Application of the Tellus Border soil chemistry data to the identification of sources of diffuse pollution in Ireland

Mark Cave and Louise Ander

British Geological Survey Keyworth Nottingham NG12 5GG

Introduction

The aim of this study is to use the geochemical survey data of the 3475 soil samples collected across the Tellus border project area (Figure 1) along with data on soil properties (pH and Loss on Ignition), land classification information (solid geology, quaternary geology and land use) as well as geographic information (population density and elevation) to help identify chemical inputs from anthropogenic sources.

Methodology

The study uses a three tiered approach:

i. Examination of the shape of the distributions of individual elements in a similar manner to Zhang (2008) through skew and kurtosis statistics and use a comparison of these parameters along with other literature to identify pathfinder elements for anthropogenic contamination.

ii. Use multivariate analysis of the geochemistry data to identify anthropogenic inputs. Specifically, a self-modelling mixture resolution (SMMR) algorithm (Cave 2008) which identifies the underlying geochemical sources and gives a quantitative estimate of how much of the anthropogenic pathfinder elements are associated with each source.

iii. Model the anthropogenic pathfinder elements concentration using soil properties, land classification information and geographic information using the Random Forest robust machine learning model approach. This allows for both continuous and categorical predictors and non-linear relationships (Breiman 1996) to be assimilated into the model. The model will produce additional evidence about the parameters which control the geochemistry of the soils.

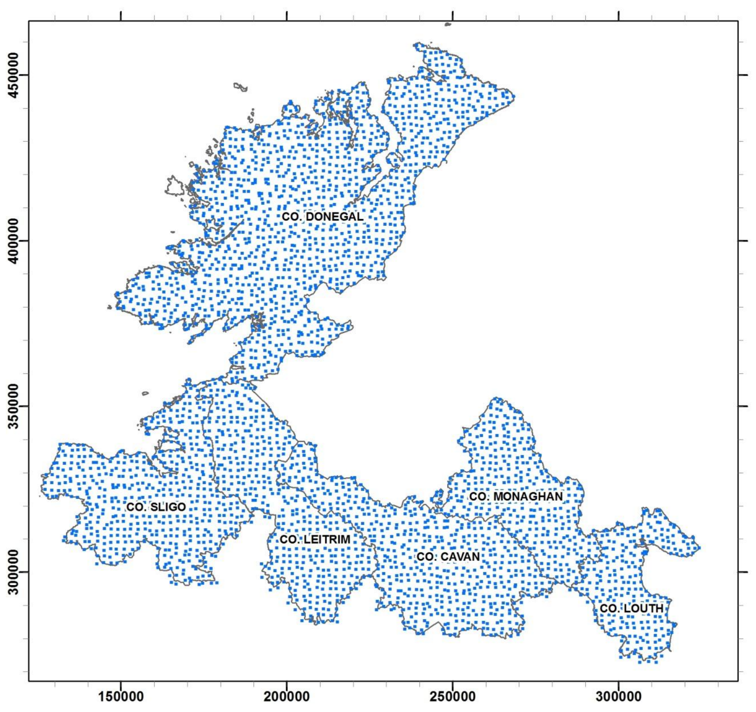

Figure 1 Tellus border soil sampling locations

Element Distribution Shape

A plot of the skewness (measure of how much the distribution deviates from a symmetric distribution) against the kurtosis (a measure of the distribution "peakedness") for the data distributions for each element was used a guide to the source of the elements in the soil samples (Figure 2). Elements with high positive skew and broader tails are more likely to be derived from anthropogenic sources. Using this data and other studies (e.g. Ander *et al* 2013 , Coggins *et al* 2006) six elements were assigned as pathfinders for diffuse anthropogenic pollution. – **As, Cd, Hg, Pb, Sb and U**.

