Spatial modelling of concentration in topsoil using random and systematic uncertainty components: comparison against established techniques

Article  (Accepted Version)
Spatial modelling of concentration in topsoil using random and systematic uncertainty components: Comparison against established techniques

Ricardo J. N. Bettencourt da Silva*, Ariadne Argyraki, Carlos Borges, Carla Palma, Michael H. Ramsey*

a – Centro de Química Estrutural, Institute of Molecular Sciences, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal
b – Department of Geology and Geoenvironment, National and Kapodistrian University of Athens, Panepistimiopolis, Zographou, 15784, Athens, Greece
c – Instituto Hidrográfico, R. Trinas 49, 1249-093 Lisboa, Portugal
d – School of Life Sciences, University of Sussex, Falmer, Brighton, BN1 9QG, UK

* Corresponding authors:
Ricardo J. N. Bettencourt da Silva; Centro de Química Estrutural, Institute of Molecular Sciences, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal; +351 217500959; rjsilva@fc.ul.pt
Michael H. Ramsey; School of Life Sciences, University of Sussex, Falmer, Brighton, BN1 9QG, UK; +44 (0)1273 678085; m.h.ramsey@sussex.ac.uk

Abstract

The assessment contaminated land often requires the collection and analysis of soil samples, and the discrete or continuous modelling of the contamination. The continuous modelling involves interpolating analyte concentrations between sampled positions. The contamination can be expressed either as a map showing the spatial distribution across the site, or as a frequency distribution, both of which can be compared with a threshold. The characterisation of land contamination is affected by random and systematic uncertainty components of the sampling, chemical analysis, sampling positioning and modelling. This work describes the application and comparison of three techniques for developing continuous models of the contamination of a site supported on the quantification of none, some or all relevant uncertainty components. The 7.3ha site was characterised using 100 soil samples and the lead contamination modelled by “Inverse Distance Weighting” (IDW), “Ordinary Kriging” (OK) and a new Monte Carlo simulation method (MCM). The IDW only uses the positions and concentrations of the samples without their uncertainty, the OK also requires the “measurement error” and other parameters to select the variogram. The MCM uses the measurement uncertainty (including random and systematic effects arising from both sampling and analysis) and ‘GPS coordinates uncertainty’. The measurements of Pb concentration across the site were log-normally distributed, and therefore log-transformed prior to modelling. The output of the models was compared against a discrete model by ‘Probabilistic Block Mapping’ (PBM) that also considers measurement uncertainty. The OK produced a smoothed spatial variation of Pb concentration that appears the more realistic. However, IDW and OK underestimate the land contamination, while MCM prediction most closely matches that of the measured concentrations including the impact of uncertainty components. The MCM allowed a metrologically sound modelling of the contamination and the linear interpolation of data reduced assessment subjectivity. Both MCM and PBM made realistic estimates of the proportion of the site that was contaminated over the threshold and included all relevant uncertainty components in the continuous or discrete modelling, respectively. The PBM identifies areas where the
true value of the contaminant concentration could exceed the threshold value, even though the single measured value did not.

**Keywords:** Spatial modelling; Geostatistics; Monte Carlo Method; Probabilistic modelling; Land contamination; Measurement Uncertainty

**Introduction**

Areas of land can have concentrations of toxic substances that are so high as to make them unfit for agricultural production, or even for the development of residential or commercial activities. This contamination can have either a natural or an anthropogenic origin. In some countries, the national authorities for environment protection have developed procedures for the assessment of contaminated land that defines a maximum level or threshold for the most frequent soil contaminants, and most common uses of the land (Agência Portuguesa do Ambiente 2019; Department of Environmental, Food and Rural Affairs 1990). If the soil has concentrations of contaminants greater than the threshold, the risk to the human health and/or the environment is considered relevant and the land cannot be used for the proposed purpose before being remediated or subjected to alternative methods of risk reduction.

The assessment of land contamination can involve determining the fraction of its area, at a relevant depth, that presents contaminant concentrations greater than the threshold. Once the level of contaminant concentration has been mapped across the whole site, it becomes possible to identify the particular areas that require action such as remediation.

Adequate tools and procedures should be defined, therefore, both to determine the percentage of a land with contaminant concentrations greater than the threshold, and also to map the contamination for remediation planning.

The contamination modelling is affected by random and systematic uncertainty components responsible for the difference between the estimated and the “true” value of the contaminant concentration. The random uncertainty components are responsible for estimates dispersion observed from replicate sampling, sample analysis or modelling. However, the contamination model is also affected by systematic uncertainty components that cannot be quantified from replicate analysis, such as “sampling bias” and “analytical bias”. These components are responsible for a correlation between collected information on the studied site, and should be quantified and their impact on the modelling determined. The “sampling bias” component has rarely been quantified, but can be estimated using a study with multiple samplers at that test site (Ramsey et al. 2011).

The assessment of land contamination, in particular the comparison with a threshold, requires that all relevant uncertainty components are considered to guarantee the comparability of independent assessments performed on different occasions or by different experts.

In order to characterise the contamination of the soil in these two ways, soil samples should be collected and analysed from defined positions within the land. The sampling strategy should allow an even and economically feasible coverage of the whole area of land.

The accuracy of the classification of land contamination depends on the number, positions and measured contamination concentration in the collected samples, and on how these data are processed and modelled. The classification is likely to be less accurate for land with a more heterogeneous spatial distribution of contaminant, and when it has been characterised with a smaller number of samples.

