# How does temporal variation affect the value of stream water as a medium for regional geochemical survey?

R.M. Lark*, J.M. Bearcock, E.L. Ander

*British Geological Survey, Keyworth, Nottinghamshire NG12 5GG, U.K.*

## Abstract

Stream water is a key medium for regional geochemical survey for mineral exploration and environmental protection. However, stream waters are transient, and measurements are susceptible to various sources of temporal variation. In a regional geochemical survey stream water data comprise 'snapshots' of the state of the medium at a sample time. For this reason the British Geological Survey (BGS) has included monitoring streams in its regional geochemical baseline surveys (G-BASE) at which daily stream water samples are collected, over variable time intervals, to supplement the spatial data collected in once-off sampling events.

In this study we present results from spatio-temporal analysis of spatial stream water surveys and the associated monitoring stream data. We show that the variability of monitoring stream data from the G-BASE surveys has a temporally correlated component which can be treated as independent between streams, and therefore as a component of the nugget (spatially uncorrelated variance) of the spatial variograms of stream water survey data. For the variables examined this component was small relative to the spatial variability, which indicates that the value of stream water data to provide spatial geochemical information is not compromised by temporal variability. However, these conclusions are conditioned on the particular data set which was collected only in the summer months, specifically to limit temporal variability. Temporal variation in stream water analyses may be less tractable in wetter conditions. We show how the spatial data from stream water surveys can be mapped by ordinary kriging, with the predictions interpreted as an estimate of the temporal (summer months) mean, and the kriging variance reflecting the partition of the nugget variance of the spatial variogram between spatial and temporal components.

*Keywords:* Geochemical survey; stream water; spatial variation; temporal variation

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

## 1. Introduction

Geochemical mapping entails the sampling of surface materials, notably soils, stream sediments and stream waters. It is generally recognized that regional scale survey of all these media can provide information on both geogenic and anthropogenic sources of geochemical variability (De Vivo et al., 2008), and this information can be useful for the investigation of mineral resources and for managing potentially harmful elements whether these arise from naturally occurring mineralizations or pollution (Cocker, 1999; Simpson et al., 1993). For this reason geochemical surveys at regional and national scale have included sampling of all three media (De Vivo et al., 2008; Birke et al., 2015). The Geochemical Baseline Survey of the Environment (G-BASE), conducted in the United Kingdom by the British Geological Survey (BGS), included sampling of stream waters for a limited set of determinands at its inception, and since 1988 routine sampling of both stream sediments and stream waters for multi-element analysis (Johnson et al., 2005).

Surveys of stream waters provide general geochemical information, and are also informative about issues of water quality of direct relevance for policy, management and regulation. For example, G-BASE stream water data have been used to estimate exposure of non-human species to naturally occurring radionuclides (Jones et al., 2009), to understand the significance of geogenic sources of arsenic (Breward, 2007) and to estimate carbon dioxide fluxes from surface waters (Rawlins et al., 2014). Geochemical surveys of stream waters have been used to investigate pollution associated with industrial activity (Vaisenen, et al., 1998) and to investigate the combined effects of geology and anthropogenic factors on water quality (Reimann et al., 2009).

While data on stream water are useful, it is, at least potentially, more transient than soil or sediment. In a regional survey a stream is visited once, and the sample that is collected represents a snapshot of its geochemical composition at a particular time. The water chemistry of a particular stream is subject to variation over time over a range of temporal scales. Kirchner and Neal (2013) report studies on detailed analysis of the streamwater chemistry from two headwater catchments at Plynlimon in Wales. These showed fractal scaling of solute concentrations consistent with a model of randomly varying inputs across the catchment followed by dispersion driven by water transport across the landscape (Kirchner et al., 2001). The concentration of

an analyte in stream water may vary in response to flow rate. One reason for this is a dilution effect. An increase in flow rate may also be associated with an increase in the influence of the distinctive chemistry of rainwater on the composition of the stream (Appelo and Postma, 2007; Drever, 1997), contributions from overland flow or increased leaching of solutes into shallow groundwater. Over longer periods stream water composition may respond to seasonal differences in rainfall and to anthropogenic inputs, such as artificial fertilizers, which may include various trace elements along with the principal nutrients, and slurries and manures which may contribute both organic components, macronutrients such as P and trace elements such as Cu.

These sources of temporal variation must be accounted for when stream water geochemical data are interpreted to understand regional spatial variation. Hutchins et al. (1999) compared the spatial variability of stream water data from G-BASE sampling in Wales with temporal observations made in a single catchment within the country at 2- to 4-week intervals. They did not attempt any spatio-temporal statistical modelling of these data, but noted that geological, meteorological and anthropogenic effects could be seen in the spatial variation. They concluded that more observations on temporal variability of stream water data were needed in combination with the spatial sampling for robust inference. In 1997 BGS modified the field sampling procedures of the G-BASE survey to include repeated sampling from a small number of monitoring sites, sampled daily while the regional survey was conducted nearby. As a result the monitoring-site data consist of relatively short local time series, from a few days up to 30 or 40. This provides information on the short-scale temporal variability of the variables measured on stream water in the G-BASE survey.

While there have been detailed studies on the temporal variation of streamwater chemistry within one or two associated catchments (e.g. Neal et al., 2013; Kirchner and Neal, 2013) we require a more extensive study of spatio-temporal variability in order to understand how temporal variation affects the interpretation of data from spatial surveys with one-off sampling of individual streams. In this paper we analyse the data on some key variables from monitoring stream sites in the G-BASE survey of part of the English Midlands and the East Anglia region. We use a linear mixed model to examine the within-stream variation over time, including the extent to which this variation is temporally correlated over short intervals. We then analyse the survey sample data (restricted to first-order streams) using statistical models for spatio-

temporal variability to examine how the temporal variation, examined at the monitoring sites, and the spatial variation interact. On the basis of this we can quantify the implications of temporal variation of stream water properties for the spatial interpretation of data from the regional survey which comprises only spot samples from any given stream.

## 2. Materials and Methods

### *2.1 Sampling and data*

The data used in this study were collected in part of the East Midlands and the East Anglia region of England from 1996 to 2007, with no sampling in 2001 due to an outbreak of foot and mouth disease. In each year sampling was undertaken during the period from June through to September, that is to say in summer months. This was a deliberate decision to avoid wetter periods of the year and so to sample, as far as possible, when base flows dominate the stream flow. We used the stream water survey data and the data from monitoring sites collected in this period. Figure 1 shows the spatial distribution of both data sets. The sampled region is a lowland area, predominantly under agriculture but with some urban centres. Figure S1 in the supplementary material (journal website) shows the solid geology according to a generalized lithological classification. The aquifers are almost exclusively sedimentary, with Triassic and Jurassic mudstones, Cretaceous chalk, Palaeogene clays and poorly consolidated Pleistocene sediments dominating the area. These give rise to a generally subdued topography and so streams are relatively slow-flowing. Figure S9 shows the stream water survey sample sites collected in each year.

The data were collected according to the standard G-BASE procedures (Johnson et al., 2005). Drainage sample sites, at which both sediment and water specimens were collected, were identified in advance on small streams (first or second order). The target sample density was one sample per 1.5–2.0 $km^2$, but sample density varied in accordance with drainage density. Figure 1 shows, for example, that samples were absent or very sparse in a band running approximately south-west–north-east where the bedrock is Cretaceous chalk. Filtered samples for major- and trace-element analysis were collected from mid-stream using a syringe, and were passed through a 0.45-μm filter into sample bottles. Unfiltered samples were also collected for the analyses

including pH and conductivity. Johnson et al. (2015) give details of the protocols that were followed.

The locations of forty monitor sites are shown in Figure 1. Sampling was undertaken by one or two teams at any one time, and a monitor site was established near the base that the team was using at any given time. The monitor site was sampled daily according to the same protocols used for the field survey. Out of the forty sites, four were sampled in two successive years of the survey, although none for more than 30 to 40 days at a time, and most for a shorter period of a week to 10 days.

Sample analysis was conducted on location for pH and conductivity, with calibration and drift checks carried out for each run (Johnson et al., 2005). The trace elements were measured by inductively coupled plasma mass spectrometry (ICP-MS) in the BGS laboratories using the filtered acidified (1% v/v $HNO_3$) aliquot, whilst DOC was measured as non-purgeable organic carbon from the filtered, unacidified aliquot, also in the BGS laboratories. Field duplicates and blanks were used as control samples to ensure that errors were not introduced by sampling and sample handling. Laboratory analyses of field and 'blind' control samples were conducted using certified standards and reference materials within an ISO 17025 certified framework (Johnson et al., 2005).

In this study we restricted the survey stream sample sites to those on first-order streams. This ensures that, among the data we used, no two sample sites were connected by a flow path. Any spatial dependence in the data can therefore be attributed to variation between the catchments of the first-order streams.

In this paper we report on the analysis of seven variables. We considered four elements (As, Cu, Co and Ni) which are of potential interest for water quality considerations, and which may reflect both anthropogenic and geogenic inputs. We also examined data on water conductivity and dissolved organic carbon, which may reflect variation over time in inputs from different sources, including overland flow. Finally we included pH which is an important property of stream water for understanding its overall geochemistry.

We have deliberately not considered major constituents of water in this study. Data on water constituents are compositional in that the quantity of a component of interest is expressed relative to others (Buccianti and Pawlowsky-Glahn, 2005). Statistical analysis of such data

requires that they are transformed to an appropriate scale, e.g. by the log-ratio transform (Aitchison, 1986). In the analysis of major constituents of water subsets are typically selected and rescaled to a subcomposition by dividing the concentration of each constituent by the sum of concentrations over the subset. Three-component subcompositions can be represented on a ternary diagram; for example, Baca and Threlkeld (2000) considered $Mg^{+}$, $Ca^{2+}$ and ($Na^{+}$ + $K^{+}$) as three constituents of a major cation subcomposition, and $Cl^{-}$, $SO_4^{2-}$ and ($HCO_3^{-}$+$CO_3^{2-}$) as a major anion composition to distinguish waters from continental and coastal precipitation. The statistical analysis of such compositions requires special treatment, for example by the additive log-ratio transformation when we are interested in spatial modelling for prediction. For minor constituents the log-transformation of concentrations generally suffices prior to analysis because it differs negligibly, up to a constant, from the additive log-ratio transformation (see Pawlowsky-Glahn and Olea, 2004). A further advantage of using only the log-transformation is that it is straightforward to compare the transformed variable to regulatory thresholds on concentrations of potentially harmful elements. This defines the scope of the study reported here. We outline in the discussion section the necessary development of our methodology to extend it to the analysis of major ions.

### *2.2 Statistical analysis*

#### *2.2.1 Exploratory analysis*

Summary statistics were calculated for both the monitoring site data and the first-order stream data, considering both the variables on their initial scales and (apart from pH) after a transformation to natural logarithms. The log-transformation was used because the variables of interest (apart from conductivity), while not major consituents of water are nonetheless compositional and the transformation is therefore appropriate as discussed at the end of the previous section (Pawlowsky-Glahn and Olea, 2004). In the case of conductivity a log-transformation was considered as an option to make the assumption of normality a plausible one. Plots of the data time series at monitoring sites were prepared.

These data were collected over a period of eleven years, with sampling concentrated in local blocks each year (see Figure S8 in the supplementary materials). For this reason long-range spatial variation in the variables of interest is confounded with any long-term temporal trends.

To examine any potential effect of long-term temporal trends (as opposed to the within-season temporal variation that we can observe at monitoring sites), we computed spatial variograms of the data. A variogram is a function that expresses the spatial dependence of variability, it is half the expected squared difference between two observations separated by some interval (lag) expressed as a function of the lag (Webster and Oliver, 2007). More information on the variogram is provided in section 2.2.3 below. For exploratory purposes we estimated two spatial variograms. Each variogram was computed from all available data for lags restricted to no more than about 40 km, about one third the largest dimension of the study region, to avoid edge effects (Webster and Oliver, 2007). The first used all comparisons among observations, the second used only comparisons between observations from sites sampled in the same season. If there were large temporal trends over the 11-year period then one might expect the variogram estimated from all pair comparisons to be systematically larger than the variogram estimated only from within-season comparisons.

*2.2.2 Temporal analysis of monitoring site data*

The daily monitoring site data on the variables of interest were analysed as a combined set of time series. The analysis was done by fitting a linear mixed model (Verbeke and Molenberghs, 2000). In the model used for analysis there was an overall mean value across the domain. This is a space-time mean for the summer months (late June to early September) over the period of sampling (1997–2007). The model allowed the mean value of data to vary randomly between sites, and treated the variation within sites as a temporally correlated random variable. It was assumed that the within-site variability was homogeneous across the study region (many more data would be needed to fit a more complex model without this assumption). Note that the between-site variation will combine both spatial variation and any temporal variation over the period of sampling, with the components confounded to some extent as sampling earlier in the period was in the west of the region. This makes the between-site component of the model difficult to interpret in isolation. However, the focus of our interest in this part of the analysis was on the within-site temporal variation.

The set of $N$ observations of some variable in the monitoring site data was modelled as an

$N \times 1$ random variate, $\mathbf{Y}$ defined by a linear mixed model where

$$\mathbf{Y} \;=\; \mathbf{X}\boldsymbol{\beta} + \mathbf{U}_{\mathrm{s}}\boldsymbol{\upsilon}_{\mathrm{s}} + \boldsymbol{\eta} + \boldsymbol{\varepsilon}. \tag{1}$$

The term $\mathbf{X}\boldsymbol{\beta}$ defines a $N \times 1$ vector of mean values for the variable $y$. In the simplest case $\mathbf{X}$ is just a vector of ones, and $\boldsymbol{\beta}$ is a vector length 1 which contains the overall mean, $\mu$. This is the model we used in this analysis, but a more general model could be used, for example to specify the mean as a function of some covariate. The remaining terms of the model are random effects. The matrix $\mathbf{U}$ is an $N \times P$ design matrix which associates each of the $N$ observations with one of the $P$ monitoring sites, all terms in the $m$th row of $\mathbf{U}$ are zero except for the $p$th column which takes the value 1 indicating that the $n$th observation is at the $p$th monitoring site. The term $\boldsymbol{\upsilon}_{\mathrm{s}}$ is a normal $p \times 1$ random variate with distribution

$$\boldsymbol{\upsilon}_{\mathrm{s}} \;\sim\; \mathcal{N}\left\{\mathbf{0}_P, \sigma^2_{\mathrm{S}}\mathbf{1}_P\right\},$$

where $\mathbf{0}_N$ is a $P \times 1$ vector of zeroes, $\sigma^2_{\mathrm{S}}$ denotes a variance, the between-site variance, and $\mathbf{1}_P$ is an $P \times 1$ vector of ones. Note that we treated the difference between the mean value for the $p$th monitoring site and the mean over the domain as an independent random variable with variance $\sigma^2_{\mathrm{S}}$. This variable captures spatial variation between sites, and any long-term temporal variation over the period of sampling.

The term $\boldsymbol{\eta}$ is an $N \times 1$ random variate which expresses the temporally correlated variation of the variable of interest within any monitoring site. That is to say, it expresses about the site mean which is due to factors which operate at different temporal scales and so gives rise to components of temporal variation which are, on average, more similar if compared over a short time interval than over a long time interval. The variance of this component is $\sigma^2_{\mathrm{T,c}}$, and the values of $\boldsymbol{\eta}$ for any two of the $N$ observations have a correlation which is expressed in the correlation matrix $\mathbf{R}_{\mathrm{T}}$. The term $\boldsymbol{\eta}$ is a normal random variate with distribution

$$\boldsymbol{\eta} \;\sim\; \mathcal{N}\left\{\mathbf{0}_N, \sigma^2_{\mathrm{T,c}}\mathbf{R}_{\mathrm{T}}\right\}.$$

Consider two observations sampled on days $t_i$ and $t_j$. The correlation of the corresponding values of $\boldsymbol{\eta}$ is zero if the observations are at different monitoring sites but otherwise is given by an autocorrelation function which, under an assumption of second-order stationarity is assumed

to depend only on the difference between the two times, the time-lag $\tau = |t_i - t_j|$. In this study we used a Matérn autocorrelation function (Matérn, 1960):

$$\rho_M(\tau|\kappa,\phi) \;=\; \frac{(\tau/\phi)^{\kappa} K_{\kappa}(\tau/\phi)}{2^{\kappa-1}\Gamma(\kappa)} \tag{2}$$

where $K_{\kappa}(\cdot)$ denotes the modified Bessel function of the second kind of order $\kappa$, $\Gamma(\cdot)$ is the gamma function, $\phi$ is a time parameter, and $\kappa$ is a parameter which determines the smoothness of the spatial process. This autocorrelation function decays with increasing temporal lag. The smoothness of the temporal variation is described by the parameter $\kappa$, with the variation appearing smoother the larger this parameter. The autocorrelation decays asymptotically to zero, and is typically characterized by the 'effective temporal range', the temporal lag at which the autocorrelation decays to 0.05. The effective temporal range depends on the values of the parameters $\kappa$ and $\phi$.

The final term in Eq. [1] is $\boldsymbol{\varepsilon}$ which is an independently and identically distributed, and hence temporally uncorrelated, error term with variance $\sigma^2_{\mathrm{T,n}}$. This is modelled as a normal variate:

$$\boldsymbol{\varepsilon} \;\sim\; \mathcal{N}\left\{\mathbf{0}_N, \sigma^2_{\mathrm{T,n}}\mathbf{1}_N\right\}.$$

This component of the model corresponds to all sources of temporal variation which are not resolved by the daily sample interval. It will therefore include short-range components of variation (e.g. any source of variation correlated over less than 24 hours or so) and any analytical error.

