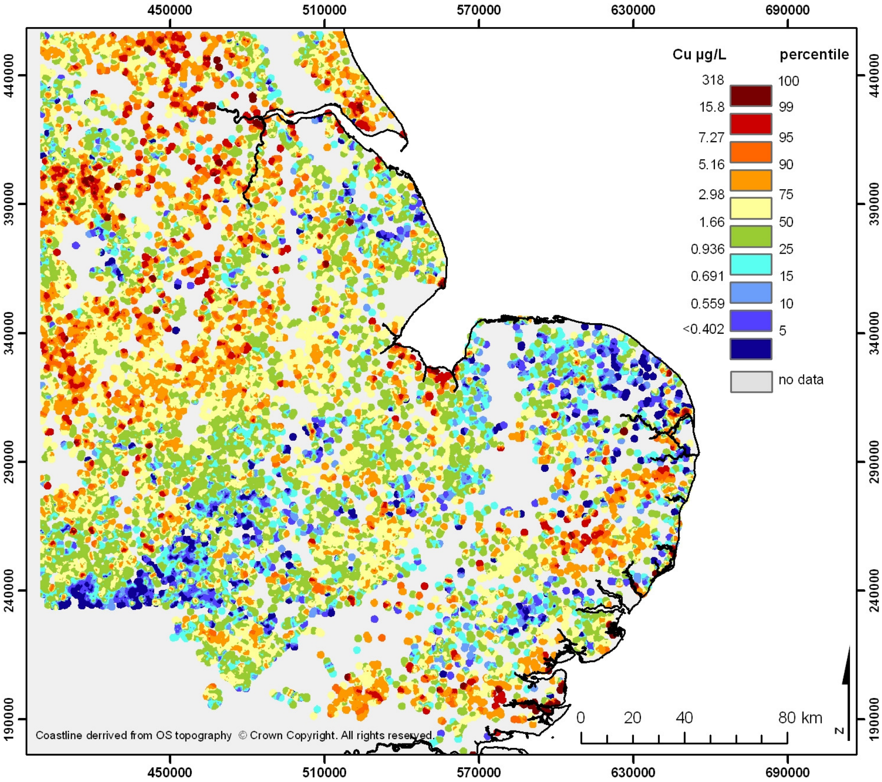
British
Geological Survey
NATURAL ENVIRONMENT RESEARCH COUNCIL

Comparison of spatial Cu stream water concentrations with intra- and inter- annual monitoring data

Environmental Geoscience Baselines Programme

Internal Report IR/10/017



BRITISH GEOLOGICAL SURVEY

ENVIRONMENTAL GEOSCIENCE BASELINES PROGRAMME
INTERNAL REPORT IR/10/017

Comparison of spatial Cu stream water concentrations with intra- and inter- annual monitoring data

J Bearcock and E L Ander

Keywords

geochemical baseline, hydrochemistry, copper, monitor site, G-BASE.

National Grid Reference

SW corner 399000,227000
Centre point 527000,342000
NE corner 653000,457000

Front cover

Copper in stream waters from Eastern England

Bibliographical reference

BEARCOCK, J. AND ANDER, E. L.. 2010. Comparison of spatial Cu stream water concentrations with intra- and inter- annual monitoring data. *British Geological Survey Internal Report*, IR/10/017. 42pp.

Keyworth, Nottingham British Geological Survey 2010

BRITISH GEOLOGICAL SURVEY

The full range of our publications is available from BGS shops at Nottingham, Edinburgh, London and Cardiff (Welsh publications only) see contact details below or shop online at www.geologyshop.com

The London Information Office also maintains a reference collection of BGS publications, including maps, for consultation.

We publish an annual catalogue of our maps and other publications; this catalogue is available online or from any of the BGS shops.

The British Geological Survey carries out the geological survey of Great Britain and Northern Ireland (the latter as an agency service for the government of Northern Ireland), and of the surrounding continental shelf, as well as basic research projects. It also undertakes programmes of technical aid in geology in developing countries.

The British Geological Survey is a component body of the Natural Environment Research Council.

British Geological Survey offices

BGS Central Enquiries Desk
Tel 0115 936 3143 Fax 0115 936 3276
email enquiries@bgs.ac.uk

Kingsley Dunham Centre, Keyworth, Nottingham NG12 5GG
Tel 0115 936 3241 Fax 0115 936 3488
email sales@bgs.ac.uk

Murchison House, West Mains Road, Edinburgh EH9 3LA
Tel 0131 667 1000 Fax 0131 668 2683
email scotsales@bgs.ac.uk

Natural History Museum, Cromwell Road, London SW7 5BD
Tel 020 7589 4090 Fax 020 7584 8270
Tel 020 7942 5344/45 email bgslondon@bgs.ac.uk

Columbus House, Greenmeadow Springs, Tongwynlais, Cardiff CF15 7NE
Tel 029 2052 1962 Fax 029 2052 1963

Forde House, Park Five Business Centre, Harrier Way, Sowton EX2 7HU
Tel 01392 445271 Fax 01392 445371

Maclean Building, Crowmarsh Gifford, Wallingford OX10 8BB
Tel 01491 838800 Fax 01491 692345

Geological Survey of Northern Ireland, Colby House, Stranmillis Court, Belfast BT9 5BF
Tel 028 9038 8462 Fax 028 9038 8461

www.bgs.ac.uk/gsni/

Parent Body

Natural Environment Research Council, Polaris House, North Star Avenue, Swindon SN2 1EU
Tel 01793 411500 Fax 01793 411501
www.nerc.ac.uk

Website www.bgs.ac.uk
Shop online at www.geologyshop.com

Foreword

This report provides an initial quantitative comparison of G-BASE spatial data for a trace element (Cu) in stream waters with short-term daily monitoring data. The regional data were collected over a period of 13 field seasons.

Acknowledgements

The authors would like to thank the large number of BGS staff and students who have contributed to the organisation, collection, analysis, and quality control of the data used in this study. This report has benefited from the technical review of Dr Pauline Smedley.

Contents

FIGURES

TABLES

Summary

The Geochemical Baseline Survey of the Environment (G-BASE) is the long established high resolution geochemical baseline mapping project of the British Geological Survey. The geochemical mapping is based on the systematic sampling and analysis of soils, stream sediments, and stream waters, and this study concerns the latter.

The central aim of this study is to establish whether the spatial variation, caused by geology, topography etc, predominates over temporal variations; and establish whether such temporal variations in trace element concentrations limit the representativeness of spatial distribution maps. Water concentrations are known to vary on a diurnal basis, and in response to external factors such as rainfall. This report addresses an assumption that stream water chemistry significantly varies over prolonged sampling periods.

This report describes the temporal stream water Cu data from samples collected during the summer field seasons in central and eastern England between 1997 and 2007. These temporal samples were collected in parallel with the primary samples used for mapping. The temporal data are obtained from sampling “monitor sites”: a carefully selected site sampled each day from each temporary fieldbase used by the field teams.

Comparison of these data showed that the variations of Cu concentrations sampled over time at monitor sites, were less than spatial variations determined by factors such as geology, topography and landuse. The concentration of Cu in the monitor site samples varied by as much as 13 mg/L at one site, so the relevance of outliers, and their effect on the interpretation and mapping of spatial data were examined. In comparison to the spatial variations, temporal variations were limited.

1 Introduction

Stream water samples are part of the suite of samples collected for the BGS's national Geochemical Baseline Survey of the Environment (G-BASE) geochemical mapping project. Full details of the G-BASE project's history and methods can be found in Johnson (2005) and Johnson et al. (2005). For the G-BASE project, a single stream water sample is taken at any one site, which provides a "snapshot" of the water chemistry at the time of sampling. Samples are primarily collected from first and second order streams. Analyses are obtained for dissolved major, minor and trace elements. In order to do this two subsamples are filtered to <0.45 µm, one of which is acidified to 1% (v/v) HNO_3 with high purity acid for cation analysis.

The chemistry of stream water at a sampling site is dependent on numerous chemical, biological and hydrological factors. The water that ends up in streams is a mixture of waters which have passed through different environments before reaching the stream. Factors which influence the proportions of each component include catchment properties such as topography, rock permeability, and hydraulic connectivity; as well as short-term climatic variations (British Geological Survey, 1999). Once the water has reached the stream, its chemistry continues to be modified as it flows downstream (Drever, 1997). As the G-BASE sample represents one sample at each location it cannot account for any natural variability over that summer sampling season, or indeed over a number of years. For this reason a "monitor site", in the vicinity of the sampling campaign, is used to assess short-term temporal variation over the duration of the sampling campaign.

The G-BASE survey is conducted from temporary field bases as it progresses. These are generally occupied for 3-4 weeks. At each field base a monitor site is selected, which is a first or second order stream to be consistent with the spatial sample sites. The monitor site is sampled over the duration of the stay at the field base, preferably at a similar time each day. Monitor sites are carefully selected to ensure that, as far as possible, the stream water chemistry is only affected by natural processes and not affected by industrial or urban inputs. In addition the access must be easy and the site safe, to make the daily sampling as efficient as possible. The monitor site samples were analysed to provide a temporal context for the spatial data. The availability of a temporal data set provides evidence to help understand the dynamic temporal variability of the chemical composition of local stream water in relation to the spatial data. The collection of these temporal data is therefore used to test the validity of the national mapping of the data.

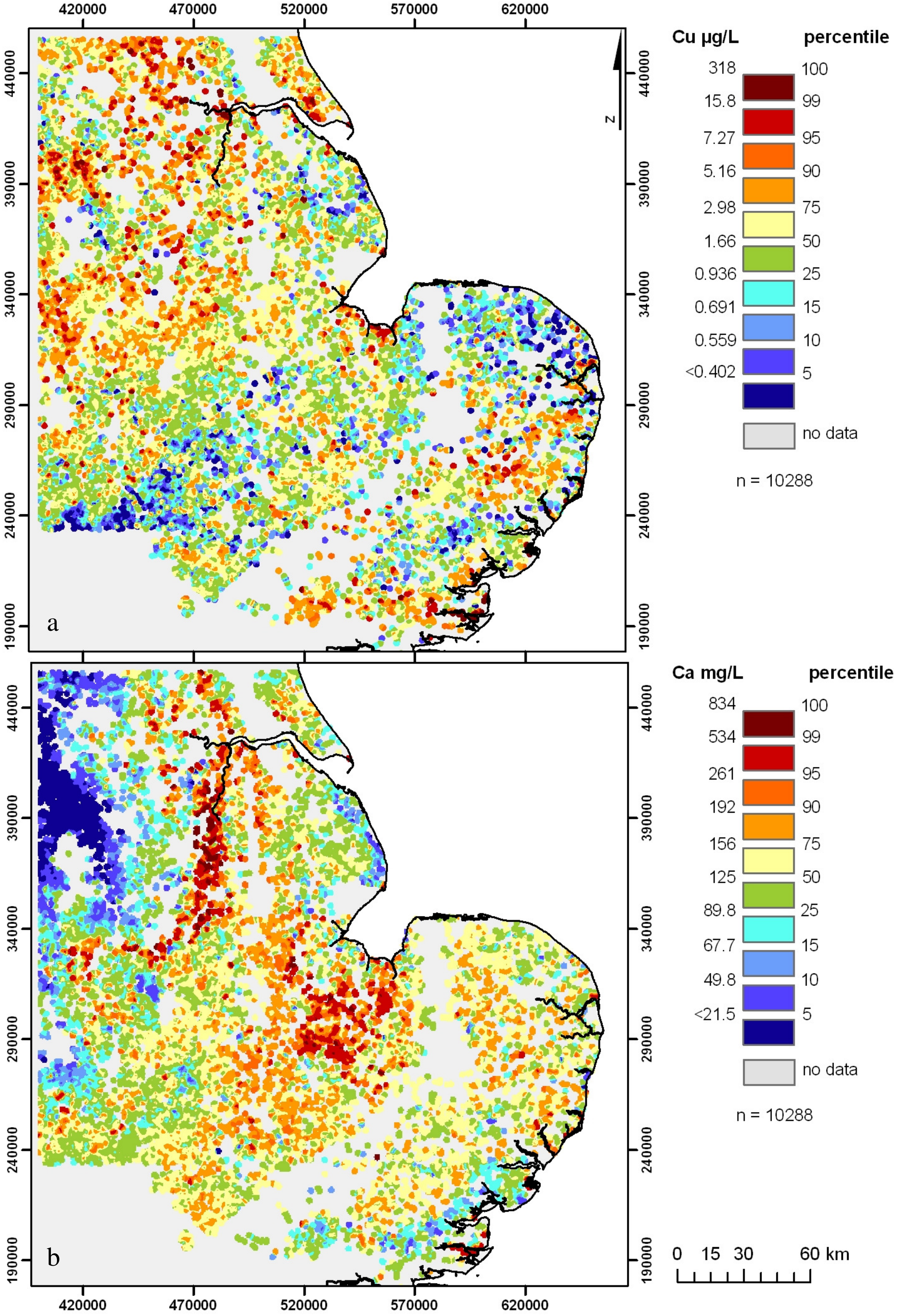
Temporal major ion data from stream waters in Wales were validated as part of the "Regional Geochemistry of Wales and Part of West-Central England Stream Water" atlas (British Geological Survey, 1999). The data reported there showed that there was little variation in the G-BASE monitor sites over time, although longer term data from a permanent monitoring site showed significant variation through the year. The variation was less in summer months and antecedent rainfall (totalled from the previous fortnight) played a major role in these variations. The summer months are therefore chosen to take G-BASE samples owing to the lower rainfall, and dominance of baseflow conditions.

This report describes the monitor site data sampled on field campaigns between 1997 and 2007 and spatial sample site data sampled on field campaigns between 1994 and 2007. This represents all the currently available monitor and spatial sample site data in the "Humber Trent", "Central and Eastern England", and "South East England" areas by which sampling campaigns were organised.

Copper was selected for this study to examine how temporal variations affect the spatial data set, and was chosen because the map produced for this area has a less clear concentration correlation to geology, topography, or landuse than most major and many trace elements. This gridded image (Figure 1.1) shows some systematic variations, but also much short-range spatial variability and it is also affected by variable sampling density. Copper is a potentially harmful element (Drever, 1997) that is present in these stream waters almost universally above the detection limit. Some systematic environmental controls on the concentration distribution can be seen. Copper has many applications in modern society and many of these uses also result in its dispersal in the environment (e.g. Martens et al., 2002; Maeda, 2003; Matkovic et al., 2004).

Figure 1.1 presents the Cu distribution compared to that of Ca. Calcium concentrations in stream water in Wales demonstrated that geological, topographical and landuse controls dominated the mapped data over any intra- or inter- annual variations (British Geological Survey, 1999). The Ca map is ideal to provide a comparison to Cu because it is a major element which, in natural waters, is very soluble and its presence is dominated by the availability of Ca-rich minerals and by solution and gas-phase equilibria involving CO_2 species. When calcium carbonate minerals are present in rocks and soils at a level of 1% or more, they will tend to dominate the aqueous chemistry, thus the regional Ca map presents a typical example of elemental data that correlates very well to the underlying geology (Hem, 1992; Langmuir, 1997; Faure, 1998). By comparison, Cu is a trace element which is most commonly sourced from sulphide mineralisation or anthropogenic inputs. The mobility of Cu is controlled by pH and sorption onto organic matter or Fe and Mn oxyhydroxides. There is thus greater scope for short-range spatial variability of Cu than Ca.

