MICA
Minerals Intelligence Capacity Analysis

FACT SHEET

Geochemical Mapping for Mineral Exploration

This factsheet describes how geochemical mapping is used to explore for economic minerals. The basic geostatistical steps used to interpret raw geochemical data to create geochemical maps are discussed along with potential errors and uncertainty.

Scope (conceptual model & main characteristics)

Geochemical mapping provides a means of visualising spatial variations in the chemical composition of the Earth's surface. The chemical signature of any specific mineral deposit will reflect the commodities that it contains, and is likely to contrast significantly with that of surrounding rocks. Geochemical maps display and quantify these geochemical contrasts, and are therefore an important line of evidence from which to guide mineral exploration.

Geochemical maps are typically produced using data collected by chemical analysis of soil or stream sediment samples, but other media may be used, such as stream water, ground water or rock chips. Soils and stream sediments are generally favored as they strike a good balance between ease of collection (low cost) and the quality and relevance of information obtained. Stream sediment data may be more useful than soils at a first-pass reconnaissance scale as the samples represent material from their entire upstream catchment area, and therefore with careful planning are capable of providing complete representation of a study area, albeit in a topographically-aggregated format. Soil data on the other hand is simpler to work with as in most cases it can be assumed that the sampled material did not originate a great distance from the collection site, and thus contains information reflecting the bedrock at that point. Soil samples can be collected quickly, easily, and consistently using a hand auger. Stream sediment samples are subject to greater compositional inconsistencies as a result of local variations in stream flow, though this can be minimised by using sieving to target finer grain sizes at the expense of collection time. Johnson *et al.* (2005) describe the collection of both media in more detail.

Regardless of the chosen media, interpolation is central to the process of geochemical map production because the high cost of chemical analysis prevents exhaustive sampling. Interpolation is therefore required to produce a continuous surface from data collected at a relatively coarse sampling density. Interpolation is generally conducted using one of three main approaches:

1) Naïve interpolation, e.g. Inverse Distance Weighting (IDW; Shepard, 1968), to predict values between geochemical observations using a standard simplistic model for spatial autocorrelation.
2) Geostatistical interpolation, e.g. Ordinary Kriging (OK; Cressie, 1988), to predict values between geochemical observations by modelling the spatial autocorrelation of the data.
3) Regression / machine learning, e.g. Random Forest (RF; Breiman, 2001), to predict values between geochemical observations based on the values of spatially continuous auxiliary variables that have been measured across the region, such as from geophysical survey and other remotely sensed data sets.

Naïve interpolation is often favored for its simplicity, but it can be expected to be less accurate than geostatistical interpolation provided that the necessary assumptions of the geostatistical model are met, namely that the input variable is normally distributed and exhibits second-order stationarity, i.e. that the mean and autocorrelation of the data do not exhibit regional trend. The regression / machine learning approaches are becoming increasingly viable as the world becomes more data-focused: more auxiliary variables are being collected and machine learning techniques are improving. Regression approaches can in fact be combined with geostatistical approaches; for example the residuals of a regression model may be geostatistically interpolated in a procedure known as Regression-Kriging (Hengl *et al.*, 2007).

Contexts of use, application fields	-> contexts (e.g., environmental, economic, social assessment) -> which types of stakeholder questions are concerned? -> link to published studies that implement the method

Geochemical mapping is generally implemented at the earliest stages of mineral exploration as it provides a cost-effective line of evidence from which to hone in on targets for subsequent drilling. The high cost of drilling means that it generally pays to be thorough at the geochemical mapping stage in order to increase the chances of success at the exploratory drilling stage. Geochemical mapping may therefore be conducted iteratively: An initial regional scale survey is generally used to identify target areas which may then be resampled at a higher density in order to produce more accurate and precise geochemical maps of the individual targets.