The temporal statistical model for the monitoring site data has been stated. It has two sets of parameters. The first set are variance parameters, which characterize the three random components. These are the variance of the between-site random effect, $\sigma^2_{\mathrm{S}}$, the variance of the temporally correlated within-site random effect, $\sigma^2_{\mathrm{T,c}}$, the parameters of the autocorrelation function, $\kappa$ and $\phi$, and the variance of the uncorrelated random effect, $\sigma^2_{\mathrm{T,n}}$. In addition to the variance parameter there are fixed effects, the values in the vector $\boldsymbol{\beta}$ which in this study is a single constant, the mean $\mu$. We used residual maximum likelihood (REML) to estimate the variance parameters, and then estimated $\boldsymbol{\beta}$ by weighted least squares. This reduces bias in the estimates of the variance parameters because these do not depend on some estimate of the unknown mean. Under the model in Eq. [1] the random variate $\mathbf{Y}$ has an $N \times N$ covariance

matrix $\mathbf{C}$. We treat the three separate random effects as mutually independent, and so $\mathbf{C}$ is the sum of three separate covariance matrices:

$$\mathbf{C} \;=\; \sigma^2_{\mathrm{S}}\mathbf{U}\mathbf{U}^{\mathrm{T}} + \sigma^2_{\mathrm{T,c}}\mathbf{R}_{\mathrm{T}} + \sigma^2_{\mathrm{T,n}}\mathbf{1}_N. \tag{3}$$

The negative log residual likelihood can be computed given a set of $N$ observations of our variable of interest in the vector $\mathbf{y}$ as

$$\ell\left(\boldsymbol{\psi}|\mathbf{y}\right) \;=\; \frac{1}{2}\left(\ln|\mathbf{C}| + \ln\left|\mathbf{X}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{X}\right| + \mathbf{y}^{\mathrm{T}}\mathbf{P}\mathbf{y}\right), \tag{4}$$

where $\mathbf{P}$ is

$$\mathbf{P} \;=\; \mathbf{C}^{-1} - \mathbf{C}^{-1}\mathbf{X}\left(\mathbf{X}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{C}^{-1}, \tag{5}$$

and $\boldsymbol{\psi}$ denotes a set of variance parameters, $\boldsymbol{\psi} = \left\{\sigma^2_{\mathrm{S}}, \sigma^2_{\mathrm{T,c}}, \kappa, \phi, \sigma^2_{\mathrm{T,n}}\right\}^{\mathrm{T}}$.

In REML estimation we find a set of variance parameters, $\widehat{\boldsymbol{\psi}}$, which minimizes $\ell$ as defined in Eq. [4], given a set of data, $\mathbf{z}$. This must generally be done by a numerical optimization. We used the optim procedure in the R package (R core team, 2014) to find REML estimates of the random effects parameters, the values that minimize the negative log residual likelihood as defined in Eq. [4]. The L-BFGS-B (Low-memory Broyden-Fletcher-Goldfarb-Shanno, bounded algorithm) optimization method was selected (Byrd et al., 1995) as this allows us to put bounds on parameters and so to avoid numerical problems if the optimizer considers a negative value for a variance. We followed Diggle and Ribeiro (2007) in not attempting to find the $\kappa$ smoothness parameter by allowing it to vary freely with the others, but rather specifying a small set of discrete values of the parameter and finding the REML estimates of the other parameters conditional on each of these values and selecting the one for which $\ell$ is smallest. We considered values of $\kappa = 0.1$, 0.5, 1.0, 1.5 and 2.0, although we did not compute the REML estimate with $\kappa = 2.0$ if there was a clear minimum of the negative residual log-likelihood for some smaller value.

The definition of the negative log residual likelihood in Eq. [4] is explicitly based on an assumption of normality. For all monitoring site data the transformation of values to logarithms (apart from pH) gave a value of the skewness coefficient within the interval $[-1, 1]$, except for cobalt which was slightly larger than 1.0. A skewness coefficient within this interval is a general

rule of thumb for success in data transformation (Webster and Oliver, 2007). On this basis, and examination of histograms, we regarded an assumption of normality as plausible for the data for pH and for the other variables with log transformation. The Box-Cox transformation, of which the log is a special case, might be considered to achieve still smaller skewness, particularly for cobalt, but this was not done because of the difficulty of comparing variance components for variables under different non-linear transformations; the log transform is widely used in geochemistry and, as noted above, is particularly suitable for data on small components of compositional sets of variables. Furthermore, likelihood-based inference is relatively robust to small departures from normality, as shown by Kitanidis (1985), and Pardo-Igúzquiza (1998) offers an argument based on maximum entropy for the use of maximum likelihood as a criterion for inference even in the non-normal case.

The fitted model assumes that there is temporal dependence in the random effects (i.e. that we require the term $\boldsymbol{\eta}$). Under an alternative model this term could be dropped, and the within-site temporal variance treated as entirely uncorrelated. It is not possible to test directly the significance of the $\boldsymbol{\eta}$ term in the model specified in Eq. [1] by comparing the model with one in which the term is dropped, but these two models can be compared with respect to their respective Akaike information criterion, AIC (Akaike, 1973):

$$\text{AIC} \;=\; 2\ell + 2\mathcal{P}, \tag{6}$$

where $\mathcal{P}$ is the number of parameters in the model and $\ell$ is the negative residual log-likelihood. The second term in Eq. [6] can be regarded a penalty for model complexity. Although the AIC is not a formal significance test, if one selects the model with the smallest AIC one minimizes the expected information loss through the selection decision (Verbeke and Molenberghs, 2000). AIC was computed for the linear mixed model with and without the correlated term $\boldsymbol{\eta}$ for each variable.

*2.2.3 Spatio-temporal analysis of first order stream data*

The first-order stream data were analysed in conjunction with the monitoring site data by the computation of space-time variograms. In the previous section we introduced the temporal lag, or interval between two observations in time, denoted by $\tau$. Similarly we can define a spatial lag by a vector $\mathbf{h}$, an interval between two observations in space. For a random variable

$Z$ sampled in space and time the space time variogram is defined as

$$\gamma(\mathbf{h},\tau) \;=\; \frac{1}{2}\mathrm{E}\left[\{Z(\mathbf{x},t)-Z(\mathbf{x}+\mathbf{h},t+\tau)\}^2\right], \tag{7}$$

where $\mathrm{E}[\cdot]$ denotes the statistical expectation of the term in square brackets. For practical purposes, with irregular sampling in space, one may define a lag bin $(\mathbf{h}_i,\tau)$ by the time lag $\tau$ and some range of tolerance about the central direction and scalar distance of the lag $\mathbf{h}$, see Webster and Oliver, 2007). A method of moments estimator of the space time variogram is then

$$\widehat{\gamma}(\mathbf{h}_i,\tau) \;=\; \frac{1}{2N_{\mathbf{h}_i,\tau}}\sum_{(\mathbf{h},\tau)\in(\mathbf{h}_i,\tau)}\left[\{z(\mathbf{x},t)-z(\mathbf{x}+\mathbf{h},t+\tau)\}^2\right], \tag{8}$$

where the summation denotes summing over all comparisons between pairs of data points separated by time-lag $\tau$ and a spatial lag in the lag-bin $\mathbf{h}_i$ and where there are $N_{\mathbf{h}_i,\tau}$ such pair comparisons.

Exploratory analysis of spatial variograms ($\tau = 0$) for lag bins centered on four directions (North–South, North-East–South-West, East–West and North-West–South-East) for all variables gave no evidence of directional dependence (see Figure S10 in the supplementary material). For this reason isotropic space-time variograms (defined on lag distances with all directions pooled) were used in further analysis. We considered lag bins centered on distances $5, 10, \ldots 40$ km with tolerance $\pm$ 2.5km. We limited the maximum lag to 40km, approximately one quarter the maximum extent of the region, to avoid edge effects (Webster and Oliver, 2007).

We estimated two sets of space-time variograms. The first were variograms for a full set of spatial lags and a subset of four temporal lags at intervals from zero days up to a maximum lag comparable to the effective range of temporal dependence identified for the variable in question from the analysis of the monitoring site data described in section 2.2.3. The second set were variograms for a full set of temporal lags and a subset of spatial lags: 5, 10, 20 and 40km. These were plotted to allow an assessment of the spatio-temporal variability of the variables.

A standard stochastic model for space-time data, on which the interpretation and modelling of space-time variograms is based, is presented by Cressie and Wikle (2011). For a random variable, Z, observed at location $\mathbf{x}$ at time $t$

$$Z(\mathbf{x},t) \;=\; \mu(\mathbf{x},t)+\zeta(\mathbf{x})+\xi(t)+\vartheta(\mathbf{x},t)+\delta(\mathbf{x},t). \tag{9}$$

The term $\mu(\mathbf{x}, t)$ is the space-time mean, which may be a constant or may vary with time or space or both. The remaining terms are random components of mean zero. The first, $\zeta(\mathbf{x})$, is a spatial random effect, common to all times, whereas the second random term, $\xi(t)$, is a temporal random effect common to all locations. The term $\vartheta(\mathbf{x}, t)$ is a random effect representing a space-time interaction, and $\delta(\mathbf{x}, t)$ is an independent and identically distributed random effect (i.e. uncorrelated in space and time). In the case with a constant mean, the space-time variogram $\gamma(\mathbf{h}, 0)$, i.e. for spatial lags with temporal lag fixed at zero, shows the spatial dependence of the term $\zeta(\mathbf{x})$, with a nugget effect (i.e. apparent intercept at $|\mathbf{h}| = 0$) equal to the variance of $\delta(\mathbf{x}, t)$. Similarly the variogram $\gamma(0, \tau)$ shows the temporal dependence of $\xi(t)$ with a nugget due to $\delta(\mathbf{x}, t)$. For the case with both spatial and temporal lag not equal to zero the variogram $\gamma(\mathbf{h}, \tau)$ may be modelled as a combination of spatial and temporal effects either with no interaction (a so-called separable model) or with some kind of interaction term.

In this paper we also consider an alternative model:

$$Z(\mathbf{x}, t) \;=\; \mu(\mathbf{x}, t) + \zeta(\mathbf{x}) + \eta_{\mathbf{x}}(t) + \delta(\mathbf{x}, t). \tag{10}$$

In this model the temporal random variable of mean zero, $\eta$, is specific to the location of interest, $\mathbf{x}$, rather than being a temporal random effect common to all locations, which is the key property of $\xi(t)$ in Eq. [9]. It is therefore assumed that any two values of the temporal term, $\eta_{\mathbf{x}_i}(t_k)$ and $\eta_{\mathbf{x}_j}(t_l)$ are independent if $i \neq j$. A stationarity assumption is made by which the variance of $\eta_{\mathbf{x}}(t)$ is a constant, $\sigma^2_\eta$, irrespective of the location, $\mathbf{x}$, and that the autocorrelation of the temporal random effect is a function of the temporal lag only, also independent of location:

$$\mathrm{Corr}\left[\eta_{\mathbf{x}_i}(t), \eta_{\mathbf{x}_i}(t+\tau)\right] \;=\; \rho(\tau), \qquad \forall \mathbf{x}_i, \tag{11}$$

where Corr[$\cdot$] denotes the correlation of the two variables in square brackets.

Under this model the temporal variation at any sampled stream is dominated by local processes at the scale of the stream's own catchment, and so is not common to any two sample sites in different streams (assuming that all streams are of first order and so are not connected by any flow path). This could be a plausible model for stream water composition. The problem is to evaluate the evidence for the two models. One way to do this would be to fit appropriate specific forms of each model by REML and compare them with respect to the maximized likelihood

(although this would not be a standard case for likelihood ratio tests). However, the data sets on first order stream chemistry comprise of the order of 4000 observations, and likelihood estimation of models with correlated random effects in this setting is computationally demanding. Furthermore, there are various alternative forms of the general model in Eq [9], each making different assumptions about the joint behaviour of the spatial and temporal variation (separable models and many possible forms on non-separable model). For this reason we based our comparison on evaluation of method-of-moment estimates of the space-time variogram, obtained with Eq. [8].

Under the model in Eq. [10] the temporal variogram for any non-zero constant spatial lag, $\tilde{\mathbf{h}}$, $\gamma\left(\tilde{\mathbf{h}}, \tau\right)$ would be a constant:

$$\gamma\left(\tilde{\mathbf{h}}, \tau\right) \;=\; \frac{1}{2}\mathrm{E}\left[\left\{\zeta(\mathbf{x}) - \zeta(\mathbf{x}+\tilde{\mathbf{h}})\right\}^2\right] + \sigma_\eta^2 + \sigma_\delta^2, \tag{12}$$

In the event that the temporal random effect in the model for the monitor site data shows evidence of temporal correlation, one would expect the temporal variogram for any non-zero constant spatial lag to show evidence of temporal dependence under the full space-time model in Eq. [9], but not under the alternative model in Eq. [10]. To evaluate the evidence for one model overagainst the other we examined the spatial variograms estimated for a set of time lags, and fitted models by weighted least squares (Webster and Oliver, 2007) to the temporal variograms for fixed spatial lags (5, 10, 20 and 40 km). The temporal variograms were either a pure nugget (i.e. a constant), consistent with the model in Eq. [10] given Eq. [12] or a Matérn with the parameter $\kappa$ set to the value selected for the temporal model for the variable of interest in the analysis of the monitor-site data. One may compare models fitted in this way on a statistic $\widehat{A}$ which is smallest for the model with smallest AIC (Webster and McBratney, 1989; Webster and Oliver, 2007):

$$\widehat{A} \;=\; n_p \log(R) + 2\mathcal{P}, \tag{13}$$

where $n_p$ is the number of time lags at which the temporal , $R$ is the mean squared residual from the fitted model and $\mathcal{P}$ is the number of model parameters. In this case $\mathcal{P}$ was 1 for the pure nugget model and 3 for the Matérn (the $\kappa$ parameter was treated as fixed).

*2.2.4 Kriging the first order stream data*

Under the alternative spatial model, in which the within-stream temporal variation is independent between separate first-order streams, the temporal variation, which we modelled as a combination of a temporally correlated random effect and an independent and identically distributed random effect in the analysis of the monitoring site data according to Eq. [1], can be thought of as representing temporal fluctuation about a local mean. One could think of a single sample from a stream as providing an estimate of that mean (for the sampling domain of interest here, i.e. the summer months), with an error variance represented by the sum of the variances of the within-site random effects in Eq. [1]. We call this variance $C_{\mathrm{T}}$ where

$$C_{\mathrm{T}} \;=\; \sigma^2_{\mathrm{T,c}} + \sigma^2_{\mathrm{T,n}}.$$

If one were to estimate the spatial variogram of a set of first-order stream data, pooling all temporal lags, then, because of the independence of this spatial component between streams, it, in effect, contributes a quantity $C_{\mathrm{T}}$ to the nugget variance of the spatial variogram. The nugget variance, denoted by $C_0$, is the apparent intercept of the empirical variogram, representing sources of variation which are not spatially correlated at scales resolved by sampling. We would expect analytical uncertainty to contribute to the variance of the uncorrelated temporal random effect, $\sigma^2_{\mathrm{T,n}}$, and so this is part of $C_{\mathrm{T}}$. The difference between the nugget variance of the spatial variogram and $C_{\mathrm{T}}$ will arise from sources of spatial variation in the mean value of the variable of interest at a sampling site at finer spatial scales than are resolved by the spatial sampling.

We estimated spatial variograms of all variables of interest, with all time lags pooled, using the variog procedure of the geoR package for the R platform (Diggle and Ribeiro, 2007; R Core Team, 2014). We then fitted spatial Matérn models to these variograms using the variofit procedure in the same package, specifying weighted least squares as proposed by Cressie (1985). This provided estimates of the nugget variance $C_0$, the spatially correlated variance $C_1$, a distance parameter $\phi$ and a smoothness parameter $\kappa$.

One may map properties of the water of first-order streams by ordinary kriging (OK) to show the spatial variation of regional geochemistry that they reveal. In the standard case for OK (e.g. Webster and Oliver, 2007) the prediction at some unsampled site is the best linear unbiased predictor of a new measurement made at that location (with any measurement error), and the OK procedure also computes the kriging variance, the expected squared error of the OK

prediction, which is minimized. However, if one can partition the nugget variance of the target variable of interest into a component for measurement error, and a component that corresponds to spatial variation unresolved by sampling ('microscale variance'), then the OK prediction of the 'signal' can be obtained, i.e the variable without measurement error (Diggle and Ribeiro, 2007). The OK predictions are identical at sites which do not correspond exactly to the locations of observations, but the kriging variances differ. The kriging variance for the signal is smaller than for a new observation. In the context of this work, we can regard $C_{\mathrm{T}}$ as measurement error, since it contains both analytical error variance and the temporal variation around the mean value at the sample site, and the difference between the fitted nugget variance for the spatial variograms and $C_{\mathrm{T}}$ is an estimate of the microscale variance.

We used the krige.conv procedure from geoR to obtain kriging predictions of As and Co concentrations in first order streams (log scale) by ordinary kriging. We considered both kriging of new measurments and of the signal, computing the microscale variance component from the fitted nugget as described above.

## 3. Results

Time series for each monitoring site, and the year of measurement, are shown (log arsenic concentration) in Figure 2. A full set of plots for all variables is presented in the supplementary material (Figures S2–S7). Table 1 gives summary statistics for all data, and histograms of the first-order stream survey data.

