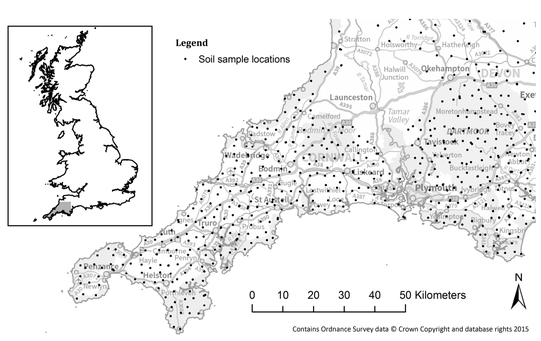
British Geological Survey
NATURAL ENVIRONMENT RESEARCH COUNCIL

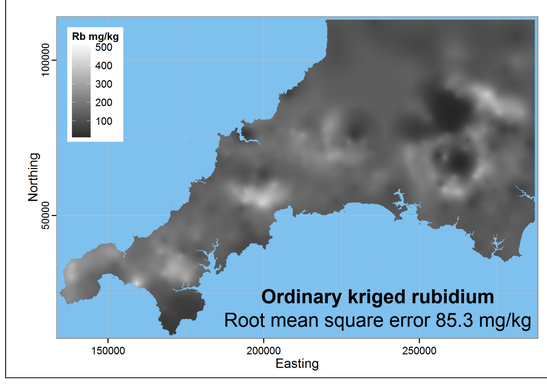
Geological mapping using high resolution regression modelled soil geochemistry

Geochemical map production:

1. For this study we make use of element concentration data from 517 shallow soil samples collected during the Geochemical Baselines Survey of the Environment's 2012 field season in south west England.

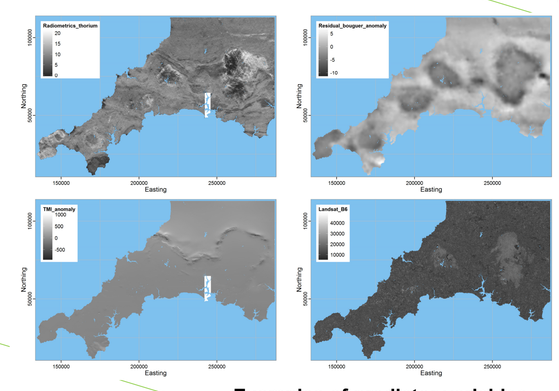


Shallow soil sample sites – south west England, 2012

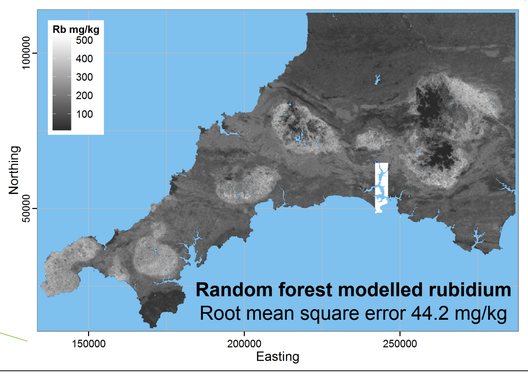


2. Traditional geochemical maps produced by simple univariate interpolation methods such as ordinary kriging are limited by the spatial sampling density of the survey, and are blind to other environmental information.

3. Instead, we build a compositional regression model using random forest (Breiman 2001) that is able to predict element concentrations in soil given high resolution geophysics and remote sensing data. This approach is able to provide increased resolution, prediction accuracy and interpretability compared to ordinary kriged maps.



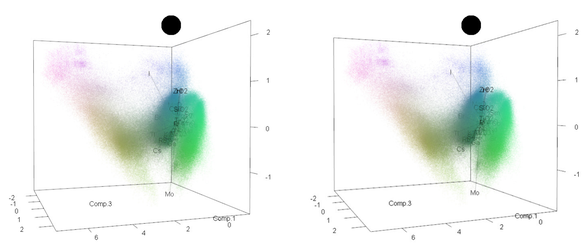
Examples of predictor variables used by the model



Geological mapping:

1. Soil geochemistry contains a huge amount of information about both subsurface and surface conditions. Thanks to the quality of modern geochemical analysis the sensitivity and consistency of this information far exceed anything that could be obtained by direct human observation. However, the breadth of information means that careful processing is required in order to extract only the relevant information for any given task: in this case to map lithology using soil geochemistry.

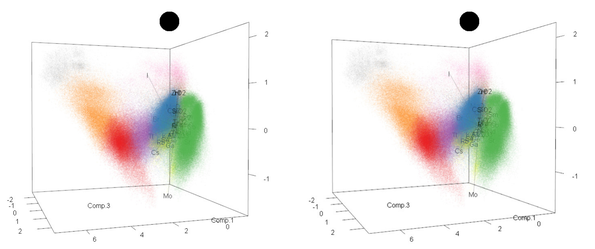
2. For the first processing step we select a sub-composition of elements which can be observed to be mostly unaffected by hydrothermal activity: Hydrothermal activity transcends lithological boundaries and minimising its influence is therefore necessary in order to cleanly resolve lithological boundaries. Compositional principal component analysis is then used to further boost the signal in favour of lithological discrimination by removing uncorrelated noise.



Cross your eyes so that the two black circles become one.

3. As can be seen from the map and stereoscopic triplot above, the first three principal components of our tailored sub-composition capture lithological variation well. However, there is contamination from peat (pink) due to the peat and lithological signals varying on the same principal axes. Another imperfection is the somewhat blurred boundaries of the granites (brown) – a combination of down-slope soil migration and perhaps a failure to fully omit hydrothermal signals.

4. We then classify the map using hierarchical clustering of our sub-composition's non-noise principal components. We ask for nine clusters in order that the map remains easily comprehensible. The clusters are assigned on the basis of compositional variation alone. As a result we find that three classes are given to granite derived soils (including one class for peat) because intra-granite variation exceeds that of any other lithology. There are many possible approaches to classification, the key point is that geochemical models derived from multiple high resolution datasets present us with a new level of geological detail.



Cross your eyes so that the two black circles become one.

Closing thoughts:

High resolution geochemical data highlight the fact that geology is highly heterogeneous. The act of producing classified maps, necessary as it may be for concisely summarising a region, seems like a worrying oversimplification when the scale of the intra-unit variation can clearly be seen. The important properties of the ground beneath our feet vary according to our purposes – the engineer may be interested in mechanical properties, the explorer in commodity element concentrations, and the epidemiologist in potentially harmful element concentrations. All of these vary on continuous scales.

As we enter an era of high resolution information, what purpose do our classifications serve?

[a] British Geological Survey, Environmental Science Centre, Keyworth, Nottingham, NG12 5GG, UK * Corresponding author. Email: cwk@bgs.ac.uk

Charlie Kirkwood [a,*], David Beamish [a], Bob Lister [a], Mark Cave [a], Antonio Ferreira [a], Paul Everett [a]