Multivariate analysis of regional-scale geochemical data for environmental monitoring

McKinley, J. M., Grunsky, E., & Mueller, U. (2016). Multivariate analysis of regional-scale geochemical data for environmental monitoring. In Proceedings of the 11th International Conference on Geostatistics for Environmental Application Conference

Published in:
Proceedings of the 11th International Conference on Geostatistics for Environmental Application Conference

Document Version:
Peer reviewed version

Multivariate analysis of regional-scale geochemical data for environmental monitoring

Jennifer M McKinley*[1], Eric Grunsky[2], Ute Mueller [3]

[1]School of Geography, Archaeology and Palaeoecology, Queen's University Belfast, BT7 1NN, UK

[2] Department of Earth and Environmental Sciences, University of Waterloo, Waterloo, Ontario, Canada, N2L 3G1

[3] School of Science, Edith Cowan University, Joondalup, Western Australia, WA 6027

Abstract

A compositional multivariate approach is used to analyse regional scale soil geochemical data obtained as part of the Tellus Project generated by the Geological Survey Northern Ireland (GSNI). The multi-element total concentration data presented comprise XRF analyses of 6862 rural soil samples collected at 20cm depths on a non-aligned grid at one site per 2 km^2. Censored data were imputed using published detection limits. Using these imputed values for 46 elements (including LOI), each soil sample site was assigned to the regional geology map provided by GSNI initially using the dominant lithology for the map polygon. Northern Ireland includes a diversity of geology representing a stratigraphic record from the Mesoproterozoic, up to and including the Palaeogene. However, the advance of ice sheets and their meltwaters over the last 100,000 years has left at least 80% of the bedrock covered by superficial deposits, including glacial till and post-glacial alluvium and peat. The question is to what extent the soil geochemistry reflects the underlying geology or superficial deposits. To address this, the geochemical data were transformed using centered log ratios (clr) to observe the requirements of compositional data analysis and avoid closure issues. Following this, compositional multivariate techniques including compositional Principal Component Analysis (PCA) and minimum/maximum autocorrelation factor (MAF) analysis method were used to determine the influence of underlying geology on the soil geochemistry signature. PCA showed that 72% of the variation was determined by the first four principal components (PC’s) implying “significant” structure in the data. Analysis of variance showed that only 10 PC’s were necessary to classify the soil geochemical data. To consider an improvement over PCA that uses the spatial relationships of the data, a classification based on MAF analysis was undertaken using the first 6 dominant factors. Understanding the relationship between soil geochemistry and superficial deposits is important for environmental monitoring of fragile ecosystems such as peat. To explore whether peat cover could be predicted from the classification, the lithology designation was adapted to include the presence of peat, based on GSNI superficial deposit polygons and linear discriminant analysis (LDA) undertaken. Prediction accuracy for LDA classification improved from 60.98% based on PCA using 10 principal components to 64.73% using MAF based on the 6 most dominant factors. The misclassification of peat may reflect degradation of peat covered areas since the creation of superficial deposit classification. Further work will examine the influence of underlying lithologies on elemental concentrations in peat composition and the effect of this in classification analysis.

