Journal of Geochemical Exploration 168 (2016) 36–48

Contents lists available at ScienceDirect

Journal of Geochemical Exploration

journal homepage: www.elsevier.com/locate/gexplo

Assessing dataset equivalence and leveling data in geochemical mapping

Benoît Pereira [a,*], Aubry Vandeuren [a], Bernadette B. Govaerts [b], Philippe Sonnet [a]

[a] Earth and Life Institute, Université catholique de Louvain, Croix du Sud 2/L7.05.10, 1348, Louvain-la-Neuve, Belgium
[b] Institut de Statistique, Biostatistique et sciences Actuarielles, Université catholique de Louvain, Voie du Roman Pays 20, 1348, Louvain-la-Neuve, Belgium

ARTICLE INFO

Article history:
Received 26 October 2015
Revised 20 March 2016
Accepted 24 May 2016
Available online 02 June 2016

Keywords:
Geochemical mapping
Data leveling
Quality control
BLS regression

ABSTRACT

Combining data originating from two or more geochemical surveys can be highly beneficial for establishing geochemical maps with increased resolution and/or coverage area. However this practice requires assessing the equivalence between datasets and, if needed, applying data leveling to correct possible biases between datasets. Here we propose two original methods for assessing equivalence and for leveling data when datasets contain records that are located within the same perimeter. The first method is designed for datasets that are similarly spatially distributed and is based on the Kolmogorov-Smirnov test and quantile regression. The second method does not require datasets to be similarly spatially distributed and is based on prior knowledge about the factors explaining the geochemical concentrations and on BLS (Bivariate Least Squares) regression. The scope of application, pros, cons and detailed practical recommendations are presented for each method. Both methods were applied to a case study involving Fe, V and Y datasets originating from two European geochemical mapping projects: the Geochemical Mapping of Agricultural Soils of Europe (GEMAS) and the Baltic Soil Survey (BSS). Both methods for assessing the equivalence and obtaining leveling equations yielded comparable results thereby illustrating their effectiveness and their feasibility.

1. Introduction

A steadily increasing number of geochemical mapping projects originating from governmental or industrial organizations have appeared within the last few decades. This growing interest comes from the many social benefits of geochemical mapping applications that include geochemical prospecting, environmental monitoring, land use planning and medical geology. Therefore, a large number of geochemical datasets are currently available, which range from geochemical atlases datasets (examples listed in Garrett et al., 2008) to datasets originating from regulations requiring sample analysis (e.g. in Europe: Van Meirvenne et al., 2008; Vandeuren et al., 2013; Baize et al., 2006). As a consequence, in many parts of the world, multiple geochemical datasets are available for cartographers when mapping a region. Using multiple geochemical datasets allows the cartographer to extend the area that can be mapped and/or to improve the resolution of the map by increasing the density of available data. However, even small differences in sample materials, sampling methods, sample preparations or analytical procedures can have a major impact on the measured chemical concentrations (Reimann and de Caritat, 2012). Therefore, these differences can make it problematic to use multiple datasets for geochemical mapping.

Biases between geochemical survey datasets and procedures for dataset leveling are issues that have been discussed for a long time. Darnley et al. (1995) provide, to our knowledge, the first study that exhaustively discusses and formalizes these issues. Considering that sample preparation and chemical analysis phases can be the most expensive part of a geochemical mapping project (Johnson, 2011), the proper use of already-existing datasets should be considered as an urgent subject for study. This subject was previously addressed by the authors as they provided a generic method for creating new datasets for geochemical mapping based on multiple pre-existing geochemical datasets (Pereira et al., 2015). The proposed generic method first presents a series of quality checks that evaluate the relevance of existing data for the intended geochemical map. It then outlines a procedure to determine if the datasets should be considered as equivalent, and if not, a method to level the datasets. The purpose of the present paper is to provide better insight into the assessment of dataset equivalence and data leveling. In the literature, the assessment of geochemical dataset equivalence is usually based on expert judgment. Here, we propose to use methods based on statistical procedures for assessing the equivalence.

In some scientific fields, such as pharmacology, "equivalence" generally refers to open intervals used to define the meaning of equivalence (Ennis and Ennis, 2010). In the statistical tests proposed for assessing this kind of equivalence, the null hypothesis is the non-equivalence and the alternative hypothesis is that the items tested are equivalent within the defined intervals. The equivalence is proved if this null

[*] Corresponding author.
E-mail address: benoit.pereira@uclouvain.be (B. Pereira).

http://dx.doi.org/10.1016/j.gexplo.2016.05.012
0375-6742/© 2016 Elsevier B.V. All rights reserved.

hypothesis is rejected. By contrast, in other scientific fields like measurement/methods comparisons by linear statistical relationships, "equivalence" refers to similar measures notwithstanding the random measurement errors (Francq and Govaerts, 2016). The null hypothesis is the equivalence and the alternative hypothesis is that there is a bias between the measurement methods. If the null hypothesis is rejected, the two methods are considered as not equivalent. If it is not rejected, equivalence is not proven, but assessed using a confidence interval approach. The size of the interval is directly linked to the random measurement errors and the sample size. In our work we choose to use this second definition of equivalence. When several datasets are chosen to be combined for geochemical mapping, we consider them as equivalent unless the contrary can be proven. Predefining intervals to define the equivalence as in the other equivalence definition is not relevant here because the magnitude of the tolerable interval will depend on several factors such as the analytical and sampling procedure used as well as on the type of sample material which varies from case to case.

Two methods for assessing the equivalence between datasets and for leveling data in geochemical mapping are proposed in this paper. The first method, referred to as the SSD method hereafter, derives from the formulation of general principles found in the literature. This method is based on the comparison of all the records located within a geographical area in which geochemical datasets are similarly spatially distributed. The second method, referred to as the SCU method hereafter, derives from the same general principles, but having undergone some improvements and adaptations that make its application more suited to real-life case studies. This method is based on the comparison of datasets on several geographical entities called Spatial Comparison Units (SCUs) by linear regression analysis. Both methods are then applied to a case study focusing on three European geochemical datasets. Finally, advantages and limitations as well as fields of application of the two methods are discussed.

2. Comparing and leveling of geochemical datasets

2.1. What do we mean by "geochemical dataset equivalence"?

Answering this question requires us to define, as a first step, what is actually being compared. In this article, a geochemical dataset will designate a data table where each record comprises at least the XY location and the concentration of the sampled material for one targeted element or chemical compound. In a particular dataset, each record concerns the same type of sampled material, and the samples were collected according to the same sampling protocol. Also, the measurements of the chemical content in the samples must have been performed by the same laboratory with the same analytical measurement method. Finally, measurements must have been carried out over a relatively short and well defined time period. Based on a combination of the terminology and notations commonly used in geostatistics (e.g. Goovaerts, 1998) and chemistry (e.g. IUPAC, 2014), we can define a dataset as a set of n measurements $z(u_i)$ for the target element or chemical compound, where u_i is the vector of spatial coordinates $[x_i\ y_i]$ of the measurement $i\ \ (i=1,\ldots,n)$. The relation between the n measurement results z with the true value τ, the error e, the bias Δ, the random error δ, and the limiting mean μ can be formulated as follows:

$$z(u_i) = \tau(u_i) + e(u_i) = \tau(u_i) + \Delta(u_i) + \delta(u_i) = \mu(u_i) + \delta(u_i) \qquad (1)$$

The true value $\tau(u_i)$ is always unknown, because there is no certified reference value for a particular u_i location. The error $e(u_i)$, is the difference between the measured value $z(u_i)$ and the true value $\tau(u_i)$. The total error may be decomposed into two components, the bias $\Delta(u_i)$, and the random error $\delta(u_i)$. The random error $\delta(u_i)$ is centered on zero. It depends on the measurement uncertainty which, in turn, is proportional to the sampling and analytical variances (see Demetriades, 2011).

The limiting mean $\mu(u_i)$ is the value that is approached as the number of measurement z at the same u_i location approaches infinity. This can be stated as

$$\mu(u_i) = E[z(u_i)] \qquad (2)$$

i. e. the limiting mean corresponds to the expectation of the measurement results z at location u_i.

The bias $\Delta(u_i)$ is the systematic error, i.e. the difference between $\tau(u_i)$ and $\mu(u_i)$. $\Delta(u_i)$ remains unknown because $\tau(u_i)$ is unknown. Basing on this general terminology, we define as "equivalent" two datasets Z_1 and Z_2 containing, respectively, n records $z_1(u_i), i=1,\ldots,n$ and m records $z_2(u_j), j=1,\ldots,m$, for which at any XY location the limiting means $\mu_1(u_i)$ and $\mu_2(u_i)$ are equal:

$$\forall u_i = u_j\ ,\ \ \mu_1(u_i) = \mu_2(u_j) \qquad (3)$$

notwithstanding the measurement errors $\delta(u_i)$ and $\delta(u_j)$. In practice, comparing datasets and assessing their equivalence would thus require that both datasets contain records located at the same XY locations. However, this situation is infrequent in real-life situations and will thus not be considered here. In this paper the comparison of datasets will only require that datasets contain records located within the same geographical area. As a basic principle, geochemical concentrations are explained by space-dependent factors. These factors can be geology, lithology, topography, proximity to ocean, climate, broad changes in vegetation, and contamination from anthropogenic sources (Garrett et al., 2008). Depending upon each specific study area, various factors will be of differing influence. Datasets with sampled materials located within the same geographical area thus have similar factors explaining the geochemical concentrations. If the sets of sampled material have been similarly influenced by the factors explaining concentrations, they should show similar geochemical population distributions provided they are equivalent. The proposed methods discussed hereafter are premised on this rationale. Several papers in the literature rely on the same rationale, without being strictly formalized. For example, Daneshfar and Cameron (1998) and Appleton et al. (2008) have leveled geochemical datasets based on the assumption that data distribution from two datasets should be similar when the data of both datasets were located in the same geological context or in the same area. Most recently, Reimann et al. (2014) discussed the effect of land use on geochemical concentrations by comparing two geochemical datasets where each dataset corresponds to a land use (agricultural soils and grazing land soils) and covers approximately the same area (Europe).