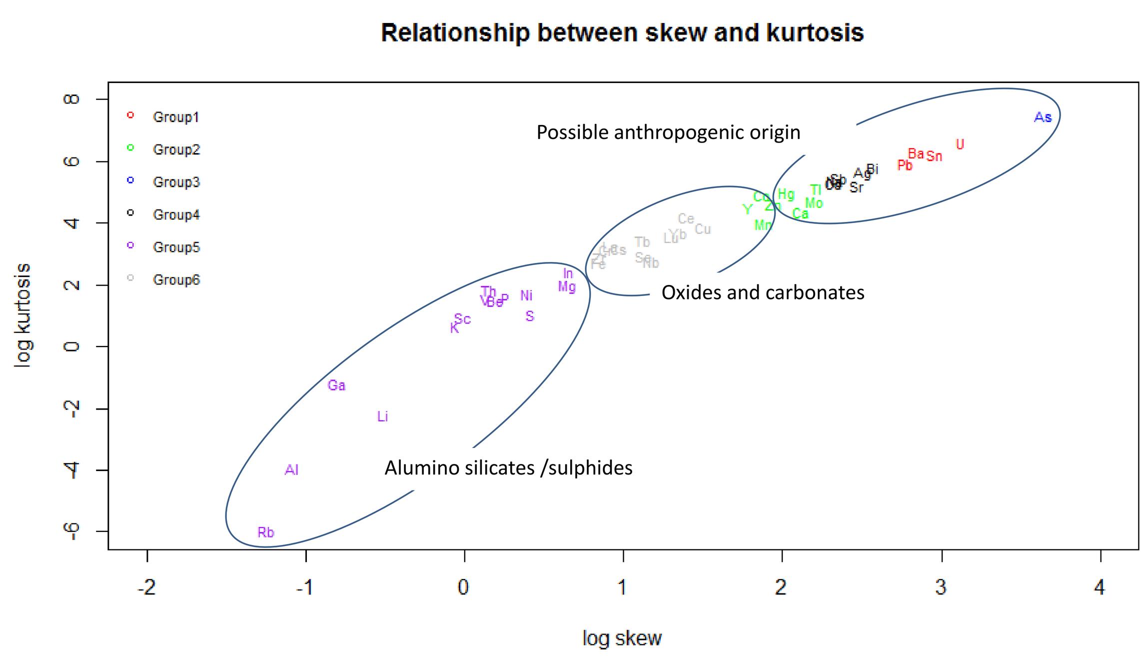

Figure 2 Plot of skewness against kurtosis for the element data distributions in the Tellus soil samples; used as a screening tool to identify elements associations

Source Apportionment

The SMMR algorithm identified 19 separate geochemical components in the Tellus soil data set. The spatial distributions are shown in Figure 3. Each component has been named using a combination of the elements that make up more than 10% of the geochemical component. The SMMR procedure also supplies the composition of each component and the proportion of each element that is associated with each component .