Introduction

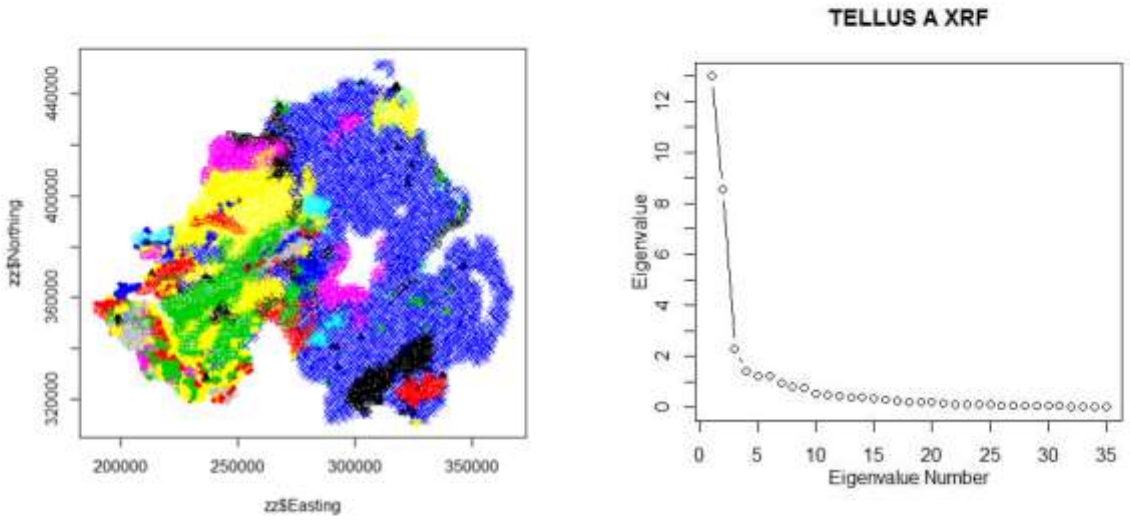
A diversity of rock types observed across Northern Ireland represents three basement terranes (Mitchell 2004). The Grampian Terrane and associated rocks in the northwest have metamorphic igneous and sedimentary origins spanning the Proterozoic Era. Psammites and semi-pelites are the dominant rock type, with subordinate sandstones and conglomerates. The Midland Valley Terrane hosts Palaeozoic igneous formations and Late Devonian-Early Carboniferous sedimentary rocks. Rock types comprise red sandstones, limestone and mudstones with a less common conglomerates. The

Southern Uplands-Down-Longford Terrane consists of granitic igneous intrusives and Lower Palaeozoic Ordovician and Silurian marine sedimentary rocks (lithic arenites and sandstones). Palaeogene flood basalts and lava-derived sedimentary clays cover a large portion of the Midland Valley and Grampian basement rocks in the northeast of the country. The advance of ice sheets and their meltwaters over the last 100,000 years has left at least 80% of the bedrock covered by superficial deposits, including glacial till and post-glacial alluvium and peat. Monitoring peat coverage has become an important in calculating soil carbon stocks due to the relatively high carbon density of peat and organic-rich soils. This is particularly important for Ireland (and other western European countries), where some 16% of the land surface is covered by peat bog. In Northern Ireland, previous work has estimated the total amount of carbon stored in vegetation to be 4.4Mt compared to 386Mt stored within soils such as peat (Cruickshank et al. 1998).

A number of comprehensive regional and national soil sampling programmes have been completed across the UK including the British Geological Survey (BGS) G-BASE survey and regional soil surveys for Northern Ireland (Jordan et al. 2001). These have been used to assess baseline element concentrations in soils and normal background concentrations of contaminants (e.g. Ander et al. 2013). The Northern Ireland Tellus Survey (GSNI 2007; Young and Donald 2013) included a ground based geochemical survey in which 6862 rural soil samples were collected between 1994 and 2006. Samples were collected on a grid of one sample site every 2km^2, with soils being collected at depths of 20 and 50cm. The samples were analysed by X-ray fluorescence spectrometry (XRFS) for 60 elements and inorganic compounds. Tellus Survey field methods and analytical methodology are described in Smyth (2007) and Young and Donald (2013). Soils and parent geology across Northern Ireland are typical for the geological and pedological conditions across the UK (Jordan et al. 2001), therefore the Tellus data set provides the basis for a comprehensive study with relevancy for the whole of the UK. This paper explores the extent to which soil geochemistry can be used to classify the underlying geology and moreover differentiate superficial deposits such as peat.

Methods

A compositional multivariate approach is adopted in this paper to analyse the rural Tellus soil data sampled at 20cm depths. Many geochemical datasets contain values that are reported at less than the lower limit of detection and these values are generally termed “censored” Grunsky (2010). For this study, published detection limits (Young and Donald 2013) were used to impute censored data, this resulted in 46 elements (including LOI%) available for further analysis. Each soil sample site was assigned to the regional geology map provided by GSNI using the dominant lithology for the map polygon. This provided 30 lithologies with a suitable number of sample sites (Fig. 1a). A process discovery approach (Grunsky 2010) involves the use of unsupervised multivariate methods such as principal component analysis (PCA). An essential part of the process discovery phase is a suitable choice of transform to overcome the problem of closure. Therefore to account for the requirements of compositional data analysis and avoid closure issues, the geochemical data were transformed using centered log ratios (clr, Aitchison 1986). To provide a comparison, minimum/maximum autocorrelation (MAF) factors based on the clr data were computed from the 46 variables (Bandarian and Mueller, 2008; Mueller and Grunsky, 2016). This produced typicalities, which can be explained as a measure of the Mahalanobis distance (MD) of a given observation to the centroid of each class. Based on the F distribution and a chosen threshold of 0.95, an observation is allocated to a specific class if the MD is less than the critical value of the F distribution for the given degrees of freedom. The associated probability (typicality) is determined according to Garrett (1989). If the MD exceeds the critical value then there is a typicality of zero.

