Estimating specific surface area of fine stream bed sediments from geochemistry

B. G. Rawlins[a], G. Turner[a], I. Mounteney[a] & G. Wildman[a]

[a]British Geological Survey, Keyworth, Nottingham, NG12 5GG U.K.

Running heading: *Estimating sediment surface area from geochemistry*

Correspondence: B. G. Rawlins. E-mail: bgr@bgs.ac.uk

ABSTRACT

Specific surface area (SSA) of headwater stream bed sediments is a fundamental property which determines the nature of sediment surface reactions and influences ecosystem-level, biological processes. Measurements of SSA – commonly undertaken by BET nitrogen adsorption – are relatively costly in terms of instrumentation and operator time. We present a novel approach for estimating fine ($< 150\mu$m) stream bed sediment SSA from their geochemistry – after removal of organic matter – for agricultural headwater catchments across 15 400 km^2 of central England. From a regional set of 1972 stream bed sediment sites with common characteristics for which geochemical data were available, we selected 60 samples – based on maximising their variation in aluminium concentrations – and measured their BET SSA by nitrogen adsorption. After careful selection of potential regression predictors following a principal component analysis and removal of subset of samples with the largest Mo concentrations (> 2.5 mg kg^{-1}), we identified four elements as significant predictors of SSA (ordered by decreasing predictive power): V>Ca>Al>Rb. Our optimum model from these five elements accounted for 73% of the variation in bed sediment SSA (range 6 to 46 m^{2}g^{-1}) with a root mean squared error of prediction – based on leave-one-out cross validation – of 6.3 m^{2}g^{-1}. We believe V is the most significant predictor because its concentration is strongly correlated both with the quantity of Fe-oxides and clay minerals in the stream bed sediments, which dominate sediment SSA. Sample heterogeneity in SSA – based on triplicate measurements of subsamples – was a substantial source of variation (standard error=2.2 m^{2}g^{-1}) which cannot be accounted for in our regression model.

We used our model to estimate bed sediment SSA at the other 1792 sites and at 30 duplicate sites where an extra sediment sample had been collected, 25 metres from the original site. By delineating sub-catchments for our headwater sediment sites we selected only those sub-catchments with a dominant ($> 50\%$ of the sub-catchment area) bedrock formation and land use type; the bedrock and land use classes accounted for 39 and 7% of the variation in bed sediment SSA, respectively. Variation in esti-

mated, fine bed sediment SSA from the paired, duplicate sediment sites was small (2.7 m^{2}g^{-1}), showing that local variation in SSA at stream sites is modest when compared to that between catchments. We discuss how our approach might be applied in other environments and its potential limitations.

1. Introduction

Specific surface area (SSA) is a fundamental material property of solids often expressed as total surface area per unit of mass. In both soil and stream sediments, SSA is a dominant factor controlling particle surface reactions and is closely related to a range of chemical properties that also influence biological processes. For example, surface area is strongly correlated with cation exchange capacity (Petersen et al., 1996), the dominant factor determining the magnitude and type of adsorption-desorption reactions between dissolved ions and mineral surfaces in soil pore waters and stream beds. Specific surface area may exert an important control on phosphate sorption. Where sediments are dominated by fine-fraction minerals – clay-size particles and associated amorphous Fe and Al hydr(oxide) coatings – they also have a larger SSA for phosphorus sorption (Horowitz and Elrick, 1987) which has significant implications for water quality in agricultural catchments (Withers and Jarvie, 2008).

The heterogeneity of physical stream bed habitats – dominated by the size and surface texture of sediment – has been shown to have a significant impact on primary productivity of stream algae and respiration of benthic biofilms (Cardinale et al., 2002) thus controlling ecosystem-level processes. Also, larger quantities of organic matter tend to occur in sediments with greater SSAs (Mayer, 1994), although this is by no means a universal relationship (Galy et al., 2008). Interactions between organic matter and mineral surfaces leads to occlusion of the latter and the formation of aggregates

(Horowitz and Elrick, 1987); the SSA of sediment before and after removal of associated organic matter may be quite different (Wagai et al., 2009). In this study, we focus on the SSA of the mineral component of fine sediments – collected beneath the oxic layer of the stream bed – after the removal of organic matter.

For weathered granular materials encountered at the Earth's surface, one approach for estimating sediment SSA is to measure the particle size distribution and make assumptions about particle shape, but this does not account for the area associated with surface textures which can be very large for certain minerals such as iron-oxides and expansible clay. The SSA of a material which includes the fine structure and texture of the particles is typically measured by gas adsorption using the BET isotherm (Brunauer et al., 1938). Specific surface area measurement based on the BET isotherm – commonly applied using nitrogen gas – is relatively costly in terms of instrumentation, and requires a reasonable amount of time for sample preparation and instrument operation. Alternative, cost-effective approaches for accurate estimation of SSA include visible and near infra red diffuse reflectance spectroscopy (VNIR-DRS; Ben-Dor and Banin, 1995); in their study the authors accounted for 70% of the variation in SSA in 91 soil samples using multiple linear regression analysis of the soil spectra. An advantage of VNIR-DRS is that quartz has no spectral signature in the visible and near infra red range; those mineral phases with larger surface areas – such as amorphous iron-oxides and clay minerals – are dominant features of the spectra.

A recent study demonstrated that the particle size fractions of soil can be estimated accurately from its geochemistry (Rawlins et al., 2009). The fact that the SSA of fine bed sediment is related to the naturally occurring elements in that sediment suggest that it might also be predicted from its elemental composition. Our knowledge of geochemistry and mineralogy would suggest certain elements are likely to be strongly correlated with SSA; aluminium because of its occurrence in clay minerals and Fe and Mn associated with their amorphous oxides (Wang et al., 1997). Strong linear correlations ($r > 0.9$) between several elements (e.g. Al, Fe and Cr) in fine ($< 125\mu$m)

4

bed sediments and their SSA were demonstrated for 17 samples from a range of sites across the USA (Horowitz and Elrick, 1987). However, the diverse range of chemical processes – Fe and Mn oxyhydroxide precipitation, changes in redox potential with depth from the sediment water interface – operating in the stream environment may be too complex for consistent relationships to be observed.

The increasing availability of comprehensive geochemical data for fine bed sediments from international (Salminen et al., 2005) and national-scale (Johnson et al., 2005) surveys of both large and small catchments makes assessment of the relationship between bed sediment geochemical composition and SSA achievable. The availability of accurate estimates of SSA for a large number of sites could help to determine the relative importance of landscape factors – such as bedrock geology and land use – on the SSA of fine bed sediment. Based on the strong positive relationships observed between SSA and the proportion of the finest ($< 2\mu$m) sediment size fractions (Horowitz and Elrick, 1987), estimates of SSA for many headwater streams could help to identify significant sources of the finest sediment to large catchment systems and the mitigation of their impacts through strategic management (Owens et al., 2005).