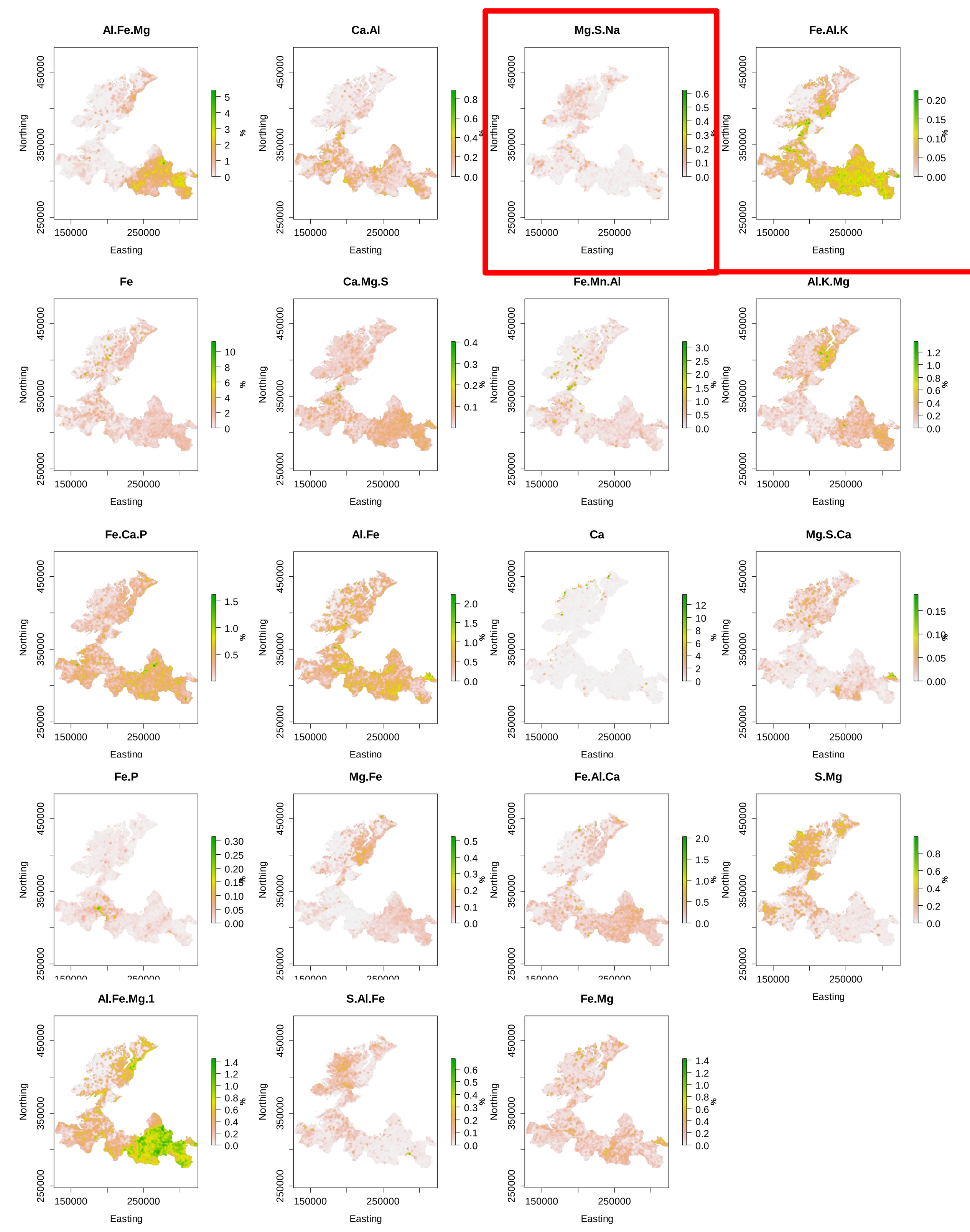

Figure 3 Spatial distribution of the geochemical sources identified by the SMMR

Source Apportionment (continued)

Examination of the composition of the 19 geochemical components shows one with particularly high concentrations of pathfinder elements (Mg.S.Na component, Figure 4). Figure 5 shows bar plots of the concentration of each of the pathfinder elements associated with each geochemical component. The Mg.S.Na component has the highest individual contribution for Pb and Sb and is an important source for Cd and Hg but is has less contribution to As and U.

