*Manuscript (including figures, tables, text graphics and associated captions)
Click here to view linked References

APPLICATION OF DATA FUSION TECHNIQUES TO DIRECT GEOGRAPHICAL TRACEABILITY INDICATORS

Michele Silvestri; Lucia Bertacchini; Caterina Durante, Andrea Marchetti, Elisa Salvatore, Marina Cocchi,

Department of Chemical and Geological Sciences, University of Modena and Reggio Emilia, Via Campi, 183 – 41125 Modena (Italy).

* corresponding author: marina.cocchi@unimore.it Tel. +390592055029; Fax: +39059373543

Abstract

A hierarchical data fusion approach has been developed proposing Multivariate Curve Resolution (MCR) as a variable reduction tool.

The case study presented concerns the characterization of soil samples of the Modena District. It was performed in order to understand, at a pilot study stage, the geographical variability of the zone prior to planning a representative soils sampling to derive geographical traceability models for Lambrusco Wines. Soils samples were collected from four producers of Lambrusco Wines, insisting in in-plane and hill areas. Depending on the extension of the sampled fields the number of points collected varies from three to five and, for each point, five depth levels were considered.

The different data blocks consisted of X-ray powder diffraction (XRDP) spectra, metals concentrations relative to thirty-four elements and the $^{87}Sr/^{86}Sr$ isotopic abundance ratio, a very promising geographical traceability marker.

A multi steps data fusion strategy has been adopted. Firstly, the metals concentrations dataset was weighted and concatenated with the values of strontium isotopic ratio and compressed. The resolved components describe common patterns of variation of metals content and strontium isotopic ratio. The X-ray powder spectra profiles were resolved in three main components that

can be referred to calcite, quartz and clays contributions. Then, an high-level data fusion approach was applied by combining the components arising from the previous data sets.

The results show interesting links among the different components arising from XRDP, the metals pattern and to which of these $^{87}Sr/^{86}Sr$ Isotopic Ratio variation is closer. The combined information allowed capturing the variability of the analyzed soil samples.

Keywords

Multivariate Curve Resolution, Hierarchical data fusion, XRDP, soils characterization, geographical traceability markers.

1. Introduction

High-throughput methodologies, megavariate database, fast fingerprinting and profiling techniques are words that nowadays are commonly present in recent literature. The recent development of analytical techniques able to describe, in a fast way and from different points of view, several features of the samples under investigations has resulted, in lot of cases, in the paradox of having a huge amount of data without an effective interpretation [1, 2]. Data fusion approaches [3-8] are oriented towards the simultaneous use of the information arising from data of different nature. The joint evaluation of the analytical results allows, at the same time, to better describe the investigated system and to answer different questions pertaining to which information it is expected (it is possible) to gain from the different sets/blocks of data, such as highlighting the consensus, common and distinctive information carried by each block, and the linkage among them [8].

Data fusion techniques can be classified in three main groups: a) *low-level data fusion* consists of the simple concatenation of the data of different nature *b) mid-level data fusion* is based on features extraction or variable selection prior to multivariate analysis *c) high-level or hierarchical data fusion* is based on the concatenation of the scores, extracted by means of multivariate projection techniques such as PCA, PLS, *etc.* [3, 7, 9] or wavelet transform [4], i.e. models are built separately on the different data blocks and the derived latent variables (or meta variables in a broad sense) are fused to obtain a final high-multivariate-model.

This last approach could be particularly effective in the case of data blocks, which are difficult to render comparable/commensurable, i.e. a suitable preprocessing procedure may not be available or completely solve the issue.

To the best of our knowledge, multivariate curve resolution (MCR) methodology has not yet been used as data reduction technique for extraction of data blocks information in high-level data fusion. The possibility to obtain chemically meaningful components, e.g. that can be

characterized in terms of chemical concentration and spectra profiles, allows a better understanding and highlighting, in the data fusion process, of the correlation between the resolved profiles of the different analytical techniques.

Most often, in data fusion context, MCR has been used combining the information acquired by the different analytical techniques in the multi-sets structure [10-12]. This is surely a sound approach, however it may be non optimal when the data sets to be fused all share the samples mode and each data block is constituted of different kind of variables, e.g. metal contents and spectral fingerprint for the same set of samples, but there is not a varying condition for each sample such as time of measurement, pH, or a second spectral dimension. In other word, when data augmentation limits to variables concatenation and there is not real replicate information for the same sample to assist the resolution of the underlying components.

Here, we present a case study where MCR was used as variable reduction tools for the development of hierarchical data fusion model in a study aimed at achieving information about the geochemical variability of soils samples.

In particular, this work is a part of a pilot study belonging to a project concerning assessment of geographical traceability models for Lambrusco wines of protected denomination of origin (PDO), a typical food product of the Province of Modena (Italy).

Food geographical traceability studies are targeted to establish the correlation between the soils of origin and the final products, hence, one of the main aspect to face is the representativeness of the sampling of the territories under investigation. To characterize soils samples the determination of several metals content together with the isotopic abundance ratio $^{87}Sr/^{86}Sr$ (a very promising geographical traceability marker) were used jointly with the X-Ray Powder diffraction profile to obtain information about the inter and intra-site variability, including depth, of soil samples at few selected locations, as well as to evaluate the link among main mineralogical phases, metals and isotopic abundance ratio $^{87}Sr/^{86}Sr$.

The MCR data fusion approach was preferred instead of the multiset based one, for the reason explained above taking into account the great difference among data blocks in terms of number of variables and measurement scales.

Several examples are present in literature for the identification of patterns of variation of pollutants or metal sources based on MCR [13, 14] whereas it is the first time that an approach based on multivariate curve resolution is proposed to attempt a partial resolution of XRDP components. In particular, we were interested on one hand to fully exploit soil samples differences and similarity, by building a comprehensive model merging the resolved MCR components from metals and isotopic ratio dataset with the XRDP one, and on the other hand to focus on the linking relations between the fused data blocks.

2. Materials and Methods

2.1. Experimental

The sampling procedure and the analytical determinations concerning the data reported in each data block have been described in our previous works [15 and references therein]. Here we will briefly report only the salient information.

2.1.1. Soils Sampling

Production of PDO Lambrusco wines is subjected to stringent regulations [16, 17] allowing grapes cultivation in the whole district of Modena. The territory of the Province of Modena varies from in-plain area (centre-north) to moderate hill area in the south.

The extension of the area (more than 90 km^2) and the amount of Lambrusco producers insisting on it (more than four thousand) made it mandatory to develop a pilot sampling to evaluate, on a reduced scale, variability of soils in the district, sampling conditions and operating procedures [15]. Thus, four long chain producers were considered, three of these producers, named here on as A, B and D are located in in-plain region, where the majority of the production of Lambrusco

114 wines insists, the fourth one, C producer, insists in the hill area. For each producer, depending on
115 the dimension of the field, from three to five coring where collected. In order to obtain
116 information about both horizontal and vertical variability, each core was split in five aliquots of
117 10 cm of length, starting from 10 cm of depth to 60 cm. All depths were analyzed for the hill
118 field and only lower and upper aliquots for the plain ones for a total of 47 samples.