Figure 5 shows the spatial variograms for all variables estimated from either all pairs of observations or only from pairs of observations within the same sample season. The differences between these are all very small, with the exception of Cu. This indicates that there might be some long-term temporal trend in Cu concentrations, although these are also confounded with spatial variations. A plot of all values against sample date does not show any dominant temporal trend for any variable (Figure S8 in the supplementary material). These results indicate that, with the possible exception of Cu, the data collected over this period can be combined to show spatial variability, with no reason to expect that this is confounded with a long-term trend at regional scale.

Table 2 presents results for the linear mixed model for the monitoring site data. Note that

in all cases $A_{\mathrm{M}}$ was notably smaller than $A_{\mathrm{IID}}$, i.e. on the basis of the AIC one would select the model with a temporally correlated temporal random effect. For three of the seven variables (As, Cu, DOC) the best-fitting correlation model for the temporal effect had a value of $\kappa$ of 0.5. In the case of pH the temporally correlated random variation was rougher than this ($\kappa = 0.1$), and in three cases it was smoother (Conductivity and Co with $\kappa = 1.0$ and Ni with $\kappa = 2.0$. The effective range of correlation are all fairly similar for the elements As, Cu, Co and Ni (21–26 days). The effective ranges are shorter for conductivity (6 days) and DOC (13 days), but longer for pH (58 days). Note that the effective range for pH is rather larger than the longest single time series in the data set, and this behaviour may indicate that there are temporal trends in pH at seasonal scale over the summer months.

In all cases the between-site variance, $\sigma^2_{\mathrm{S}}$ is notably larger than the within-site variance components. The between-site variance comprises spatial and temporal variation, and for this reason we interpret the within-site variances in the context of the spatial analysis of the first-order stream survey data.

Figures 6–12 show space-time variograms for selected lags for all variables. Note that there is no systematic increase in the variance with temporal lag for the spatial variograms (the top graph in each Figure) and the temporal variograms at spatial lags 5, 10, 20 and 40km appear mostly flat. Table 3 presents values of the statistic $\widehat{A}$ computed from Eq. [13] for the fitting of a Matérn or a pure nugget (constant) model to the temporal variograms. For four of the variables the pure nugget model is preferred (smaller $\widehat{A}$) for the temporal variogram at all spatial lags. For the other three variables the Matérn model is favoured at no more than one spatial lag. In summary, there is no systematic evidence for a common temporally correlated component in the space-time model for the first-order stream survey data, and the model given in Eq. [10] is favoured. Under this model the temporally correlated variation at one sample site is independent of the variation at sites on other first-order streams. On this basis one can compute a pooled spatial variogram of the first-order stream survey data, with all times combined, and treat the sum of the two within-site variance components presented in Table 2 as components of the spatial nugget variance.

Figure 13 shows these pooled estimated spatial variograms and the fitted models, the parameters of which are in Table 4. The nugget variance, $C_0$, is shown as a horizontal dotted

line on each variogram in Figure 13. Also shown is $C_{\mathrm{T}}$, the sum of the temporally correlated and independently and identically distributed within-site random effects of the models fitted to the monitoring site data. Recall that this is regarded as a component of the nugget variance of the spatial variograms, combining analytical error and temporal variation around the mean concentration at a site over the summer months. We therefore can regard $C_{\mathrm{T}}$ as a measurement error when the field spot samples from first-order streams are treated as estimates of the summer mean concentration. Note in Table 4 that $C_{\mathrm{T}}$ varies from 20% to a little over 40% of the nugget of the spatial variogram. Furthermore, with the exception of Cu, $C_{\mathrm{T}}$ is also smaller than the spatially correlated variance, $C_1$, although the values are close in the case of pH. This indicates that, while the temporal variation of stream water data is significant, it is small relative to the spatial variation between first-order stream measurements, including the variation at scales too fine to be resolved at the sampling density used here.

Figure 14(a) shows the kriged map of As (log scale), and Figure 14(b) shows the kriging standard error (square root of the kriging variance). Note that the latter is computed treating $C_{\mathrm{T}}$ as measurement error variance and the remainder of the nugget as microscale variance, i.e,. it is the standard error of the kriged map treated as a prediction of the local mean value over the summer months. The standard error largely represents the density of the observations, it is largest where the sampling is sparse, particularly over the Chalk bedrock. Figure 15(a) shows the upper limit of the 95% confidence interval for log As (i.e. the prediction plus 1.96 times the standard error), where the nugget is all treated as microscale variance (no measurement error). This is the confidence interval for the prediction of a new observation. Figure 15(b) shows the upper limit of the confidence interval for the prediction of the local summer mean. Note that these values are generally smaller. The contour line corresponds to log 10 $\mu$g $\mathrm{l}^{-1}$, a somewhat arbitrary value (corresponding, in fact, to the regulatory limit for drinking water) selected for illustration. The region inside the contour (where the shading is lighter) is where the upper confidence limit exceeds this threshold. Note that the region within which the limit of the prediction of the summer temporal mean exceeds the threshold is more restricted than the region where the limit of the prediction of a new measurement does so. In a regulatory context the former may be more relevant. This shows how quantification of the temporal component of variability of stream water measurements can help the interpretation of spatial data for practical

purposes.

Figures 16 and 17 correspond to 14 and 15 but represent the data on Co concentration. The contour on Figure 17 corresponds to log 3 $\mu$g $l^{-1}$, which is the Statutory regulatory threshold (annual mean) for the protection of freshwater life in the UK (Environment Agency, 2011). Note again that the subregion where the upper confidence limit exceeds this threshold is more restricted when we treat the temporal variance as a measurement error with the spot data an estimate of the temporal summer mean.

## 4. Discussion

All variables showed evidence of temporal correlation in the data at monitoring streams, indicating that these determinands show variability arising from temporal processes at scales coarser than 24 hours or so. Table 2 shows that for all variables the temporally correlated variance was larger than the component of variance attributable to a combination of measurement error and variation at shorter temporal scales than the sampling interval. The effective range of temporal correlation, also shown in Table 2, was similar for all the elemental analyses (21–26 days), and shorter for conductivity and DOC, and larger for pH.

The comparison between spatial variograms of the stream survey data, estimated from either all pairs of observations or only from pairs of observations within the same sample season (Figure 5) show that, with the possible exception of copper, there is no evidence of a long-term temporal trend in any of the variables confounded with regional-scale spatial variation. The long apparent range of temporal dependence of the data on stream water pH at the monitoring site may therefore indicate some within-season temporal trend.

We considered different models for the spatio-temporal variability of all determinands in the stream survey data. The space-time variograms of the survey stream data (see Table 3) support the alternative space-time model, Eq. [10], under which the temporal variation is correlated within sites but not between. This means that we can treat the temporal variance as, effectively, the measurement error variance of a spot observation in a data set such as the G-BASE spatial stream survey data, if our interest is in the mean value of some variable (at least over the summer months). This is based on the assumption that analytical error is a component of the temporal nugget in the model for the monitor site data.

Given this conclusion, we can observe that the temporal variance is a component of the nugget variance if we pool all the survey stream data into a single spatial set, ignoring sample time. The temporal variance ranges from 20 to just over 40% of this variance, indicating that a good deal of the nugget variability in the spatial variograms can be attributed to spatial variation at short spatial scales. For all the variables we considered, the spatially correlated variance of the stream water data is larger than the within-site temporal variance (although these two quantities are very similar for pH). In this region, and for these variables, we can therefore be confident that the temporal variability is small relative to the total spatial variation. The value of the stream water data is not undermined by temporal variation.

Note, however, that this conclusion, while encouraging with respect to the use of the G-BASE water data for examining spatial variation, at least in this region, may not hold more generally for stream water surveys. G-BASE sampling was specifically planned for the summer months, when rainfall is generally smaller which means that stream waters are dominated by baseflows and stream water chemistry is therefore dominated by that of shallow groundwater, making it less variable. One could not extrapolate our conclusion about the temporal stability of stream water analyses to other surveys in which waters were collected at wetter times of year. Similarly, the conclusions cannot necessarily be extrapolated to contrasting environments such as upland catchments. That said, the data set we have examined is substantial, representing ten years of field work and we are not aware of any comparable study in which the examination of temporal variation of water properties has been set in the context of such extensively-sampled spatial data. The primary significance of this study is that we develop and present a method of analysis that can be used to examine such effects.

We noted that it is possible to account for the temporal variability of stream waters, under the model we selected for these data, when kriging the spatial variation of stream water measurements and interpreting these as estimates of the temporal mean (over the summer months) rather than as a prediction of a point measurement. This is advantageous because the uncertainty under this interpretation is somewhat smaller, as indicated in Figure 15 and 17.

There is scope for further research on questions raised by this study. There are some specific questions raised by features of these data. We noted the possibility that there are within-season trends in pH. Examining these would require sampling on longer time series than those available

here. Similarly, it would be interesting to investigate whether there is indeed a between-season temporal trend in Cu, as suggested by the within and between-season spatial variograms shown in Figure 5.

More general questions of methodology require further work. As noted above, the analysis of major components of stream water is often of scientific interest. This is addressed by computing subcompositions of the major cations and anions (e.g. Otero et al., 2005). The analysis of data in this form would require a different approach, based on multivariate linear mixed models for the constituents after a transformation such as the additive log-ratio (Aitchison, 1986). Such analyses have been done in a purely spatial context (e.g. Pawlowsky-Glahn and Olea, 2004; Lark and Bishop, 2007; Lark et al., 2015). Lark et al (2015) proposed a linear mixed model for additive log-ratio transformed data on the composition of seabed sediments by particles in different size classes. This was a multivariate model which included a spatially correlated linear model of coregionalization for the random component. To apply this model to data on major water constituents from this study would require its extension to an appropriate space-time model. This would be a substantial development, but of considerable interest. Finally, we have assumed stationarity of the temporal variability of stream water properties, but this might not hold, and the temporal variability might vary in space. A larger data set would be needed to examine this, perhaps using models with a tempered temporal spectrum, similar to the approach taken to non-stationary spatial statistical modelling by Haskard and Lark (2009) and Lark(2016).

## 5. Conclusions

To conclude, our analysis has shown that key stream water variables show temporally correlated spatial variation within stream which, at least for the summer months of the G-BASE surveys, appears to be independent between sites. This means that we can treat the temporal variation as a component of the nugget variance of the spatial variograms of stream water survey data, and account for this partition of the nugget when computing the variance of the ordinary kriging estimate, treated as an estimate of the temporal mean (summer months). The within-site temporal variation appears to be small relative to the spatial variability of the variables, which means that the temporal variation should not mask the spatial geochemical variation that

we hope to examine through stream water surveys. However, the G-BASE surveys are restricted to the summer months. Greater temporal variability, interacting in a more complex way with spatial effects, might be expected in data sets collected in wetter times of the year.

## 6. Acknowledgements

We acknowledge the contribution of colleagues in the G-BASE field teams and in BGS's analytical laboratories who generated the data that we analyse here. This work is published with the permission of the Executive Director of the British Geological Survey (NERC).

## References

Aitchison, J. 1986. The Statistical Analysis of Compositional Data. Chapman and Hall, London.

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Second International Symposium on Information Theory (ed B.N. Petov and F. Csaki), pp. 267–281. Akademia Kiado, Budapest.

Appelo. C.A.J., Postma, D., 2007. Geochemistry, groundwater and pollution. 2nd Ed. A.A. Balkema, Leiden.

Baca, R.M., Threlkeld, S.T. 2000. Inland dissolved salt chemistry: statistical evaluation of bivariate and ternary diagram models for surface and subsurface waters Journal of Limnology, 59, 156–166.

Birke, M., Rauch, U., Stummeyer, J., 2015. How robust are geochemical patterns? A comparison of low and high sample density geochemical mapping in Germany. Journal of Geochemical Exploration 154, 105–128.

Breward, N., 2007. Arsenic and presumed resistate trace element geochemistry of the Lincolnshire (UK) sedimentary ironstones and revealed by a regional geochemical survey using soil, water and stream sediment sampling. Applied Geochemistry, 22, 1970–1993.

Buccianti, A., Pawlowsky-Glahn, V. 2005. New perspectives of water chemistry and compositional data analysis. Mathematical Geology, 37, 703–727.

607 Byrd, R.H., Lu, P., Nocedal,J., Zhu,C., 1995. A limited memory algorithm for bound con-
608 strained optimization. SIAM Journal of Scientific Computing, 16, 1190–1208.

609 Cocker, M.D., 1999. Geochemical mapping in Georgia, USA: a tool for environmental studies,
610 geologic mapping and mineral exploration. Journal of Geochemical Exploration 67, 345–
611 360.

612 Cressie, N., 1985. Fitting variogram models by weighted least squares. Mathematical Geology,
613 17, 563–586.

614 Cressie, N., Wikle, C.K., 2011. Statistics for spatio-temporal data. John Wiley & Sons,
615 Hoboken, New Jersey.

616 De Vivo, B., Lima, A., Bove, M.A., Albanese, S., Cicchella, D., Sabatini, G., Di Lella, L.A.,
617 Protano, G., Riccobono, F., Frizzo, P., Raccagni, L., 2008. Environmental geochemical
618 maps of Italy from the FOREGS database. Geochemistry: Exploration, Environment,
619 Analysis 8 (3–4), 267–277.

620 Diggle, P.J., Ribeiro, P.J., 2007. Model-based geostatistics. Springer, New York.

621 Drever, J.I., 1997. The geochemistry of natural waters: surface and groundwater environments.
622 3rd Ed. Prentice Hall, New Jersey.

623 Environment Agency, 2011. Chemical Standards Report CAS RN: 7440-48-4, Cobalt.
624 http://evidence.environment-agency.gov.uk/ChemicalStandards/report.aspx?cid=46

625 Haskard, K.A., Lark, R.M., 2009. Modelling non-stationary variance of soil properties by
626 tempering an empirical spectrum. Geoderma, 153. 18–28.

627 Hutchins, M.G., Smith, B., Rawlins, B.G., Lister, T.R., 1999. Temporal and spatial variability
628 of stream waters in Wales. the Welsh borders and part of the West Midlands, UK — 1.
629 Major ion concentrations. Water Research, 33, 3479–3491.

630 Johnson, C.C., Breward, N., Ander, E.L., Ault, L., 2005. G-BASE: baseline geochemical
631 mapping of Great Britain and Northern Ireland. Geochemistry : exploration, environment,
632 analysis, 5, 347–357

633 Johnson, C.C., 2005. 2005 G-BASE Field Procedures Manual. British Geological Survey
634 Interal Report, IR/05/097. British Geological Survey, Keyworth.
635 http://nora.nerc.ac.uk/5190/1/2005proceduresmanual.pdf

636 Jones, D.G., Appleton, J.D., Breward, N., Mackenzie, A.C., Scheib, C., Beresford, N.A., Bar-
637 nett, C.L., Wood, M.D., Copplestone, D., 2009. Assessment of naturally occurring ra-
638 dionuclides around England and Wales: application of the G-BASE dataset to estimate
639 doses to non-human species. Radioprotection 44 (5), 629–634.

640 Kirchner, J.W., Neal, C. 2013. Universal fractal scaling in stream chemistry and its implications
641 for solute transport and water quality trend detection. PNAS, 110, 12213–12218.

642 Kirchner, J.W., Feng, X., Neal, C. 2001. Catchment-scale advection and dispersion as a mecha-
643 nism for fractal scaling in stream tracer concentrations. Journal of Hydrology, 254, 82–101.

644 Kitanidis, P.K., 1985. Minimum-variance unbiased quadratic estimation of covariances of re-
645 gionalized variables. Journal of the International Association of Mathematical Geology.
646 17, 195–208.

647 Lark, R.M., 2016. Changes in the variance of a soil property along a transect, a comparison of
648 a non-stationary linear mixed model and a wavelet transform. Geoderma, 266, 84–97.

649 Lark, R.M., Bishop, T.F.A. 2007. Cokriging particle size fractions of the soil. European Journal
650 of Soil Science, 58, 763–774.

651 Lark, R.M., Marchant, B.P., Dove, D., Green, S.L., Stewart, H., Diesing, M. 2015. Combining
652 observations with acoustic swath bathymetry and backscatter to map seabed sediment
653 texture classes: the empirical best linear unbiased predictor. Sedimentary Geology, 328,
654 17–32.

655 Matérn, B. 1960. Spatial variation. stochastic models and their application to some problems
656 in forest surveys and other sampling investigations. Meddelanden från Statens Skogs-
657 forskningsinstut, 49, 1–144.

658 Neal, C., Reynolds, B., Kirchner, J.W., Rowland, P., Norris, D., Sleep, D., Lawlor, A., Woods,
659 C., Hacker, S., Guyatt, H., Vincent, C., Lehto, K., Grant, S., Willimas, J., Neal, M., Wick-

660 ham, H., Harman, S., Armstrong, L. 2013. High-frequency precipitation and stream water
661 quality time series from Plynlimon, Wales: an openly accessible data resource spanning
662 the periodic table. Hydrological Processes, 27, 2531–2539.

663 Otero, N., Tolosana-Delgado, R., Soler, A., Pawlowsky-Glahn, V., Canals, A. 2005. Relative
664 vs. absolute statistical analysis of compositions: a comparative study of surface waters of
665 a Mediterranean river. Water Research 9, 1404–1414.

666 Pardo-Igúzquiza, E., Dowd, P.A., 1998. Maximum likelihood inference of spatial covariance
667 parameters of soil properties. Soil Science, 163, 212–219.

668 Pawlowsky-Glahn, V., Olea, R.A. 2004. Geostatistical Analysis of Compositional Data. Oxford
669 University Press, New York.