The aim of this report is to assess and understand the statistical distribution of the monitor site data. Where there are outliers in the monitor site data these are assessed to establish whether the corresponding spatial sample site data are affected, thus compromising the integrity of the geochemical stream water maps. In the context of regional mapping, single data points are not interpreted separately. The trends in the majority of the data are more important: where these spatial controls dominate, significance is not placed on individual data points, but on the systematic and regional variations seen in the data set as a whole.



Coastline derrived from OS topography © Crown Copyright. All rights reserved.

Figure 1.1 Gridded geochemical images of a) Cu and b) Ca in stream waters. The image was generated in ArcGIS 9.2 using an Inverse Distance Weighting (IDW) algorithm with a cell size of 250m and a fixed search radius of 1500m.

2 Stream water chemistry

Stream flow forms an integral part of the water cycle between areas of precipitation and evaporation, or discharge to other water-bodies. While the ultimate source of stream waters is precipitation, the water entering streams is derived from several sources:

- When it has not rained for some time the water in the stream consists of base flow, which is derived from the groundwater system.
- During and immediately after heavy rain the base flow is augmented by additions from soil water, overland flow and direct input from rainfall.
- When a stream rises, water is transferred into its banks; during drier weather this "bank storage" can flow back again (Drever, 1997).

A number of factors contribute to the chemical composition of stream water, which the rigorous G-BASE quality control procedures are designed to level (see Johnson, 2005). The daily sampling of the monitor site is designed to provide a measure for any changes encountered over the sampling season, and a measure of temporal variation, which can be compared to spatial variation.

2.1 PROCESSES CONTROLLING STREAM WATER CHEMISTRY

The "quasi static" processes described here are those which are predominantly controlled by spatially fixed variables. These are the processes which will dominate the mapped distribution of an element where temporal factors are not significant. In the absence of major landscape change (e.g. deforestation, changes in farming practices) these processes should not affect data collected over the period in which the G-BASE programme operates.

2.1.1 Bedrock and soil chemistry

The main influence on stream water chemistry is the geological distribution of soluble minerals in bedrock and overburden of the catchment above the sample site. The soil geochemistry refelects the underlying regolith, therefore the soil can represent drift or solid geological formations. In addition soil chemistry is modified over time, particularly by leaching. The effects of leaching vary according to drainage conditions of the soil and the rainfall regime (British Geological Survey, 1999).

Chemical weathering is one of the most important processes that controls the global hydrochemical cycling of elements (Stumm and Morgan, 1996). It affects minerals in the bedrock and soil. These minerals, many of which were formed at higher pressures and temperatures, are transformed into stable secondary minerals and solute species. Minerals either completely dissolve (congruent dissolution) or leave a residual mineral deposit (incongruent dissolution). This is facilitated by carbonic acid in the rain water (described in Section 2.1). For example $CaCO_3$ dissolves congruently to form Ca^{2+} and HCO_3^-. In contrast silicate minerals generally dissolve incongruently in natural waters and produce solute cations and silica, and residual clay minerals. Minerals have different weathering rates, meaning that solutes derived from the most reactive minerals (e.g. Ca from $CaCO_3$) are disproportionately abundant in stream water. Such minerals may dominate the chemistry, even if it comprises only a small proportion of the upstream geology (British Geological Survey, 1999).

2.1.2 Processes controlling chemistry of soil solution

The composition of soil solutions is predominantly controlled by ion exchange, anion adsorption, solution and hydrolysis of soil minerals, redox processes, and solution, deprotonation and degradation of organic acids. The ability of soils to neutralise acidic rainwaters is controlled by the carbonate and weatherable silicate content, the cation exchange capacity and base saturation (the proportion of exchange sites occupied by Ca^{2+}, Mg^{2+}, Na^+ and K^+). Soil organic matter and clay minerals tend to control processes in the soil, the latter becoming more important with depth (British Geological Survey, 1999).

2.1.3 Terrestrial organisms

Living matter plays an important role influencing the stream water chemistry. While mineral soils contribute solutes to the stream water (see Section 2.1.1), biota in organic portions of the profile can assist weathering by providing organic ligands, acids, and locally, CO_2 (Stumm and Morgan, 1996). Solutes are also extracted from pore water by plants supported in the soil profile. These tend to be nitrogen compounds and phosphates (Drever, 1997).

2.1.4 Occurrence of drift deposits

Drift deposits can be composed of material comprising of local or distant source materials, and can often be distinctly different from the underlying local geology. This may produce local stream water chemistry variations in areas where the bedrock geology is continuous and uniform, and the drift is discontinuous and of a contrasting composition. Locally derived drift deposits may produce variations if the re-working of the deposit gives rise to the production of weathering products via different processes. For instance different grain sizes mean that weathering rates will be different (British Geological Survey, 1999).

2.1.5 Mineral weathering and groundwater composition

In general a long aquifer residence time produces groundwaters closer to equilibrium with the surrounding rocks. It should be noted that preferential dissolution may occur where there are smaller grain sizes or where there is increased dissolution of mineral coatings or a soluble matrix. Baseflow of local groundwater can provide significant inputs to stream water flows, particularly during drier periods (British Geological Survey, 1999).

Groundwater often contains minor concentrations of potentially harmful elements. These can be indicative of local sources of contamination, but may also correlate to natural contamination sources such as natural pyrite dissolution (Shand et al., 2007).

2.1.6 Catchment hydrology

The boundaries of surface and groundwater catchments are often similar in small upland areas, but in large lowland basins a groundwater catchment area may cover multiple surface catchments. The hydrology of the catchment may greatly affect the stream water chemistry by mixing, either with “quickflow”, a rapid routing of rainfall, or a tributary of differing chemistry. Natural changes caused by the river flowing downstream are otherwise limited, as the residence time in a stream is generally short and there is little contact between the water and any reactive soil and rock surfaces (Drever, 1997; British Geological Survey, 1999). The G-BASE programme takes stream water samples from first and second order streams during the summer months. This ensures the rivers generally represent small catchments and baseflow conditions with a low quick flow component (British Geological Survey, 1999).

2.1.7 In-stream processes

There are a range of processes and controlling factors that are active within the stream channel, which influence stream water chemistry. For instance degassing of CO_2 can cause the stream’s H^+ concentration to reduce by up to 2 orders of magnitude. Stream particulate matter can play a vital role in element cycling, reacting with the water both organically and inorganically. Particulates, whether suspended or on the stream bed, are also involved in exchange reactions.

Organisms can play a role in in-stream processes, as vegetation takes up solutes and returns decomposing organic matter, and microbes can catalyze redox reactions (British Geological Survey, 1999). Upon collection samples are filtered to <0.45 µm to remove particulate and microbial matter. This process prevents mobilisation of ions from particulate exchange sites and microbial redox reactions (British Geological Survey, 1999).

2.2 FACTORS CAUSING SHORT-TERM FLUCTUATIONS

Short-term fluctuation factors are important because they could lead to short-term variations in the data, which could affect the legitimacy of the presentation of the data as a geochemical map.

2.2.1 Rainfall and atmospheric deposition

Rainwater is the ultimate source of most surface waters, and groundwaters, providing the initial input of solutes (Appelo and Postma, 2007). Rainwater chemistry is dominated by the influence of marine vapour, and resembles strongly diluted seawater containing carbonic acid and a sea salt aerosol with a mixture of HCl, HNO_3, H_2SO_4, and usually some NH_4 (Kinniburgh and Edmunds, 1986; Drever, 1997; Appelo and Postma, 2007).

The composition of rainwater is determined by the source of the water vapour and the ions that are acquired or lost in the atmosphere. Rainwater is naturally acidic owing to the dissolution of atmospheric CO_2, which forms a weak solution of carbonic acid with an equilibrium pH of around 5.7. Areas closer to industrial atmospheric sources are likely to have rainfall with higher concentrations of S and N species and to have a low pH. The pH of groundwater recharge may be even lower than that measured in the rainfall owing to the oxidation of NH_4^+ ions to NO_3 which results in the release of H^+ ions and concentration by evapotranspiration (Smedley and Allen, 2004; Shand et al., 2007). In addition to dissolution of airborne particles, dry deposition and throughfall are thought to contribute to a three times enrichment of dissolved salts in rainwater. Even if it is deposited directly into the stream, the rainwater can contribute a variety of salts to the overall stream water chemistry (Stumm and Morgan, 1996; Drever, 1997).

The quantity and frequency of rainfall can influence the stream water chemistry by affecting the input of salts as described above, or diluting the stream water. During dry periods soluble salts from the substrate can become more concentrated, subsequent heavy rainfall leaches these creating a higher stream water concentration.

2.2.2 Anthropogenic influences

Anthropogenic activities can have a significant impact upon the chemistry of stream water. These include activities that introduce air pollution and alter rainfall chemistry (as described in Section 2.2.1) as well as direct inputs from diffuse or point sources. Human activities can accelerate natural processes. For instance, the oxidative weathering of sulphide minerals, that can produce dissolved potential harmful elements and acidity, is completely natural. However, mining breaks up large quantities of rock containing such minerals, and the problem is exacerbated as the reactive surface area has been increased (Faure, 1998). In areas of agriculture, fertilisers added to increase plant productivity may leach into stream waters. These contribute inputs of N, P, and K in particular, along with other major and trace elements (such as C, Cd, Cl, S, and U) (British Geological Survey, 1999).

2.3 CONTROLS ON COPPER HYDROCHEMISTRY

In order to interpret the spatial Cu distribution, it is critical to understand the controls on Cu occurrence and solubility. In nature, Cu is found as either the native element or in sulphide deposits (Shand et al., 2007). Copper can occur in solution as either Cu^{2+} or Cu^+, although Cu^{2+} tends to dominate in oxygenated water owing to favourable redox conditions and the tendency of

Cu^+ to disproportionate ($2Cu^+ \rightarrow Cu^0 + Cu^{2+}$) (Hem, 1992). Divalent Cu can be free or complexed. Where the solution is oxidising, which is typical of most natural stream waters, Cu is most soluble under acidic conditions. As the pH increases Cu can become adsorbed to organic matter or Fe and Mn oxyhydroxides. Sorption of Cu^{2+} to Fe and Mn oxyhydroxides occurs over a typical range of around 2 pH units: at pH 4 adsorption of Cu is minimal, but at pH 6 Cu^{2+} sorption is essentially complete. At pH values higher than this the Cu concentration is controlled by the solubility of carbonate or oxide/hydroxide species (Drever, 1997).

Stumm and Morgan (1996) presented a speciation diagram showing the behaviour of 3 µg/L Cu(II) in a carbonate bearing water. They showed the pH control on Cu species, where Cu^{2+} tends to dominate below pH 6 and $CuCO_3^0$ should dominate between pH 6 and 10. Figure 2.1a shows the relationship between the Cu concentration and pH for the majority of the spatial sample site data, values >100 µg/L are not presented. Two distinct peaks can be seen around pH values of 4 and 7.5. These correspond to the dominance of the Cu^{2+} and $CuCO_3^0$ species defined by Stumm and Morgan (1996), and the HCO_3 peak seen in the spatial sample site data (Figure 2.1b). However the largest Cu concentrations correspond to lower HCO_3 concentrations (Figure 2.1c). The majority of the Cu sample site data behave as predicted by Stumm and Morgan (1996) and Drever (1997).

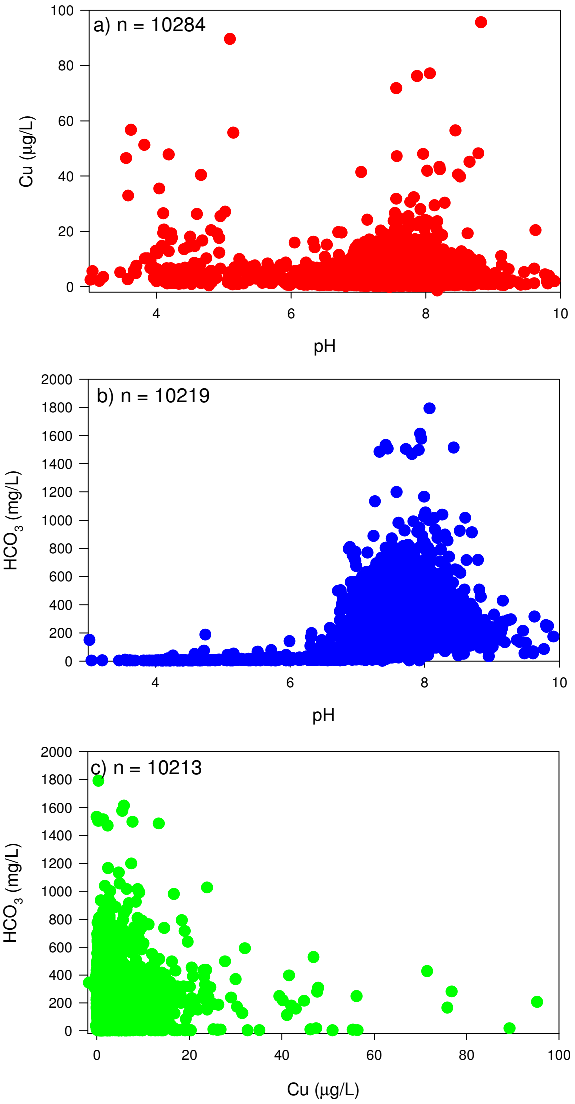


Figure 2.1 Relationship between a) pH and Cu concentration, b) pH and HCO_3 concentration, and c) Cu and HCO_3 concentrations

3 Study area

Between 1994 and 2007, an area of approximately 43,000 km^2, covering central and eastern England, was sampled by G-BASE teams. There was no regional G-BASE sampling campaign during the summer of 2001 because of the nationwide problems of Foot and Mouth Disease. During the eastern England field seasons of 1997 to 2007, 42 monitor sites were set up. Of these, 35 were sampled over a period of 10 days or more. In three cases the same location was used in subsequent field seasons, and therefore these provide temporal data spanning more than one field season.

Monitor sites were named in the field, predetermined according to whether there are two teams ("A" and "B") or by pre allocation of site names as described by Johnson (2005) in the G-BASE field procedures manual (page 11-7). This means that the original monitor site names are in different formats, and, as some site names do not incorporate the year, the same name is often replicated. The original monitor site details are permanently held in the G-BASE field chemistry books and all available details are published in the QC reports. For the purposes of this report, however, a unique site identifier has been used. These site names take the format of "YEAR_SITE ID". The site ID begins with M to indicate it is a monitor site. This is followed by "A" or "B" if there were two teams working in the field simultaneously. The site ID ends with a number indicating the order in which the monitor sites were used in each year. For example the third site to be sampled by team B in 1998 is called "1998 MB3". Where the sites were sampled over more than one year, and in this report are considered as one site, they were termed MS 1,2, and 3. Whilst these data have been plotted on an annual basis as part of the routine project QC

Topography © Crown Copyright. All rights reserved

Figure 3.1: Location of monitor sites. Labels marked in blue indicate sites used on more than one year to within 40 m. Where the distance between sites is greater than 40 m they are considered separate.

procedures, they have not previously been studied systematically by comparing an analyte across these years of sampling.