In this paper we present our investigation of the relations between SSA and elemental composition of fine fraction, headwater stream bed sediments across a large region of central England. We explored the geochemical data and established that there are strong linear relations with SSA. We sought the most effective significant multiple regression equations for prediction from the data that we have from the regional surveys of stream sediments undertaken by the British Geological Survey. We have used our knowledge to select regressors so as to avoid redundancy and misleading inference from chance inclusions of elements.

We present our predictions of fine fraction bed sediment SSA and their errors, and discuss the limitations of our approach. The equations are used to investigate local variation in bed sediment SSA, and the importance of bedrock geology and land use type for headwater bed sediment SSA.

2. Theory

2.1 Multiple regression

Multiple linear regression is a standard statistical method for prediction. The regression model, expressed in matrix form is

$$y = \mathbf{z}^{\mathrm{T}} \boldsymbol{\beta} + \varepsilon , \tag{1}$$

in which $\mathbf{z}$ is vector of length $p+1$ containing the values of p predictor variables plus a first element, a dummy, set to 1, $\boldsymbol{\beta}$ is a vector of coefficients, the first of which is a constant, and ε is a random residual. So, to predict an unknown value, say y_0, we compute its estimate as

$$\widehat{y}_0 = \mathbf{z}_0^{\mathrm{T}} \boldsymbol{\beta} , \tag{2}$$

where $\mathbf{z}_0$ is the vector of known values for the same sample of bed sediment.

We must first have obtained the coefficients $\boldsymbol{\beta}$. If we have data on both the variable y and the predictors $z_1, z_2, \ldots, z_p$ for a sample of size n then we can estimate $\boldsymbol{\beta}$ by

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{Z}^{\mathrm{T}}\mathbf{Z}\right)^{-1} \mathbf{Z}^{\mathrm{T}}\mathbf{y}, \tag{3}$$

where $\mathbf{Z}$ is a matrix of dimensions $n \times (p+1)$ and vector $\mathbf{y}$ is of length n. The prediction variance is

$$s^2(\widehat{y}_0) = s^2(y.\mathbf{z}) + s^2(y.\mathbf{z})\mathbf{z}_0 \left(\mathbf{Z}^{\mathrm{T}}\mathbf{Z}\right)^{-1} \mathbf{z}_0^{\mathrm{T}} , \tag{4}$$

where $s^2(y.\mathbf{z})$ is the variance of the residuals from the regression of y on $\mathbf{z}$.

2.2 Selection of regressors

The British Geological Survey has determined the concentrations of 51 elements in its surveys of the headwater stream sediments across central England (see Figure 1). Some of these were unlikely to be related to SSA and so these were ignored. This left 21 elements to consider (see Methods). To have attempted to predict SSA from all these

6

21 elements in a single model would have entailed two risks, (i) the propagation of errors in the regression coefficients of weak predictors into the final predicted values and (ii) numerical instability caused by strong correlation among any two or more predictors. A common way of avoiding this difficulty is stepwise addition or elimination of the predictors, but this too carries risks in that the final model can depend on the order in which the variables are considered for addition to (in forward selection) or deletion from (in backward elimination) the model. Thus stepwise regression is not necessarily stable.

To minimize the above risks we combined a principal component analysis (PCA) of the 21 variables with judgement based on experience and general understanding of geochemistry. The PCA showed which groups of variables were strongly correlated to one another and enabled us to select only one from each such group so that we should avoid instabilities from that source. Using our experience and understanding we chose elements that we thought most likely to be correlated with fine sediment SSA and describe the combination of these in Methods.

2.3 Centre log-ratio transformation

Compositional data, such as comprehensive geochemical analyses which include silicon and aluminium, give rise to a problem in that the components have a constant sum, 1 (or 100%), with distributions that are curtailed at the limits of 0 and 1 (or 0 and 100%). Standard statistical techniques devised for unconstrained random variables – including principal components analysis based on the correlation matrices of vectors of observations – cannot be used to analyse compositional data in its raw form. Aitchison (1986) proposed a way to avoid this difficulty by using the centred log-ratio (clr) transformation. If there is a composition X of D elements:

$$X = [x_1, x_2, \ldots, x_D], \tag{5}$$

such that

$$x_i > 1, 2, \ldots, D, \tag{6}$$

7

the clr transformation of an observation (x_i) is:

$$clr(X) = ln((x_i)/g\{(x_i)\}) \tag{7}$$

and where g is the geometric mean of the elements of the composition:

$$g = \sqrt[D]{x_1 \times x_2 \ldots \times x_D}. \tag{8}$$

We can analyse the transformed data as we can any other multivariate normal data.

3. Methods

3.1 Study regions and surveys

Figure 1 shows the stream sediment sampling locations of the study region which covers around 15 400 km^2 of central England. The stream sediment samples were collected by the G-BASE project of the British Geological Survey (Johnson et al., 2005). The elevation range across the region is $20 - 255$ m above sea level, with undulating topography; the mean slope angle is around 2.5 °. The stream sediment sampling was undertaken in the summers of 1997, 1998 and 1999 in rural and peri-urban areas. Bedrock in the study region ranges in age from Precambrian to Cretaceous with a wide range of predominantly sedimentary lithologies including limestones, sandstones, siltstones, mudstones, shale, coal measures, marls, ironstones and chalk (Figure 2). There are also extensive superficial deposits including glacial tills, glacial sands and gravels, marine and river alluvium, river terrace deposits, and to the east of the region, peat deposits (Sylvester-Bradley and Ford, 1968). The soils are dominated by Brown Earths, Surface Water Gleys, Pelosols, Ground Water Gley Soils (Soil Survey of England and Wales, 1983a; Soil Survey of England and Wales, 1983b). The land use types include arable (48%), grassland (21%), built-up areas (13%), woodland (7%) and small areas of a range of other land use types (11%.)

Potential stream sampling sites were identified using Ordnance Survey maps. Mainly first and second order streams were selected, either avoiding or located upstream of obvious sources of contamination such as road intersections and farm buildings.

Where possible, sediment was collected from central areas of active stream beds after removal of the upper layer of oxidised sediment. Between 15 and 25 kg of sediment was wet-screened on site to collect the fraction finer than 150μm typically yielding a final mineral mass of approximately 500 g. All samples were returned to a local field base for air-drying. The location of each stream bed sampling site was recorded using a handheld GPS with an accuracy of around 5 m. At site, the width of each stream, stream order and a classification of flow conditions during sampling were recorded. An estimate of the organic matter content of the sediment (low, medium or high) was also recorded.

In total there were 5047 stream sediment sampling sites across the study region; at sixty-five of these sites, an extra or duplicate sample was collected from a location within 25 m of the original sampling site; the same sediment sampling protocol was applied at the duplicate sites. The sites across the study region (n=5047) comprise a range of : i) stream orders (1st to 4th), ii) flow conditions (no flow to bank full), and iii) stream channel sizes (widths of a few feet to several metres). We wished to make comparisons of fine bed sediment for streams of similar sizes and flow conditions to avoid any bias such variation might introduce into our estimates of SSA. We therefore restricted the number of stream sediment sampling sites to: i) first and second order streams, ii) low to moderate flow conditions, and iii) streams with channels of between 1 and 3 metres width. Of the original 5047 stream sediment sites, a total of 1972 sites met these conditions. We limited our study to these sites and included 30 of the 65 duplicate sites which also met these conditions. We use the paired sites (original and duplicate) to provide an estimate of the local variation in fine bed sediment composition. Their locations are shown in Figure 1.

