ARTICLE IN PRESS

Spatial Statistics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Spatial Statistics

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

Using third-order cumulants to investigate spatial variation: A case study on the porosity of the Bunter Sandstone

R.M. Lark *

British Geological Survey, Keyworth, Nottinghamshire NG12 5GG, UK

ARTICLE INFO

Article history:
Received 14 October 2013
Accepted 7 February 2014
Available online xxxx

Keywords:
Cumulants
Non-multivariate Gaussian
Bunter Sandstone

ABSTRACT

The multivariate cumulants characterize aspects of the spatial variability of a regionalized variable. A centred multivariate Gaussian random variable, for example, has zero third-order cumulants. In this paper it is shown how the third-order cumulants can be used to test the plausibility of the assumption of multivariate normality for the porosity of an important formation, the Bunter Sandstone in the North Sea. The results suggest that the spatial variability of this variable deviates from multivariate normality, and that this assumption may lead to misleading inferences about, for example, the uncertainty attached to kriging predictions.

1. Introduction

Geostatistical analysis of spatially variable geological data allows us to quantify the uncertainties in inferences made from partial samples by treating data as realizations of a random field. In most cases the underlying model is multivariate Gaussian, and the plausibility of this assumption is usually judged from the marginal distribution of observations (e.g. Webster and Oliver, 2007). Where necessary the data may be transformed, for example to logarithms or, more generally, by the Box–Cox transformation. However, it is recognized that the assumption of a Gaussian or trans-Gaussian (Gaussian after transformation) distribution is not always safe, and, particularly, that it might not hold even when it seems plausible for the marginal distribution of the data. Of particular

* Tel.: +44 0 115 9363242.
E-mail addresses: mlark@bgs.ac.uk, mlark@nerc.ac.uk.

http://dx.doi.org/10.1016/j.spasta.2014.02.001
2211-6753/

concern is the recognition that, under the multivariate Gaussian model, the first and second order moments entirely characterize the spatial distribution of a variable since all odd moments larger than the first are zero and all even moments larger than the second can be written in terms of it. However, it is known that the complex geometries that may be encountered in geological data, the strongly-connected patterns of coarse-textured alluvium in former braided streams are a *locus classicus*, might not be fully characterized by the first and second moments, and more complex spatial distributions are necessary (e.g. Guardiano and Srivastava, 1993).

It is therefore necessary to develop exploratory methods to examine the higher-order behaviour of spatially variable data. Dimitrakopoulos et al. (2010) have shown how higher order spatial cumulants of random variables can capture features of dense training images that are not compatible with the assumption of an underlying multivariate-Gaussian variable. The objective of the present paper is to show how such a cumulant can be used in an inferential framework to test the strength of evidence against the null hypothesis that, possibly relatively sparse, observations are drawn from a variable in which these cumulants take values expected in the Gaussian case; and to identify exploratory statistics that might be used to judge whether a Gaussian assumption is plausible. The approach is illustrated using data on porosity of an important sedimentary formation under the North Sea. A sound spatial stochastic model for this variable is necessary because the pore-space in this unit may be important as a site for future carbon capture and storage (Holloway, 2009).

2. Cumulants

A real-valued random variable, Z, with a probability density function $f_Z(z)$, has a moment-generating function:

$$M(v) \;=\; \mathrm{E}\left[\exp\{vZ\}\right] \;=\; \int_{-\infty}^{\infty} \exp\{vz\} f_Z(z)\mathrm{d}z. \tag{1}$$

If $M(v)$ has a Taylor series expansion about the origin then it may be written as

$$M(v) \;=\; \mathrm{E}\left[\exp\{vZ\}\right] \;=\; \mathrm{E}\left[1 + vZ + \frac{v^2}{2!}Z^2 + \cdots + \frac{v^r}{r!}Z^r + \cdots\right]. \tag{2}$$

Note that the rth non-centred moment of Z,

$$\mu'_r \;=\; \mathrm{E}\left[Z^r\right],$$

is the coefficient of $\frac{v^r}{r!}$ in the rth term in this expansion, hence the name of the function.

Cumulants of the random variable may be defined in a similar and related way. The cumulant generating function is

$$K(v) \;=\; \ln\left(\mathrm{E}\left[\exp\{vZ\}\right]\right),$$

so we may write

$$1 + \mu'_1\frac{v}{1!} + \mu'_2\frac{v^2}{2!} + \cdots + \mu'_r\frac{v^r}{r!} + \cdots = \; \exp\left\{\kappa_1\frac{v}{1!} + \kappa_2\frac{v^2}{2!} + \cdots + \kappa_r\frac{v^r}{r!} + \cdots\right\}, \tag{3}$$

where κ_r is the rth cumulant of Z.

The cumulants and moments of a distribution are related, for example (Kendall and Stuart, 1977)

$$\begin{aligned} \mu'_1 &= \kappa_1, \\ \mu'_2 &= \kappa_1^2 + \kappa_2, \\ \mu'_3 &= \kappa_1^3 + 3\kappa_1\kappa_2 + \kappa_3. \end{aligned} \tag{4}$$

However, cumulants have certain properties which can make them more useful than moments. In particular they generalize simply to the multivariate case (McCullagh and Kolassa, 2009). Consider an