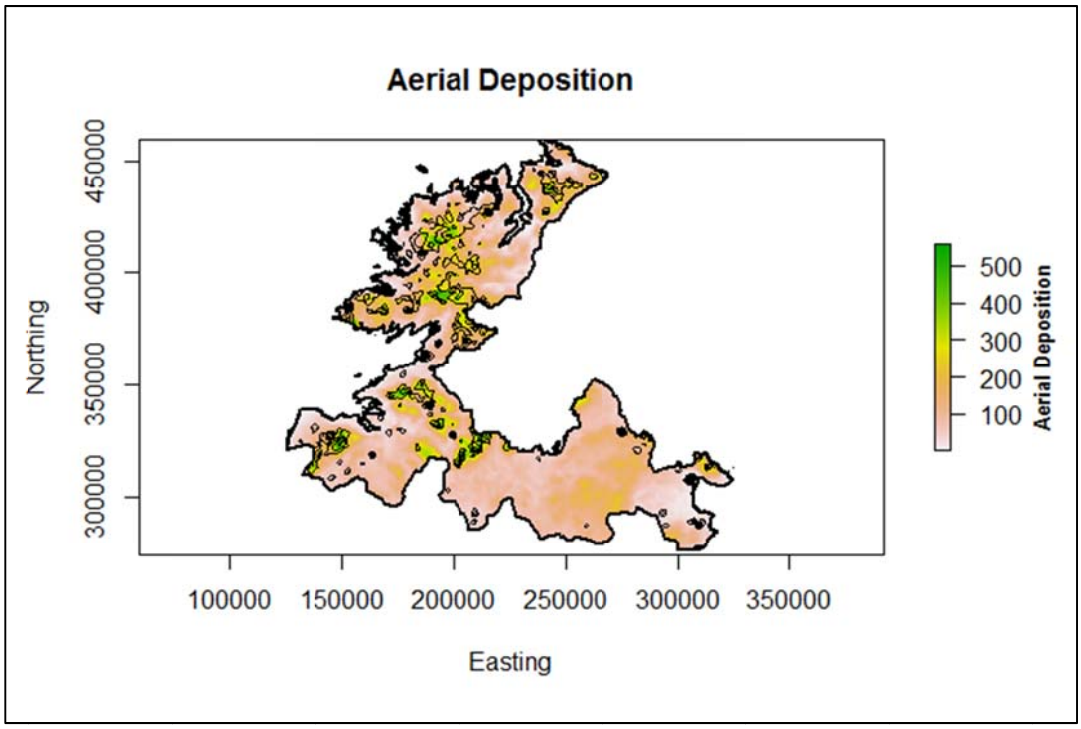
British Geological Survey
NATURAL ENVIRONMENT RESEARCH COUNCIL

TELLUS BORDER
Understanding Underground

European Union
European Regional Development Fund
Investing in your future
A project supported by the EU's INTERREG IVA Programme managed by the Special EU Programmes Body

End of Project Report: Application of the Tellus Border soil chemistry data to the identification of sources of diffuse pollution in Ireland



Tellus Border is a €5 million cross-border project to map the environment and natural resources in the border region of Ireland and continue the analysis of data in the border counties of Northern Ireland. It is a joint initiative between the Geological Survey of Northern Ireland, the Geological Survey of Ireland, Queen's University, Belfast and Dundalk Institute of Technology.

This research is supported by the EU INTERREG IVA-funded Tellus Border project managed by the Special EU Programmes Body (SEUPB) .

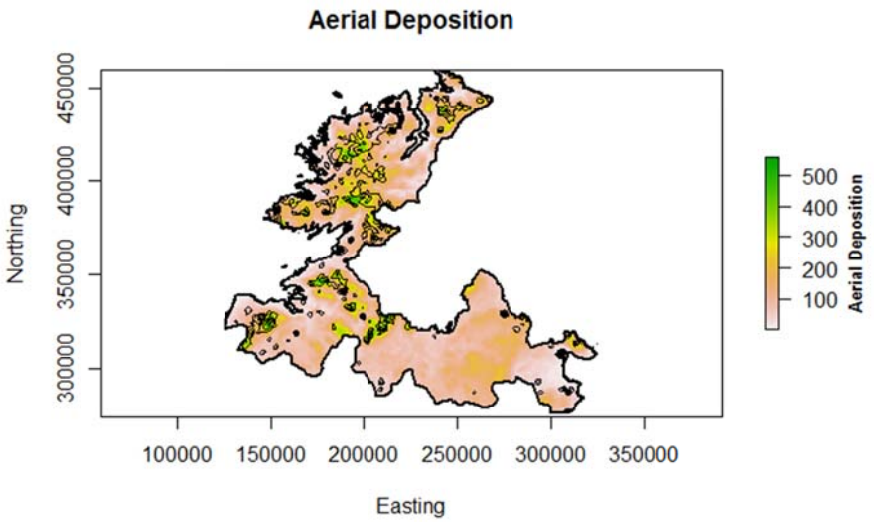
The views and opinions expressed in this research report do not necessarily reflect those of the European Commission or SEUPB.

BRITISH GEOLOGICAL SURVEY

Geochemistry Baselines and Medical Geology Team

End of Project Report: Application of the Tellus Border soil chemistry data to the identification of sources of diffuse pollution in Ireland

Cave MR and Ander EL



Bibliographical reference

CAVE MR AND ANDER EL. 2013. End of-Project Report: Application of the Tellus Border soil chemistry data to the identification of sources of diffuse pollution in Ireland *Report by the British Geological Survey,Keyworth, Nottingham*. 54pp.

Keyworth, Nottingham British Geological Survey 2013

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

Environmental Science 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

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

Contents

Contents ..... **i**

1 Introduction **1**

2 Tasks **3**

2.1 Stakeholder Engagement 3

2.2 Literature Search 6

2.3 Statistical Methods 6

3 Summary of Findings **9**

3.1 EDA 9

3.2 SMMR 9

3.3 Identification of Diffuse Pollution 15

3.4 Random Forest Modelling 17

4 Conclusions **25**

5 Suggestions for future research **26**

6 Report references **27**

FIGURES

Figure 1: Project Gantt chart 1

Figure 2: Title slide of PowerPoint presentation prepared for stakeholder meetings 3

Figure 3 PowerPoint presentation given at the Tellus Border Conference 4

Figure 4 Poster presented at the Tellus Border Conference 5

Figure 5 Map of the Tellus border sampling locations 7

Figure 6 Comparison of element distribution shapes 7

Figure 7 Pictorial representation of the SMMR procedure 8

Figure 8 Spearman correlation of the mass of material associated with each component to populaton, LOI, pH and Elevation 12

Figure 9 Mass distribution of Pb between the SMMR source components 13

Figure 10 Mass distribution of Sb between the SMMR source components 13

Figure 11 Mass distribution of Hg between the SMMR source components 14

Figure 12 Mass distribution of Cd between the SMMR source components 14

Figure 13 Mass distribution of As between the SMMR source components 15

Figure 14 Mass distribution of U between the SMMR source components 15

Figure 15 Suspected anthropogenic component (black contour lines) superimposed on the elevation data (colour ramp) showing a close spatial match 16

Figure 16 Plot actual versus predicted log transformed concentrations for the six anthropogenic pathfinder elements along with the line of equivalence 17

Figure 17 Univariate plots showing the behaviour of Pb with respect to geographical variables and soil properties 19

Figure 18 Univariate plots showing the behaviour of Sb with respect to geographical variables and soil properties 20

Figure 19 Univariate plots showing the behaviour of Hg with respect to geographical variables and soil properties 21

Figure 20 Univariate plots showing the behaviour of Cd with respect to geographical variables and soil properties 22

Figure 21 Univariate plots showing the behaviour of As with respect to geographical variables and soil properties 23

Figure 22 Univariate plots showing the behaviour of U with respect to geographical variables and soil properties 24

TABLES

Table 1: Summary status of Project items 2

Table 2 Summary of Geological sources identified in the soil chemistry, the names of the sources are made up from those elements that make up more than 10% of the source composition 10

Table 3 Summary of sources identified in the soil chemistry related to secondary processes, the names of the sources are made up from those elements that make up more than 10% of the source composition 10

Table 4 Summary of sources identified in the soil chemistry related to peat bogs, the names of the sources are made up from those elements that make up more than 10% of the source composition 11

Table 5 Summary of sources identified in the soil chemistry whose sources are less clearly identified, the names of the sources are made up from those elements that make up more than 10% of the source composition 11

1 Introduction

This report is a required deliverable for the BGS project to investigate the "Discriminating diffuse anthropogenic input of selected contaminants to soil in the Tellus Border area" as detailed in the Project Plan (Tellus Border project reference 10761).

For the purposes of this report diffuse pollution is defined as pollution derived from widespread activities with no one discrete source, e.g. aerial deposition of particulates, agricultural practices, vehicular emissions etc.

The scientific aim of this project is:

-To identify geogenic and anthropogenic chemical signatures and their potential sources (e.g. geology, land use, industry) within the Tellus Border soils data;

This has been achieved by meeting the following objectives:

- To develop a robust data processing methodology using data mining, spatial statistics and GIS methodologies which gives both spatial location information and quantitative compositional data on the underlying chemical signatures;

- To give a preliminary interpretation on the identified chemical signatures

- To produce a peer-reviewed scientific publication.

The Project Gantt chart is shown in Figure 1. The purpose of this report is to provide the evidence for completion of tasks detailed in Figure 1. This is summarised in Table 1.

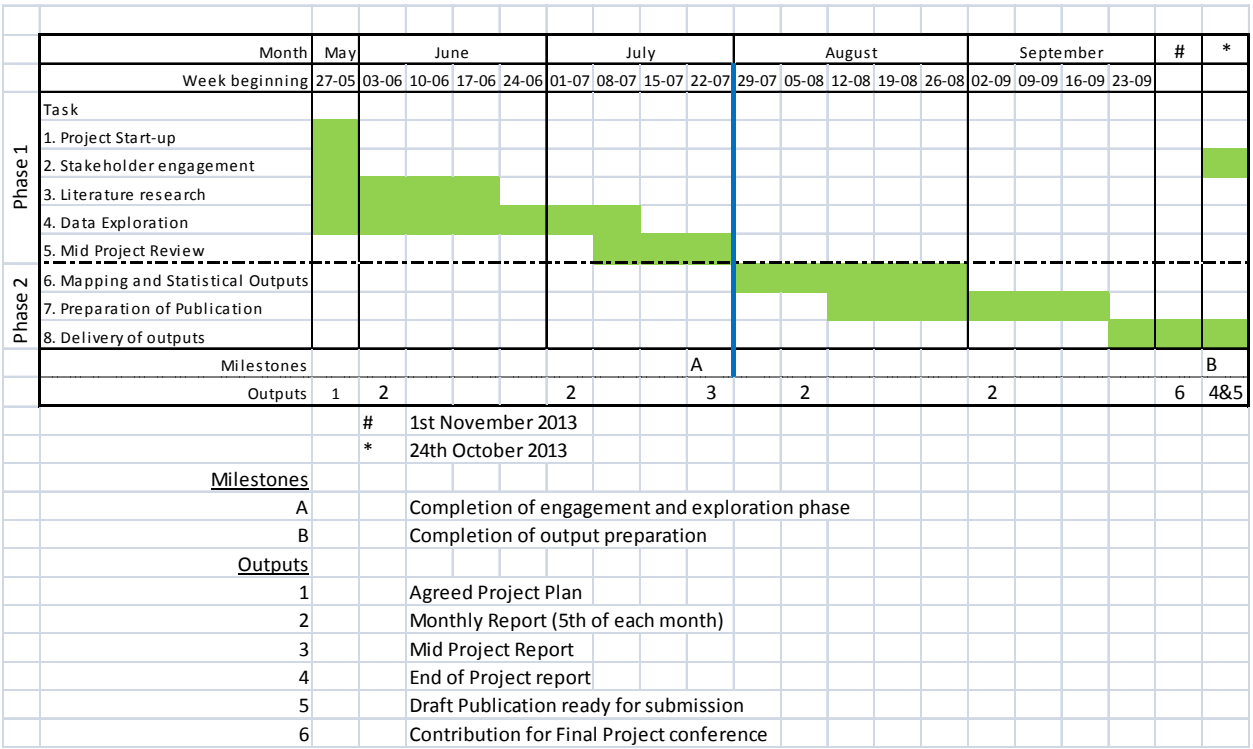


Figure 1: Project Gantt chart

Table 1: Summary status of Project items

Item	Status	Comment
TASKS: 1. Project Start-up	COMPLETE	Videoconference 29th April 2013
2. Stakeholder engagement	COMPLETE	Completed end of May with visit to Ireland .
3. Literature search	COMPLETE	Initial phase complete though as publication is prepared new references will be added to bibliography
4. Data exploration	COMPLETE	Tellus Border data received and uploaded to ArcGIS project. Suitable elements explored.
5. Mid project review	COMPLETE	Teleconference 31st June 2013
6. Mapping and statistical outputs	COMPLETE	Following on from the data exploration the statistical and mapping outputs for the project have been clarified and are now being produced (see examples in the mid-project report)
7. Preparation of publication and delivery of outputs	COMPLETE	On schedule to be completed in November
OUTPUTS: 1. Agreed Project Plan	COMPLETE	Approved 10/05/2013 – email from Mairéad Glennon
2. Monthly reports	COMPLETE	Complete .
3. Mid-project report	COMPLETE	This report, submitted 26th July 2013
4. End-project report	COMPLETE	Initially to be prepared in September, final version to be received by 1st November 2013
5. Draft publication	ONGOING	Paper drafting is ongoing.
6. Contribution to Tellus Border Conference	COMPLETE	Undertaken 24th October 2013
MILESTONES: A. Completion of engagement and exploration phase	COMPLETE	Completion coincident with mid-project review
B. Delivery of all outputs	COMPLETE	This report completes required outputs.

2 Tasks

2.1 STAKEHOLDER ENGAGEMENT

The main phase of engagement was carried out 29th & 30th May 2013 when the BGS team (Chris Johnson, Louise Ander and Mark Cave) visited Dublin and Co. Wexford. During this time meeting were held as follows:

- 29th May 2013: GSI team involved in Tellus Border Geochemistry Task (Ray Scanlon, Mairéad Glennon and Kate Knights)
- 30th May 2013: Rachel Creamer, Soil Quality Research Officer, Teagasc, Johnstown Castle, Wexford
- 30th May 2013: Jane Brogan, Environmental Protection Agency, Johnstown Castle, Wexford

During the time in Dublin the team also got a chance to meet Howard Fox and Maria Cullen at the GSI to discuss their Tellus Border Research project on lichens and how these may relate to the chemical surface environment.

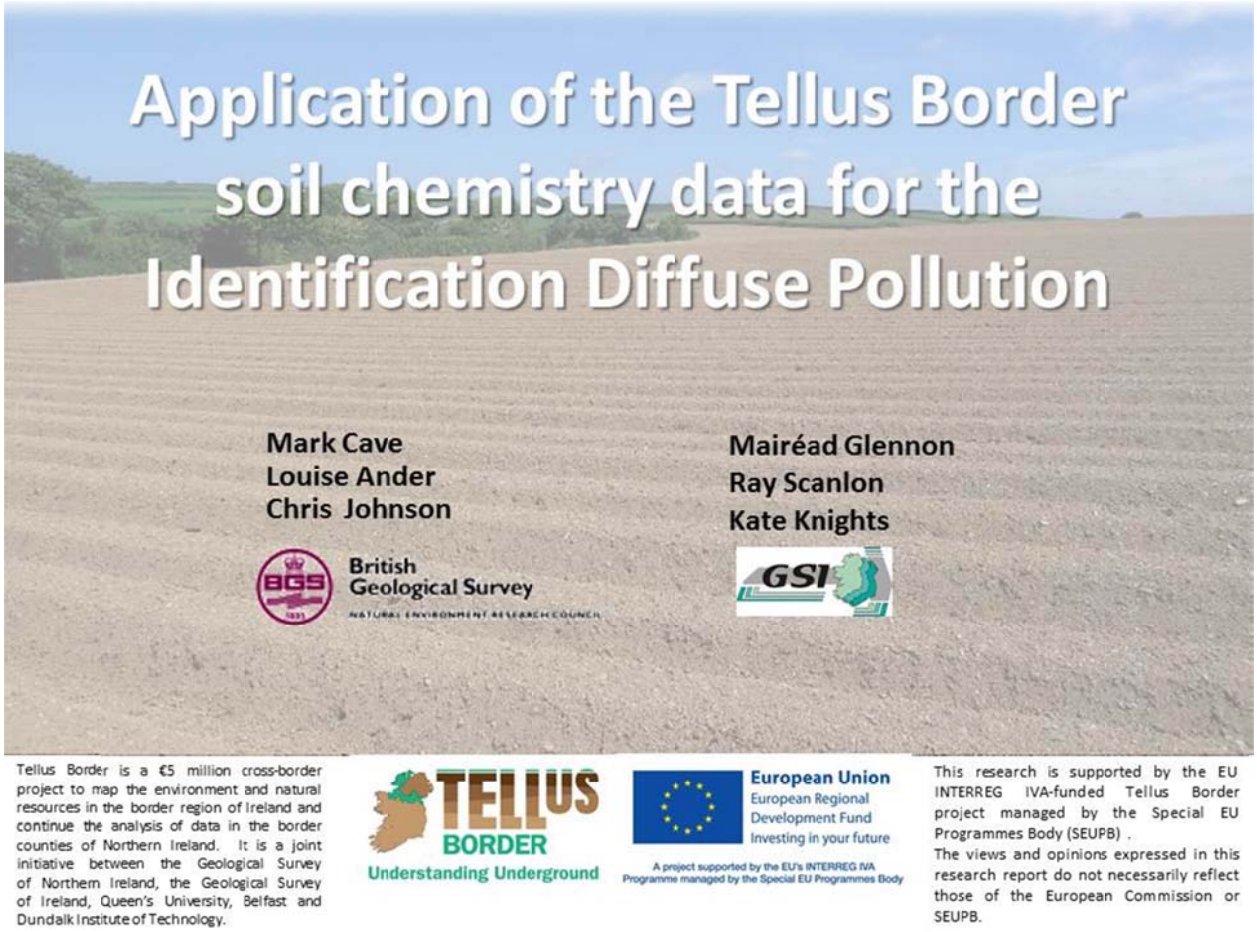


Figure 2: Title slide of PowerPoint presentation prepared for stakeholder meetings

At all meetings a short PowerPoint presentation was given by Mark Cave (Figure 2) which showed some preliminary exploratory analysis and outlined the approach to be used on the data. This presentation has been made available to staff from GSI and GSNI.