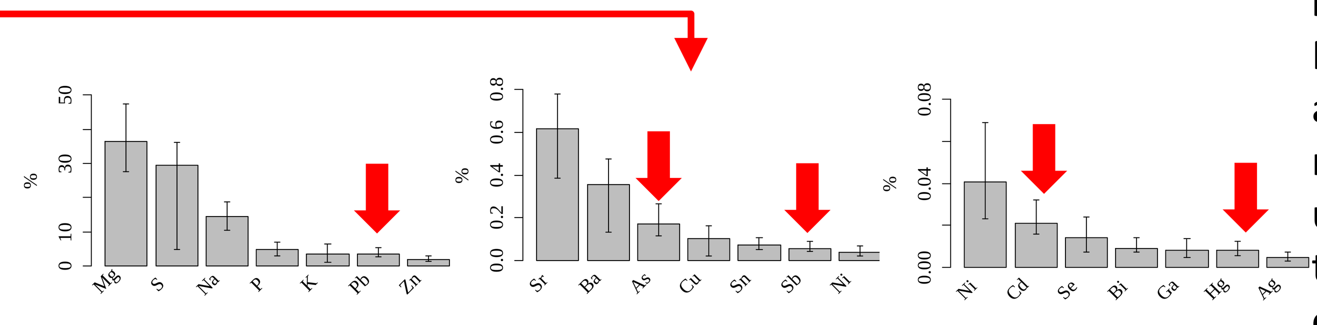

Figure 4 Chemical composition of the Mg.S.Na geochemical component showing high concentrations of pathfinder elements