At every stage, geochemical mapping provides data on the chemical composition of the Earth's surface, within which is contained information on the composition of the subsurface. For mineral exploration, geochemical maps provide evidence of the locations of subsurface ore deposits by highlighting concentrations of commodity elements. Additionally, geochemical maps provide information on the concentrations of environmentally harmful elements which may co-occur with

commodities; an important consideration when assessing the viability of ore extraction. Examples of geochemical maps include the UK Geochemical Baseline Survey of the Environment (G-Base) project (BGS, 2016) and the Geochemical Atlas of Europe (FOREGS, 2005). Several similar databases and projects exist in Europe and beyond.

Type(s) of related input data or knowledge needed and their possible source(s)	-> which types of data are needed to run the method, from which sources could they come… -> could be qualitative data or quantitative data, and also tacit knowledge, hybrid, etc.

Both naïve and geostatistical interpolation methods (e.g. Inverse Distance Weighting and Ordinary Kriging) require only geochemical observations and their coordinates as input data in order to produce continuous-surface geochemical maps (though geostatistical methods do also require data-derived model parameters to be chosen by the operator). In addition, the regression / machine learning approaches to geochemical mapping require continuous observations of auxiliary variables throughout the desired mapping extent.

Geochemical observations are obtained from soil and stream sediment samples (e.g. Johnson and Breward, 2004) using a variety of analytical methods, but most commonly either x-ray fluorescence spectrometry (XRF) or inductively coupled plasma mass spectrometry (ICP-MS) is used, with additional fire-assay for precious metals such as gold. Depending on the specifics of the equipment used, concentrations may be reported for more than 50 elements, effectively quantifying the entire chemical composition. Coordinates are generally measured using handheld GPS at the site of sample collection, though they may still be map-read in areas of forest cover.

While it is simplest to produce geochemical maps using the concentration data for individual elements this practice has come under criticism because it does not respect the compositional nature of the data (McKinley *et al.*, 2016). In compositional data the variables are not independent of one another because they are confined together within the total sum of the closed composition, whether or not all components have been measured. The concentration of a single element therefore does not necessarily reflect the amplitude of the underlying process through which it was concentrated, but may simply reflect the absence of (or dilution by) other elements. In these compositional data sets each variable is said to carry only relative information, and it is the ratios between elements, rather than their individual concentrations, that are meaningful (Pawlowsky-Glahn and Egozcue, 2006). For effective mineral exploration it is therefore recommended that suitable log-ratios and compositional components are identified and mapped, rather than single element concentrations.

Model used (if any, geological mathematical, heuristic…)	-> e.g., geological model for mapping -> e.g., mathematical model such as mass balancing, matrix inversion, can be stepwise such as agent -based models, dynamic including time or quasidynamic specifying time series… -> can also be a scenario

The most common naïve interpolation method, Inverse Distance Weighting (Fig. 1, top), predicts new values as an inverse distance weighted average of surrounding observations, i.e. a predicted value will be more similar to nearby observations than to distant observations, and will not extrapolate beyond the range of observed values. This simple method adheres to Tobler's first law of geography: that "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970).

The core geostatistical method, Ordinary Kriging (Fig. 1, middle), builds on the logic of Tobler; nearby observations are given greater weights than those far away, but the function which assigns these weights is statistically fitted according to the spatial autocorrelation of the data. This fitting increases the accuracy of the interpolation over IDW provided the model fit is good. In Kriging the weights are adjusted to account for spatial dependence of the observations; observations within clusters are down weighted to provide overall uniformity of observation weight across the study area.

There are many possible models that can be used for regression / machine learning approaches, but in general the predictions will purely be made according to the values of auxiliary variables present at the prediction location, rather than according to nearby observations of the variable to be predicted. The success of these methods therefore depends on the quality and relevance of available auxiliary datasets, but can produce very good results with sufficient data. For example, Kirkwood *et al.* (2016) demonstrated the effectiveness of the Random Forest algorithm for geochemical map production supported by high resolution geophysics and remotely sensed auxiliary data (Fig. 1, bottom).