The final stakeholder engagement was at the Tellus border Conference which took place at Monaghan, Co. Monaghan on 24th October 2013. A four minute PowerPoint presentation was given outlining the project and its findings, with greater detail given in an accompanying poster (Figure 3 and Figure 4)

Figure 3 PowerPoint presentation given at the Tellus Border Conference

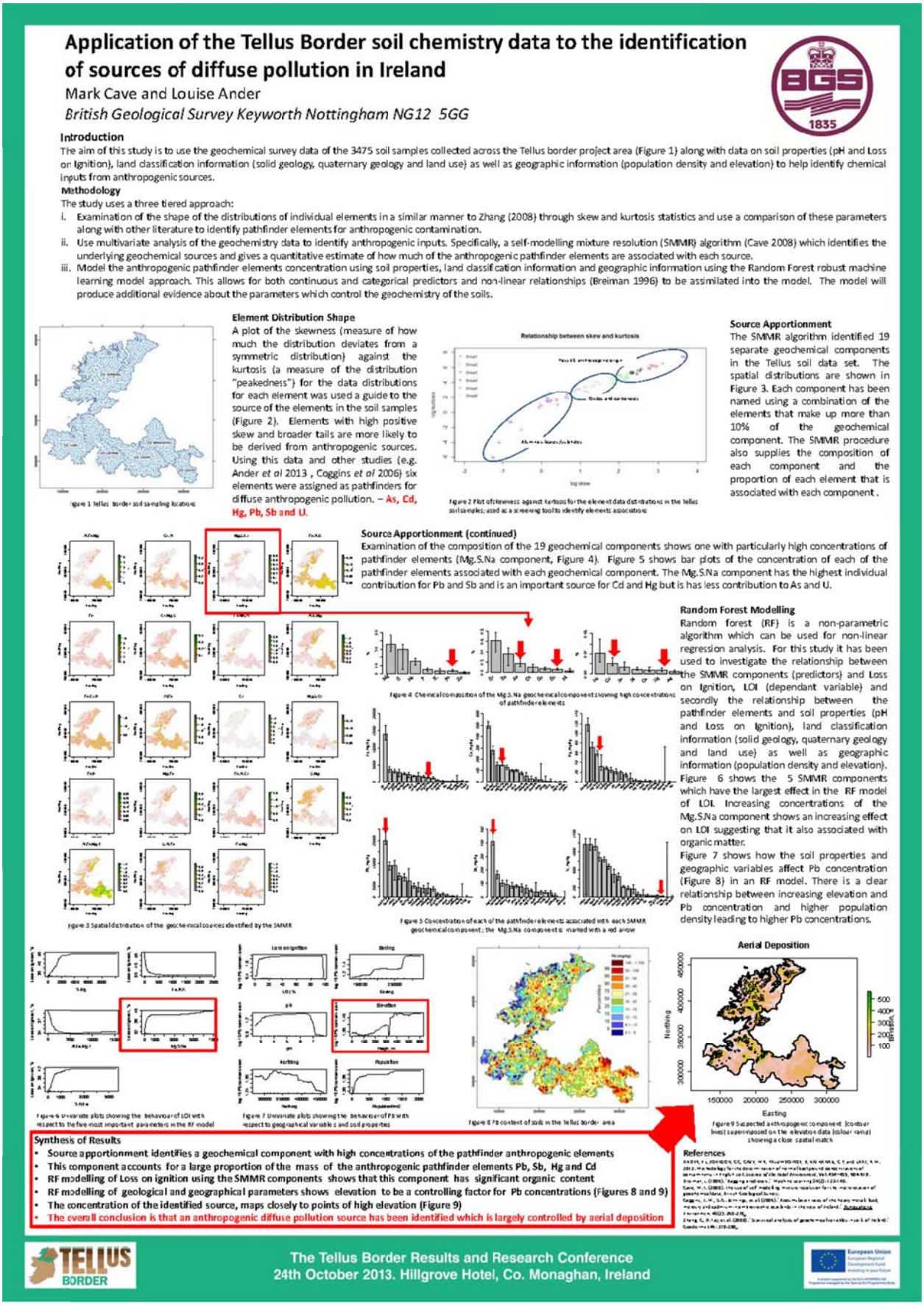


Figure 4 Poster presented at the Tellus Border Conference

2.2 LITERATURE SEARCH

An important part of the project has been to gather together literature relevant to the application of geochemical survey soil chemical data, with a specific focus on the Republic of Ireland.

Without producing a full review of the literature here, a very brief overview of the findings on the literature showed that some of the older data on potentially harmful elements in soils provides some useful indicators, but the most relevant publications which use a multivariate approach to inorganic analysis of soils in Ireland are the studies of Dempster et al. (Dempster et al., 2013) and the work of Zhang et al. (Zhang, 2006; Zhang et al., 2008). Zhang's studies on the soils of Galway (Zhang, 2006) and one using the national soil data base (Zhang et al., 2008) make particular use of the comparison of element distribution shapes to categorise elements followed by multivariate statistical methods, including principal component analysis and cluster analysis. In their study on soils in Galway, they identified that Cu, Pb, Zn and As were coming principally from anthropogenic sources. In the study of the national soil database they identified Ni, Zn, Cd, Y, Mo, Hg, Pb and Sb as probably coming from anthropogenic sources.

This diffuse pollution study used a similar approach to that successfully applied by Zhang to study Irish soils. An important difference is that the self-modeling mixture resolution methodology described below, using principal component analysis and cluster analysis as part of the modeling process, has an end point which is a quantitative rather than a qualitative output.

2.3 STATISTICAL METHODS

This first step has been carried out as follows. R code was written to read in and select the appropriate element data from the Tellus Border aqua-regia soil extractions data set (Figure 5). Only those elements with more than 60% of the data greater than detection limit were chosen. This consisted of 45 elements (Al, Ba, Ca, Cr, Cu, Fe, K, Li, Mg, Mn, Na, Ni, P, S, Sr, V, Zn, Zr, Ag, As, Be, Bi, Cd, Ce, Co, Cs, Ga, Hg, In, La, Lu, Mo, Nb, Pb, Rb, Sb, Sc, Se, Sn, Tb, Th, Tl, U, Y, Yb).

An initial study of the shape of the data distributions for each element was carried out using a combination the skew, octile skew and kurtosis of the data distributions. A visual plot of the element groupings was obtained through an R code program, Figure 6.

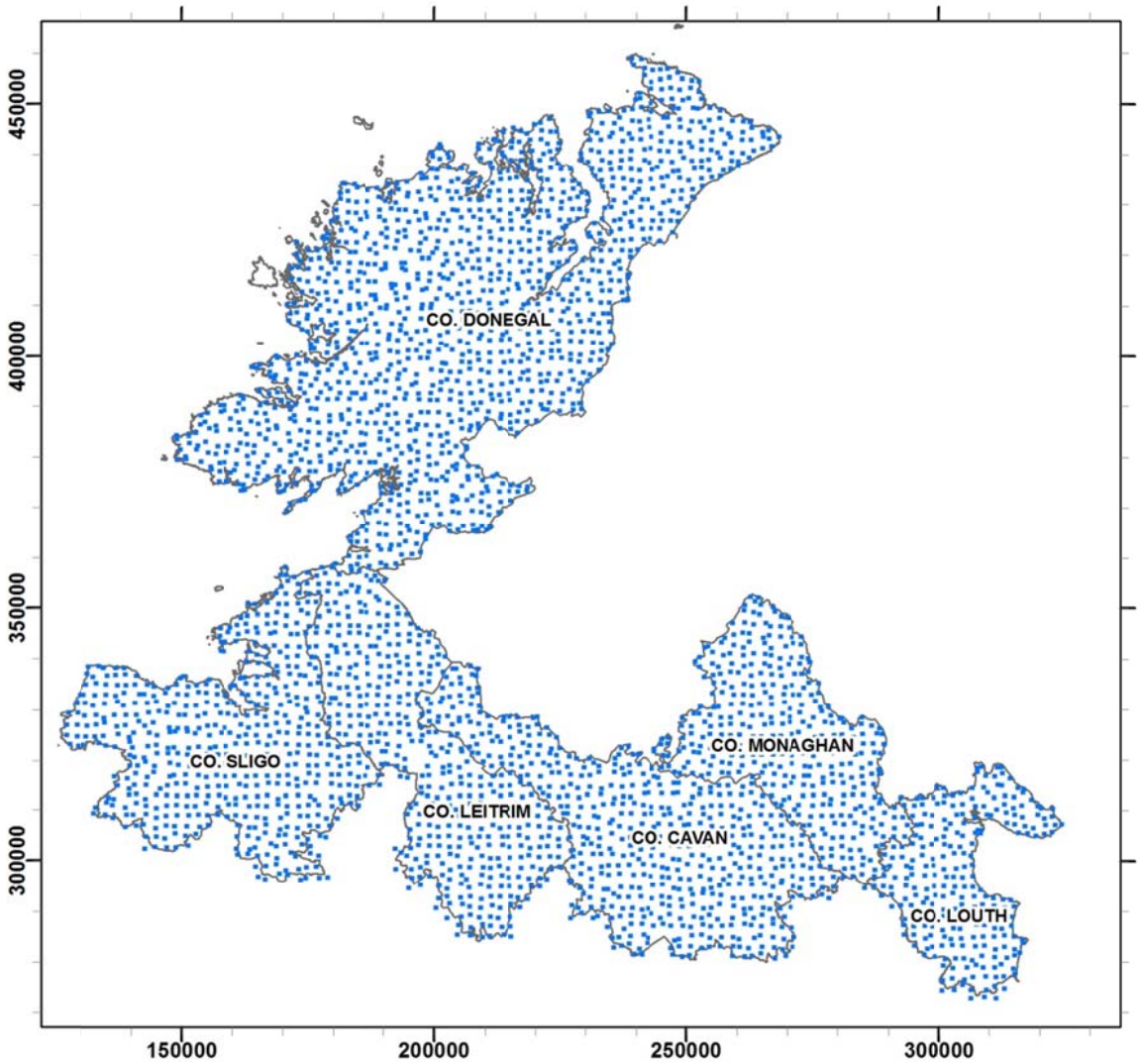


Figure 5 Map of the Tellus border sampling locations

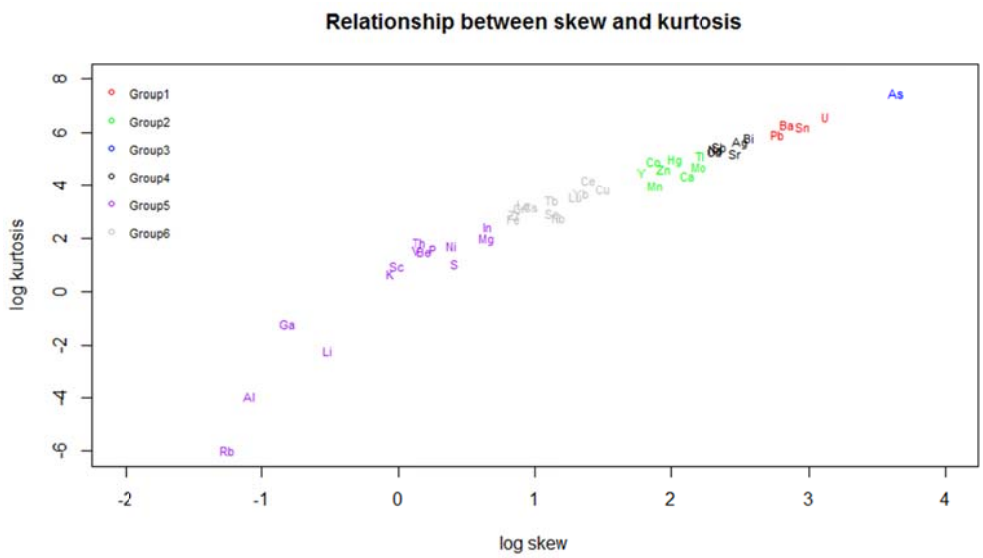


Figure 6 Comparison of element distribution shapes

Moving from the bottom left to the top right of Figure 6 the data distributions change from being broad symmetrical distributions to narrower more positively skewed distributions with possible outliers in the distribution tail. The elements in the bottom left are more associated

with geogenic, silicate and sulphide sources, those in the centre with carbonates and iron oxides and those to the top right with long right hand tails are more likely to have a diffuse pollution origin.

2.3.1 Statistical Analysis

2.3.1.1 Self-modelling mixture resolution (SMMR)

The SMMR methodology is fully described by (Cave, 2009) but a brief overview follows.

As the terminology ''self-modelling'' implies, SMMR, in principle, does not require a priori any specific information concerning the data to resolve the pure variables. The basis for these techniques can be explained by reference Figure 7.

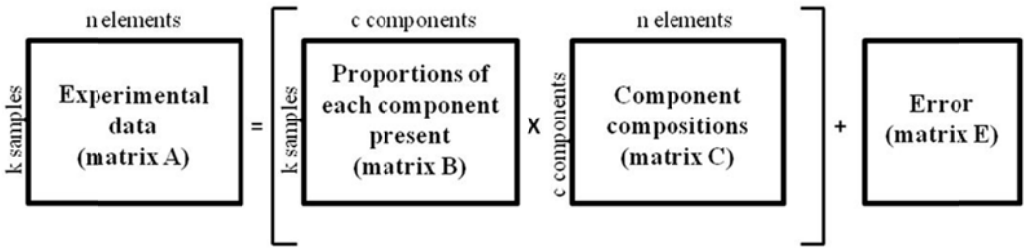


Figure 7 Pictorial representation of the SMMR procedure

The data matrix A represents the experimental mixture data (in this case the combined total element data for the Tellus border soil samples consisting 3475 samples by 45 elements). The aim of SMMR is to arrive at the pure component information consisting of the proportions matrix (B) and the component composition matrix (C) without any other information apart from that contained in the experimental data (A). In matrix notation this is:

$$A=BC + E \qquad \textit{Equation 1}$$

The method uses Varimax rotation of the Principal Component Analysis (PCA) scores matrix, of the soils elemental compositions data (matrix A, equation 1), as a first approximation of the pure components proportions matrix. This is followed by iterative refining of the pure components matrices (B and C, equation 1), applying non-negativity and mass balance constraints at each step.

2.3.1.2 Random Forest modelling

Random Forest (RF) is an ensemble method that was developed as an extension of classification and regression trees (CART) to improve the prediction performance of the model (Breiman, 1996). The model building process is the same as in CART with the difference that many trees are built resulting in a "forest of models". For each tree, only a subset of the predictor variables is used. The number of predictors to be used in each tree-building process and the number of trees to be built in the forest can be varied depending on the data set. Each tree is built from a bootstrap sample of the original data set which allows for robust error estimation.

The advantages that arise from this procedure are (a) higher prediction performance, (b) no overfitting, (c) low correlation of individual trees since the diversity of the forest is increased through the usage of a limited number of predictors, (d) low bias and low variance due to averaging over a large number of trees and (e) robust error estimates. In addition RF models can use the pH, loss on ignition (LOI), elevation, population density, geology and land use information in an efficient manner; being able to handle a mixture of both continuous and categorical data. One major disadvantage of random forest is its "black box" nature that

prohibits easy interpretation of the relationships between the response and predictor variables, since it is impractical to investigate the structure of all trees in the forest. In order to overcome this shortcoming, the procedure allows estimation of the variable importance, measured by the mean decrease in prediction accuracy before and after permuting a variable. The RF modelling has been carried out using "randomForest" package of the R programming language.

RF models have been used in many different applications but recent studies have shown that they can be particularly useful in examining the relationships between spatial variables in environmental science applications e.g. (Birant, 2011; Li et al., 2011; Schwarz et al., 2013; Wiesmeier et al., 2011).

3 Summary of Findings

3.1 EDA

The data exploration has been informed by the literature search particularly recent work using multivariate approaches (Glennon et al., 2013; Zhang, 2006; Zhang et al., 2008; Zhang et al., 2011), which has initially been focused on research in the Irish environment and has highlighted the most important elements potentially associated with diffuse pollution.

Following on from the exploratory data analysis (EDA) the final output has been resolved into a more tightly defined output which focusses on processes controlling a limited suite of contaminants. These include As, Sb, Hg and Pb identified by literature and through the initial EDA phase of this work (e.g. Figure 6).

3.2 SMMR

The SMMR analysis of the soils data sets suggests there are 19 underlying source components. This was determined by carrying out the SMMR algorithm on the data using increasing number of components. The model with the minimum number components required to reproduce the original data (i.e. not significantly different from the original data) was chosen. The SMMR method produces an idealised model assuming there are a discrete number of source components with distinct chemical compositions mixing together over the whole Tellus region. Whilst the natural situation is unlikely to be this straightforward, the SMMR model can provide an insight into the underlying processes governing the soil geochemistry. Barplots of the chemical composition of each component and the interpolated spatial distribution of the component are given in Appendix 1. In order to ensure the SMMR source components have physical meaning and are not artefacts of the data processing algorithm, it is necessary to put the SMMR source components in to a geochemical context, making sure they are consistent with additional soil properties (pH and LOI), geology and the related physical and human geography of the region. Each component was screened according to the following criteria:

i) Does the SMMR source have a clearly defined chemical composition? – large uncertainties on the major and trace elements suggests that this may just be describing a "noise component";
ii) Does the spatial extent of the SMMR source align with the underlying geology of the region and is the geochemical composition consistent with the geology?
iii) Does the SMMR source show a correlation to soil properties or physical or human geographic parameters?

Examination of the SMMR sources with respect to criterion i) removed two components having poorly defined chemical compositions and having relatively low overall contribution to the overall extracted mass (< 1.4%). Under criterion 2 it was possible to clearly define the source of 8 of the components as having been derived from the underlying geology. These are summarised in Table 2.

Table 2 Summary of Geological sources identified in the soil chemistry, the names of the sources are made up from those elements that make up more than 10% of the source composition

SMMR source	Geology	Comments
Al.Fe.Mg 1	Lower Palaeozoic, mostly associated with greywacke formation	
Al.Fe.Mg	Same as Al.Fe.Mg 1 but with higher trace elements resulting from additional mineralisation	
Fe.Al.K	Clay	High in K and Rb
Ca.Al	Mostly associated with the Tournasian limestone formation	High in Cd
Al.K.Mg	Probably derived from weathering of Dalradian aluminosilicates	High concentration of Zr
Al.Fe	Underlying granites	High in Li Th and U
Fe.P	Namurian Shales	High in Mo
Mg.Fe	Derived from weathering of Dalradian aluminosilicate rocks	High in REEs

In addition to the underlying geology three other components were identified from their chemical composition as being derived from secondary processes (Table 3): a source made up of over 90% iron, suggesting that this an iron oxide; a source containing 60% Fe 20% Mn which is probably a mixed Fe/Mn oxide; and, a source made up of over 90% Ca occurring almost exclusively along the shoreline of Co. Donegal which suggests a marine calcium carbonate origin.

Correlations of the SMMR sources with soil properties and physical or human geographic parameters (Figure 8) found 4 sources with high sulphur content and positive correlations to LOI and negative correlations to pH which, along with their spatial distributions; these are clearly identified as components associated with peat bogs (Table 4).

The origin of two of the source components were less easily defined (

Table 5) but the Fe.Ca.P component is probably derived from mineralised underlying geology and the Fe.Al.Ca is possibly derived from a mixed Fe oxide phase in the soil.

