Geoderma 209–210 (2013) 65–74

Contents lists available at SciVerse ScienceDirect

Geoderma

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

Land use and lead content in the soils of London

R.M. Lark *, C. Scheib

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

ARTICLE INFO

Article history:
Received 18 March 2013
Received in revised form 4 June 2013
Accepted 9 June 2013
Available online 4 July 2013

Keywords:
Lead content
Urban soils
Land use
Orthogonal contrasts
Cokriging

ABSTRACT

It is important to understand how and where pollution and other anthropogenic processes compromise the ability of urban soil to serve as a component of the natural infrastructure. An extensive survey of the topsoil of the Greater London Area (GLA) in the United Kingdom has recently been completed by a non-probability systematic sampling scheme. We studied data on lead content from this survey. We examined an overall hypothesis that land use, as recorded at the time of sampling, is an important source of the variation of soil lead content, and we examined specific orthogonal contrasts to test particular hypotheses about land use effects. The assumption that the residuals from land use effects are independent random variables cannot be sustained because of the non-probability sampling. For this reason model-based analyses were used to test the hypotheses. One particular contrast, between the lead content in the soil of domestic gardens and that in the soil under parkland or recreational land, was modelled as a spatially dependent random variable, predicted optimally by cokriging.

We found that land use is an important source of variation in lead content of topsoil. Industrial sites had the largest mean lead content, followed by domestic gardens. Detailed contrasts between land uses are reported. For example, the lead content in soil of parkland did not differ significantly from that of recreational land, but the soil in these two land uses, considered together, had significantly less lead than did the soil of domestic gardens. Local cokriging predictions of this contrast varied substantially, and were larger in outer parts of the GLA, particularly in the south west.

1. Introduction

Traditionally soil survey and inventory has been focussed in the rural environment, to provide information for agricultural extension, catchment management, etc. In the last 25 years or so there has been increasing interest in the soil of urban environments (Bullock and Gregory, 1991). It is recognized that these soils are influenced by human activity to a unique extent (Craul, 1985) and that urban soil quality can have a direct influence on human health. In particular humans are exposed to contaminants present in the soil both through inhalation and ingestion (Mielke et al, 2007), and this may have direct consequences for health and wellbeing (Miranda et al., 2007).

This growing awareness of the importance of urban soils has resulted in increasing activity to survey them. Johnson et al (2011) provide accounts of geochemical surveys of urban soils in several European cities, and urban sites have been introduced into the network of long-term ecological research sites in the United States (Bain et al., 2012). In 1992 the British Geological Survey (BGS) extended its geochemical baseline survey of the environment, which includes soil sampling, to include urban soils (Fordyce et al., 2005). Since then more than 20 major urban centres in the UK have been surveyed. The resulting data have been used to study the spatial distribution of potentially hazardous elements in soil, notably metals (e.g. Marchant et al., 2011; Rawlins et al., 2005).

Lead in urban soil is of considerable interest. It is strongly influenced by human activity, and diffuse lead pollution of the soil arises from industrial use of lead and from the past use of lead tetraethyl as an additive in petrol. In the past lead piping was extensively used for plumbing and lead was a major constituent of paints. All these anthropogenic sources of lead are particularly dense in the urban environment, and so lead is commonly elevated both in urban soils (Clark et al, 2006) and in periurban soils (Rawlins et al, 2012). Furthermore, lead is persistent in soils and sediments because it is strongly bound by various soil minerals (Maurice, 2009), so historical land use may have a pronounced 'signature' in the contemporary content of soil lead. This lead is a threat to human health. The resuspension of soil lead is an important source of atmospheric lead, which may be inhaled and so absorbed (Laidlaw and Filippelli, 2008). The ingestion of soil is another important pathway by which soil lead can be absorbed, particularly by children (Guney et al., 2010). Another pathway for the uptake of lead is possible if some urban soils, such as domestic gardens, are used to grow vegetables for human consumption (Davies et al, 1979).

It is therefore important to understand the origins and distribution of lead, as well as other metals, in the urban soil. In the present paper we use data from a recent extensive survey of the topsoil across the whole of London to examine specific orthogonal hypotheses about

* Corresponding author.
E-mail address: mlark@nerc.ac.uk (R.M. Lark).

0016-7061 © 2013 The Authors. Published by Elsevier B.V. Open access under CC BY license.
http://dx.doi.org/10.1016/j.geoderma.2013.06.004

how land use, recorded at the time of sampling, accounts for variations in total soil lead content. This requires a model-based analysis because the survey, like many others carried out to provide geochemical baseline data, was not conducted by probability sampling. Such an analysis provides estimates of mean effects (such as the difference in lead content between two particular land uses) but these effects may vary spatially in ways that give insight into the nature of the land use effect. We therefore examine how a particular land use effect on soil lead content varies spatially across the Greater London Area using an optimal cokriging estimator.

2. Materials and Methods

2.1. The London Earth survey

From 2005 to 2009 BGS undertook soil sampling in the Greater London Authority (GLA) area which comprises the City of London and the London Boroughs (Knights and Scheib, 2010). Details of the survey methodology are provided by Johnson (2005) but we summarize the key points here.

