Supplemental Material

Analyzing the Spatial Distribution of PCB Concentrations in Soils Using Below–Quantification Limit Data

Thomas G. Orton, Nicolas P. A. Saby,† Dominique Arrouays, Claudi C. Jolivet, Estelle J. Villanneau, Jean-Baptiste Paroissien, Ben P. Marchant, Giovanni Caria, Enrique Barriuso, Antonio Bispo, and Olivier Briand

Monte Carlo Maximum Likelihood parameter estimation with censored data

In the following, we will use $f(\cdot | \cdot)$ to denote a conditional distribution for the variables or parameters on the left of the vertical line, given the variables or parameters on the right of the line. The likelihood function when some of the data are censored, Equation (3) of the main text, can be written as:

$$L_C(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{X}, \kappa; D) = L(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{X}, \kappa; \mathbf{s}_o) \int_{s_c=-\infty}^{\infty} f(s_c | \mathbf{s}_o, \boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{X}, \kappa) \, g(s_c) \, ds_c. \quad (S1)$$

where $g(s_c)$ is the indicator function, $g(s_c) = 1$ if all values of $s_c$ are less than $QL_S = \ln QL$, $g(s_c) = 0$ otherwise: this means that we can write $\infty$ for the upper limit of integration. We write $D$ for the total dataset (i.e. both the observed and the censored data). The first term on the right-hand-side, $L(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{X}, \kappa; \mathbf{s}_o)$, is the likelihood function for the parameters given $\mathbf{s}_o$ only, defined by Equation (2) of the main text. The term inside the integral, $f(s_c | \mathbf{s}_o, \boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{X}, \kappa)$, is a conditional distribution for $s_c$, given the parameters and models on the right-hand-side of the vertical line. In this case, the conditional distribution is defined by the by a multivariate Gaussian probability density function (pdf), with mean $\mathbf{s}_{c;SK}$ and covariance matrix $V[\mathbf{s}_{c;SK}]$; $\mathbf{s}_{c;SK}$ is the (multi-point) simple kriging prediction of $s_c$ given
observed data $s_o$ and the parameters and models defined to the right-hand side of the vertical line, and $V[s_{c,SK}]$ is the associated (multi-point) simple kriging covariance matrix.

In the MCML method we approximate and maximize the likelihood, Equation (S1). To do this, we define a simulating distribution, $f^*(s_c)$, which is designed to be close to the distribution for $s_c$ given the data and ML parameters. In this work, we define $f^*(s_c)$ using a constant mean, $\beta^*$, (i.e. $X^*$ is a vector of ones), and a Gaussian covariance function, $\kappa^*$, parameterized by $\theta^*$. (We choose the simulating parameters, $\beta^*$ and $\theta^*$, based on a short initial run of the MCML algorithm.) Our simulating distribution is then defined as the conditional distribution for $s_c$ given the data (both the observed and the censored data) and these simulating parameters:

$$f^*(s_c) = \frac{f(s_c | s_o, \beta^*, \theta^*, X^*, \kappa^*) g(s_c)}{\int f(s_c | s_o, \beta^*, \theta^*, X^*, \kappa^*) g(s_c) \, ds_c}.$$  

(S2)

This distribution is a multivariate Gaussian distribution with mean $\hat{s}_{c,SK}^*$ and $V^*[\hat{s}_{c,SK}^*]$, which has been truncated so that only values less than $QL_S$ are possible; $\hat{s}_{c,SK}^*$ and $V^*[\hat{s}_{c,SK}^*]$ are defined similarly to $\hat{s}_{c,SK}$ and $V[\hat{s}_{c,SK}]$, but using the simulating parameters and models (rather than the parameters and models for which likelihood function $L_C$ is being evaluated, $X$ and $\kappa$).

To help visualize this distribution, we provide supplemental Figure S1. This is based on a simple example, using all 68 of the observed data in our study, and the censored data at just two locations. The figure displays 5000 simulated samples from the simulating distribution, Equation (S2), defined based on the same simulating parameters we use in the paper: $X^*$ is a vector of ones, $\kappa^*$ a Gaussian covariance function.