670 R Core Team 2014., R: A language and environment for statistical computing. R Foundation
671 for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

672 Rawlins, B. G., Palumbo-Roe, B., Gooddy, D.C., Worrall, F., Smith, H. 2014. A model
673 of potential carbon dioxide efflux from surface water across England and Wales using
674 headwater stream survey data and landscape predictors. Biogeosciences 11, 1911–1925.

675 Reimann, C., Finne, T.E., Nordgulen, Ø., Sæther, O.M., Arnoldussen, A.M., Banks, D., 2009.
676 The influence of geology and land-use on inorganic stream water quality in the Oslo region,
677 Norway. Applied Geochemistry 24 , 1862–1874.

678 Simpson, P.R., Edmunds, W.M., Breward, N., Cook, J.M., Flight, D., Hall, G.E.M., Lister,
679 T.R., 1993. Geochemical mapping of stream water for environmental studies and mineral
680 exploration in the UK. Journal of Geochemical Exploration 49, 63–88.

681 Vaisanen, U., Misund, A., Chekushin, V., 1998. Ecogeochemical investigation: stream water
682 quality as an indicator of pollution in the border aresa of Finland, Norway and Russia.
683 Water, Air and Soil Pollution 104 (1–2), 205–219.

684 Verbeke, G., Molenberghs, G., 2000. Linear Mixed Models for Longitudinal Data. Springer,
685 New York.

686 Webster, R., McBratney, A.B., 1989. On the Akaike Information Criterion for choosing models
687 for variograms of soil properties. Journal of Soil Science, 40, 493–496.

688 Webster, R., Oliver, M.A. 2007., Geostatistics for Environmental Scientists. 2nd Edition John
689 Wiley & Sons, Chichester.

**Table 1**. Summary statistics of stream water data.

| | | | Original scale | | | | $\log_e$-transformed | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Variable | Units | n | Mean | Median | Standard deviation | Skewness | Mean | Median | Standard deviation | Skewness |
| Survey sites | | | | | | | | | | |
| As | μg $l^{-1}$ | 4138 | 2.65 | 1.40 | 7.18 | 18.4 | 0.43 | 0.34 | 0.88 | 0.74 |
| Cu | μg $l^{-1}$ | 3908 | 2.27 | 1.47 | 3.79 | 20.3 | 0.39 | 0.38 | 0.90 | 0.05 |
| Co | μg $l^{-1}$ | 4140 | 0.81 | 0.49 | 2.56 | 33.3 | −0.59 | −0.71 | 0.67 | 1.62 |
| Ni | μg $l^{-1}$ | 4140 | 4.75 | 3.73 | 6.23 | 23.1 | 1.37 | 1.32 | 0.53 | 1.10 |
| Conductivity | μS $cm^{-1}$ | 4242 | 1051.0 | 764.0 | 2522.2 | 16.3 | 6.69 | 6.64 | 0.53 | 1.29 |
| DOC | mg $l^{-1}$ | 3969 | 7.22 | 5.40 | 6.85 | 4.3 | 1.69 | 1.69 | 0.74 | 0.02 |
| pH | | 4249 | 7.80 | 7.84 | 0.36 | −1.94 | | | | |
| Monitor sites | | | | | | | | | | |
| As | μg $l^{-1}$ | 711 | 1.86 | 1.59 | 1.47 | 1.9 | 0.33 | 0.46 | 0.79 | −0.20 |
| Cu | μg $l^{-1}$ | 713 | 2.08 | 1.38 | 2.41 | 4.9 | 0.37 | 0.32 | 0.81 | 0.33 |
| Co | μg $l^{-1}$ | 713 | 0.61 | 0.43 | 0.66 | 4.1 | −0.75 | −0.84 | 0.63 | 1.18 |
| Ni | μg $l^{-1}$ | 713 | 4.09 | 3.22 | 3.54 | 4.0 | 1.22 | 1.17 | 0.55 | 0.99 |
| Conductivity | μS $cm^{-1}$ | 695 | 821.6 | 741.0 | 352.5 | 2.9 | 6.65 | 6.61 | 0.33 | 0.94 |
| DOC | mg $l^{-1}$ | 688 | 5.06 | 4.30 | 3.42 | 2.6 | 1.43 | 1.46 | 0.62 | −0.01 |
| pH | | 698 | 7.84 | 7.86 | 0.29 | −0.2 | | | | |

**Table 2**. Results for temporal models fitted by REML.

| Variable | $\kappa$ | $\ell$ | $\sigma^2_{\mathrm{S}}$ | $\sigma^2_{\mathrm{T,n}}$ | $\sigma^2_{\mathrm{T,c}}$ | $\phi$ /days | Range[a] /days | $A_{\mathrm{M}}$[b] | $A_{\mathrm{IID}}$[c] |
|---|---|---|---|---|---|---|---|---|---|
| As | 0.1 | −694.8 | | | | | | | |
| (log) | 0.5 | −742.0 | 0.505 | 0.008 | 0.117 | 8.76 | 26 | −1472.0 | −1009.3 |
| | 1.0 | −737.8 | | | | | | | |
| | 1.5 | −734.8 | | | | | | | |
| Cu | 0.1 | −243.6 | | | | | | | |
| (log) | 0.5 | −244.6 | 0.468 | 0.099 | 0.140 | 8.00 | 24 | −477.2 | −364.7 |
| | 1.0 | −243.7 | | | | | | | |
| | 1.5 | −242.9 | | | | | | | |
| Co | 0.1 | −739.7 | | | | | | | |
| (log) | 0.5 | −751.8 | | | | | | | |
| | 1.0 | −751.9 | 0.349 | 0.023 | 0.047 | 5.20 | 21 | −1491.8 | −1330.0 |
| | 1.5 | −751.5 | | | | | | | |
| Ni | 0.1 | −818.6 | | | | | | | |
| (log) | 0.5 | −821.4 | | | | | | | |
| | 1.0 | −822.1 | | | | | | | |
| | 1.5 | −822.2 | | | | | | | |
| | 2.0 | −822.3 | 0.238 | 0.022 | 0.037 | 4.81 | 26 | −1632.5 | −1521.8 |
| Conductivity | 0.1 | −986.1 | | | | | | | |
| (log) | 0.5 | −987.8 | | | | | | | |
| | 1.0 | −991.6 | 0.085 | 0.010 | 0.012 | 1.44 | 6 | −1970.5 | −1870.4 |
| | 1.5 | −991.2 | | | | | | | |
| DOC | 0.1 | −514.7 | | | | | | | |
| (log) | 0.5 | −515.1 | 0.309 | 0.039 | 0.049 | 4.30 | 13 | −1018.2 | −929.9 |
| | 1.0 | −514.5 | | | | | | | |
| | 1.5 | −514.0 | | | | | | | |
| pH | 0.1 | −1096.8 | 0.056 | 0.000 | 0.021 | 41.60 | 58 | −2181.5 | −2076.0 |
| | 0.5 | −1094.0 | | | | | | | |
| | 1.0 | −1087.7 | | | | | | | |
| | 1.5 | −1083.7 | | | | | | | |

The terms in the first 5 columns are defined in section 2.2.2.

a) Effective temporal range, the number of days over which the temporal correlation of the correlated within-site random effect decays to 0.05.

b) $A_{\mathrm{M}}$ is the AIC, Eq. [6], of the full linear mixed model with a temporally correlated component.

c) $A_{\mathrm{IID}}$ is the AIC, Eq. [6], of a linear mixed model in which the within-site variation has no temporal correlation.

**Table 3**. Values of $\widehat{A}$, computed from Eq [13], for the comparison of alternative models for temporal variograms for different spatial lag bins. The smallest value for each pair of models is shown in bold.

| | Spatial lag | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 5 km | | 10 km | | 20 km | | 40 km | |
| | Temporal model[a] | | | | | | | |
| | Correlated | Nugget | Correlated | Nugget | Correlated | Nugget | Correlated | Nugget |
| As | 85.9 | **83.2** | 75.6 | **73.3** | **59.0** | 64.2 | 67.2 | **65.7** |
| Cu | 94.9 | **91.0** | 76.1 | **72.4** | 40.4 | **36.4** | **38.0** | 52.7 |
| Co | 49.8 | **45.8** | 51.2 | **47.2** | 9.2 | **7.2** | 69.2 | **65.2** |
| Ni | 29.7 | **25.9** | 20.9 | **17.0** | −25.6 | **−29.6** | 32.6 | **28.6** |
| Conductivity | 8.7 | **4.7** | −1.6 | **−5.6** | 26.6 | **22.6** | 52.8 | **48.8** |
| DOC | **45.2** | 46.7 | 10.5 | **6.5** | 18.1 | **14.1** | 40.2 | **36.2** |
| pH | 21.1 | **20.4** | 10.6 | **6.6** | −4.9 | **−8.9** | 26.8 | **22.8** |

a). The correlated temporal model is the Matérn with the parameter $\kappa$ set at the value selected for the variable in the model of the monitor site data.

**Table 4**. Parameters for models fitted to empirical spatial variograms (all times pooled) for each analyte and the total variance of the temporal component of variation estimated from the data at monitoring sites, $C_{\mathrm{T}}$. All terms are defined in section 2.2.4.

| Variable | $C_0$ | $C_1$ | $\kappa$ | $\phi$/metres | $C_{\mathrm{T}}$ | $\frac{C_{\mathrm{T}}}{C_0}$ |
|---|---|---|---|---|---|---|
| As | 0.324 | 0.473 | 0.5 | 18 490 | 0.125 | 0.386 |
| Cu | 0.539 | 0.196 | 1.5 | 9 882 | 0.239 | 0.443 |
| Co | 0.279 | 0.113 | 1.5 | 8 895 | 0.070 | 0.251 |
| Ni | 0.149 | 0.129 | 0.5 | 19 506 | 0.059 | 0.396 |
| Conductivity | 0.113 | 0.103 | 2.0 | 8 785 | 0.022 | 0.195 |
| DOC | 0.320 | 0.203 | 1.0 | 9 989 | 0.088 | 0.275 |
| pH | 0.092 | 0.026 | 2.0 | 9 761 | 0.021 | 0.228 |

**Figure Captions**

1. First order stream sample sites (small light grey symbols) and monitoring stream sites (large dark grey symbols with numbers). Coordinates are in metres relative to the origin of the British National Grid. The location of the sampled region is seen in an inset map of Great Britain.

2. Time series plots for log concentration of arsenic at monitoring streams. Plot numbers refer to monitoring streams as shown in Figure 1, with sampling year also indicated.

3. Histograms of first-order stream data on original scales.

4. Histograms of first-order stream data with log-transformation applied to all but pH.

5. Spatial variograms of all first-order stream data computed from (•) comparisons among all observations and (∘) comparisons only from observations collected in the same season.

6. Spatial and temporal variograms of first-order stream data on arsenic.

7. Spatial and temporal variograms of first-order stream data on copper.

8. Spatial and temporal variograms of first-order stream data on cobalt.

9. Spatial and temporal variograms of first-order stream data on nickel.

10. Spatial and temporal variograms of first-order stream data on conductivity.

11. Spatial and temporal variograms of first-order stream data on dissolved organic carbon.

12. Spatial and temporal variograms of first-order stream data on pH.

13. Empirical spatial variograms (all times pooled) for each analyte with fitted models (parameters in Table 4). The broken horizontal lines show the nugget variance, $C_0$, and the total variance of the temporal component of variation estimated from the data at monitoring sites, $C_{\mathrm{T}}$.

14. (a) Kriging prediction of total As content of first order stream water and (b) standard error of the prediction treated as the summer mean.

15. Upper bound of the 95% confidence interval of predicted log As content of first order stream water treated as (a) a point sample or (b) the summer mean. The contour line on each map encloses the region where this value exceeds log 10μg $l^{-1}$.

16. (a) Kriging prediction of total Co content of first order stream water and (b) standard error of the prediction treated as the summer mean.

17. Upper bound of the 95% confidence interval of predicted log Co content of first order stream water treated as (a) a point sample or (b) the summer mean. The contour line on each map encloses the region where this value exceeds log 3μg $l^{-1}$.

Fig. 1.

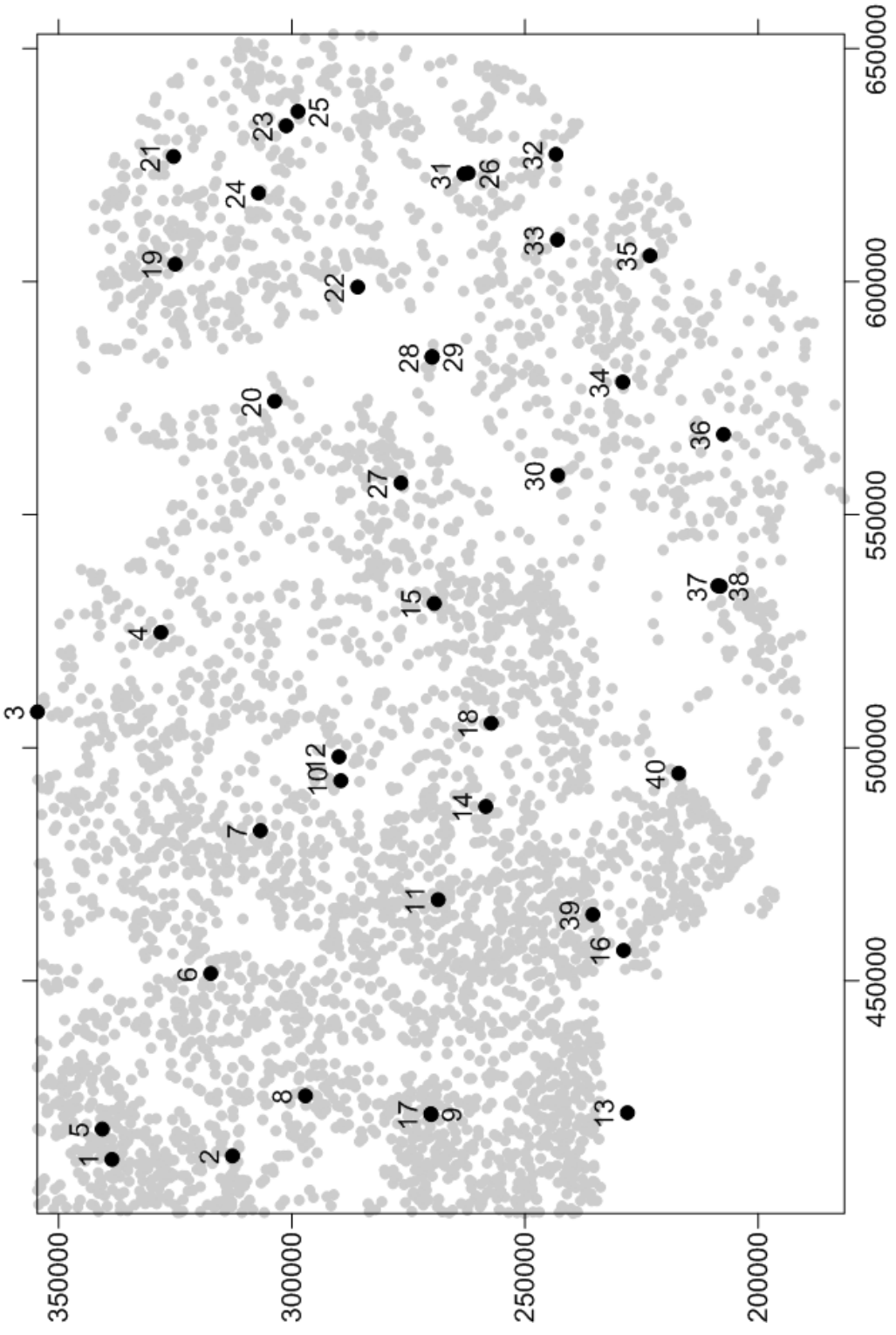

650000
600000
550000
500000
450000
350000
300000
250000
200000
1
5
2
8
17
9
13
6
39
16
11
7
14
40
10
12
18
3
4
15
37
38
27
30
36
20
28
29
34
22
33
35
19
21
24
23
25
31
26
32

Fig. 2.