Figure 3.1 shows the location of monitor sites in the study area, and Figure 3.2 shows this in context to the sampling sites (shown in blue). Labels marked in red indicate sites used on more than one year, which lie within 40m of each other. Where the distance between sites is greater than 40 m they are considered separate. At one site, 2004 M1, the sample site was moved 200 m upstream after two days of sampling owing to a contamination concern. There was little difference in Cu concentration in these first two samples and those taken 200 m upstream, so they are considered as a single site. The location details of each monitor site are summarised in Table 3.1. From 2003 additional site information (such as drainage type, drainage condition and landuse) was collected in order to be consistent with the spatial G-BASE samples. Additionally the monitor site samples were assigned a sample site number and integrated into the normal spatial sample analysis procedures. This was to ensure that their analysis was consistent with the field samples with which they were collected (Ander, 2009). Table 3.2 presents details of the

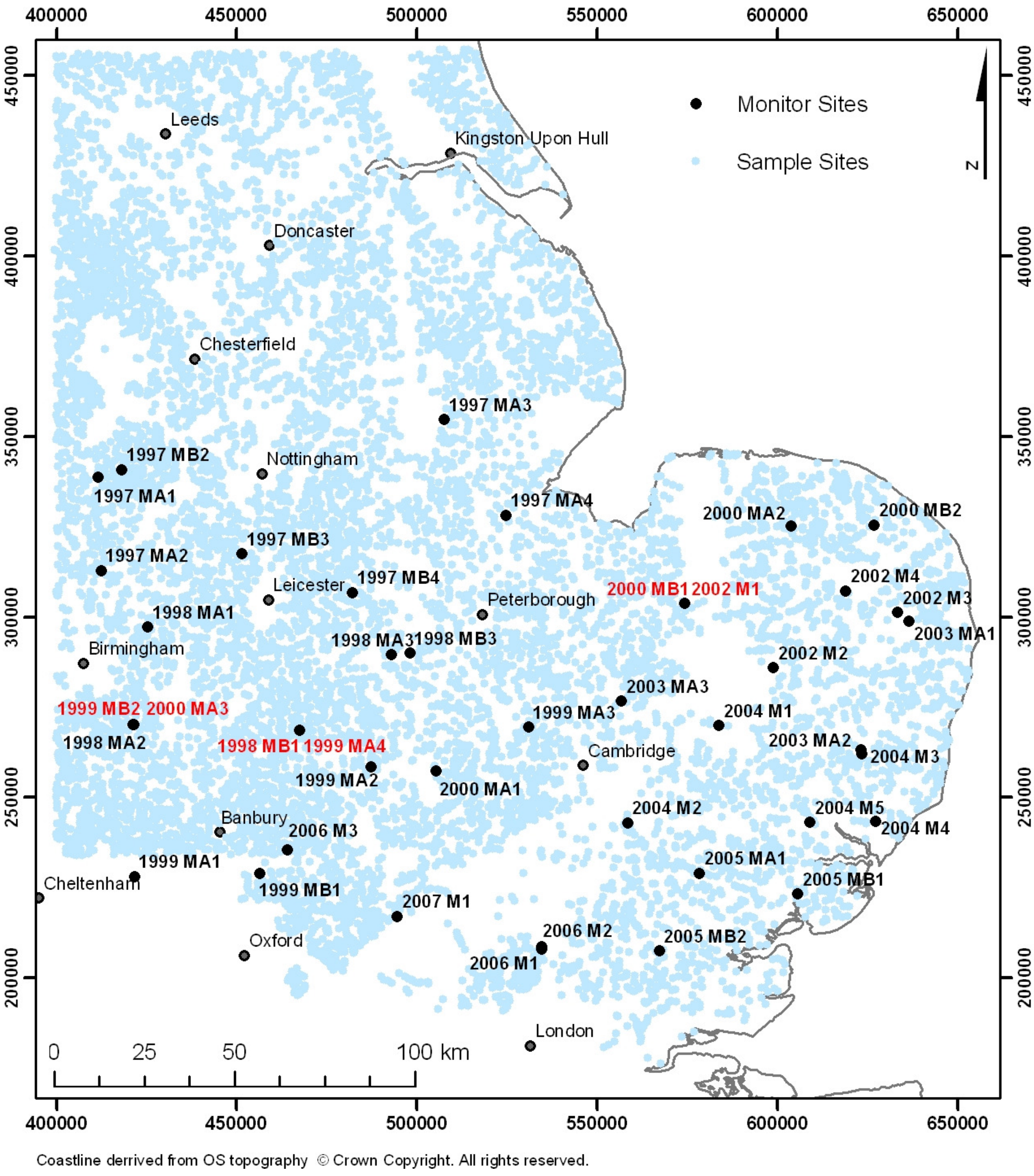


Figure 3.2 Location of spatial sample sites (blue circles) and monitor sites (black circles). Labels marked in red indicate sites used on more than one year

monitor sites which were sampled on two field campaigns. These provide more samples at each site, but they span a period of over a year, showing any variation between years. Only the sites with 10 or more samples taken will be used in this study, and where sites were sampled over two field campaigns, these data are combined.

Table 3.1 Location and sampling dates of the monitor sites of central and eastern England and East Anglia.

Site	Easting	Northing	Sample Count	Start	End
1997 MA1	411670	338590	11	17/06/1997	27/06/1997
1997 MA2	412430	312730	12	30/06/1997	11/07/1997
1997 MA3	507680	354580	19	14/07/1997	01/08/1997
1997 MA4	524720	328090	16	03/08/1997	19/08/1997
1997 MB2	418180	340680	13	29/06/1997	11/07/1997
1997 MB3	451570	317400	19	14/07/1997	01/08/1997
1997 MB4	482210	306750	15	04/08/1997	19/08/1997
1998 MA1	425310	297090	17	06/07/1998	31/07/1998
1998 MA2	421450	270200	5	13/07/1998	17/07/1998
1998 MA3	492920	289490	39	03/08/1998	10/09/1998
1998 MB1	467380	268620	25	22/06/1998	17/07/1998
1998 MB3	498060	289900	33	02/08/1998	10/09/1998
1999 MA1	421690	228020	18	15/06/1999	02/07/1999
1999 MA2	487380	258400	12	05/07/1999	16/07/1999
1999 MA3	530950	269450	17	19/07/1999	06/08/1999
1999 MA4	467400	268650	25	09/08/1999	02/09/1999
1999 MB1	456510	228850	12	21/06/1999	02/07/1999
1999 MB2	421310	270110	25	05/07/1999	29/07/1999
2000 MA1	505253	257240	19	03/07/2000	21/07/2000
2000 MA2	603730	325030	26	24/07/2000	18/08/2000
2000 MA3	421310	270110	12	20/08/2000	31/08/2000
2000 MB1	574330	303700	19	10/07/2000	28/07/2000
2000 MB2	626820	325370	18	31/07/2000	17/08/2000
2002 M1	574330	303700	18	24/06/2002	11/07/2002
2002 M2	598820	285870	19	15/07/2002	02/08/2002
2002 M3	633410	301210	9	05/08/2002	22/08/2002
2002 M4	618970	307160	4	12/08/2002	15/08/2002
2003 MA1	636500	298700	19	14/07/2003	01/08/2003
2003 MA2	623283	262151	19	04/08/2003	22/08/2003
2003 MA3	556800	276600	16	25/08/2003	10/09/2003
2004 M1	583749	269913	2	28/06/2004	29/06/2004
2004 M1a	583907	269932	17	30/06/2004	16/07/2004
2004 M2	558415	242972	19	19/07/2004	06/08/2004
2004 M3	623085	263071	13	07/08/2004	19/08/2004
2004 M4	627296	243353	14	23/08/2004	02/09/2004
2004 M5	608947	243063	6	06/09/2004	11/09/2004
2005 MA1	578436	229017	18	27/06/2005	15/07/2005
2005 MB1	605552	223187	17	27/06/2005	15/07/2005
2005 MB2	567198	207419	10	18/07/2005	27/07/2005
2006 M1	534644	208081	7	26/06/2006	02/07/2006
2006 M2	534735	208615	9	03/07/2006	13/07/2006
2006 M3	464197	235445	18	17/07/2006	03/08/2006
2007 M1	494497	217004	38	02/07/2007	09/08/2007

Table 3.2 Location and sampling dates of the monitor sites sampled over more then one year

Sites	Unique site name	Easting	Northing	Sample Count	Start	End
1998 MB1 and 1999 MA4	MS 1	467380	268620	50	22/06/1998	02/09/1999
1999 MB2 and 2000 MA3	MS 2	421310	270110	37	05/07/1999	31/08/2000
2000 MB1 and 2002 M1	MS3	574330	303700	37	10/07/2000	11/07/2002

4 Data analysis

4.1 DATA DISTRIBUTION

The G-BASE spatial sample site data are routinely presented according to standard percentile values (Johnson et al., 2005). This is a non parametric method of presenting a summary of the data spread, and the percentiles chosen emphasise high and low values of the data set, with the central 50% of the data in only two classes around the median. The spatial G-BASE data are usually not normally distributed, which is typical of data that represent unique sample sites, and the reason for using non-parametric methods. Figure 4.1 demonstrates the spatial Cu data used in this report fitted to a modelled normal distribution. Although most of the data conforms to the modelled normal distribution, this breaks down at high and low values, which supports the use of non parametric statistics to investigate the spatial sample site data. It is also intuitive that the data shown in Figure 1.1a represents more than one population.

The monitor site data represent multiple samples of the same site, which should vary around a mean value if there are no external factors affecting the site. Therefore it would be expected that these data have a normal distribution within sites. However in order to approach the data consistently, the monitor site data will be treated the same as the spatial sample site data, with the exception that the mean standard deviation of all monitor sites will be used to compare the spread of the concentration data at individual monitor sites. Sample and monitor site data are presented on box and whisker plots, which simplify the data and provide a useful method of displaying data spread.

When presented in the same box and whisker plot the concentration within each monitor site and how this compares between the monitor sites and sample population can be seen clearly. Box plots were created using Minitab, where the default components of a box plot follow the

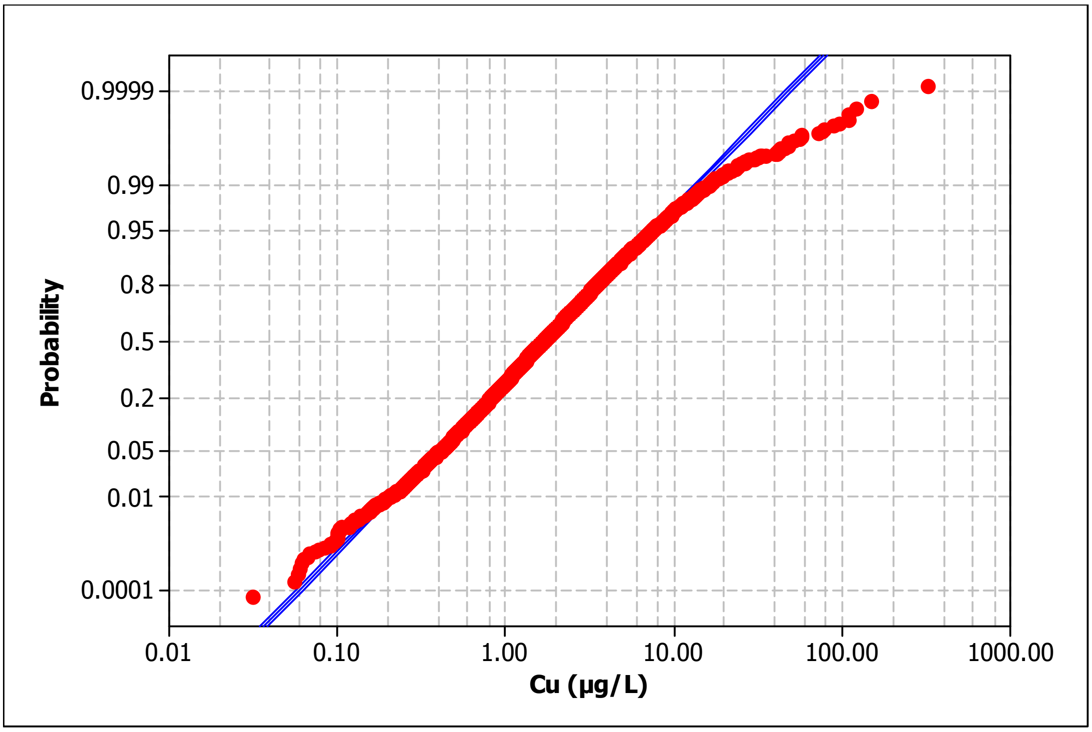


Figure 4.1 Probability plot of all spatial sample site data (red points) plotted on a modelled normal distribution (blue line).

protocols described by Tukey (1977). Figure 4.2 shows how data are presented in a box plot. The boxes are defined as the interquartile range (Q3-Q1), the intermediate horizontal black line depicts the median. The upper limit of the whisker extends to the highest point within 1.5 times the interquartile range from the top of the box, and the lower limit of the whisker extends to the lowest point within 1.5 times the interquartile range from the bottom of the box. The asterisks represent outliers, which exist beyond the upper or lower limits of the whiskers. The outliers may identify data points in the monitor sites which have been affected by external processes and potentially present a useful method to establish which data may require further investigation. Complementary data such as pH, HCO_3 or conductivity may offer a method of identifying if the sample has suffered contamination after it has been removed from the stream. This may occur by mishandling by the sampler, or during analysis. Outliers may be of limited use when describing the spatial sample site data because any data set of such a great size (n=10288) will have outliers. The relative significance of outliers in relation to number of sites and type of data is critical.

The data are compared to assess the spread of Cu data within sites and between sites in the sample area. The spread of data is assessed using descriptive statistics. To assess the relative significance of the spread of data between monitor sites the actual G-BASE regional sample data are used. This gives the full range of data and puts the monitor sites in context, i.e. for the spatial distribution to be meaningful the concentration variation between spatial sample sites should be greater than that within sites described at each of the monitor sites.

4.2 SPATIAL DATA

Individual spatial sample sites are examined further using the time series data collected from each monitor site. Where outliers occur in the monitor site data the spatial sample sites taken on this day are examined in relation to the surrounding sample sites using G-BASE colour maps to establish if there are corresponding effects in the mapped data. The maps produced by G-BASE present colour classified images based on class boundaries set at the 5, 10, 15, 25, 50, 75, 90, 95, and 99 percentile levels (British Geological Survey, 1999). The gridded images produced show regional trends and local variations and it is these, rather than specific individual data points that are the focus of the high density regional data.