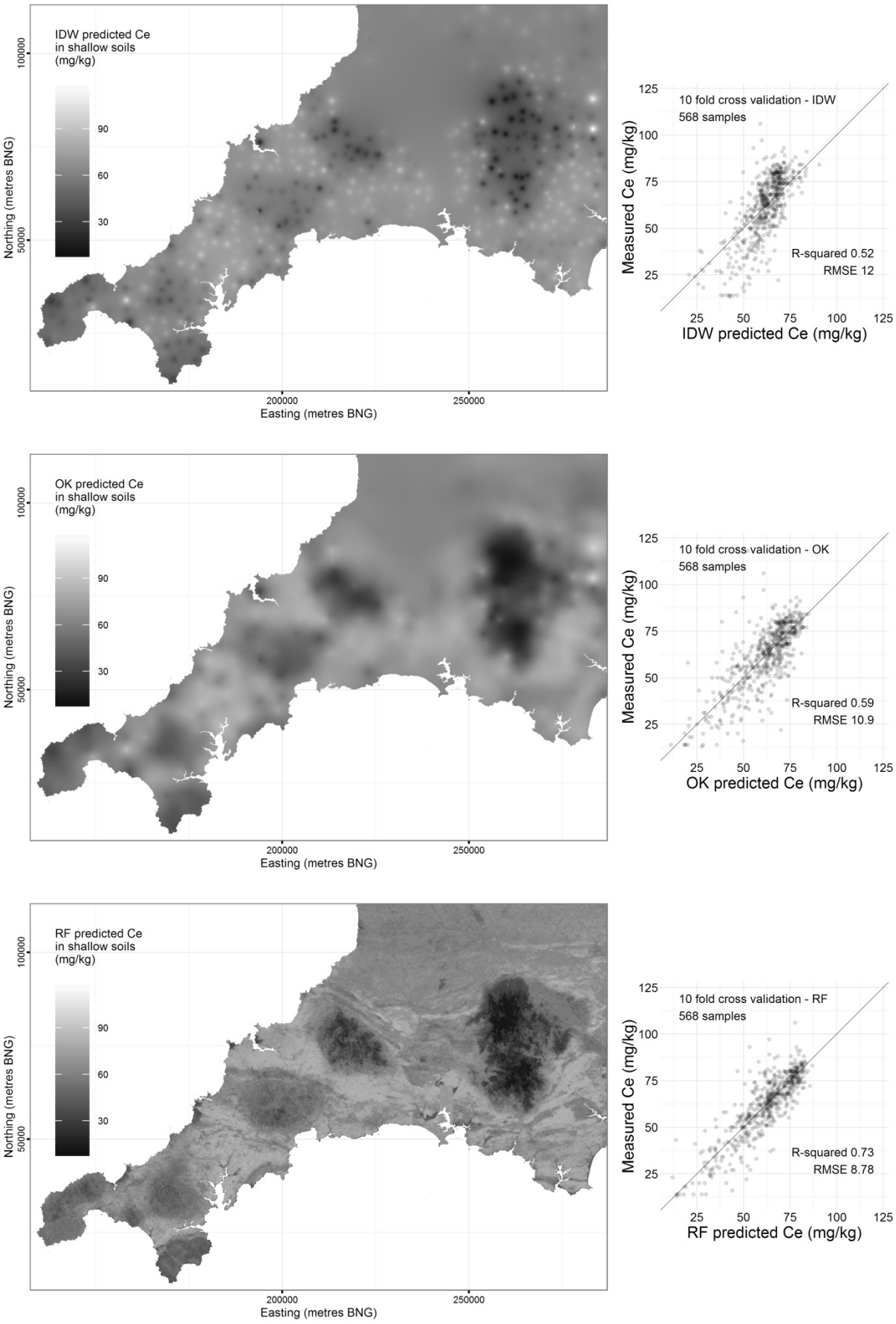

Figure 1. Comparison of cerium maps for south west England produced by IDW, OK and RF, with cross-validation plots.