We can multiply and divide the integrand in Equation (S1) by $f^*(s_c)$:
\[
L_C(\beta, \theta, X, \kappa; D) = L(\beta, \theta, X, \kappa; s_o) \int_{s_c=-\infty}^{\infty} \frac{f(s_c | s_o, \beta, \theta, X, \kappa) g(s_c)}{f^*(s_c)} f^*(s_c) \, ds_c
\]

\[
= L(\beta, \theta, X, \kappa; s_o) \int_{s_c=-\infty}^{\infty} \frac{f(s_c | s_o, \beta, \theta, X, \kappa) g(s_c)}{\int f(s_c | s_o, \beta^*, \theta^*, X^*, \kappa^*) g(s_c) \, ds_c} f^*(s_c) \, ds_c , \quad (S3)
\]

and because the term, \( \int f(s_c | s_o, \beta^*, \theta^*, X^*, \kappa^*) g(s_c) \, ds_c \), in the denominator does not depend on \( \beta, \theta, X \) or \( \kappa \), and is effectively a constant (which we rewrite as \( A^{-1} \)), we have:

\[
L_C(\beta, \theta, X, \kappa; D) = A^{-1} L(\beta, \theta, X, \kappa; s_o) \int_{s_c=-\infty}^{\infty} \frac{f(s_c | s_o, \beta, \theta, X, \kappa)}{\int f(s_c | s_o, \beta^*, \theta^*, X^*, \kappa^*) g(s_c) \, ds_c} f^*(s_c) \, ds_c . \quad (S4)
\]

The integral in Equation (S4) defines the expectation of the quotient in the equation with respect to the simulating distribution, \( f^*(s_c) \). Therefore, if we can draw \( J \) samples, \( s_c(j) \), \( j = 1, \ldots, J \) of \( S_c \) from the simulating distribution, then we can compute a Monte Carlo approximation to Equation (S4):

\[
L_C(\beta, \theta, X, \kappa; D) \approx A^{-1} L(\beta, \theta, X, \kappa; s_o) \frac{1}{J} \sum_{j=1}^{J} \frac{f(s_c(j) | s_o, \beta, \theta, X, \kappa)}{\int f(s_c(j) | s_o, \beta^*, \theta^*, X^*, \kappa^*)} . \quad (S5)
\]

The \( J \) samples, \( s_c(j) \), from the simulating distribution can be drawn using a Markov chain Monte Carlo (MCMC) method — see Gilks et al. (1996) for a useful introduction to the theory and applications of MCMC.

Parameters are chosen to maximize the Monte Carlo approximation, Equation (S5). The constant term \( A^{-1} \) need not be calculated. For parameter estimation and model comparison, we found that we obtained sufficiently stable results with \( J=5000 \). We tested the sensitivity to the choice of simulating distribution by comparing results with those from an alternative
choice of $\mathbf{\beta}^*$, $\mathbf{\theta}^*$, $\mathbf{X}^*$ and $\kappa^*$. We defined this alternative using the models selected by AIC as the best explanation of the PCB-187 data, $\mathbf{X}^* = \mathbf{X}^E$ and $\kappa^* = \kappa^E$, and the associated fitted parameters, $\mathbf{\beta}^* = \mathbf{\beta}$ and $\mathbf{\theta}^* = \mathbf{\theta}$.

Bayesian prediction based on censored data

In the following description, we drop the terms for the fixed effects and covariance models, $\mathbf{X}$ and $\kappa$, from our notation. These were required in the previous section to distinguish between the simulating model ($\mathbf{X}^*$ and $\kappa^*$), and the model for which $L_C$ was being evaluated ($\mathbf{X}$ and $\kappa$). In this section, we presume that $\mathbf{X}$ and $\kappa$ are known. In the following, we will use $f(\cdot | \cdot)$ to denote a general conditional distribution, rather than its explicit functional form.

To implement the Bayesian approach for spatial prediction, we must set a prior distribution for the parameters, $f_0(\mathbf{\beta}, \mathbf{\theta})$, which we define in a similar way to Orton et al. (2009). We assume independent priors for the parameters that represent limited prior knowledge. We use a flat improper prior (i.e. cannot be made to integrate to one) for the $p$-vector of regression parameters, $\mathbf{\beta}$. We parameterize the Gaussian and exponential covariance functions by $\mathbf{\theta} = \{ \sigma^2, \xi, r \}$. For the total variance, $\sigma^2$, we adopt an improper prior that is proportional to $1/\sigma^2$ (or equivalently, that is flat for ln $\sigma^2$); Berger et al. (2001) showed that this prior for $\sigma^2$ and $\mathbf{\beta}$ gives rise to proper posterior distributions provided that proper prior distributions are used for other covariance parameters. For the proportion of the variance that is spatially correlated, $\xi$, we use a uniform prior on the interval $[0, 1]$, and for the effective range parameter, $r$, we consider an inverse Gamma prior whose mode is at an a priori guess of the value of $r$ (from inspection of the experimental variogram of the raw data). Thus we write the prior distribution for the parameters as:
where the subscript ‘0’ indicates that this is the prior distribution for the arguments.