119 *2.1.2 Analytical Determination*

120 X-ray diffraction of powder, XRDP, was carried out by a θ/θ PANalytical X'Pert Powder
121 diffractometer equipped with a Real Time Multiple Strip (RTMS) detector (PANalytical
122 X'Celerator). Metals quantification needed a preliminary phase of sample pretreatment. Analytes
123 were transferred from soils into solution, in order to perform the analytical measurement, by
124 means of an acid leaching, with concentrated Suprapur® HNO_3, assisted by microwave. The
125 quantification of the Ca, Mg, K, Na content was evaluated by means of a F-AAS (SpectrAA
126 220FS, supplied by Varian, equipped with a sample introduction/dilution system, SIPS 10) . An
127 inductively coupled plasma mass spectrometer, ICP/qMS, XSeriesII from Thermo Fisher
128 Scientific (Bremen, Germany), was used for the determination of the following isotopes: ^{7}Li,
129 ^{51}V, ^{52}Cr, ^{60}Ni, ^{63}Cu, ^{66}Zn, ^{68}Zn, ^{71}Zn, ^{85}Rb, ^{88}Sr, ^{109}Ag, ^{114}Cd, ^{133}Cs, ^{137}Ba, ^{139}La, ^{140}Ce, ^{146}Pr,
130 ^{149}Sm, ^{151}Eu, ^{158}Gd, ^{163}Dy, ^{165}Ho, ^{167}Er, ^{169}Tm, ^{172}Yb, ^{175}Lu, ^{205}Tl, ^{208}Pb, ^{232}Th, ^{238}U. The
131 evaluation of isotopic abundance ratio $^{87}Sr/^{86}Sr$ was achieved by means of a multi-collector high-
132 resolution ICP/MS Neptune® provided by Thermo Scientific after the separation of the isobaric
133 interference of rubidium via SPE with Eichrom's Sr Resin (4,4'(5')-di-t-butylcyclohexano 18-
134 crown-6 crown ether).

135 *2.2 Data Analysis*

136 *2.2.1 Multivariate Curve Resolution*

137 MCR is based on bilinear decomposition of the data matrix [18, 19] according to the model:

$$\mathbf{D} = \mathbf{C}\,\mathbf{S}^{T}$$

The model is calculated by alternating least squares algorithm (MCR-ALS).

Since MCR is not an orthogonal decomposition such as PCA, it needs constraints to resolve the system in a way that the S (spectra) matrix corresponds to a real chemical behavior.

Constraints can be applied both to the spectra (S) and the concentration (C) matrices, in order to reduce the rotational ambiguity of the model since MCR-ALS has not a unique solution. Constraints can be considered as the translation in mathematical formulae of a characteristic of the investigated system.

Two types of constraints can be implemented in an MCR model: i) soft constraints such as non-negativity, unimodality, selectivity and closure constraints that allow reducing rotational ambiguities; ii) hard constraints, based on physicochemical models able to describe the system under investigation such as kinetic or equilibrium model are able to reduce in the same time both rotational and intensity ambiguities.

In the resolution of a chemical system, non-negativity constraints are very common. Furthermore within the family of constraints [20] other usefully adopted for the reduction of the ambiguities are: unimodality (i.e. the resolved profiles are imposed to have only a maximum), closure (i.e. the total amount of the species within the system is constant) and selectivity (i.e. imposition of the presence or absence of a species in a mixture or a region of the spectrum).

Here, we adopted soft constrains, such as non-negativity constraints both for concentrations and spectral profiles. Constrains motivation and application details for each data block are reported in the section 3, Discussion.

2.2.2 Datasets and Preprocessing

As pointed out in the Introduction, we decided to follow a hierarchical data fusion approach, analyzing separately by multivariate curve resolution methodology the two data sets and then merging the extracted concentration profiles by each of them. In fact, the way in which variables are weighted is one of the most important tasks in data-fusion. In particular, in the case of fusion of blocks with very different number of variables and measurement scales, as in this case, low-level data fusion could be difficult. Forcing a block of few variables to have the same variance as a more numerous one can lead to a concatenated dataset having the majority of variable with "intensities" very low and few ones with high values. This is more troublesome in the case of the use of multivariate curve resolution instead of orthogonal methods, such as PCA.

The data fusion approach we adopted addresses two purposes: meaningful components are extracted and the two data blocks are made comparable as more or less the same number of variables (resolved component) each one having a high variance with respect to the original data are concatenated.

The whole data fusion process is illustrated in figure 1 and described here after.

The collected diffractograms consist of 6882 2θ intensities (from 0° to 120° on 2 theta scale). The last part of the signals was cut at 79.99 2θ and diffractograms were then preprocessed in order to reduce noise and background effects and to minimize horizontal shift [15]. The signals were finally arranged in a 47x4488 matrix called *"XRDP dataset"*. By means of MCR-ALS three concentration profiles were extracted and arranged in a 47x3 matrix call *"MCR-XRDP dataset"*.

On the other side, the concentrations (mg kg^{-1}) of the 34 metals were scaled to unit variance (without centering) and merged together with the isotopic abundance ratio values. The obtained matrix, namely *"Met-I.R. dataset"* of dimension 47x35 was block-scaled in a way that the 15% of the total variance correspond to $^{87}Sr/^{86}Sr$ in order to assign to this important primary indicator a value six times higher with respect to each other metals concentration. Strontium abundance

isotopic ratio is an important parameter in geochemistry and geochronology, its value depends on several factors such as the age of formation of soils and the initial concentration of ^{86}Sr, ^{87}Sr and ^{87}Rb. ^{87}Rb is not a stable isotope and decays to ^{87}Sr with an half-life time ($t_{1/2}$) of more than ten to the tenth years. Given all these properties, we decided to enhance the importance of this variable with respect to all the others.

Four concentration profiles were extracted with MCR-ALS and arranged in a 47x4 dataset called *"MCR- Met-I.R. dataset"*

Finally the two concentration profiles matrices were concatenated and block-scaled in order to give the same variance to each block and then MCR-ALS were applied giving the bilinear decomposition of the arranged dataset *"DF MET-I.R.-XRDP Dataset***"** based on three components.

2.2.3 Software

Multivariate Curve Resolution was carried out by MCR-ALS GUI (http://www.ub.edu/mcr/web_mcr/mcrals.html). Arrangement of dataset and fusion of data was obtained by homemade routine written with MATLAB (Mathworks MA, USA) and PLS Toolbox 6.0 (distributed by Eigenvector Research Inc WA, USA).

3 Results and Discussion

This section is divided in three parts describing the application of MCR-ALS to the different datasets: 3.1 results on *"XRDP dataset"*, 3.2 results on *"MET-I.R. dataset*", 3.3 results on *"DF MET-I.R.-XRDP Dataset***"**

3.1 XRDP Dataset

208 Diffractograms of soils powders are very complex signals characterized by a high number of
209 peaks of different intensities. The number of mineralogical phases contributing to the whole
210 signal (chemical rank) depends on the complexity of the soils. In this work the application of
211 MCR-ALS on the XRDP signals is not oriented to a complete quantitative resolution of the
212 system under investigation but it is proposed as tool for the extraction of the meaningful
213 information able to characterize the investigated soils. In fact, here the aim is not the
214 quantification of the single mineralogical phases, but rather a qualitative identification of the
215 phases responsible of the main sources of variability for soil samples differentiation. The use of a
216 rank deficient model for the description of a system leads to a not complete resolution of the
217 components that results in overlapped spectra profiles. If few major components are presents and
218 able to characterize adequately the system under investigation, as in this case, the choice of a
219 rank deficient model is sufficient but the so called "pure component" probably contains part of
220 the information of other components not considered in the curve resolution process.

221 Several attempts were tried for the application of MCR-ALS on the *"XRDP dataset"*, varying
222 the number of components, initial estimates, constraints and normalization strategies.

223 Preliminarily a MCR model with three components was chosen, based on singular value
224 decomposition (SVD) results. Initial estimates were determined by SIMPLISMA [21] on spectra,
225 non-negativity constraints were applied both for concentration and spectra, and spectra profiles
226 were height normalized. Two of the three resolved spectra profiles (not reported) by this
227 preliminary model resulted to be very similar to signals of the pure compounds calcite (the most
228 stable polymorph of calcium carbonate) and quartz (silicon dioxide). In the third spectra profile
229 most of the bands were present in the low 2θ region and many clays related peaks could be
230 identified in the middle region of the spectrum.

231 Quartz, clays and calcite, in order of abundance, are the main constituents covering the majority
232 of the compositional profile of soils insisting in the Modena district, so it seems reasonable to

refine the model described above implementing selectivity constraints in order to reduce the rotational ambiguity of the system.