2.2. Leveling biased datasets by linear transformation

Dataset leveling is a concept that should not be confused with dataset alignment. In computer science, the term "alignment" is used to refer to the integration of heterogeneous database that have to be consistent and coherent with one another. A closer usage can be found is geophysics, where the term "leveling" relates to a step of the processing of airborne magnetic data, which aims at removing measurement errors due to the effects of temporal variations in the earth's magnetic field (Luyendyk, 1997). In geochemical mapping, "leveling" is used when two geochemical datasets Z_1 and Z_2 are not equivalent, and the values of one dataset must be adjusted to the values of the other one (Grunsky, 2010). Leveling a dataset first requires selecting a dataset (called the "reference dataset" in this paper) against which the other dataset will be leveled. Choosing this dataset is ultimately a matter of judgment although several reasoning examples for this choice are presented in Grunsky (2010) and in Pereira et al. (2015). Choosing a dataset as the reference dataset, e.g. Z_1, is tantamount to deciding that for this Z_1 dataset, at any u_i location, the bias $\Delta_1(u_i)$ is null and the expected value $\mu_1(u_i)$ is the true value $\tau(u_i)$. In the theoretical situation where Z_1 and Z_2 have records located at the same u_i locations, the true value $\tau(u_i)$ may be decomposed into the expected value of Z_2 and a bias:

$$\mu_1(u_i) = \tau(u_i) = \mu_2(u_i) + \Delta_2(u_i) \quad (4)$$

Secondly, leveling a dataset involves choosing an appropriate leveling computation method. The leveling computation method considered here is a linear transformation as it is the computation method most widely used in the literature. As leveling through linear transformation is considered here, $\mu_1(u_i)$ and $\mu_2(u_i)$ are assumed to be linked by a linear regression model:

$$\mu_2(u_i) = \alpha + \beta\mu_1(u_i) \quad (5)$$

where α the intercept and β the slope can be estimated respectively by α and β through an adapted regression method, such as Ordinary Least Square regression (OLS). The estimated parameters α and β provide the information to assess the equivalence. In case of equivalence, a straight line is expected with a slope equal to 1 and an intercept equal to 0. This means that the bias $\Delta_2(u_i)$ is assumed to be composed of α, which is a constant value corresponding to a systematic bias, and β which is a value multiplying the expected value of Z_1, corresponding to a proportional bias:

$$\Delta_2(u_i) = \mu_1(u_i) - \mu_2(u_i) = \mu_1(u_i) - \alpha - \beta\mu_1(u_i) \quad (6)$$

Z_1 and Z_2 distributions and corresponding quantiles should be identical in case of dataset equivalence. A leveling method widely used is the computation of a leveling equation by fitting a regression line on paired quantiles from dataset distributions (e.g. in Daneshfar and Cameron, 1998; Appleton et al., 2008; Grunsky, 2010; Pereira et al., 2015). When a systematic bias of a constant value is observed between two datasets, the biased dataset can be leveled by a linear equation with a slope of 1 and an intercept of the value of the systematic bias (Case A, Fig. 1). When a proportional bias between two datasets is observed, the distribution of values of the biased dataset shows a different spread around the mean value than the spread produced by the reference dataset. In this case, the biased dataset can be leveled by a linear equation with a slope different from 1 (Cases B and C, Fig. 1).

In this paper, we have decided to confine the discussion to the dataset comparison without choosing a reference dataset containing the values considered as “true”. The term “biased” will be used to designate a dataset that cannot be considered as equivalent to another dataset. The term “negatively biased” will refer to the dataset where the measured concentrations tend to be lower than the concentrations of the dataset to which it is compared (such as the datasets drawn in blue in Fig. 1), and the opposite is true for the term “positively biased”.

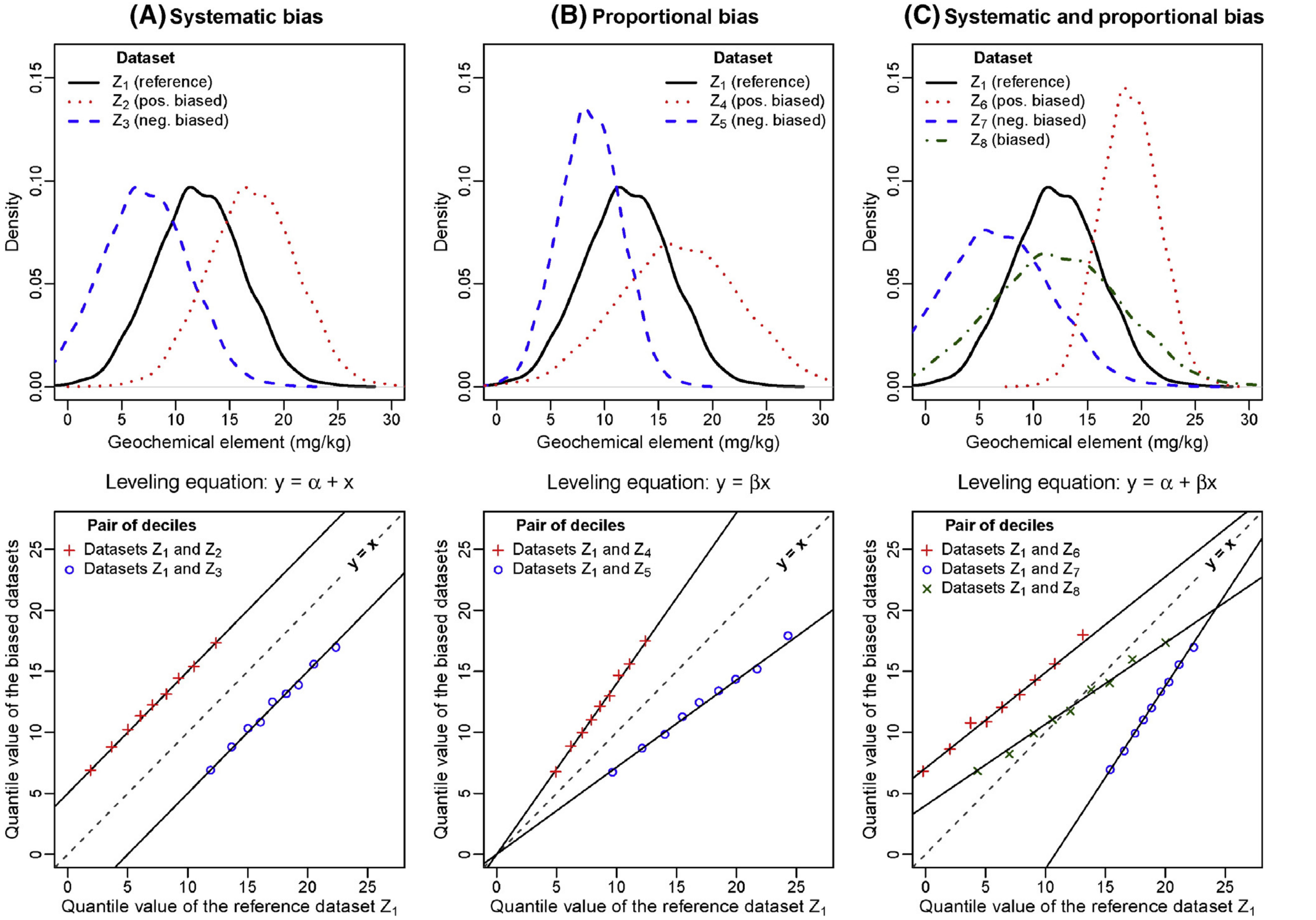

Fig. 1. Bias between datasets and leveling based on the quantile regression line. Upper row: density function of the measurements originating from a reference dataset (Z_1) and variously biased datasets (Z_2 to Z_7). The fictive reference dataset Z_1 was produced by random generation of values drawn from a normal distribution. Lower row: quantile-quantile plot for reference and biased datasets. Each of the nine points represents a pair of deciles (the first to the ninth decile) of the measurement in the reference (X-axis) and biased (Y-axis) datasets. Column A: Biased datasets Z_2 and Z_3 show a shift for a constant value which can be leveled by a linear equation of the type “$y = \alpha + x$”. Column B: Biased datasets Z_4 and Z_5 can be leveled by a multiplier, with a leveling equation of the type “$y = \beta x$”; Column C: Biased datasets Z_6, Z_7 and Z_8 involve both a shift and a multiplier and can be leveled by a linear equation of the type “$y = \alpha + \beta x$”.

3. Methods for assessing the equivalence and leveling

Let A and B be two geochemical datasets (as defined in Section 2.1) which contain records that are located within a perimeter for which a geochemical map is needed. Assume that the relevance of A and B for geochemical mapping has been confirmed either by using information about the quality control related to these datasets or by quality checks such as described in Pereira et al. (2015). Here we propose two methods for assessing the equivalence of A and B through a statistical test where the null hypothesis corresponds to conditions in which A and B should be considered as equivalent. If datasets are biased (rejection of the null hypothesis), each method contains a procedure to level the data using linear transformation.

3.1. The method based on an area wherein datasets are similarly spatially distributed (SSD method)

3.1.1. Outline

This method applies to any situation where two geochemical datasets, A and B, contain records similarly spatially distributed within the same geographical area. In practice, the method consists of three steps:

Step 1 consists in delineating a Geographical Area (abbreviated as GA hereafter) where datasets A and B both have records similarly spatially distributed and then in extracting the records from A and B that are located within the GA. The subsets from A and B that are located within the GA will from now on be referred to as subset A and subset B. Datasets A and B are similarly spatially distributed when they exhibit either a similar sampling density (e.g. 1 record/100 km^2) or a similar relative sampling density (e.g. both dataset sampling schemes are based on a grid but one dataset contains twice as many records per km^2 as the second dataset). A similar sampling density can also correspond to a similar variable sampling density (e.g. both datasets have 1 record/100 km^2 except in the northern part of GA where both have 1 record/50 km^2).

