doi: 10.1255/tosf.67

Proper sampling for archeometric discrimination of Bronze-age fields on Bornholm, Denmark – Archaeology meets TOS meets Chemometrics

Bastian Germundsson[1,3], Anders Pihl[2] and Kim H. Esbensen[3,4]
[1]**Department of Geosciences and Natural Resource Management (IGN), Copenhagen University, Denmark.**
E-mail: bgermundsson@gmail.com
[2]**Bornholm Museum, Rønne, Denmark Email: ap@bornholmsmuseum.dk**
[3]**Geological Survey of Denmark and Greenland (GEUS). Copenhagen. Denmark. Email: ke@geus.dk Phone: +45 20214525**
[4]**ACABS Research Group, Aalborg University, Denmark**

In archaeology it is of interest to ascertain whether a particular Bronze-age field has been cultivated or not based on traditional archaeological evidences, but these often deal with one chemical element only, Phosphorous. We here augment this endeavour to include multi-element geochemistry characteristics. A pilot study sampling campaign was carried out (2014) on the island of Bornholm with the objective to discriminate between well-documented cultivated and un-cultivated Bronze-age agricultural fields based on multivariate data analysis (chemometrics) of soil chemistry (metal concentrations, ICP-MS). All samples originate from the same soil depth corresponding to the paleo-cultivated layer, or the equivalent depth in uncultivated fields. The experimental design focused on proper field sampling (Theory of Sampling), including replicate sampling at two levels. Applying Principal Component Analysis (PCA), the first three components corresponds to 68 % of the most discriminative variance in the 15 variable/41 sample array. The first and third PC-component reveals a complete discrimination of un-cultivated vs.3 cultivated fields; it is likely that general soil chemistry features are compensated for by the second component in the PCA solution. We present the specifics pertaining to the field sampling procedure, including the hierarchical two-level experimental design, which allow assessment of the local vs. field-wide heterogeneities in order better to understand the successful discrimination achieved. Five elements appear to be particularly involved in the discrimination [P, Fe, Mn, Zn, Pb], currently undergoing paleo-agricultural/geochemical interpretation. Based on these first results we plan a full test-set validation campaign in 2015 which will be the ultimate performance test for this type of archeometric discrimination. This contribution illustrates the versatility and power of multivariate data analysis (chemometrics) applied to data with a substantial proportion of potential sampling errors, in need of effective management (TOS).

Introduction

Bornholm is a minor Danish island in the Baltic Sea known for a diverse, interesting geology – and a magnificent archaeological venue with a great number of Celtic (Bronze age) agricultural fields, which date back to the first century BC. The Celtic field systems have been recognized and documented for more than 100 years, but little is known to how the fields were cultivated and what crops were grown. The primary knowledge is related to different indirect evidence in the form of Agricultural tools, and crops, found on secondary locations, typically settlements. The aim of this pilot project was to gain information from the primary sources, the cultivated fields themselves. By introducing geochemical fingerprinting of the subsoil from both cultivated and pristine fields, and applying Multivariate Data Analysis (MVDA), this project entertains whether it is possible to discriminate between cultivated and uncultivated fields of Celtic age on this basis. It is hoped that this may contribute to increased insight into the different agricultural methods and strategies that were used in Late Bronze Age and Early Iron Age. For this purpose Bornholm is an obvious location due to a comprehensive documentation of Celtic fields and due to the geomorphology, which allows archaeologists to distinguish between cultivated and uncultivated areas with ease and certainty, which is important classification information to be used in training a data analytical discrimination facility.

Data analysis – from univariate to multivariate

Traditional archeological data analysis in this context has overwhelmingly been an univariate approach, i.e. a directed focus on just one element, Phosphorous, which has been used extensively as a 'signal element' due to its increased concentration in manure that has been used as fertilizer. In the present study this univariate approach shows severe limitations however, for which reason a multivariate approach may act better in discriminating between fields based on a full series of 15 geochemical elements. General knowledge as to which elements might correlate with Phosphorous in cultivated fields is sparse however; Nielsen et al. (2014) showed in a similar multivariate study that Sr conceivably correlate with cultivated fields due to addition of bone fragments and household waste. Information is also scanty regarding how the geochemical fingerprints of uncultivated fields might appear in this context. We have therefore adopted a multivariate data analysis approach (Chemometrics) without any prerequisites or assumptions, letting the data speak for themselves. The archeological field use discrimination is an important piece of the puzzle.