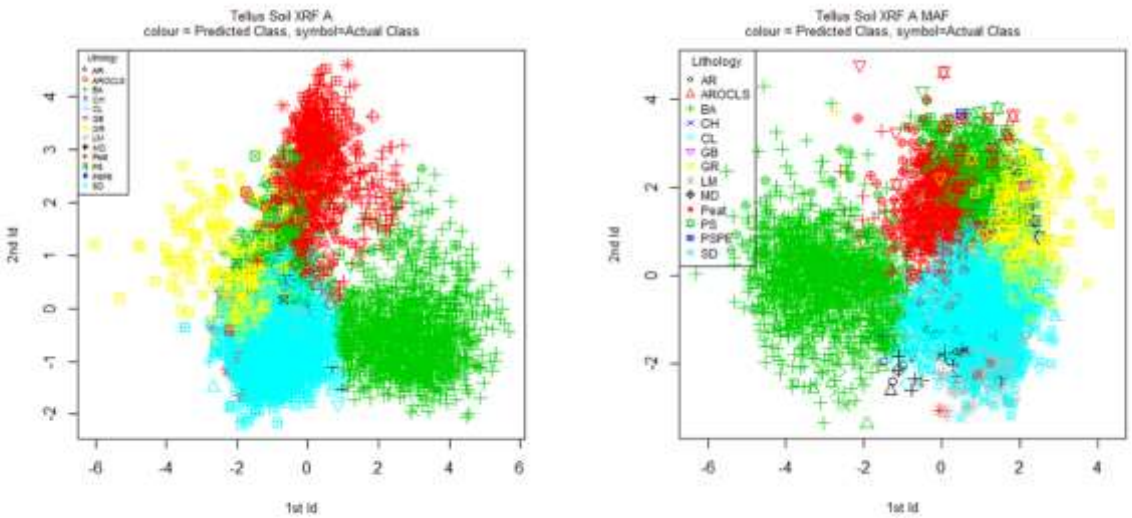


A) B)

Figure 1: A) Soil sample sites assigned to regional geology (GSNI) using the dominant lithology for the map polygon; B) PCA scree plot.

Results and discussion

PCA showed that 72% of the variation was determined by the first four principal components (PCs) implying "significant" structure in the data (Figure 1b). Analysis of variance showed that only 10 PC's were necessary to classify the soil geochemical data. To consider an improvement over PCA that uses the spatial relationships of the data, a classification based on MAF analysis was undertaken using the first 6 dominant factors. As the aim of this study was to elucidate the relationship between soil geochemistry and superficial deposits for environmental monitoring, the next stage of the analysis explored whether peat cover could be predicted from the classification. For this, the lithology designation was adapted to include the presence of peat, based on GSNI superficial deposit polygons and linear discriminant analysis (LDA) undertaken (Fig. 2).



A) B)

Figure 2: Linear discriminant analysis (LDA) including the presence of peat (shown as red symbols) using A) PCA and B) MAF analysis.

Using LDA the presence of peat is clearly differentiated from the other classifications. Prediction accuracy for LDA classification was found to be improved from 60.98% based on PCA using 10 principal components (Fig. 2a) to 64.73% using MAF based on the 6 most dominant factors (Fig.2b).