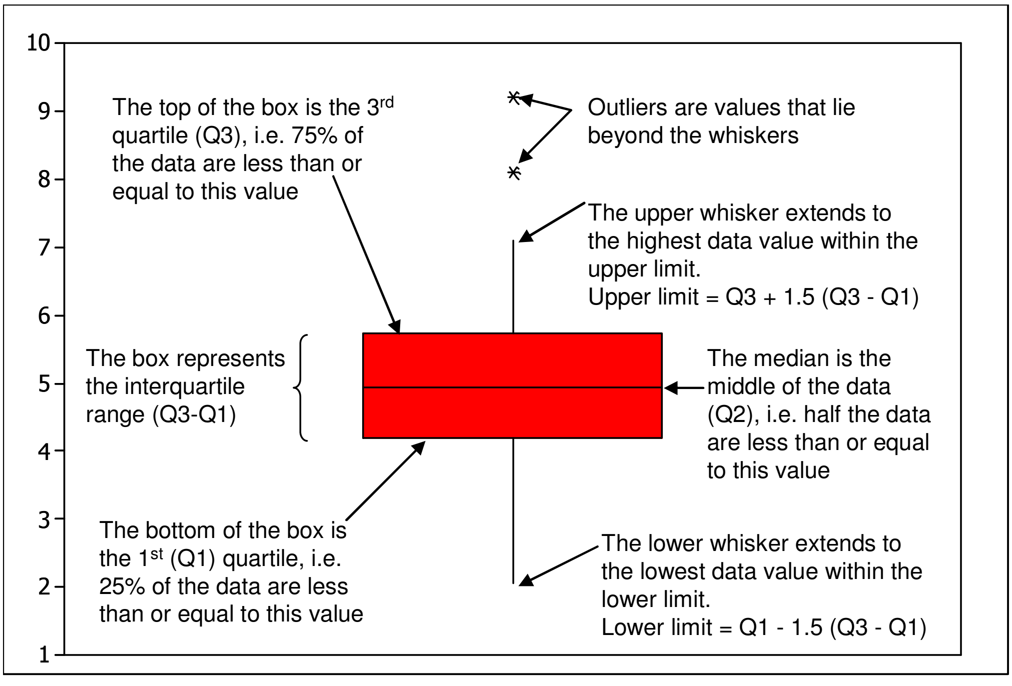


Figure 4.2 Example box plot showing data presentation

5 Temporal and spatial variations

5.1 REGIONAL AND MONITOR SITE DATA DISTRIBUTION

Table 5.1 presents summary statistics for Cu in samples from the 33 monitor sites where n≥10, and the spatial sample sites. Percentiles are used to demonstrate the spread of the Cu concentration data. The standard deviation of the monitor sites is generally <1 µg/L, although 6 monitor sites have a standard deviation >2 µg/L. The highest standard deviations correspond to sites which have a large outlier. The mean standard deviation of the monitor sites is 1.3 µg/L. This figure is a good representation of expected variation within each site. The standard deviation of the samples is 5.39 µg/L, this is greater than the majority of the monitor sites' standard deviations, but is exceeded by 1997 MA3 and 2005 MB2, which each have one large outlier.

The median and mean of the data for each monitor site are generally very similar indicating a near normal distribution. Where there are large outliers, however, this offsets the mean from the median indicating a non normal distribution. The mean of these data is always greater than the median, indicating that the skew is always towards the higher concentrations.

Figure 5.1 presents a summary of the selected Cu monitor site and all the spatial sample site data. The outliers are not shown; meaning the spread of the majority of the data can be more clearly compared, because the y-axis is hence reduced in range. The horizontal black lines represent the sample site data quartiles to allow comparison of the monitor site data to the spatial sample data distribution. The dotted red line represents the detection limit of Cu (0.25 µg/L). It can be seen that this is generally exceeded; only five monitor site and 165 spatial sample site measurements are below the detection limit. While data below the detection limit is not verified, artificially setting the lower limit of the whisker at 0.25 µg/L gives a false impression of the data distribution, and values below the detection limit are retained, but treated with caution. The exception to this is where the reported value is <0, which gives an artificially large range. Six spatial sample site values were reported to be <0. These were removed meaning the lower limit of the whisker was set to the lowest positive value.

The concentration of Cu within all the monitor site samples is generally low (<5 µg/L), although values range up to a maximum value of 28 µg/L. It can be seen that the spread and the median of data vary between monitor sites providing a representative range of concentration data with respect to the regional sample data. The spread of the spatial sample site data between the upper and lower limits of the whiskers is greater than the majority of that of the monitor sites temporal variation. Three monitor sites, however, have a greater spread between the upper and lower whiskers: 1997 MA1; 1997 MB2; and 1998 MA3. All these sites have a high median concentration (>4.5 µg/L) which puts these sites in the upper 25% of the spatial sample site data. Many of the data from these monitor sites extend beyond the upper limit of the whisker of the spatial sample sites box plot, indicating that they represent a relatively small proportion of the sample site data. Monitor sites 1998 MB3 and 2005 MA1, however, span the entire interquartile range of the spatial sample site data, and thus represent a more significant spread in terms of number of mapped data points over the duration of their collection.

It should be noted, however that the box and whisker plots are constructed from a set of rules that consider percentiles of data. It therefore occurs that where a data set contains so many independent values (in this case n = 10288) many will be greater than the extent of the upper limit (see section 4), defining these data points as outliers, despite there being no obvious distinction from the main data set (see Figure 4.1).

Figure 5.2 and Figure 5.3 present the same data as Figure 5.1, with the outliers included, showing the actual spread of all the data. The spatial sample site data are removed from Figure 5.2 to show the monitor site data range more clearly. Figure 5.2 and Figure 5.3 demonstrate how the monitor site outliers are few, but in most cases are significantly distinct from the majority of the data for each monitor site. The spatial sample site data, however, have many outliers, which form a continuum from the top of the upper limit (6 µg/L) to 35 µg/L Cu. Out of the 10288 sample sites, 741 are between 6 and 35 µg/L, and there are 27 sites which exceed 35 µg/L. Most of the spatial sample site outliers are thus outliers by statistical definition alone because of the large number of samples.

The majority of the spatial sample site data show a greater spread than the majority of the temporal monitor site data, which to some extent confirms the validity of the mapped data. However the spatial variation needs to be examined to establish if the spatial sample site data were affected on days where there were monitor site outliers and whether this needs to be accounted for when using and presenting the regional spatial sample data.

Site	Mean	σ	5	10	15	25	50 (Median)	75	90	95	99
1997 MA1	5.36	2.41	2.92	2.93	3.19	3.62	4.59	7.02	8.59	9.09	9.48
1997 MA2	5.74	0.83	4.82	4.90	5.01	5.13	5.89	5.98	6.12	6.91	7.68
1997 MA3	3.18	6.09	1.42	1.49	1.51	1.57	1.66	1.87	2.41	6.88	23.41
1997 MA4	1.40	0.36	0.69	0.81	0.99	1.36	1.49	1.58	1.68	1.75	1.89
1997 MB2	5.66	2.03	3.48	3.62	3.80	3.99	4.80	7.13	8.29	8.88	9.39
1997 MB3	2.55	0.39	2.23	2.25	2.27	2.35	2.43	2.64	2.84	2.96	3.74
1997 MB4	3.47	0.65	2.73	2.78	2.81	3.01	3.33	3.81	4.35	4.62	4.76
1998 MA1	1.08	0.51	0.83	0.85	0.86	0.89	0.92	1.04	1.26	1.71	2.71
1998 MA3	6.47	3.14	3.61	3.88	4.24	4.58	5.38	6.83	10.94	14.35	15.41
1998 MB3	1.19	0.95	0.49	0.52	0.53	0.54	0.65	1.50	2.83	3.08	3.29
1999 MA1	0.40	0.14	0.30	0.30	0.31	0.32	0.36	0.41	0.49	0.59	0.82
1999 MA2	1.99	0.23	1.75	1.79	1.84	1.90	1.95	2.07	2.09	2.33	2.57
1999 MA3	1.15	0.13	1.01	1.06	1.08	1.09	1.13	1.19	1.28	1.37	1.48
1999 MB1	4.01	0.86	3.10	3.23	3.25	3.38	3.97	4.36	4.77	5.38	5.93
2000 MA1	2.19	0.33	1.82	1.85	1.86	1.91	2.16	2.39	2.61	2.77	2.84
2000 MA2	1.73	3.49	0.83	0.86	0.87	0.89	0.95	1.18	1.50	1.64	14.52
2000 MB2	0.67	0.87	0.25	0.27	0.28	0.30	0.45	0.57	0.86	1.61	3.55
2002 M2	1.27	1.11	0.58	0.72	0.75	0.79	0.88	1.03	2.06	3.95	4.59
2003 MA1	0.95	0.54	0.52	0.55	0.57	0.59	0.68	1.22	1.65	1.85	2.32
2003 MA2	0.71	0.65	0.23	0.23	0.25	0.28	0.35	1.03	1.88	1.93	1.95
2003 MA3	1.71	0.63	1.23	1.33	1.40	1.45	1.56	1.67	2.04	2.63	3.63
2004 M1	0.42	0.11	0.25	0.30	0.32	0.36	0.42	0.45	0.55	0.58	0.67
2004 M2	1.98	0.60	1.19	1.19	1.21	1.27	2.21	2.42	2.63	2.64	2.75
2004 M3	3.05	1.96	1.43	1.47	1.56	1.72	2.62	3.40	4.92	6.40	7.98
2004 M4	0.46	0.06	0.38	0.39	0.39	0.39	0.46	0.49	0.53	0.56	0.59
2005 MA1	2.92	1.41	1.15	1.40	1.57	1.91	2.80	3.68	4.48	5.73	5.86
2005 MB1	1.82	0.81	1.28	1.40	1.43	1.49	1.67	1.80	2.05	2.74	4.38
2005 MB2	4.03	7.76	1.24	1.26	1.30	1.39	1.46	1.97	4.65	15.37	23.94
2006 M3	1.46	1.63	0.85	0.90	0.92	0.93	0.99	1.10	1.58	3.33	6.91
2007 M1	1.13	0.64	0.82	0.86	0.87	0.93	1.01	1.07	1.19	1.60	3.68
MS1	2.14	1.13	1.30	1.38	1.48	1.64	1.81	2.29	2.83	3.50	6.82
MS2	1.11	0.31	0.83	0.86	0.87	0.94	1.02	1.11	1.63	1.82	2.03
MS3	0.93	0.16	0.78	0.79	0.80	0.83	0.88	1.00	1.10	1.16	1.43
Samples	2.65	5.39	0.40	0.56	0.69	0.94	1.66	2.98	5.16	7.27	15.78

Table 5.1 Summary statistics of monitor sites and samples (Cu µg/L). 5,10,15,25,50,75,90,96, and 99 are percentiles.σ = standard deviation

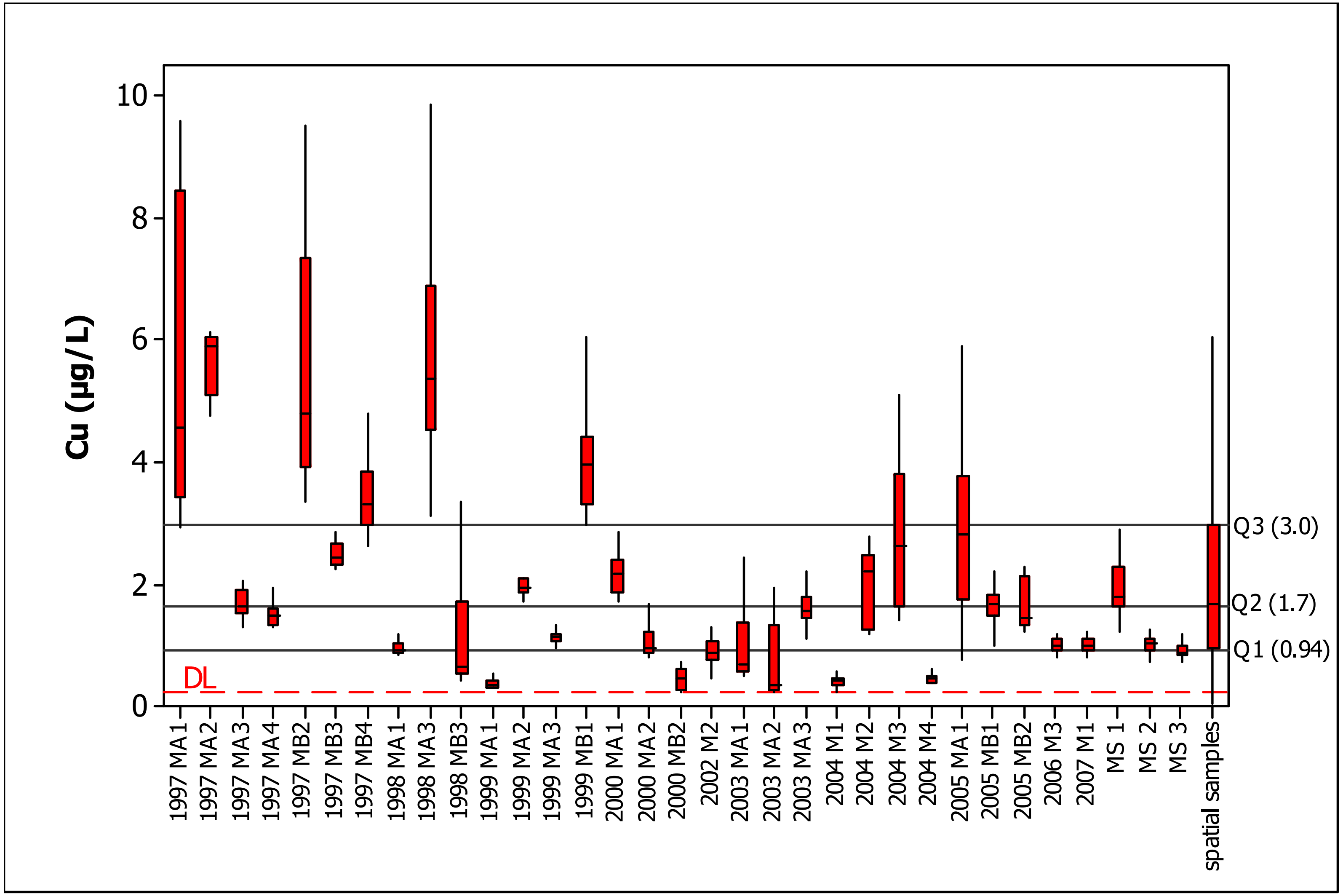


Figure 5.1 Box and whisker plot showing variation of Cu concentration in monitor sites (where n≥10) and samples. The interquartile ranges presented on the graph refer to the regional spatial sample site data. DL = detection limit (0.25 µg/L)