Theory of Sampling (TOS)

TOS is also a critical agent in this endeavor: The validity of analytical results is exactly as good, or bad, as the validity of the primary sampling, as well as of all sub-samples produced in the laboratory on the pathway to the analytical aliquots. The primary – and secondary sampling in this project was in complete TOS-control, DS 3077 (2013), with a strong emphasis on unbiased field sampling and subsequent mass-reduction (riffle-splitting). The tertiary sampling consisting of spatula extraction of the analyte (0,5-1,0 gram) was carried out in and by the analytical laboratory involved (interesting minor sampling error effects were detected here, fully reported elsewhere in the first authors M.Sc. thesis; luckily these were detected early and were not of a magnitude to interfere with the first order conclusions reported below).

In order to quantify the Total Sampling Error (TSE) and to evaluate the magnitude of the soil heterogeneity on different levels, two experimental designs were embedded in the field sampling plan.

Methods

Primary sampling was conducted in August 2014, where mild weather resulted in dry soils, giving optimal conditions to distinguish between different soil horizons, and in general making field sampling easier. Due to the need for comparison between the final data, the entire sampling campaign was carried out under identical conditions.

Two cultivated fields, A & B and one uncultivated field, X were sampled on the same day, in which a total of 41 samples were collected. The three fields are located in the now forested area "Vestre Indlæg", Figures 1 and 2a, and have never been involved in previous studies. The stratum of interests, according to archeological experiences, manifests itself as a yellow quartz-rich sand underlying a purple heather-rich sandy topsoil, which was found just under the contemporary O horizon. The purple, heather-rich topsoil was used as an upper boundary demarcation due to its marked, recognizable characteristics, while the lower boundary of the target stratum was not identified (generally located 45-55 cm below the surface in the area).

The fields were prepared with 9 sampling locations for the uncultivated field X and 10 sampling locations for each of the two cultivated fields A & B. For the latter two, different sampling plans were chosen: Cultivated field A was sampled along a transect while the cultivated field B was sampled in a random grid within the archeologically delineated boundary. The uncultivated field X was also sampled along a transect which constituted an extension of the

Figure 2 a,b. Line transect (left) and expanded local embedded sampling ("box") (right), see text for details.

transect for field A, Figure 1. The experimental design thus totals 29 samples. Each single sample was collected as a 4-increment composite sample as explained below.

Primary sampling

Each cross in Figure 1 denotes a sample location, approx. 20 × 20 × 20 cm. The vertical dimension of the sample dug outs was constant in order not to incur unnecessary Increment Delimitation Error (IDE). A four-increment composite sample was manually collected from each box with a combined use of a garden shovel and a trowel, Figure 2b. Each increment was composed of an equal volume scrape-off material from one side of the box. In the field, when aggregated these four increments were deemed to constitute a representative, Incorrect Sampling Error (ISE)-eliminating and Correct Sampling Error (CSE)-minimizing composite sample. Identical use of the sampling tools allowed a minimum Increment Delimitation Error, Increment Extraction Error (IEE) & Increment Preparation Error (IPE) (the precise trowel was used to scrape off material into the garden shovel which was used to allow all the scraped-off material to be carefully collected – eliminating spillage and/or contamination). This sampling procedure also honors the Fundamental