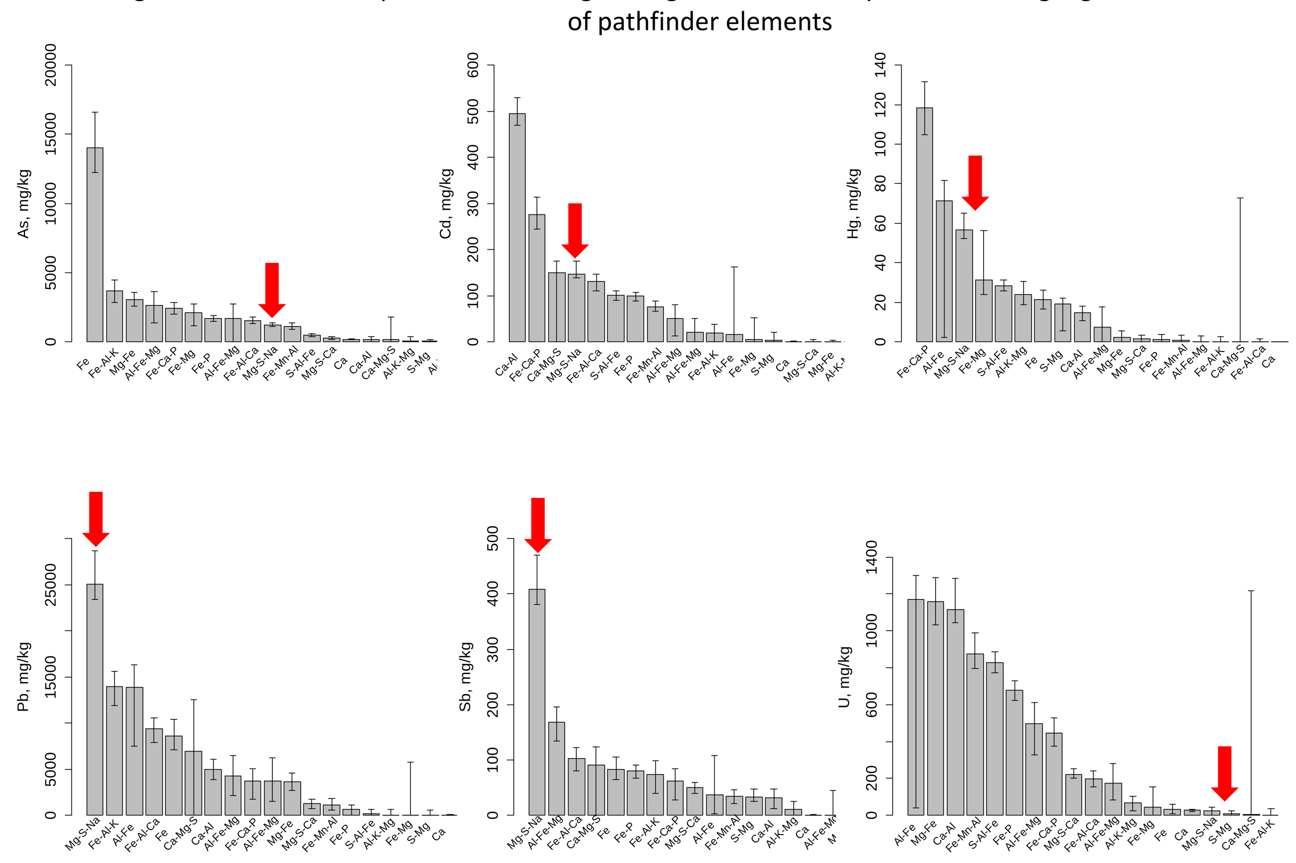

Figure 5 Concentration of each of the pathfinder elements associated with each SMMR geochemical component ; the Mg.S.Na component is marked with a red arrow

Random Forest Modelling

Random forest (RF) is a non-parametric algorithm which can be used for non-linear regression analysis. For this study it has been used to investigate the relationship between the SMMR components (predictors) and Loss on Ignition, LOI (dependant variable) and secondly the relationship between the pathfinder elements and soil properties (pH and Loss on Ignition), land classification information (solid geology, quaternary geology and land use) as well as geographic information (population density and elevation). Figure 6 shows the 5 SMMR components which have the largest effect in the RF model of LOI. Increasing concentrations of the Mg.S.Na component shows an increasing effect on LOI suggesting that it also associated with organic matter.

Figure 7 shows how the soil properties and geographic variables affect Pb concentration (Figure 8) in an RF model. There is a clear relationship between increasing elevation and Pb concentration and higher population density leading to higher Pb concentrations.

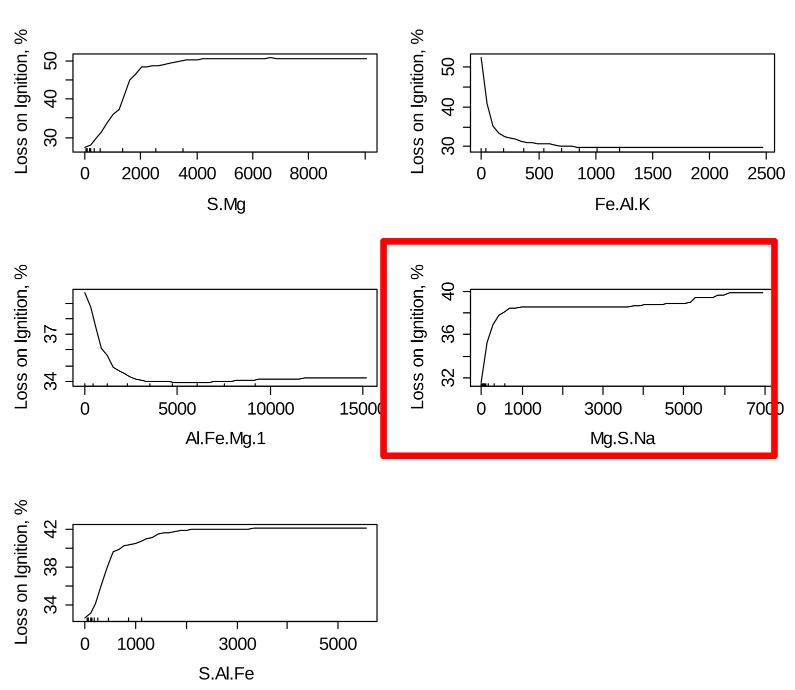

Figure 6 Univariate plots showing the behaviour of LOI with respect to the five most important parameters in the RF model