On return to the laboratory, all samples were freeze-dried, coned and quartered and a 50-g sub-sample was ground in an agate planetary ball mill. The total concentrations of major and trace elements were determined in each sample by wavelength and energy dispersive XRFS (X-ray fluorescence spectrometry).

9

We wished to select a subset of 60 samples for SSA measurement. We ranked all 1972 samples from smallest to largest by their total Al content because we considered that Al was the element most likely to be closely related to SSA. We randomly selected one sample from equally sized groups of 33 samples from our ranked set to ensure that we would capture the full range of Al concentrations. We plotted the locations of the 60 samples across the study region to check that they were representative of the entire region, shown in Figure 1. We analysed each of the 60 selected samples for their SSA (see below), and used the remaining samples to estimate SSA from their geochemistry.

3.2 Total organic carbon and BET specific surface area analysis

Two sub-samples each weighing approximately 10 g were taken from each of the sixty selected sediments; one for the estimation of total organic carbon (TOC), the other for SSA. For TOC, each 300 mg sub-sample was treated with a small quantity of HCl (5.7 M) to remove inorganic carbon and total organic carbon was estimated on the remaining sample by combustion in a Costech ECS4010 Elemental Analyser (EA) calibrated against an Acetanilide standard. Replicate analysis of well-mixed samples indicated a precision of $\pm < 0.1\%$.

Prior to BET SSA analysis, all organic matter was removed from the 60 sub-samples. Approximately 150 ml of hydrogen peroxide (H_2O_2; 30% by weight) was added to each 10 g subsample in a vessel and placed in a water bath at 60 °C for 3 hours. Ethanol was added to reduce effervescence where necessary to prevent loss of sample material. The solution was left overnight and in the morning its temperature was raised to 100 °C for a few minutes. When the solution had cooled to leave a residue the same procedure was repeated one or two times until no further reaction occurred. The final sample residue was gently disaggregated with a pestle and mortar. For each of the sixty samples, approximately 1 g of each sub-sample was weighed out using an accurate, four decimal place mass balance. All samples were prepared using a Micromeritics Gemini VacPrep Degasser; the samples were degassed overnight at 60 °C prior to SSA analysis. The BET specific surface area of each sample was

10

₂₃₆ determined using a multi pressure point analysis using a Micromeritics (Norcross, GA,
₂₃₇ USA) Gemini VI 2385C series physisorption system. A carbon black standard was
₂₃₈ analysed with each batch of samples to monitor accuracy and precision. Triplicate
₂₃₉ analyses were undertaken on three separate sub-samples of six representative samples
₂₄₀ to estimate the degree of sub-sampling heterogeneity. Repeated, triplicate analyses
₂₄₁ were also undertaken on six samples to estimate precision.

₂₄₂ *3.3 Landscape analysis*

₂₄₃ We used a 5 m resolution digital terrain model (DTM; Intermap, 2009) of the study re-
₂₄₄ gion and hydrological functions in ArcMap9.3$^{\text{TM}}$(ESRI) to generate drainage networks.
₂₄₅ We then superimposed on this the locations of each of the 1972 sediment sampling sites
₂₄₆ and snapped these to the nearest stream, ensuring that where two streams were close
₂₄₇ together, the correct stream had been selected by reference to positions which had been
₂₄₈ recorded on hardcopies of Ordnance Survey maps during sample collection. We then
₂₄₉ used hydrological functions and the DTM to delineate polygons of the sub-catchment
₂₅₀ areas draining to each of the sampling sites.

₂₅₁   We used digital versions of the 1:50 000 maps of bedrock geology of England,
₂₅₂ part of DigiMap GB of the British Geological Survey (2006) to determine the types
₂₅₃ and proportions of each bedrock formation which outcrop in each sub-catchment. We
₂₅₄ overlaid the sub-catchment polygons onto a layer of bedrock polygons with codes for
₂₅₅ each class in a GIS system, and cut the former into sections using the latter. We
₂₅₆ then used the GIS to calculate the proportions of each bedrock type in each sub-
₂₅₇ catchment based on the codes for each bedrock formation polygon. We considered
₂₅₈ including the superficial deposits as a further classification but when we added this
₂₅₉ as another layer the number of polygon classes increased to such an extent that we
₂₆₀ decided that statistical analysis and interpretation of the results would have been too
₂₆₁ complex. Using the same procedure, we used a digital version of the 25 m resolution
₂₆₂ Land Cover Map 2000 (Fuller et al., 2002) for the study region with codes for each
₂₆₃ dominant habitat type to determine their proportions in each sub-catchment of the

₂₆₄ sediment sampling sites.

₂₆₅   To determine the importance of bedrock formation on estimated bed sediment
₂₆₆ SSA, we first needed to identify those sub-catchments which were dominated by one
₂₆₇ lithology or land-use type and remove any catchments which were anomalously large.
₂₆₈ We selected those sub-catchments where a single bedrock formation accounted for more
₂₆₉ than 50% of the catchment area (n=1340) and where the sub-catchment area was < 40
₂₇₀ km^2. We applied the same approach to the sub-catchments based on the land cover
₂₇₁ data, selecting those which were dominated ($> 50\%$ of total area) by one habitat type
₂₇₂ (n=415). We then used the bedrock and habitat codes for these sub-catchments in
₂₇₃ a one-way analysis of variance to determine their importance for bed sediment SSA
₂₇₄ based on our estimates.

₂₇₅ *3.4 Exploratory analysis and choice of predictor variables*

₂₇₆ Of the 51 elements recorded across the study region we chose 21 which seemed likely
₂₇₇ to be related to SSA. They were as follows (with their detection limits in mg kg^{-1}):
₂₇₈ Al (530), As(1), Ba(5.1), Ca(720), Co(1.2), Cr(1.3) , Fe(70), Ga(0.7), Hf(0.7), K(415),
₂₇₉ Mg(600), Mn(77), Ni(0.6), Rb(0.5), Si(470), Sr(0.6), Ti(120), Th(0.7), U(0.6), V(1.3),
₂₈₀ and Zr(0.8). In those cases where element concentrations in each sample were below
₂₈₁ the detection limit, the concentrations were set to half this value. Table 1 lists their
₂₈₂ means, medians, standard deviations and skewness coefficients, and also their values
₂₈₃ after clr transformation.

₂₈₄   Prior to a principal component analysis, we first closed the composition of the 21
₂₈₅ geochemical elements – ensuring that the sum of each sample is one – using the closure
₂₈₆ function (clo) in the R package *compositions* (van den Boogaart et al., 2008). After
₂₈₇ converting the data to a compositional data structure using the 'acomp' function we
₂₈₈ applied the clr transformation (Equation **??** and **??**). We then analysed the correlations
₂₈₉ between the principal components of the transformed data based on the correlation
₂₉₀ matrix, avoiding the problems typically associated with compositional data analysis.