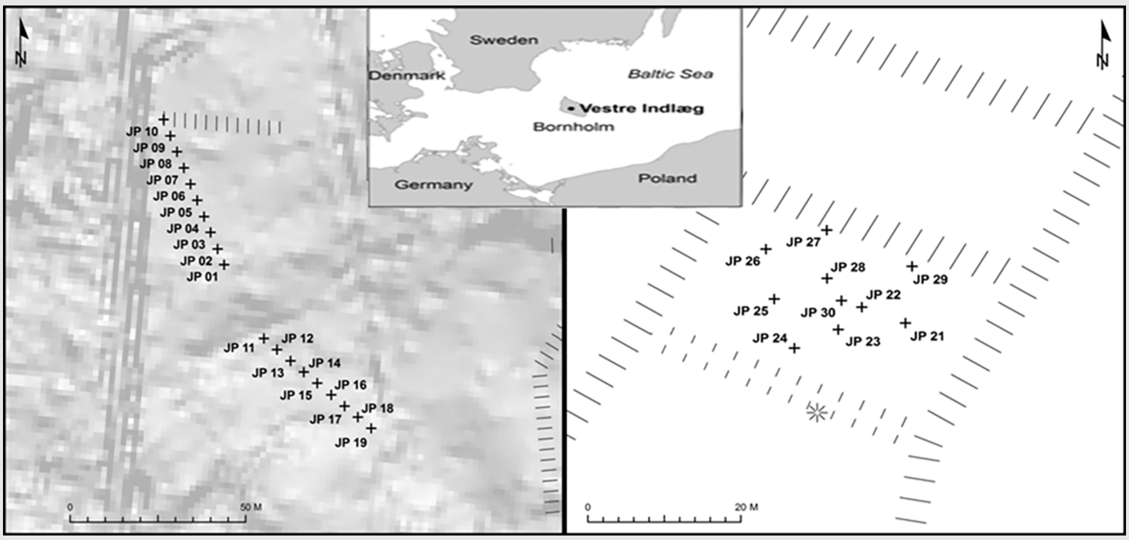

Figure 1 a,b. Location and pilot study area on the Danish island of Bornholm (Baltic Sea)

Sampling Principle (FSP) because the soil underlying the entire location has an equal possibility to end up as a part of the sample (the edge direction of the sample dug out was chosen at random; there should always be some random element in all good sampling procedures).

Local embedded sampling

In order to be able to quantify the heterogeneity at different field scales, an additional experiment was embedded in the overall experimental plan described above. For each of the three fields (at a randomly selected location along the transects or within the grid), a small-scale replication experiment, DS 3077 (2013) was carried out in the form of four additional field samples arranged in a "box-like" pattern, Figure 2b.

The standard dugout was here expanded in size to approx. 100 × 100 × 40 cm which allowed improved pedological characterization as a basis for a larger sample size. These special samples were collected using the same general protocol as previous, but all four sides were now specifically not combined to form a composite sample. Instead each side was sampled separately along the entire wall face. This resulted in four individual 'parallel' samples (designated w, x, y and z) + the existing composite sample (c) belonging to the transect (or grid). This set up allow quantification of the local heterogeneity for each field commensurate with dimensions 100 × 100 cm (termed embedded "boxes"). This replication scheme added 12 samples, the entire pilot project now totaling 41 samples.

Laboratory sample processing

Laboratory sample preparations comprised drying, homogenization and sieving through a 2 mm sieve. The sieving process was carried out with an effort to minimize spillage (IPE). After sieving, the samples were mass reduced using a RAKO Riffle Splitter (32 chutes) to a sample size of 2-3 gram. Laboratory mass reduction meets all the requirements for representative mass reduction as laid out by Petersen et al. (2004). Finally the samples were analyzed for 15 geochemical elements by Inductively Coupled Plasma analysis (ICP), courtesy of Aalborg University, campus Esbjerg.

Lot characterization

Field heterogeneity

Traditionally it has been argued, that for comparison between the geochemistry of different fields, only one single 'representative' sample is needed from each. There are countless examples in the literature where 'representativity' is only assumed for a single grab sample however, very often without proper documentation. But from even a cursory examination of this approach, in the light of TOS' understanding of heterogeneity, it is extremely likely that this can never result in reliable conclusions. A single sample is a grab sample w.r.t. the field it is supposed to represent; there is no way this can express both the local as well as the "global" field heterogeneity in a valid fashion; such an approach is therefore not to be considered trustworthy. Any singular grab sample from any one field cannot be representative hereof without specific proof.

Therefore the primary field sampling constitutes a replication experiment with respect to the full heterogeneity within each field. The overall heterogeneity can be regarded as a specific signature, characteristic of the scale pertaining to cultivated as well as uncultivated fields, but it cannot necessarily be assumed to be identical between fields, Figures 3 and 5. Thus 9 (or 10) composite samples from each field constitutes a replication experiment allowing reliable aggregated results, and also to detect, and remove, outliers, whether defined by the heterogeneity or TSE (one analytical outlier was detected only because of this type of inter-leaved replication experiments in the ultimate laboratory stage). Each field is at the outset considered as a unique sampling target characterized by 9(10) samples covering the specific lot geometry. 9/10 were chosen based on the available logistical constraints (this number could alternatively had been higher, e.g. 20 if no economical, practical, or logistical sampling limitations had existed).

The sample plans were laid out at random – either as a randomly selected transect direction, or as a randomly oriented grid. Replicate samples from each field are hypothesized to correlate stronger within-group than with respect to between-group (between-fields). The two cultivated fields A & B are also assumed to correlate stronger between themselves contra the uncultivated field X. Such relationships would be expected if the geochemical discrimination hypothesis is to be substantiated. But does this hold for all geochemical elements analyzed for? Or just for a few?