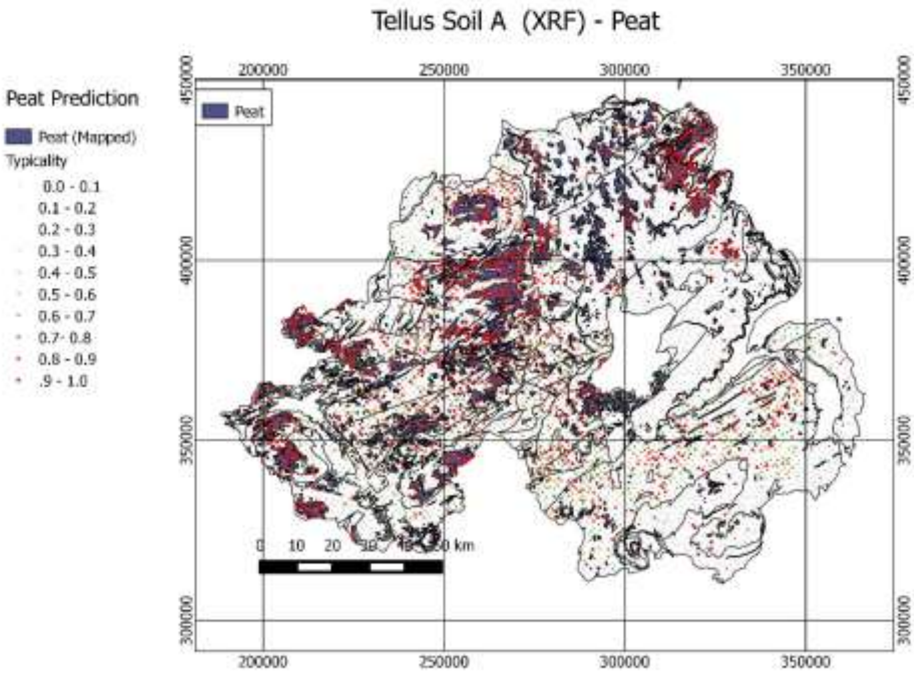


Figure 3: Plotted PCA typicalities overlain on areas of reported peat.

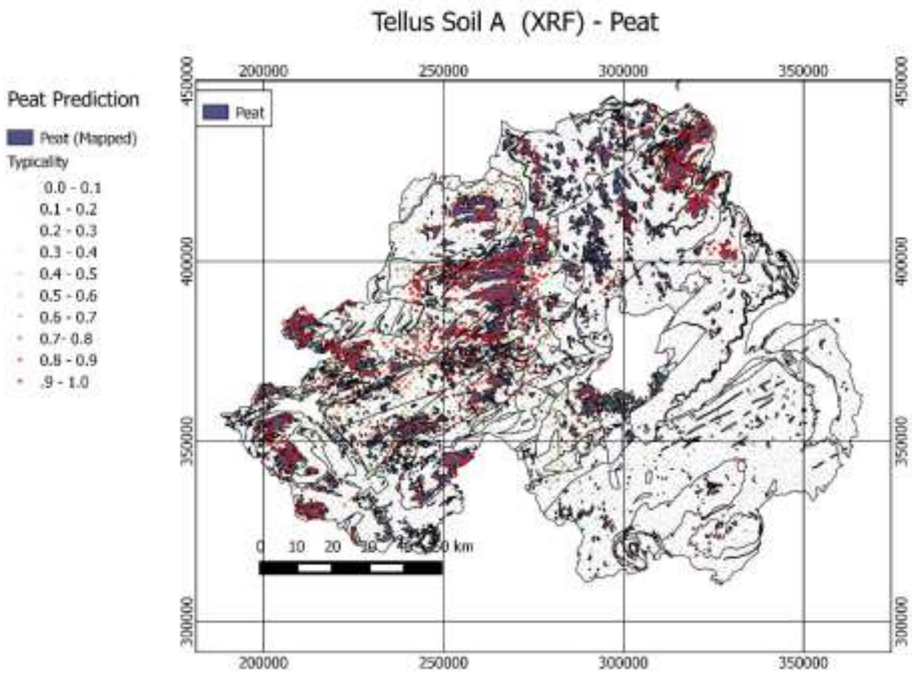


Figure 4: Plotted MAF typicalities overlain on areas of reported peat.