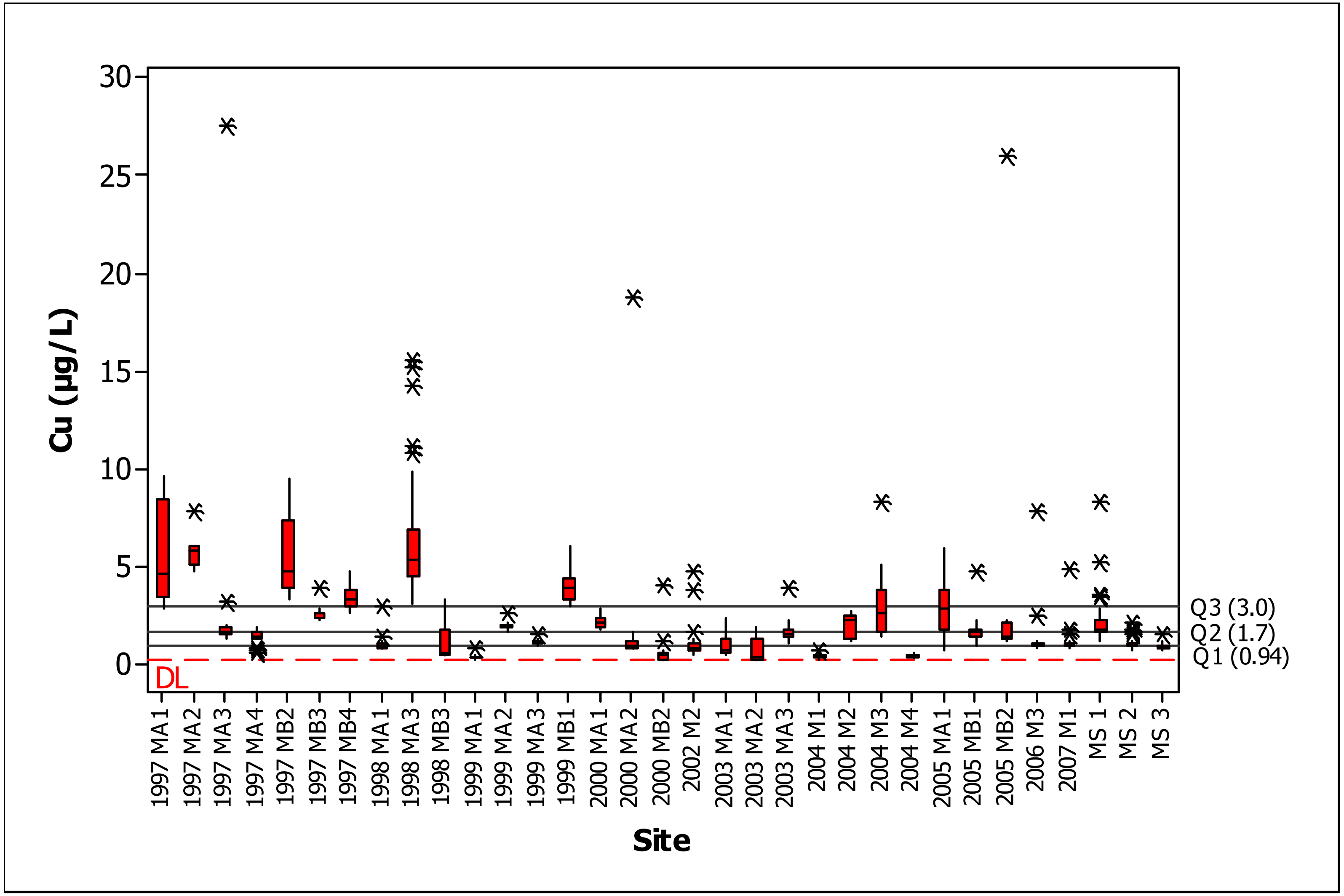


Figure 5.2 Box and whisker plot, including outliers, showing variation of Cu concentration in monitor sites (where n≥10). The interquartile ranges presented on the graph refer to the spatial sample site data. DL = detection limit (0.25 µg/L)

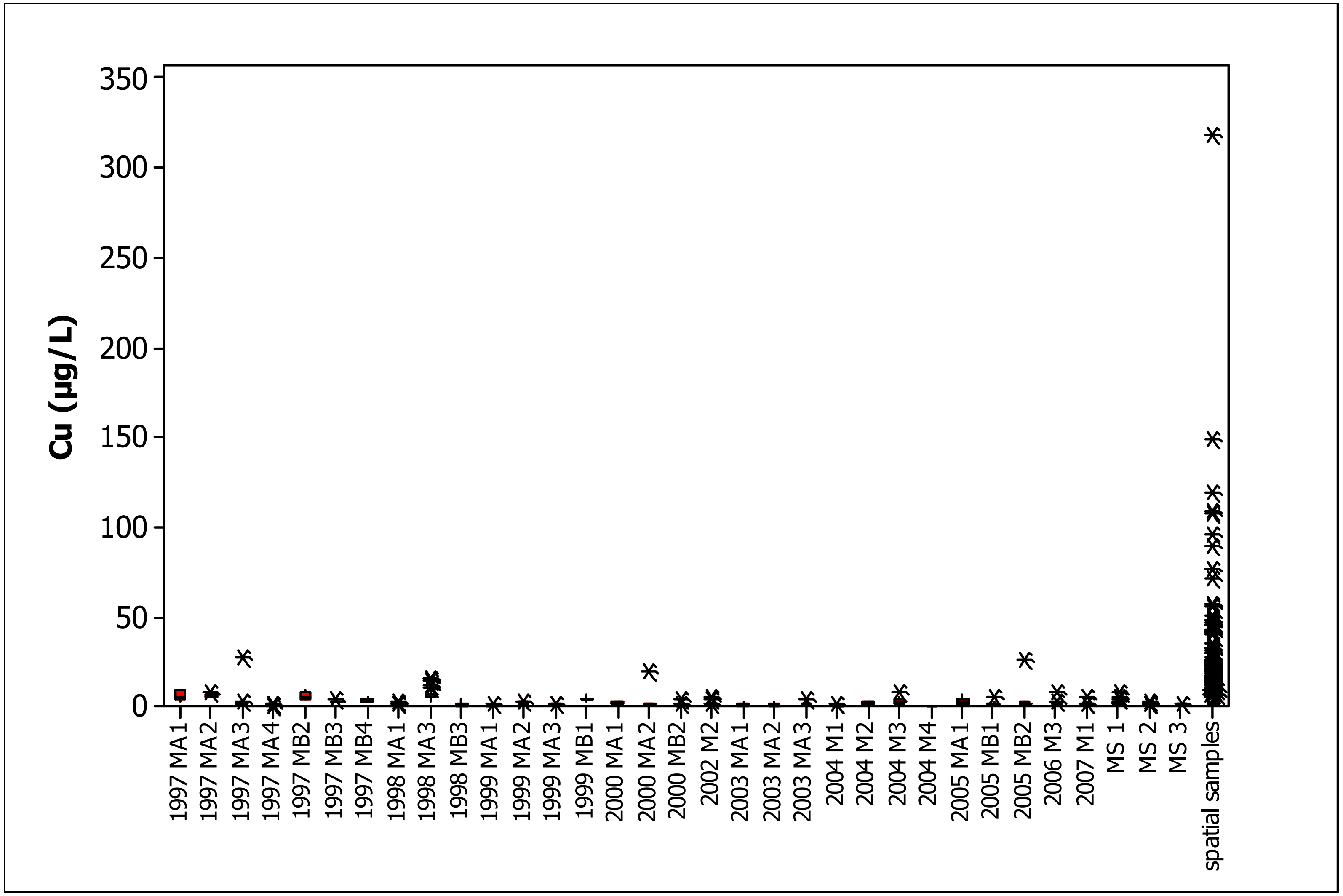


Figure 5.3 Box and whisker plot, including outliers, showing variation of Cu concentration in monitor sites (where n≥10) and spatial sample sites. The scale differs from Figure 5.2 to display all the spatial sample site outliers.

5.2 MONITOR SITE DATA OUTLIERS

Outliers have been defined in Section 4 as any data falling outside the limits of Q3 + 1.5×(Q3-Q1) and Q1 – 1.5×(Q3-Q1) and this section concentrates on those above the main body of the data. This however may mean that data points that are not necessarily distinct from the main data set are defined as outliers. These can be biased by the number of data points within the data set and the numerical spread of those data. Where there are more samples in a data set there is greater confidence that outliers are significant.

Nine of the 35 (n ≥10) monitor sites have been used for further interpretation. The rationale for these choices is presented in Table 5.2. These are presented as time-series graphs to demonstrate the behaviour of Cu concentrations at each monitor site over time. Where appropriate the y-axis scale is uniform for easy comparison between sites. Where outliers exist in the monitor site data the spatial sample site data are examined to establish whether a process can be recognised that may have caused this. It is common for sampling to be taking place at great distances (>50 km) from the monitor site, so different weather conditions and drainage regimes may prevail between the monitor site and corresponding spatial sample sites. The time series data are discussed with reference to relevant weather events recorded by the field team and and local baseflow index (BFI) conditions.

Marsh and Hannaford (2008) computed BFIs from the archive record of daily mean river flows from a network of gauging stations around the UK. The BFI can be used as an indicator of the proportion of a river's run off that is derived from stored (groundwater) sources. The BFI increases with permeability of the bedrock, superficial deposits and soils within the catchment. A river with a high BFI is therefore dominated by baseflow, which will sustain the river flow during drier periods (Marsh and Hannaford, 2008). In this report the BFI which best represents the site in terms of proximity of gauging station and underlying geology is reported. It is the closest match derived from Marsh and Hannaford (2008), so it can only present an indication of the likely flow regime at each monitor site. The location of each site associated with monitor site outliers is overlaid onto the G-BASE regional stream water copper map to see if these correspond to anomalies on the map.

5.2.1 2004 M4

Figure 5.4 presents the time series data for 2004 M4, which represents an example of a site which has a small spread of data. This is typical behaviour of most of the Cu concentrations at each monitor site (see Figure 5.1). This site overlies gravel, sand, and clay, with glacial sand and gravel. These sediments are generally well drained and would provide a steady input of any rainfall events. The most representative BFI for this site is 0.92 (Marsh and Hannaford, 2008) indicating that groundwater is expected to comprise the majority of the stream flow. This is reflected in the monitor site data which show no fluctuations in concentration despite three prolonged periods of heavy rain over the duration of sampling at this site.

Table 5.2 rationale for monitor site discussion

Site	Rationale
2004 M4	Typical of monitor site data
MS3	Typical example of monitor site data over multiple field seasons
1997 MA1	Second largest range between upper and lower limits
1997 MA3	Small range but with significant outliers
1997 MB2	Third largest range between upper and lower limits
1998 MA3	Largest range, many outliers
1998 MB3	Spans Q1-Q3, same time scale as 1998 MA3, but with a smaller range and no outliers
2005 MA1	Spans the interquartile range, no notable rainfall events
MS1	Monitor site over multiple field seasons with outliers

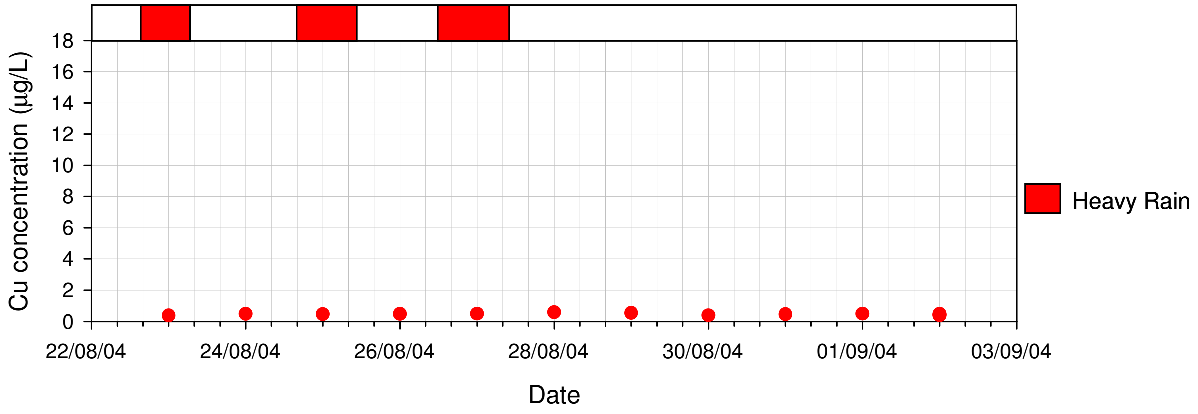


Figure 5.4 Time series data for 2004 M4

5.2.2 MS 3 (2000 MB1 and 2002 M1)

Figure 5.5 shows the time series data for MS3, comprising 2000 MB1 and 2002 M1. This demonstrates the temporal variations spanning three field seasons. The site overlies chalk and the most representative BFI for this site is 0.88 (Marsh and Hannaford, 2008) indicating the importance of the groundwater inputs. There are minor fluctuations in the data which are not affected by heavy rainfall inputs. There are no outliers and the 2000 MB1 data are comparable with the 2002 M1 data, with mean concentrations of 0.88 µg/L Cu, and 0.97 µg/L Cu respectively. Most of the data lie between the 15th and 25th percentile of the spatial data set, with the remainder ranging no higher than the 50th percentile. Even with a sampling gap of over two years the difference in concentration is minimal and the data do not vary by more than one G-BASE percentile class. This emphasises that collecting samples during predominantly baseflow conditions provides comparable data between different field seasons, despite any changes in flow regime throughout the preceding months. If any sample from this monitor site were used as spatial sample site data there would be no significant difference on the map regardless of which data point was selected.

The remaining time series graphs represent the minority of sites where outliers or significant variations in the monitor site data set were evident.

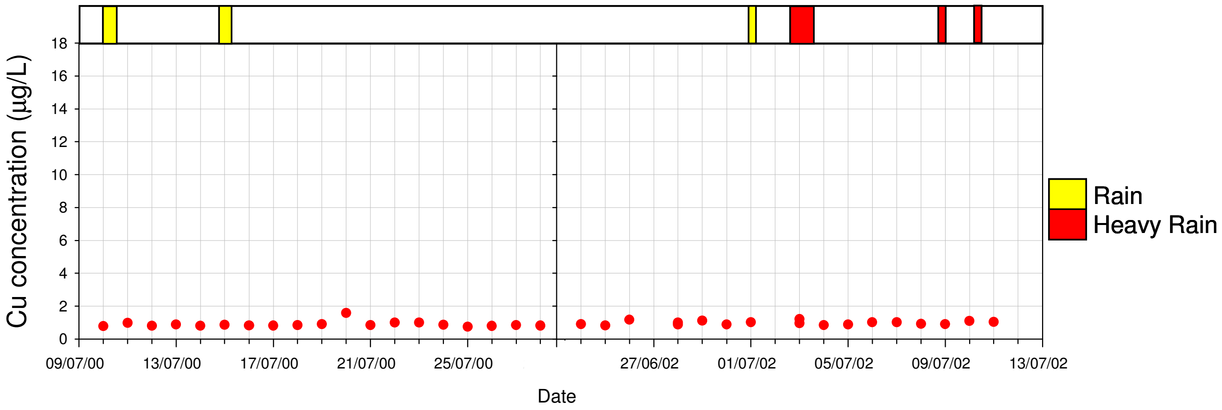


Figure 5.5 Time series data for MS 3

5.2.3 1997 MA1

Figure 5.6 presents the time series data for 1997 MA1, which has one of the largest spreads of data. However these are only equivalent to the upper 25% of the spatial sample site data. The data range is between 3 and 10 µg/L although there are no outliers. There are peaks at 21/06/97 and 25-26/06/97. These peaks correspond to two heavy rainfall events. The first peak was preceded by two days of showers and two and a half days of heavy rain. The showers had little effect on the monitor site data; although it is likely the heavy rain caused the peak on 21/06/97. It seems likely that heavy rain all day on 25/06/97 and 26/06/97 caused the higher concentrations on these days. This monitor site overlies Mercia Mudstone, and the closest measured BFI is 0.62 (Marsh and Hannaford, 2008). This indicates a major portion of the river flow at the monitor site should be maintained by groundwater baseflow. The response of such rivers is generally not "flashy"; the concentration peaks corresponded with prolonged heavy rain fall, rather than occurring with each shower. The large variations mean the data span a large part of the range shown by the spatial data (upper 50%). This means there would be some mapping variations if these data were treated as spatial sample sites. This shows a limitation of simply using the box plots to identify potentially important sites.