In particular, few selected peaks were forced to be modeled only by one of the component on the basis of the XRPD signal of the pure quartz and calcite respectively, as retrieved by the database powder diffraction files version 4 (ICDD). The XRPD region corresponding to 50.1 2θ was imposed to be selective for quartz. The regions corresponding to 29.5 2θ, 39.4 2θ, 48.5 2θ were imposed to be selective for the "calcite" component, in these regions calcite responds, and not any other interfering species, or at least a contribution may be given only by species that are extremely rare in the investigated territory.

The three components model built implementing selectivity constraints explains 92% of the total variance with a lack of fit with respect to PCA of 1.1%. In figure 2 are reported the three resolved spectra profiles, The tentative identification of the three resolved profiles, as clays, calcite and quartz does not pretend we are referring to completely resolved profiles for these pure components, as pointed out above, we are aware of having chosen to fit a rank deficient model leading to not complete resolution. However, the three MCR components, as shown by the results obtained by the MCR model not forced with selectivity constraints and then confirmed by the constrained MCR model, are quite close to the pure spectra of these mineralogical phases, that represent the main constituents of soils of the Province of Modena. Hence, for convenience the terms "clays", "calcite" and "quartz" will be associated to the resolved spectra profiles.

In figure 3 is highlighted a region of the spectra in which is clear the good resolution of other peaks belonging to the calcite component that were not forced by the applied selectivity constraints.

In order to inspect the information about samples, pertaining to the concentration matrix C, a scatter plot representation of its columns is proposed in figure 4. As MCR is not an orthogonal decomposition, this representation, while offering an easier way, with respect to separate bar

graphs for each C column, to formulate considerations about the similarities or differences of samples on the basis of two components at time, does not have the same meaning as samples distance in PCA spaces. However, the coordinates of samples in the concentration space of the MCR factors are directly related to the concentration and should have a more directly interpretable meaning than the distances between samples in the space of orthogonal factors.

The most striking observation that emerges in figure 4 is the separation of the hill samples (producer C) that are separated for low concentration values for the clays component with respect to all the in-plane samples (producers A, B and D). Moreover, the hill samples are split in two groups due to a different amount of both calcite and quartz. The distinction in two clusters of the hill samples is in agreement with a certain degree of soils variability at the site of producer C (hill area) that was also noticed during the on field sampling (15) and confirmed by preliminary results from texture analysis of the same soil samples where the whole content of sand, clay, silt and $CaCO_3$ were determined. The distribution of these components in the different holes shows a great complexity and heterogeneity of these soils. Thus the intra site location of the collected samples may reflect the presence of a more calcareous soil fraction with respect to others characterized by a high amount of quartz. At variance, all in-plain samples are slightly more homogeneous and only the samples of the producer A present a lower amount for the calcite component and higher for the quartz ones with respect to the samples of the producers B and D.

3.2 MET-I.R. Dataset

The dataset *"MET-I.R. dataset"* containing the concentrations of the 34 metals determined and the isotopic abundance ratio $^{87}Sr/^{86}Sr$ was pretreated as described in the previous section, also in this case several attempts were tried for the selection of the best parameters for the application of MCR-ALS.

A four components model was calculated using SIMPLISMA on concentration direction for the evaluation of initial estimates, applying non-negativity in both concentration and "spectra" (variables) direction (99% of total variance explained, lack of fit with respect to PCA 0.8%).

The resolved component profiles (variables mode) are reported in figure 5. Each profile represents a common pattern of variation of the investigated soils resolved by MCR-ALS. The information that can be extracted from these profiles is a characterization of the samples on the basis of groups of metals that vary likewise for all the soils analyzed. It is possible to denote that some variables present high values for almost all the components (Cr, U, Ni).

On the other hand, other variables are not present in some components, for example Calcium presents values near or equal to zero for the first, second and fourth component and the highest one for the third one, as magnesium, which shows the same behavior. Sodium and potassium have high contribution in the second and third components but not in the other two, the same trend is observed for rare earths elements in the first, second and fourth component. Isotopic abundance ratio of $^{87}Sr/^{86}Sr$ presents high values for the first component and low or very low for all the other ones.

Considering the samples concentration profiles obtained for the MET-I.R. dataset, which are reported in figure 6, they share similarities with respect to the trends shown above in figure 4. In particular the second component arising from the MET-I.R. model is associated to a distribution of the samples, as regards concentrations values, very similar to the clays component of the XRDP dataset. The same consideration can be extended for the third MET-I.R. component and the calcite ones. The first component, on the other hand, presents higher values for all the hill samples with respect to the in-plane ones. The fourth component is not able to highlight differences on the groups of soils, only one hill sample is present at very high values indicating a great amount for this soil of one or more of the variables with high contribution on the fourth component (e.g. cadmium).

307 *3.3 DF MET-I.R.-XRDP Dataset*

308 The three resolved concentration profiles from XRDP model were concatenated with the four
309 from the analysis of metal and isotopic ratio dataset in a way that the variance explained by each
310 block was the same, since the aim the data fusion process is to extract the correlation between
311 the two blocks of variables in order to better explain the variability of the soils samples.

312 The resulting "DF MET-I.R.-XRDP dataset" of dimensionality 47x7 was evaluated by a three
313 component MCR-ALS model, using SIMPLISMA on concentration direction for the evaluation
314 of initial estimates and applying non-negativity constraints for both concentration and "spectra"
315 (variables mode). The model is characterized by 95.9% of variance explained, lack of fit with
316 respect to PCA 0.3%.

317 In figure 7a, 7b and 7c are shown the bar graphs of the three MCR components respectively, for
318 the S matrix (variables mode). To better describe the linkage among the components extracted by
319 the different data blocks in figure 7d the three-dimensional scatter plot is shown. In particular, it
320 can be observed the grouping of the quartz component from the XRDP model and the first
321 component from MET-I-R that get the highest value on the first resolved component of the fused
322 dataset. The first MET-I.R. component is principally related to the values of the isotopic
323 abundance ratio of $^{87}Sr/^{86}Sr$, *U, Cr,* (that presents the maximum value, figure 5 top left) and to
324 the contribution of all the rare earths patterns. Thus, in the investigated soil samples to higher
325 quartz content corresponds as well a higher isotopic abundance ratio $^{87}Sr/^{86}Sr$.

326 On the second resolved component of the fused dataset, clays (XRDP data set) and the second
327 and fourth of the MET-I.R. dataset show the highest values. The correlation between one XRDP
328 component and two components from the metal dataset (mostly influenced by transition metals:
329 zinc, nickel, cobalt, vanadium, cadmium, monovalent elements such as rubidium, potassium,
330 sodium, and by the rare earths pattern) highlight the complexity of the possible relation between
331 the different variables (metals composition and lattice structures) and it can also be due to the not

complete resolution of the clays component, that is a combination of several crystalline structure of different clays (muscovite, illite, chlorite, serpentine) typically presenting the inclusion of different metals in the lattice.

The third component shows the correlation between the calcite component and the third MET-I.R. component that, as expected, is characterized by the highest contribution of calcium with respect to all the other spectra profiles and other bivalent elements that commonly are presents in carbonate soils, such as magnesium, strontium, zinc, nickel and by alkaline metals such as sodium, potassium, rubidium.

In figure 8 are reported the concentration profiles. The distribution of the samples presents similarities with the separate models presented above. All in-plain samples have the highest values for the second component (as they showed on the second component in MET-I.R model and the clays component in the XRDP model) but are more differentiated in the first and third component. Samples from the A producer present more positive values for the first component and lower values for the third one with respect to all other in-plain samples. This indicates that the field of producer A has higher amount of quartz and strontium isotopic ratio, the most important variable for the second component of the MET-I.R model.