Local embedded replicate experiment

The sampling process of the "local" box replication experiment was described above, allowing quantification of the heterogeneity pertaining to this local scale. It is a fair assumption that with five samples it should be possible to express the local heterogeneity with reasonable resolution; these samples should be correlated stronger with each other than with respect to the whole field data, see Figure 5 below.

Univariate data analysis

A traditional univariate data analysis, visualized as a box plot, Figure 3, is carried out for Phosphorous based on data from the three local replication experiments and full field data, allowing to characterize and compare the local and global heterogeneity of each field. Figure 3 will also show to which degree it is possible to distinguish between cultivated and uncultivated fields within this traditional univariate regime.

Table 1 presents the relevant averages and standard deviations. Comparing the three sets of "Box characterizations" it is not

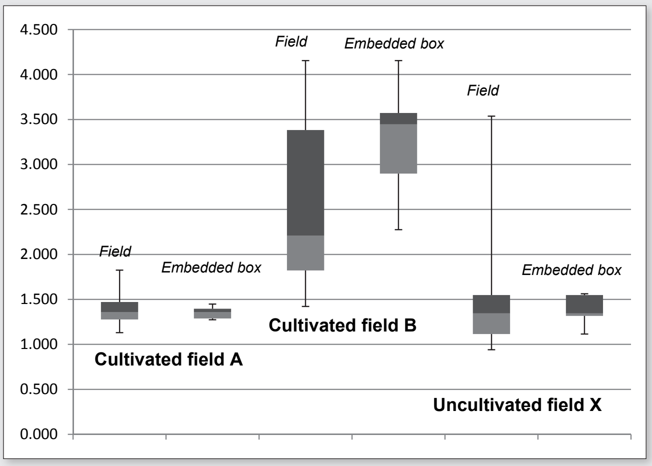

Figure 3. Box-plots of field vs. box heterogeneity characteristics for the element Phosphorous

Table 1. Phosphorous data characterisation (ICP)

ICP results for the three boxes					
Cultivated field A					
	9	9w	9x	9y	9z
Phosphor (mg/L)	1,396	1,288	1,448	1,360	1,274
Average	1,353				
Std. Deviation	0,073125				
Cultivated field B					
	30	30w	30x	30y	30z
Phosphor (mg/L)	2,275	3,573	2,898	4,155	3,446
Average	3,269				
Std. Deviation	0,713165				
Uncultivated field X					
	13	13w	13x	13y	13z
Phosphor (mg/L)	1,344	1,548	1,563	1,317	1,115
Average	1,377				
Std. Deviation	0,185182				

possible to conclude that fields A & B are in fact both cultivated. On the contrary, if this approach is used one would probably conclude that field B was cultivated (because of its elevated P levels) while field A and field X represent uncultivated areas. This is manifestly incorrect however – and the univariate approach fails. The box plot evidence also pictures the difference between the local and global phosphor heterogeneity and, as assumed a priori, the local heterogeneity constitutes but a fraction of the global field heterogeneity.

From the standard deviations one can conclude that the variability for phosphor is largest in cultivated field B.

Multivariate Data Analysis (MVDA)

Clear limitations and attending misinterpretations were found by the univariate approach. This is due to the fact that cultivated fields apparently do not have the same levels of elevated phosphor concentration. But even though the P concentration is low, field A is actually cultivated as shown by irrefutable archeological evidences. Perhaps such relationships can be better appreciated from a multivariate approach when considering a range of 15 elements simultaneously?

MVDA is an approach in which the covariance structure of different datasets is modeled and visualized based on the correlations between the variables included. MVDA contains different methods that can handle different data analysis objectives. One of the powerful tools is Principal Component Analysis (PCA), which reveals data structures (exploratory data analysis) in two complementary plots, the so-called scores and loadings plots (Esbensen (2010), Martens & Næs (1989)). Results of PCA carried out on soil metal concentrations are depicted in Figures 4–6 (41 objects and 15 variables). PCA on this data set will also allow to survey heterogeneity in the different fields due to the two different experimental designs.