The Bayesian predictive distribution is given by an integral (De Oliveira, 2005):

\[ p(S(x_{\text{pred}}) = s_{\text{pred}} \mid D) = \int \int f(s_{\text{pred}} \mid s_o, s_c, \beta, \theta) f(s_c, \beta, \theta \mid D) \, d\beta \, d\theta \, ds_c, \]  

where \( D \) represents the full (censored and observed) data. The term \( f(s_{\text{pred}} \mid s_o, s_c, \beta, \theta) \) is the Gaussian pdf with mean and variance given by simple kriging at \( x_{\text{pred}} \), conditional on the observed data, \( s_o \), the values of \( S \) at the censored data locations, \( s_c \), and the parameters on the right-hand-side of the vertical line. The other term inside the integral, \( f(s_c, \beta, \theta \mid D) \), represents the uncertainty in the conditioning variables and parameters, \( s_c, \beta \) and \( \theta \), and is called the ‘posterior’ distribution, given the data, \( D \). From the expression on the final line of Equation (4.1) of De Oliveira (2005), this posterior distribution is given by:

\[ f(s_c, \beta, \theta \mid D) \propto f_0(\beta, \theta) f(s_c, s_o \mid \beta, \theta) g(s_c), \]  

where \( g(s_c) \) is the indicator function defined in the previous section. Since \( g(s_c) = 0 \) if any element of \( s_c \) is greater than \( QL_S \), we write \( \infty \) for the upper limit of the integral in Equation (S7).

It is possible to approximate directly the predictive distribution, Equation (S7), by a Monte Carlo average:

\[ p(S(x_{\text{pred}}) = s_{\text{pred}} \mid D) = \frac{1}{M} \sum_{m=1}^{M} f(s_{\text{pred}} \mid s_o, s_c(m), \beta(m), \theta(m)), \]  

where \( f_0(\beta, \theta) \) is the prior distribution for the arguments.
where \( \{ \mathbf{s}_c(m), \mathbf{\beta}(m), \mathbf{\theta}(m) \} \) is the \( m \)th of \( M \) sets of samples drawn from the posterior distribution, Equation (S8). These samples could be drawn by MCMC sampling. However, to improve the efficiency of the MC approximation, we can integrate \( \mathbf{\beta} \) and \( \sigma^2 \) out of the predictive distribution analytically. We show how in the next section (Analytic integration of Equation (S7) with respect to \( \mathbf{\beta} \) and \( \sigma^2 \)), following the working of De Oliveira et al. (1997).

The analytic integration of \( \mathbf{\beta} \) and \( \sigma^2 \) leads to:

\[
p\left( S(x_{\text{pred}}) = s_{\text{pred}} \mid D \right) = \int_{s_c = -\infty}^{\infty} \int_{\xi, r} f\left( s_{\text{pred}} \mid s_o, s_c, \xi, r \right) f\left( s_c \mid D \right) \, d\xi \, dr \, ds_c , \tag{S10}
\]

where \( \mathbf{\theta} = \{ \sigma^2, \xi, r \} \) were the full covariance function parameters. The conditional distribution, \( f\left( s_{\text{pred}} \mid s_o, s_c, \xi, r \right) \), in Equation (S10) is given by the three-parameter version of Student’s \( t \)-distribution, with pdf:

\[
f\left( s_{\text{pred}} \mid s_o, s_c, \xi, r \right) = \frac{\Gamma\left( \frac{N - p + 1}{2} \right)}{\Gamma\left( \frac{N - p}{2} \right) \sqrt{N \pi \text{var}\left[ \hat{s}_{\text{pred;UK}} \right]}} \left[ 1 + \frac{(s_{\text{pred}} - \hat{s}_{\text{pred;UK}})^2}{N \text{var}\left[ \hat{s}_{\text{pred;UK}} \right]} \right]^{-\frac{N-p+1}{2}} . \tag{S11}
\]