The modelling of contaminated land is never totally accurate in terms of the value and spatial location of contaminant concentrations, because it is affected by uncertainty arising from at least four sources: contaminant distribution heterogeneity, spatial uncertainty of sampling positioning, and measurement uncertainty arising from both the sampling process and the chemical analysis of the laboratory sample.

These uncertainties propagate into uncertainty in the classification of the contamination across the site. Sampling is defined here as the process of collecting a sample from a specific location, with the aim of characterising that location, for instance over an area which is defined by the positioning
uncertainty. The soil at the studied location, to a specified depth, is referred to here as the “sampling target”, which is defined as the ‘portion of material, at a particular time, that the sample is intended to represent’ (Analytical Methods Committee 2005). This definition agrees with more widely used terminology (Ramsey et al. 2019), that also gives definitions of further terms that are useful for this study. Hence, sampling usually aims to characterise the specified area around the sampling location (i.e. the sampling target), rather than the whole site.

The measurement uncertainty that arises from the chemical analysis is independent from that arising from the sampling process, which originates largely from the heterogeneity of the analyte concentration within the sampling target. The measurement uncertainty can be quantified either by modelling (i.e. combining the separately quantified uncertainty from all of the measurement steps and additional effects responsible for random or systematic effects on the measured value) or empirically (e.g. by duplicated sampling and chemical analysis, followed by Analysis of Variance, ANOVA, combined with the uncertainty component arising from the estimate of bias) (Ramsey et al. 2019).

Measurement results from several different sampling targets can be used to characterise a larger area (i.e. redefining a larger or extended sampling target). The uncertainty on the mean of the estimated concentrations of the extended target will simply be the standard error on the mean combined with any relevant systematic effect. However, this mean value does not necessarily accurately estimate the mean analyte concentration for the whole site, since it does not include the analyte concentration between the sampling points. To include the area between the sampling points within the value of the site mean, some method of interpolation is required. However, interpolation will not always improve the estimate of the site mean if small ‘hot spots’ of contamination have been missed by the sampling grid.

The selection of the method of interpolation, and often modelling, will have a major impact on the inference made and on the accuracy of the interpolation. A wide range of interpolation methods have been described and compared, including inverse distance weighting, nearest neighbour methods, geostatistical interpolation methods, Kalman filter methods and Bayesian Maximum Entropy methods (Knotters et al. 2010). Some methods minimize the assumptions that are made, and are therefore more objective. Other more complex interpolation methods are based upon the assumed spatial variation of analyte concentration, and are therefore more subjective. The spatial modelling and interpolation should also allow for the fact that the input data, i.e. sampling coordinates and sample concentrations, are affected by several uncertainty sources already mentioned.

This work compares different methods and techniques that can be used to produce continuous models of the contamination of land by interpolating between georeferenced data points obtained by analysing soil samples collected at specific coordinates. The advantages and the disadvantages of these continuous techniques, and one discrete technique, are discussed through the comparison of their application to the characterisation of the Pb contamination of a land located in West London. The ability of the various techniques to model the variation in spatial composition, and to determine the percentage of the land with a concentration greater than a threshold, will be discussed.

**Continuous characterisation of the contamination of a land**

**Spatial mapping**

One of the simplest methods that is often applied to estimate analyte concentration between sampling locations, and hence create a geochemical map, is Inverse Distance Weighting (IDW). This is a deterministic technique of interpolation in which the assigned values for intermediate points are calculated using a weighted average of the measured values at the sampling locations. It is often compared against kriging (Qiao et al. 2018), discussed below. One advantage of IDW over kriging, is that requires very few decisions to be made regarding model parameters. No assumptions are required of the frequency distribution of the data for the application of IDW, giving it flexibility to deal with complex and unspecified distributions. However, it does not include assessment of prediction errors. Geostatistics is the name given to a group of statistical techniques that aim to describe the spatial (or temporal) variation of analyte concentration with distance (or time). It was initially developed for ore
grade assessment by Danie Krige (Krige 1951) in the 1950s, but has since been applied to a wide range of environmental settings (Isaaks and Srivastava 1989), including soils (Goovaerts 2001; Van Meirvenne and Goovaerts 2001; Cattle et al. 2002). The first step is the modelling of how the analyte concentration varies with distance, using a variogram. The selection of the most suitable basic shape of variogram model is somewhat subjective, but visual inspection of the scatter of the data points (ideally at least 100, (Webster and Oliver 1992)) around the model allows to goodness of fit to be judged. The second step uses the variogram to produce a map showing the interpolated estimates of concentration between all of the sampling locations using a technique called kriging, named after Krige. It is also possible to calculate the uncertainty in the interpolated estimates of concentration, called the kriging standard error of estimation. Kriging is an interpolator resulting in a more or less smoothed surface depending on the measurement error model. It uses statistical models that allow a variety of output surfaces including predictions, prediction standard errors, probabilities and quantiles. Kriging assumes the data come from a stationary stochastic process, and some kriging methods assume normally-distributed data. It is flexible, but thereby requires some subjective judgements, such as the selection of the general type of model to fit. Ordinary Kriging (OK) is the simplest form, but many more complex forms have been developed for particular purposes (Knotters et al. 2010).