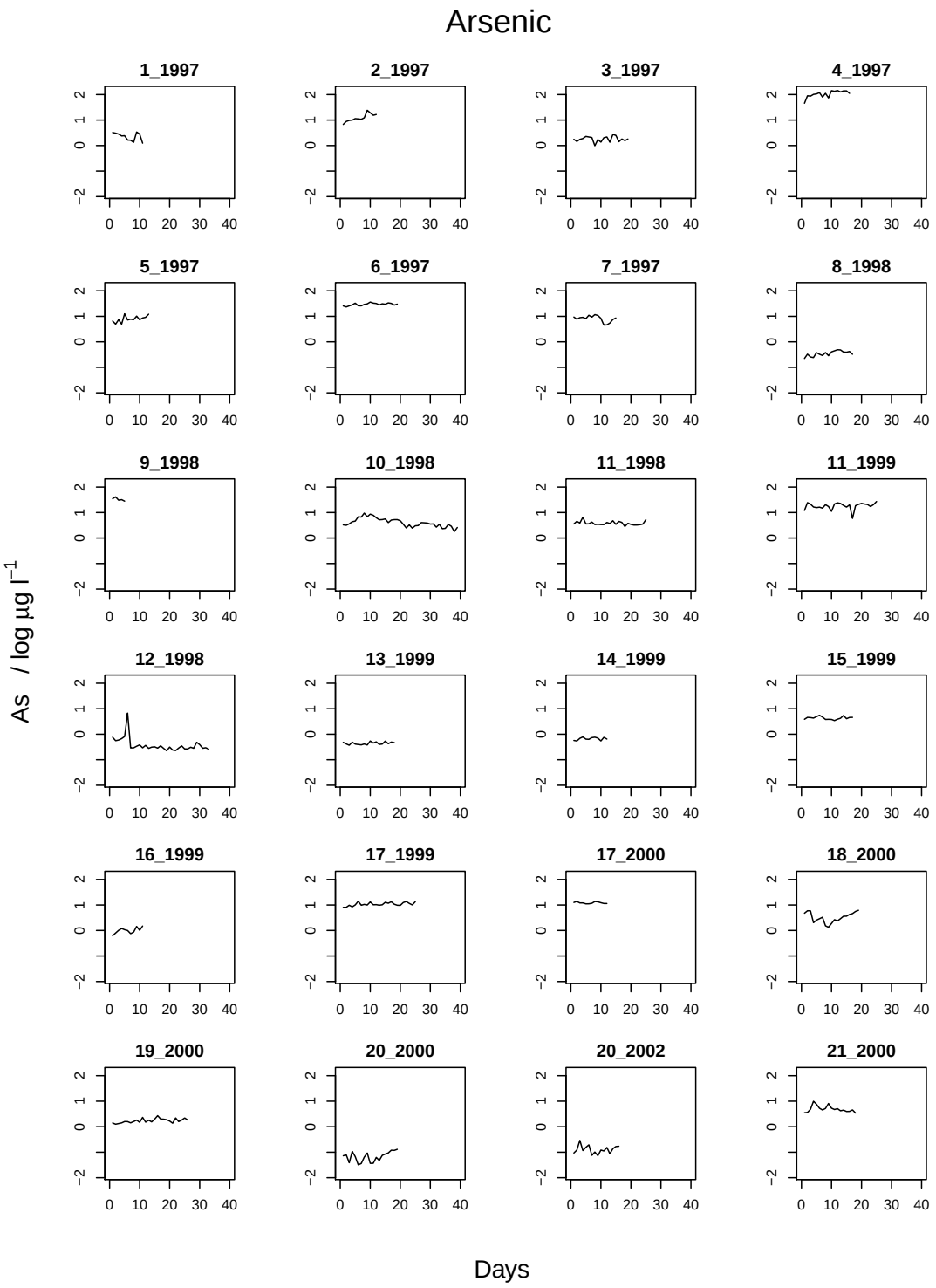

Arsenic
1_1997
2_1997
3_1997
4_1997
5_1997
6_1997
7_1997
8_1998
9_1998
10_1998
11_1998
11_1999
12_1998
13_1999
14_1999
15_1999
16_1999
17_1999
17_2000
18_2000
19_2000
20_2000
20_2002
21_2000
As / log µg l−1
Days

Fig. 2.

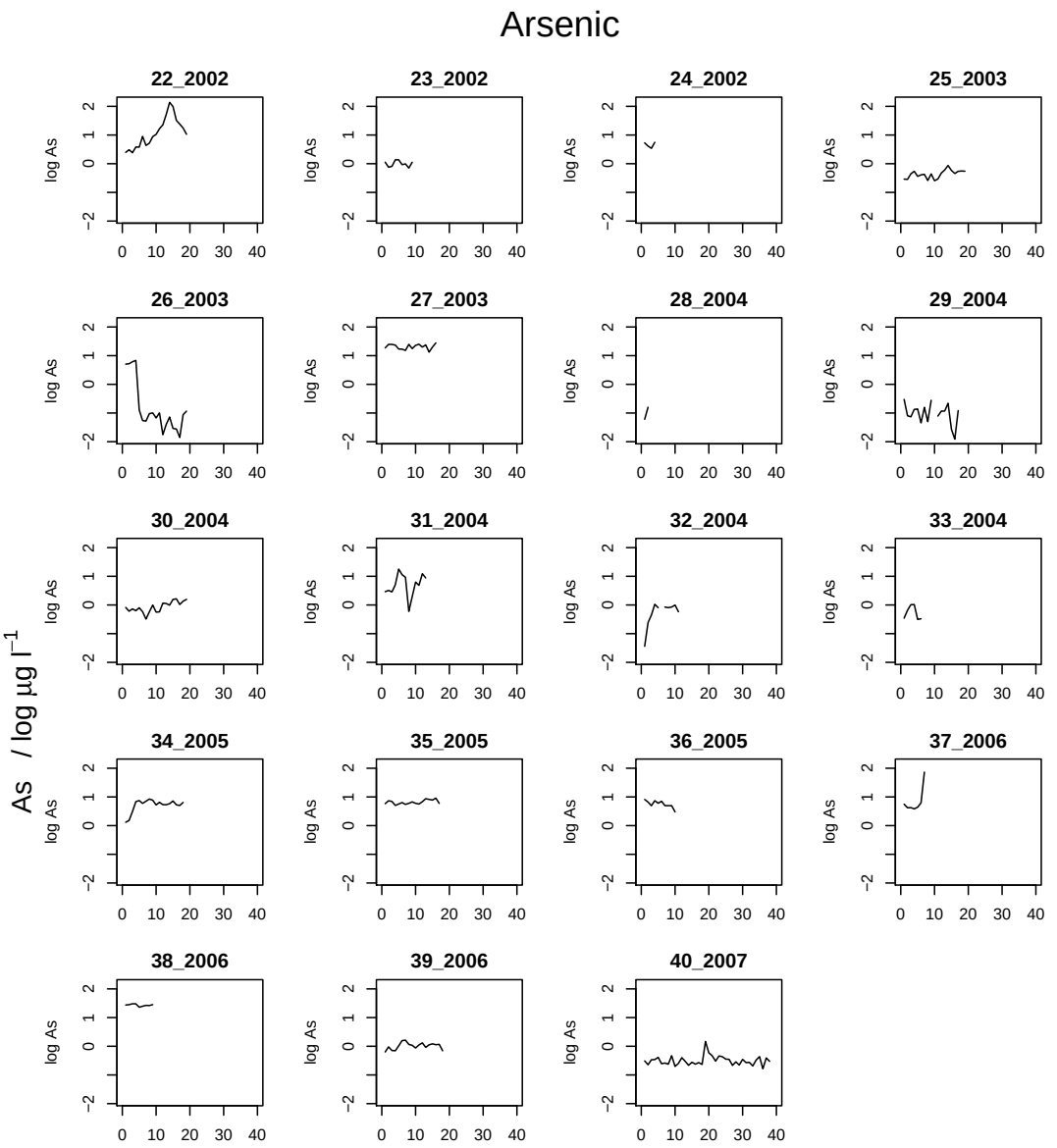

Arsenic
22_2002
23_2002
24_2002
25_2003
26_2003
27_2003
28_2004
29_2004
30_2004
31_2004
32_2004
33_2004
34_2005
35_2005
36_2005
37_2006
38_2006
39_2006
40_2007
log As
As / log µg l−1
Days

Fig. 3.

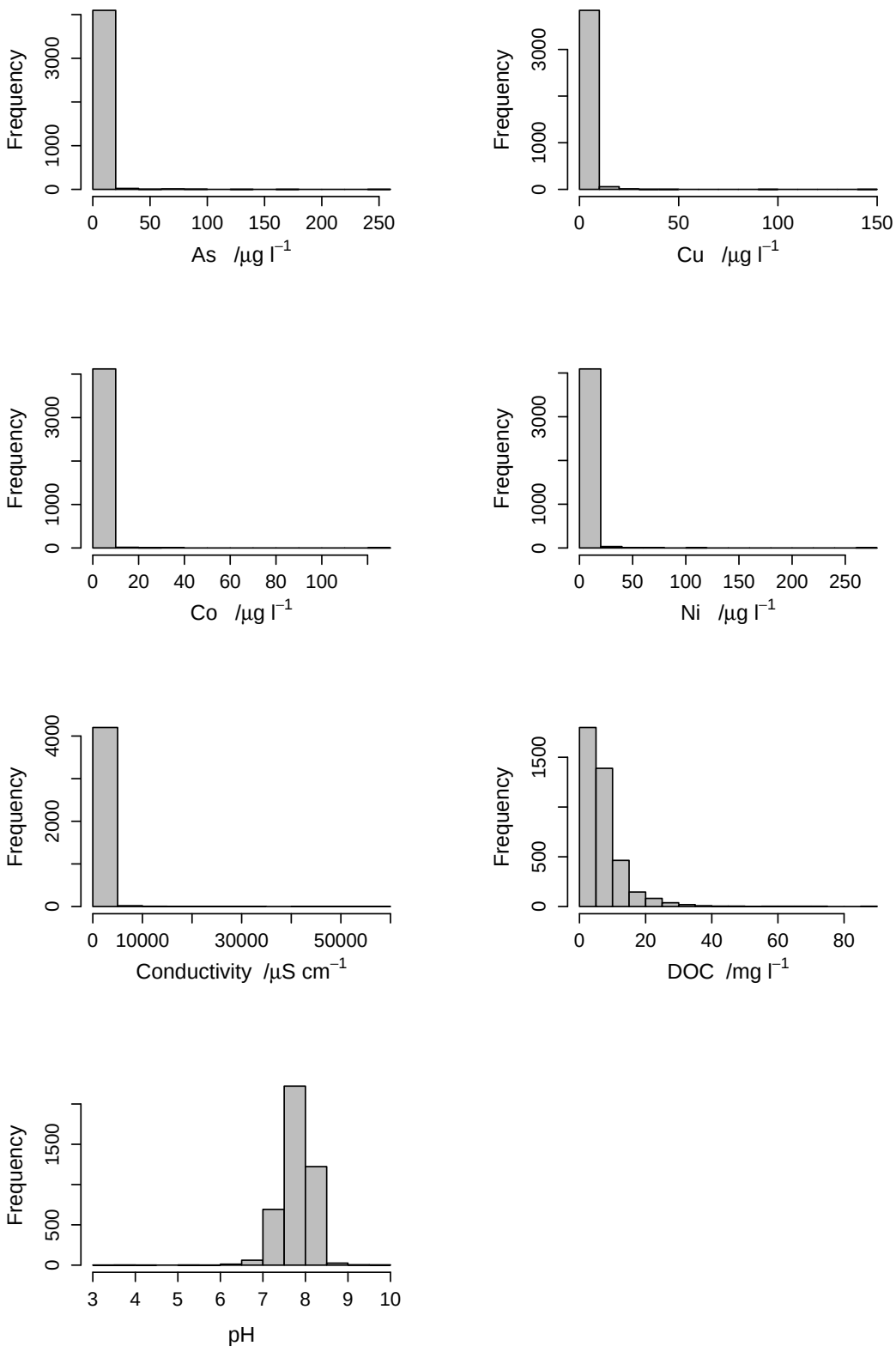

Frequency
0
1000
3000
0
50
100
150
200
250
As /µg l−1
Frequency
0
1000
3000
0
50
100
150
Cu /µg l−1
Frequency
0
1000
3000
0
20
40
60
80
100
Co /µg l−1
Frequency
0
1000
3000
0
50
100
150
200
250
Ni /µg l−1
Frequency
0
2000
4000
0
10000
30000
50000
Conductivity /µS cm−1
Frequency
0
500
1500
0
20
40
60
80
DOC /mg l−1
Frequency
0
500
1500
3
4
5
6
7
8
9
10
pH

Fig. 4.

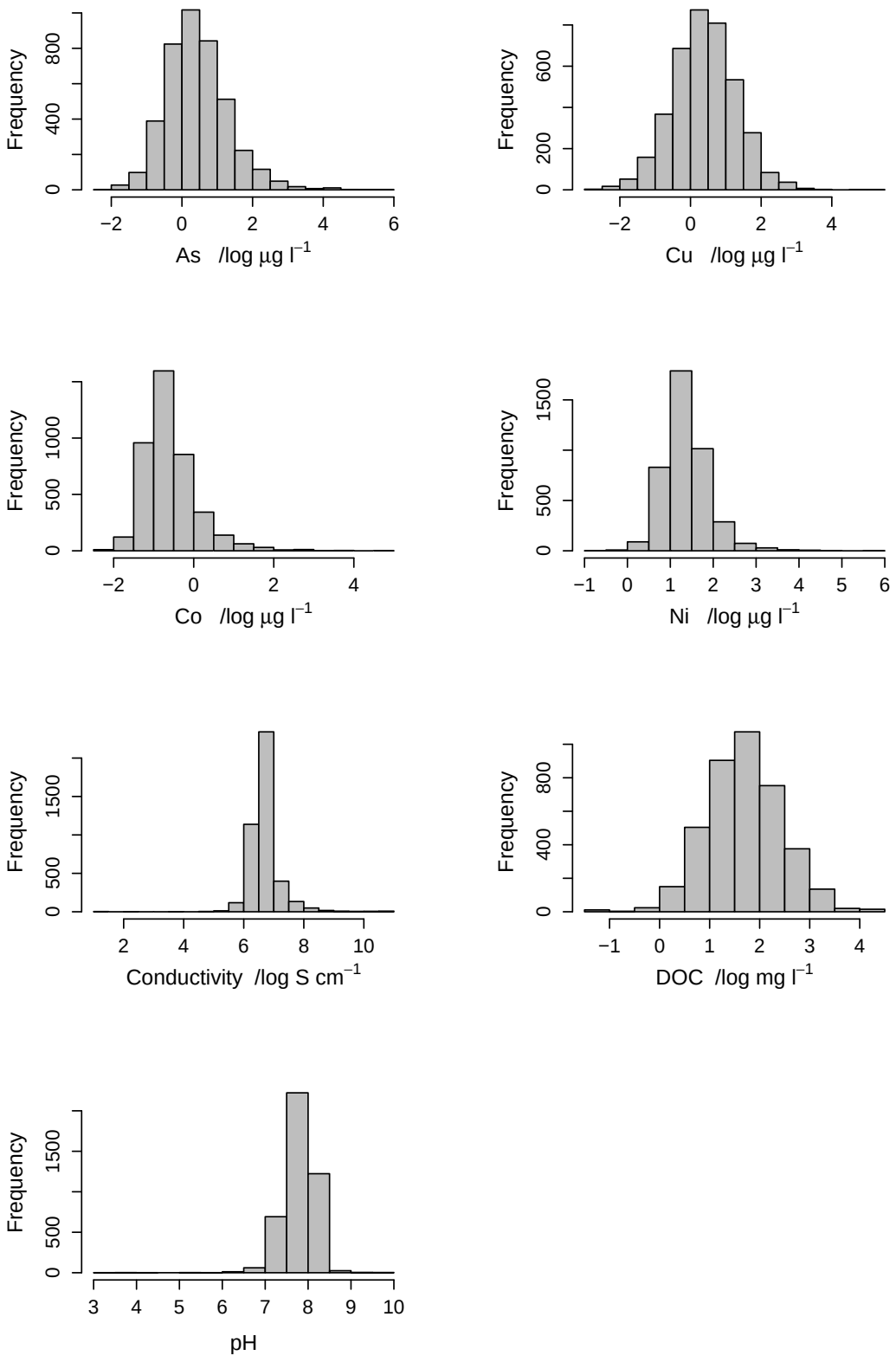

Frequency
As /log µg l$^{-1}$
Cu /log µg l$^{-1}$
Co /log µg l$^{-1}$
Ni /log µg l$^{-1}$
Conductivity /log S cm$^{-1}$
DOC /log mg l$^{-1}$
pH

Fig. 5.

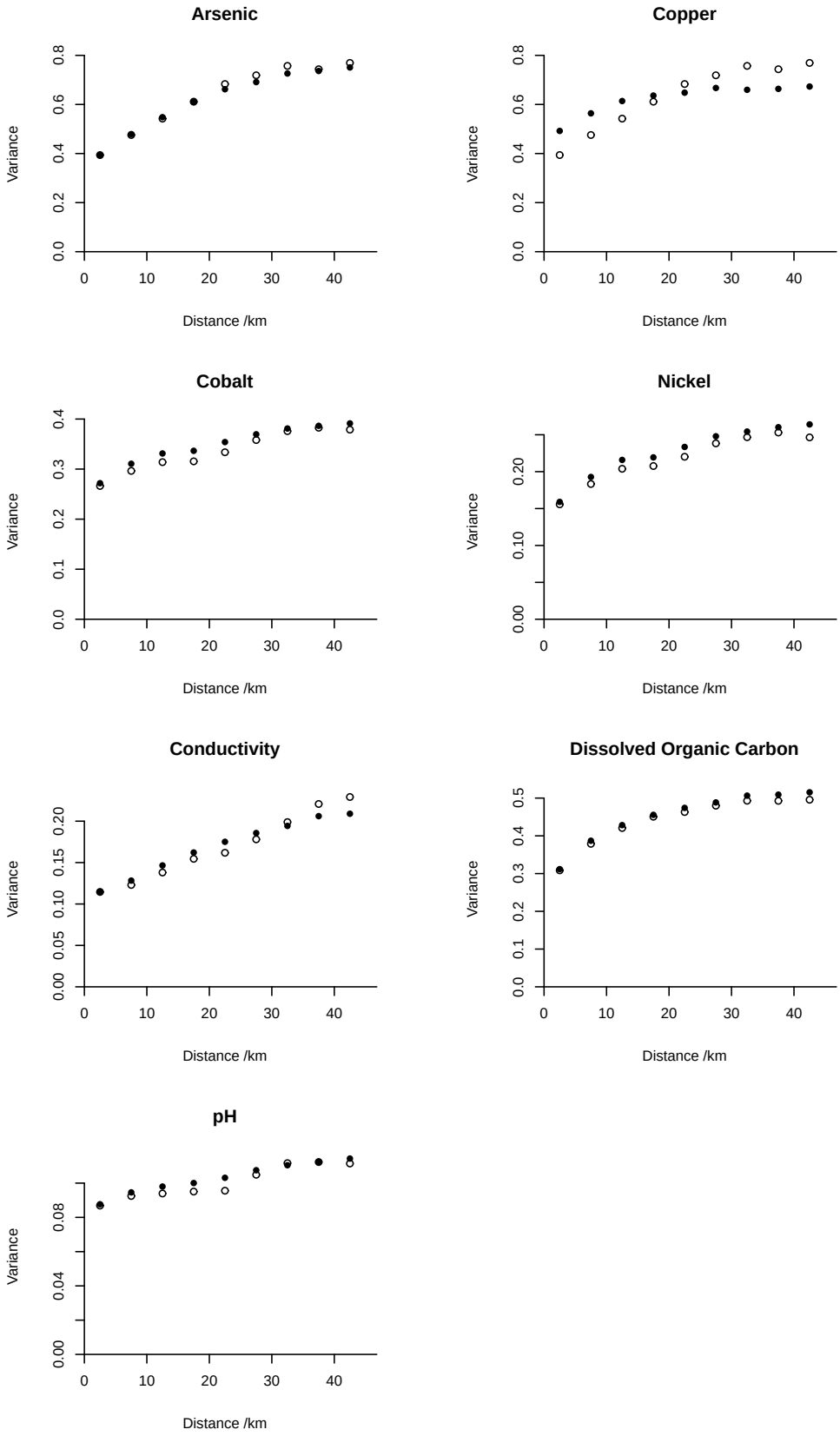

Arsenic
Copper
Cobalt
Nickel
Conductivity
Dissolved Organic Carbon
pH
Variance
Distance /km

Fig. 6.