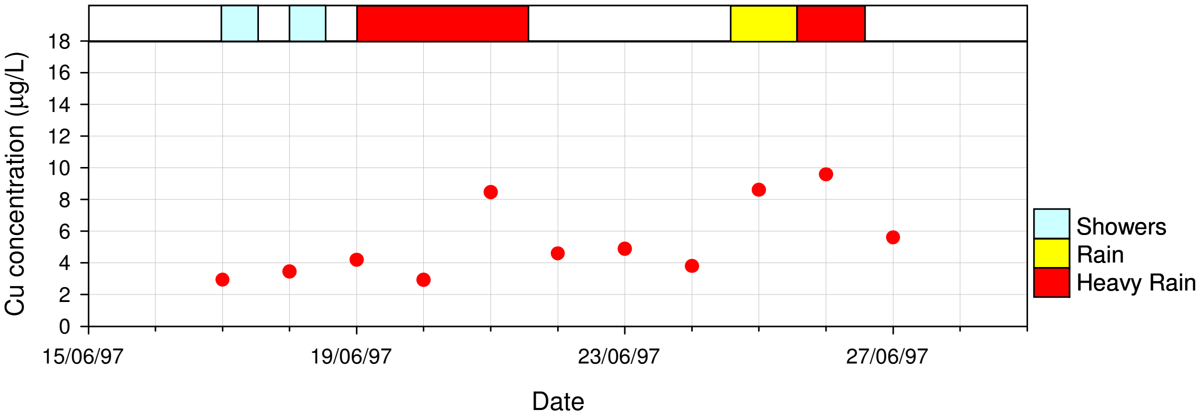


Figure 5.6 Time Series data for 1997 MA1

5.2.4 1997 MA3

Figure 5.7 presents the time series data for 1997 MA3, which is representative of a site with a small range, yet one large outlier. Most of the data ranges from 1.3 to 2.1 µg/L, and therefore they all lie between the 50th and 75th percentile classes of the spatial data. There are two outliers: one of 3.2 µg/L sampled on 22/07/97, which is one percentile class greater than the bulk of the data; and one of 27.5 µg/L sampled on 14/07/97, which is four percentile classes greater than the bulk of the data.

The larger outlier does not correspond to any weather events and the magnitude and isolation from any other sample implies that this may be caused by post-sampling contamination or analytical error. Further examination of the monitor site data confirms this. The conductivity is 910 µS/cm, which is the highest recorded at 1997 MA3, but it is not elevated in comparison to the rest of the conductivity data to the same extent that the Cu data is. However the upper 99th percentile of spatial sample site Cu data (>15.7 µg/L Cu) have an average conductivity of almost 8000 µS/cm. The pH and HCO_3 of this one sample are comparable to the rest of the samples at this monitor site. It is therefore likely that either particulates contaminated the sample at the time of sampling or post-collection contamination occurred.

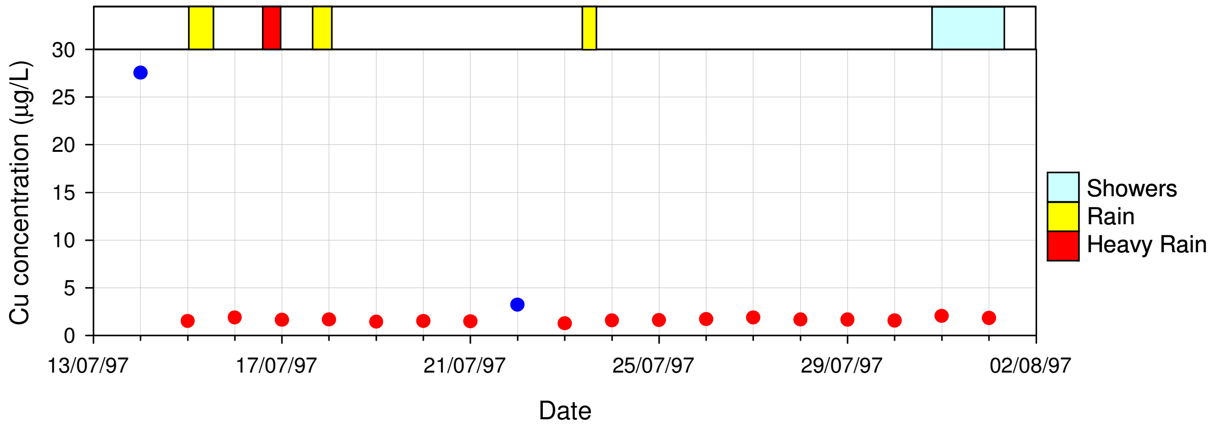


Figure 5.7 Time series data for 1997 MA3. Blue circles represent outliers

The second outlier (Figure 5.7) also does not correspond to any weather events, and can be seen to be a minor increase in relation to the rest of the data. Figure 5.8 shows the location of sites sampled on 14/7/97 and 22/07/97 in relation to the surrounding samples. It can be seen that the samples taken on these days do not represent anomalous data when compared to the surrounding spatial data. For example these sites do not stand out for having particularly higher or lower concentrations than the surrounding data, nor are they particularly associated with areas where the interpolated map appears particularly variable. It is therefore unlikely that the cause of these outliers had any measurable effect on the day's spatial sample sites. Figure 5.9 shows the spatial sample site data corresponding to the 1997 MA3 data. It shows that there are no temporal

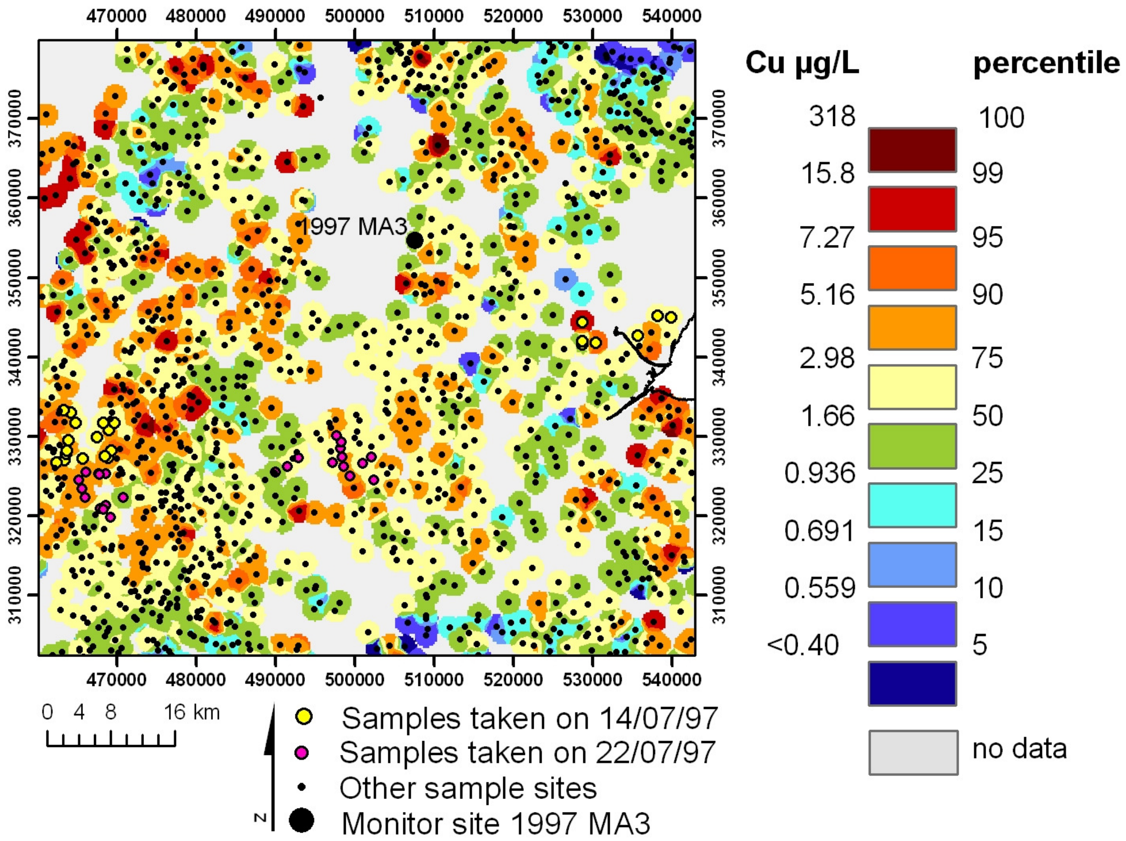


Figure 5.8 Location of sites sampled on 22/07/97

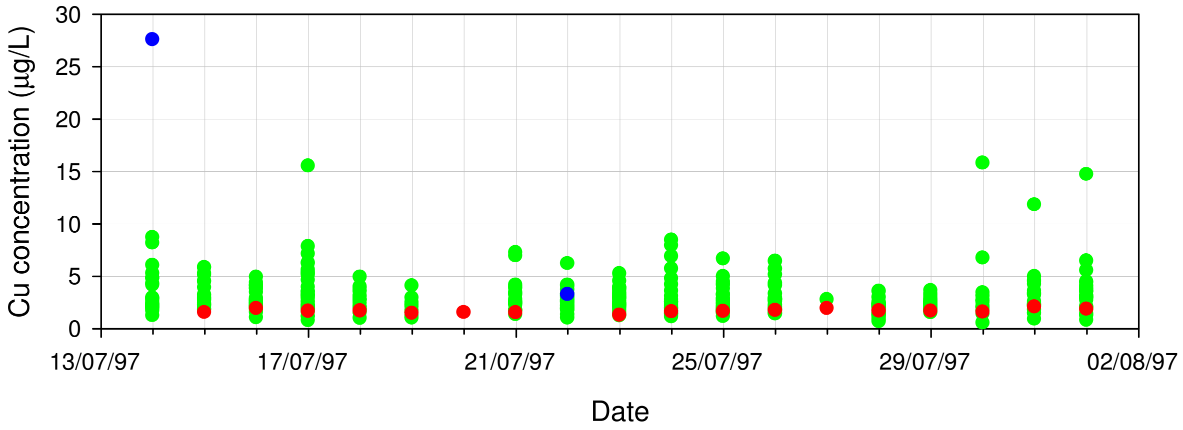


Figure 5.9 Time series data for 1997 MA3 (red circles, blue circles represent outliers) with corresponding spatial sample site data (green circles)

relationships to the outliers: that is the spatial sample site data corresponding to the monitor site outliers are not distinct from the remaining spatial sample site data. While the monitor site has shown these outliers, their cause does not appear to significantly affect the spatial sample site data. It also shows the unexpected nature of the high Cu concentration on 14/07/97.

5.2.5 1997 MB2

Figure 5.10 presents the time series data for 1997 MB2, which also has one of the largest data spreads. These data also lie within the upper 25th percentile of the spatial sample site data. The data range between 3 and 10 µg/L and there are no outliers. It is not surprising that these data are similar to 1997 MA1, as the sites are only 7km apart, and also overlie Mercia Mudstone Group geology. There appear to be peaks at 30/06/97 and 03/07/97, which correspond to showers and short-lived rainfall events. The closest recorded BFI is 0.49 (Marsh and Hannaford, 2008), which implies moderate groundwater influence. The effects of shorter lived rainfall events are evident at this site. The data span three G-BASE percentile classes, which would have some effect on the regional map. This again shows that there are no outliers to highlight this due to the intrinsic dataset variability.

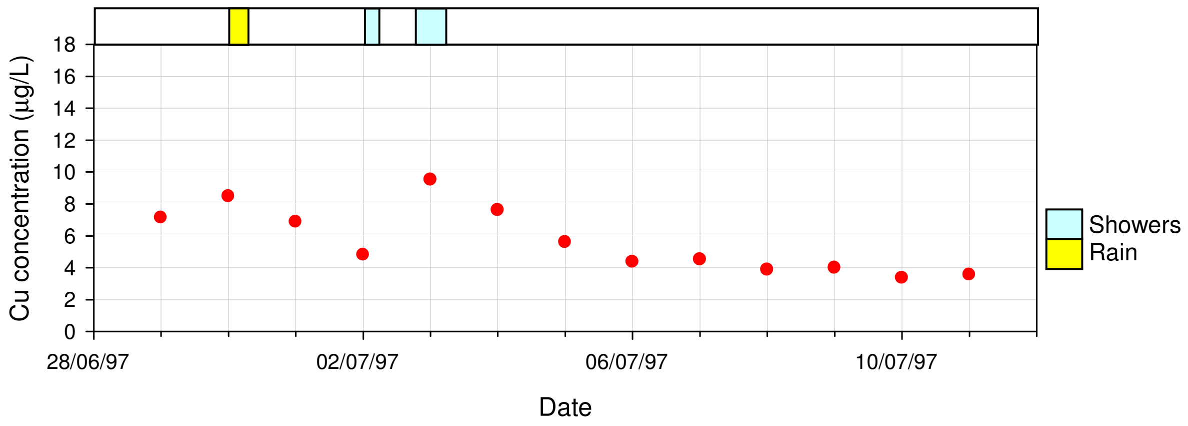


Figure 5.10 Time series data for 1997 MB2

5.2.6 1998 MA3

Figure 5.11 presents the time series data for 1998 MA3. This monitor site has the largest spread of Cu values, ranging from 3 to 16 mg/L, although this does lie within the upper 25th percentile of the spatial sample site data. This monitor site overlies limestones and sandstones of the Inferior Oolite Group. The closest recorded BFI, which overlies the same geology, is 0.72 (Marsh and Hannaford, 2008), which implies a strong groundwater baseflow influence. Despite this there are a series of small peaks centred on 16/08/98, 21/08/98, 25/08/98 which correspond to rainfall events within the preceding 24 hours. In contrast the largest and second largest peaks (30/08/98-04/09/98 and 08/08/98 respectively) are not preceded by any large rainfall event. It is therefore unlikely that weather events are the sole cause of the variable concentration. The monitor site is located in a built up area and is only ~10 m upstream of a footbridge. Furthermore the stream emerges near an industrial estate and crosses under major roads prior to the monitor site location. Impermeable built up areas allow over land discharge, and it would be easy for contamination to affect the temporal variation. The site selection here is poor, and it is unlikely that it represents the majority of spatial G-BASE samples. However it was selected to demonstrate the effect of such a location in contrast to 1998 MB3 (see Section 5.2.7).

Between 30/08/98 and 04/09/98 most of the 1998 MA3 data were defined as outliers. When compared to the spatial sample site data these outliers are above the 95th percentile. The spatial sample site data were examined to establish if the 1998 MA3 outliers corresponded to any anomalous spatial sample site data. Unfortunately there was only one day in this period (03/09/98) where spatial samples were taken. Figure 5.12 shows the location of sites sampled on 03/09/98 in relation to the surrounding spatial samples. This figure shows that these spatial sample sites range from the 25th to the 90th percentile, which is quite a large spread of data for only eight sample sites. However the area around the sites sampled on 03/09/98 demonstrates the variable concentrations typical of the regional Cu map (see Figure 1.1), and it is not the samples taken on this particular day that cause inconsistencies any more than samples taken on other days. The samples taken on 03/09/98 show no systematic artificial increase in the concentration that the monitor site data may imply. It can be seen that these sample sites are fairly close to the monitor site (within 15 km). However, the sample sites overlie Kellaways Formation and Oxford Clay Formation which comprise mudstones, siltstones, and sandstones. The closest BFI measurement to the sites sampled on 03/09/98 of corresponding geology is 0.48 (Marsh and Hannaford, 2008). This BFI indicates that the ground water influence is proportionally less than that at the monitor site, meaning these rivers may have a more "flashy" response to rainfall events. These streams may therefore be more readily affected by any weather events than the monitor site. As the spatial sample site data do not show evidence of such a disturbance, it is likely that the outliers seen in 1998 MA3 do not necessarily correspond to anomalies in the spatial data set.