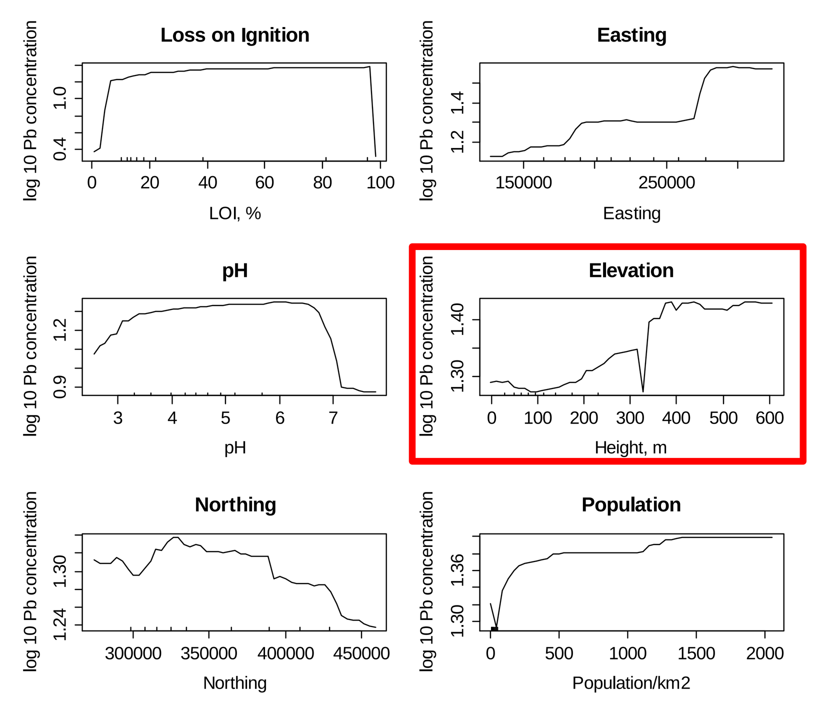

Figure 7 Univariate plots showing the behaviour of Pb with respect to geographical variable s and soil properties