Table 3 Summary of sources identified in the soil chemistry related to secondary processes, the names of the sources are made up from those elements that make up more than 10% of the source composition

SMMR source	Source description	Comments
Fe	Made up of over 90% iron suggests that this an iron oxide source	Occurs primarily in localised regions of the highlands of co Donegal
Fe.Mn.Al	Mixed Fe Manganese oxides (60% Fe 20%Mn)	Occurs primarily in localised regions the highlands of co Donegal and Co Sligo and Co Leitrim
Ca	Made up of over 90% Ca suggests that this Ca carbonate source	Probably a marine carbonate as it occurs along the shoreline of Co Donegal

Table 4 Summary of sources identified in the soil chemistry related to peat bogs, the names of the sources are made up from those elements that make up more than 10% of the source composition

SMMR source	Source description	Comments
S.Mg	Made up of over 60% S and 15% Mg with highest concentrations found along the coast of Co. Donegal. High correlation with LOI and elevation	Relatively low concentrations of inorganic elements suggests that this is a high organic matter content
S.Al.Fe	Made up of 30% S and Al. U content of 0.1% High correlation with LOI and elevation	Higher concentration of inorganic elements suggests higher ash content
Mg.S.Na	Contains 30% Mg and S and 3% Pb. High correlation with LOI and elevation	High concentrations of anthropogenic pathfinder elements suggests aerial deposition
Mg.S.Ca	Contains 30% Mg and S and occurs principally on the peninsula above Dundalk in Co. Louth. Has the lowest correlation with LOI and elevation out of the 4 components associated with peat sources.	

Table 5 Summary of sources identified in the soil chemistry whose sources are less clearly identified, the names of the sources are made up from those elements that make up more than 10% of the source composition

SMMR source	Source description	Comments
Fe.Ca.P	Made up of 45% Fe, 25% Ca.With high concentrations of Zn, Pb and Hg. Highest concentrations found principally along the edge of the graywacke formation in counties Monaghan and Cavan	Probably a mineralised source
Fe.Al.Ca	Made up of 40% Fe, 30% Al with high concentrations of As. Fairly widespread over the Tellus border region	Probably a mixed Fe oxide

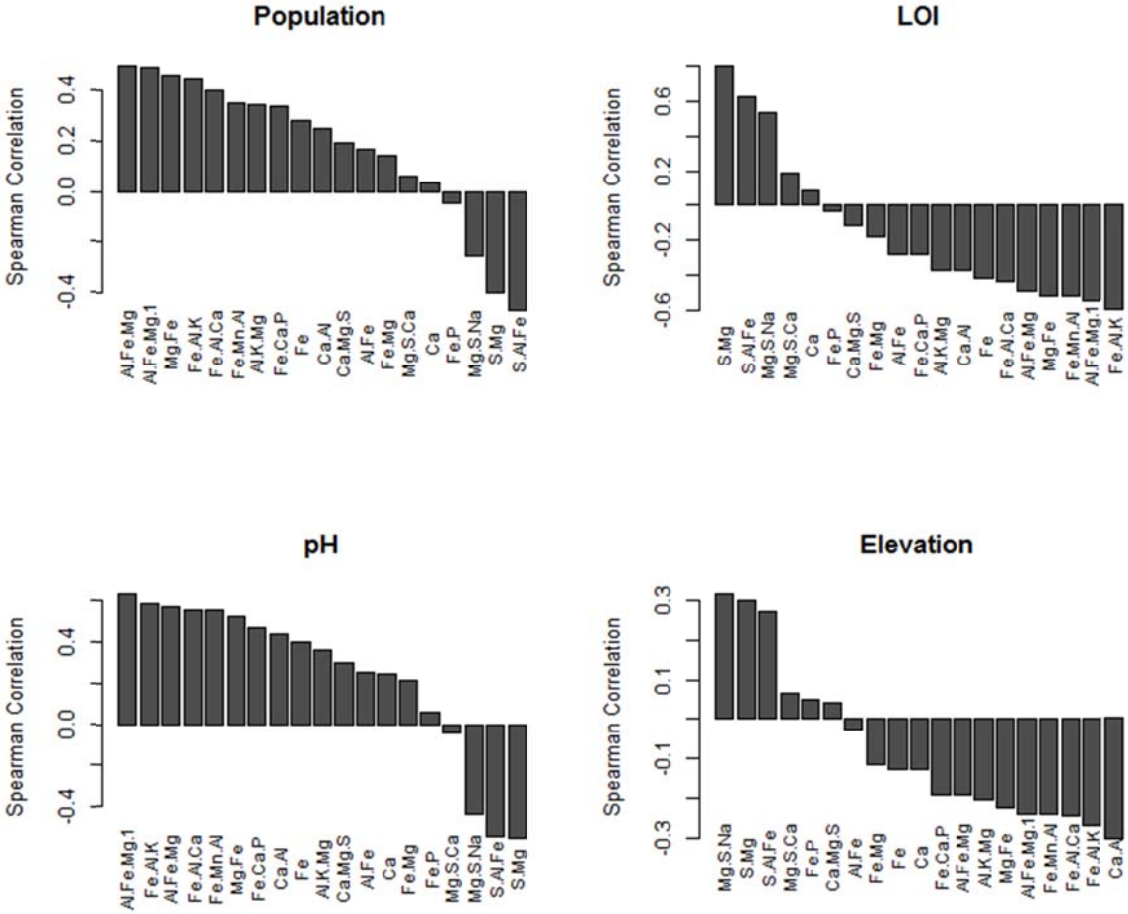


Figure 8 Spearman correlation of the mass of material associated with each component to populaton, LOI, pH and Elevation

Figure 9 to Figure 14 show how the mass of each pathfinder element summed over all soil samples is distributed between the source components identified by the SMMR modelling along with 95th percentile error bars.

For Pb and Sb (Figure 9 and Figure 10) the Mg.S.Na component identified as the possible diffuse pollution source in Table 4 is the has the highest single contribution to the overall mass inventory for these elements. For Hg (Figure 11) the mineralised component (Fe-Ca-P) has the highest contribution taking into account the uncertainty the next highest contribution is the diffuse pollution Mg.S.Na component.

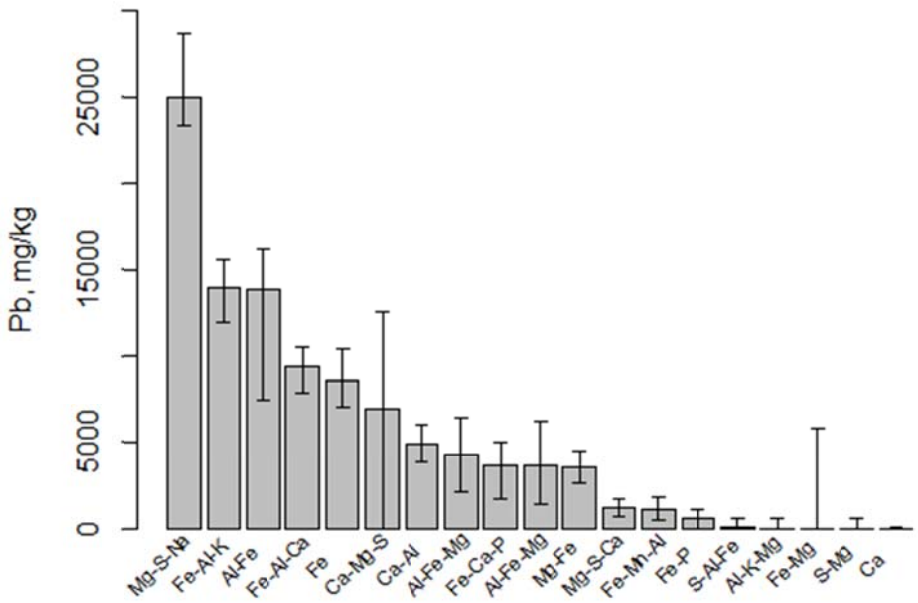


Figure 9 Mass distribution of Pb between the SMMR source components

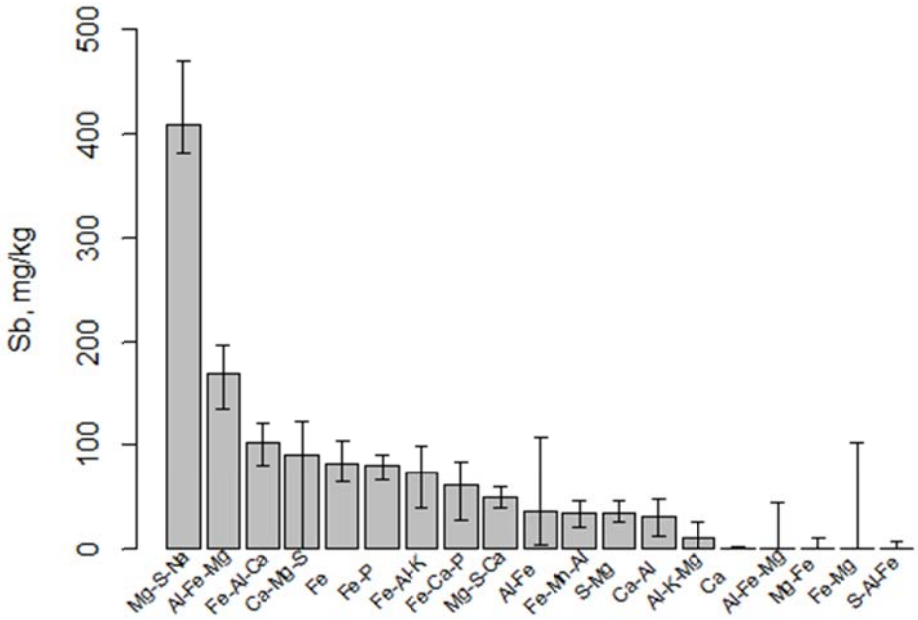


Figure 10 Mass distribution of Sb between the SMMR source components

For Cd (Figure 12) the top two mass contributors are the Ca-Al limestone derived component and the Fe-Ca-P mineralised component with the next most significant being the Mg.S.Na diffuse pollution component.

Arsenic (Figure 13) shows a very different but probably predictable mass distribution with Fe oxide as the main host for As in the soils, which is similar to studies in England which show that As is most often associated with Fe oxides (Palumbo-Roe et al., 2005). There is, however, a significant portion of the As associated with the diffuse pollution component.

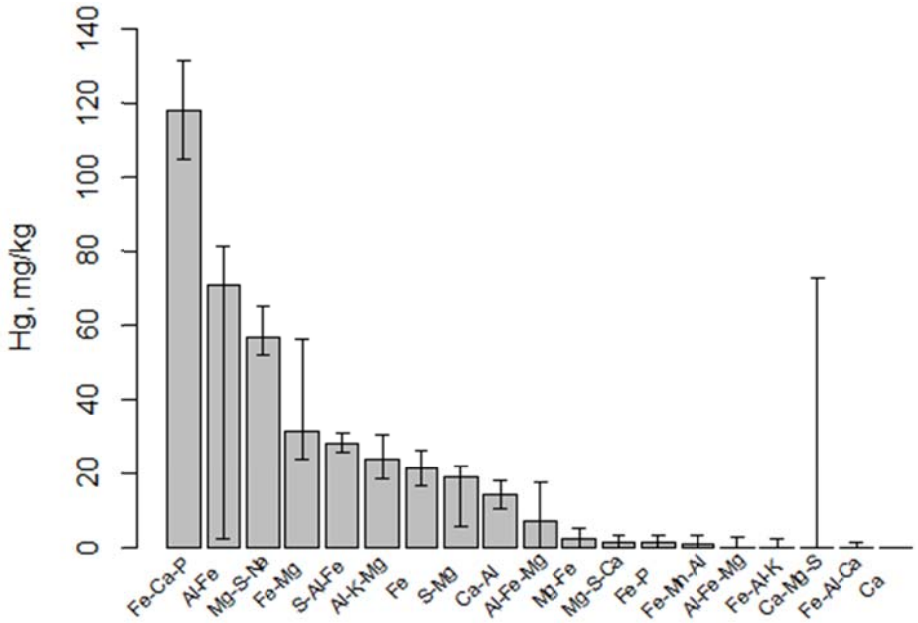


Figure 11 Mass distribution of Hg between the SMMR source components

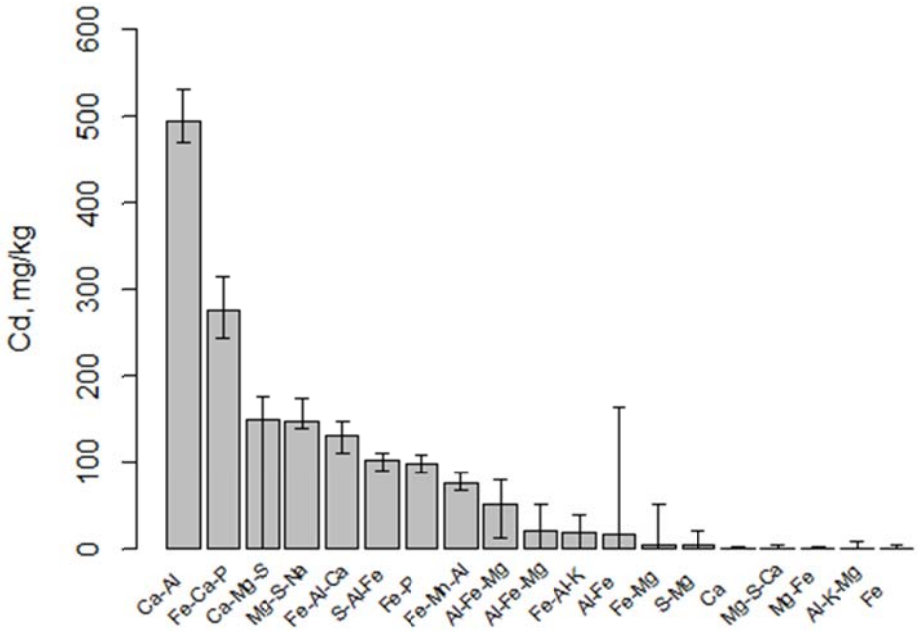


Figure 12 Mass distribution of Cd between the SMMR source components

Finally U (Figure 14) shows quite a different pattern to the other pathfinder elements with the highest concentrations associated with granites (Al.Fe), Dalradian aluminosilicates (Mg.Fe) and limestones (Ca.Al) and Fe/Mn oxides (Fe.Mn.Al). Interestingly, the next highest contribution is to the mineralised peat component (S.Al.Fe) but there is no significant U content associated with the peat derived diffuse pollution component.

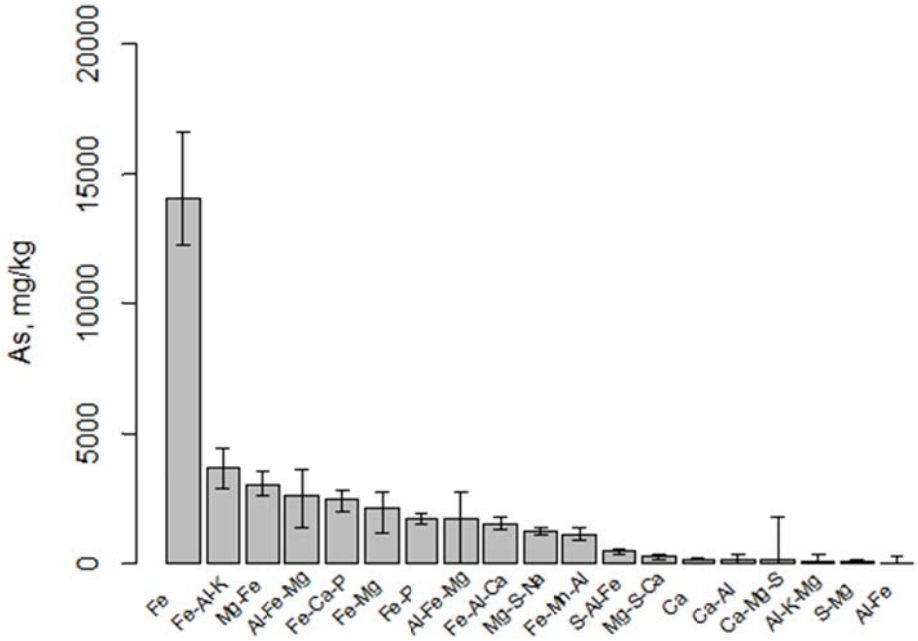


Figure 13 Mass distribution of As between the SMMR source components

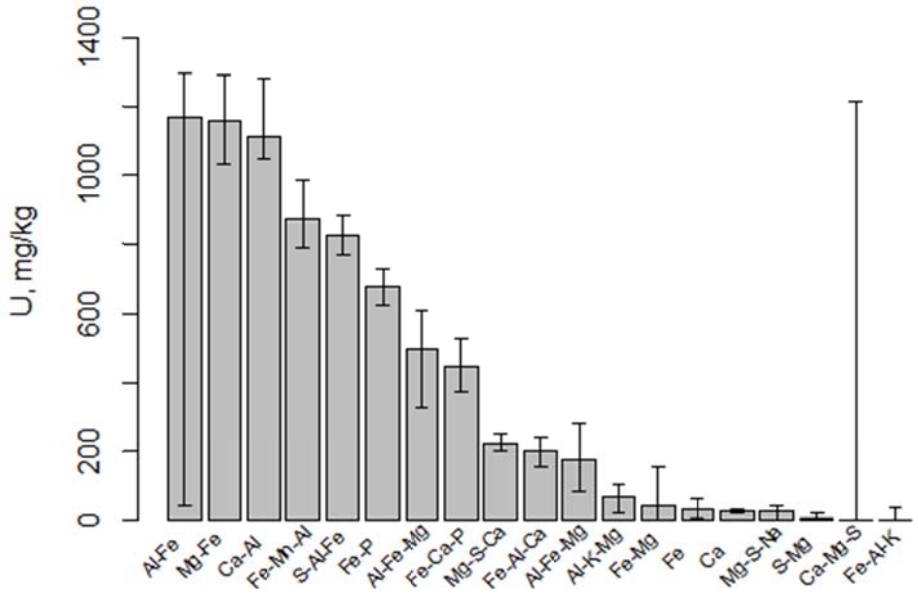


Figure 14 Mass distribution of U between the SMMR source components

3.3 IDENTIFICATION OF DIFFUSE POLLUTION

The results of the SMMR analysis of the Tellus border soils data suggests that majority of chemical sources which contribute to the soil chemistry are derived from three broad categories: The underlying geology; secondary processes such as the formation of iron oxides and carbonates; and from peat bogs. The anthropogenic pathfinder elements can be found in occurring naturally in geologically derived components particularly where mineralisation occurs and in the case of As it occurs principally in iron oxides.