Input parameters	-> which parameters are needed to run the method

Input parameters vary according to the method used. Inverse Distance Weighting has only one adjustable parameter: power. The weightings are derived from the inverse of distance raised to this power. Increasing the power decreases the influence of distant observations relative to nearby observations. IDW tends to be run with a default power value of 2. For all interpolation types it is generally possible to set a maximum distance and a maximum number of samples to be used at each prediction, which may be desirable to reduce computation time.

Ordinary Kriging requires the user to select an appropriate model type and parameters to represent the relationship between the distance between observations and the difference between their values. This relationship is visualised using the variogram (Fig. 2). In principal there are three parameters to decide; nugget, sill, and range. Nugget is the semivariance value at which the model intercepts the y axis. The nugget represents variation in the data that is not spatially autocorrelated on the scale of the survey, and may be due to measurement error or fine scale processes. The sill is the semivariance value at which the model levels off, and the range is the distance at which the sill is reached, representing the distance beyond which observations are no longer related.

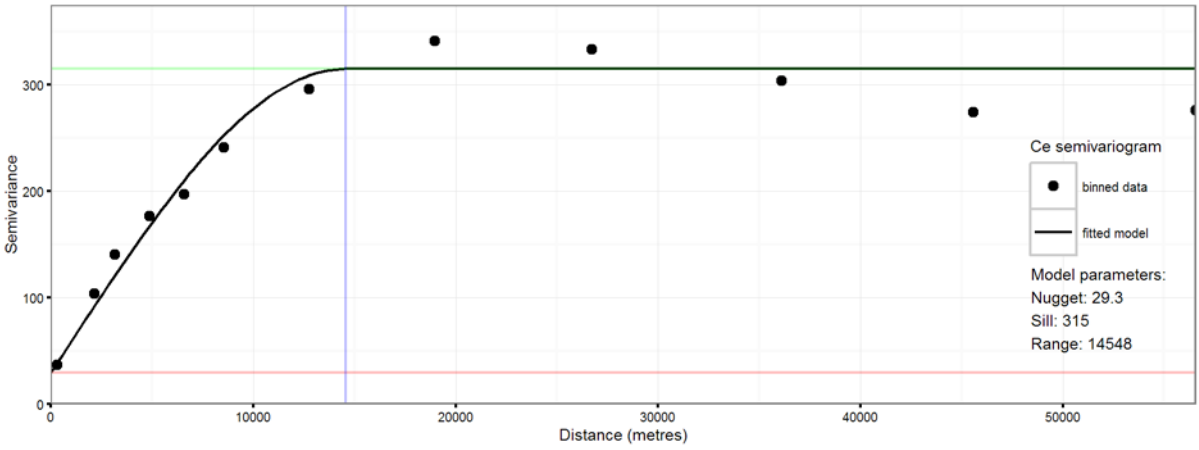

Figure 2. Example variogram using cerium data from south west England. The horizontal red and green lines mark the nugget and sill, while the vertical blue line marks the range.

Classical regression requires parameterisation in terms of an intercept and coefficients for each predictor variable; however in all modern software packages this process is automated and so little user input is required. However, selection of predictor variables and specification of any supposed interactions still requires user discretion. Machine learning approaches are highly automated, but may have tuning parameters that allow generalisation (resilience to over-fit) to be optimised for the data in hand. Again, provision of suitable predictor variables is down to the user.

Time / Space / Resolution /Accuracy / Plausibility…	-> to which spatio-temporal domain it applies, with which resolution and/or accuracy (e.g., near future, EU 28, 1 year, country/regional/local level…) -> for foresight methods can also be plausibility, legitimacy and credibility…

Geochemical maps for mineral exploration are produced as static models. Repeating observations through time may reveal some seasonal changes in surface processes but the mineral deposits of interest are likely to be static on human timescales and so geochemical mapping for mineral exploration does not typically deal with the dimension of time.