₂₉₁   Table 2 lists the leading eigenvalues of the correlation matrix. The first accounts

for more than a quarter (27%) of the variance, the second more than a fifth (21%)
and the third, an eight (13%). The first four principal components accounted for 71%
of the variance. We converted the eigenvectors to correlation coefficients between the
component scores and the original variates by

$$c_{ij} \;=\; a_{ij}\sqrt{v_j/\sigma_i^2} \;, \tag{9}$$

where a_{ij} is the ith element of the jth eigenvector, v_j is the jth eigenvalue, and σ_i^2 is
the variance of the ith original variable. The scatter of the coefficients in unit circles in
the planes of the first and second dimensions is plotted in Figure 3a, and first and third
dimensions in Figure 3b. In general, the closer the points lie to the circumference of
one of these circles (i.e. larger correlation coefficients) the better are they represented
in that projection. The first principal component (PC) is one of magnitude, whilst the
second and third PC discriminate.

The elements exhibit a broad range of correlation coefficients with the first PC;
those elements with the largest positive coefficients are Fe, Cr and V; by contrast Ba,
Hf , K, Mg, Si and Zr, have large negative coefficients. Component 2 discriminates; Rb
and Mn having large positive and negative coefficients respectively. Four elements (Al,
Ti, Ga and Th) which have positive coefficients with the first and second components
cluster together in Figure 3a. Uranium has little relation to the other elements in
Figure 3a. In Figure 3b, the third PC is dominated by Zr and Hf which have large
negative coefficients. Calcium had the largest correlation coefficient with the fourth
principal component (not shown).

We chose the following eight elements as potential predictors:

1. **Fe** and **Mn** because their oxides have large surface areas,

2. **Al, Ga, K** and **Rb** because of their association with clay minerals and the different
   correlations of the latter with the first three principal components,

3. **V** because it was so strongly correlated with the first principal component,

4. **Ca** because it was strongly correlated with the fourth principal components and is a major component in feldspar minerals and carbonates

3.5 Building the regression model

Our eight predictors for estimating SSA by multiple linear regression do not constitute a closed composition; they do not sum to 100% because three of the major elements (Si, Fe, Mg) were not included. We needed to consider whether we should use the clr transformed data or the original (untransformed) data to fit linear regression models. If application of the clr transform helps to linearize the relationships between predictors and SSA there would be a basis for its use. However, our exploratory analysis did not show this, so we used the original, untransformed data to fit linear regression models to the sediment SSA values.

Initially we used all 60 analyses of SSA to fit regression models by least squares; having relatively few samples we decided to use leave-one-out cross validation to test the accuracy of the model predictions. Regression models were formed, Equation (**??**), by forward selection and backward elimination. Four samples had large residual errors which we observed in scatterplots of measured and predicted values. When we investigated these samples further we noted they all had large Mo concentrations (> 2.5 mg kg^{-1}; detection limit 0.8 mg kg^{-1}) and their stream bed sampling sites occurred over outcrops of shale bedrock. When we removed these four samples from our dataset the performance of the regression models improved substantially. We chose to fit regression models only to the 56 sediment samples for which we had measured SSA and their total Mo concentrations were below this threshold. This eliminated a The models accounted for a large proportion (71.3%) of the variance in sediment SSA.

We then repeated our regression analysis after removal of a single sample from the dataset; we used the model formed from the 55 observations to predict the SSA of the sample that had been removed and repeated this sequentially for each of the 56

14

$_{343}$ samples. We then calculated the root-mean-squared-error of prediction (RMSEP):

$$\text{RMSE} = \sqrt{\frac{1}{n_\text{V}} \sum_{i=1}^{n_\text{V}} (\widehat{z}_i - z_i)^2} \,, \tag{10}$$

$_{344}$ where z_i is the measured SSA and $\widehat{z}_i$ is its predicted value from the regression model.
$_{345}$ We also calculated the bias in the estimates.

$_{346}$ To investigate the magnitude of local variation in SSA at stream sites, we used
$_{347}$ the optimum regression model to estimate SSA for the 30 paired, original and duplicate
$_{348}$ sampling sites (see Methods). We undertook a one-way analysis of variance (ANOVA)
$_{349}$ using the paired sites as the factor to determine the standard error of SSA (m^2g^{-1}) and
$_{350}$ the coefficient of variation (%). Finally, we investigated the influence of bedrock geology
$_{351}$ and land use type on fine stream sediment SSA by examining box and whisker plots
$_{352}$ where these factors were classifiers and we undertook one-way ANOVA to calculate the
$_{353}$ amount of variation in bed sediment SSA they account for.

$_{354}$ A summary of the number of stream sediment sites used in the different compo-
$_{355}$ nents of this study and the criteria for their selection are presented in Table 3.

$_{356}$ **4. Results**

$_{357}$ *4.1 Geochemistry and SSA*

$_{358}$ The minimum and maximum TOC contents of the 60 samples prior to the removal of
$_{359}$ organic matter were 1.0 and 8.3%, respectively with a mean content of 3.2%. The SSA
$_{360}$ of the fine ($< 150 \ \mu$m) bed sediment in the sixty samples varied by almost one order
$_{361}$ of magnitude ranging from 5.98 to 46.04 m^2g^{-1} with a mean of 22.2 m^2g^{-1}. The SSA
$_{362}$ from repeat analysis of six subsamples had a maximum surface area of 52.16 m^2g^{-1}.
$_{363}$ The correlation between SSA and TOC for the 60 samples was weak (linear Pearson
$_{364}$ correlation $r = 0.02$; non-linear Spearman-rank correlation $r = 0.06$).

$_{365}$ The triplicate analyses of subsamples from six of the sixty samples had a SSA
$_{366}$ standard error of 2.2 m^2g^{-1} and a coefficient of variation 8.4%. Repeat analyses of the
$_{367}$ same subsamples showed the instrument precision was excellent; the standard error of
$_{368}$ repeat analyses was 0.035 m^2g^{-1} whilst the coefficient of variation was 0.1%. Sample

heterogeneity is therefore a significant source of variation in our fine bed sediment SSA analyses.

The optimum linear regression model accounted for 72.7% (adj R^2) of the variation in fine bed sediment surface area with a residual standard error of 5.28 m^2g^{-1}. A scatterplot of measured versus predicted values from application of the regression model is shown in Figure 4. The four elements with greatest predictive power were: V>Ca>Al>Rb, in order of decreasing absolute value of t; Table 4 lists the model coefficients. The root mean square error of prediction (RMSEP; Equation ??) from the leave-one-out cross validation was 6.3 m^2 g^{-1} and the bias was -0.32 m^2 g^{-1}. We feel that the prediction errors and bias are sufficiently small to justify estimation of SSA using the regression model for the sediments in similar streams across the study region and for the duplicate sites (Figure 5).