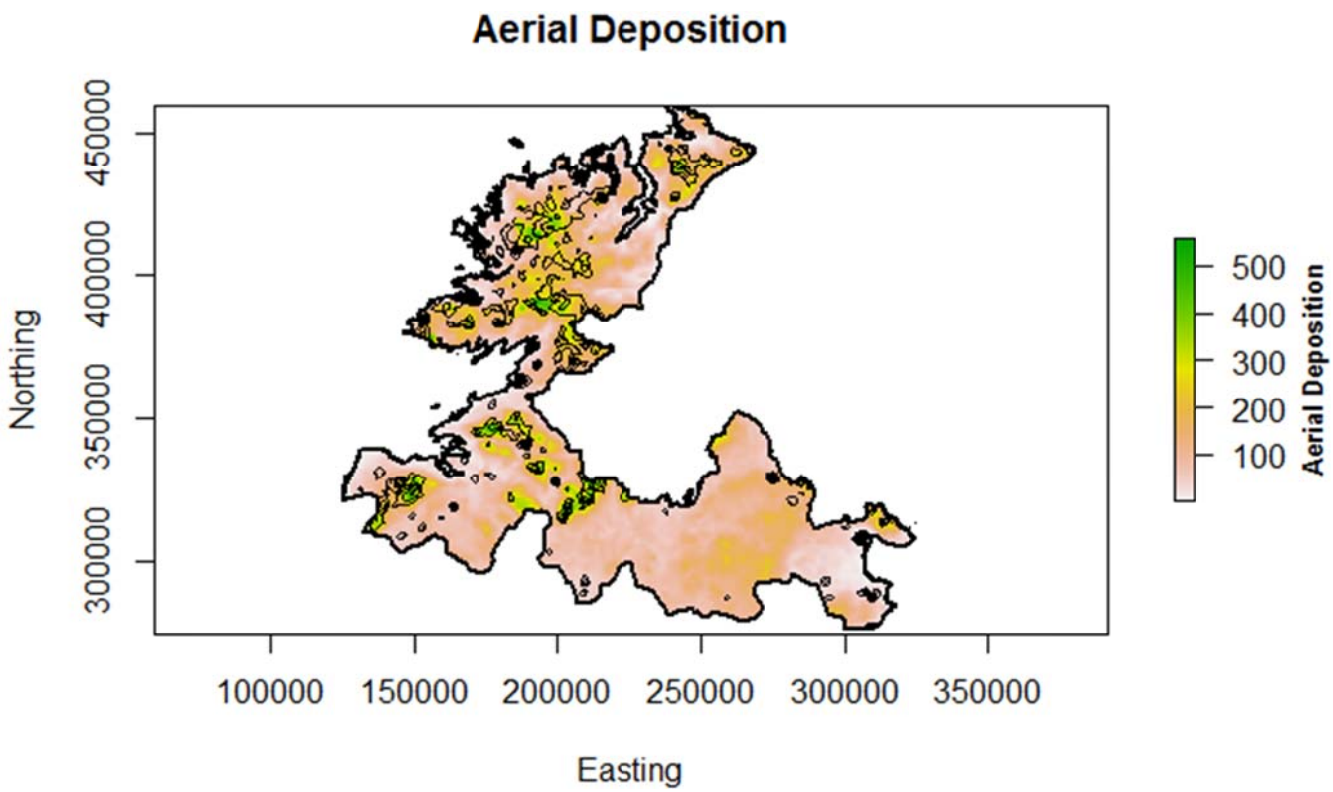


Figure 15 Suspected anthropogenic component (black contour lines) superimposed on the elevation data (colour ramp) showing a close spatial match

One source, (Mg.S.Na), with the highest positive correlation to elevation (Figure 8) shows particularly high concentrations of the pathfinder elements (Appendix 1). This source is thought to be related to peat bogs (Table 4). This source has the highest individual contribution for the total Pb and Sb found in the Tellus border soils and is an important source for Cd and Hg but has less contribution to total As and U. Plotting the concentration of this source as a contour plot overlaying a colour map of elevation Figure 15 shows a close spatial match.

Peat records from ombrotrophic bogs have been shown to provide valuable information about the atmospheric inputs of trace metals e.g. (Farmer et al., 2009; Shotyk, 1996), because they are fed with nutrients and pollutants only through atmospheric inputs. The SMMR component Mg.S.Na is therefore believed to be a peat derived component containing aerial deposition of anthropogenic particulates.

A study of peat core just to the south of the Tellus border region in Co. Mayo and Co. Galway (Coggins et al., 2006) showed elevated concentrations of Cd, Hg and Pb that which when dated using Pb and Cs isotopes were shown to be associated with recent anthropogenic pollution (1950-1970) from aerial deposition. This suggests that the component dominated by an organic matter source, which is associated with high elevations, has trace element (Pb, Sb, Hg and Cd) concentrations identified in this study which are expected to be similarly derived from aerial deposition of anthropogenic pollution onto peat uplands.

The source of the high concentrations of these pathfinder elements is subject to speculation. Coggins *et al.* (2006) points out that the prevailing wind direction across Ireland is from the Atlantic Ocean in the west not from easterly winds, which suggests that the diffuse pollution is not from Britain and Northern Europe. They point out, however, that the peat core profiles are similar to those found in North America and make the suggestion the anthropogenic pollution could be derived from this source.

The unique feature of this study is that the data have been derived from surface soil samples not from a few core samples so the spatial extent of the diffuse pollution from aerial deposition can be mapped over a large region.

3.4 RANDOM FOREST MODELLING

The RF modelling was used as an independent check of the SMMR approach to source apportionment of the pathfinder elements. The log 10 transformed concentration of each of the pathfinder elements was modelled as a function of soil properties, location, elevation, population density, geology and land use.

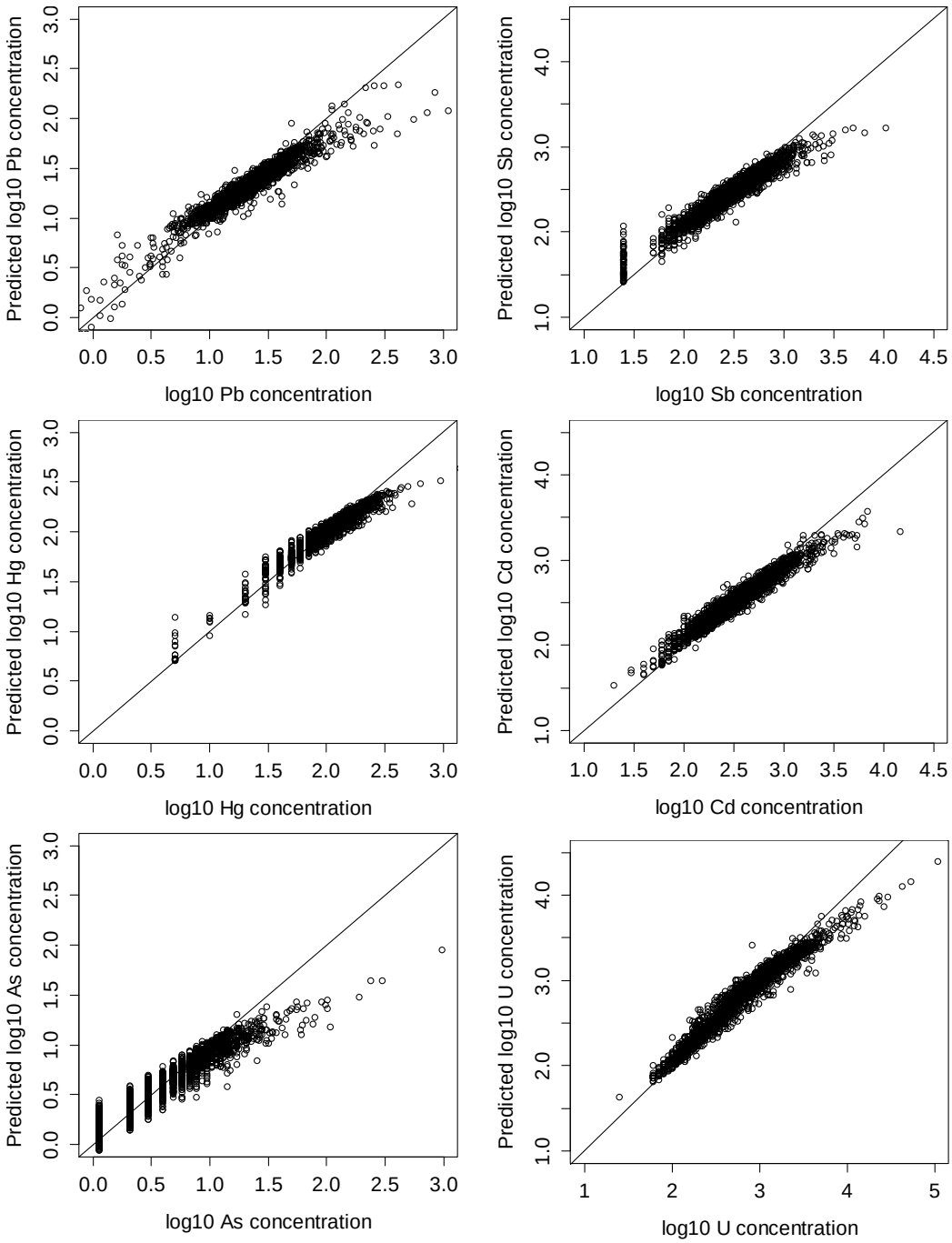


Figure 16 Plot actual versus predicted log transformed concentrations for the six anthropogenic pathfinder elements along with the line of equivalence

Figure 16 shows the actual vs the RF predicted values for each pathfinder element, illustrating that the RF models provide a reasonable predictive model, although there is some under prediction for the highest concentration of each element. The aim of this study, however, is not

to predict the concentrations of the elements but to understand how the predictor variables control the element compositions. Figures 17 to 22 show the univariate plots for each path finder element where one variable is varied and the other variables are held at their average value. The tick marks on the inside of the X axis represent the deciles of the actual values of the parameter used to set up the model. Whilst this does not show the full picture of the relationships between the element concentrations and the variables as there may be interactions between predictor variables, it does give some idea of how the predictor variables affect the element concentration.

Some points of note are:

- For all elements apart from U there are higher concentrations of the element of interest towards the east of the Tellus border region (i.e. higher Easting values) which can be explained by the eastern counties have more mineralised soils than the west where the soils have higher organic carbon content. The SMMR element distributions show that these elements have a significant contribution from the more mineralised soils (Figure 9-Figure 12);
- There is a clear indication that, in the main, higher pathfinder element concentrations are associated with high elevation, confirming the findings of the SMMR study.
- Whilst the SMMR study does not identify a source which is related to high population density (i.e. a localised anthropogenic input) the RF model suggests that there is a trend for increase in pathfinder element concentration in areas of higher population density. This would suggest that there may not be a well-defined multi-element composition from high population areas but that they do make relatively small but significant contribution to pathfinder elements in the Tellus border soils. This may be because there are few higher density areas in the Tellus Border study region in comparison to other parts of Ireland.

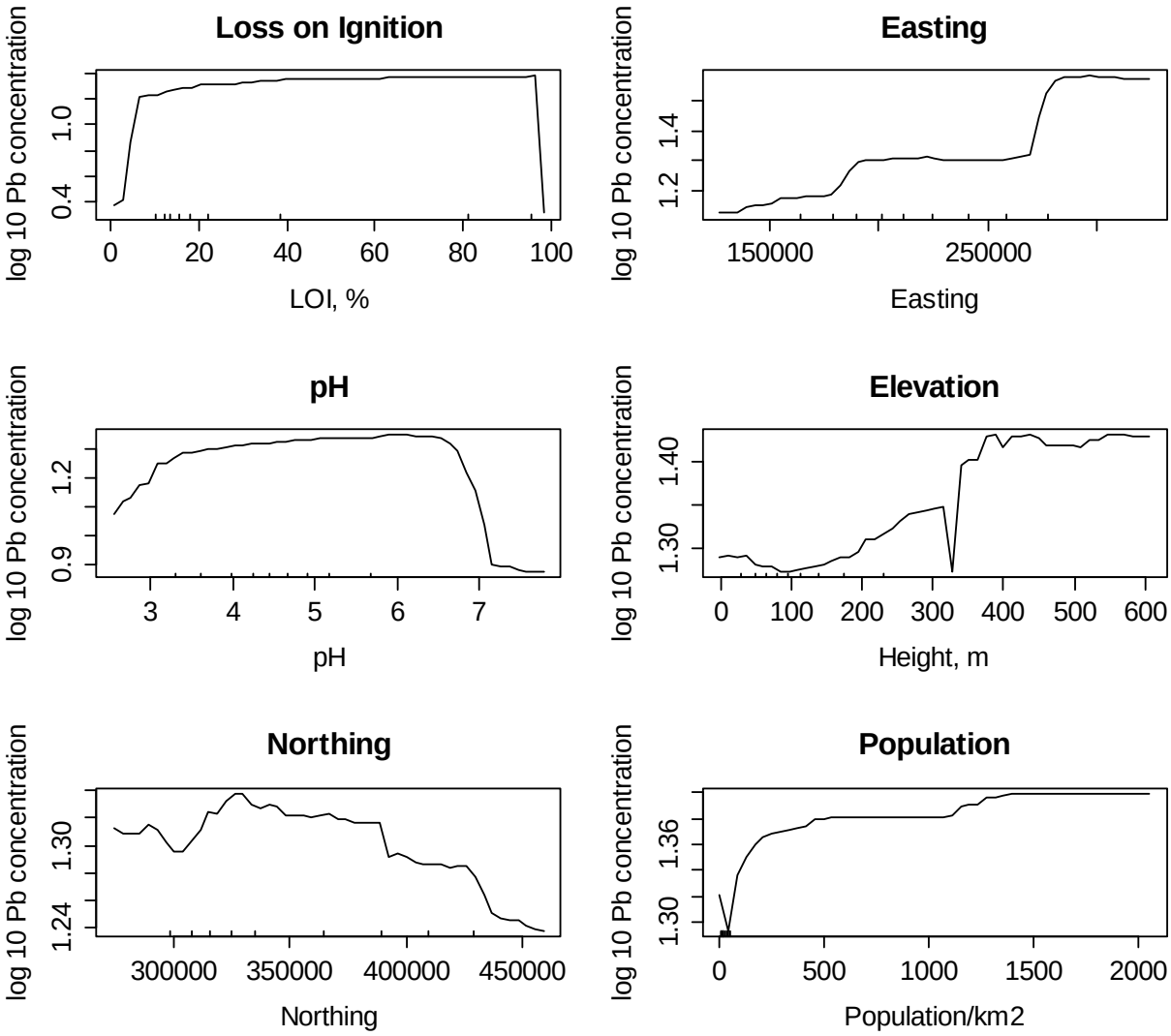


Figure 17 Univariate plots showing the behaviour of Pb with respect to geographical variables and soil properties

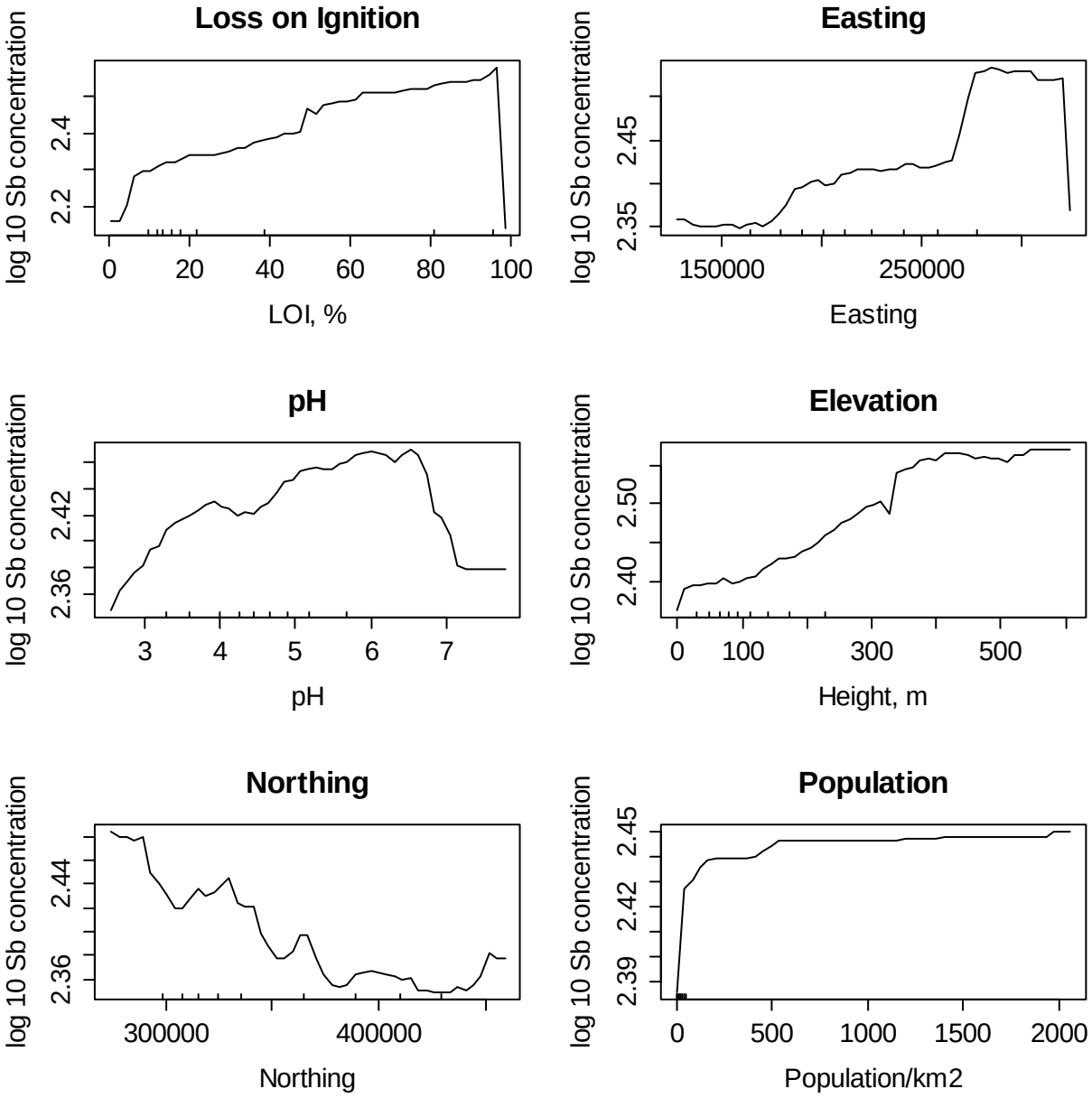


Figure 18 Univariate plots showing the behaviour of Sb with respect to geographical variables and soil properties