We put \( \mathbf{s} = [s_o^T, s_c^T]^T \) and define correlation matrix, \( \mathbf{\varphi} = \frac{1}{\sigma^2} \mathbf{C} \), explicitly using the parameters \( \xi \) and \( r \). Then, with \( \hat{\sigma}^2 = s^T[\mathbf{\varphi}^{-1} - \mathbf{\varphi}^{-1} \mathbf{X} (\mathbf{X}^T \mathbf{\varphi}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{\varphi}^{-1}]s/N \) (this is the maximum likelihood estimate of \( \sigma^2 \) given \( s \) and \( \{ \xi, r \} \)), the distribution in Equation (S11) is centered on the universal kriging prediction, \( \hat{s}_{\text{pred;UK}} \), given \( s_o, s_c, \xi, r \) and \( \hat{\sigma}^2 \). The term \( \text{var}\left[ \hat{s}_{\text{pred;UK}} \right] \) in Equation (S11) is the associated universal kriging variance. See De Oliveira et al. (1997) for more details.

The posterior distribution in Equation (S10) is defined by:
The term, \( f(s_c, s_o \mid \xi, r) \), is sometimes referred to as the ‘integrated likelihood function’ (see e.g. Berger et al., 2001), and is given by:

\[
f(s_c, s_o \mid \xi, r) \propto \frac{1}{|\varphi|^{\frac{N}{2}} |X^T \varphi^{-1} X|^{\frac{1}{2}} \left\{ s^T [\varphi^{-1} - \varphi^{-1} X (X^T \varphi^{-1} X)^{-1} X^T \varphi^{-1}] s \right\}^{-\frac{N-p}{2}}}
\]  

(S13)

As in Equation (S9), the predictive distribution can be approximated by a Monte Carlo average. We used MCMC to draw \( M \) sets of samples of the parameters, \( \{\xi(m), r(m)\} \), \( m = 1, \ldots, M \), and \( M \) samples of \( S_c, s_c(m) \), \( m = 1, \ldots, M \), from the posterior distribution, Equation (S12). If required, the full predictive pdf can then approximated by:

\[
p(S(x_{\text{pred}}) = s_{\text{pred}} \mid D) = \frac{1}{M} \sum_{m=1}^{M} f(s_{\text{pred}} \mid s_o, s_c(m), \xi(m), r(m)) .
\]  

(S14)

We use the mean of this distribution as our prediction of \( S(x_{\text{pred}}) \):

\[
\hat{s}_{\text{pred}} = \frac{1}{M} \sum_{m=1}^{M} \hat{s}_{\text{pred};UK}(m) .
\]  

(S15)

To provide predictions on the original scale, we use the median back-transform, which is recommended by Tolosana-Delgado and Pawlowsky-Glahn (2007) when kriging has been performed on a log-transformed variable. Assuming that the predictive distribution, Equation (S14), has zero skew on the transformed scale, the median will be equal to the mean, Equation (S15). Then, the exponential of Equation (S15) is used as our prediction on the original scale.
The probability of $S$ being less than a value, $s_{\text{pred}}$ say, is given by the average of the cumulative distribution function (cdf) values (corresponding to the predictive pdf), which are given by:

$$F \left( s_{\text{pred}} \mid s_o, s_c(m), \xi(m), r(m) \right) = F_t \left( \frac{s_{\text{pred}} - \hat{s}_{\text{pred;UK}}(m)}{\sqrt{N \text{ var}[\hat{s}_{\text{pred;UK}}(m)]}}; N - p \right).$$  \hspace{1cm} (S16)

where $F_t(x; \upsilon)$ denotes the cdf of the standard $t$-distribution with $\upsilon$ degrees of freedom.

We tested the robustness of the MCMC prediction algorithm for computing predictions, pdfs, and cdfs for 100 randomly chosen test locations from the prediction grid. We repeated the algorithm five times with $M=5000$, and found that the maximum standard error between the estimates was 0.00035 for the predictions, 0.000001 for the pdf values, and 0.00001 for the cdf values, indicating sufficient precision for our needs.