Previous studies have compared IDW and OK. For example Qiao et al. (Qiao et al. 2018) compared them for arsenic contamination of soil using a technique called cross validation. They found that kriging generally underestimated the concentration and number of maximum concentration points, and similarly overestimated the minimum concentration points, due to its smoothing effect. They concluded that IDW was preferable to OK because it can therefore identify areas of high pollution more comprehensively, and it requires less subjective judgement and fewer sampling locations than kriging. This seems to be an undue exaggeration, as OK can be adjusted to be more or less smoothed, depending on the measurement error model.

Cattle et al. (Cattle et al. 2002) studied the distribution of Pb contamination of land by "multiple indicator kriging" and "kriging with cumulative distribution function of order statistics" aiming at determining which one of those methods better determines the contamination level using a threshold for Pb contamination. Both methods were equivalent in determining the spatial and frequency distribution of Pb concentration in the whole site. Method performance was assessed by a cross-validation strategy where the Pb concentration of randomly selected positions were not used to model the contamination, but instead to compare model prediction with the observed concentration in those positions. Method performance was quantified by the root mean square error. These methods did not consider sample positional and analysis uncertainty. In this study, the prediction methods generally underestimated the Pb contamination.

Borges et al. (Borges et al. 2019; Borges et al. 2020) developed a numerical Monte Carlo Method (MCM) for simulating the spatial variation of water composition in a large river area at a defined depth based on the GPS coordinates and analyte concentrations of \( m \) collected water samples, including GPS and measurement uncertainty. Each sampling point is described by three variables \((x, y, z)\) where \(x\) and \(y\) represent the longitudinal and latitudinal distance to a defined GPS coordinates origin (typically the coordinates of the first sampling) and \(z\) the estimated analyte concentration. The variables \(x\) and \(y\) of the various sampling points have negligible correlation due to possible systematic effects affecting GPS coordinates. However, the estimated \(z\) values of analyte concentration in the various samples can have a significant correlation if affected by the same large bias. The \(m\) samples are used to build a 3D surface from the combination of flat triangular surfaces that connect various sets of three point without other points within the 2D \((x, y)\) projection of the surface. Each of the 3D triangular surfaces can be used to interpolate the analyte concentration at any position in the middle of the three GPS coordinates. Since the variables of each point are affect by the respective uncertainty, in the first MCM simulation series, \(m\) new 3D points are simulated given their uncertainty building a possible 3D surface from which it is selected a random position and interpolated the respective concentration. The simulation of the 3D surface and the respective interpolation in a random position is repeated more than 10,000 times.
producing the same number of interpolated analyte concentrations from random positions inside the studied area. Since, typically, the simulation of \( z \) values only takes precision components of the measurement uncertainty into account, the measurement bias components are combined with the initially simulated \( z \) by the MCM method. The obtained 3D interpolated points are used to model the gathered knowledge on the spatial variation and the frequency distribution of analyte concentration in the studied area. This model expresses more than the best estimate of the concentration distribution; it also intends to express the uncertainty of our understanding of this distribution.

This simulation method can also be used to estimate the mean concentration of the analyte in the extended sampling target, and to evaluate the uncertainty from the determination of this mean from composite sampling in a future characterisation of the same extended sampling target or of another large area with similar analyte concentration heterogeneity. The applicability of analyte concentration heterogeneity estimated on one occasion at one site to another occasion or to another site, needs to be assessed. This applicability can be evaluated from duplicate samplings at the extended location where the agreement between measurement results is compared by using a suitable precision limit. Different types of composite samplings can be studied by this simulation method. Borges at al. (Borges et al. 2019; Borges et al. 2020) studied random and linear composite samplings where samples are randomly or linearly collected from the extended sampling target.

The current work describes the comparison of the characterisation of the contamination of a land by the existing, commonly used techniques IDW and OK and the recent MCM developed by Borges et al. (Borges et al. 2019).

The comparison will be exemplified using the contamination by Pb of land on Hounslow Health, London, previously studied by Ramsey and Argyraki (Ramsey and Argyraki 1997). The contamination at this site is also described in Example A2 of the Eurachem guide on the estimation of measurement uncertainty arising from sampling, or “sampling uncertainty” (Ramsey et al. 2019). Ramsey and Argyraki (Ramsey and Argyraki 1997) and the Eurachem guide (Ramsey et al. 2019) describe the evaluation of the uncertainty of the measurement of analyte concentration in a small area of the land around the sampling location designated as the “sampling target”. By contrast, IDW, OK and the MCM are used to characterise the concentration across the whole site; i.e. of an extended sampling target.

To improve text readability by most readers, the general term “concentration” is used here for the measurement value of the analyte, but the more specific and technically correct term for this particular case study would be “mass fraction”.

**Concentration frequency distribution**

Although contamination mapping is very useful for remediation planning, the description of soil contamination as the frequency distribution of analyte concentrations in the land summarises the contamination problem more easily. The determination of the percentage of the whole site with analyte concentrations above the defined threshold can be used to decide if the land can be used for a specific purpose, and to estimate the cost of necessary remediation.

All the previously described mapping techniques can provide this information about the studied land allowing the easy comparison of their outputs by superimposing the three distribution functions. The distribution functions can be compared with the histogram of the measured concentrations of the collected samples to check the agreement between the model and observed concentrations.

**Discrete characterisation of the contamination of a land**

Ramsey and Argyraki (Ramsey and Argyraki 1997) used the results from the analysis of samples collected at a site to describe the spatial distribution of land contamination and to determined how many sampling targets have contaminant concentrations larger than the threshold without interpolating concentrations between sampling points. This assessment was performed by either not taking, or by taking, the measurement uncertainty into account in approaches designated classification as
“deterministic” or “probabilistic”, respectively. These two approaches can be understood as the most direct interpretation of the collected experimental data because do not require defining any assumption about the spatial distribution of the analyte between the sampling points.