Regarding the samples from the hill field, it is clear the greater variability compared to the in-plain ones. Analyzing the first and third component plot in figure 8b it is evident the separation in two groups of the hill samples. These components are related to variables from XRDP and MET-I.R. dataset that appear to be mutually complementary. The presence of a high amount of one of them is linked to the presence of a lower one for the other component. These results are in agreement with preliminary information regarding the textures analysis of the same soils, such as the percentage of sand and silt. This variability is so pronounced that some samples of the same hole but at different depth present high values for the third component in some cases and lower for some other.

4. Conclusions

The high-level data fusion approach adopted for the analysis of data of different nature proved to be a powerful tool for a preliminary inspection of blocks of data characterized by a very different dimensionality. In this contest, MCR-ALS proved to be a relevant variable reduction mean because the extracted components bear chemically meaningful information, hence an easier interpretation of the data in each stage of the data fusion process.

When preliminary knowledge about variability and features of samples are not present or not detailed, as in this case, the possibility of the interpretation of the results on the basis of a multifold instrumental approach helps the determination of both the characteristics of samples and of the single results of each block of variables. In particular, the linkage among the metal profile and some mineralogical phases with the isotopic ratio may furnish a preliminary estimate of what variability has to be expected depending on field location and thus aids the planning of soil sampling.

The applications of MCR-ALS on very complex data, such as diffractograms, demonstrate the capability of this tool of achieving good results when used on fingerprinting techniques data in a way similar to an exploratory analysis.

5. Acknowledgments

This work was supported by the AGER, Agroalimentare e Ricerca, cooperative project between grant-making foundations under the section “wine growing and producing” project. New analytical methodologies for varietal and geographical traceability of enological products; contract n. 2011-0285. We are also grateful to Consorzio Marchio Storico Lambruschi Modenesi for use of their facilities during sampling procedures.

Moreover, we would like to thank Prof. A. De Juan for valuable suggestions.

381 **References**

382 [1] L. Eriksson, H. Antti, J. Gottfries, E. Holmes, E. Johansson, F. Lindgren, I. Long, T.
383 Lundstedt, J. Trygg, S. Wold, Anal Bioanal Chem 380 (2004) 419-429.
384 [2] P. J. Gemperline, J. Chemometrics 2007; 21: 507–508
385 [3] J. Forshed, H. Idborg, S.P. Jacobsson, Chemometr Intell Lab, 85 (2007) 102–109
386 [4] Y. Lin, S.D. Brown, Anal Bioanal Chem, 380 (2004) 445-452
387 [5] A. K. Smilde, M. J. van der Werf, S. Bijlsma, B. J. C. van der Werff-van der Vat, R. H.
388 Jellema, Anal Chem, 77 (2005) 6729-6736
389 [6] Y. Nia , Y. Lai, S. Brandes, S. Kokot, Anal Chim Acta, 647 (2009) 149–158
390 [7] S. Wold, N. Kettaneh, K. Tjessem, J. Chemometr, 5 (1996) 463-482
391 [8] I. Van Mechelen, A.K. Smilde, Chemolab 104 (2010) 83-94
392 [9] J. A. Westerhuis, T. Kourti, J. F. MacGregror, J. Chemometrics, 12 (1998) 301–321
393 [10] E. Peré-Trepat, R. Tauler, J. Chromatogr A, 1131 (1-2) (2006) 85-96
394 [11] S. Navea, R. Tauler, A. de Juan, Anal Chem, 78 (2006) 4768-4778
395 [12] S. Mas, R. Tauler, A. de Juan, J. Chromatogr A, 1218-51 (2011) 9260-9268
396 [13] A. Gredilla, J. M. Amigo, S. Fdez-Ortiz de Vallejuelo, A. de Diego, R. Bro, J. M.
397 Madariaga, Anal Methods, 4 (2012) 676]
398 [14] E. Pere-Trepat, A. Ginabreda, R. Tauler, Chemometr Intell Lab, 88 (2007) 69-83
399 [15] L. Bertacchini, C. Durante, A. Marchetti, S. Sighinolfi, M. Silvestri, M. Cocchi, Talanta
400 (2012) http://dx.doi.org/10.1016/j.talanta.2012.06.067
401 [16] Decree of 27 July 2009 published on O.J. no. 187 of 13 August 2009.
402 [17] CE Regulation no. 813 of 17 April 2000.
403 [18] J. Jaumot, R. Gargallo, A. de Juan, R. Tauler, Chemometr Intell Lab, 76 (2005) 101-110
404 [19] S.C. Rutan, A. de Juan, R. Tauler, Comprehensive Chemometrics – Chemical and
405 Biochemical Data Analysis, Volume 2, 249-259

406 [20] A. de Juan, Y. Vander Heyden, R. Tauler, D.L. Massart, Anal Chim Acta 346 (1997) 307-
407 318

408 [21] W. Windig, Chemom. Intell. Lab. Syst. (1997), 36, 3–16.
409

Captions of figures

Figure 1: Schematization of the data fusion process

Figure 2: MCR-ALS resolved spectra profiles of XRDP dataset. a) First component, “clays”. b) Second Component, “calcite”. c) Third Component, “quartz”.

Figure 3: Zoom of the 40-49 2theta region showing superimposed the MCR-ALS resolved profiles for the three XRDP components.

Figure4: MCR-ALS resolved concentration profiles of XRDP dataset. a) “clays” versus “calcite” resolved concentration profiles. b) “calcite” resolved versus “quartz” resolved concentration profiles.

Figure5: MCR-ALS resolved “spectra” profiles (variables mode) of MET-I.R. dataset.

Figure6: MCR-ALS resolved concentration profiles of MET-I.R. dataset. a) third component versus first component; b) second component versus third component; c) first component versus fourth component.

Figure7: MCR-ALS resolved “spectra” (variables mode) of the fused dataset DF MET-I.R.-XRDP. a) First component. b) Second component. c) Third component. d) Scatter plot of three of the resolved components.

Figure8: MCR-ALS resolved concentration profiles of the fused dataset DF MET-I.R.-XRDP. a) Second component versus first component; b) first component resolved versus third component; c) second component versus third component.