The plotted PCA (Fig. 3) and MAF typicalities (Fig. 4) demonstrate a good match between the reported peat areas and the highest probability or typicality for peat. However, there are areas of mapped peat where the predictions indicate a low probability of peat. The explanation for this misclassification of peat for both the PCA and MAF analysis may be twofold: these areas reflect degradation of peat covered areas since the creation of superficial deposit classification or a further refinement in the classification of peat is required. Irish peatland is divided into blanket peatland (approximately 85%) and raised peat bogs (approximately 15%; Tomlinson and Davidson 2000). Blanket bogs typically form on gentle slopes within upland regions (>315m above SL; Hamilton et al. 1982). The distribution of blanket bogs is more spatially continuous and associated with areas of high precipitation (rainfall exceeding 1200mm). Raised bogs develop primarily in lowland areas (<200m above Sea Level (SL); Wheeler 1995) where accumulating peat in fens becomes isolated from the groundwater supply. This process of accumulation gradually forms a dome of ombrogenous peat above the fen giving raised bogs a distinct topography, with the steep margins to the main bog expanse. Raised bogs are more limited in extent and occur as isolated features. The prediction of peat covered areas using MAF analysis methods, which use the spatial relationships of the data, have been more successful in predicting the more extensive upland blanket bogs than lowland raised bogs.

Conclusions

Compositional multivariate techniques, PCA and MAF analysis methods were used to determine the influence of underlying geology on the soil geochemistry signature. The approach was explored for environmental monitoring of peat to ascertain whether peat cover could be predicted from the classification. Using LCA the presence of peat was clearly differentiated from the other lithological classifications. Moreover, prediction accuracy for LDA classification improved using MAF analysis. In an attempt to reduce the number of areas of misclassification of peat further work will examine the influence of underlying lithologies on elemental concentrations in peat composition and the effect of this in classification analysis.

Acknowledgements

The Tellus Project was carried out by GSNI and funded by The Department for Enterprise, Trade and Investment (DETINI) and The Rural Development Programme through the Northern Ireland Programme for Building Sustainable Prosperity.

References

Aitchison, J., 1986. The statistical analysis of compositional data. Chapman and Hall, New York. 416p.

Ander, E.L., Johnson, C.C., Cave, M.R., Palumbo-Roe, B., Nathanail, C.P., Lark, R.M. 2013. Methodology for the determination of normal background concentrations of contaminants in English soil. Science of The Total Environment, 454-455. 604-618.

Bandarian E.M., Mueller, U.A., 2008. Reformulation of direct minimum/maximum Autocorrelation factors as a generalised Eigenvalue problem.In J. M. Ortiz & X. Emery (Eds.), Geostats 2008 (Vol. 2, pp. 1173-1178). Santiago, Chile: Universidad De Chile.

Cruickshank, M.M., Tomlinson, R.W., Devine, P.M., Milne, R. 1998. Carbon in the vegetation and soils of Northern Ireland. Biology and Environment: Proceedings of the Royal Irish Academy 98B, 1, 9-21.

Garrett, R.G. 1989. A robust multivariate allocation procedure with applications to geochemical data. In: Agterberg, F and Bonham-Carter, G.F. (ed), Statistical Applications in Earth Sciences, Geological Survey Paper 89-9, 309-318.

Geological Survey Northern Ireland (GSNI) 2007. Tellus project overview. http://www.bgs.ac.uk./gsni/Tellus/index.html

Grunsky, E.C., 2010.The interpretation of geochemical survey data; Geochemistry, Exploration, Environment Analysis, 10(1), 27-74.

Hamilton, A.C., 1882. Peatland, in: J. G. Cruickshank, and D. N. Wilcock, (Eds) Northern Ireland environment and natural resources, QUB, Belfast, 185-206.

Jordan, C., Higgins, A., Hamill K., Cruickshank, J.G., 2001. The Soil Geochemical Atlas of Northern Ireland. Department of Agriculture and Rural Development, NI.

Mitchell, W., 2004. The Geology of Northern Ireland: Our Natural Foundation, Geological Survey of Northern Ireland, Belfast, 2004.

Mueller, U.A., Grunsky, E.C., 2016. Multivariate Spatial Analysis of Lake Sediment Geochemical Data; Melville Peninsula, Nunavut, Canada, Applied Geochemistry, doi:10.1016/j.apgeochem.2016.02.007

Smyth, D., 2007. Methods used in the Tellus geochemical mapping of Northern Ireland. British Geological Survey Open Report OR/07/022.

Tomlinson., R.W., Davidson, L., 2000. Estimates of carbon stores in four Northern Irish lowland raised bogs. Suo 51. 169-179.

Wheeler., B.D., Shaw, S.C., 1995. Restoration of Damaged Peatlands. HMSO, London.

Young, M.E., Donald, A.E., 2013. A Guide to the Tellus Data. Belfast: Geological Survey of Northern Ireland, Belfast.