Spatially, geochemical maps may be produced at a range of extents and scales. The extent is dictated by the extent of the area of interest, but sampling should extend beyond the boundaries of this extent to ensure that predictions are always interpolations rather than extrapolations. The maps are usually presented in a raster format; i.e. a grid is constructed and values are predicted for each grid cell. There are no hard specifications for the size of the grid cells, but they should be sufficiently fine to retain all useful information within the map without being so fine as to cause computational difficulties. For example national scale surveys may use 1km grid cells, while regional scale surveys may use 100m grid cells and targeted surveys may use 1m grid cells.

Indicators / Outputs / Units	-> this refers to what the method is actually meant for. Units are an important part but that is most of the time not sufficient to express the meaning. For example, **the indicators used in LCA express the cradle-to-grave environmental impacts of a product or service**. This can be expressed in kg CO_2-equivalent. But also in €. Or in millipoints. Or in m^2year land use. -> for foresight methods the outputs are products or processes

The output is a geochemical map, a raster in which each grid cell contains predicted values of the geochemical variable in question. For individual element maps, the units will either be percentages (%), parts per million (ppm, or mg/kg) or parts per billion (ppb), depending on whether they show a major element, minor element, or trace element respectively. Log-ratio maps are without units, but provide a more informative representation of geochemical composition than individual element concentrations (see ‘types of input data’).