The “probabilistic” assessment of the contamination uses estimates of the uncertainty of each measured value of contaminant concentration (including the contributions from both sampling and chemical analysis). This measurement uncertainty for each measurement value (i.e. for each sampling target) has been defined as ‘An estimate attached to a measured value which characterises the range of values within which the “true” value is asserted to lie’ (International Organization for Standardization 1993).

In this example, each sampling target is specified as the relatively small square area (with side 30 m) centred around each sampling location. This evaluation involved the duplicate collection of samples around 10 % of sampling locations, and the duplicate chemical analysis of both duplicate samples in a balanced hierarchical design. Applying Analysis of Variance (ANOVA) to the resultant measurement values, quantifies the relevant uncertainty components, including the sampling and analytical repeatability precision. These two components are subsequently combined with the uncertainty from the analytical bias, and the sampling bias when known, together with the measured concentration value for that sampling target, estimates the measurement result. The precision uncertainty at the targets actually used for duplicate sampling will be root-two times smaller than the value applicable to most of the targets that are measured using a single analysis of a single sample.

In order to protect the analysis of variance from undue effects from a small proportion of outlier values (<10 %), a robust analysis of variance (RANOVA) can be performed (Analytical Methods Committee 1989). The direct use of classical ANOVA and RANOVA for measurement uncertainty determination assumes measured concentrations are normally distributed. If results are lognormally distributed, the original measured values should be log transformed by taking its natural logarithm (loge, or ln). After processing the transformed data by ANOVA, and combining the precision uncertainty with the bias uncertainty components, typically expressed in log-space, the estimated uncertainty in log-space can be expressed as an uncertainty factor (Ramsey et al. 2015) or back transformed by using the exponential of the confidence interval limits from the log-transformed data.

**Comparison criteria of continuous characterisation of land contamination**

The three approaches for the continuous modelling of the contamination of a land will be compared using the following three criteria:

i) The closeness of the model to the estimated concentrations of collected samples having in mind the evaluated measurement uncertainty.

ii) The plausibility of the mapping by the extent to which it produces “bull’s eyes” or other artificial model features introduced by modelling process.

iii) The frequency distribution of analyte concentrations, including the determination of the proportion of the land area with an analyte concentration greater than a specified threshold.

**Experimental**

**Studied site and used analytical method**

The investigation was undertaken of a 7.3 ha site within Hounslow Heath, West London, UK, as part of the assessment of the land for potential housing development (Ramsey and Argyraki 1997). According to the United Kingdom guidance at the time of that original survey, the Pb concentration in top soil should be below a threshold of 500 mg kg⁻¹ for domestic gardens in residential developments (Inter-departmental Committee on the Redevelopment of Contaminated Land 1987). At the time of the later interpretation (Ramsey et al. 2019), the current threshold value used was then 450 mg kg⁻¹, which was the Soil Guideline Value for Pb in residential developments with or without plant uptake (Environment Agency 2002). The primary samples, taken to a depth of 150 mm, were dried, prepared, sieved, sub-sampled and the Pb dissolved in nitric and perchloric acids, prior to analysis by ICP-AES (Inductively coupled plasma atomic emission spectroscopy).
The measured concentration of Pb at the 100 sampling targets, laid out on an orthogonal 10 × 10 grid (with 30 m spacing) across the site, are shown in tabular form in Table 1. The frequency distribution of the measured Pb concentration values is clearly positively skewed (Fig 1a). According to the Kolmogorov Smirnov test on log-transformed data, the measured values are log-normally distributed for p > 0.05. Figure 1b illustrates the symmetric shape of the histogram of log transformed concentrations.

Results and discussion
This section is divided in the discrete description of the observed land contamination by using a block mapping that takes the measurement uncertainty into account (“probabilistic block mapping”), and the continuous description of the contamination across the whole site by using “Inverse Distance Weighting”, “Ordinary Kriging”, and “Monte-Carlo Method simulations”. At the end of this section, the later three approaches are compared with each other and with the discrete description of the contamination.

Discrete description of the contamination
Ramsey and Argyraki (Ramsey and Argyraki 1997) studied the impact of measurement uncertainty on the comparison of the estimated concentrations in the various sampling targets with the threshold and classified sampling targets as “uncontaminated” or “contaminated” if the upper or lower limit of result confidence interval for 95% confidence level is below or above the threshold, respectively. In these cases, the comparison with the threshold is performed with a risk of false decision not greater than 2.5%. These authors also classified sampling targets where the confidence interval of results overlap the threshold as “probably contaminated” or “possibly contaminated” when the measured value is greater or lower than the threshold, respectively, because there is a probability of greater than 50% that the classification is correct.