Step 2 consists in assessing the equivalence of subsets A and B using the two-sample Kolmogorov-Smirnov test (KS test). Let $a_1, \ldots, a_j$ and $b_1, \ldots, b_k$ be the records from A and B that are located within the GA. For every real number t, $F_m(t)$ and $G_n(t)$ are the empirical distribution functions for the subsets A and B:

$$F_m(t) = \frac{\text{number of sample } a' s \leq t}{m} \qquad G_n(t) = \frac{\text{number of sample } b' s \leq t}{n} \tag{7}$$

The KS test (Hollander et al., 2014) assesses whether there are any differences whatsoever between $F_m(t)$ and $G_n(t)$:

$$\begin{aligned} &H_0 : F_m(t) = G_n(t), \text{for every } t \\ &H_1 : F_m(t) \neq G_n(t), \text{for at least one } t \end{aligned} \tag{8}$$

If the null hypothesis H_0 must be rejected, subsets A and B cannot be considered as equivalent and step 3 should be applied for the leveling of the datasets.

Step 3 consists in leveling datasets using a leveling equation obtained by fitting a line on quantiles pairs from subsets A and B by orthogonal regression. Orthogonal regression on quantile pairs can be used to obtain a regression line equation which can then be applied to correct the records of the biased dataset (see Section 2.2). Orthogonal regression (also known as total least squares regression, see Francq and Govaerts, 2014b) computes a regression line so that the sum of the square of orthogonal distances (i.e. both the X-axis and Y-axis distances are taken into account) between each point (quantile pair) and the leveling line is minimized. This method is appropriate here as there are errors in quantile estimations for both subsets.

3.1.2. Comments and practical recommendations

The method can only be applied to compare geochemical datasets with similar sampling strategy. For example, the SSD method cannot be applied if subsets A and B are based on a similar sampling strategy inside a GA except for its northern part, wherein subset A has twice as many records as subset B. In this example, a bias could be observed between subsets A and B if the northern part of the GA shows a different concentration distribution than the southern part of the GA. Furthermore, the SSD method can only be applied to compare datasets with similar spatial data coverage. If subset A covers a geographical area GA_1, and subset B covers both GA_1 and another geographical area GA_2, a bias between subsets A and B may be observed due to the difference between the GA_1 and GA_2 concentration distributions. This is an important remark: in Chapter 8 of the European atlas of the GEMAS (Reimann et al., 2014), a method close to the SSD method was used to compare geochemical concentrations of datasets corresponding to agricultural soils (Ap) and grazing land soils (Gr). Differences between Gr and Ap were assessed using a Wilcoxon rank sum test (also called the Mann and Whitney test) in order to discuss the land use effect on geochemical concentrations. However, although Gr and Ap datasets have the same spatial sampling density (1 site/2500 km^2), they do not have the exact same spatial coverage. For example, Finland is covered by Gr but only partially covered by Ap. Applying our method, i.e. delineation of a GA (step 1), and comparing Gr and Ap subsets (step 2) provided different results for several elements and this remained true even after applying the Wilcoxon rank sum test (instead of the KS test used in our method). This suggests that several highlighted statistical differences between agricultural soil and grazing land soil can be explained by the difference in the spatial coverage of the Gr and Ap datasets. In other words, taking into account the Gr records located on areas not covered by Ap (and inversely) impacts the result of the Wilcoxon test and therefore subsequent conclusions.

In our method, the use of the KS test instead of the Wilcoxon rank sum test is justified as follows. The Wilcoxon rank sum test is a non-parametric test which is based on the “location shift model” and is generally used for assessing whether population A is the same as population B except that it is shifted by a constant value (i.e. for assessing differences in the population medians/means, see Hollander et al., 2014). The test is based on the calculation of a statistic, usually called U, which depends on the rank of each observation (Mann and Whitney, 1947). The use of the Wilcoxon rank sum test with the location shift model assumes that A and B are identically distributed (identical shapes for the data distributions). However, this assumption cannot be made in all real-life case studies. In cases where datasets are not identically distributed, it is possible for two datasets to have different rank sums (U statistic) and yet have equal medians. Moreover, two datasets with different spreads and equal shapes (and means) can have equal rank sums (e.g. $A \sim N(mean = 20, sd = 4)$ and $B \sim N(mean = 20, sd = 2)$; Fig. 2). Contrary to the Wilcoxon test, the KS test assesses whether there are any differences between the distributions of subsets A and B such as differences in shape (skewness, kurtosis, etc.) or in location (median, quartiles, etc.). The KS test is based on the calculation of a statistic, usually called J, which is the maximum distance between the Empirical Cumulative Distribution Functions (ECDF) of subsets A and B (Hollander et al., 2014; Fig. 3). Since it is based on a non-parametric procedure, the KS test is relatively insensitive to outliers (unusually high or low concentrations).

Datasets which contain a significant proportion of data not characterized by a true measured value (due to factors such as detection limits; data called “censored data” hereafter) must be processed with particular care. Each censored data has to be taken into account before implementing the KS test. This is usually simply done by replacing all censored data in both datasets by the same value (a value below the

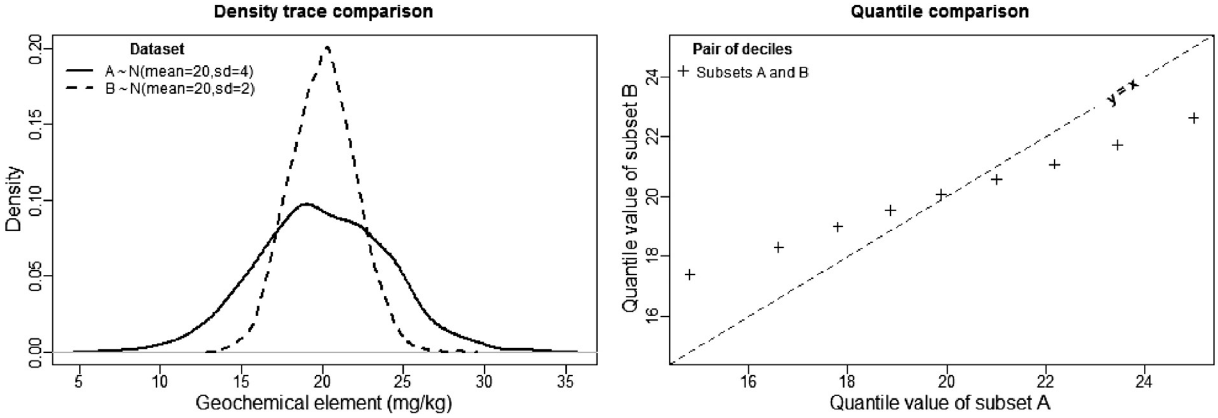

Fig. 2. Comparison of two fictitious geochemical subsets $A \sim N(mean=20, sd=4)$ and $B \sim N(mean=20, sd=2)$. Subsets A and B would have been considered as equivalent if assessed by a Wilcoxon rank sum test.

censoring limit, e.g. 50% of the limit value). However, if subsets A and B have different censoring limits, for instance "<2" and "<4" respectively, a preprocessing of the data from the subset with the lower censoring limit must be performed. This preprocessing consists first in replacing all data values that are below the higher censoring limit by a censored data ("<4" for instance) corresponding to the higher censoring limit. Then the replacement of all censored data in both datasets by the same value (as explained before) has to be done. This operation should always be carried out because even a small difference between the censoring limits of subsets A and B can lead to a large difference between the ECDF of subsets A and B, and thereby in the computation of the J statistic for the KS test. Finally, it should be noted that substantial differences between the censoring limits of datasets A and B can sometimes be considered (by expert judgment that does not require any particular data processing) as sufficient proof to decide that datasets A and B are not equivalent.

In step 3, applying leveling only makes sense when there is a good linear relationship between pairs of quantiles and if the regression line significantly differs from the $Y = X$ line. To the authors' knowledge, no established statistical test exists for regression on quantile pairs to assess if the regression line significantly differs from the $Y = X$ line. That is why we suggest to rule by expert judgment following the custom of previous authors who resorted to regression on quantile pairs for leveling (see e.g. Appleton et al., 2008; Grunsky, 2010).