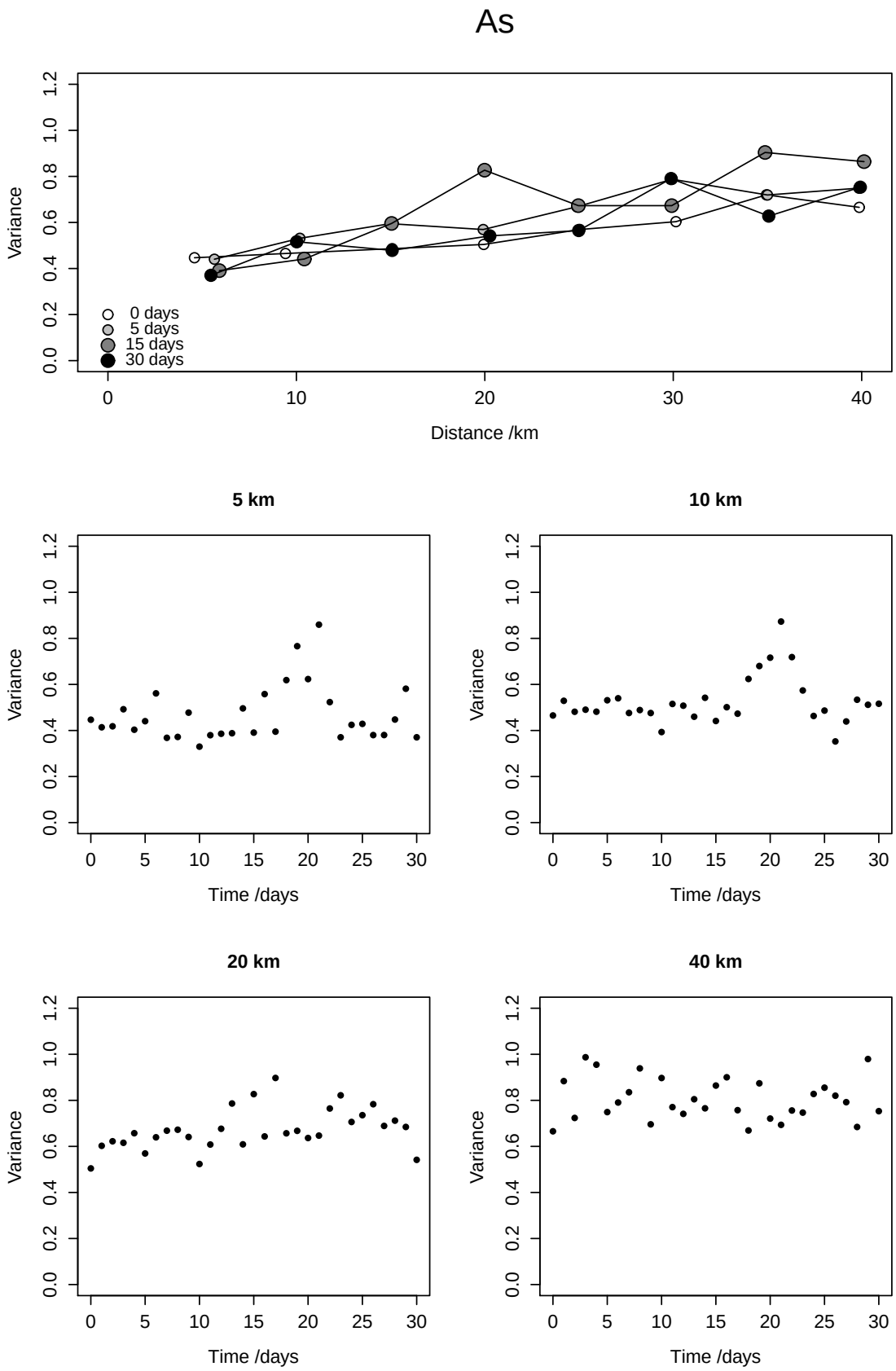

As
Variance
Distance /km
0 days
5 days
15 days
30 days
5 km
10 km
20 km
40 km
Time /days

Fig. 7.

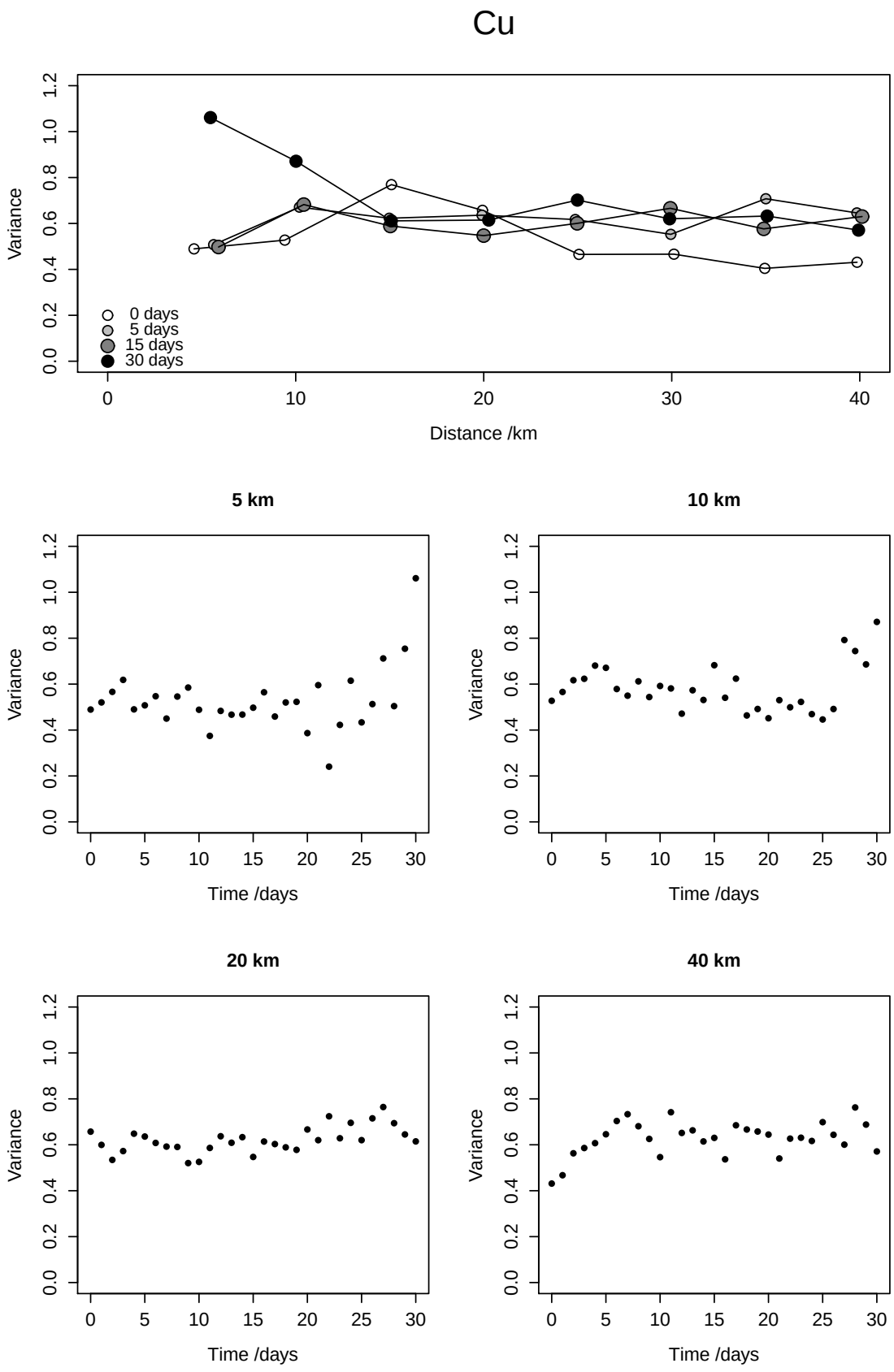

Cu
Variance
Distance /km
0 days
5 days
15 days
30 days
5 km
10 km
20 km
40 km
Time /days

Fig. 8.

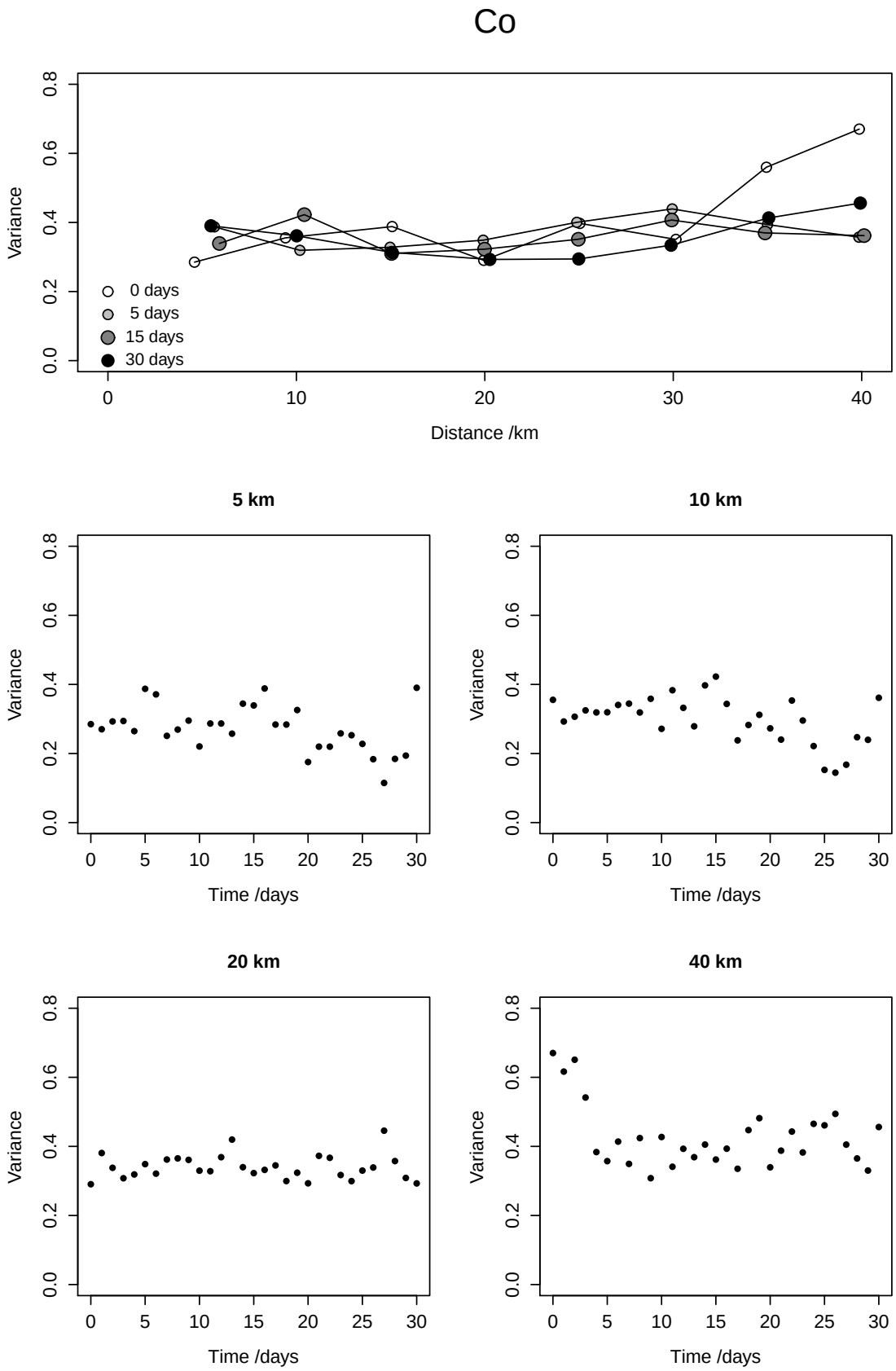

Co
Variance
Distance /km
0 days
5 days
15 days
30 days
5 km
10 km
20 km
40 km
Time /days

Fig. 9.

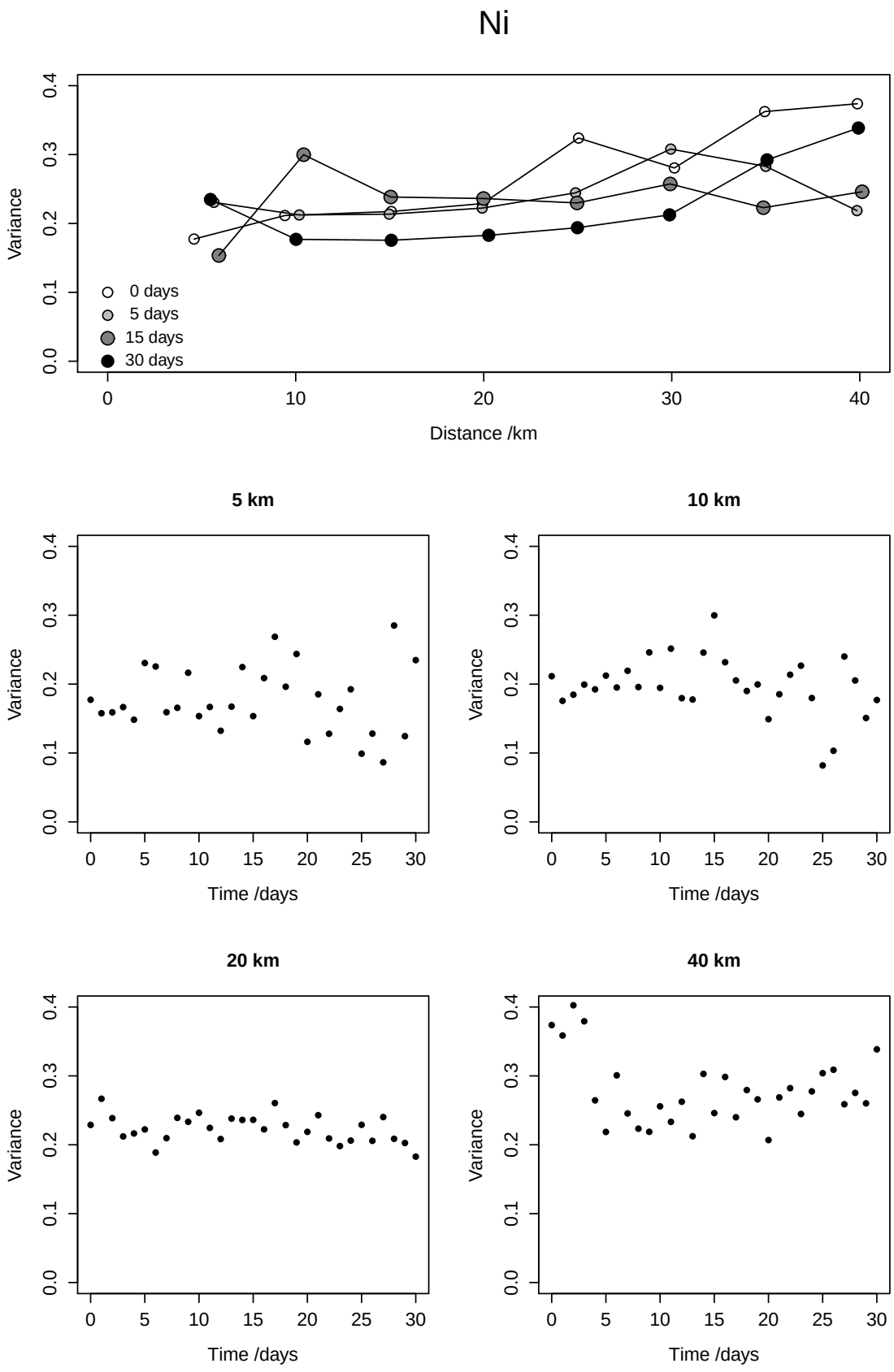

Ni
Variance
Distance /km
0 days
5 days
15 days
30 days
5 km
10 km
20 km
40 km
Time /days

Fig. 10.