Sampling was undertaken according to a non-probability systematic design. Each 1 × 1-km square of the British National Grid within the GLA area was sampled, with a sample site in each of the four 500 × 500-m quadrants. The sample site was placed as close as possible to the centre of the quadrant avoiding any obvious sources of point contamination such as spoil heaps. The surveying team identified the land use from a series of land use codes (see Table 1) at each sample site. At each site a soil specimen was collected with an auger at the centre and vertices of a 20 × 20-m square centred on the site. The soil specimen was collected with an auger from depth 0 to 15 cm after removal of any surface litter. The five specimens were then aggregated to form a single bulk specimen for the site. At one sample site in every 100 a duplicate aggregate specimen was formed from cores collected at the centre and vertices of an adjacent 20 × 20-m square.

The aggregated material from each site was subsequently air-dried, disaggregated and sieved through a nylon sieve to pass 2 mm and sub-sampled by coning and quartering. A 50-g sub-sample was ground in an agate planetary ball mill until 95% of the material was finer than 53 μm. Concentrations (totals) of 50 major and trace elements were determined for each sample by wavelength dispersive X-ray fluorescence (XRF) spectrometry. Data were obtained for a total of 6245 sites.

2.2. Exploratory data analysis

Summary statistics were computed for the data on lead content and a transformation to logarithms was considered. The residuals of lead content from land use class mean were also examined, both with and without transformation of the original data. In general if the conventional coefficient of skewness is out with the range [−1.0, 1.0] then a transformation is deemed necessary (Webster and Oliver, 2007). However, the conventional coefficient of skewness is very susceptible to outlying values, and decisions on transformation should be based on the shape of the distribution of the underlying variable, separate from the effects of any outlying values. For this reason we used the octile skewness, which is a robust measure of the asymmetry of the distribution of a variable, insensitive to outliers (Brys et al., 2003). An equivalent rule of thumb to that of Webster and Oliver (2007) for interpretation of the conventional skewness coefficient is to consider transformations of a variable if the octile skewness is not in the range [−0.2, 0.2] (Lark et al., 2006). By basing decisions on transformation on the octile skewness we aim to use transformations only when this is necessary to justify the assumption that the data are drawn from an underlying normal random variable, possibly with outliers present. Note that, because some of the recorded lead concentrations were zero it was necessary to add a small constant (0.1) to the data before the transformation.

Table 1
Land use classes and numbers of observations in each.

Class	Number of observations
Arable	270
Commercial and residential	195
Domestic garden	1554
Industrial	64
Other	197
Park	828
Pasture	144
Recreational	657
Road verge	597
Rough grazing	346
Urban open space	1119
Woodland or forest	274

2.3. Overall land use effects

2.3.1. The model, estimation and inference

Table 1 shows the land use classes that were considered in this study. The objective was to examine the evidence for overall differences between the classes with respect to lead content, and to consider more specific questions about contrasts between particular land uses or groups of land uses.

One method to address such a question is the analysis of variance, with the partition of the sum-of-squares for differences between p land use classes into components that correspond to particular contrasts. That method cannot be used in this case because the analysis of variance is based on the assumption that the residuals from land use class means can be regarded as independent random variables. That assumption is justified by the use of an appropriate probability sampling design, such as simple random sampling or stratified random sampling (de Gruijter et al., 2006). Such an assumption cannot be made with these data because the sample sites are selected according to a systematic rule with no element of randomization. For this reason a model-based analysis is necessary. We regard the $n \times 1$ vector of observations $\mathbf{z}$ as a realization of a random function, $\mathbf{Z}$ which is described by the linear mixed model

$$\mathbf{Z} = \mathbf{M}^T \boldsymbol{\beta} + \boldsymbol{\eta} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\mathbf{M}$ is a $n \times p$ vector that indicates the land use class present at each location so that the element in the ith column and jth row is 1.0 if the ith class occurs at the jth sample location, and is zero otherwise. The $p \times 1$ vector $\boldsymbol{\beta}$ contains the fixed effects coefficients which here are mean values of the variable of interest within each land class. The terms $\boldsymbol{\eta}$ and $\boldsymbol{\varepsilon}$ are, respectively, a spatially correlated and an independently and identically distributed random variable. Ideally the components of a linear mixed model are estimated by finding the residual maximum likelihood (REML) estimates of the variance parameters of the random variables, and then generalized least squares estimates of the fixed effects coefficients (e.g. Lark and Cullis, 2004). However, this is not practical with substantial data sets such as this one, with $n = 6245$, since the REML estimation requires repeated inversion of a $n \times n$ matrix, which is computationally demanding. It was therefore decided to estimate the parameters of the random components from ordinary least squares (OLS) residuals from the class means. REML is generally preferred because estimates based on OLS residuals are prone to bias because of estimation error of the class means. However, in a large data set this estimation error, and so the resulting bias, will be small.

For this reason we used an iterative generalized least squares procedure to fit the model (Webster and Oliver, 2007). The procedure is outlined below.