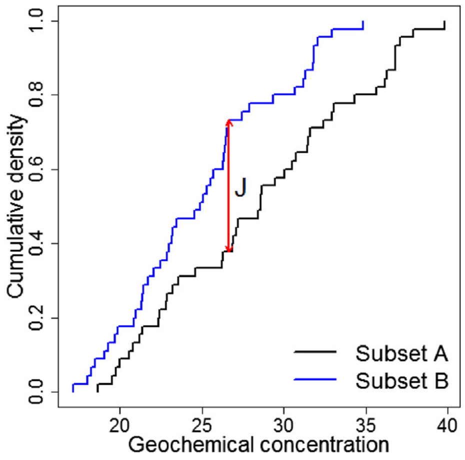

Fig. 3. Illustration of the Kolmogorov-Smirnov statistic J for two fictitious geochemical subsets (A and B). Black line is the ECDF of subset A ($F(t)$), blue line is the ECDF of subset B ($G(t)$), and the red arrow is the maximum distance between ($F(t)$) and ($G(t)$), i.e. the J statistic. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The quantiles to be used for the computation of the leveling equation is also a matter of judgment, although we propose here some indicative guidelines. If a preprocessing of the subsets had to be carried out to account for the presence of censored data, the quantile pairs corresponding to values below the highest censoring limit should not be used for the computation of the regression line. This is because the values resulting from the preprocessing should not influence the leveling parameters obtained by the regression. The best number of quantile pairs to be used for the computation of the leveling equation is difficult to determine. Obviously, there should be a sufficient number of points (quantile pairs) to perform the regression. However, if we attempt to compute too many quantile pairs, the estimation error for each quantile will be excessive because there will not be a sufficient amount of data on which to rely. Quantile estimation error depends on several other factors such as the shape of the histograms, which may vary from one dataset to another. As a rough guide, we recommend to use a number of quantile pairs that is a function of *nss*, which is the total number of records of the smaller subset (or smallest in cases of multiple datasets). While the chosen number of quantile (nq) should not exceed one quarter of the number of records of the smaller subset (or $nq \leq nss/4$), the ideal number of quantile should be equal to one thirtieth of the number of records comprised in the smaller subset (or $nq = nss/30$). If occurrence of outliers is suspected, using quantiles below the 10th percentile or above the 90th percentile should be avoided. This is because outliers have a strong impact on the extreme quantiles and thus can adversely affect the parameters of the leveling line resulting from the regression analysis. Also, the precision of quantiles estimation can be taken into consideration in the regression. Central quantiles (quantiles located around the median of the subset) are indeed estimated with more precision than the outer quantiles. This can be achieved by the use of weighted linear regression models that favor quantile pairs near the median (see e.g. Daneshfar and Cameron, 1998; Grunsky, 2010). However according to our experience (as well as that described in Appleton et al., 2008), more sophisticated models do not significantly alter the parameters α and β obtained by the linear regression.

Concerning the regression technique, orthogonal regression may not be suited to situations where the quantile estimation errors for subsets A and B cannot be considered as equivalent. For example, if A comprises ten times more records than B, one could choose to ignore the error in the quantile estimation of A by using Ordinary Least Squares regression. A must then be located on the X-axis, since this procedure does not take into account the errors in the variable located on this axis.

3.2. The method based on Spatial Comparison Units (SCU method)

3.2.1. Outline

This method applies to any situation where (1) two geochemical datasets, A and B, contain records within the same GA and (2) spatial information about factors governing the geochemical concentrations is available. In practice, the method consists of three steps:

Step 1 consists in partitioning the GA into SCUs (illustrated schematically in Fig. 4) and extracting the subsets of records. As with the SSD method, this step first requires delineating the GA for which both geochemical datasets contain records. However, as opposed to the SSD method, there is no requirement concerning the spatial distribution of the records located within the GA. The SCU delineation therefore consists in partitioning the GA into spatial entities that are deemed to be homogeneous with respect to factors known to govern the geochemical concentrations. In each single SCU, all samples should be explained by the same factors and therefore belong to the same population (same central value and dispersion). For example, if the nature of the underlying geological material is the main factor known to explain the geochemical concentrations, geological maps covering the study area offer the best support for creating the SCUs (for further information on SCU delineation, see Pereira et al., 2015). Only SCUs that contain at least 10 records from each of the datasets will be used in further steps. In what follows, the records from datasets A and B located in these SCUs will be referred as the subset A and the subset B respectively.

Step 2 consists in checking the correlation between A and B concentration values based on the graphical inspection of the scatterplot of the SCU mean concentrations. This scatterplot is a 2-axis graph where each axis corresponds to a dataset and each point represents one SCU. The points coordinates are the mean values of the concentration observed for A and B in this SCU. If this check results in the conclusion that the scatterplot exhibits strong positive linear correlation between the two datasets SCU mean concentrations, BLS regression (step 3) can be applied. A weak correlation between datasets indicates that the two datasets do not have similar spatial patterns or “spatial data structures” (see Reimann, 2005). In other words, this means that areas with high or low concentration values are not the same in both datasets. In this case, step 3 (assessing the dataset equivalence and leveling data) should not be applied because (i) the non-equivalence of the datasets is obvious and the use of a statistical procedure for assessing equivalence is not needed and (ii) the leveling of datasets through linear transformation only makes sense when a good linear relationship exists between datasets.

Step 3 consists in assessing datasets equivalence and calculating the leveling equation by BLS regression (Francq and Govaerts, 2014a). This approach is based on a regression analysis (a linear relationship) of the scatterplot of the SCU mean concentrations. Let us suppose that a number of s SCU have been delineated in step 1. Let a_{ij} and b_{ik} be the values measured in SCU$_i$ ($i = 1, 2, \ldots s$) that belong to subsets A and B

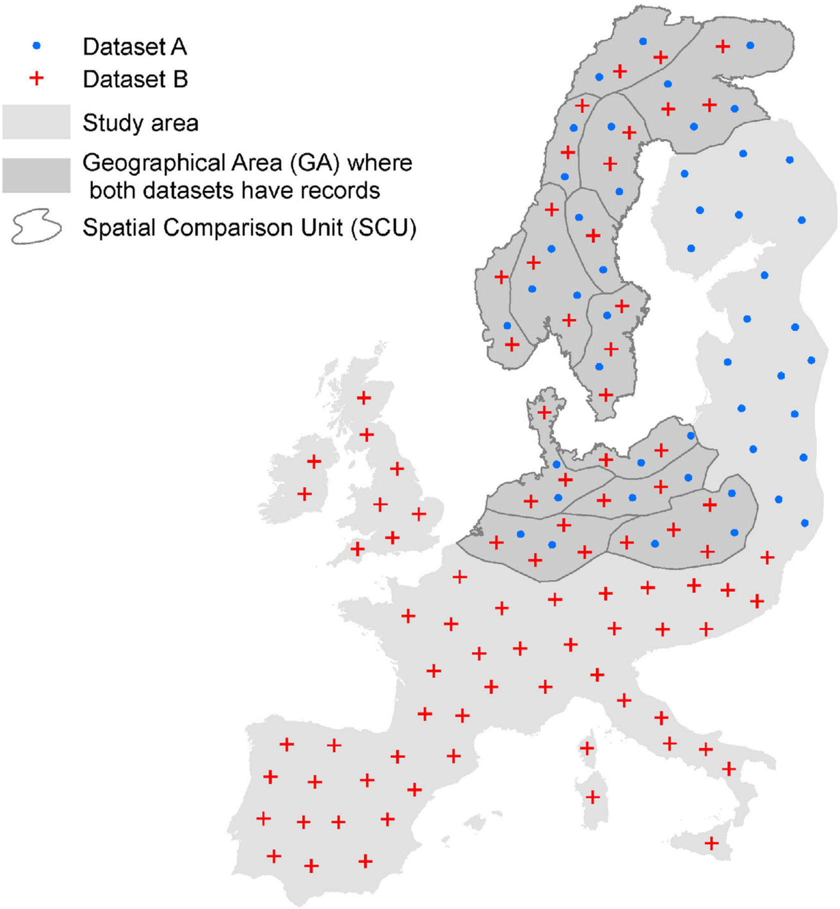
● Dataset A
+ Dataset B
Study area
Geographical Area (GA) where both datasets have records
Spatial Comparison Unit (SCU)

Fig. 4. Schematic representation of step 2 of the SCU method for a hypothetical study area (light grey) surveyed by two fictitious geochemical datasets (A and B). A geographical area where both geochemical datasets contain records (dark grey) is partitioned into 13 Spatial Comparison Units (outlined spatial polygons).

respectively. The measurements can be modeled as:

$$\begin{aligned} a_{ij} &= \zeta_i + \tau_{ij} (j = 1, 2, .., n_{a_i}) \\ b_{ik} &= \eta_i + \nu_{ik} (k = 1, 2, .., n_{b_i}) \end{aligned} \tag{9}$$

where n_{a_i} and n_{b_i} are the number of data located in SCU$_i$ that belong to subset A and B respectively. The measurement errors, τ_{ij} and ν_{ik} are supposed to be normally distributed. The true but unobservable mean concentration of the dataset A and B on the SCU$_i$, ζ_i and η_i are assumed to be related by a linear equation:

$$\eta_i = \alpha + \beta \zeta_i$$

Therefore, in case of equivalent datasets, the α and β parameters of the BLS regression line should be close to 0 and 1 respectively. The equations used in the BLS regression model as well as a comparison with other regression models are described in Francq and Govaerts (2014a). Parameters α (intercept) and β (slope) estimated by the BLS regression line are denoted as α and β respectively. The joint-Confidence Interval (hereafter referred to as joint-CI) for α and β takes the form of a confidence ellipse centered on (α, β). The joint-CI can be calculated to test the null hypothesis H_0: $\alpha = 0$ and $\beta = 1$ to check jointly whether there is constant and/or proportional bias between the two subsets. This hypothesis is rejected if the point $(\alpha = 0, \beta = 1)$ lies outside the joint-CI. Rejection of the null hypothesis indicates that the datasets cannot be considered as equivalent and should be leveled. The BLS regression line equation can be used for this purpose.

3.2.2. Comments and practical recommendations

BLS regression takes into account the heteroscedastic errors on both axes for every point (i.e. SCU mean concentration). Practically, this means that for a given dataset, the distribution of data inside two different SCUs can show unequal variances. BLS regression is especially useful for real-life case studies because SCUs can comprise more or less heterogeneous types of sampled materials leading to a more or less wide concentration variability. Moreover, inside a particular SCU, A and B distribution of data could show different variances. This situation can occur for example when the instruments used for the geochemical measurements of the two datasets are different.

