is positive definite if b_n is non-negative and the sums of the b_ns are finite, and presented ways to determine b_n. It is not sufficient to show that a covariance matrix is positive definite, however, at a specific location, as the validity might change with different configurations (Gotway and Young, 2008).

In this study, we propose a solution to this problem by using isometric embedding. In isometric embedding, the non-Euclidean distance function is computed with the original coordinates and then embedded into the Euclidean space by transforming the spatial coordinates in such a way that the distance between points is preserved in the Euclidean space. Isometric embedding is theoretically difficult to implement, and it is approximated by means of multidimensional scaling (Mardia et al., 1979). A simple multidimensional scaling example is given by Schabenberger and Gotway (2005), and its use in a spatial context is demonstrated with a simulation study by Curriero (2006). Boisvert et al. (2009) determined a nonlinear path between geological deposits, resulting in a non-Euclidean distance metric. They implemented multidimensional scaling to ensure positive definiteness of the resulting kriging system. In the case of large datasets, embedding using multidimensional scaling may create computational difficulties. Boisvert and Deutsch (2011) overcome this by using a small subset of the total number of points to perform the embedding while the remaining points are filled in with trilateration, a technique based on determining the locations of points by measurement of distances, using the geometry of circles.

Our data form part of the Gondwana Geoscience Indexing Database (GO-GEOID) (Wilsher, 1995), which contains information on the minerals, geology, and tectonics of Gondwana. During the period 1999–2006, information about the mineral layer of Africa was updated, using three different sources, and today it contains commodity information from 7571 deposits. For each deposit, the commodity and the size of the commodity are reported. The size of the commodity is on a ordinal interval scale, and the boundaries between class intervals are not comparable across the three sources of the database. More than 40% of deposits have gold as a commodity, whereas the other deposits have some other mineral as a commodity. In this study, we will focus on the commodity gold.

The main objective of this paper is to produce a map of Africa displaying the potential of the commodity gold for a deposit from the GO-GEOID database. To this end we have to select an appropriate distance metric, combine the ordinal interval size classes across the three sources of the database into a unified scale, to use this new interval scale as a proxy for data reliability, and to include the geology