**Analytic integration of Equation (S10) with respect to $\beta$ and $\sigma^2$**

We begin by using Equations (9) and (11) of De Oliveira et al. (1997) to write:

$$\int_{\beta, \theta} \int f(s_{\text{pred}}, s_c \mid s_o, \beta, \theta) f(\beta, \theta \mid s_o) \, d\beta \, d\theta = \frac{\int_{\xi, r} f(s_{\text{pred}}, s_c \mid s_o, \xi, r) f(s_o \mid \xi, r) f_0(\xi, r) \, d\xi \, dr}{\int_{\xi, r} f(s_o \mid \xi, r) f_0(\xi, r) \, d\xi \, dr}. \hspace{1cm} (S17)$$
where \( \Phi = \{ \sigma^2, \xi, r \} \). We note that this is based on the prior specified in De Oliveira et al. (1997), which has the same form as ours for \( \beta \) and \( \sigma^2 \). Applying Bayes theorem to the term \( f(\beta, \Phi | s_o) \) on the left-hand-side of Equation (S12) gives:

\[
\frac{\int_{\beta, \Phi} f(s_{\text{pred}}, s_c | s_o, \beta, \Phi ) f(s_o | \beta, \Phi ) f_0(\beta, \Phi ) d\beta d\Phi}{\int_{\beta, \Phi} f(s_o | \beta, \Phi ) f_0(\beta, \Phi ) d\beta d\Phi} = \frac{\int_{\xi, r} f(s_{\text{pred}}, s_c | s_o, \xi, r ) f(s_o | \xi, r ) f_0(\xi, r ) d\xi d r}{\int_{\xi, r} f(s_o | \xi, r ) f_0(\xi, r ) d\xi d r}.
\]

(S18)

Now, returning to the predictive distribution, we can substitute the posterior distribution into Equation (S10) to give:

\[
p(S(x_{\text{pred}}) = s_{\text{pred}} | D )
= \int_{s_c = -\infty}^{\infty} \int_{s_o} f(s_{\text{pred}}, s_c | s_o, \beta, \Phi ) f(s_c | \beta, \Phi ) g(s_c ) d\beta d\Phi ds_c.
\]

(S19)

We can apply the fundamental law of probability to rewrite the first two terms, giving:

\[
p(S(x_{\text{pred}}) = s_{\text{pred}} | D )
= \int_{s_c = -\infty}^{\infty} \int_{s_o} f(s_{\text{pred}}, s_c | s_o, \beta, \Phi ) f(s_o | \beta, \Phi ) g(s_c ) d\beta d\Phi ds_c.
\]

(S20)

The integral with respect to \( \beta \) and \( \Phi \) here is in the same form as the integral in the numerator on the left-hand-side of Equation (S18); we can therefore use this expression to write Equation (S20) as:
\[ p(S(x_{\text{pred}}) = s_{\text{pred}} \mid D) \]
\[ = \int_{s_{\text{c}} = -\infty}^{\infty} \int_{\beta, \theta} f(s_o, \theta) \, d\beta \, d\theta \frac{\int_{\xi, r} f(s_{\text{pred}}, s_c \mid s_o, \xi, r) f(s_o \mid \xi, r) f_0(\xi, r) \, d\xi \, dr}{\int_{\xi, r} f(s_o \mid \xi, r) f_0(\xi, r) \, d\xi \, dr} g(s_c) \, ds_c . \quad \text{(S21)} \]

The integrals with respect to \( \beta \) and \( \theta \), and to \( \xi \) and \( r \) (in the denominator) of Equation (S21) are now independent of \( s_c \) and \( s_{\text{pred}} \). Therefore, they act as constants and can be dropped:

\[ p(S(x_{\text{pred}}) = s_{\text{pred}} \mid D) \]
\[ \propto \int_{s_{\text{c}} = -\infty}^{\infty} \int_{\xi, r} f(s_{\text{pred}}, s_c \mid s_o, \xi, r) f(s_o \mid \xi, r) f_0(\xi, r) \, d\xi \, dr \, g(s_c) \, ds_c . \quad \text{(S22)} \]

Once again applying the fundamental law of probability allows us to write:

\[ p(S(x_{\text{pred}}) = s_{\text{pred}} \mid D) \]
\[ \propto \int_{s_{\text{c}} = -\infty}^{\infty} \int_{\xi, r} f(s_{\text{pred}} \mid s_o, s_c, \xi, r) f(s_o, s_c \mid \xi, r) f_0(\xi, r) \, g(s_c) \, ds_c \, dr . \quad \text{(S23)} \]

This is now is exactly the form of the predictive distribution given in Equation (S10), with Equation (S12) for the posterior distribution (up to a normalization constant) inserted.

\[ \text{References} \]