The BLS regression assumes that inside each particular SCU, the data from dataset A as well as from dataset B are normally distributed. It is an important limitation which requires special care when undertaking step 3. The normality of the dataset's distributions in each SCU can be visually inspected (e.g. with a Quantile-Quantile plot) or statistically tested (e.g. with a Shapiro-Wilk test). In general, geochemical datasets are not normally distributed because distributions are usually right-skewed and contain outliers. Applying a logarithmic transformation is often sufficient to approach a normal distribution by reducing the extreme values and spreading out the low values (Reimann et al., 2008). Moreover, geochemical distributions corresponding to trace element analyses are often plagued by detection limits problems. However, one can decide to use the SCU method only to compare datasets for SCUs where both datasets contain a very small proportion of censored data (e.g. cases where, for both datasets, censored data represent less than 10% of the data in the SCU). Practically, this can be done by replacing the censored data by the value of (or half of) the censoring limit, provided the censoring limit value is the same in both datasets in order to avoid artificially introducing a bias between the two datasets. Note that this practice could lead to a poor estimation of the variance for SCUs that contain censored data and can thus slightly affect the result of the test.

The way in which the GA is partitioned into SCUs will determine the number of SCUs and the number of records available for each SCU. Note that the statistical power of the BLS regression is low when based on an insufficient number of points. Ideally, the size of each SCU must be small enough to capture the spatial patterns of geochemical concentrations but large enough to include a sufficient number of records. Ensuring that there are sufficient records from each dataset in each SCU is required to properly represent the SCU data population for the computation of the mean and the variance in the BLS regression. This is why we propose in step 1 to consider only SCUs that contain at least 10 records. However, choosing the minimum number of records per SCU will require expert judgment which depends on each particular case study, the type of sampled material and the level of uncertainty that the user of the method considers acceptable.

Finally, for step 2, note that there is a particular situation which demonstrates that the use of the BLS regression can be legitimate even when there is only a weak correlation between datasets. This situation occurs when the geochemical concentrations observed within the GAs do not exhibit any clear spatial patterns. This may be due to the fact that variations among the SCU mean values happen to be smaller than the variations among data values observed within the SCUs. In this circumstance, it is still relevant to resort to BLS regression to assess equivalence. However, in case of rejection of the null hypothesis (the datasets cannot be considered as equivalent), the BLS regression line equation should not be used to level the datasets. As a matter of fact, leveling of datasets through linear transformation only makes sense when a good linear relationship exists between datasets.

4. Case study

In order to provide an example of how the SSD method and the SCU method outlined above can be applied to real-life situations, we will now consider three geochemical datasets existing for agricultural topsoil in Europe. The datasets come from two low density geochemical mapping project results: the GEMAS project and the BSS project.

4.1. Datasets

The GEMAS (Geochemical mapping of agricultural and grazing land soils) project focused on agricultural soils from 33 European countries from an area covering approximately 5,600,000 km^2. Two agricultural soil types were surveyed separately: arable soil and grazing land soil. The sample depth is 0–20 cm for arable soil and 0–10 cm for grazing land soil. The sample density is one site per 2500 km^2 (50 × 50 km grid). More than 60 elements were analyzed by up to 4 different methods including X-ray fluorescence (XRF). Details on sample

Table 1
Summary statistics of the subsets (all values in mg/kg excepting the two first rows).

	Iron			Vanadium			Yttrium		
	Ap	Gr	BSS	Ap	Gr	BSS	Ap	Gr	BSS
n	518	514	520	518	514	520	518	514	520
Censored data	0	0	0	1	0	13	0	0	19
Minimum	1609	1469	1818	<5	7	<5	6	4	<3
Q_{25}	10,701	10,631	9704	29.2	27	21	16	14	8
Median	19,129	16,960	17,310	50	43	37	23	20	15
Q_{75}	31,387	28,990	30,004	80.7	73	70.2	29	26	21
Maximum	120,932	79,871	85,677	253	601	258	78	110	47

Table 2
Results of the pairwise comparison of the Gr, Ap and BSS datasets performed by the SSD method and the SCU method. P-values less than the significance level of 5% are in bold text.

	SSD method			SCU method		
	Gr vs Ap	Gr vs BSS	Ap vs BSS	Gr vs Ap	Gr vs BSS	Ap vs BSS
Fe	0.058	0.816	0.077	0.06	0.384	0.954
V	0.076	**0.006**	<**0.001**	0.145	<**0.001**	<**0.001**
Y	<**0.001**	<**0.001**	<**0.001**	<**0.001**	<**0.001**	<**0.001**

preparation, analytical and quality control procedures of the GEMAS project can be found in Reimann et al. (2014).

The BSS (Baltic Soil Survey) project focused on agricultural soils from northwestern Russia and Belarus and 8 European countries in an area of about 1,800,000 km^2. The sample density is one site per 2500 km^2 (50 × 50 km grid). Topsoil and subsoil layers were sampled at about 750 sampling sites in agricultural soils (indistinctly arable and grazing land). Topsoil was collected at a depth of 0–25 cm and subsoil was collected at an approximate depth of 50–75 cm. More than 60 elements were analyzed by up to 4 different methods including XRF. Details on sample preparation, analytical and quality control procedures of the BSS project can be found in Reimann et al. (2003).

Iron (Fe), vanadium (V) and yttrium (Y) topsoil XRF analyses from the GEMAS project and from the BSS project were selected as datasets to be compared. Equivalence between datasets was assessed and when a bias was detected between two datasets, a leveling equation was computed.

4.2. Application of the SSD method

4.2.1. Step 1: delineation of the GA and extraction of subsets

The GA is delineated so that it corresponds to the geographical area located at a maximum distance of 35.35 km from the dataset sampling points. Supposing that the points are located exactly in the center of the 50 × 50 km cell from the sampling grid, 35.35 km is the distance from the center of a cell to a corner ($\sqrt{(25^2+25^2)}$). The dataset records located inside the GA were extracted in order to create the representative subsets of records. Hereafter, the subset corresponding to the arable soil of the GEMAS will be called Ap, the subset corresponding to the grazing land soil of the GEMAS will be called Gr and the BSS dataset will be called BSS. Table 1 summarizes the statistics for the 9 datasets.

4.2.2. Step 2: assessing the equivalence

There are censored records in the Ap and BSS datasets (reported as “detection limit” in the corresponding originating atlases). Both datasets have the same censoring limit values (5 mg/kg for vanadium and 3 mg/kg for yttrium). We thus replaced the censored records by 50% of the limit value. We then applied the KS test to compare the three datasets pairwise. P-values of the tests are shown in the three first columns of Table 2. For iron, the three datasets cannot be considered to be significantly different for the chosen significance level of 5%. For vanadium, the two datasets from the GEMAS project were not significantly different, while these two datasets were significantly different from the BSS dataset. Finally, the three yttrium datasets were significantly different from one another.

4.2.3. Step 3: computing the leveling equations

We computed a leveling equation by using orthogonal regression to fit a line on quantile pairs from datasets that cannot be considered as equivalent in the previous step. The leveling lines for the vanadium and yttrium datasets are illustrated in Figs. 5 and 6 respectively. In these figures, the legends mention the parameters of the leveling line equations. For vanadium, these parameters are quite similar in both compared datasets: α and β are close to about -8 and 1 respectively. BSS is negatively biased compared to both GEMAS datasets since we found that BSS exhibits a systematic bias of about -8 mg/kg. For yttrium, the Ap comparison with Gr exhibits a slight proportional ($\beta = 0.94$) and systematic ($\alpha = 1.4$ mg/kg) bias. The BSS - Ap and BSS - Gr regression line parameters are quite similar: the slopes reflect a slight proportional bias (1.04 for Gr and 0.97 for Ap) and the intercept reflects a proportional bias (5.2 mg/kg for Gr and 6.6 mg/kg for Ap) between the datasets. Note that the difference between the intercept of the two leveling lines is close to 1.4 mg/kg, which corresponds to the systematic bias found between Gr and Ap.

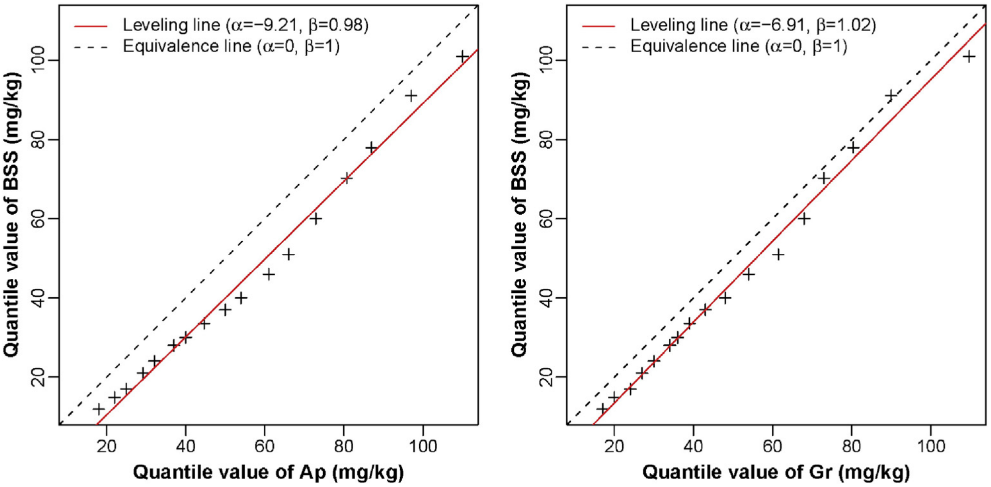

Fig. 5. Leveling line for removing the bias between datasets for vanadium obtained by orthogonal regression on pairs of quantiles (crosses: from 10th to 90th percentile with increments of 5).