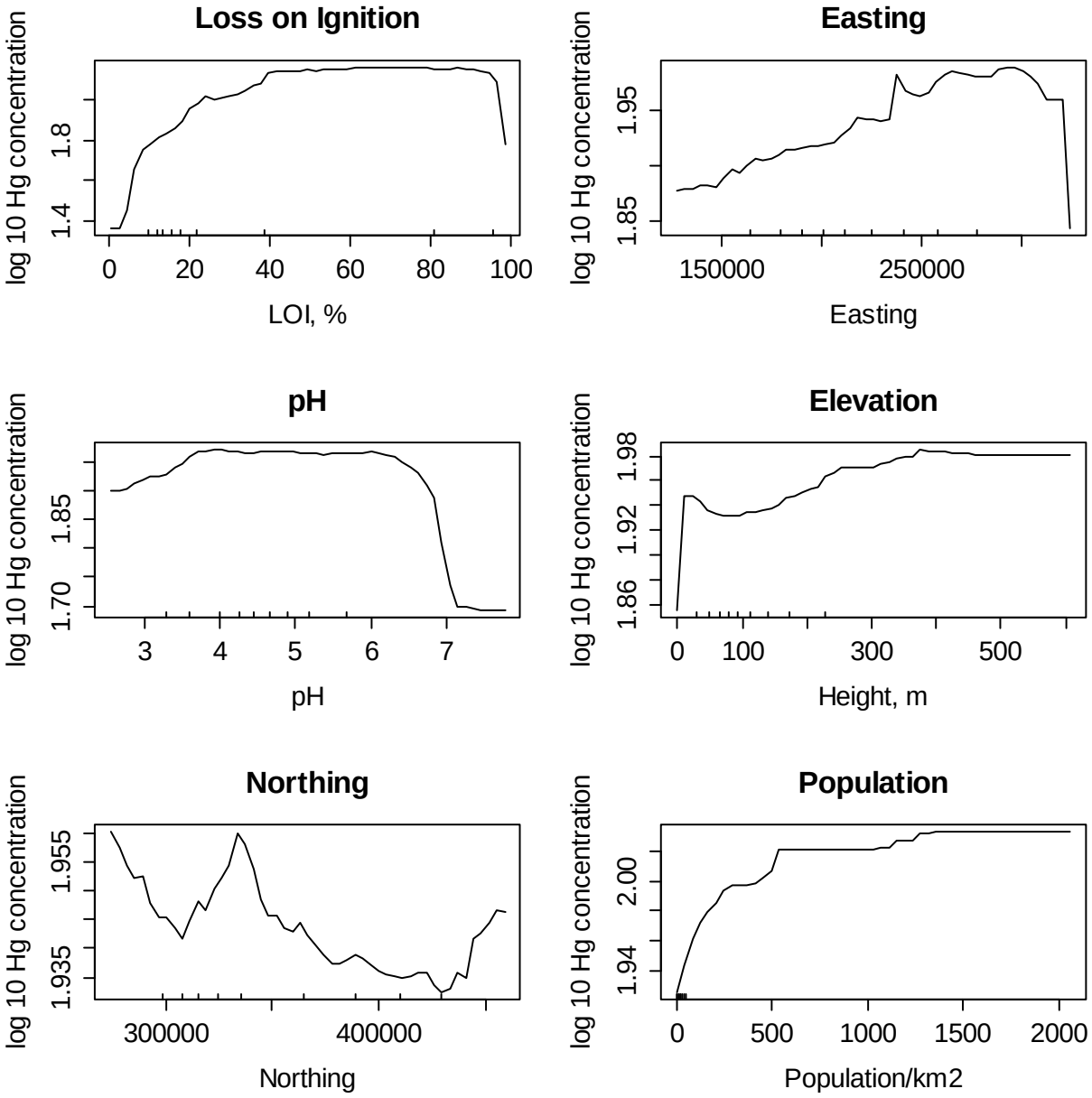


Figure 19 Univariate plots showing the behaviour of Hg with respect to geographical variables and soil properties

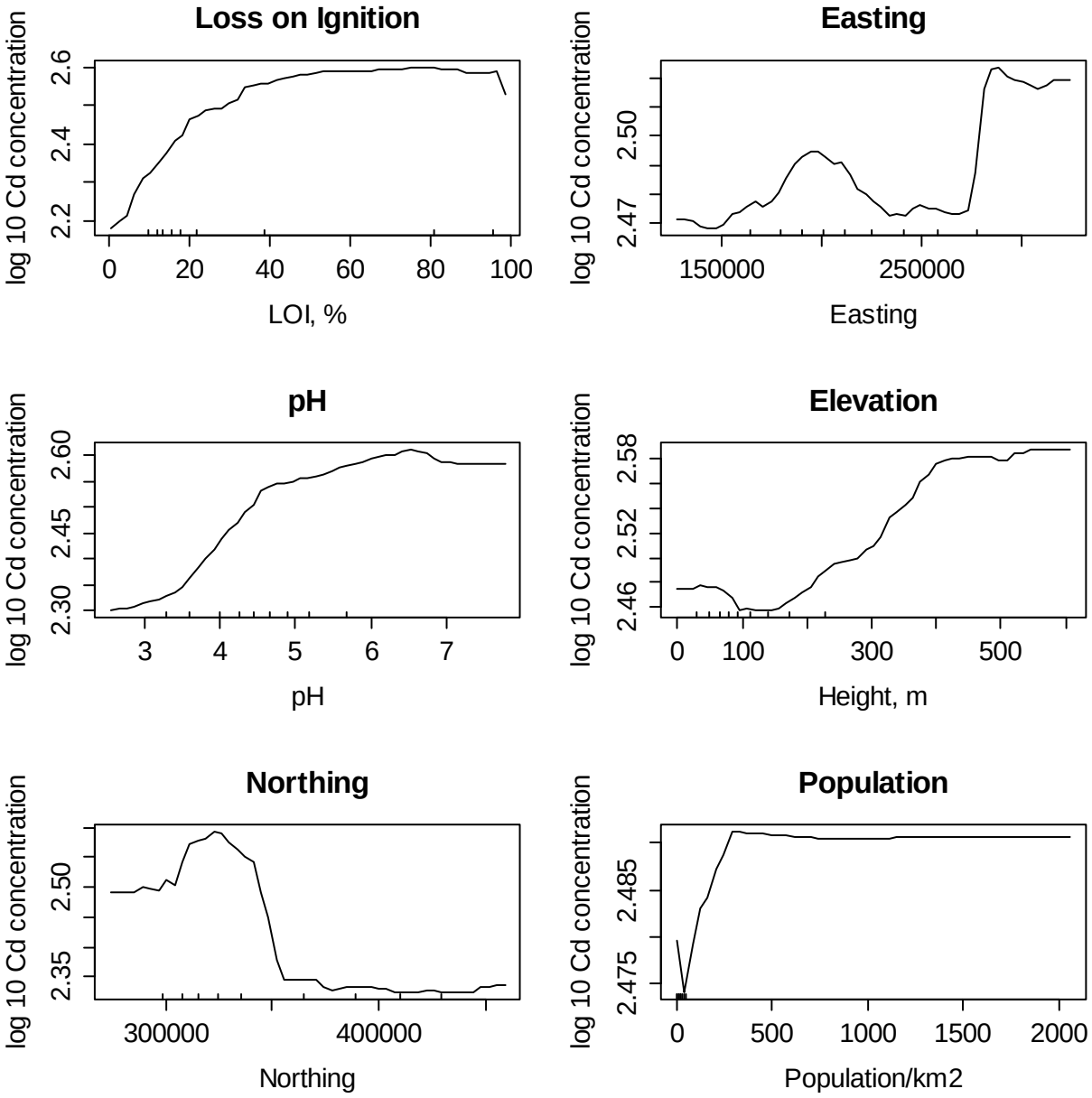


Figure 20 Univariate plots showing the behaviour of Cd with respect to geographical variables and soil properties

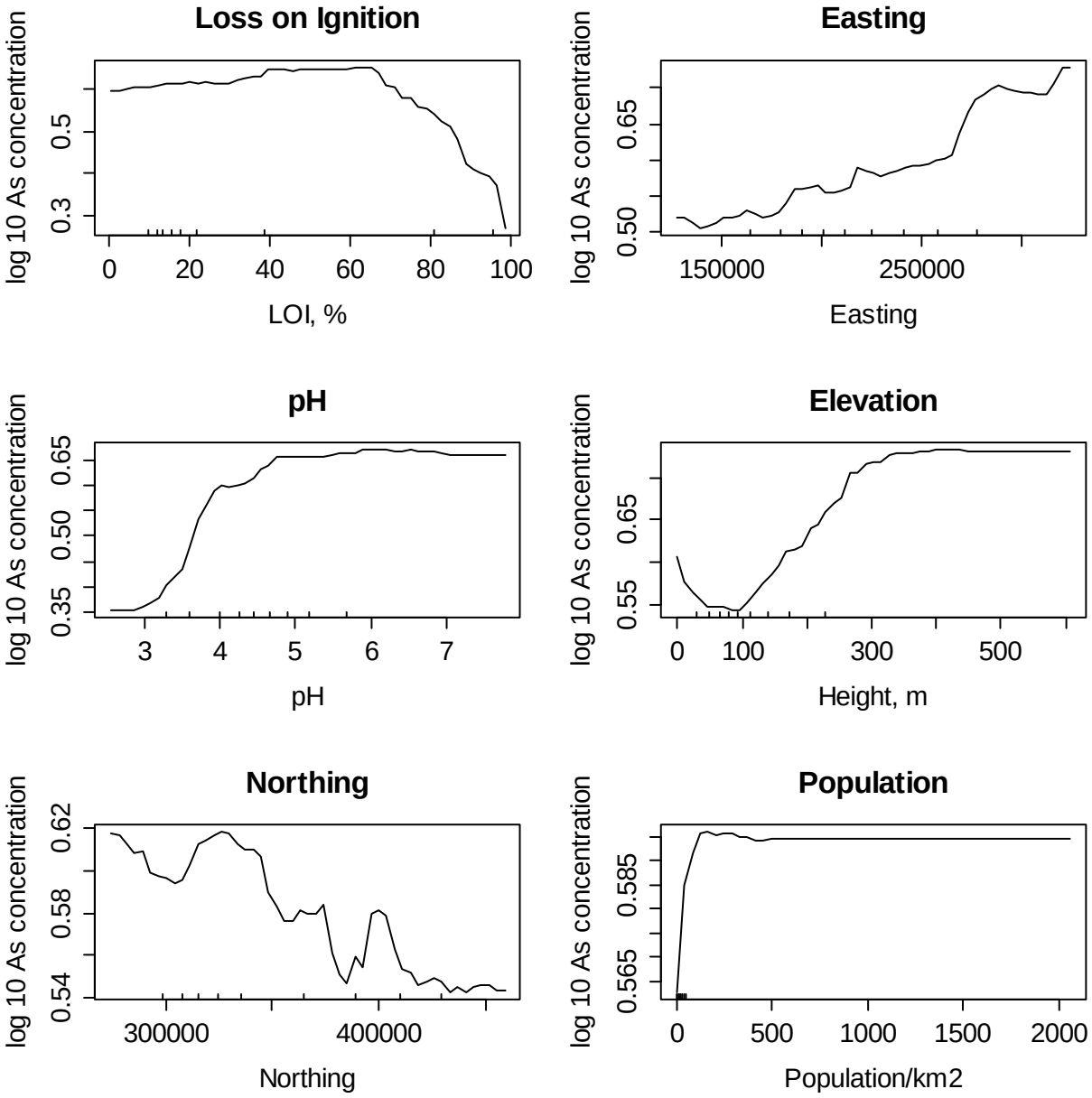


Figure 21 Univariate plots showing the behaviour of As with respect to geographical variables and soil properties

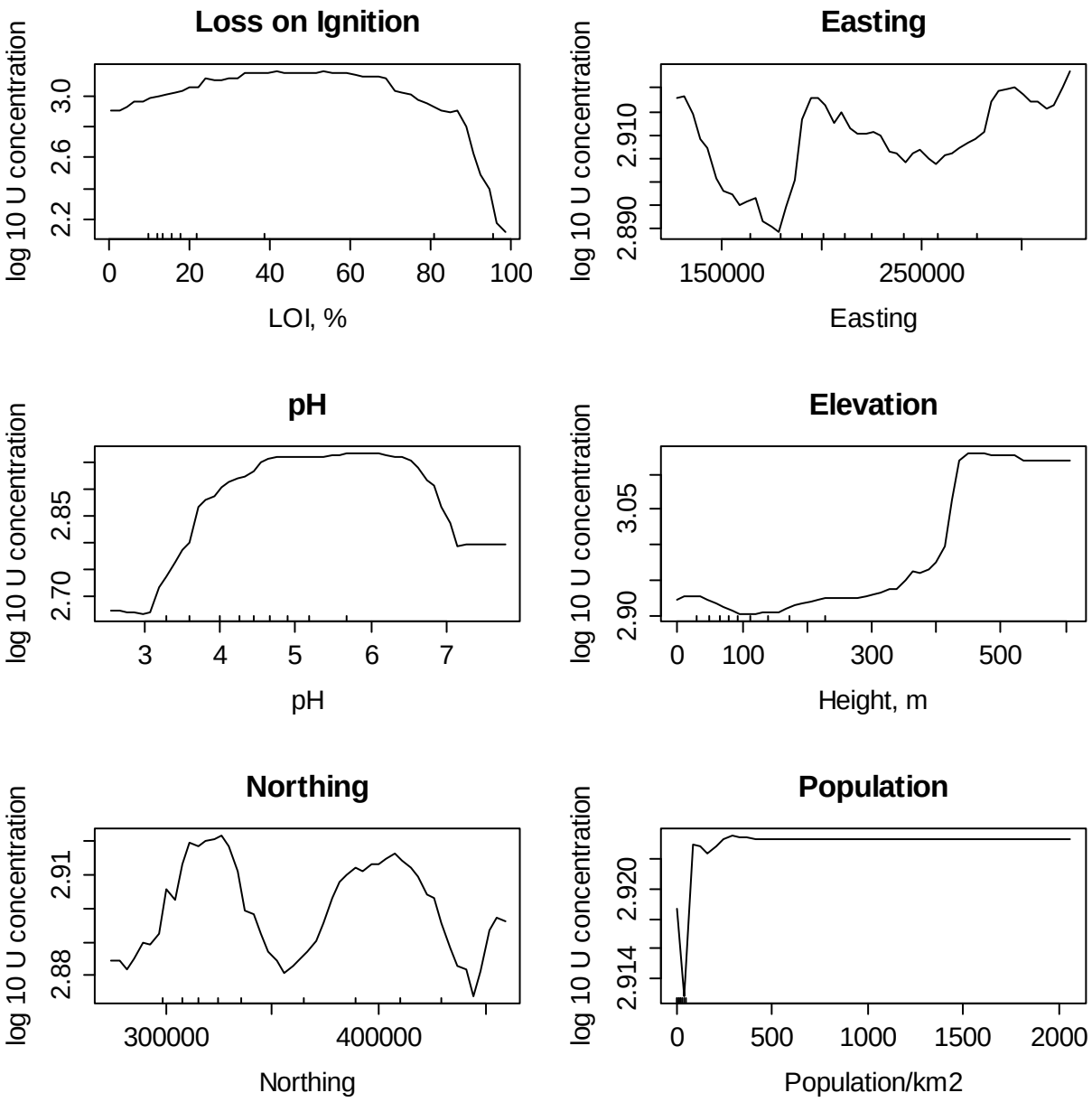


Figure 22 Univariate plots showing the behaviour of U with respect to geographical variables and soil properties

4 Conclusions

Drawing together and summarising the main findings of the report:

- Self-modelling mixture resolution has proved to be a useful tool for identifying chemical signatures in the Tellus border soils data;
- The SMMR methodology identifies 17 geochemical components of which 9 are derived from the underlying geology (high concentrations of trace metals suggest 2 of these are from mineralised sources), 4 are from secondary processes (Fe oxides and carbonates) and 4 associated with peat bogs;
- One of the peat bog components, which has the highest positive correlation with elevation, has high concentrations of the anthropogenic pathfinder elements. This component is thought to be derived from aerial deposition of anthropogenic particulates derived from modern industrialisation;
- The unique feature of this study is that the data have been derived from surface soil samples not from a few core samples so for the first time the spatial extent of the diffuse pollution from aerial deposition can be mapped over a large region;
- Random Forest modelling provides an independent check of the outcome of the SMMR model confirming that the anthropogenic pathfinder elements are found at the higher elevations where aerial deposition is thought to be the source of diffuse pollution;
- RF modelling shows that there is an increase in pathfinder elements with increasing population density which suggests that there may a small contribution to soil chemistry from diffuse pollution from urbanisation.

5 Suggestions for future research

Understanding soil quality, including inorganic parameters, is important to manage the soil resource most effectively. The Tellus Border soil geochemical data provide a baseline against which to understand natural and historical inputs of elements to soils, as well as a baseline to monitor future changes against.

This research project has identified diffuse polllution as well as natural geological sources of trace elements to soils in the Tellus Border research area. We expect to submit this research for peer-reviewed publication in the near future, because we use novel statistical approaches to address the question of identifying diffuse soil pollution. This will help to publicise the Tellus Border data and outputs to the research community as well as ensure peer-review of the findings.

In addition we suggest that this opens up future research options which may further the understanding of diffuse pollution in the Tellus Border area and/or more widely in Ireland:

- comparison of outputs with another Tellus Border research project "*Twig epiphyte communities as indicators of environmental quality in the border region of Ireland*.", Principle Investigator Howard Fox, to establish whether there are any consistencies in observations that may give rise to greater understanding;
- isotope analyses of archived residual material from the soil samples, particularly for lead (Pb), may help identify proportions of local vs longer-range lead deposition sources; *and*
- understanding soil quality diffuse pollution burdens around more extensively and/or densely populated areas of the Republic of Ireland would be useful (e.g. to complement the Dublin SURGE study (Glennon et al., 2013)), in any future sampling scheme. This is because somewhat greater diffuse inputs may be expected in those areas, based on international experience.

6 Report references

BIRANT, D. 2011. Comparison of Decision Tree Algorithms for Predicting Potential Air Pollutant Emissions with Data Mining Models. *Journal of Environmental Informatics*, Vol. 17, 46-53.

BREIMAN, L. 1996. Bagging predictors. *Machine Learning*, Vol. 24, 123-140.

CAVE, M R. 2009. The use of self modelling mixture resolution for the interpretation of geochemical data. *British Geological Survey*, IR/08/035.

COGGINS, A M, JENNINGS, S G, and EBINGHAUS, R. 2006. Accumulation rates of the heavy metals lead, mercury and cadmium in ombrotrophic peatlands in the west of Ireland. *Atmospheric Environment*, Vol. 40, 260-278.

DEMPSTER, M, DUNLOP, P, SCHEIB, A, and COOPER, M. 2013. Principal component analysis of the geochemistry of soil developed on till in Northern Ireland. *Journal of Maps*, 1-17.

FARMER, J G, ANDERSON, P, CLOY, J M, GRAHAM, M C, MACKENZIE, A B, and COOK, G T. 2009. Historical accumulation rates of mercury in four Scottish ombrotrophic peat bogs over the past 2000 years. *Science of the Total Environment*, Vol. 407, 5578-5588.

GLENNON, M M, HARRIS, P, OTTESEN, R T, SCANLON, R P, and O'CONNOR, P J. 2013. The Dublin SURGE Project: geochemical baseline for heavy metals in topsoils and spatial correlation with historical industry in Dublin, Ireland. *Environmental Geochemistry and Health*.

LI, J, HEAP, A D, POTTER, A, and DANIELL, J J. 2011. Application of machine learning methods to spatial interpolation of environmental variables. *Environmental Modelling & Software*, Vol. 26, 1647-1659.

PALUMBO-ROE, B, CAVE, M R, KLINCK, B A, WRAGG, J, TAYLOR, H, O'DONNELL, K, and SHAW, R A. 2005. Bioaccessibility of arsenic in soils developed over Jurassic ironstones in eastern England. *Environmental Geochemistry and Health*, Vol. 27, 121-130.