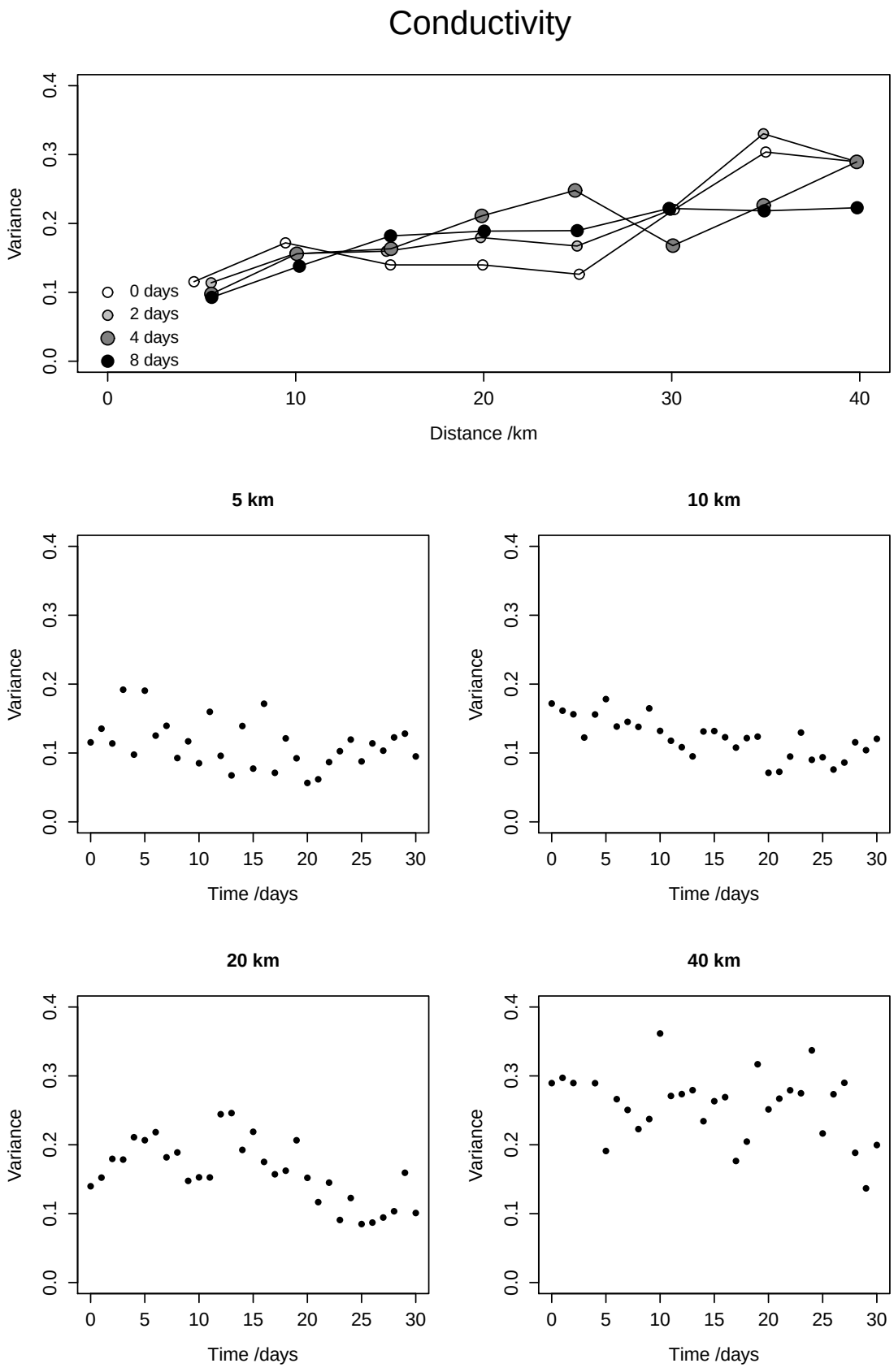

Conductivity
Variance
Distance /km
0 days
2 days
4 days
8 days
5 km
10 km
20 km
40 km
Time /days

Fig. 11.

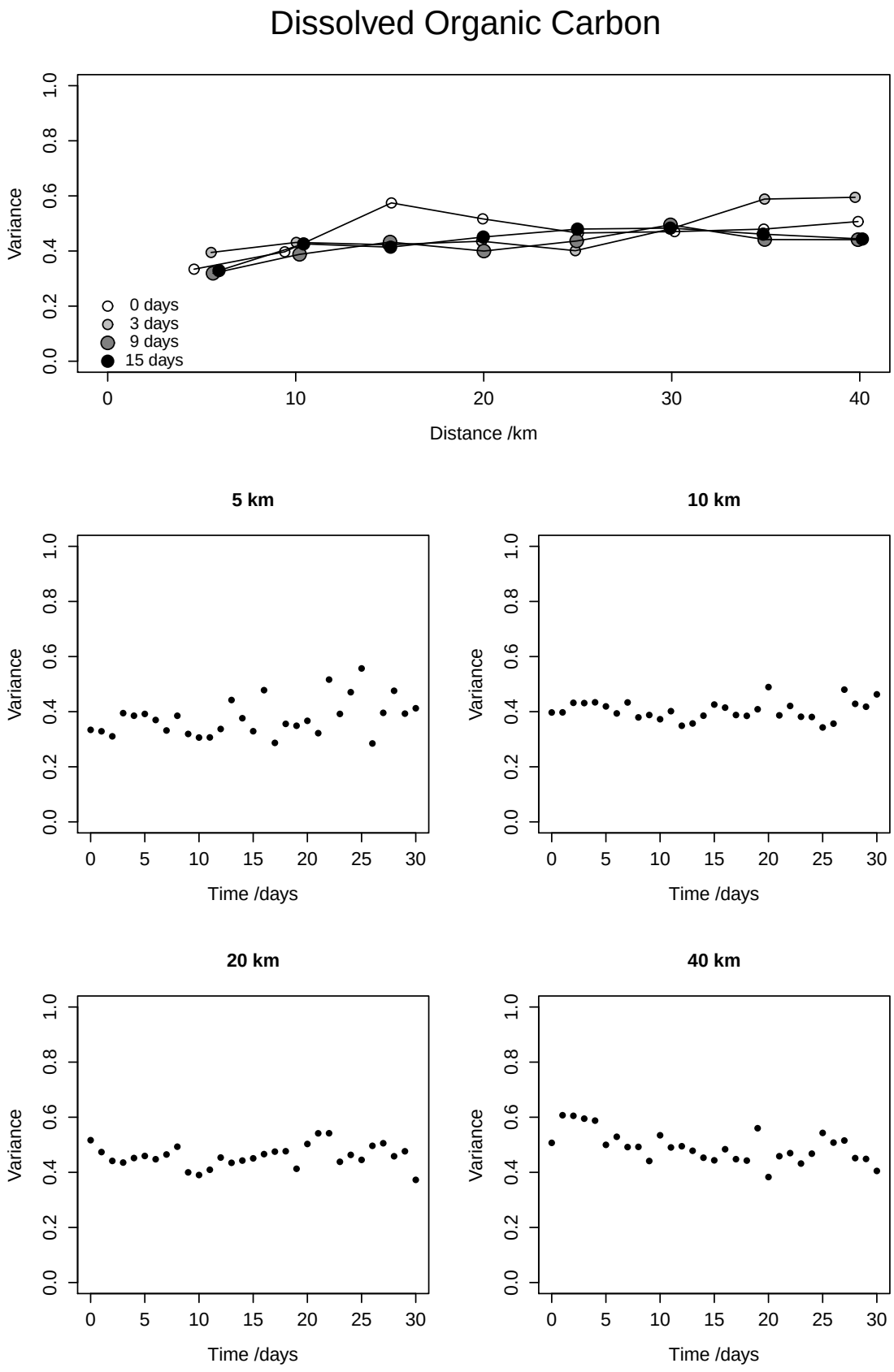

Dissolved Organic Carbon
Variance
Distance /km
0 days
3 days
9 days
15 days
5 km
10 km
20 km
40 km
Time /days

Fig. 12.

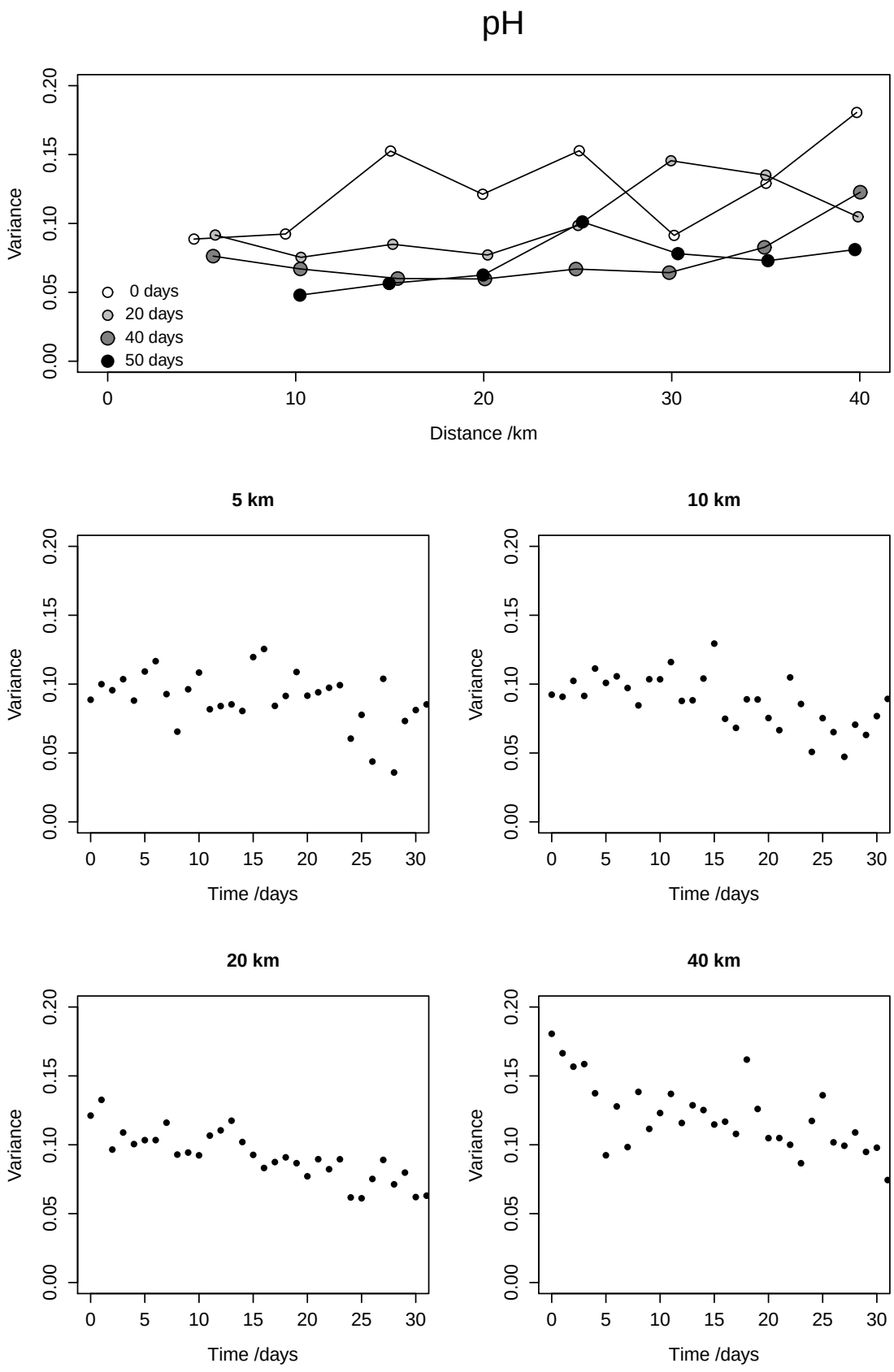

pH
Variance
Distance /km
0 days
20 days
40 days
50 days
5 km
10 km
20 km
40 km
Time /days

Fig. 13.

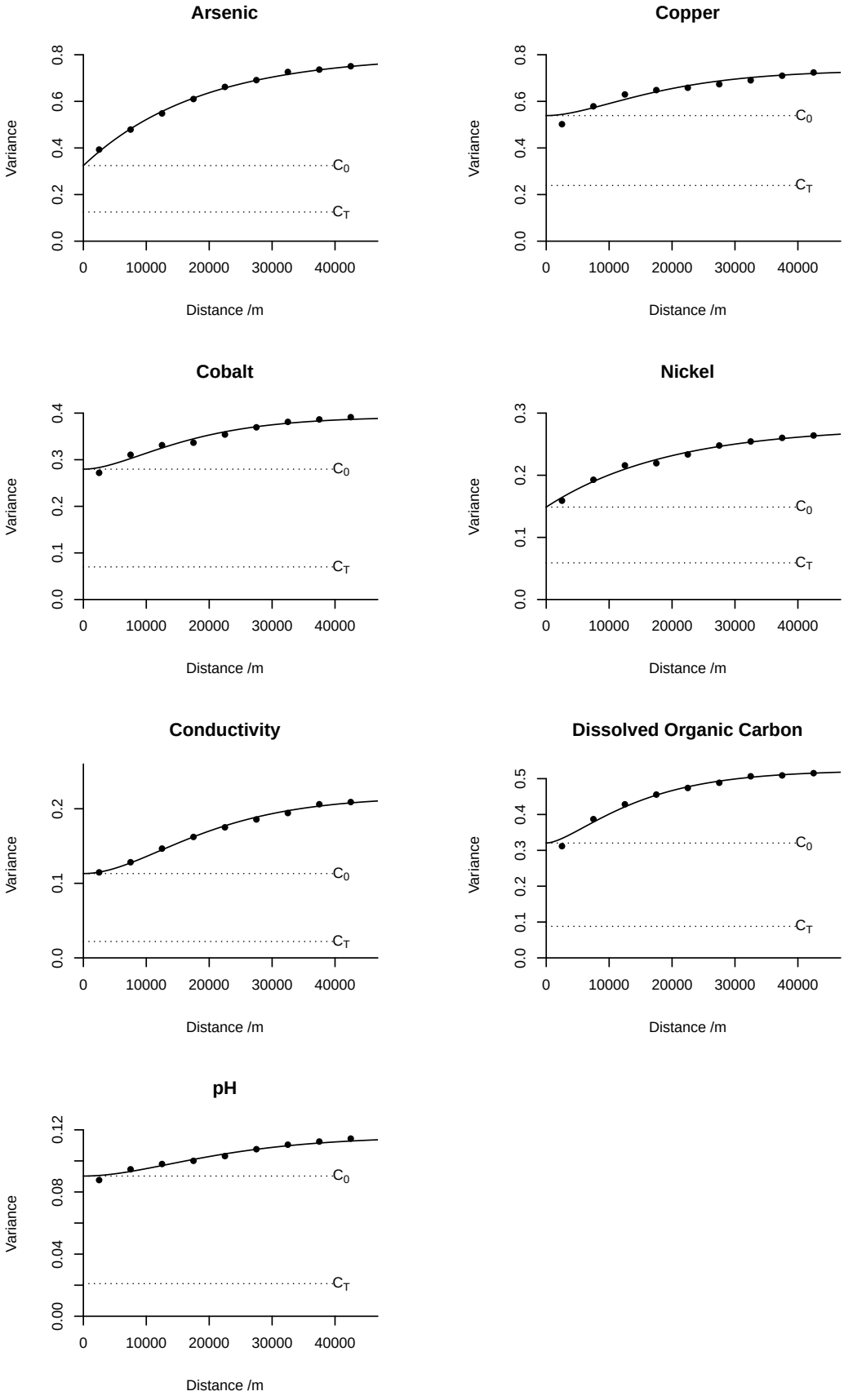
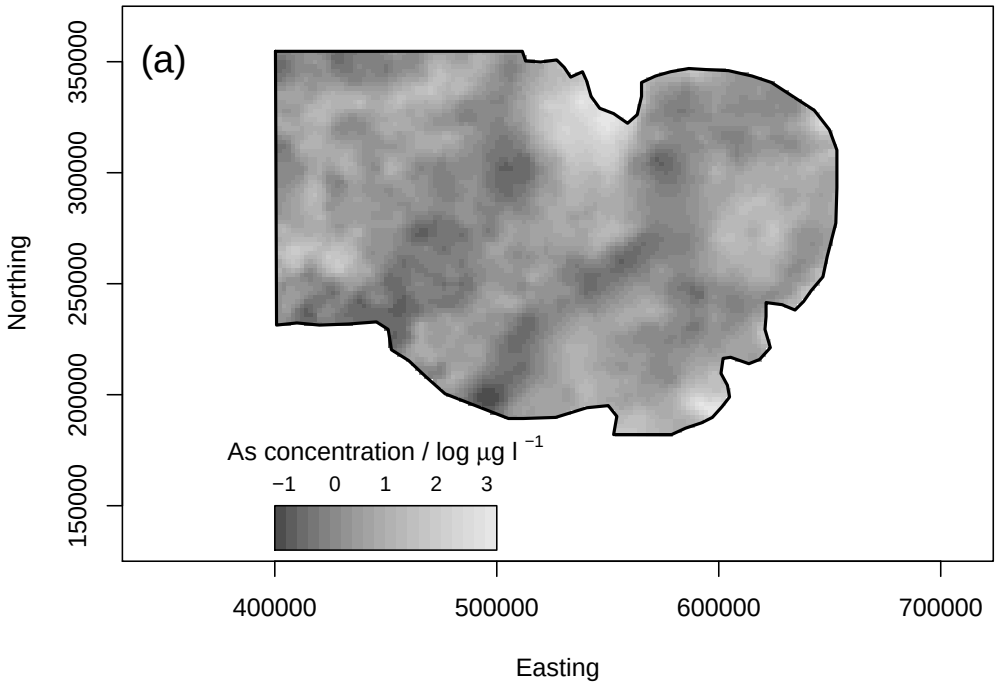

Arsenic
Copper
Cobalt
Nickel
Conductivity
Dissolved Organic Carbon
pH
Variance
Distance /m
$C_0$
$C_T$

Fig. 14.