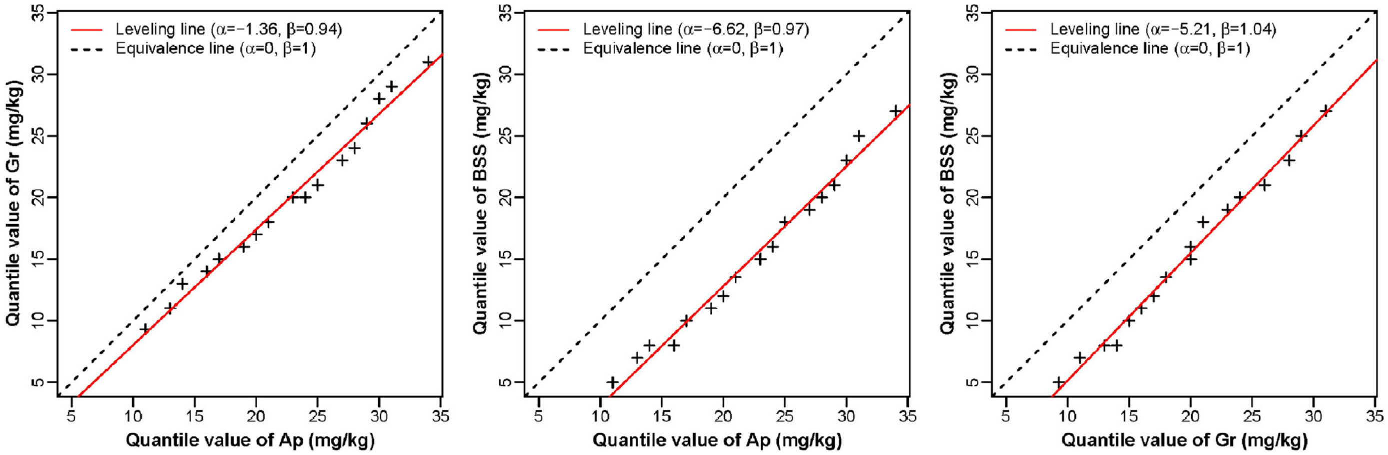

Fig. 6. Leveling line for yttrium (legend as in Fig. 5 above).

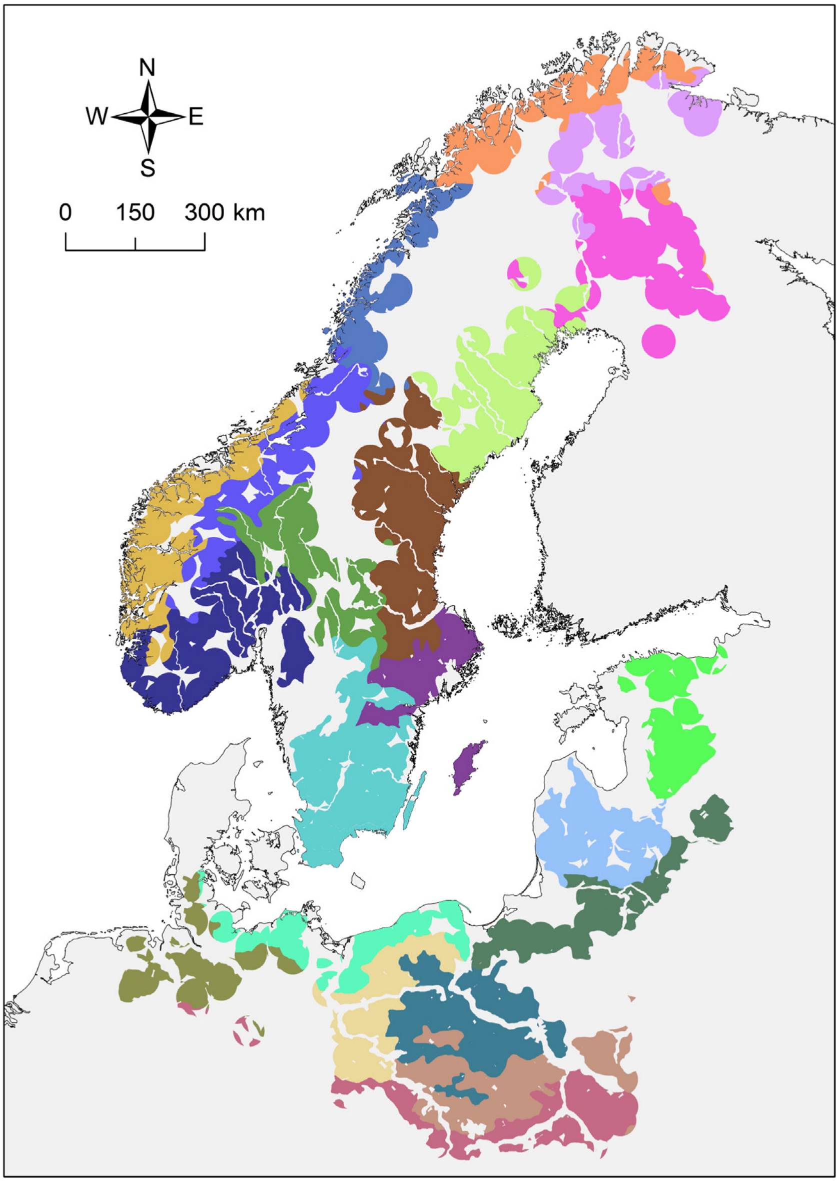

Fig. 7. Map of the SCUs (based on information from BGR, 2005). Each color corresponds to a particular SCU. The circular shapes of the SCU delineation are due to the 35.35 km search radius applied around the sampling points in the delineation of the GA. Only SCUs that contain more than 10 records from each of the three compared datasets are represented on the map.

Table 3
Summary statistics about the compared subset of records (all values in mg/kg excepting the two first rows).

	Iron			Vanadium			Yttrium		
	Ap	Gr	BSS	Ap	Gr	BSS	Ap	Gr	BSS
n	393	396	411	393	396	411	393	396	411
Censored data	0	0	0	0	0	8	0	0	15
Minimum	1609	1469	1818	5	8	<5	6	4	<3
Q_{25}	10,911	10,701	10,036	30	28	21	16	14	9
Median	18,885	16,960	17,625	49	43	37	22	19	16
Q_{75}	31,125	29,392	31,193	79	73	71	28	26	21
Maximum	120,932	79,871	85,677	253	601	218	78	110	47

4.3. SCU method application

4.3.1. Step 1: delineation of the SCUs and extraction of subsets

Here we assume that the distributions of Fe, V and Y in soil at the continental scale are mostly explained by the nature of the soil parent material, while anthropic or other influences are of local concern. This assumption is consistent with the information found in several European atlases (e.g. Reimann et al., 2003, 2014; Salminen et al., 2005). SCU delineation according to the types of soil parent material is therefore most appropriate for this case study. The map "Soil Regions of the European Union and Adjacent Countries" (BGR, 2005) containing valuable information about soil parent material helped us to delineate the GA intoSCUs (Fig. 7).

We only take into consideration SCUs that contain at least 10 records from each of the compared datasets in order properly represent the SCU data population for the computation of the mean and the variance in the BLS regression. Table 3 summarizes the descriptive statistics concerning the 9 compared subsets of records located within the SCUs taken into consideration.

This case study involves 21 SCUs. The number of records per SCU ranges from 10 to 31 with a mean of 19 for the Ap and Gr datasets and ranges from 10 to 39 with a mean of 20 for the BSS datasets.

4.3.2. Step 2: spatial correlation check between datasets

The upper rows of Figs. 8, 9 and 10 illustrate the correlation between the mean concentrations of the SCUs: all Pearson correlation coefficients are strongly positive and range from 0.89 to 0.97. This confirms that the Ap, Gr and BSS datasets feature the same spatial patterns and therefore the BLS regression (step 3) can be applied.

4.3.3. Step 3: assessing the equivalence and calculating the leveling equation line

As there are very few censored data, and only in the BSS datasets (Table 3), we have thus assumed that it could not greatly affect the BLS regression results. For the computation, we replaced the censored records by 50% of the limit value. The normality of the dataset distributions has been assessed. The records were first transformed by standardization (i.e. by subtracting the mean and dividing by the standard