Field characterization

Figure 4 is a first PCA visualization of the overall structure (score plot t1-t3), which depicts the variance-maximized relationships between the three fields. Based on the information modeled by PC 1 and

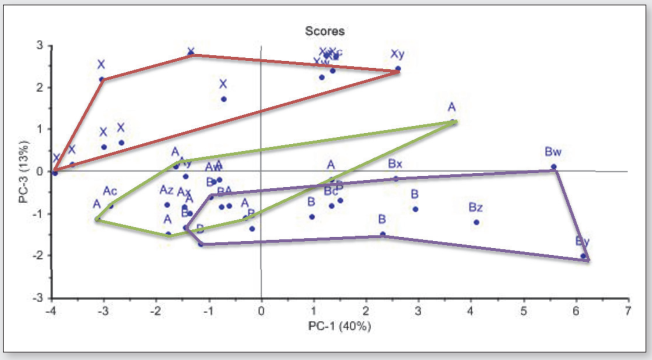

Figure 4. PCA score plot t_1-t_3 of the complete data set. Convex polygons surround data from each of the three fields (X, A, B), see text for details.

PC 3, three clear data groupings (data classes) can be identified, helped along with the known archeological field assignment annotation (A, B, X), which is used to draw convex polygons enveloping the fields. The three fields outline a trend from the uncultivated field X to the two cultivated fields A and B. The latter two fields are only very slightly overlapping due to their distinct geochemistry fingerprints. Based on this simple score plot it is possible to discriminate fully between these two agricultural groups with ease and certainty, but no information about the geochemistry and which elements are causing the between-group trend has been identified – yet. For this the complementary loading plot is needed, in which is depicted the correlation relationships between all the variables involved in the data analysis.

The loading plot, Figure 6, reveal that the strong mutual correlation between [P, Zn, Fe, Mn] is the defining feature for the two cultivated fields A & B, allowing one to conclude that the levels of these elements are elevated in these fields. Due to the group trend from uncultivated to cultivated, which is most pronounced in the vertical direction along PC 3, Mn would appear to be the element that correlates strongest with the cultivated group. Conversely Pb is correlating strongest with the uncultivated field X along PC 1 and B along PC 3.

The multivariate approach is clearly useful for distinguishing between cultivated and uncultivated fields employing 15 geochemical elements instead of one.

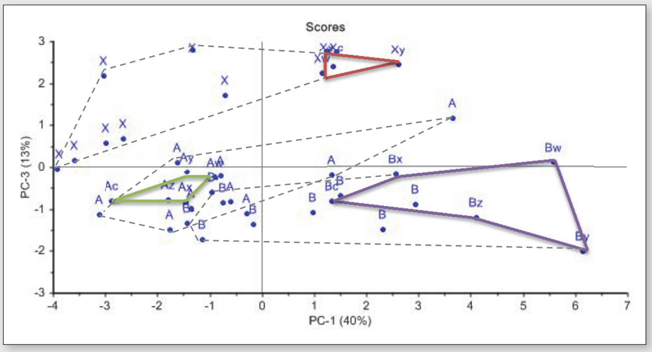

Figure 5. PCA score t1-t3 plot, annotated with convex polygons for the three local box experiments only (heavy lines). Relationships to field-wide classifications (in Fig. 4) indicated by stippled lines

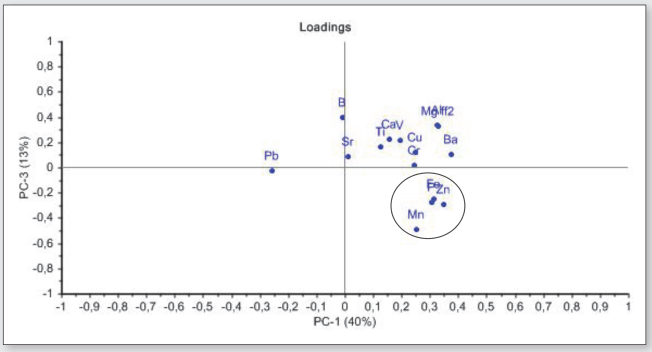

Figure 6. PCA loading plot t_1-t_3 (all data). The most influential correlated elemental group relative to the discriminations seen [Fe, Zn, Mn] in Figures 4 and 5 is indicated (circle); see text for details.

The field replicate experiment also allows a display of the global heterogeneity variations within each field. By use of "connecting lines" one can direct attention to the convex heterogeneity envelope for each field and compare them, which is the annotation used in Figure 4. From this one can argue that they are displaying almost the same degree of field heterogeneity but with different elements as the largest contributors, which can be studied by a more detailed interpretation of Figure 6.