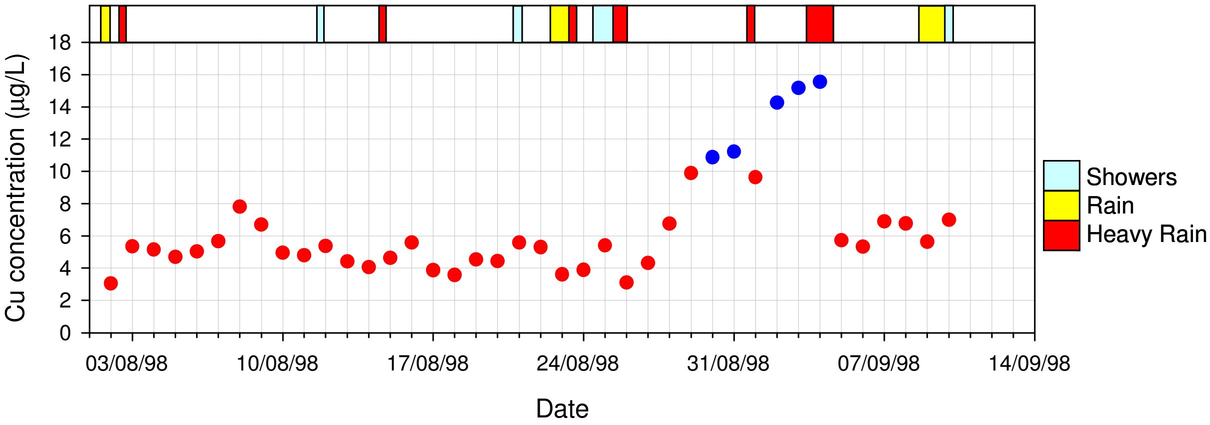


Figure 5.11 Time series data for 1998 MA3. Blue circles represent outliers

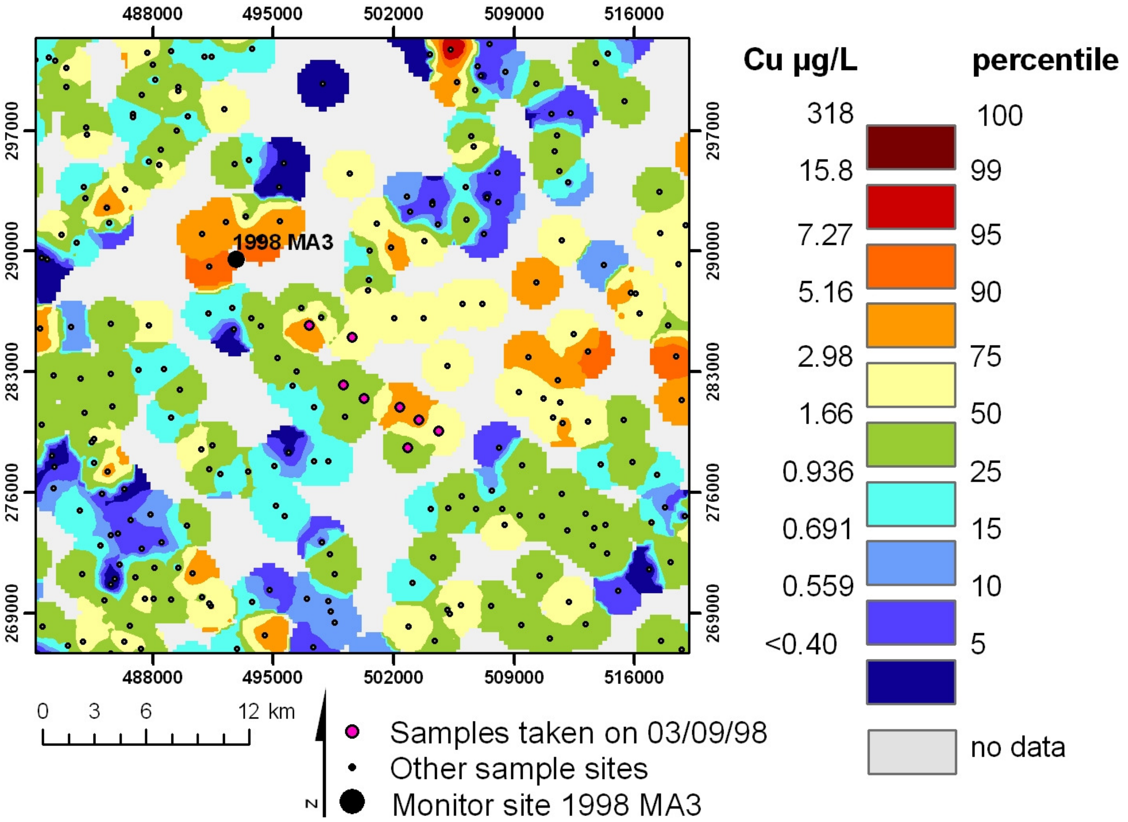


Figure 5.12 Location of sites sampled on 03/09/98

5.2.7 1998 MB3

Figure 5.13 represents time series data for 1998 MB3. The concentrations at this site range from 0.4-3 µg/L, which spans the interquartile range of the spatial data set. This site was sampled over the same time period as 1998 MA3 and is situated only 5 km away from that site. The underlying geology is similar to 1998 MA3, comprising sandstones and limestones of the Great Oolite Group. The closest estimation of BFI based on distance from the site and similar geology is also 0.72 (Marsh and Hannaford, 2008) The main geological difference between 1998 MB3 and 1998 MA3 is the presence of diamicton at 1998 MB3. According to Marsh and Hannaford (2008) this could mean the actual BFI is likely to be higher than 0.72, as the gauging station from which this measurement derives is not underlain by diamicton. In contrast to 1998 MA3, the site selection is ideal, 1998 MB3 is a first order stream which does not flow through any urban areas and the site

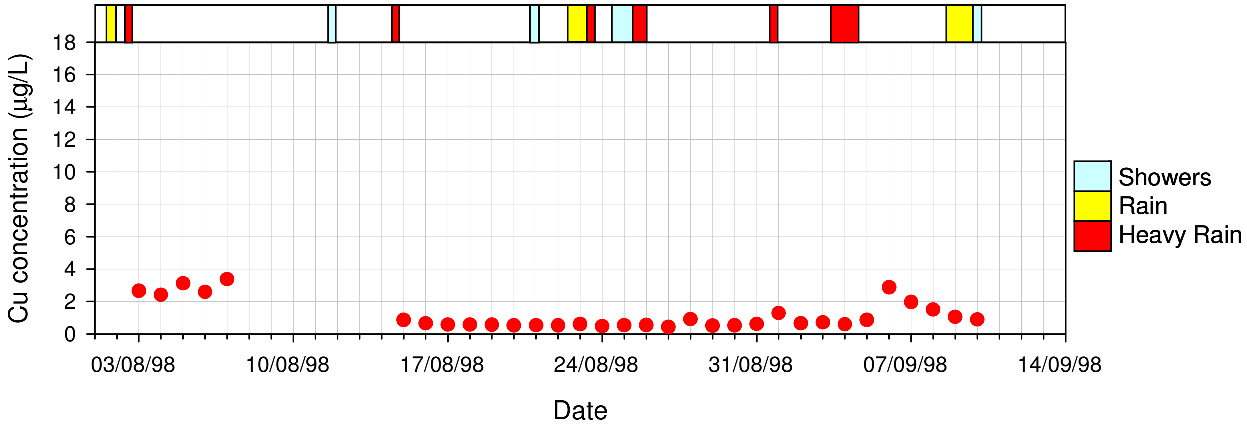


Figure 5.13 Time series data for 1998 MB3

is ~140 m upstream from the first road crossing. These site conditions are reflected in the monitor site data, which, while it spans the entire interquartile range it has no outliers. Interestingly, despite the site being so close to 1998 MA3 there is no corresponding peak in the data in the period 30/08/98 to 04/09/98. There is a small peak around 06/09/98, which is preceded by a rainfall event. The monitor site samples taken between 03/08/98 have Cu concentrations twice those of the remaining data, but there is no clear reason for this. None of the other rainfall events which occurred during the sampling of this monitor site seem to have any effect on the 1998 MB3 data.

Both 1998 MB3 and 1998 MA3 are presented with corresponding spatial sample site data on Figure 5.14. This shows how the 1988 MB3 data do not correspond to the 1998 MA3 outliers, and critically how the sample data seems unaffected by the processes causing these, which is probably the effluent or the presence of storm drains increasing run off from the built up area.

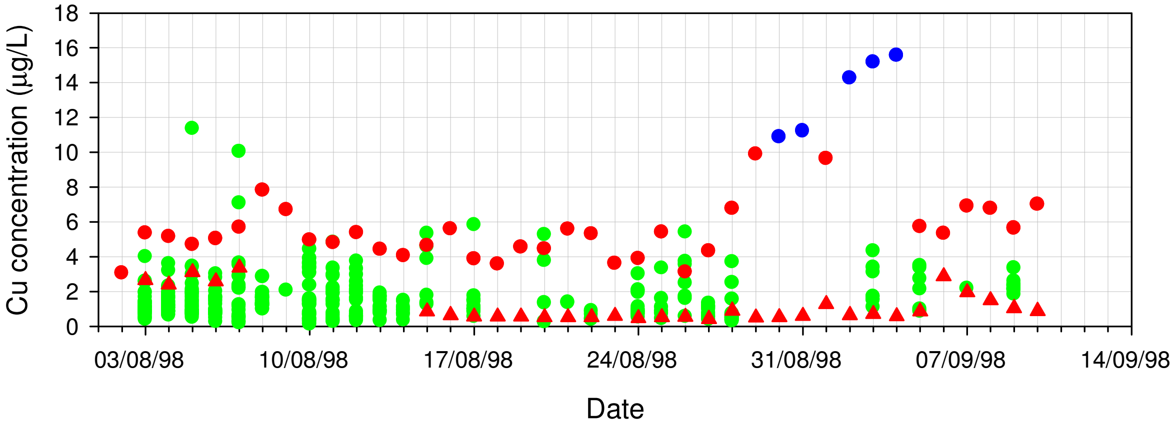


Figure 5.14 Time series data comprising 1998 MA3 (red circles, and blue circles represent outliers) and 1998 MB3 (red triangles) and corresponding spatial sample site data (green circles).

5.2.8 2005 MA1

Figure 5.15 presents the time series data for 2005 MA1. The concentrations at this site range from 1-6 µg/L and span the entire interquartile range. The site overlies London Clay comprising

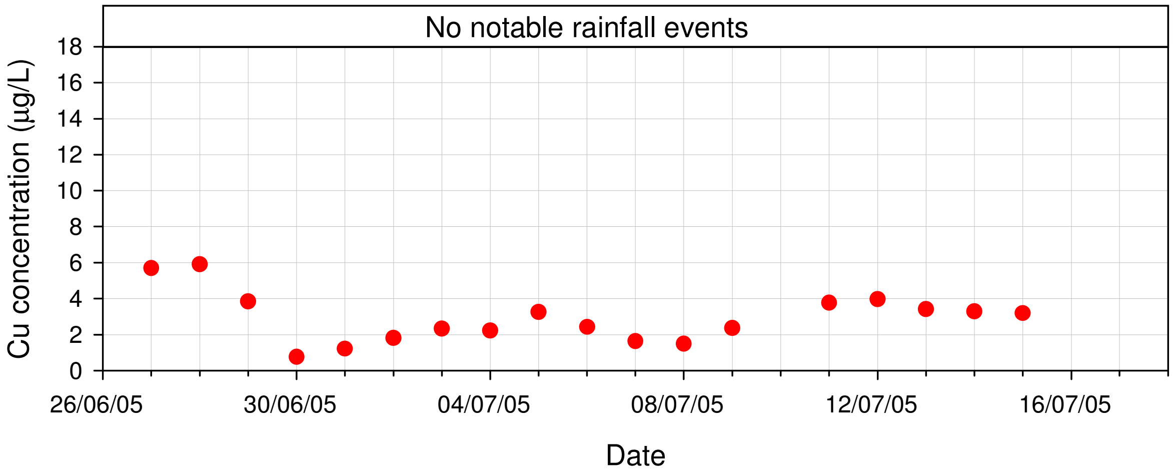


Figure 5.15 Time series data for 2005 MA1

clay, silt, sand and gravel. The closest representative BFI is 0.28 (Marsh and Hannaford, 2008) indicating a low groundwater influence. During sampling at this site there were no rainfall events. This is notable, because a lack of precipitation can also give rise to changes in stream water chemistry due to the changes in relative contribution of a longer residence time in the groundwater body. There are peaks in the data which may be caused by the lack of percolating water. Soluble salts leached from the substrate can become more concentrated during dry periods, creating a higher concentration, although the flow, and hence total metal loading, may be lower. Figure 5.16 presents the monitor site data with corresponding spatial sample site data, showing that any relationship of spatial sample site data with fluctuations in the monitor site data caused by the dry weather is minimal.

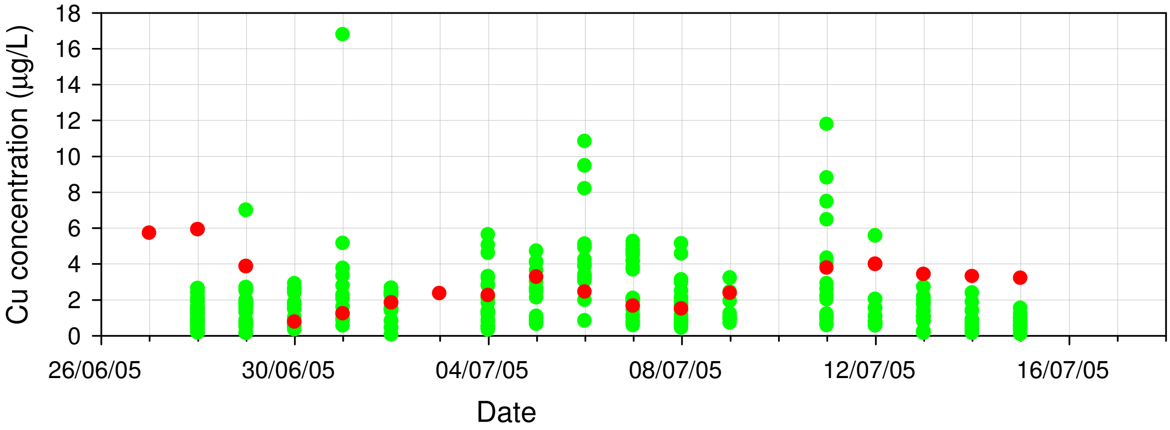


Figure 5.16 Time series data for 2005 MA1 (red circles) with corresponding spatial sample site data (green circles)

5.2.9 MS1 (1998 MB1 and 1999 MA4)

Figure 5.17 represents the time series data for MS1, comprising 1998 MB1 and 1999 MA4. This demonstrates the temporal variations spanning two field seasons. If the outliers are ignored the 1998 MB1 data averages 1.7 µg/L Cu, and the 1999 MB4 data averages 2.1 µg/L Cu. These data lie between the 50th and 75th percentile which is indistinguishable on a G-BASE percentile classified map. The individual data values span two G-BASE percentile classes, which would represent minimal difference on the regional map. The site overlies mudstones, and the most representative BFI is 0.59. There is therefore a moderate groundwater influence. Fluctuations in