We can attempt to interpret the regression coefficients of these predictors using our knowledge of mineralogy and geochemistry. Vanadium (V) is the most significant predictor with a positive coefficient; this may be because Fe-oxyhydroxides formed in soil environments incorporate V^{III} into their structure (Schwertmann and Pfab, 1996). So V may be a useful indicator of the quantity of large SSA amorphous Fe-oxyhydroxides (Wang et al., 1997) which have been transported into the stream environment. Another potential contribution to larger SSA is the occurrence of V in micaceous minerals such as roscoelite or V^{III} in the octahedral position of clay minerals with substantial SSAs (Breit and Wanty, 1991). Calcium has a negative coefficient (Table 4) and so is negatively correlated with SSA; calcium concentrations may be dominated by its occurrence in carbonate minerals which have a relatively small SSA. Aluminium and rubidium (Rb) were also strong predictors and this may be due to their occurrence in clay minerals, the latter because of its similar ionic radii with – and replacement of – potassium (K) in K-bearing clay minerals.

After taking natural logarithms of the measured total P concentrations to reduce the positive skewness of its distribution (log P skewness= 0.52), log total P had a small

16

397 positive correlation with estimated SSA ($r = 0.06$); we discuss this further below.

398 *4.2 Local variation in SSA*

399 The standard error from a one-way analysis of variance for SSA estimated at the 30

400 duplicate sites was 2.7 m^2g^{-1} with a coefficient of variation of 12.3%. This suggests

401 that local variation at stream sites is modest when compared to the range of SSAs,

402 considering the standard error of SSA for sub-sample analyses was 2.2 m^2g^{-1}. This

403 suggests the vast majority of the variation in SSA occurs either between headwater

404 catchments or at larger scales along stream reaches, not locally (<25 m) within stream

405 beds.

406 *4.3 Landscape factors*

407 The 10th and 90th percentiles of the cumulative distribution of the sub-catchment areas

408 across the study region were 0.6 and 13.3 km^2, respectively; the median was 2.43 km^2.

409 The results of one-way ANOVAs based on bedrock geology and land cover type are

410 shown in Table 5. There were a total of 66 bedrock types which accounted for a

411 large proportion (38.7%) of the variation in bed sediment SSA. This is shown in the

412 form of a boxplot for the 20 dominant bedrock formations (Figure 6). The median

413 SSA of bedrock types with larger amounts of coarse quartz (sandstones and siltstones)

414 generally have smaller median SSAs than the clay and mudstone lithologies. The

415 Northampton Sand is an exception; both it and the Marlstone Rock Formation which

416 outcrop in this region are known to contain large quantities of iron mineralisation and

417 associated arsenic (Appleton et al., 2008). Soils developed over Jurassic ironstones in

418 central England are known to contain sedimentary iron oxide phases (Breward, 2007)

419 which account for the larger SSAs in the stream bed sediments derived from them.

420 The land cover types for 90% of the selected sub-catchments were either arable

421 (n=222) or improved grassland (n=174); these were selected because they were dom-

422 inated (area $> 50\%$) by one land cover type. In total there were eight classes of

423 dominant land cover type which accounted for 7% of the variation in bed sediment

17

SSA. Based on an unpaired *t*-test, the median SSA of the fine bed sediment in the arable dominated catchments (21.2 m^{2}g^{-1}) was significantly larger ($P = 0.0016$) than the median for the improved grassland dominated catchments (18.1 m^{2}g^{-1}) which is consistent with studies that have shown greater erosion of topsoil occurring in the former (Fullen, 1997).

5. Discussion

We developed a regression model which can make reasonably accurate estimates of fine ($< 150\mu$m) bed stream sediment SSA based on four geochemical predictors in predominantly agricultural headwater catchments over a substantial area of lowland England, with a broad range of sedimentary bedrock types. Our approach needs to be tested in more diverse geochemical and geomorphological environments, such as upland areas with more acidic streamwater and also for different lithologies such as igneous and metamorphic rocks. The use of sediments sieved to a pre-determined size fraction provides us with only a partial understanding of the surface area properties of stream bed sediments. If we knew the mass of material required to produce a known mass of $< 150\mu$m sediment, we could determine the relative contributions of fine versus more coarse ($>$2mm) sediment for SSA of each stream bed.

We chose not to include the distribution of superficial deposits as a further means of classifying headwater catchments to understand the factors which account for variation in bed sediment SSA; we took the view it would lead to too many classes – each with a small number of catchments – for our statistical analysis. The amount and distribution of superficial deposits across the study area is likely to account for part of the variation we observe in bed sediment SSA for individual bedrock types. This could be explored further using a more complex approach in which the proportions of the bedrock types and superficial deposits are used as predictors in a multiple linear regression framework, but we considered this was beyond the scope of the current study.

In areas of substantial environmental contamination, approaches to the estimation

18

of bed sediment SSA based on geochemical predictors might be prone to significant error or bias. With the exception of calcium (Ca) – of which large quantities are dispersed on land through both agricultural lime and cement used in the construction industry – our predictors are not typically released in large quantities by anthropogenic activities. Hence, we believe our approach should be applicable in most earth surface environments.

Our approach for estimating fine sediment SSA is only likely to be cost-effective where geochemical data are available from large scale surveys such as the recent continental-scale geochemical survey of Europe (Salminen et al. 2005) in which bed sediments were collected using the same size threshold ($< 150\mu$m). Where some or all of the geochemical predictors are unavailable, estimation of fine bed sediment SSA by mid infra-red diffuse reflectance spectrometry (MIR-DRS) may be an effective, alternative approach. When we applied MIR-DRS to the ground powders to the 56 in our study and used partial least squares regression with spectral wavelengths as predictors to estimate SSA, the RMSEP based on leave-one-out cross validation using four model components was 6.8 m^2g^{-1}. This is not a great deal larger than the RMSEP from the four geochemical predictors (6.3 $\text{m}^2 \text{ g}^{-1}$).

Previous studies have shown positive correlations between suspended sediment SSA and total phosphorus in lowland streams of England (Evans et al., 2004); we found a small positive correlation between log total P and estimated sediment SSA for 1792 sites in headwater catchments. Our high-resolution estimates of SSA could help to highlight catchments which have the potential to store large quantities of phosphorus in their fine bed sediment (Ballantine et al., 2009) or act as sources of fine sediment to larger systems, aiding catchment appraisal and the identification of areas for targeted management (Mainstone et al., 2008).

6. Conclusions

The main conclusions from our study are:

1. The total concentrations of four elements are sufficient to estimate reasonably accurately the BET SSA of the mineral component of fine bed sediments in predominantly agricultural headwater catchments covering 15 400 km^2 of central England. The four geochemical predictors – in order of decreasing predictive power – were: V>Ca>Al>Rb. We believe V is the most significant predictor because it occurs in the structure of both Fe-oxides and certain clay minerals.

2. Our model accounted for 73% (adjusted R^2) of the variation in fine ($< 150\mu$m) bed sediment SSA; the standard error was 5.3 m^{2}g^{-1} across the range 5.98 to 46.0 m^{2}g^{-1}. The RMSEP from leave-one-out cross validation was 6.7 m^{2}g^{-1} with a small negative bias (-0.32 m^{2}g^{-1}).