This approach has been used to make a probabilistic block map (PBM) of the Pb contamination at this site when compared against the threshold value, using these same four categories (Ramsey and Argyraki 1997) (Figure 2). In this case, it is assumed the estimated Pb concentration of the sampling target is applicable to an area of 900 m² around the sampling position. In this discrete ‘block mapping’ approach, different colours are used to represent the measured Pb concentration and its classification, at each sampling location that represents a 30 m × 30 m sampling target. No interpolation method was employed to estimate the concentration between the sampling locations. When applied to this case study, a deterministic classification resulted in 8% of the target sub-areas as being contaminated over the threshold (as shown in Table 1). The probabilistic classification has been applied to the log-transformed measurements (Figure 2), at 95% confidence, using uncertainty expressed as twice the standard deviation of loge-transformed value (sG) as UG = 2sG. The value of sG (0.4817) was determined from a classical ANOVA of loge-transformed duplicated analyses of duplicated samples (see Table A2.3 of the Eurachem Guide6). The uncertainty due to the analytical bias has been estimated as a relative standard deviation of 0.0524, or 5.24% ⁴. This systematic effect in the linear domain, as it is less than 0.2, can be added into the measurement uncertainty in the log domain (Ramsey and Ellison 2017), to give a combined value of sG of 0.4845, using the equation (Ellison and Williams 2012; Cordeiro et al. 2018):

\[ s_{G,\text{meas}} = \sqrt{s_{G,\text{meas,random}}^2 + s_{\text{rel,anal,bias}}^2} \]

The class intervals can be calculated by effectively transferring the uncertainty to the threshold value (T). The concentration, cu, equal to the upper limit of the category ‘Uncontaminated’ (cu = T – UG) is 190 mg kg⁻¹ (calculated in the log-domain as 5.25, which is 6.21 – 2 × 0.48). The lower limit of ‘Contaminated’ (cₐ = T + UG) is 1318 mg kg⁻¹ (calculated in the log-domain as 7.17, which is 6.21 +
2 \cdot 0.48). These two values can also be calculated in the linear domain using the expanded uncertainty factor \((U = \exp(2s_e))\) of 2.64, as \(T/\overline{U}\) and \(T \cdot \overline{U}\).

This map shows that only 2% of the targets have a true value of Pb concentration definitely over the threshold, i.e. “Contaminated”, 6% are “Probably contaminated”, and a further 46% of the site Possibly Contaminated. Another 46% of the site was classified as definitely “Uncontaminated” using this approach. Such block maps do not resolve the variations in contaminant concentration within each sampling target. However, the implementation of such maps typically uses the sampling targets that are contaminated to initiate a follow-up survey with greater spatial resolution in order to more finely delineate the area of contamination.

**Inverse Distance Weighting modelling**

Inverse Distance Weighting (IDW) was applied to the log\(_e\)-transformed Pb concentration data (originally expressed as mg kg\(^{-1}\)), using the software ArcMap GIS with the input parameters listed in the caption to Fig. 3.

According the estimated frequency distribution from the IDW model (Figure 4), the Pb concentration varies between 89.6 mg kg\(^{-1}\) and 500 mg kg\(^{-1}\) for 95% confidence level (see P2.5 and P97.5 of Table 2). According to this model, 2.4% of the land area has a Pb concentration above 500 mg kg\(^{-1}\).

**Geostatistical modelling using Ordinary Kriging**

Ordinary kriging assumes a constant mean over the whole site. The first step was to fit a variogram model to the spatial variation (as proportion of total variance) against distance between sampling locations in meters (Fig 5).

Using ArcMap GIS, the type of model that was judged to fit the data best was stable, and the spatial variance was assumed not to be different in different directions (i.e. zero anisotropy). The parameters for the model are given in the caption of Figure 5. The measurement error was set at 34%, which is the measurement uncertainty when expressed as a proportion of the total variance in the log-domain (Table A2.3 of (Ramsey et al. 2019)).

Ordinary kriging was then applied on log\(_e\)-transformed Pb data (to fulfil assumption of normality) using this variogram model, to produce one map with the predictions of concentration at all locations, including those that were unsampled (Fig 6a). A second map was produced of kriging standard error (Fig 6b), to express the uncertainty in the predicted concentration values.

According to the frequency distribution estimated from the performed Ordinary Kriging (Figure 7), the Pb concentration varies between 56 mg kg\(^{-1}\) and 3590 mg kg\(^{-1}\) (min, max) and 1.2% of the land presents a Pb concentration greater than 500 mg kg\(^{-1}\). The 2.5\(^{th}\) and 97.5\(^{th}\) percentiles of Pb concentrations estimated from the kriging are 102 mg kg\(^{-1}\) and 438 mg kg\(^{-1}\), respectively (Table 2).

**Monte Carlo Method modelling**

Since the report of the characterisation of the site under investigation (Ramsey and Argyraki 1997) does not provide the GPS coordinates of the sampling locations, and the software developed for the MCM simulations requires that information (Borges et al. 2019), the GPS coordinates of the centre of Hounslow Heath used were those stated in Wikipedia (accessed on 2020/06/14). It was assumed that the first sample was taken at those GPS coordinates and the other sampling locations were defined by adding longitudinal degrees and subtracting latitudinal degrees to represent a square 10 by 10 array of 100 positions with orthogonal 30 m distances.

Since the Pb concentration in the land has a log-normal distribution, the measured concentrations where log\(_e\) transformed before the simulation and considered measurement precision was expressed in the log-domain. The simulated spatial Pb concentration variation in the log-domain was back transformed and combined with the estimated bias uncertainty by the MCM.

The parameters for the spatial distribution simulations were:

1. Log\(_e\)-transformation of the measured concentrations.
2. Expanded uncertainty of the coordinates: 3 m. The sample positioning was performed by using tapes and lines with an estimated spatial uncertainty for the relative positioning of points of 3 m.
3. The measurement uncertainty used to simulate possible measured concentrations was expressed in the log(e)-scale as the standard deviation of log(e)-transformed concentrations, \( s_c = 0.4817 \).
4. Bias relative standard uncertainty equal to 5.24 %.