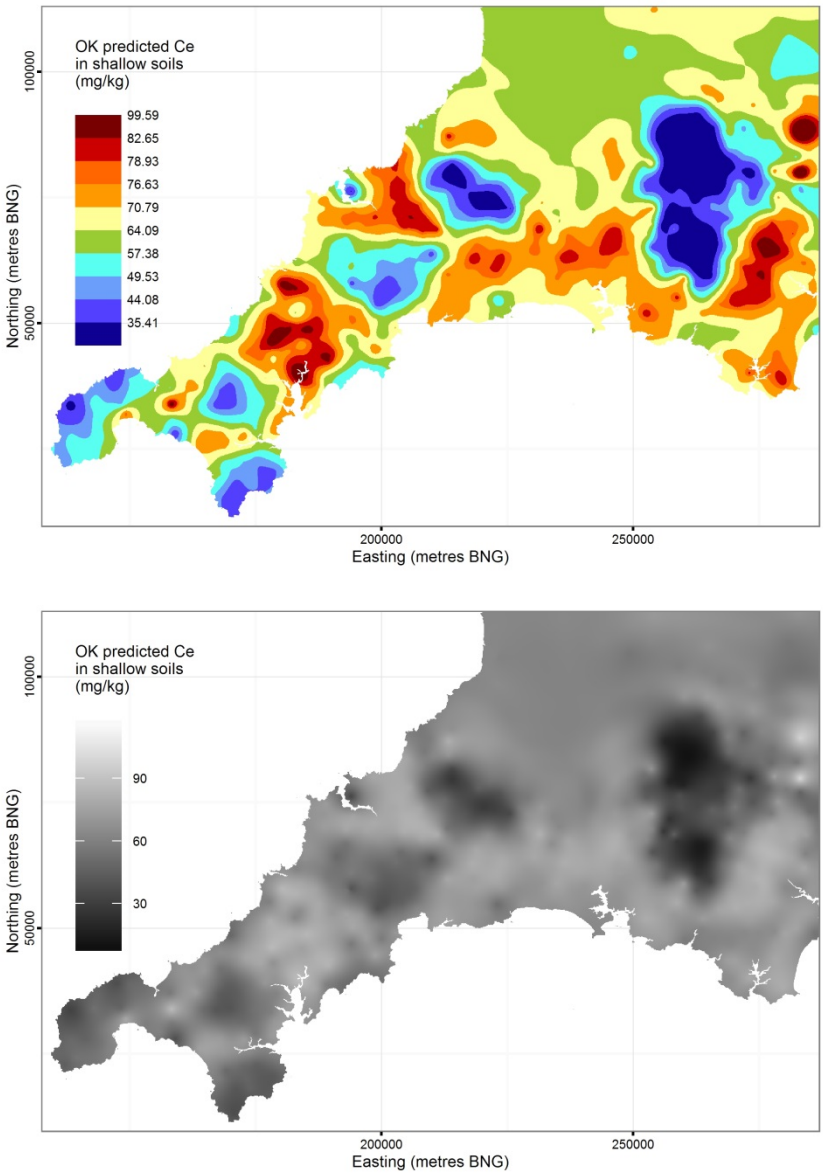

Figure 3. The same map of cerium in south west England, symbolised using both quantile-classified rainbow and continuous greyscale colour schemes. The continuous map is instinctively more intelligible. Arbitrary classification and rainbow colours only serve to impede the clear conveyance of information, even to the fully colour-sighted.

The grid cell values of a geochemical map are symbolised with a colour scheme of the producer's choosing. Geochemical maps are often displayed using a classified renderer, wherein different colours are used to represent a range of quantile classes in the data. Such visualisations sacrifice a lot of detail and introduce misleading hard boundaries in what is fundamentally continuous data, and so should be avoided unless there is a genuine reason for classification. Even in continuous colour scales, 'rainbow' colour schemes should be avoided as they obscure the information in the data (Borland and Taylor II, 2007, Moreland, 2016).

Single hue continuous colour scales, or at least perceptually uniform colour scales, are recommended for geochemical maps as they provide the most natural representation of the detail in the data, and

offer the best chance to understand the features in the data (Fig. 3). If the data is highly skewed, histogram equalisation can be used to improve detail across the map.

Treatment of uncertainty, verification, validation	-> evaluation of the uncertainty related to this method, how it can be calculated/estimated

Treatment of uncertainty depends on the modelling method used. Naïve interpolators such as IDW have no statistical basis and are unable to provide prediction intervals. Geostatistical methods such as Ordinary Kriging provide variance as an output of the interpolation; allowing the production of an accompanying uncertainty map, which displays how uncertainty increases with increasing distance from observations. Regression / machine learning approaches will also offer prediction intervals according to the particular methodology used. Prediction intervals are a useful tool for iterative mapping, as they identify locations with the greatest uncertainty, which should therefore be targeted in later rounds of sampling.

All geochemical maps should be validated to provide users with information on their accuracy. K-fold cross-validation is the most commonly accepted method for doing this (Kohavi, 1995). The value of k can be chosen by the user, but it is generally accepted that 10 provides a good balance between the high bias of using too few folds and the high variance of using too many. In 10-fold cross-validation the data is split into 10 separate folds of approximately equal distribution using stratified sampling. The chosen model is then trained using the data in 9 of these folds, and used to predict values for the locations of the observations in the remaining 'test' fold. By repeating this process 10 times, so that each fold is used as test data, the accuracy of the model can be assessed by comparing the predicted and observed values. Accuracy will often be reported using cross-validated root mean square error (RMSE) in map units, or coefficient of determination (R^2) for unitless comparison between the accuracy of maps for different variables.