SCHWARZ, K, WEATHERS, K C, PICKETT, S T A, LATHROP, R G, JR., POUYAT, R V, and CADENASSO, M L. 2013. A comparison of three empirically based, spatially explicit predictive models of residential soil Pb concentrations in Baltimore, Maryland, USA: understanding the variability within cities. *Environmental Geochemistry and Health*, Vol. 35, 495-510.

SHOTYK, W. 1996. Natural and anthropogenic enrichments of As, Cu, Pb, Sb, and Zn in ombrotrophic versus minerotrophic peat bog profiles, Jura Mountains, Switzerland. *Water Air and Soil Pollution*, Vol. 90, 375-405.

WIESMEIER, M, BARTHOLD, F, BLANK, B, and KOEGEL-KNABNER, I. 2011. Digital mapping of soil organic matter stocks using Random Forest modeling in a semi-arid steppe ecosystem. *Plant and Soil*, Vol. 340, 7-24.

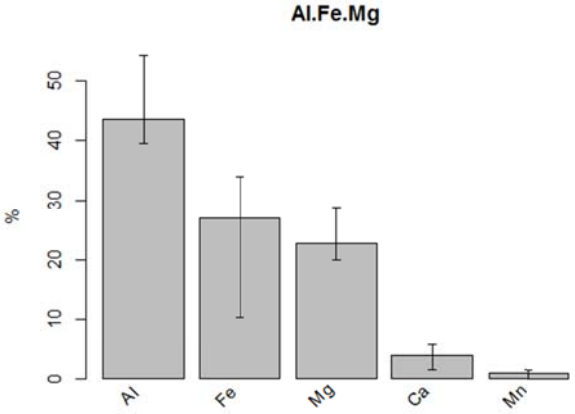
ZHANG, C. 2006. Using multivariate analyses and GIS to identify pollutants and their spatial patterns in urban soils in Galway, Ireland. *Environmental Pollution*, Vol. 142, 501-511.

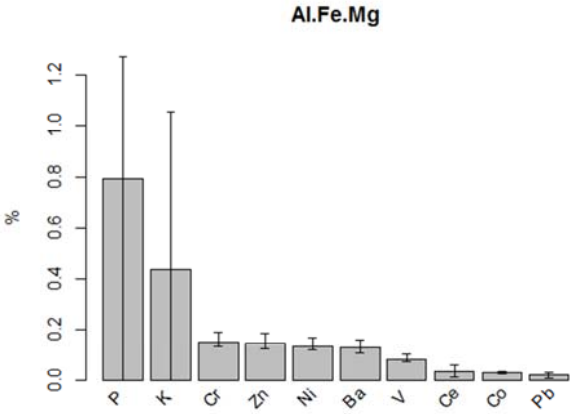
ZHANG, C, FAY, D, MCGRATH, D, GRENNAN, E, and CARTON, O T. 2008. Statistical analyses of geochemical variables in soils of Ireland. *Geoderma*, Vol. 146, 378-390.

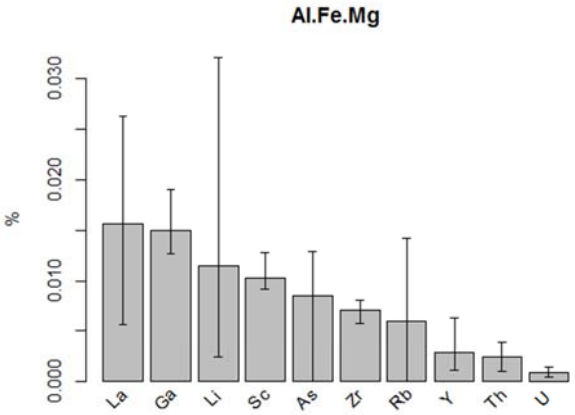
ZHANG, C, TANG, Y, XU, X, and KIELY, G. 2011. Towards spatial geochemical modelling: Use of geographically weighted regression for mapping soil organic carbon contents in Ireland. *Applied Geochemistry*, Vol. 26, 1239-1248.

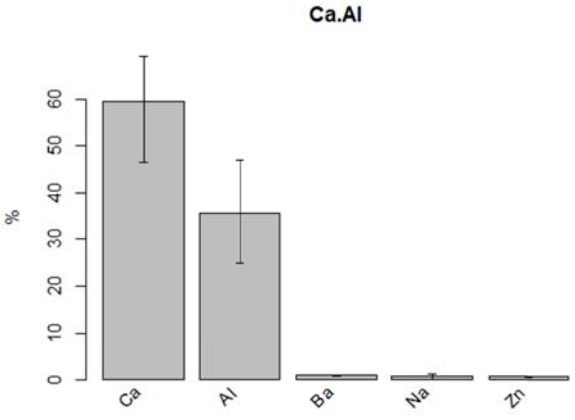
Appendix 1

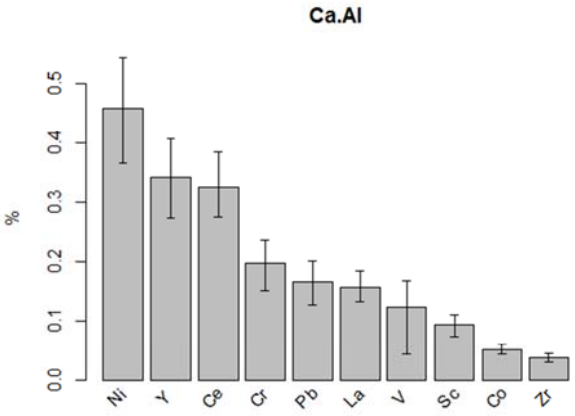
Chemical composition and interpolated spatial distribution of the SMMR soil components

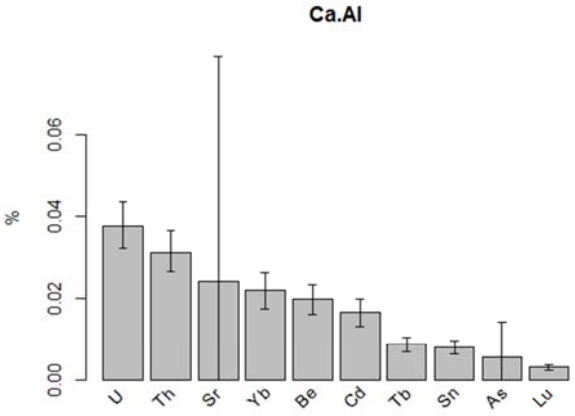


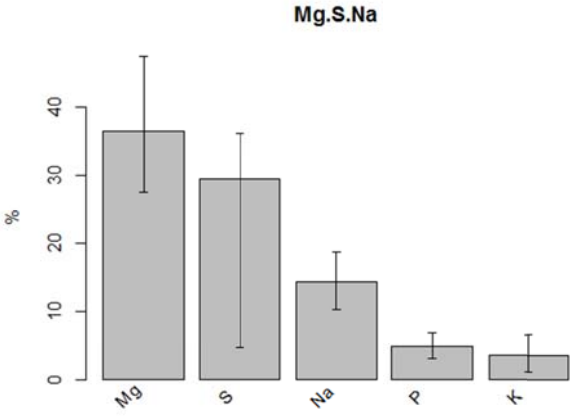


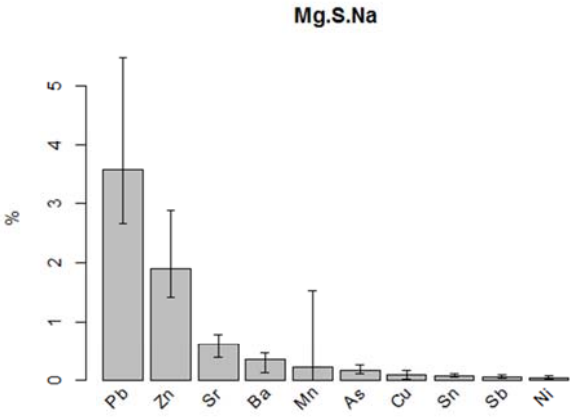


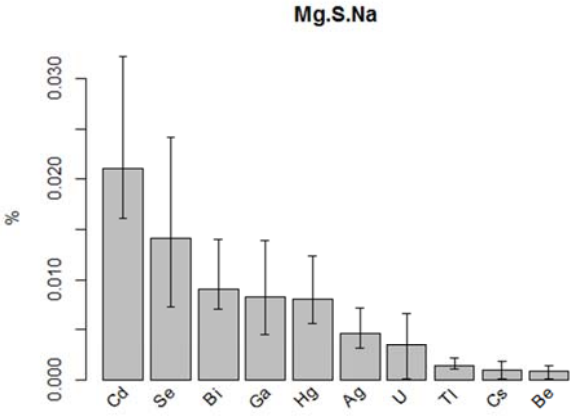

Ca.Al
%
0
10
20
30
40
50
60
Ca
Al
Ba
Na
Zn

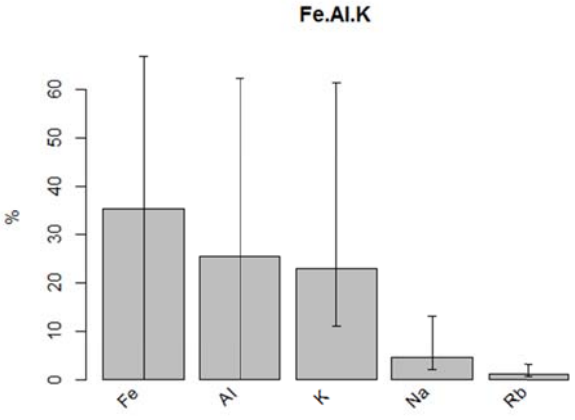

Ca.Al
%
0.0
0.1
0.2
0.3
0.4
0.5
Ni
Y
Ce
Cr
Pb
La
V
Sc
Co
Zr

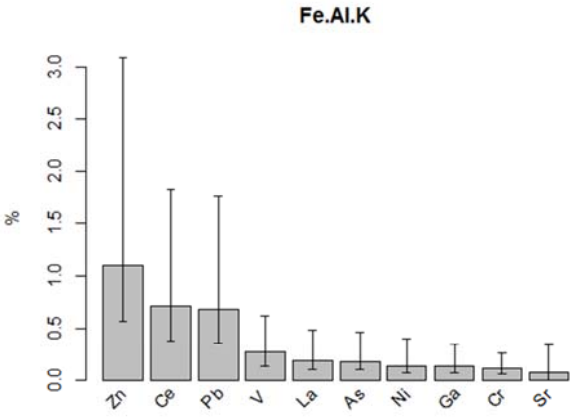

Ca.Al
%
0.00
0.02
0.04
0.06
U
Th
Sr
Yb
Be
Cd
Tb
Sn
As
Lu

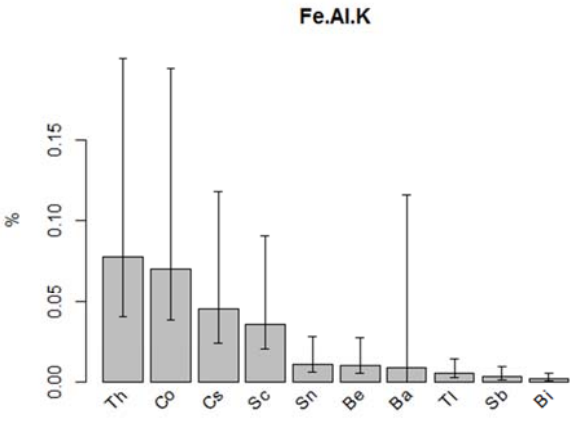

Mg.S.Na
%
0
10
20
30
40
Mg
S
Na
P
K

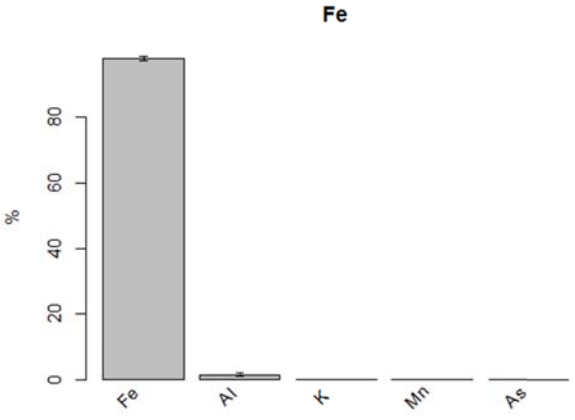

Mg.S.Na
%
0
1
2
3
4
5
Pb
Zn
Sr
Ba
Mn
As
Cu
Sn
Sb
Ni

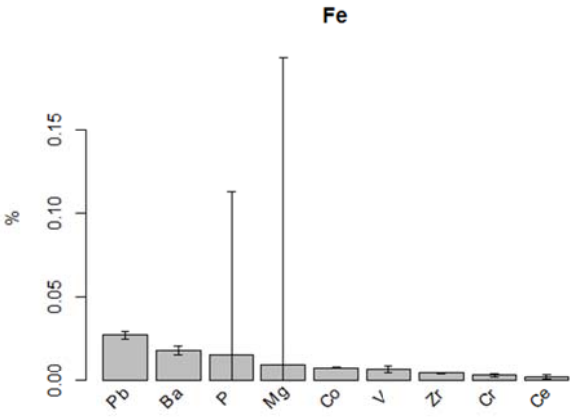

Mg.S.Na
%
0.000
0.010
0.020
0.030
Cd
Se
Bi
Ga
Hg
Ag
U
Tl
Cs
Be

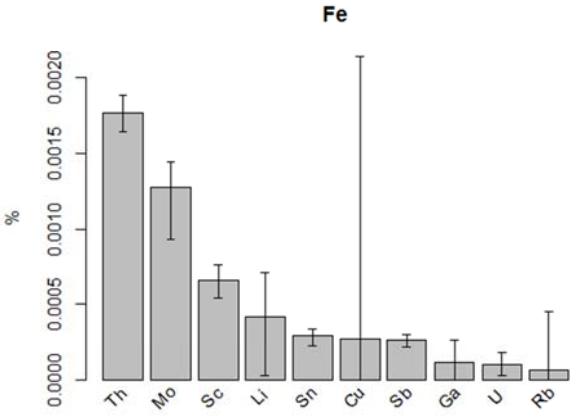

Fe.Al.K
%
0
10
20
30
40
50
60
Fe
Al
K
Na
Rb

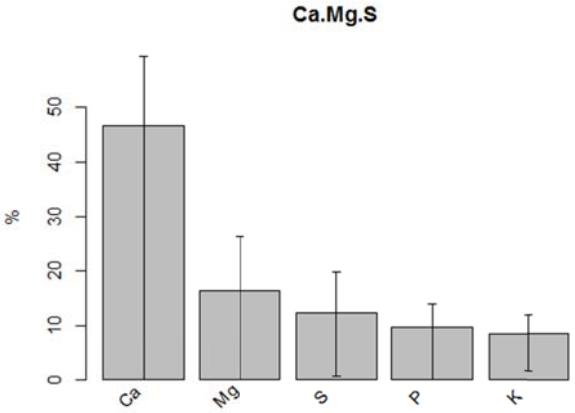

Fe.Al.K
%
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Zn
Ce
Pb
V
La
As
Ni
Ga
Cr
Sr