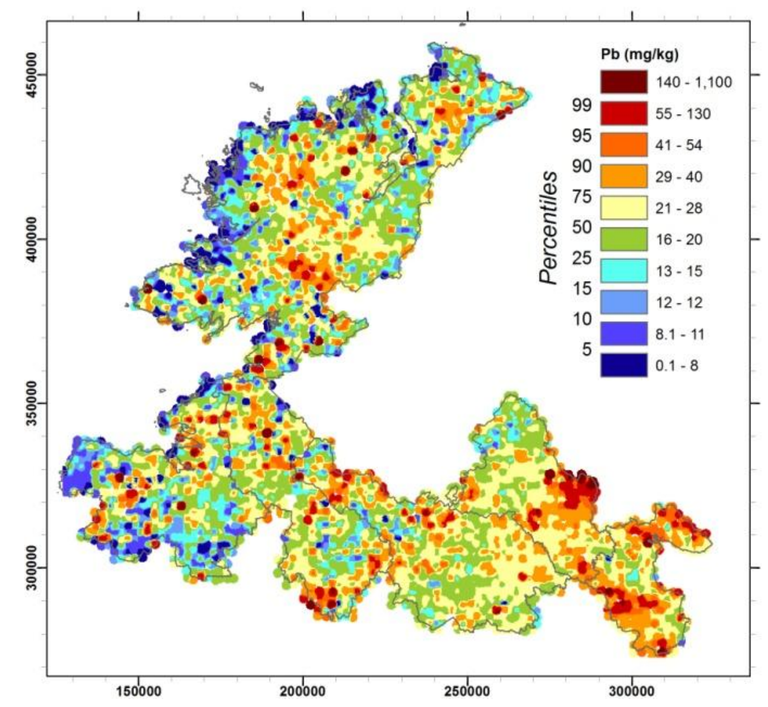

Figure 8 Pb content of soils in the Tellus border area

Aerial Deposition

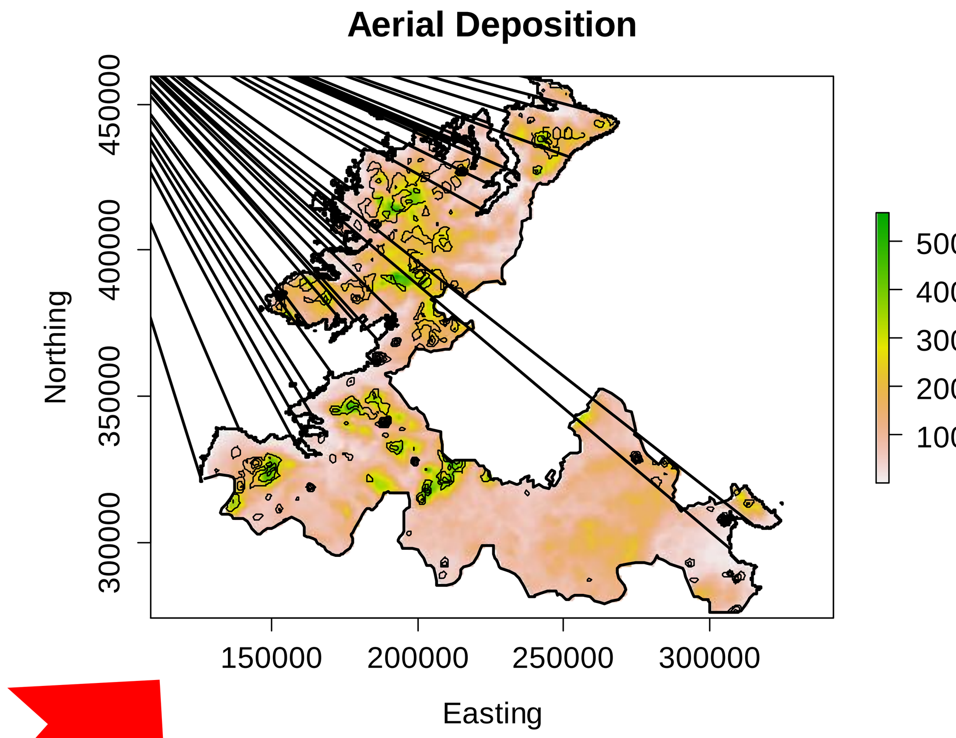

Figure 9 Suspected anthropogenic component (contour lines) superimposed on the elevation data (colour ramp) showing a close spatial match

Synthesis of Results

- **Source apportionment identifies a geochemical component with high concentrations of the pathfinder anthropogenic elements**
- **This component accounts for a large proportion of the mass of the anthropogenic pathfinder elements Pb, Sb, Hg and Cd**
- **RF modelling of Loss on ignition using the SMMR components shows that this component has significant organic content**
- **RF modelling of geological and geographical parameters shows elevation to be a controlling factor for Pb concentrations (Figures 8 and 9)**
- **The concentration of the identified source, maps closely to points of high elevation (Figure 9)**
- **The overall conclusion is that an anthropogenic diffuse pollution source has been identified which is largely controlled by aerial deposition**

References

ANDER, E L, JOHNSON, C C, CAVE, M R, PALUMBO-ROE, B, NATHANAIL, C P, and LARK, R M. 2013. Methodology for the determination of normal background concentrations of contaminants in English soil. *Science of the Total Environment*, Vol. 454–455, 604-618.

Breiman, L. (1996). "Bagging predictors." Machine Learning **24**(2): 123-140.

Cave, M. R. (2009). The use of self modelling mixture resolution for the interpretation of geochemical data, British Geological Survey.

Coggins, A. M., S. G. Jennings, et al. (2006). "Accumulation rates of the heavy metals lead, mercury and cadmium in ombrotrophic peatlands in the west of Ireland." Atmospheric Environment **40**(2): 260-278.

Zhang, C., D. Fay, et al. (2008). "Statistical analyses of geochemical variables in soils of Ireland." Geoderma **146**: 378-390.

TELLUS BORDER

The Tellus Border Results and Research Conference
24th October 2013. Hillgrove Hotel, Co. Monaghan, Ireland

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