The spatial distribution of Pb concentration in the land modelled by MCM is shown in Figure 8.

‘Hot spots’ are again evident at the two locations B2 and F4, that are also visible in the raw values shown in Table 1. Figure 9 presents the variation of the frequency of simulated Pb concentrations plotted against the concentration value in a log-scale axis. Table 2 also presents various percentiles of this distribution.

According to the MCM simulation, the Pb concentration varies between 64.2 mg kg\(^{-1}\) and 784 mg kg\(^{-1}\) for 95 % confidence level (see P2.5 and P97.5 of Table 2). A proportion of 8.7 % of the simulated Pb concentrations are greater than the threshold of 500 mg kg\(^{-1}\).

The MCM simulation method is applicable to any type of frequency distribution of the results from analysis of the collected samples due to the flexibility of the simulation method.

**Comparison of the three methods of interpolation modelling and mapping**

The choice between the three methods of geochemical mapping and/or spatial distribution modelling (IDW, OK and MCM) depends on the criteria used for assessment. One possible criterion is the closeness of agreement between the concentration values predicted by the model and the true values at all points on the map. The obvious limitation of this criterion is that the true values of concentration are unknown. The closest suitable information available is the concentration values measured at the 100 sampling location, taken together with their measurement uncertainty (which aims to include the true value). For this investigation the expanded measurement uncertainty was estimated as an uncertainty factor of 2.62, which accounts for 34 % of the total variance of the measured Pb concentrations across the site (Borges et al. 2019). It is possible therefore to judge whether the predicted concentration value of the sampling location falls within the range of the measurement uncertainty at that location. However, this is not an option for the majority of points on the map that are between the sampling locations.

The high level of measurement uncertainty is primarily caused by a large amount of analyte heterogeneity at the small-scale (i.e. within target/location). Such heterogeneity has been reported for Pb down to the centimetre scale (Ramsey et al. 2013). This suggests that a map which accurately reflected the true value of the Pb concentration at each point, down to the centimetre scale would be impossibly ‘noisy’. Some level of spatial smoothing is therefore required. This smoothing was demonstrated above for kriging, by assuming ‘measurement error’ (as measurement uncertainty) accounts for 34 % of the total variance (Fig. 6). The smoothing must necessarily remove the representation of the smaller scale variability which is known to be present, but produces a clearer picture of the larger scale trends in spatial variability.

Regarding this modelling method characteristic, the MCM directly uses the relevant measurement uncertainty for the spatial modelling making this tool accurate regarding the impact of this uncertainty source in the modelling.

A second assessment criterion could be the tendency of the model to identify and emphasise ‘bulls eyes’ of both high contamination (i.e.‘hot spots’), and low contamination. It is clear from Fig. 3 that IDW has this tendency (e.g. locations B2 and F4), although its visibility is strongly affected by the choice of class intervals. MCM also show this tendency to produce ‘bulls eyes’ to some extent (Fig. 8), at some of the same locations as seen in the IDW maps (i.e. B2 and F4). The observation that these ‘bulls eyes’ are almost perfectly circular (e.g. Fig. 3 and Fig. 8), and coincidently all centred on
of kriging, normality, kriging models, As previously observed, previously reported f
The observed underestimation of probabilistic mapping does therefore underestimated this measurement values judge interp contaminated or gave mapping estimate that 8 one particular application of this third measurement values are explained by the measurement uncertainty.
One particular application of this third assessment criteria for the models, is the proportion of the site that is classified as contaminated above the threshold value. The raw measurement values for each sampling target (Table 1), and deterministic block mapping (Ramsey and Argyraki 1997), give an estimate that 8 % of the site is over the threshold value of 500 mg kg⁻¹. When probabilistic block mapping (PBM) was applied to these targets in the log-domain without spatial interpolation, (Fig. 2) gave the proportion ranging from 2 % as definitely ‘contaminated’, 8 % as definitely and probably contaminated or 54 % at least ‘possibly contaminated’. The lowest proportion estimated by interpolation modelling is 1.2 % by ordinary kriging, followed by 2.4 % by IDW, and then 8.7 % by MCM. Whist this criterion enables a comparison to be made, there is no true value against which to judge which is the most accurate option. The prediction by MCM is closest to the 8 % estimated from the measurement values (deterministic approach), or by the probabilistic approach (PBM) taking the cases likely to be above the threshold, overestimating by 35 %. Using this same criterion, OK and IDW therefore underestimated this contaminated proportion by factors of 7 and 3 respectively. The probabilistic mapping does have the additional advantage of emphasising, that a further 46 % of targets have the probability that the true value of the Pb concentration could be over the threshold value, even though the measurement value is below.
The observed underestimation of maximum contamination level by IDW and OK confirms the previously reported findings by Qiao et al. (Qiao et al. 2018) and Cattle et al. (Cattle et al. 2002).
As previously observed, IDW does not rely on the assumption of a normal distribution as do many kriging models, but also requires none of the subjective judgements used for kriging. IDW can therefore provide a good way to take a first look at an interpolated surface, without testing for normality, but it does ignore the assessment of prediction errors. By contrast, kriging is very flexible and allows the investigation of graphs of spatial auto- and cross-correlation. However, the flexibility of kriging inevitably requires a lot of decision-making.
The MCM overcomes some of the limitation of the IDW by using the measurement uncertainty in the modelling, and it does that without the need to define subjective modelling parameters. The MCM uses the known uncertainty sources from the measurement process (sampling and chemical analysis), and from positioning, to model the spatial distribution. In addition, the MCM (and PBM) also allows for the consideration of the uncertainty from systematic effects in the modelling, such as the effects of analytical bias and sampling bias. Although an estimate of the sampling bias was not available for this case study, it could be estimated in future using one round of a sampling proficiency test conducted at this site. (Ramsey et al. 2013) The option within MCM for the linear interpolation of simulated concentrations, or log-transformed concentrations, given positional and concentration uncertainty, avoids the need for the subjective selection of nonlinear modelling parameters. Usually, not enough data is available to avoid the subjectivity of the selection of the modelling parameters in kriging. The MCM usually assumes measured values from the analysis of collected samples are normally or log-normally distributed, but it does not predefine the distribution of analyte concentration across the whole site. The MCM is flexible enough to allow the use of another frequency distribution for the measured values. It does, however, have the same tendency to produce ‘bull’s eyes’ and ‘hot lines’ around location of high and low concentrations.

