Spatial Statistics 4 (2013) 1–13

Contents lists available at SciVerse ScienceDirect

Spatial Statistics

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

Optimized multi-phase sampling for soil remediation surveys

B.P. Marchant[a,*], A.B. McBratney[b], R.M. Lark[a,c], B. Minasny[b]

[a] *Rothamsted Research, Harpenden, Hertfordshire, AL5 2JQ, UK*

[b] *Faculty of Agriculture, Food and Natural Resources, The University of Sydney, NSW 2006, Australia*

[c] *British Geological Survey, Keyworth, Nottingham, NG12 5GG, UK*

ARTICLE INFO

Article history:
Received 28 June 2012
Accepted 9 November 2012
Available online 11 December 2012

Keywords:
Contamination
Copula
Geostatistics
Multi-phase design
Survey

ABSTRACT

We develop an algorithm for optimizing the design of multi-phase soil remediation surveys. The locations of observations in later phases are selected to minimize the expected loss incurred from misclassification of the local contamination status of the soil. Unlike in existing multi-phase design methods, the location of multiple observations can be optimized simultaneously and the reduction in the expected loss can be forecast. Hence rational decisions can be made regarding the resources which should be allocated to further sampling. The geostatistical analysis uses a copula-based spatial model which can represent general types of variation including distributions which include extreme values. The algorithm is used to design a hypothetical second phase of a survey of soil lead contamination in Glebe, Sydney. Observations for this phase are generally dispersed on the boundaries between areas which, according to the first phase, either require or do not require remediation. The algorithm is initially used to make remediation decisions at the point scale, but we demonstrate how it can be used to inform over blocks.

1. Introduction

Human-health and environmental concerns require the remediation of contaminated soils near former industrial sites throughout the world. In many cases, thresholds have been defined for

Abbreviations: AIC, Akaike information criterion; AEIL, Australian Environmental Investigation Limit; EBLUP, empirical best linear unbiased predictor; ML, maximum likelihood; pdf, probability density function; SSA, spatial simulated annealing.

* Correspondence to: British Geological Survey, Keyworth, Nottingham, NG12 5GG, UK. Tel.: +44 0 115 9363100.
E-mail address: benmarch@bgs.ac.uk (B.P. Marchant).

2211-6753/$ – see front matter
doi:10.1016/j.spasta.2012.11.001

the permissible concentration of metals and other contaminants in the soil (see e.g. Environment Agency, 2008). If the contamination is localized then spatial surveys can be conducted to suggest where concentrations are greater than these thresholds and hence remediation is required (van Meirvenne and Goovaerts, 2001). Uncertainty is inevitably attached to the results of such surveys and geostatistical techniques are used to assess the probability that a particular location is falsely designated as contaminated or not contaminated. This information, combined with an understanding of the costs of an incorrect remediation decision at a site, permits an informed decision about the extent of the remediation.

The accuracy and cost of soil-remediation surveys increase with the number of observations made. It has previously been suggested (see e.g. van Meirvenne and Goovaerts, 2001; Marchant and Lark, 2006; Verstraete and van Meirvenne, 2008; Meerschman et al., 2011) that the efficiency of surveys can be improved if they are split into a number of phases. The initial phase yields a low resolution map of soil contamination. It might show that no further measurements are required in much of the study area where the soil can be designated with great certainty as either contaminated or not contaminated. The later phases concentrate observations in parts of the study region where the contamination status is in doubt. As the survey progresses the resolution of the contamination map in these regions increases until eventually it is suitable for selecting the locations which are to be remediated. Heuvelink et al. (2010) consider a related problem in the design of mobile radioactivity monitoring networks. Normally the network is fairly coarse but in the event of a nuclear accident more sensors are required close to the accident site.

There are two key issues to address before such a multi-phase strategy can be used in practice. The first is the amount of additional sampling. How many observations should be made, how should they be divided between phases and how should the practitioner decide when a survey is adequate? The second issue is the selection of the locations of observations within a single phase of the survey. We consider the situation where a phase of sampling has been conducted and kriging (Webster and Oliver, 2007) has been used to predict the contamination across the study region. Two factors dictate whether further sampling is advantageous at a particular location $\mathbf{x}$. The first is how close the local prediction of the soil contamination $\widehat{z}(\mathbf{x})$ is to the threshold z_c. The second is the uncertainty of this prediction. This uncertainty can be expressed in terms of the kriging variance $\sigma^2(\mathbf{x})$. Juang et al. (2008) and van Meirvenne and Goovaerts (2001) considered how the proximity of predictions to the threshold could be incorporated into a design algorithm. They suggested that the most beneficial locations for making additional observations are where $|\widehat{z}(\mathbf{x}) - z_c|/\sigma$ is small. Thus they could order every potential observation location according to this criterion. This approach led to clusters where it was desirable to observe the contamination because existing observations were sparse and predictions were close to z_c. However they could not forecast the effect that additional sampling would have on this criterion because the new value of $\widehat{z}(\mathbf{x})$ depended on the new observations. Therefore they had to make intuitive decisions about the intensity with which each cluster was sampled and the total number of observations.

Demougeot-Renard et al. (2004) addressed this problem in a multi-phase survey of soil contamination at a former smelter in France. Following the initial survey, they selected additional sampling sites which greatly reduced the cost of misclassifying the remediation requirements of the soil. They simulated an observation, conditional upon the existing observations, at each site in their proposed design. They then used these simulated observations to estimate the cost and to determine whether the design was fit for purpose. However, because their updated objective function was calculated from a single realization of the new design they could not determine the uncertainty associated with it or know whether it was truly representative of the proposed design. Also rather than using a numerical algorithm to optimize their additional sampling they compared the values of their objective function for different designs which were selected according to intuitive rules.

We develop a Monte Carlo multi-phase sampling strategy. Later phases of the survey are optimized to minimize the expected total loss from misclassifications of the contamination status of the soil. The expected total loss is estimated through multiple conditional simulations from a parametric model of spatial variation that is fitted to available data. The expected loss is referred to as the objective function of the optimization and it is minimized by a numerical procedure called spatial simulated annealing (SSA; van Groenigen et al., 1999). Our algorithm is an advance upon existing techniques for the