Figure 1
Click here to download high resolution image

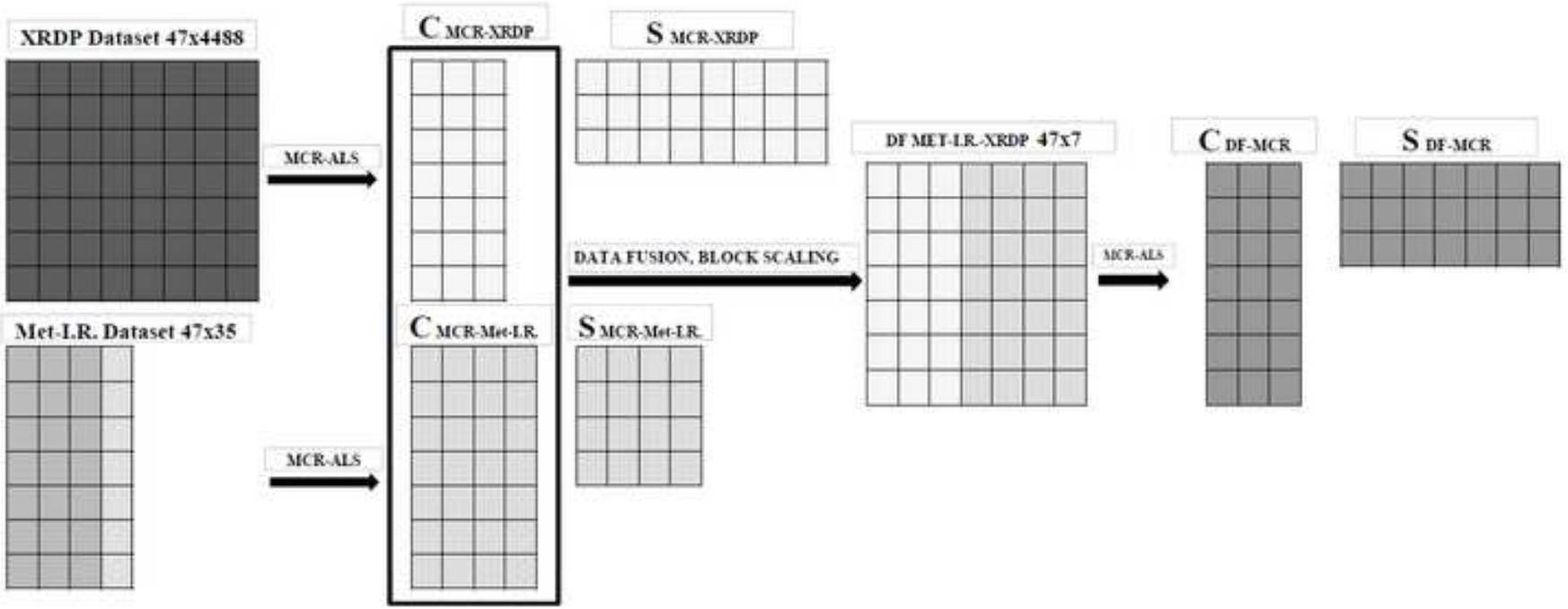

Figure 2
Click here to download high resolution image

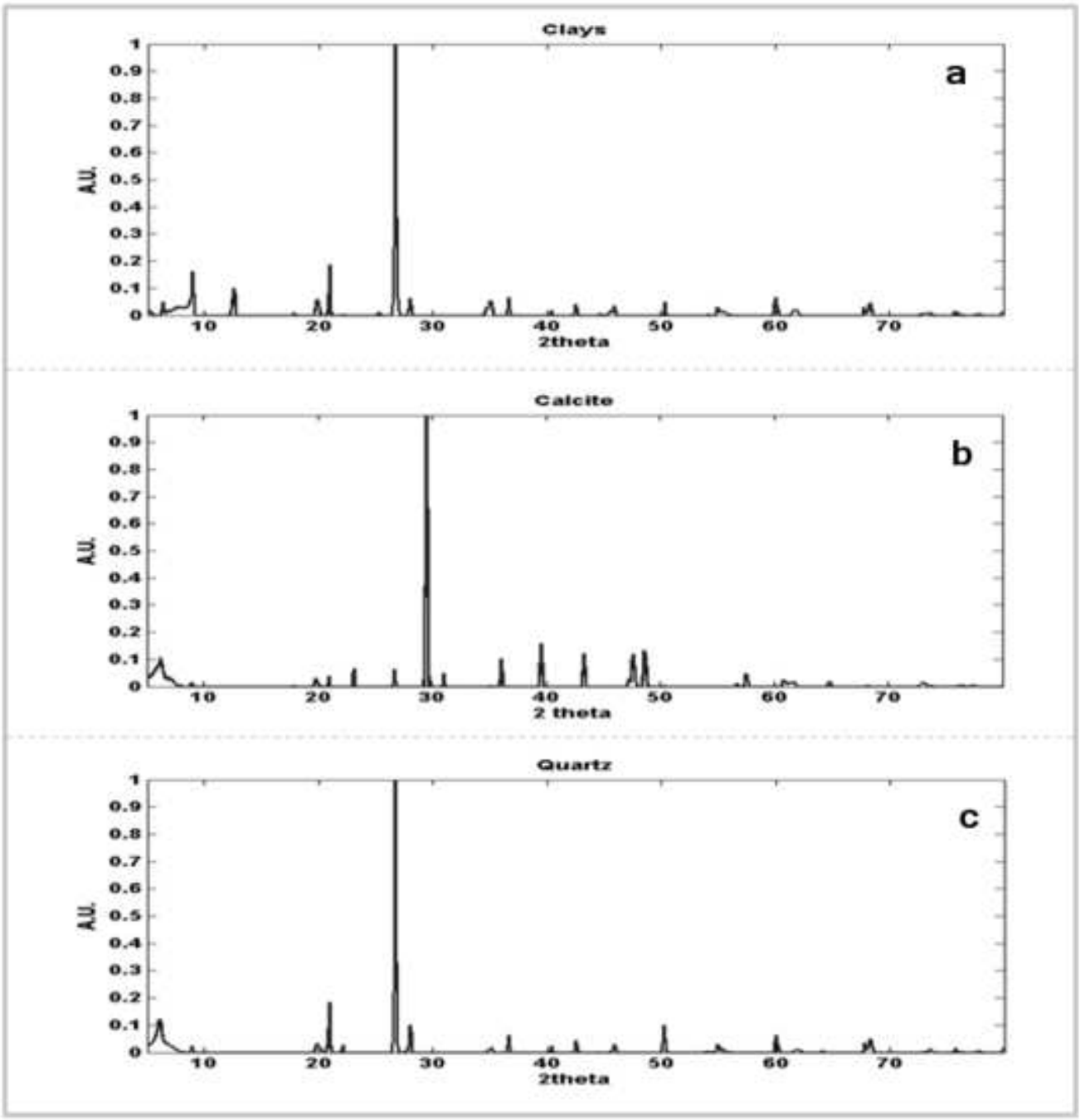

Figure 3
Click here to download high resolution image