(a)
Northing
150000
200000
250000
300000
350000
As concentration / log µg l −1
−1 0 1 2 3
400000
500000
600000
700000
Easting

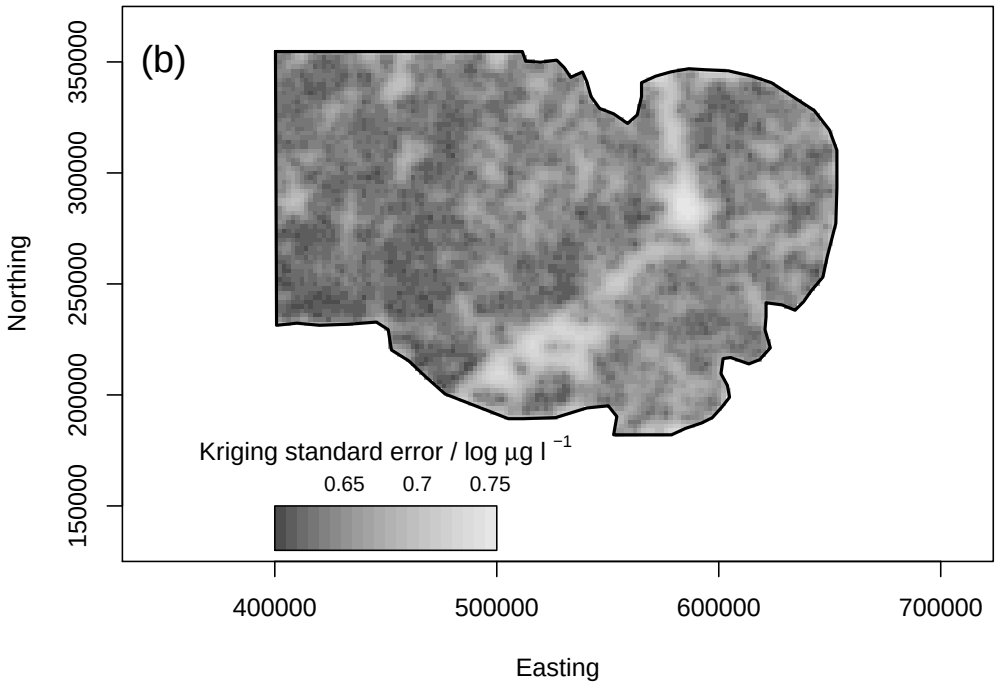
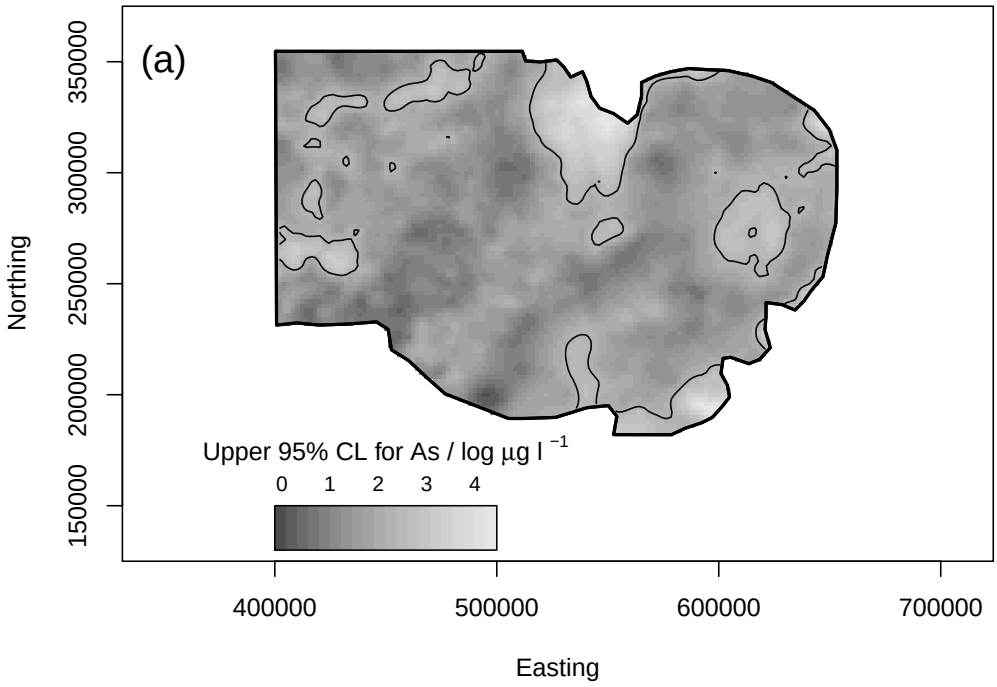

(b)
Northing
150000
200000
250000
300000
350000
Kriging standard error / log µg l −1
0.65 0.7 0.75
400000
500000
600000
700000
Easting

Fig. 15.


(a)
Northing
150000
200000
250000
300000
350000
Upper 95% CL for As / log µg l −1
0 1 2 3 4
400000
500000
600000
700000
Easting

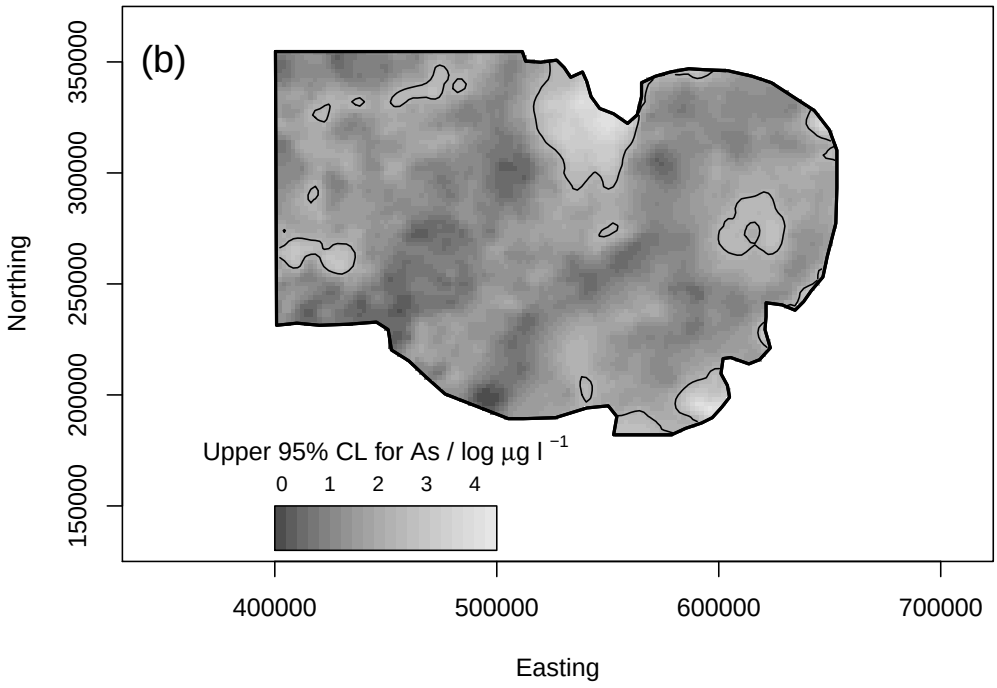
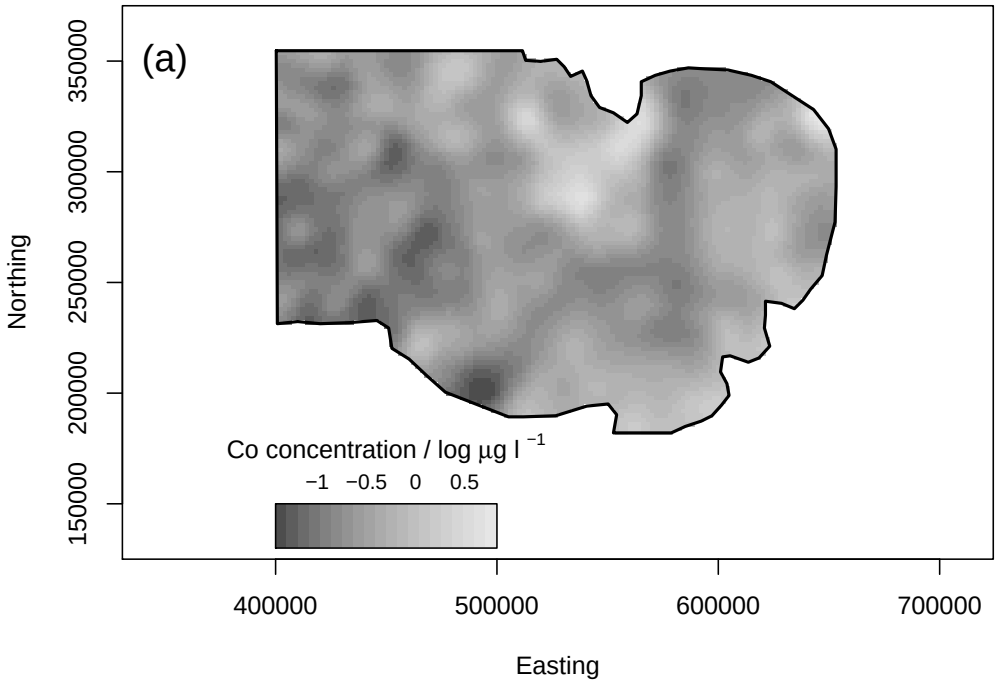

(b)
Northing
150000
200000
250000
300000
350000
Upper 95% CL for As / log µg l −1
0 1 2 3 4
400000
500000
600000
700000
Easting

Fig. 16.


(a)
Northing
150000
200000
250000
300000
350000
Co concentration / log µg l −1
−1 −0.5 0 0.5
400000
500000
600000
700000
Easting

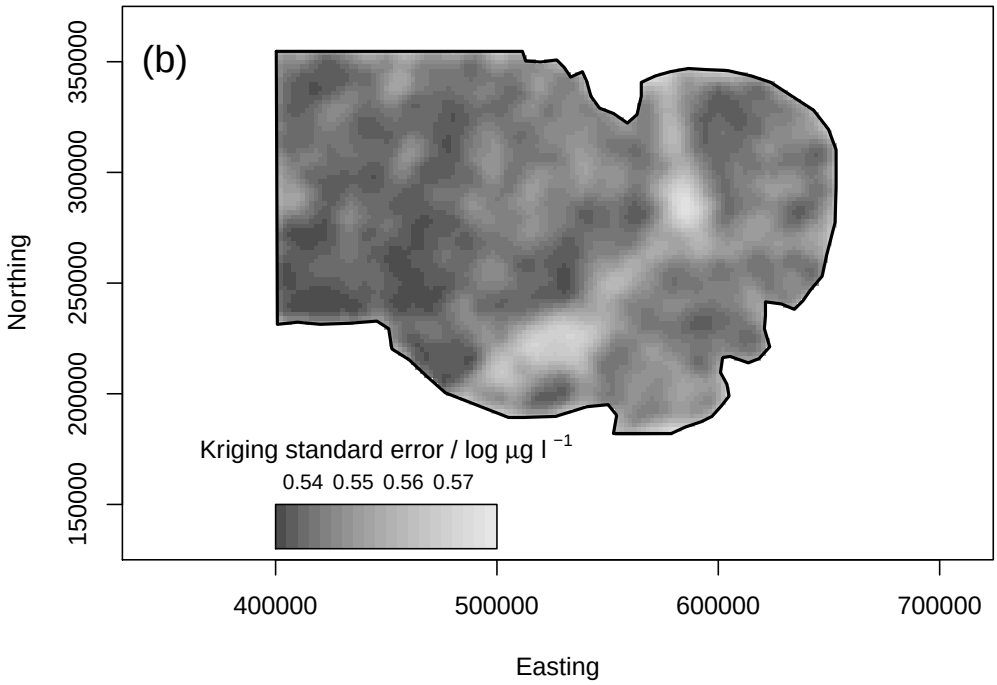
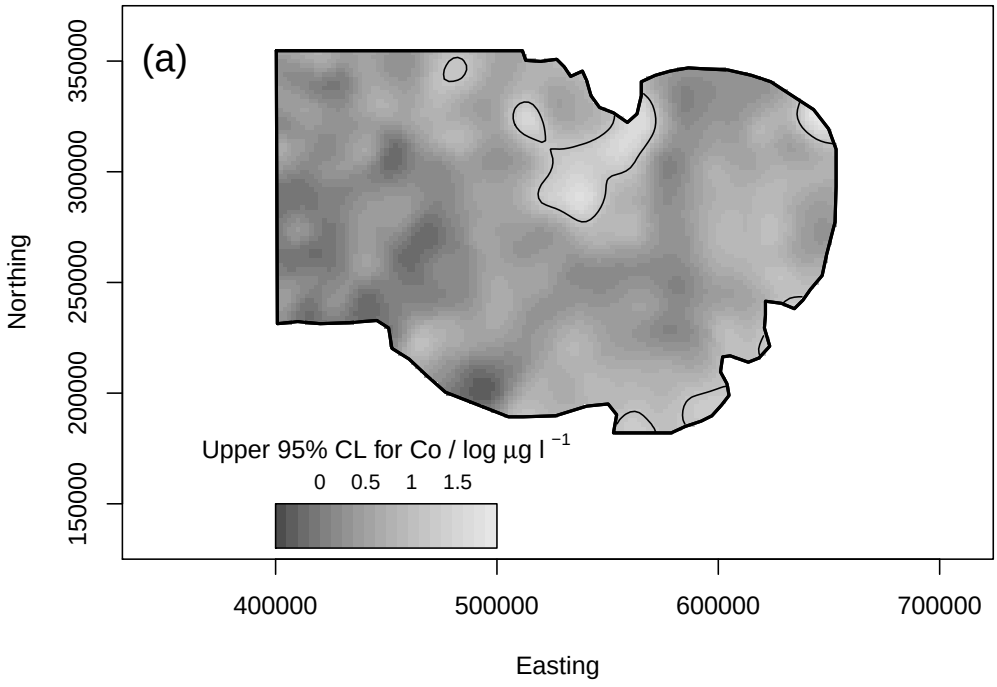
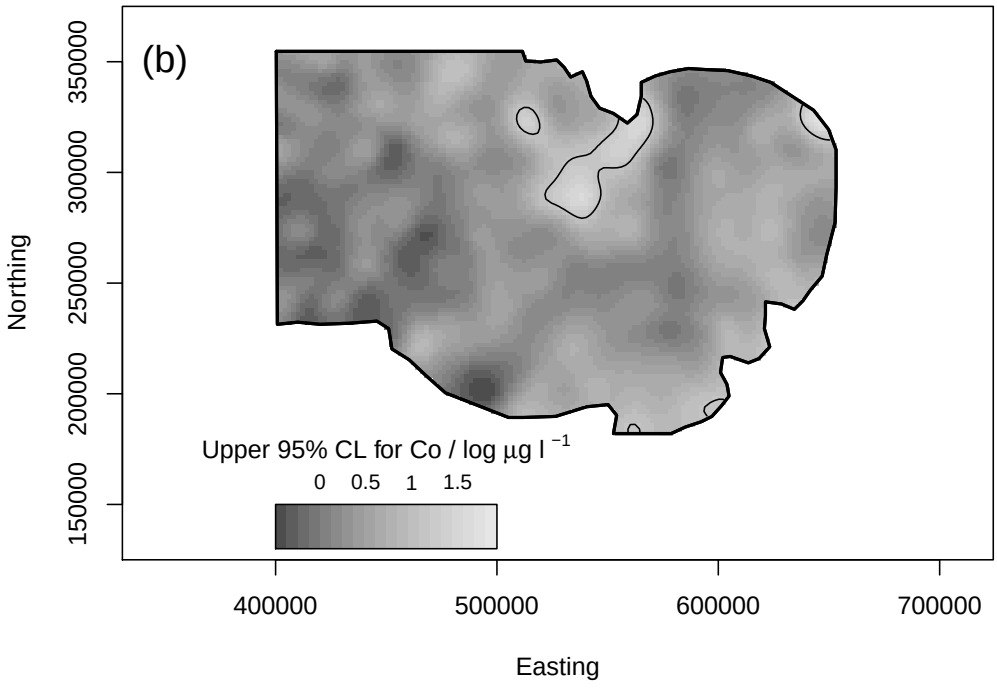

(b)
Northing
150000
200000
250000
300000
350000
Kriging standard error / log µg l −1
0.54 0.55 0.56 0.57
400000
500000
600000
700000
Easting

Fig. 17.


(a)
Northing
150000
200000
250000
300000
350000
Upper 95% CL for Co / log µg l −1
0 0.5 1 1.5
400000
500000
600000
700000
Easting


(b)
Northing
150000
200000
250000
300000
350000
Upper 95% CL for Co / log µg l −1
0 0.5 1 1.5
400000
500000
600000
700000
Easting