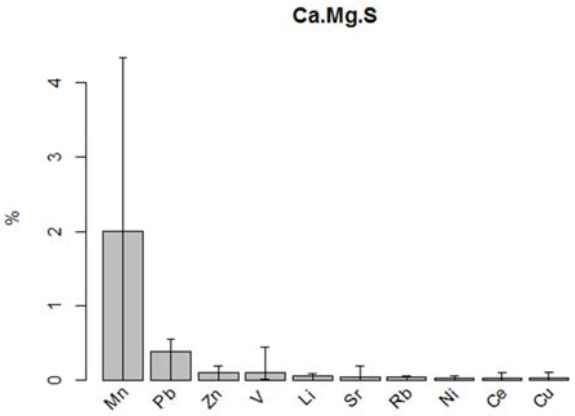

Fe.Al.K
%
0.00
0.05
0.10
0.15
Th
Co
Cs
Sc
Sn
Be
Ba
Tl
Sb
Bi

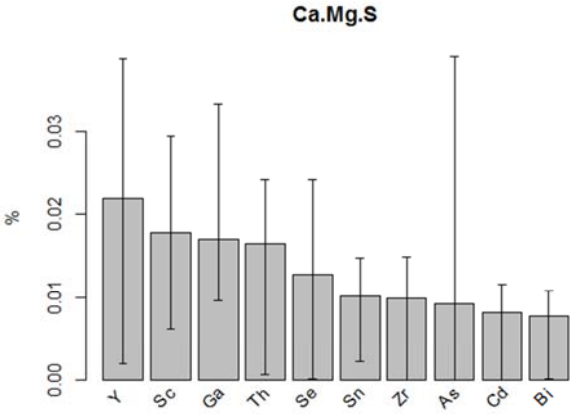

Fe
%
0
20
40
60
80
Fe
Al
K
Mn
As

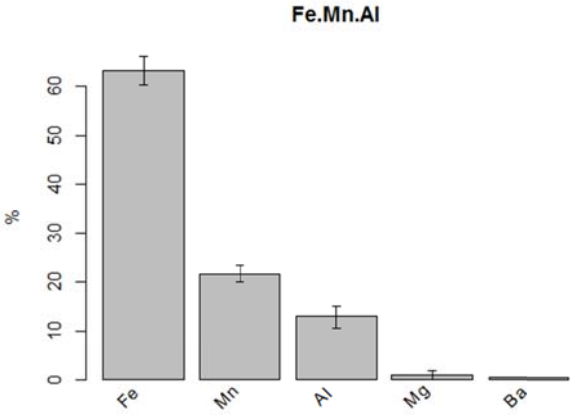

Fe
%
0.00
0.05
0.10
0.15
Pb
Ba
P
Mg
Co
V
Zr
Cr
Ce

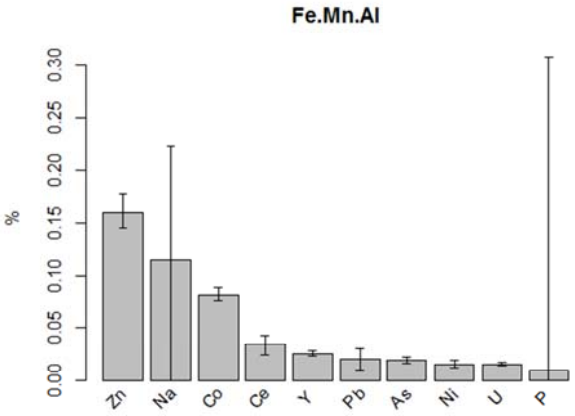

Fe
%
0.0000
0.0005
0.0010
0.0015
0.0020
Th
Mo
Sc
Li
Sn
Cu
Sb
Ga
U
Rb

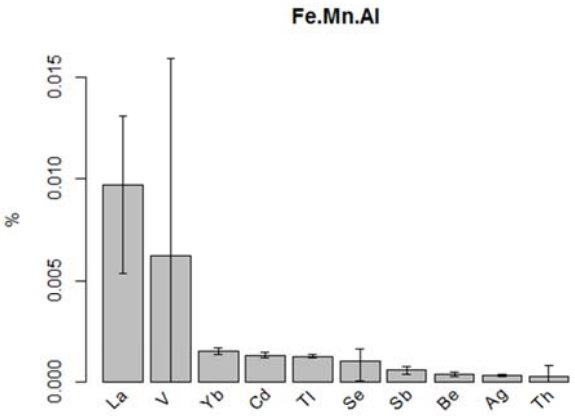

Ca.Mg.S
%
0
10
20
30
40
50
Ca
Mg
S
P
K

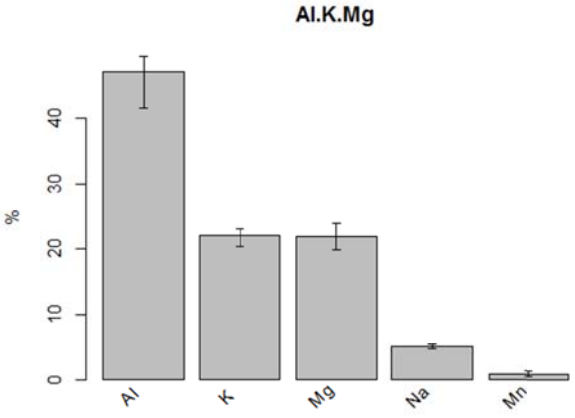

Ca.Mg.S
%
0
1
2
3
4
Mn
Pb
Zn
V
Li
Sr
Rb
Ni
Ce
Cu

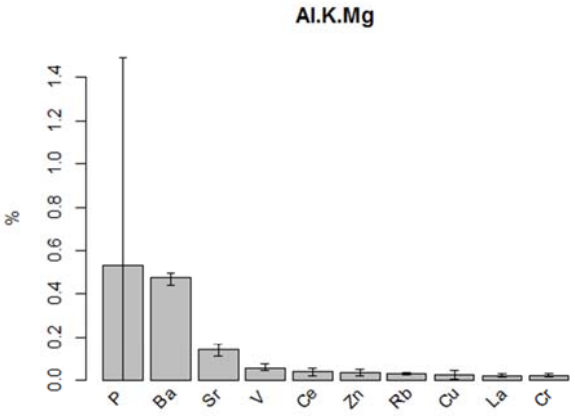

Ca.Mg.S
%
0.00
0.01
0.02
0.03
Y
Sc
Ga
Th
Se
Sn
Zr
As
Cd
Bi

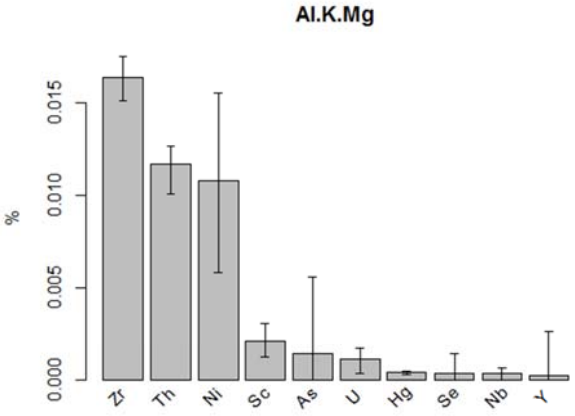

Fe.Mn.Al
%
0
10
20
30
40
50
60
Fe
Mn
Al
Mg
Ba

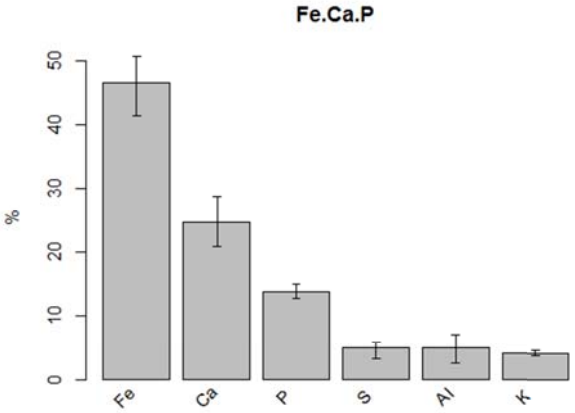

Fe.Mn.Al
%
0.00
0.05
0.10
0.15
0.20
0.25
0.30
Zn
Na
Co
Ce
Y
Pb
As
Ni
U
P

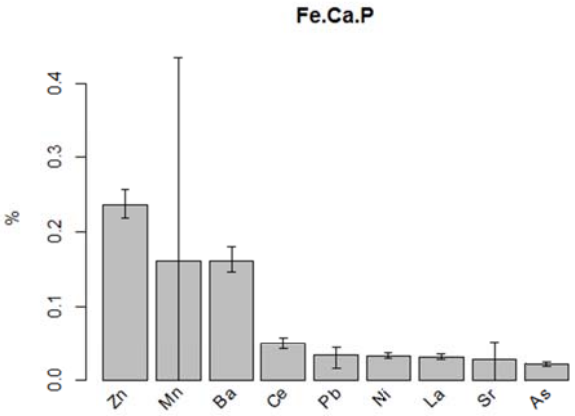

Fe.Mn.Al
%
0.000
0.005
0.010
0.015
La
V
Yb
Cd
Tl
Se
Sb
Be
Ag
Th

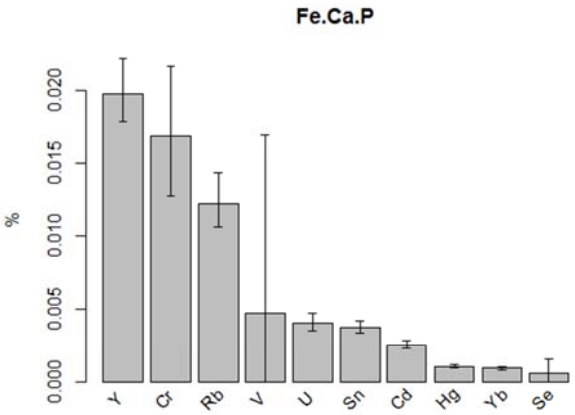

Al.K.Mg
%
0
10
20
30
40
Al
K
Mg
Na
Mn

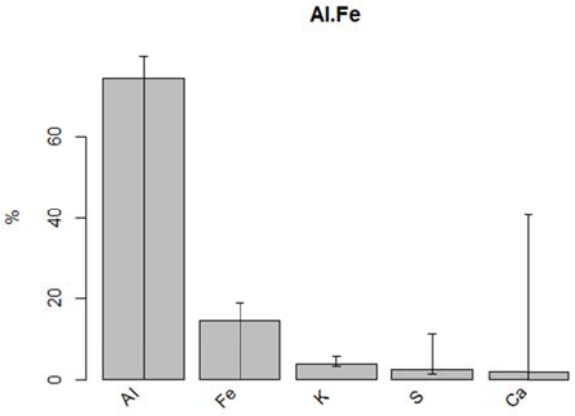

Al.K.Mg
%
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
P
Ba
Sr
V
Ce
Zn
Rb
Cu
La
Cr