Conclusions
A comparison has been made between three methods of interpolation modelling for mapping analyte concentrations in contaminated land, and a further method not using interpolation. The recently devised Monte Carlo Method of simulation (MCM) was compared against the commonly used, established methods of Inverse Distance Weighting (IDW), and Ordinary Kriging (OK), all using interpolation, and Probabilistic ‘Block’ Mapping (PBM) without interpolation. The IDW assumes collected sample results have no (or negligible) uncertainty and the OK models contamination assuming a specified random error (“measurement error”). Both MCM and PBM can take all systematic and random uncertainty component into account in the continuous or discrete modelling of the contamination, respectively.

When applied to a previously published case study of lead contaminated top soil in West London, the three criteria were used to compare these assessment techniques were the: (1) closeness of the model predictions to the measured concentration values, allowing for measurement uncertainty, (2) plausibility of the geochemical maps in avoiding modelling artefacts such as ‘bull’s eyes’, and (3) frequency distribution of analyte concentrations, and the consequent estimate of the proportion of the site contaminated above a specified threshold.

It was found that all four methods could produce maps that drew attention to the areas that had concentration values above the threshold, and would require action. However, IDW and MCM tended to produce ‘bull’s eyes’ of high, and low, modelled concentration unrealistically centred on sampling locations. OK avoided this tendency, but by smoothing the variation in concentration, which has the unwanted side effect of underestimating the high concentration values (and over estimating the low values). PBM does not model the concentration between the sampling locations, but uses the measured concentration to characterise contiguous square sampling targets centred around each location. The uncertainty on each measured value is used to assess the probability that the true value of contaminant concentration exceeds the threshold. MCM has a frequency distribution that most closely matches that of the measured concentration values, with 8.7 % of the test site predicted to be over the threshold, compared to 8 % for the measurements. IDW and OK predict the unrealistically low values of 2.4 % and 1.2 % of the site respectively, which reflects their frequency distributions that are much narrower than that for the measurement values and MCM. PBM adds the extra information that a further 46 % of the site, is likely to have a true value of Pb concentration over the threshold, which could be revealed by resampling and remeasurement. Although PBM does not aim to use interpolation to delineate the contaminated area, it does identify all of the contaminated targets that may require further closer-spaced sampling to make this delineation. Although no estimate of the sampling bias was obtained in
the studied case, this uncertainty component could be estimated through a future proficiency test conducted at the site (Ramsey et al., 2013).

The observed differences of modelling techniques can also be considered when selecting the strategy for modelling the spatial variability of other environmental media, such as sediment, water or air.

Acknowledgments
This work was supported by Fundacão para a Ciência e Tecnologia (FCT) through projects UIDB/00100/2020 and UIDP/00100/2020.
References


Table 1. Measured lead concentrations at each target on the sampling grid (mg kg\(^{-1}\)), shown by the actual coordinates used in the regular sampling grid (spacing 30 m) (Ramsey et al. 2019). Eight of the values (i.e. 8%, in bold) are above the then threshold value of 500 mg kg\(^{-1}\).