3. Sample heterogeneity was a significant source of variation in our fine bed sediment SSA analyses; based on triplicate analyses of six samples the standard error was 2.2 m^{2}g^{-1} with a coefficient of variation 8.4%. This variation cannot be accounted for by our linear regression model.

4. Based on SSA predictions for 30 paired sites – where duplicate samples from across the entire study area were collected at 25 m from their original sampling sites – the standard error from a one-way analysis of variance for SSA at these paired sites was was 2.7 m^{2}g^{-1} (coefficient of variation of 12.3%). This suggests that the majority of variation in bed sediment SSA occurs either between headwater catchments or at larger scales along stream reaches, not locally (<25 m) within stream beds.

5. Based on the selection of catchments with a dominant bedrock formation or land cover type ($> 50\%$ of the catchment area for each sediment site), these factors accounted for 39 and 7% respectively of the variation in bed sediment SSA across the study region. Catchments dominated by arable land use at this regional scale had (statistically significant) larger bed sediment SSAs than those in grassland dominated catchments.

6. For the sixty sediments where both SSA and TOC were measured, there was no linear correlation between them (Pearson correlation $r=0.02$). There was a small positive correlation between log transformed total sediment phosphorus and estimated bed sediment SSA (Pearson correlation $r=0.06$).

Acknowledgements

References

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

Appleton, J.D., Rawlins, B.G., Thornton, I. 2008. National-scale estimation of potentially harmful element ambient background concentrations in topsoil using parent material classified soil:streamsediment relationships. Appl. Geochem. 23, 2596–2611.

Ballantine, D.J., Walling, D.E., Collins, A.L, Leeks, G.J.L. 2009. The content and storage of phosphorus in fine-grained channel bed sediment in contrasting lowland agricultural catchments in the UK. Geoderma, 151, 141–149.

Ben-Dor, E., Banin, A. 1995. Near-infrared analysis as a rapid method to simultaneously evaluate several soil properties. Soil Sci. Soc. Am. J. 59, 364–372.

Breit, G. N., and Wanty, R.B. 1991. Vanadium accumulation in carbonaceous rocks. A review of geochemical controls during deposition and diagenesis. Geochim. Cosmochim. Acta. 91, 83–97.

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

British Geological Survey 2006. *Digital Geological Map of Great Britain 1:50 000 scale (DiGMapGB-50) data [CD-ROM] Version 3.14.* British Geological Survey, Keyworth, Nottingham.

Brunauer, S., Emmett, P. H., Teller, E. 1938. Adsorption of gases in multimolecular layers. J. Am. Chem. Soc. 60, 309–319.

Cardinale, B.J., Palmer, M.A., Swan, C.M., Brooks, S., Poff, N. L. 2002. The influence of substrate heterogeneity on biofilm metabolism in a stream ecosystem. Ecology, 83, 412–422.

Evans, D.J., Johnes, P.J., Lawrence, D.S. 2004. Physico-chemical controls on phosphorus cycling in two lowland streams. Part 2 - The sediment phase. Sci. Tot. Environ. 329, 165–182.

Fullen, M.A. 1997. A comparison of runoff and erosion rates on bare and grassed loamy sand soils. Soil Use Manage. 7, 136–139.

Fuller, R.M., Smith, G. M., Sanderson, J. M., Hill, R.A., Thomson, A.G. 2002. The UK Land Cover Map 2000: Construction of a parcel-based vector Map from satellite images. Cartogr. J. 39,15–25

Galy, V., France-Lanord, C., Lartiges, B. 2008. Loading and fate of particulate organic carbon from the Himalaya to the Ganga-Brahmaputra delta. Geochim. Cosmochim. Acta. 72, 1767–1787.

Horowitz, A.J., Elrick, K. A. 1987. The relation of stream sediment surface area, grain size and composition to trace element chemistry. Appl. Geochem. 2, 437–451.

Intermap (2009). NEXTMap Britain. Intermap. http://www.intermap.com/nextmap-britain . Accessed 6th December 2009.

Johnson, C. C., Breward, N., Ander, E. L., Ault, L. 2005. G-BASE: Baseline geochemical mapping of Great Britain and Northern Ireland. Geochem. Explor. Environ. Anal. 5, 1–13.

Mainstone, C.P., Dils, R.M, Withers, P.A.J. 2008. Controlling sediment and phosphorus transfer to receiving waters - A strategic management perspective for England and Wales. Journal of Hydrology, 350, 131–143.

Mayer, L. M. 1994. Surface area control of organic carbon accumulation in continental shelf sediments. Geochim. Cosmochim. Acta. 58, 1271–1284.

Owens, P.N., Batalla, R.J., Collins, A.J., Gomez, B., Hicks,D. M., Horowitz, A.J., Kondolf,G.M., Marden, M., Page,M.J., Peacock,D.H., Petticrew,E.L., Salomons, W., Trustrum, N.A. 2005. Fine-grained sediment in river systems: Environmental significance and management issues. River Res. Appl. 21, 693–717.

Petersen, L.W., Moldrup, P., Jacobsen, O.H., Rolston, D.E. 1996. Relations between specific surface area and soil physical and chemical properties. Soil Sci. 161, 9–21.

Rawlins, B.G., Webster, R., Tye, A.M., Lawley, R., OHara, S.L., 2009. Estimating particle-size fractions of soil dominated by silicate minerals from geochemistry. Eur. J. Soil Sci. 60, 116–126

Salminen, R., Batista, M.J., Bidovec, M, Demetriades, A., De Vivo, B., De Vos, W. *et al.* 2005. *Geochemical Atlas of Europe, Part 1: Background Information, Methodology and Maps.* Geological Survey of Finland, Espoo, Finland.

583 Schwertmann, U., Pfab, G. 1996. Structural vanadium and chromium in lateritic iron
584   oxides: Genetic implications. Geochim. Cosmochim. Acta. 60, 4279–4283.

585 Soil Survey of England and Wales, 1983a. *Soils and their Use in Midland and Western*
586   *England.* Ordnance Survey for the Soil Survey of England and Wales, Southamp-
587   ton.

588 Soil Survey of England and Wales, 1983b. *Soils and their Use in Eastern England.*
589   Ordnance Survey for the Soil Survey of England and Wales, Southampton.

590 Sylvester-Bradley, P.C. and Ford, T.D. 1968. *The geology of the East Midlands.*
591   Leicester University Press, 400pp.

592 van den Boogaart, K. G., Tolosana, R., Bren, M. 2008. compositions: Compositional
593   Data Analysis. R package version 1.01-1. http://www.stat.boogaart.de/compositions.

594 Wagai, R., Mayer, L. M., Kitayama, K. 2009. Extent and nature of organic coverage
595   of soil mineral surfaces assessed by a gas sorption approach. Geoderma, 149,
596   152–160.

597 Wang, F., Chen, J., Chen, J., Forsling, W. 1997. Surface properties of natural aquatic
598   sediments. Wat. Res. 31, 1796–1800.