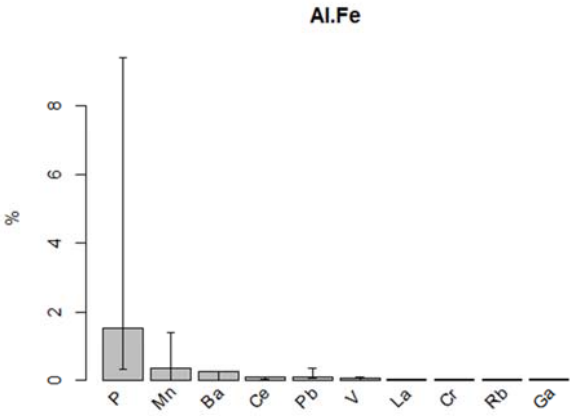

Al.K.Mg
%
0.000
0.005
0.010
0.015
Zr
Th
Ni
Sc
As
U
Hg
Se
Nb
Y

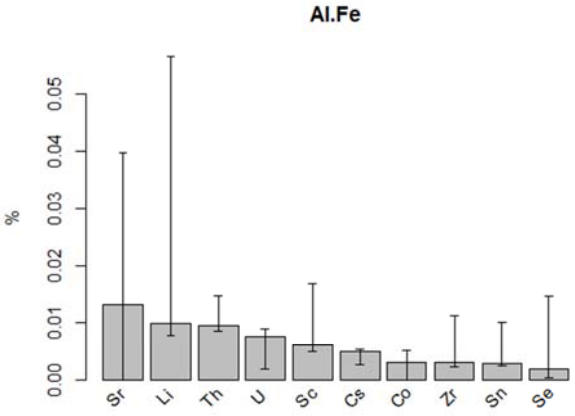

Fe.Ca.P
%
0
10
20
30
40
50
Fe
Ca
P
S
Al
K

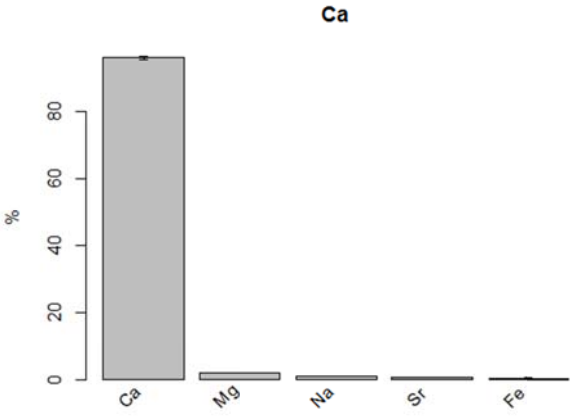

Fe.Ca.P
%
0.0
0.1
0.2
0.3
0.4
Zn
Mn
Ba
Ce
Pb
Ni
La
Sr
As

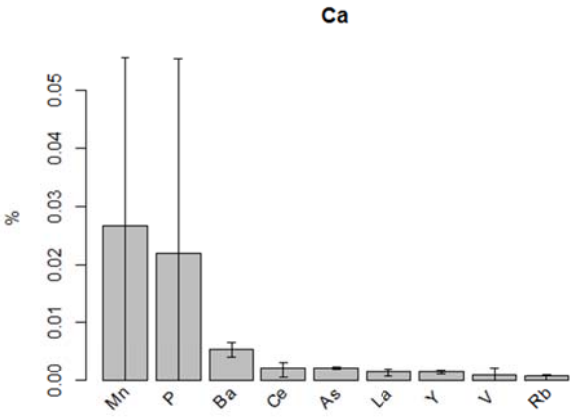

Fe.Ca.P
%
0.000
0.005
0.010
0.015
0.020
Y
Cr
Rb
V
U
Sn
Cd
Hg
Yb
Se

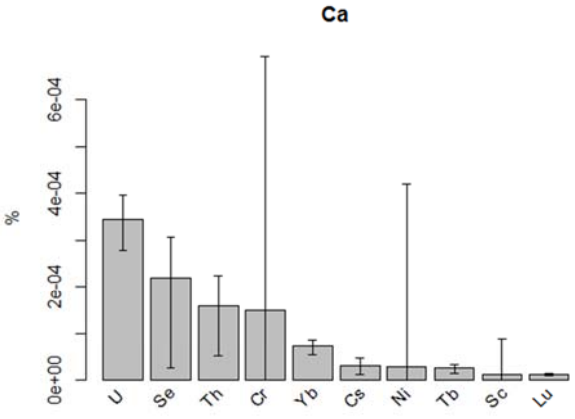

Al.Fe
%
0
20
40
60
Al
Fe
K
S
Ca

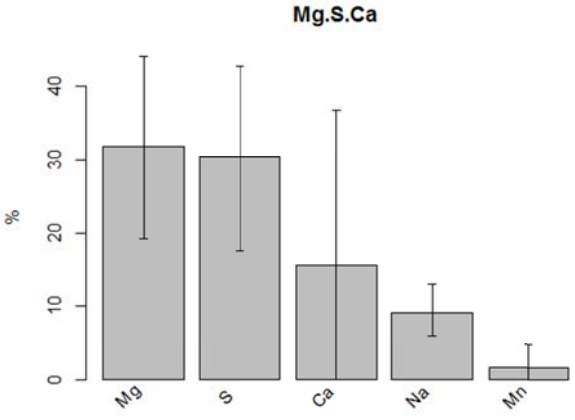

Al.Fe
%
0
2
4
6
8
P
Mn
Ba
Ce
Pb
V
La
Cr
Rb
Ga

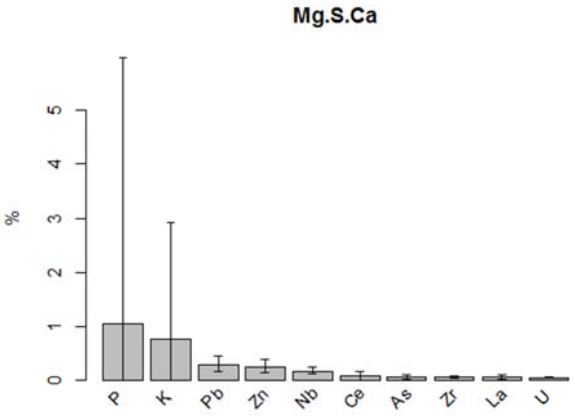

Al.Fe
%
0.00
0.01
0.02
0.03
0.04
0.05
Sr
Li
Th
U
Sc
Cs
Co
Zr
Sn
Se

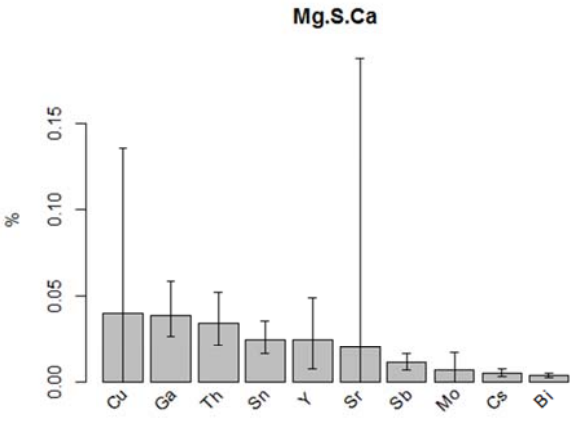

Ca
%
0
20
40
60
80
Ca
Mg
Na
Sr
Fe

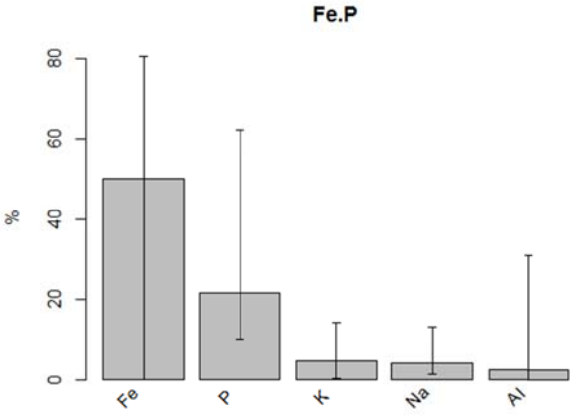

Ca
%
0.00
0.01
0.02
0.03
0.04
0.05
Mn
P
Ba
Ce
As
La
Y
V
Rb

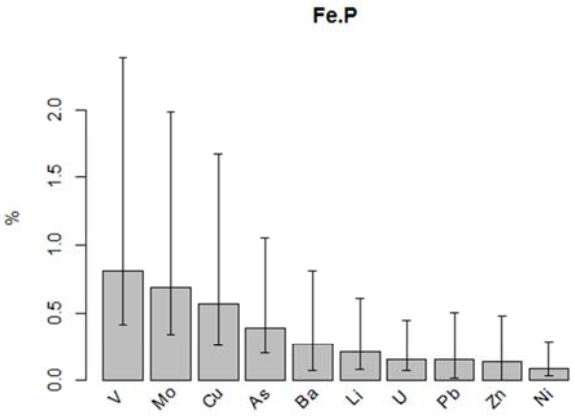

Ca
%
0e+00
2e-04
4e-04
6e-04
U
Se
Th
Cr
Yb
Cs
Ni
Tb
Sc
Lu

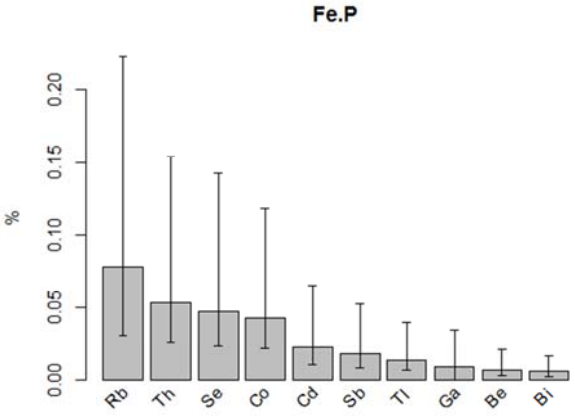

Mg.S.Ca
%
0
10
20
30
40
Mg
S
Ca
Na
Mn

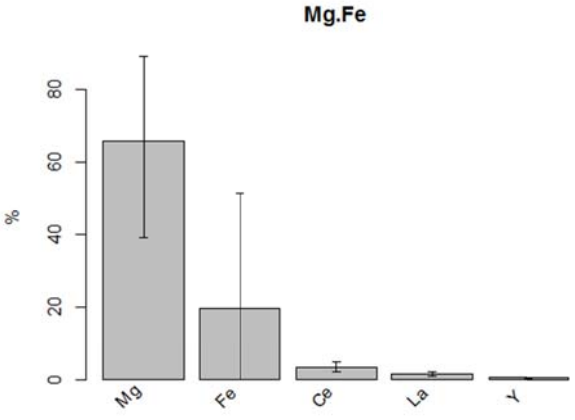

Mg.S.Ca
%
0
1
2
3
4
5
P
K
Pb
Zn
Nb
Ce
As
Zr
La
U

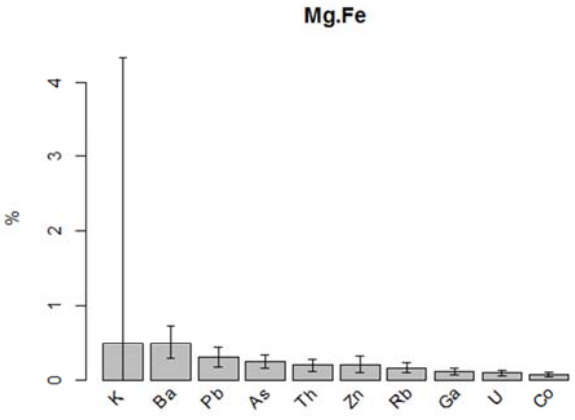

Mg.S.Ca
%
0.00
0.05
0.10
0.15
Cu
Ga
Th
Sn
Y
Sr
Sb
Mo
Cs
Bi

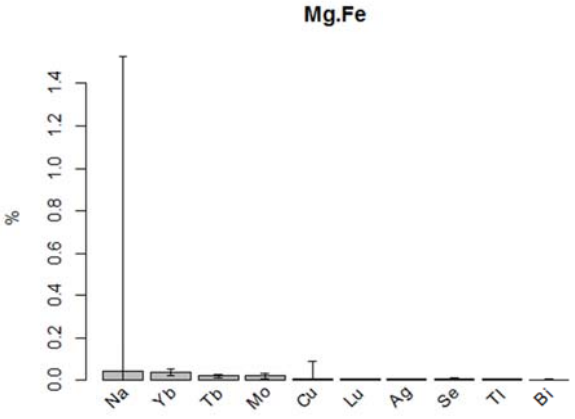

Fe.P
%
0
20
40
60
80
Fe
P
K
Na
Al

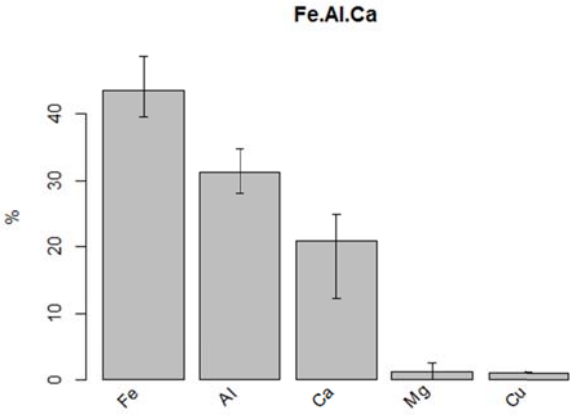

Fe.P
%
0.0
0.5
1.0
1.5
2.0
V
Mo
Cu
As
Ba
Li
U
Pb
Zn
Ni

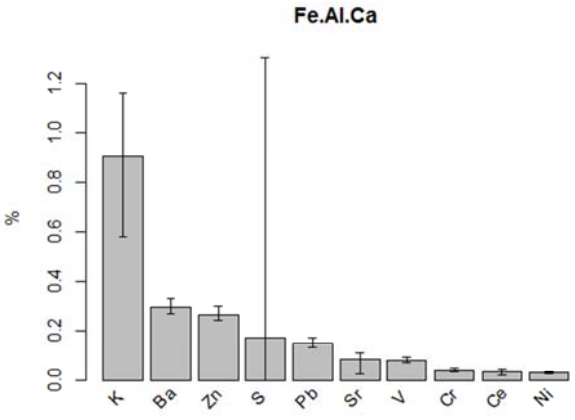

Fe.P
%
0.00
0.05
0.10
0.15
0.20
Rb
Th
Se
Co
Cd
Sb
Tl
Ga
Be
Bi

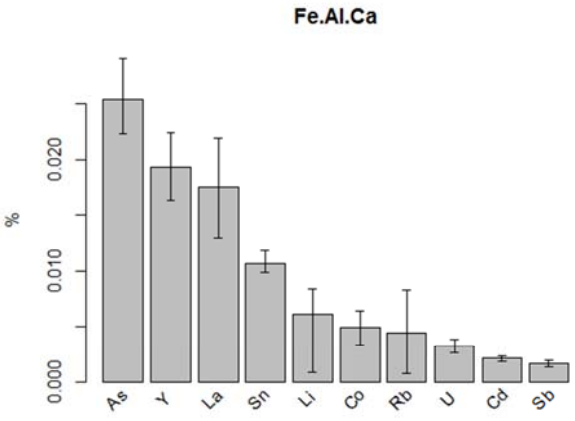

Mg.Fe
%
0
20
40
60
80
Mg
Fe
Ce
La
Y

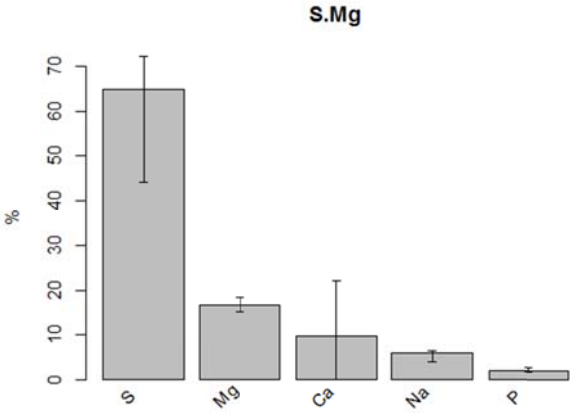

Mg.Fe
%
0
1
2
3
4
K
Ba
Pb
As
Th
Zn
Rb
Ga
U
Co

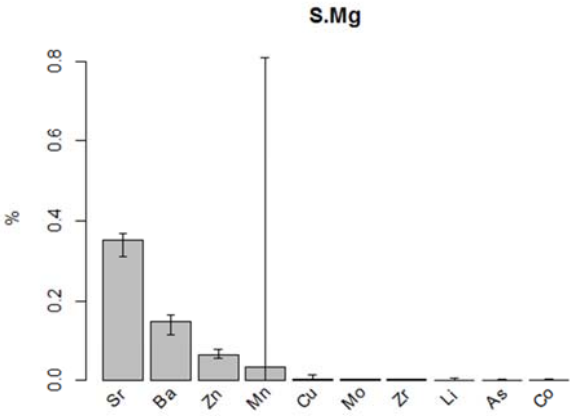

Mg.Fe
%
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
Na
Yb
Tb
Mo
Cu
Lu
Ag
Se
Tl
Bi

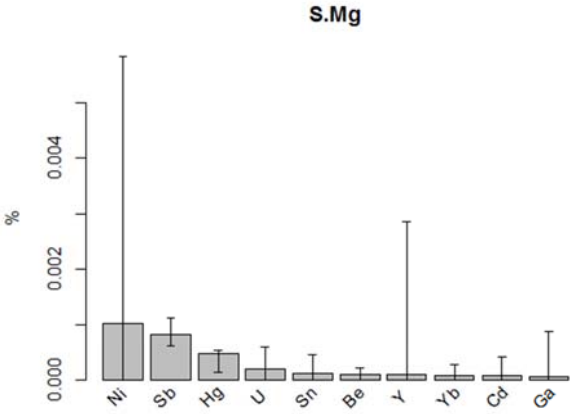

Fe.Al.Ca
%
0
10
20
30
40
Fe
Al
Ca
Mg
Cu

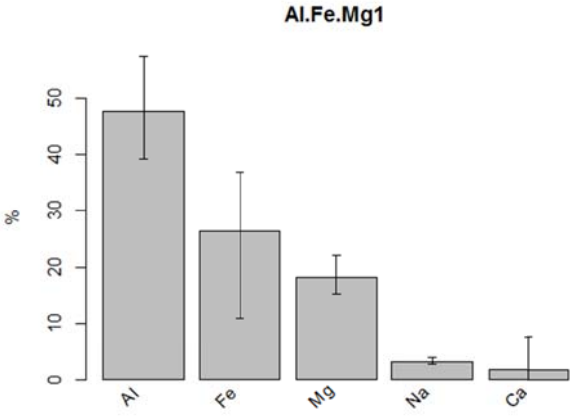

Fe.Al.Ca
%
0.0
0.2
0.4
0.6
0.8
1.0
1.2
K
Ba
Zn
S
Pb
Sr
V
Cr
Ce
Ni

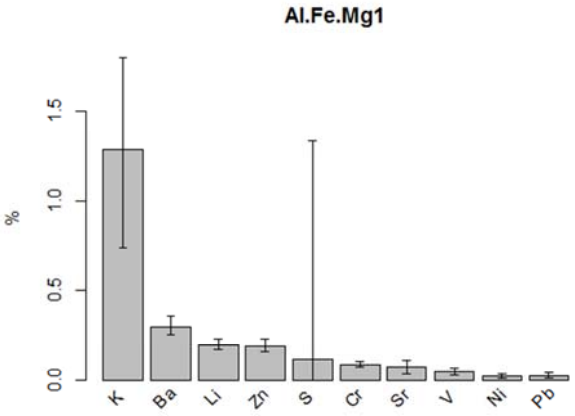

Fe.Al.Ca
%
0.000
0.010
0.020
As
Y
La
Sn
Li
Co
Rb
U
Cd
Sb

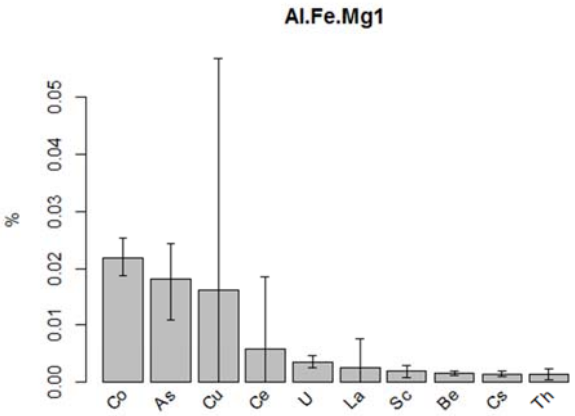

S.Mg
%
0
10
20
30
40
50
60
70
S
Mg
Ca
Na
P

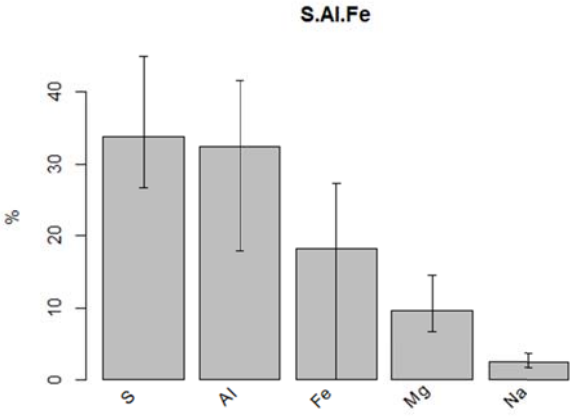

S.Mg
%
0.0
0.2
0.4
0.6
0.8
Sr
Ba
Zn
Mn
Cu
Mo
Zr
Li
As
Co

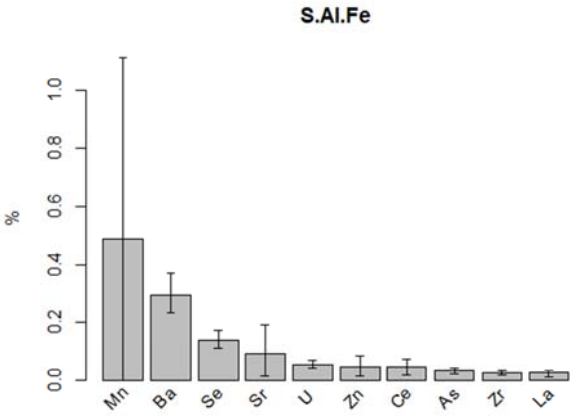

S.Mg
%
0.000
0.002
0.004
Ni
Sb
Hg
U
Sn
Be
Y
Yb
Cd
Ga

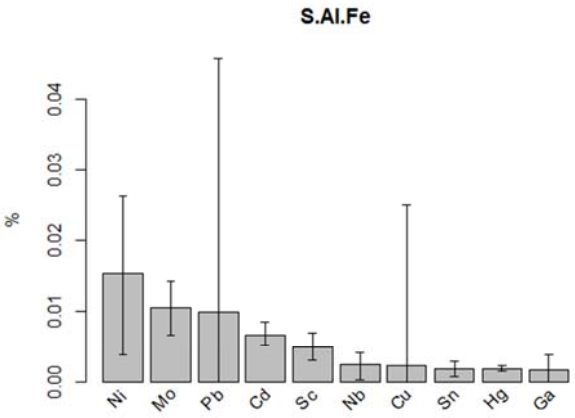

Al.Fe.Mg1
%
0
10
20
30
40
50
Al
Fe
Mg
Na
Ca

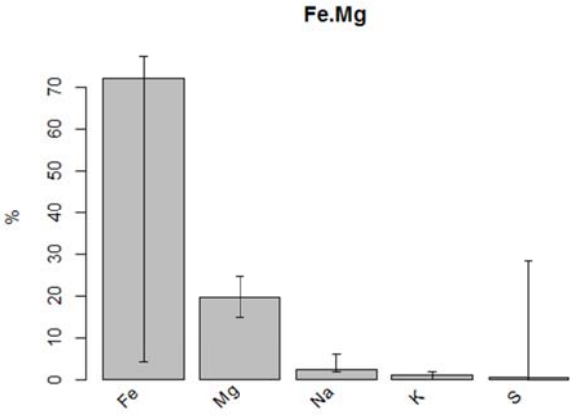

Al.Fe.Mg1
%
0.0
0.5
1.0
1.5
K
Ba
Li
Zn
S
Cr
Sr
V
Ni
Pb

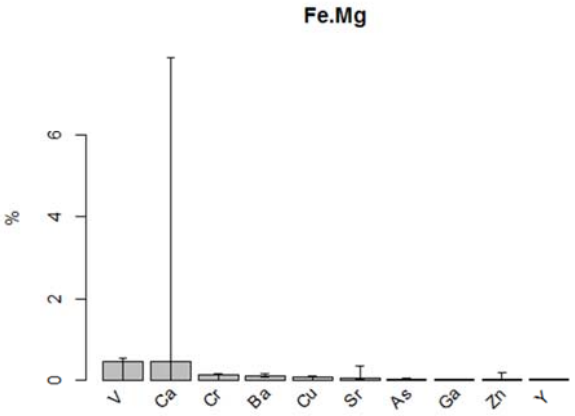

Al.Fe.Mg1
%
0.00
0.01
0.02
0.03
0.04
0.05
Co
As
Cu
Ce
U
La
Sc
Be
Cs
Th

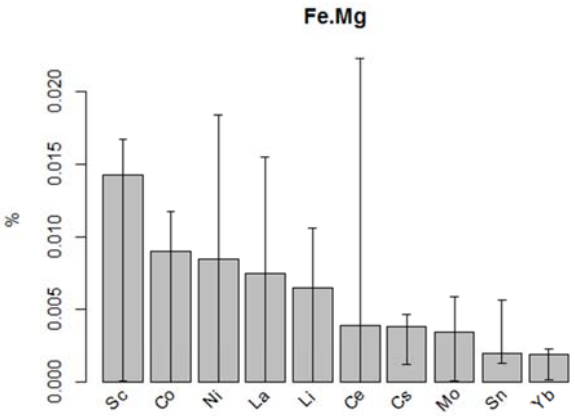

S.Al.Fe
%
0
10
20
30
40
S
Al
Fe
Mg
Na

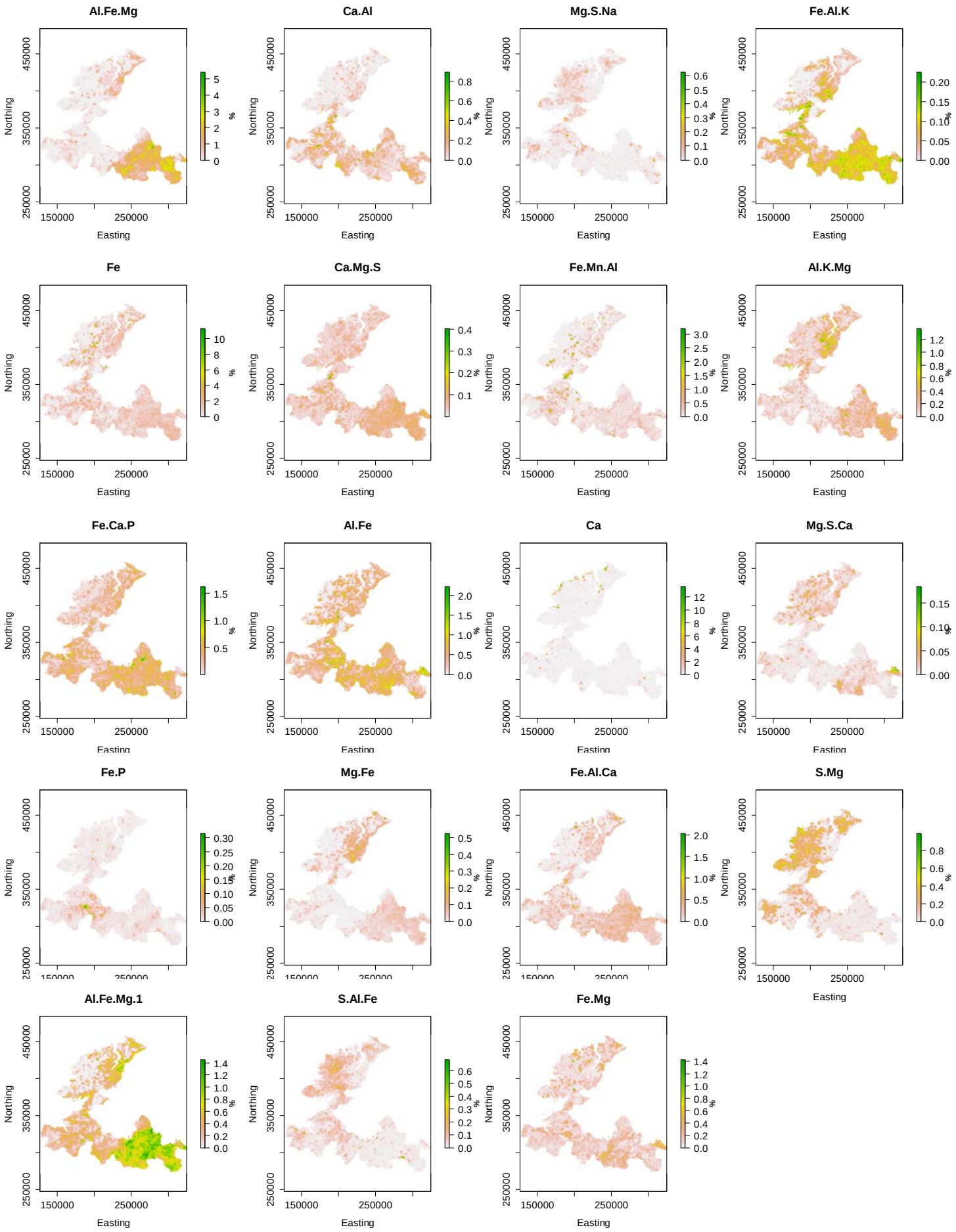

S.Al.Fe
%
0.0
0.2
0.4
0.6
0.8
1.0
Mn
Ba
Se
Sr
U
Zn
Ce
As
Zr
La


S.Al.Fe
%
0.00
0.01
0.02
0.03
0.04
Ni
Mo
Pb
Cd
Sc
Nb
Cu
Sn
Hg
Ga


Fe.Mg
%
0
10
20
30
40
50
60
70
Fe
Mg
Na
K
S


Fe.Mg
%
0
2
4
6
V
Ca
Cr
Ba
Cu
Sr
As
Ga
Zn
Y


Fe.Mg
%
0.000
0.005
0.010
0.015
0.020
Sc
Co
Ni
La
Li
Ce
Cs
Mo
Sn
Yb


Al.Fe.Mg
Ca.Al
Mg.S.Na
Fe.Al.K
Fe
Ca.Mg.S
Fe.Mn.Al
Al.K.Mg
Fe.Ca.P
Al.Fe
Ca
Mg.S.Ca
Fe.P
Mg.Fe
Fe.Al.Ca
S.Mg
Al.Fe.Mg.1
S.Al.Fe
Fe.Mg
Northing
Easting
150000
250000
250000
350000
450000
%