Main publications / references	-> e.g. , ILCD handbook on LCA, standards (e.g. , ISO) -> can include reference to websites/pages -> references to be entered with their DOI

BGS. (2016). The Geochemical Baseline Survey of the Environment (G-Base) for the UK. Available at: URL<http://www.bgs.ac.uk/products/geochemistry/GbaseUK.html>.
Borland, D. & Taylor II, R. M. (2007). Rainbow color map (still) considered harmful. IEEE computer graphics and applications 27, 14-17. DOI: 10.1109/MCG.2007.323435
Breiman, L. (2001). Random forests. Machine learning 45, 5-32. DOI: https://doi.org/10.1023/A:1010933404324
Cressie, N. (1988). Spatial prediction and ordinary kriging. Mathematical Geology 20, 405-421.
FOREGS. (2016). Geochemical Atlas of Europe. Available at: <http://weppi.gtk.fi/publ/foregsatlas/index.php>. DOI: https://doi.org/10.1007/BF00892986

Hengl, T., Heuvelink, G. B. & Rossiter, D. G. (2007). About regression-kriging: from equations to case studies. Computers & Geosciences 33, 1301-1315. https://doi.org/10.1016/j.cageo.2007.05.001
Johnson, C. & Breward, N. (2004). G-BASE: Geochemical baseline survey of the environment. Nottingham, UK, British Geological Survey, 16pp. (CR/04/016N) (Unpublished).
Johnson, C., Breward, N., Ander, E. & Ault, L. (2005). G-BASE: baseline geochemical mapping of Great Britain and Northern Ireland. Geochemistry: Exploration, Environment, Analysis 5, 347-357. DOI: https://doi.org/10.1144/1467-7873/05-070
Kirkwood, C., Cave, M., Beamish, D., Grebby, S. & Ferreira, A. (2016). A machine learning approach to geochemical mapping. Journal of Geochemical Exploration 167, 49-61. DOI: https://doi.org/10.1016/j.gexplo.2016.05.003
Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In Ijcai, pp. 1137-1145.
McKinley, J. M., Hron, K., Grunsky, E. C., Reimann, C., de Caritat, P., Filzmoser, P., van den Boogaart, K. G. & Tolosana-Delgado, R. (2016). The single component geochemical map: Fact or fiction? Journal of Geochemical Exploration 162, 16-28. DOI: https://doi.org/10.1016/j.gexplo.2015.12.005
Moreland, K. (2016). Why We Use Bad Color Maps and What You Can Do About It. Proceedings of Human Vision and Electronic Imaging (HVEI) (To appear). DOI: https://doi.org/10.2352/ISSN.2470-1173.2016.16.HVEI-133
Pawlowsky-Glahn, V. & Egozcue, J. (2006). Compositional data and their analysis: an introduction. Geological Society, London, Special Publications 264, 1-10. DOI: https://doi.org/10.1144/GSL.SP.2006.264.01.01
Shepard, D. (1968). A two-dimensional interpolation function for irregularly-spaced data. In Proceedings of the 1968 23rd ACM national conference, pp. 517-524. ACM.
Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. Economic geography 46, 234-240.

Related methods	-> List of comparable methods, their particularities… -> link to one or several other existing fact sheet(s)

Geophysical survey
Remote sensing
Geological mapping
Prospectivity analysis
Resource estimation

Some examples of operational tools (CAUTION, this list is not exhaustive)	-> e.g., software… Only give a listing and a reference (publication, website/page…) -> **should be provided only if ALL main actors are properly cited**

QGIS - QGIS Development Team, 2016. QGIS Geographic Information System. Open Source Geospatial Foundation Project. http://www.qgis.org/

SAGA - Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., and Böhner, J. (2015): System for Automated Geoscientific Analyses (SAGA) v. 2.1.4, Geosci. Model Dev., 8, 1991-2007, doi:10.5194/gmd-8-1991-2015.

GRASS - GRASS Development Team, 2015. Geographic Resources Analysis Support System (GRASS) Software, Version 7.0. Open Source Geospatial Foundation. http://grass.osgeo.org

R - R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Key relevant contacts	-> list of relevant **types** of organisations that could provide further expertise and help with the methods described above.

The geological survey of the country concerned should be contacted in the first instance; they may well have conducted their own national-scale geochemical mapping programs, which are an ideal starting point from which to plan more detailed mapping.