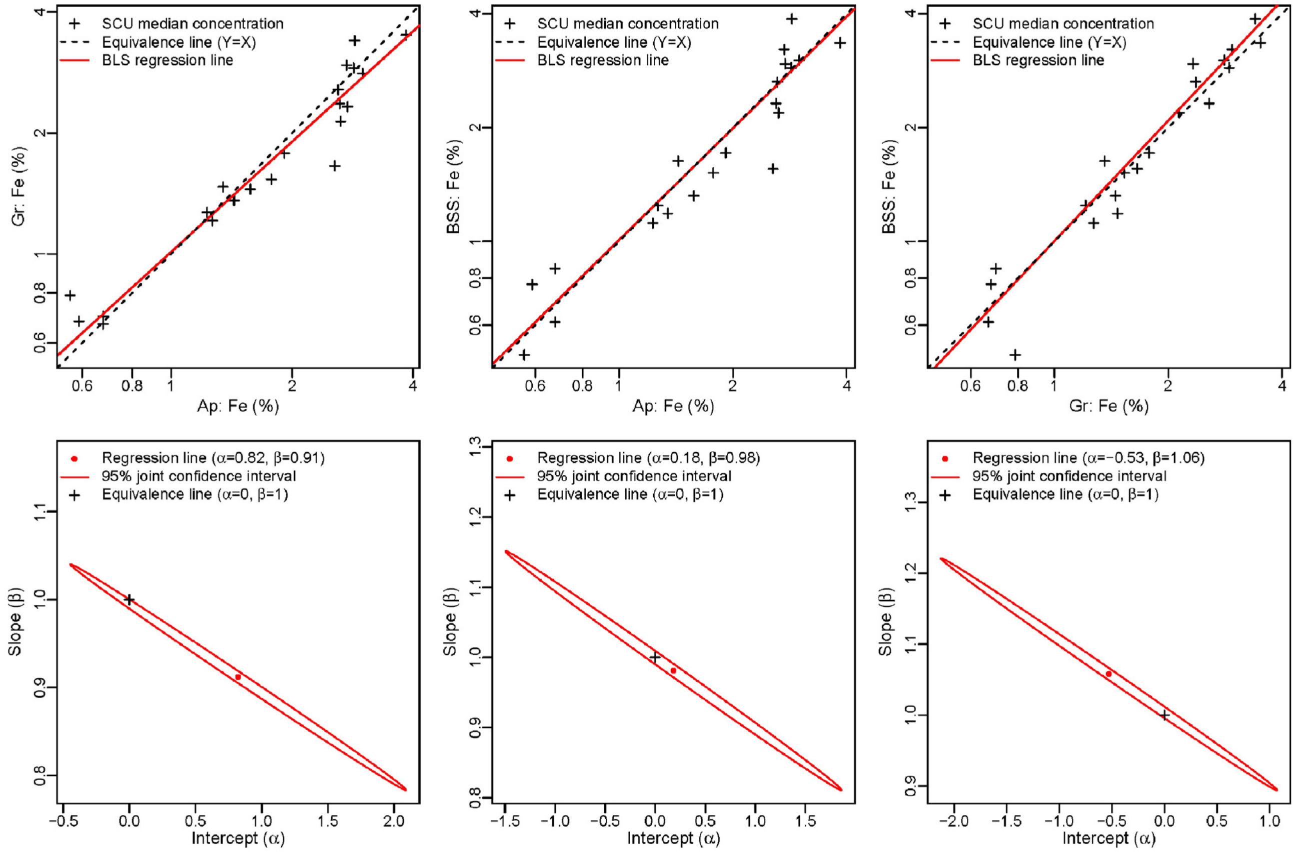

Fig. 8. Comparisons of the datasets by the SCU method (upper row) for Fe and corresponding joint confidence intervals for the α and β parameters for the BLS regression line (lower row).

deviation of the corresponding SCU), then plotted in a Q-Q plot. For each dataset, logarithmic transformation was sufficient to reach normality, as required when applying the BLS regression method. The BLS regression lines are illustrated in the graphs located in the upper rows of Figs. 8, 9 and 10, and the parameters of the lines are mentioned in the legends of the graphs. The results of the equivalence assessment are illustrated in the lower rows of Figs. 8, 9 and 10. In each graph, the black cross represents the equivalence line: if it falls inside the confidence ellipse, which is the 95% joint-CI, the datasets can be considered as equivalent. The results are identical to what was obtained with the SSD method for the same significance level (5%; see Table 2). The parameters of the BLS leveling lines (corresponding to the red points in Figs. 8, 9 and 10) cannot be interpreted in terms of systematic or proportional biases due to the logarithmic transformation of the data.

5. Discussion

The two proposed methods are based on the comparison of subsets that contain a fraction of the total number of records of the corresponding datasets. Equivalence assessment results and leveling equations provided by the proposed methods are therefore primarily relevant to these subsets. However, if the subsets sufficiently represent the factors that are supposed to be influencing the bias, like the type of sample material or the range of geochemical concentrations, it can be decided that the results produced by these methods also apply to the datasets in their entirety.

Both methods require that the compared datasets contain records located within the same geographical area. In the SSD method, geochemical datasets must have records that are evenly spatially distributed throughout the GA, as opposed to the SCU method which does not impose this requirement. In many real-life case studies, this limitation can make it difficult to apply the SSD method and the SCU method may thus provide a more suitable solution.

Both methods assess the equivalence based on a statistical procedure. In the SSD method, the statistical procedure is a two-sample Kolmogorov-Smirnov test, which is a non-parametric procedure robust to outliers and which requires very few assumptions regarding the underlying populations of datasets A and B. In the SCU-method, the statistical procedure is the BLS regression, which assumes a normality of A and B data distribution for each SCU. This normality assumption will require special care, as discussed in Section 3.2.2, and could limit the practical applicability of the SCU method.

In the SSD method, the expected benefit of the leveling through a linear transformation must be assessed by expert judgment. This is because two datasets can be deemed not equivalent according to the KS test and yet the quantile regression line may be close or equal to $y = x$. Conversely, with the SCU method, no expert judgment is needed because the diagnosis of the dataset equivalence is directly related to the distance of the BLS regression line to the $y = x$ line.

The comparison of the results provided by the two proposed methods in this case study suggests that both methods provide broadly consistent results despite the differences in the compared subsets (see Tables 1 and 3) and in the statistical procedures used. In this case study, the results of the equivalence assessment are identical for both methods although the p-values might sometimes be quite different (e.g. Ap-BSS comparison for Fe datasets). The parameters of the leveling equations obtained by both methods are not directly comparable because we applied a logarithmic transformation in the SCU method. However, the type of bias detected was identical regardless of the method. Both methods concluded that, for yttrium and vanadium, Ap and Gr are positively biased relative to BSS, while for yttrium, Ap was positively biased relative to Gr.

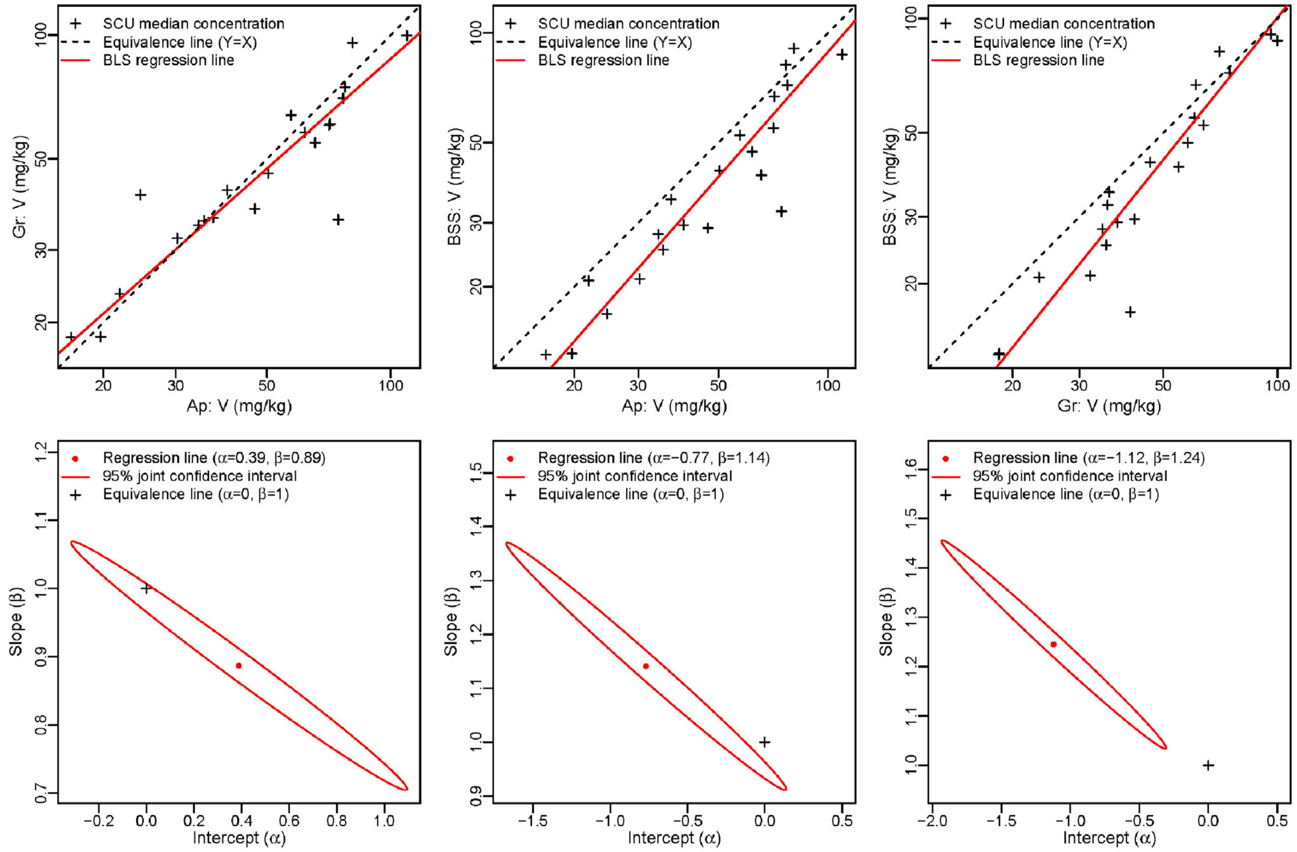

Fig. 9. Same legend as Fig. 8, but for Vanadium.