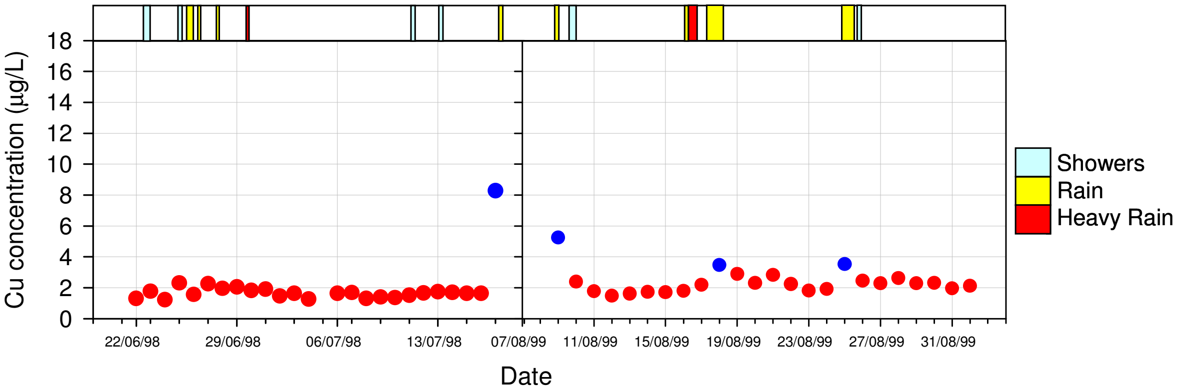


Figure 5.17 Time series data for MS1. Blue circles represent outliers

the Cu concentration seem largely controlled by the rainfall events, which precede peaks in the data. These lead to outliers in the 1999 data. The outlier in the 1998 data may also be caused by rainfall, but it is an isolated point measuring 8.2 µg/L Cu. This may imply that it is caused by either particulates entering the sample at the time of sampling, contamination during sample preparation, or analytical error. The measured pH and HCO_3 for this sample are comparable with those measured for the rest of the monitor site samples. The conductivity corresponding to this sample is the highest measurement for 1998 MB1 (899 µS/cm), but lower than comparable spatial sample site data: the average conductivity for the sample sites where the Cu concentration ranges from 7.2 to 9.2 µg/L is in excess of 2000 µS/cm. This implies that this outlier represents contamination derived from the sampling process or an analytical artefact.

The samples taken on 17/07/98, 09/08/99, 18/08/99, and 25/08/99 were defined as outliers. The spatial sample site data corresponding to these dates were examined to establish if the MS1 outliers corresponded to any anomalous spatial sample site data. Figure 5.18 presents the monitor sites superimposed on the spatial sample site data collected while MS1 was the monitor site. Where outliers are present in the monitor site data, there are no apparent anomalies corresponding in the sample site data. Figure 5.19 shows that the spatial sample site data taken on dates where there were monitor site outliers is comparable with the surrounding data. These data fit the general trend of the surrounding spatial data and show no more, or less inter-percentile class variability than the surrounding data. These spatial sample sites range from two to 65 km away from the monitor site. In many cases the weather regime and local hydrology will not necessarily be relevant. However even the spatial sample sites closest to the monitor site do not show any anomalous behaviour given the surrounding trends.

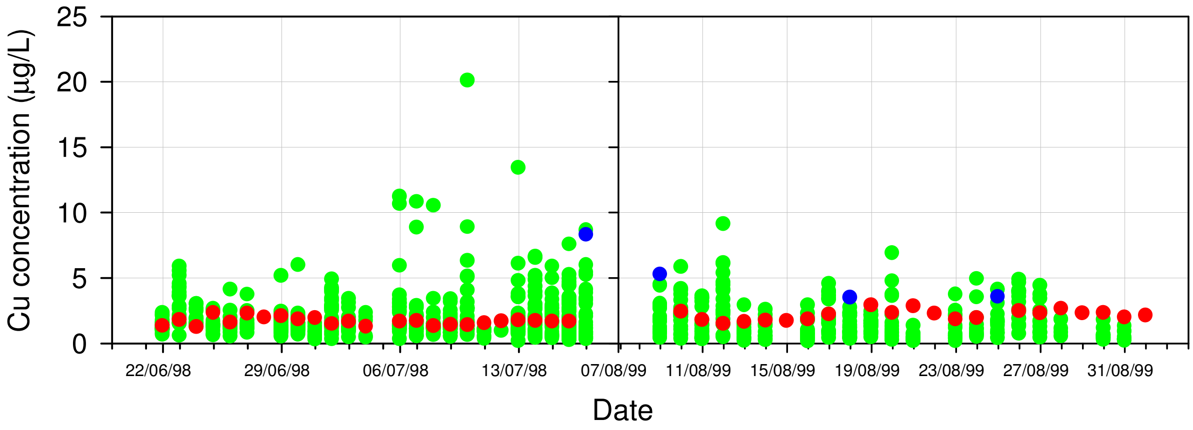
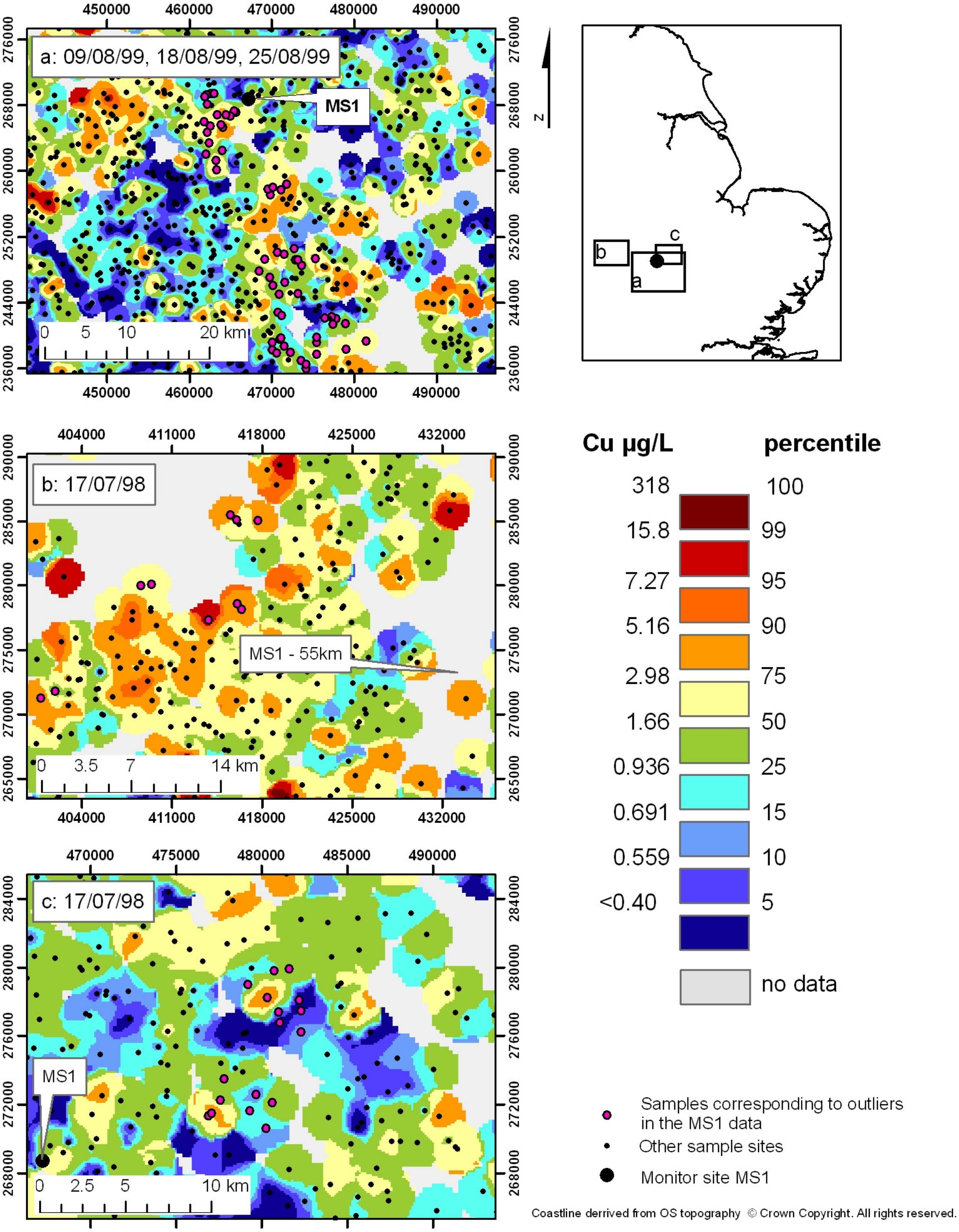


Figure 5.16 Time series data for MS1 (red circles, and blue circles represent outliers) and corresponding spatial sample site data (green circles).



Coastline derrived from OS topography © Crown Copyright. All rights reserved.

Figure 5.19 Location of sites sampled on 09/08/99, 18/08/99, 25/08/99 (a) and, 17/07/98 (b and c)

6 Conclusions

A number of significant observations have been highlighted in the above discussion:

- While examples of outliers in the monitor site data have been emphasised throughout Section 5, the majority of the monitor sites have significantly less variation within sites than the spread of spatial data between sites. It should also be noted that the greatest variations within monitor sites were seen in 1997 MA1, 1997 MB2, and 1998 MA3, which also had the highest median values.
- Where the monitor site data did present peaks and outliers this did not necessarily indicate poor spatial sample site data. There may have been small differences on these days, but not enough to change the overall trends evident on the regional map. The outliers are caused by rainfall events, sampling or handling errors, or random analytical errors. The only one of these to affect the spatial sample site data is a rainfall event, and localised rainfall may not affect the monitor site and spatial sample site data equally.
- Despite a variety of geological and hydrological conditions between each monitor site the temporal outliers represent <4% of the Cu monitor site data (See Figure 5.2), and the fluctuations in the monitor site data are generally limited to two percentile classes for G-BASE spatial data. This indicates that the overall effect on this presentation and interpretation of these Cu data is minimal.
- While broad trends can be seen in the Cu spatial data, the overall appearance of the G-BASE map is locally more variable than observed for other elements such as Ca. Of course it is possible that the occasional G-BASE spatial sample would represent one such outlier, although the sampling density is great enough, that such sites would be lost in the overall trend of the spatial data. The local concentration variations presented on the regional gridded Cu map either represent a natural variability caused by the wide dispersal of Cu in the environment, or large scale Cu contamination of many samples during the sampling and handling processes. The latter is highly unlikely, as there are no significant differences in the variable nature of the map between field seasons, and the rigorous quality control procedures (blanks, duplicates, replicates) are designed to pick up such instances.
- While individual anomalous spatial sample sites may be of significance, such as a point source of pollution, it would be expected that proximal sites may also show some effect, albeit to a lesser extent. So when interpreting the data the limitations of taking a single sample should be appreciated and caution should be exercised if trying to place significance on a single spatial sample site, as is the case with all such data.
- High Cu concentrations or upper outliers were often associated with rainfall events. Such increases could relate to increased runoff and particulate loadings in streams, with increased concentrations of colloidal materials in the analysed samples.

7 Future study

This work can be expanded by evaluating more G-BASE data to asses to the relevance of any spatial data that may also vary temporally. This will be achieved by the following:

- A suite of elements will be studied. This will establish how elements with different geochemical behaviour react to the same weather events. Their behaviour will be compared both in the temporal monitor site data, and within the spatial mapping data.
- A statistical approach will be taken, using ANOVA analysis to numerically assess variance within and between monitor sites.
- Automation of data, using R, will be used to enable rapid analysis of the large data set, enabling comparison within and between sites for a variety of elements.

References

British Geological Survey holds most of the references listed below, and copies may be obtained via the library service subject to copyright legislation (contact libuser@bgs.ac.uk for details). The library catalogue is available at: http://geolib.bgs.ac.uk.

ANDER, L. 2009. Quality control of regional geochemical stream water samples collected in 2005 in Northern Ireland for the Tellus project.

APPELO, C A J, and POSTMA, D. 2007. *Geochemistry, Groundwater and Pollution, 2nd Edition*. (Lieden: A. A. Balkema Publishers.)

BRITISH GEOLOGICAL SURVEY. 1999. *Regional geochemistry of Wales and part of west-central England: stream water*. (Keyworth, Nottingham: British Geological Survey.) ISBN 0 85272 363 6.

DREVER, J I. 1997. *The Geochemistry of Natural Waters: Surface and Groundwater Environments-Third Edition*. (New Jersey: Prentice Hall.)

FAURE, G. 1998. *Principles and Applications of Geochemistry*. (New Jersey: Prentice Hall.)

HEM, J D. 1992. *Study and Interpretation of the Chemical Characteristics of Natural Water*. (Third Edition edition). US Geological Survey Water-Supply Paper 2254. (Washington: United StatesGovernment Printing Office.)

JOHNSON, C C. 2005. 2005 G-BASE Field Procedures Manual. *British Geological Survey, Keyworth, UK*, Internal Report No. IR/05/097.

JOHNSON, C C, BREWARD, N, ANDER, E L, and AULT, L. 2005. G-BASE: Baseline geochemical mapping of Great Britain and Northern Ireland. *Geochemistry: Exploration-Environment-Analysis*, Vol. 5(4), 347-357.

KINNIBURGH, D G, and EDMUNDS, W M. 1986. The susceptibility of UK groundwaters to acid deposition. Hydrogeological Report, British Geological Survey No. 86/3.

LANGMUIR, D. 1997. *Aqueous Environmental Geochemistry*. (New Jersey: Prentice-Hall.)

MAEDA, Y. 2003. Metal recycling ang waste treatment. 535-544. In: *Metallurgical and Materials Processing: Principles and Technologies, Vol 2 - High-Temperature Metal Production*. KONGOLI, F, ITAGAKI, K, YAMAUCHI, C, and SOHN, H Y (editors).

MARSH, T J, and HANNAFORD, J, (editors), 2008. *UK Hydrometric Register*. Hydrolgical data UK series. (Wallingford: Centre for Ecology and Hydrology).

MARTENS, P N, RUHRBERG, M, and MISTRY, M. 2002. Assessing global land requirement for surface copper mining. *Mineral Resources Engineering*, Vol. 11, 337-348.

MATKOVIC, P, CRNKO, J, and MATKOVIC, T. 2004. Education of modern engineers of metallurgy in the context of the world development of metallurgical industry. 57-70. In: *Annual 2004 of the Croatian Academy of Engineering*. KNIEWALD, Z (editor). *Annual of the Croatian Academy of Engineering*.

SHAND, P, EDMUNDS, W M, LAWRENCE, A R, SMEDLEY, P L, and BURKE, S. 2007. The natural (baseline) quality of groundwater in England and Wales. *British Geological Survey & Environment Agency*, RR/07/06 & NC/99/74/24.

SMEDLEY, P L, and ALLEN, D. 2004. The Granites of south-west England. *British Geological Survey and Environment Agency*, BGS Report CR/04/255; Environment Agency Report NC/99/74/16.

STUMM, W, and MORGAN, J J. 1996. *Aquatic Chemistry: Chemical Equilibria and Rates in Natural Waters, 3rd Edition*. (New York: John Wiley & Sons, inc.)

TUKEY, J W. 1977. *Exploratory Data Analysis*. (Reading: Addison-Wesley.)