599 Withers, P.J.A., Jarvie, H. P. 2008. Delivery and cycling of phosphorus in rivers: A
600   review. Sci. Total Environ. 400, 379–395.

Figure captions

Figure 1 Stream sediment sampling locations across central England and the location of the 60 sites (black discs) where SSA was measured and the 1792 sites where SSA was estimated from geochemistry (open discs). Thirty sites where duplicate samples were collected are also shown (grey discs). Coordinates are kilometres of the British National Grid.

Figure 2 . Simplified map of bedrock geology across the study region; approximate scale 1:1.1 million.

Figure 3 Projections of the correlations between variables and the principal component (PC) scores in unit circles: top component 2 against component 1; bottom component 3 against component 1. The percentage variance accounted for by each PC is shown in Table 2.

Figure 4 Measured and predicted SSA (m^2g^{-1}) of fine stream bed sediments at 56 selected sites.

Figure 5 The spatial distribution of fine bed sediment SSA (m^2g^{-1}) in selected (n=1792) first and second order streams across the study region. Stream reaches are depicted upstream from each sediment sampling site based on the sub-catchments derived from a Digital Elevation Model. Coordinates are metres of the British National Grid.

Figure 6 Box and whisker plot of fine bed sediment SSA (n=1236) for twenty geological formations across the study region. There were between 12 and 434 sediment sampling sites where these bedrock formations accounted for $> 50\%$ of the area in their drainage catchment.

Table 1 Summary statistics of 21 elements (units $\mathrm{mg\,kg^{-1}}$) in fine bed stream sediments

(n=1972), and after centred log-ratio transformation.

Element	Mean	Median	St. Dev.	Skew	Centred log-ratio Mean	Centred log-ratio St. Dev.	Centred log-ratio Skew
Al	80430	79879	14143	-0.1	7.4	0.2	-1.0
As	20.8	16.7	16.1	5.4	-1.0	0.6	0.5
Ba	576	402	2027	25.8	2.3	0.4	2.4
Ca	16062	13007	10312	0.8	5.6	0.7	-0.1
Co	19.6	16.7	14.7	11.9	-1.0	0.4	1.5
Cr	108.0	103.0	30.7	2.3	0.8	0.2	0.4
Fe	52538	48363	23296	4.1	7.0	0.3	0.6
Ga	15.0	14.9	3.6	-0.1	-1.2	0.3	-3.4
Hf	9.8	8.3	15.1	35.6	-1.7	0.5	0.0
K	20371	18594	5811	0.9	6.0	0.3	0.2
Mg	12744	7840	10887	2.0	5.4	0.7	0.7
Mn	11723	782	3364	27.6	2.9	0.6	1.8
Ni	46.5	43.2	27.5	11.3	-0.1	0.3	0.4
Rb	90.8	90.7	19.0	0.0	0.6	0.3	-3.0
Si	246460	244708	22346	-0.5	8.6	0.2	-0.3
Sr	99.5	92.8	85.5	31.7	0.7	0.3	1.2
Th	10.6	10.4	2.5	7.2	-1.5	0.3	-4.5
Ti	4719	4706	689	0.1	4.6	0.2	-0.8
U	2.5	2.4	1.4	7.5	-3.1	0.5	-1.4
V	130.	124	47.3	1.7	1.0	0.3	0.1
Zr	415	355	249	3.5	2.1	0.5	0.0

626 **Table 2** Leading eigenvalues and cumulants from principal component analysis of the

627 correlation matrix of 21 elements after centred log-ratio transformation.

Order	Eigenvalue	Percentage of variance	Cumulative percentage
1	5.669	27.1	27.1
2	4.432	21.1	48.2
3	2.717	12.9	61.1
4	2.088	9.9	71.0

628

Table 3 Description, number, and selection criteria for stream sediment sites used in components of this study

Row	Sediment group description	Number of sites	Selection criteria	From row
1	Regional survey	5047		
2	Selection	1972	flow conditions, stream order, channel size	1
3	SSA determined by BET analysis	60	range of total Al concentrations	2
4	Regression model	56	excluding sites with large Mo concentrations	3
5	SSA estimated using regression model	1792	excluding sites with large Mo concentrations	2
6	Dominant bedrock type > 50% catchment area	1340		3 and 5
7	Dominant habitat type > 50% catchment area	415		3 and 5
8	Duplicate sediment sites	65	see text	1
9	Selected duplicate sediment sites	30	flow conditions, stream order, channel size	8

632 **Table 4** Regression coefficients of multiple linear regression model for BET specific
633 surface area using the original (untransformed) geochemical data (units mg kg $^{-1}$).

634

Element	Estimate	Std error	t	P
Intercept	-1.459	4.633	-3.148	0.003
Al	23.05×10^{-5}	70.50×10^{-6}	3.269	19.4×10^{-4}
Ca	-29.95×10^{-5}	59.36×10^{-6}	-5.046	60.9×10^{-7}
Rb	0.099	0.059	1.675	0.100
V	0.116	15.10×10^{-3}	7.667	48.3×10^{-11}

635 **Table 5**. Results from one-way analysis of variance of sediment SSA for (a) bedrock
636 formation (n=1338), and (b) land cover class (n=413).

	Degrees of freedom	Sum of squares	Mean square	F ratio
(a)				
Bedrock	65	38034	585	12.4
Residual	1273	60019	47	
Total	1338	98053		
(b)				
Land cover	7	2369	38.4	4.41
Residual	406	31166	76.8	
Total	413	33535		