<table>
<thead>
<tr>
<th>Row</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>474</td>
<td>287</td>
<td>250</td>
<td>338</td>
<td>212</td>
<td>458</td>
<td>713</td>
<td>125</td>
<td>77</td>
<td>168</td>
</tr>
<tr>
<td>2</td>
<td>378</td>
<td>3590</td>
<td>260</td>
<td>152</td>
<td>197</td>
<td>711</td>
<td>165</td>
<td>69</td>
<td>206</td>
<td>126</td>
</tr>
<tr>
<td>3</td>
<td>327</td>
<td>197</td>
<td>240</td>
<td>159</td>
<td>327</td>
<td>264</td>
<td>105</td>
<td>137</td>
<td>131</td>
<td>102</td>
</tr>
<tr>
<td>4</td>
<td>787</td>
<td>207</td>
<td>197</td>
<td>87</td>
<td>254</td>
<td>1840</td>
<td>78</td>
<td>102</td>
<td>71</td>
<td>107</td>
</tr>
<tr>
<td>5</td>
<td>395</td>
<td>165</td>
<td>188</td>
<td>344</td>
<td>314</td>
<td>302</td>
<td>284</td>
<td>89</td>
<td>87</td>
<td>83</td>
</tr>
<tr>
<td>6</td>
<td>453</td>
<td>371</td>
<td>155</td>
<td>462</td>
<td>258</td>
<td>245</td>
<td>237</td>
<td>173</td>
<td>152</td>
<td>83</td>
</tr>
<tr>
<td>7</td>
<td>72</td>
<td>470</td>
<td>194</td>
<td>82.5</td>
<td>162</td>
<td>441</td>
<td>199</td>
<td>326</td>
<td>290</td>
<td>164</td>
</tr>
<tr>
<td>8</td>
<td>71</td>
<td>101</td>
<td>108</td>
<td>521</td>
<td>218</td>
<td>327</td>
<td>540</td>
<td>132</td>
<td>258</td>
<td>246</td>
</tr>
<tr>
<td>9</td>
<td>72</td>
<td>188</td>
<td>104</td>
<td>463</td>
<td>482</td>
<td>228</td>
<td>135</td>
<td>285</td>
<td>181</td>
<td>146</td>
</tr>
<tr>
<td>10</td>
<td>89</td>
<td>366</td>
<td>495</td>
<td>779</td>
<td>60</td>
<td>206</td>
<td>56</td>
<td>135</td>
<td>137</td>
<td>149</td>
</tr>
</tbody>
</table>

Reproduced with permission from (Ramsey et al. 2019) by Publisher.
Table 2. $x$th percentiles, $P_x$, of modelled or simulated Pb contaminations by “Inverse Distance Weighting” (IDW), “Ordinary kriging” (OK) and the “Monte Carlo Method” (MCM).

<table>
<thead>
<tr>
<th>$P_x$</th>
<th>IDW Concentration (mg kg$^{-1}$)</th>
<th>OK Concentration (mg kg$^{-1}$)</th>
<th>MCM Concentration (mg kg$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P0.5</td>
<td>68.4</td>
<td>81.2</td>
<td>47.7</td>
</tr>
<tr>
<td>P2.5</td>
<td>89.6</td>
<td>102</td>
<td>64.2</td>
</tr>
<tr>
<td>P50</td>
<td>221.8</td>
<td>211</td>
<td>204</td>
</tr>
<tr>
<td>P97.5</td>
<td>500.3</td>
<td>437.5</td>
<td>784</td>
</tr>
<tr>
<td>P99.5</td>
<td>655.5</td>
<td>501.3</td>
<td>1520</td>
</tr>
</tbody>
</table>
Fig 1. Histograms of the Pb concentration (in mg kg\(^{-1}\)) measured in 100 soil samples (Table 1) shown on (a) the original linear scale (b) after natural logarithms were taken (Ramsey et al. 2019).
Fig 2. Probabilistic Block Map (PBM) classifying the contamination of 100 sampling targets (each 30 m by 30 m) at the Hounslow site. Classification compares the log$_e$-transformed measurement results, reported with an expanded uncertainty for 95 % confidence level, against a Pb threshold concentration of 500 mg kg$^{-1}$ (6.21 in the log$_e$-domain). A sampling target is designated “uncontaminated” or “contaminated” when there is, at least, a 97.5 % probability of this claim being correct. For measured concentrations closer to the threshold, sampling target is considered “possibly contaminated” and “probably contaminated” when the probability of being correct is greater than 50 %.
Figure 3. Geochemical map of Pb concentration (in mg kg$^{-1}$) in the soil at the Hounslow site, using Inverse Distance Weighting (IDW) with six geometric classes. Two ‘hot spots’ of particularly high concentration are centred on locations B2 and F4, seen in the raw values given in Table 1. Input parameters were: Power: 2; Searching neighbourhood: Standard; Neighbours to include: 15 (at least 10); Sector type: Four and 45 degrees, Major and Minor semi-axis: 95.46; Angle: 0.
Figure 4. Frequency distribution and percentiles of modelled Pb concentrations at the Hounslow site by Inverse Distance Weighting, based on 8281 point estimates on a 3 m × 3 m grid cell.

Figure 5. The stable variogram model fitted to the log$_e$-transformed Pb measured concentrations for the Hounslow site. The parameters of the model are: Number of lags: 11; Lag size: 30; Nugget: 0.18; measurement error 34%; Range: 330; Partial sill: 0.4466.
Figure 6. Maps of (a) predicted Pb concentration (mg kg\(^{-1}\)), and (b) its kriging standard error, produced by Ordinary Kriging for the Hounslow site. The ‘hot spots’ at location B2 and F4 (visible in Fig. 2, Fig.3 and Table 1) are still evident as discrete concentration peaks (Fig. 5a). The input data were: Trend type: None; Searching neighbourhood: Standard; Neighbours to include: 5 (Include at least 2); Sector type: eight and 45 degree; Major and Minor semiaxis: 330; Angle: 0.
Figure 7. Frequency distribution and percentiles of modelled Pb concentrations in the studied site estimated using Ordinary Kriging, based on 8281 point estimates on a 3 m × 3 m grid cell.

Figure 8. Spatial distribution of Pb concentration (mg kg⁻¹) at the Hounslow site simulated by the MCM, and displayed using six geometric class intervals.
Figure 9. Frequency distribution and percentiles of the simulated Pb concentrations in the studied site using the MCM.
Figure 10. Superimposition of the variation of the normalised frequency of Pb concentrations with the concentration value in a log-scale, modelled or simulated by the three different approaches, compared to the frequency of the measured concentration (Fig. 1).