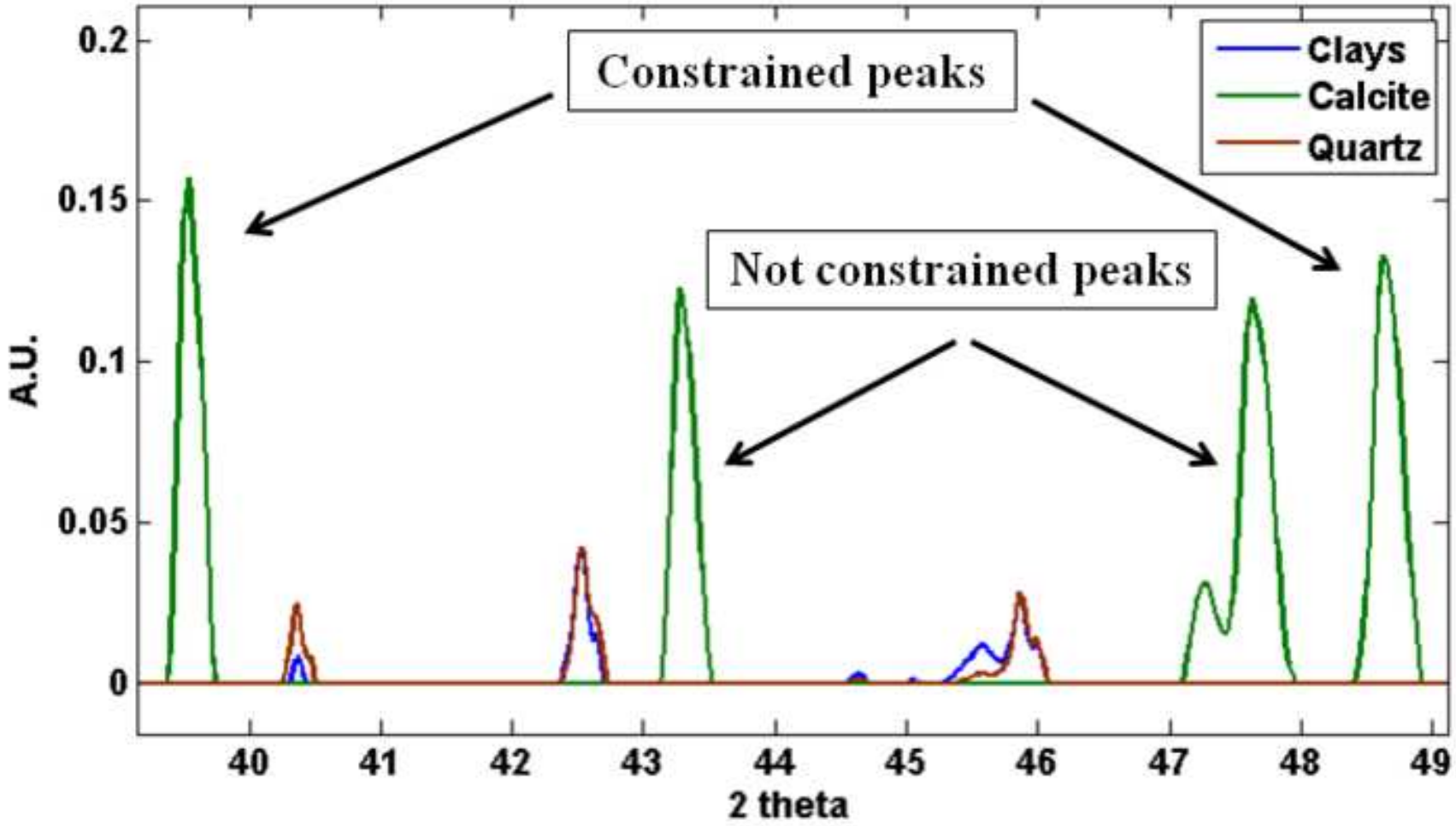

Figure 4a
Click here to download high resolution image

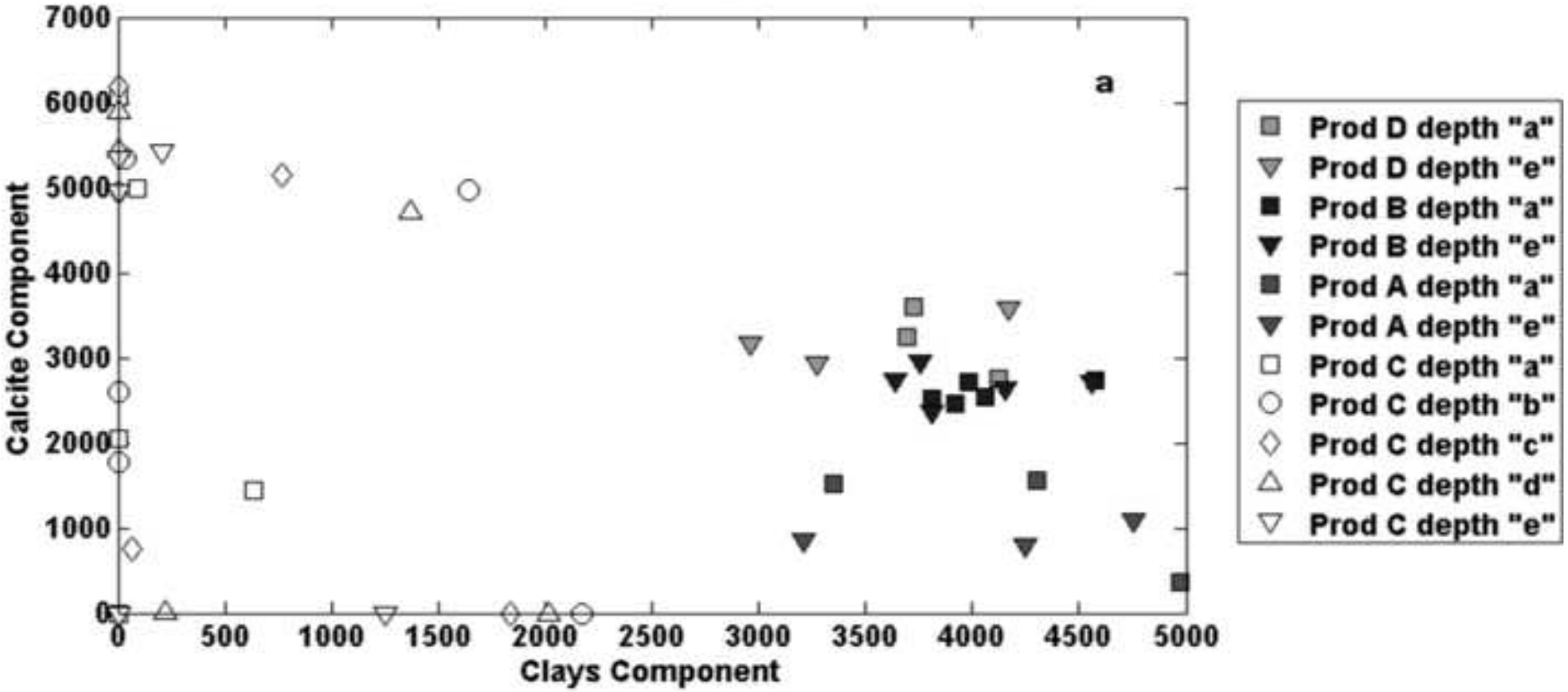

Figure 4b
Click here to download high resolution image

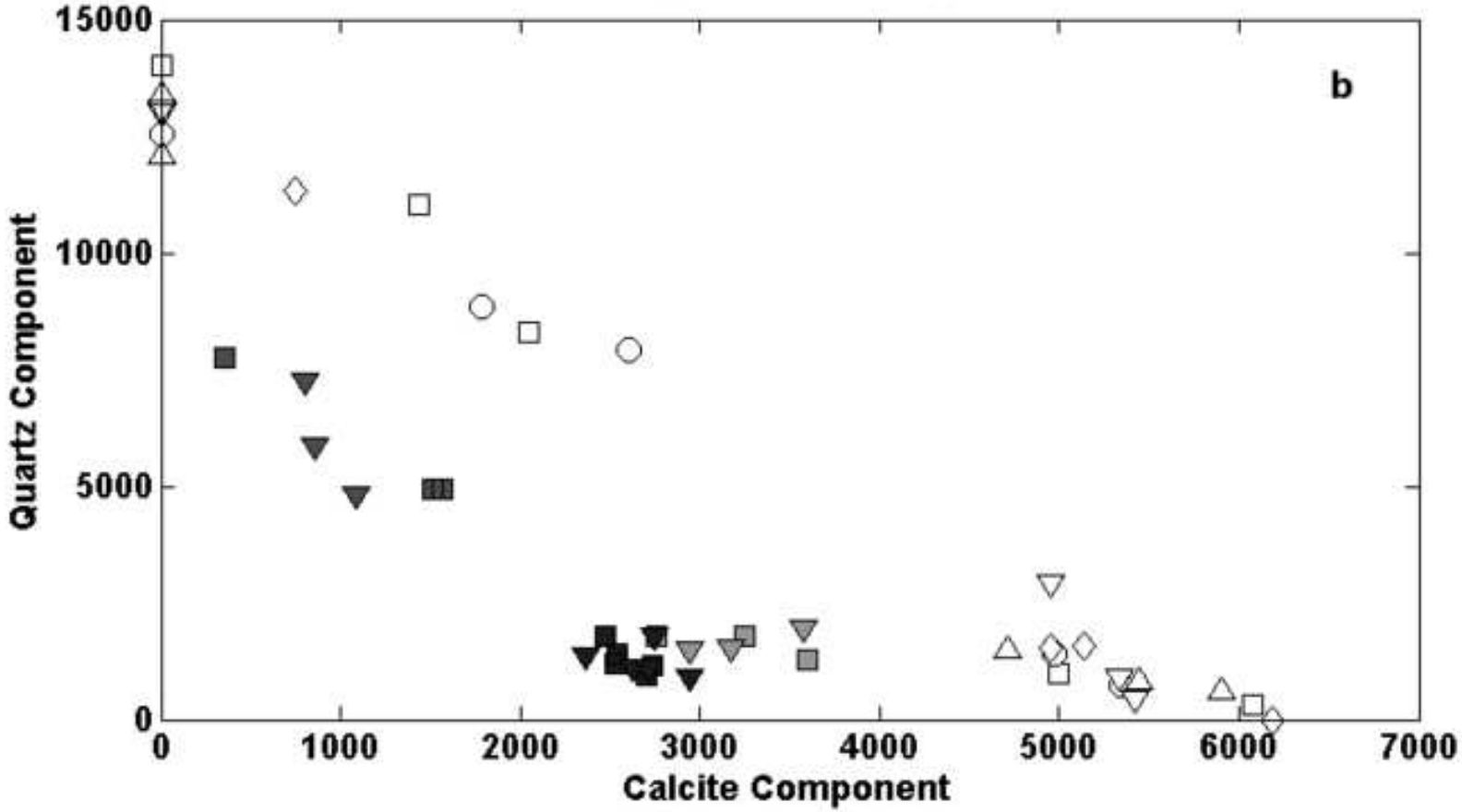