637
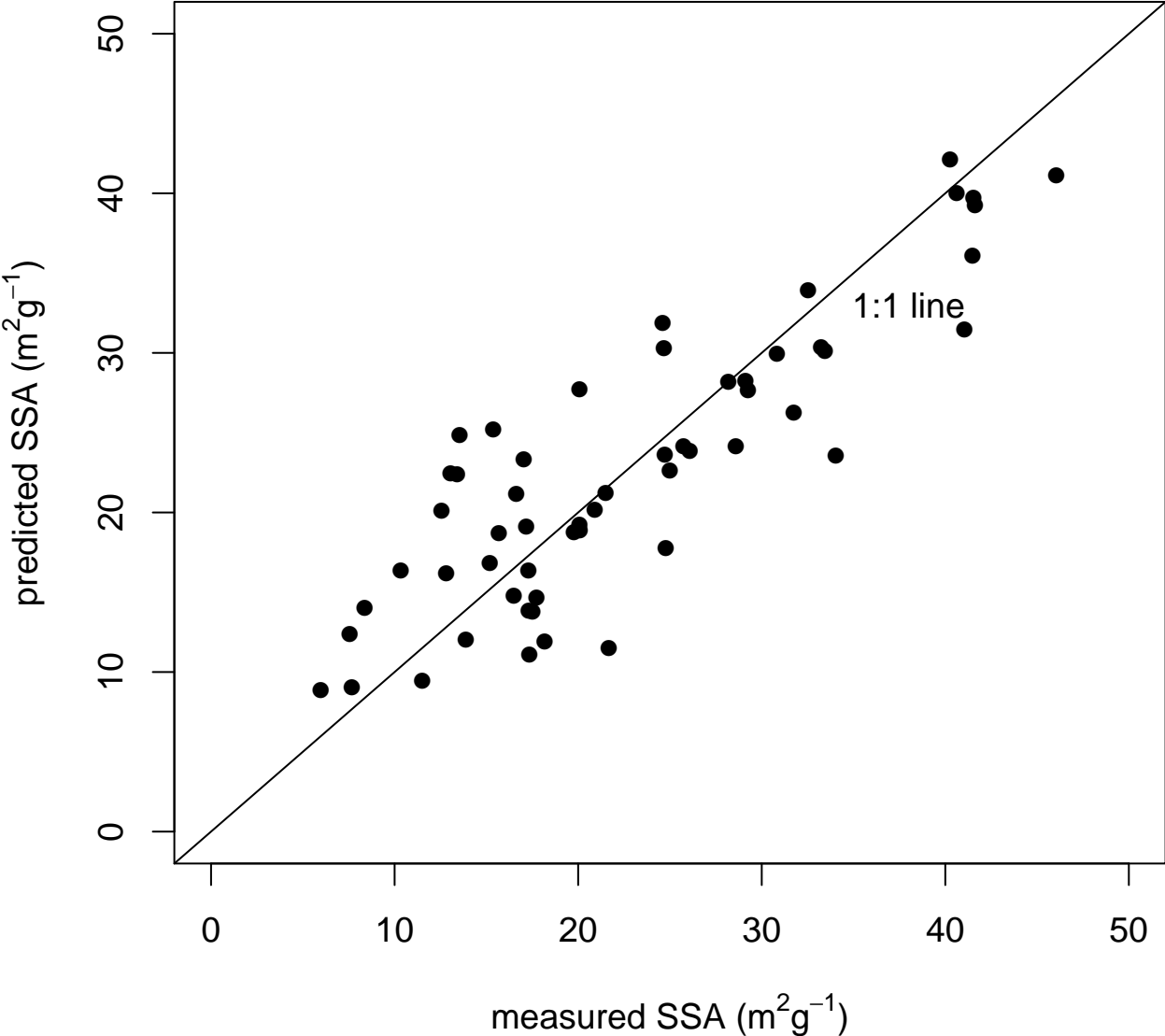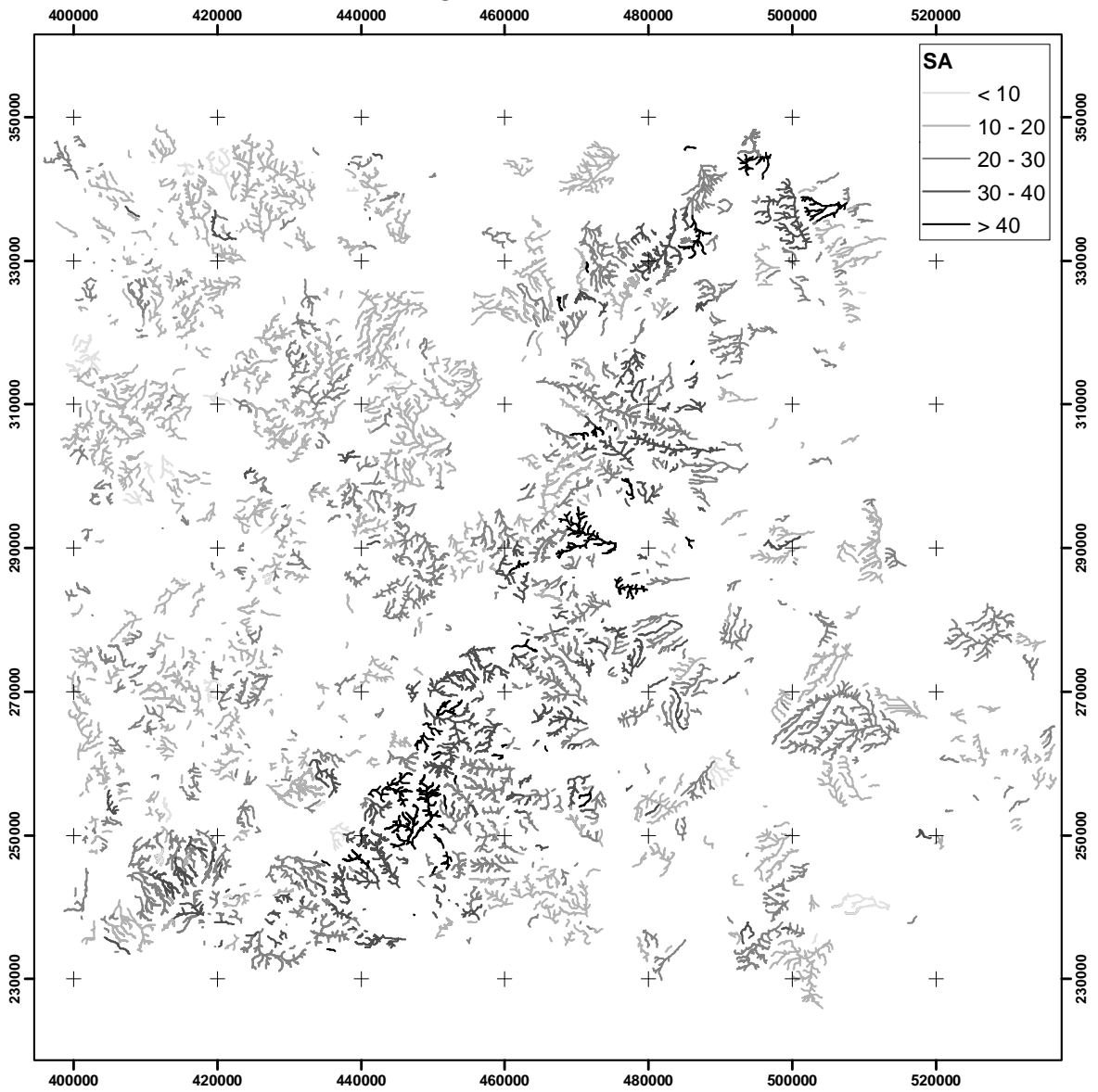
Figure 1:

1.pdf

Figure 2:

Scale is approximately 1:140,000

Cretaceous Chalk

Gault Formation and Upper Greensand Formation

West Walton, Ampthill Clay and Kimmeridge Clay Formations

Kellaways and Oxford Clay Formation

Great Oolite Group (including Blisworth Clay)

Inferior Oolite Group (including Northampton Sand)

Lias Group (including Dyrham, Blue Lias, Whitby Mudstone and Marlstone Rock)

Bowland High Group and Craven Group (including Bowland Shale)

Dinantian Rocks (Limestone)

Lower Greensand Group

Millstone Grit

Pennine Coal Measures Formation

Triassic Rocks (Mercia Mudstone and Bromsgrove Sandstone)

Warwickshire Group

Zechstein Group

Figure 3:

33

4.pdf

Figure 5:

Figure 6:

6.pdf