Local heterogeneity characterization

To illustrate the 'local' embedded replicate experiment, the same score plot, Figure 5, can be used again but for this purpose the connecting lines now only frame the sample subsets from the three embedded boxes, emphasizing the local heterogeneity. It is observed that the largest local heterogeneity is indeed found within field B, with much smaller variabilities for field A and X, which show somewhat similar local heterogeneities. From this plot it is also possible to point out potential outliers.

Interestingly the convex polygon that pictures the heterogeneity of uncultivated field X is found as an end-member of the entire field heterogeneity – without the embedded replicate experiment one could perhaps have been led to conclude that this sample could be an outlier.

The above first interpretations from a simple PCA shows the strength of replication experiments on both field and local scales and that the local heterogeneity can vary among, and between fields of different status even. Though small, the present data set is complex to a non-trivial extent, precluding meaningful data analysis based on only one, traditional parameter (P). The complexity is easily and effectively delineated in full measure however when based on the chemometric multivariate approach, PCA[1,4].

The present pilot study data set is not large enough to make a reasonable validation of the strength of PCA solutions calculated. For this it is necessary to invoke a test set, a new data set from similar fields, also taking in at least one of the present fields for re-sampling as well (to be carried out in the summer 2015). Test set validation forms an essential part of proper chemometric data analysis[1,4].

Conclusion

Based on a chemometric multivariate discrimination along PC 3, it is fully possible to distinguish between cultivated and uncultivated Celtic fields on the island of Bornholm – a task for which the traditional P-based univariate approach fails (in the areas investigated here). The present results can therefore be of significant help for archeologists, who until recently would have classified cultivated field A and perhaps many others also, as uncultivated using the traditional univariate P-approach. The multivariate approach is able to yield much more reliable and trustworthy results.

This holds true if – and only if – sampling is done in a representative fashion however, eliminating the majority of all ISE and minimize CSE. Geochemical data typically can contain up to 50% or so random data variance ('data analytical noise'), so PCA decomposition is essential ('shredding data structure from noise').

In this pilot project four elements showed the strongest correlation with the cultivated fields and especially Mn was found to be of pronounced influence. Sparse knowledge as to why Mn, Fe and Zn behave in this correlated fashion with P is raising interest in further geochemical and/or agricultural studies. These relationships could only have been discovered using the chemometric PCA. It will almost always be of interest to increase the number of elements analyzed and e.g. Cobalt should be an element that are of significance in the archeological world.

Through two different experimental designs it was found that each field is characterized by quite similar overall heterogeneities, and that the local heterogeneity (embedded box experiments) was indeed significantly less extensive, Figure 5. The largest heterogeneity was found in cultivated field B, which also had the largest levels of Fe, Zn, P and Mn, Figure 4-6. Geochemical multi-element signatures successfully define different data classes (fields) outlining their internal structures and variable correlations. Why the uncultivated field is particularly strongly correlated with Pb and B is not fully understood at present, an issue that is incorporated in the planned follow-up studies (2015).

All the above findings could only have been discovered using MVDA: Archeology meets TOS meets Chemometrics.

Biography

Bastian Germundsson is a M.Sc. student at IGN & GEUS, currently engaged in different projects with primary focus on sampling issues and with special interests in environmental – and urban geochemistry. Anders Pihl is an archaeologist at Bornholm's Museum with primary interest in Celtic agriculture. Kim H. Esbensen is research Professor at GEUS and AAU, particularly interested in applying TOS and Chemometrics in all of science, technology and industry

References

1. K. H. Esbensen, Multivariate data analysis: in practice: an introduction to multivariate data analysis and experimental design. 5.th Ed. 598 p. CAMO Publ. (2010). ISBN 82-993330-3-2
2. L. Petersen, C. K. Dahl, & K.H. Esbensen, Representative mass reduction in sampling—a critical survey of techniques and hardware.Chemometrics and Intelligent Laboratory Systems, 74(1), 95-114 (2004).
3. DS 3077 (2013). Representative Sampling – Horizontal Standard. Danish Standardisation. www.ds.dk
4. H. Martens, & T. Næs, Multivariate Calibration. Wiley. (1989). ISBN 0-471-90979-3
5. N.H. Nielsen, & S.M. Kristiansen, Identifying ancient manuring: traditional phosphate vs. multi-element analysis of archaeological soil. Journal of Archaeological Science, 42, 390-398 (2014).