Figure 5
Click here to download high resolution image

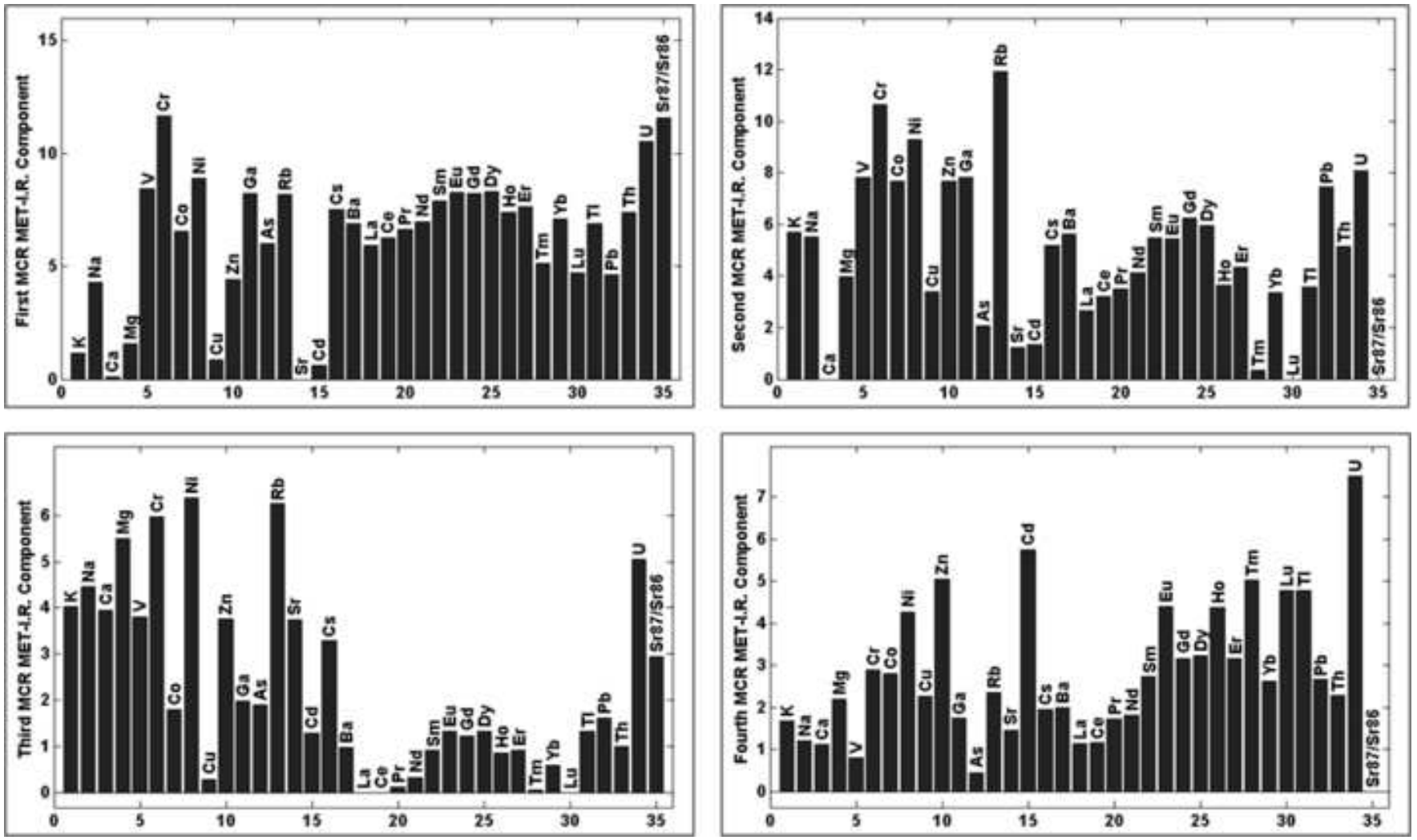

Figure 6a
Click here to download high resolution image

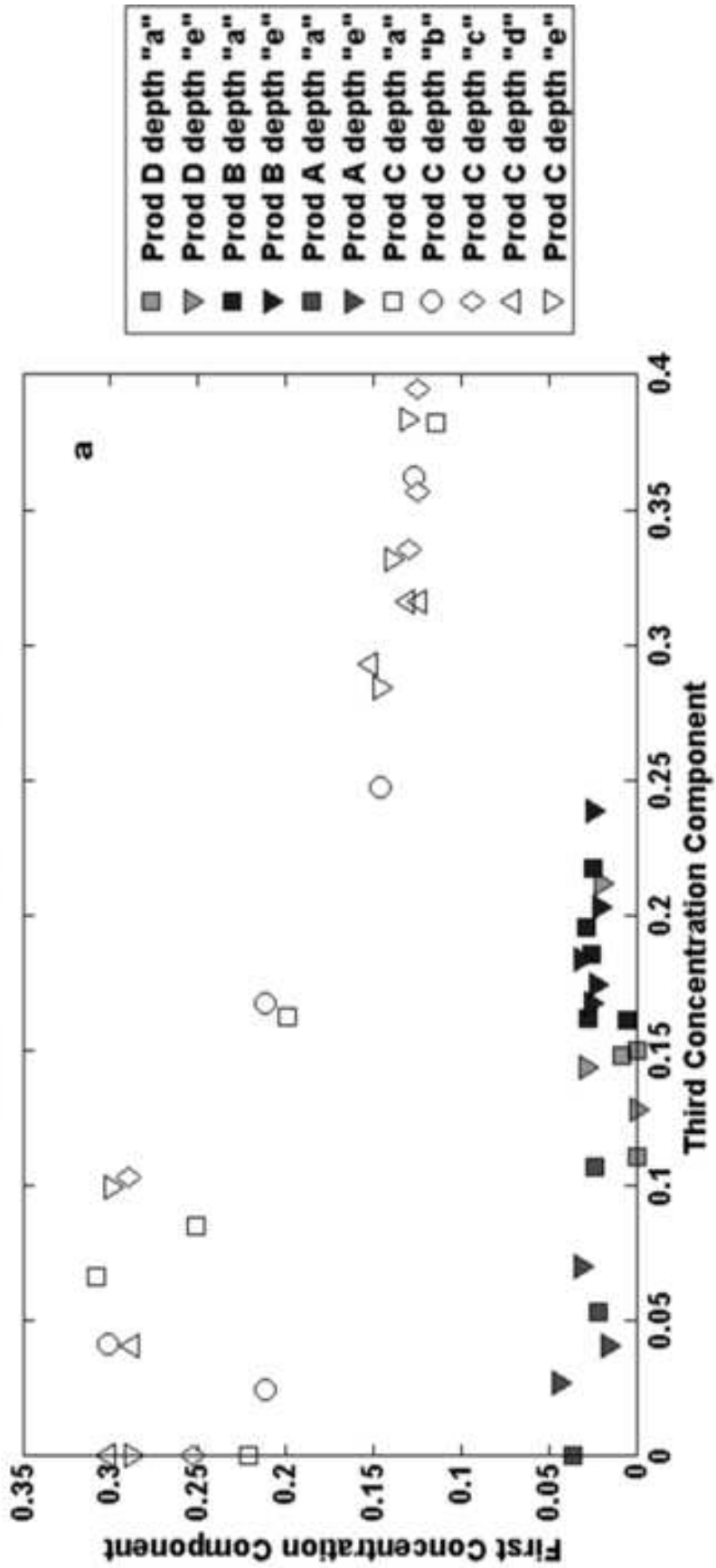

Figure 6b