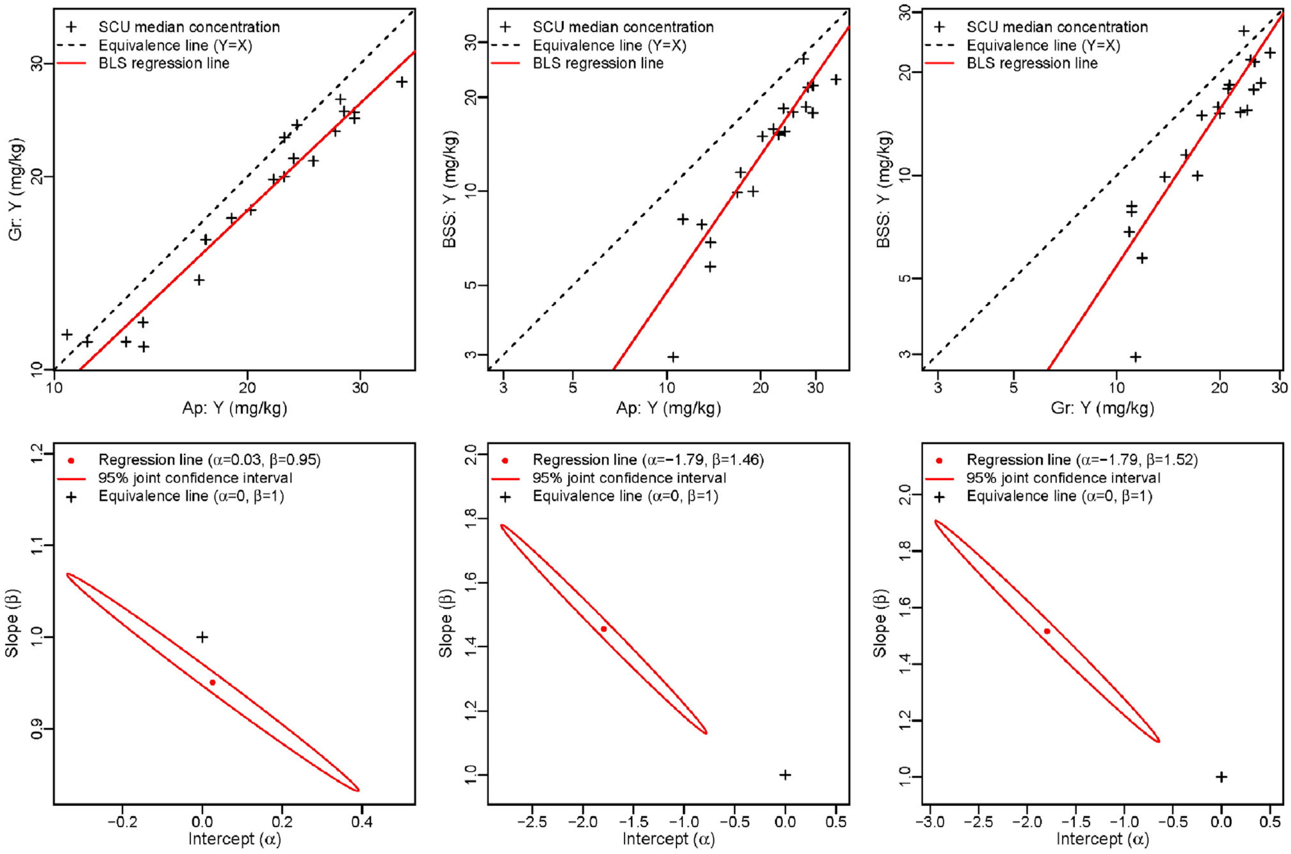

Fig. 10. Same legend as Fig. 8, but for Yttrium.

6. Conclusions

This paper establishes a methodological framework and provides practical recommendations for dealing with biases between geochemical datasets by equivalence assessment and data leveling. To our knowledge, this is the first study to properly formalize a protocol for assessing the equivalence between geochemical datasets. Our case study involving major and trace element geochemical datasets coming from two European low density geochemical mapping projects illustrates how to practically implement both proposed methods. The two methods yielded similar results thereby suggesting that they are both reliable and effective enough to take advantage of the steadily increasing number of geochemical datasets available for mapping a region.

Each of the two proposed method presents its own particular benefits and limitations and several specific questions may warrant further investigations. For example, assessing the equivalence between more than two geochemical datasets will stress the need to address the multiple testing correction problem encountered in the statistical procedures. The SCU method appears to be more suitable than the SSD method for most real-life case studies because (i) there is no condition requiring that both datasets contain records that are evenly spatially distributed and (ii) the method provides a quantitative procedure to evaluate the benefit of applying a linear transformation for leveling the datasets. However, in the SSD method, the equivalence assessment may prove more appropriate for some geochemical datasets as it requires fewer assumptions about the underlying populations.

Acknowledgements

Cynthia Rozewicz is gratefully acknowledged for proofreading the manuscript as a native English speaker. We also wish to thank Bernard Francq (Robertson Centre for Biostatistics, University of Glasgow) for the assistance provided about the BLS regression. The subsets compared in the case study are available upon request from the authors (i.e. the list of sample labels extracted from the GEMAS project and the BSS project).

References

Appleton, J.D., Rawlins, B.G., Thornton, I., 2008. National-scale estimation of potentially harmful element ambient background concentrations in topsoil using parent material classified soil: stream–sediment relationships. Appl. Geochem. 23 (9), 2596–2611.

Baize, D., Saby, N., Deslais, W., Bispo, A., Feix, I., 2006. Analyses totales et pseudo-totales d'éléments en traces dans les sols. Principaux résultats et enseignements d'une collecte nationale. Etude Gest. Sols 13 (3), 181–200.

BGR, 2005. (Bundesanstalt für Geowissenschaften und Rohstoffe), Soil Regions Map of the European Union and Adjacent Countries 1:5,000,000 (version 2.0). Tech. Rep. EU catalogue number S.P.I.05.134., Ispra.

Daneshfar, B., Cameron, E., 1998. Leveling geochemical data between map sheets. J. Geochem. Explor. 63 (3), 189–201.

Darnley, A., Bjorklund, A., Bolviken, B., Gustavsson, N., Koval, P., Steenfelt, A., Tauchid, M., Xuejing, X., 1995. A global geochemical database for environmental and resource management. Tech. rep.UNESCO Publishing, Paris

Demetriades, A., 2011. Understanding the quality of chemical data from the urban environment—part 2: measurement uncertainty in the decision-making process. In: Johnson, C.C., Demetriades, A., Locutura, J., Ottesen, R.T. (Eds.), Mapping the Chemical Environment of Urban Areas. Wiley-Blackwell, Oxford, pp. 77–98.

Ennis, D.M., Ennis, J.M., 2010. Equivalence hypothesis testing. Food Qual. Prefer. 21 (3), 253–256 (sensometrics 2008 (TBC)).

Francq, B.G., Govaerts, B.B., 2014a. Hyperbolic confidence bands of errors-in-variables regression lines applied to method comparison studies. J. Soc. Fr. Stat. 155 (1), 23–45.

Francq, B.G., Govaerts, B.B., 2014b. Measurement methods comparison with errors-in-variables regressions. from horizontal to vertical ols regression, review and new perspectives. Chemom. Intell. Lab. Syst. 134, 123–139.

Francq, B.G., Govaerts, B.B., 2016. How to regress and predict in a bland-altman plot? Review and contribution based on tolerance intervals and correlated-errors-in-variables models. Stat. Med.

Garrett, R., Reimann, C., Smith, D., Xie, X., 2008. From geochemical prospecting to international geochemical mapping: a historical overview. Geochem. Explor. Environ. Anal. 8 (3–4), 205–217.

Goovaerts, P., 1998. Geostatistical tools for characterizing the spatial variability of microbiological and physico-chemical soil properties. Biol. Fertil. Soils 27 (4), 315–334.

Grunsky, E.C., 2010. The interpretation of geochemical survey data. Geochem. Explor. Environ. Anal. 10 (1), 27–74.

Hollander, M., Wolfe, D.A., Chicken, E., 2014. Nonparametric Statistical Methods. third ed. John Wiley & Sons.

IUPAC, 2014. International Union of Pure and Applied Chemistry — Compendium of Chemical Terminology (the "Gold Book"). second ed. Blackwell Scientific Publications, Oxford (last update: 2014–02-24; version: 2.3.3).

Johnson, C.C., 2011. Understanding the quality of chemical data from the urban environment—part 1: quality control procedures. In: Johnson, C.C., Demetriades, A., Locutura, J., Ottesen, R.T. (Eds.), Mapping the Chemical Environment of Urban Areas. Wiley-Blackwell, Oxford, pp. 61–76.

Luyendyk, A.P.J., 1997. Processing of airborne magnetic data. J. Aust. Geol. Geophys. 17, 31–38.

Mann, H.B., Whitney, D.R., 1947. On a test of whether one of two random variables is stochastically larger than the other. Ann. Math. Stat. 18 (1), 50–60.

Pereira, B., Vandeuren, A., Sonnet, Ph., 2015. Geochemical mapping based on multiple geochemical datasets: a general method, and its application to Wallonia (Southern Belgium). J. Geochem. Explor. 158, 34–43.

Reimann, C., 2005. Geochemical mapping: technique or art? Geochem. Explor. Environ. Anal. 5 (4), 359–370.

Reimann, C., de Caritat, P., 2012. New soil composition data for europe and australia: demonstrating comparability, identifying continental-scale processes and learning lessons for global geochemical mapping. Sci. Total Environ. 416, 239–252.

Reimann, C., Siewers, U., Tarvainen, T., Bityukova, L., Eriksson, J., Gilucis, A., Gregorauskiene, V., Lukashev, V., Matinian, N., Pasieczna, A., et al. (Eds.), 2003. Agricultural Soils in Northern Europe: a Geochemical Atlas. Schweizerbart Science Publishers, Stuttgart, Germany.

Reimann, C., Filzmoser, P., Garrett, R.G., Dutter, R., 2008. Statistical Data Analysis Explained: Applied Environmental Statistics with R. Wiley, Chichester, UK (343 pp.).

Reimann, C., Birke, M., Demetriades, A., Filzmoser, P., O'Connor, P. (Eds.), 2014. Chemistry of Europe's Agricultural Soils, Part A: Methodology and Interpretation of the GEMAS Data Set. Bundesanstalt für Geowissenschaften und Rohstoffe (BGR) (352 pp.).

Salminen, R., Plant, J., Reeder, S. (Eds.), 2005. Geochemical Atlas of Europe. Part 1, Background Information, Methodology and Maps. Geological survey of Finland, Espoo (526 pp.).

Van Meirvenne, M., Meklit, T., Verstraete, S., De Boever, M., Tack, F., 2008. Could shelling in the first world war have increased copper concentrations in the soil around ypres? Eur. J. Soil Sci. 59 (2), 372–379.

Vandeuren, A., Pereira, B., Sonnet, Ph., 2013. Rapport final du projet de recherche CAPASOL 2: Mise en oeuvre de l'outil de prédiction et de gestion de la capacité des sols de la région wallonne à accepter l'épandage d'amendements organiques conforme à la réglementation. Tech. rep.UCL, Louvain-la-Neuve (168 pp.)