Click here to download high resolution image

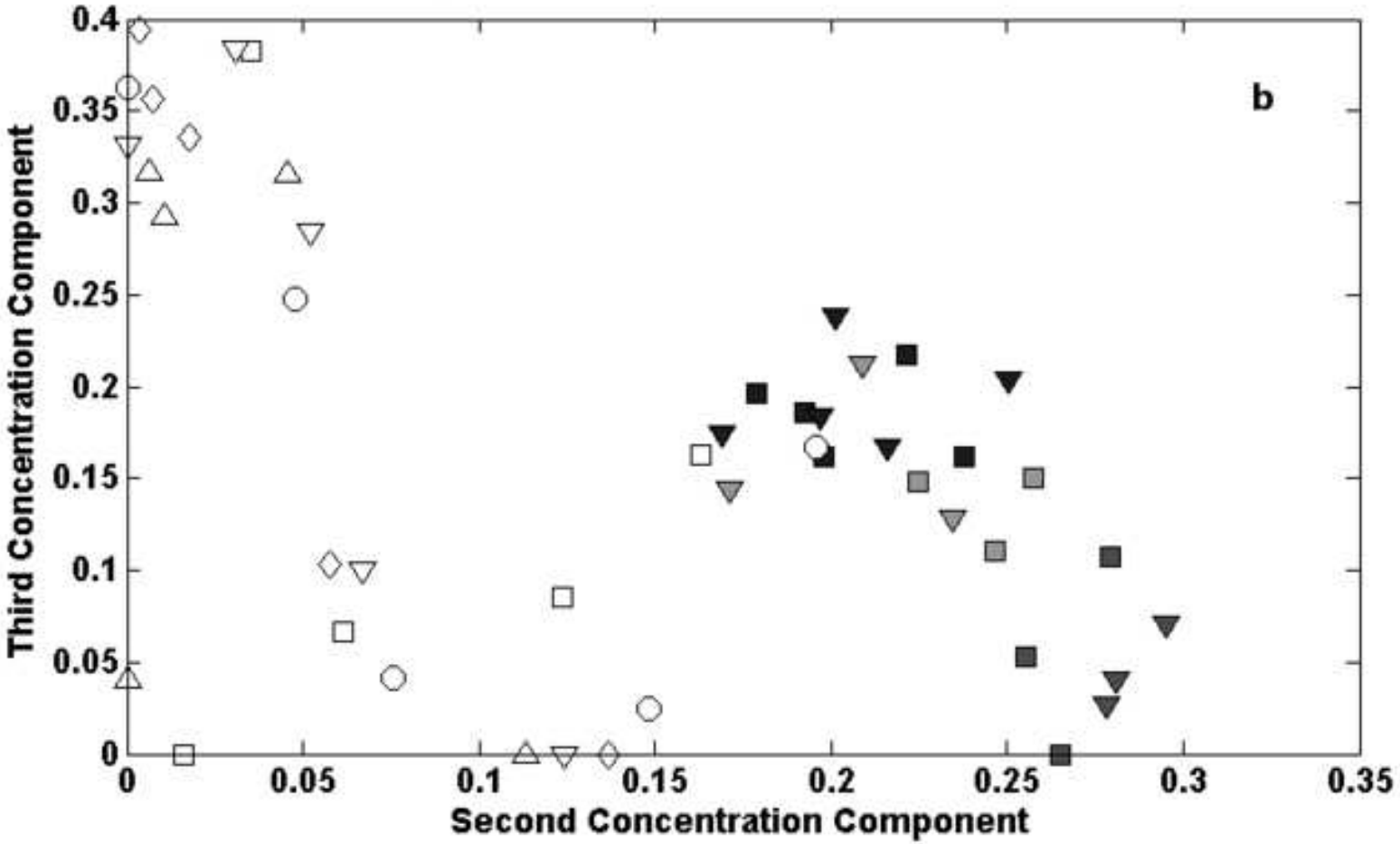

Figure 6c
Click here to download high resolution image

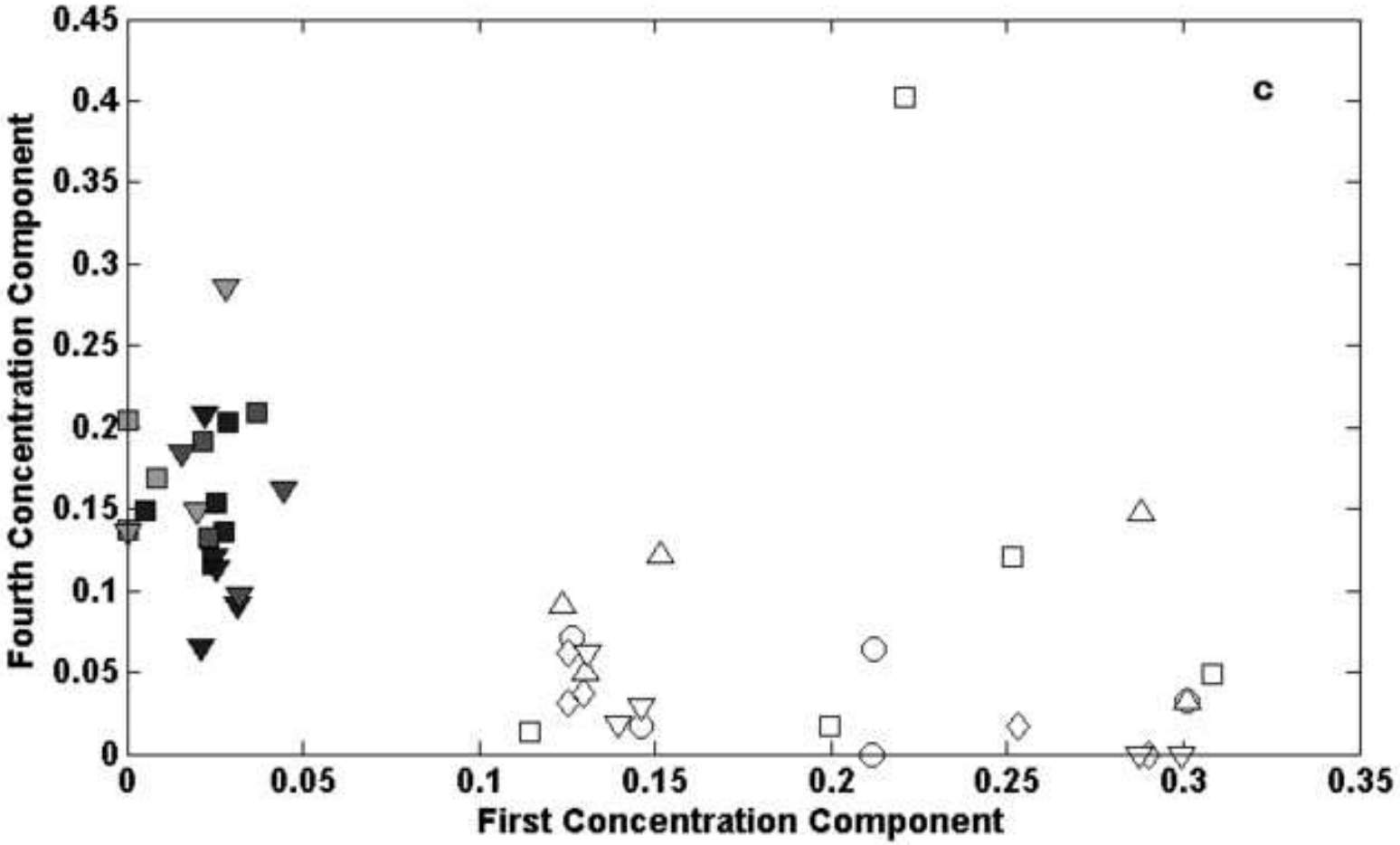

Figure 7a
Click here to download high resolution image

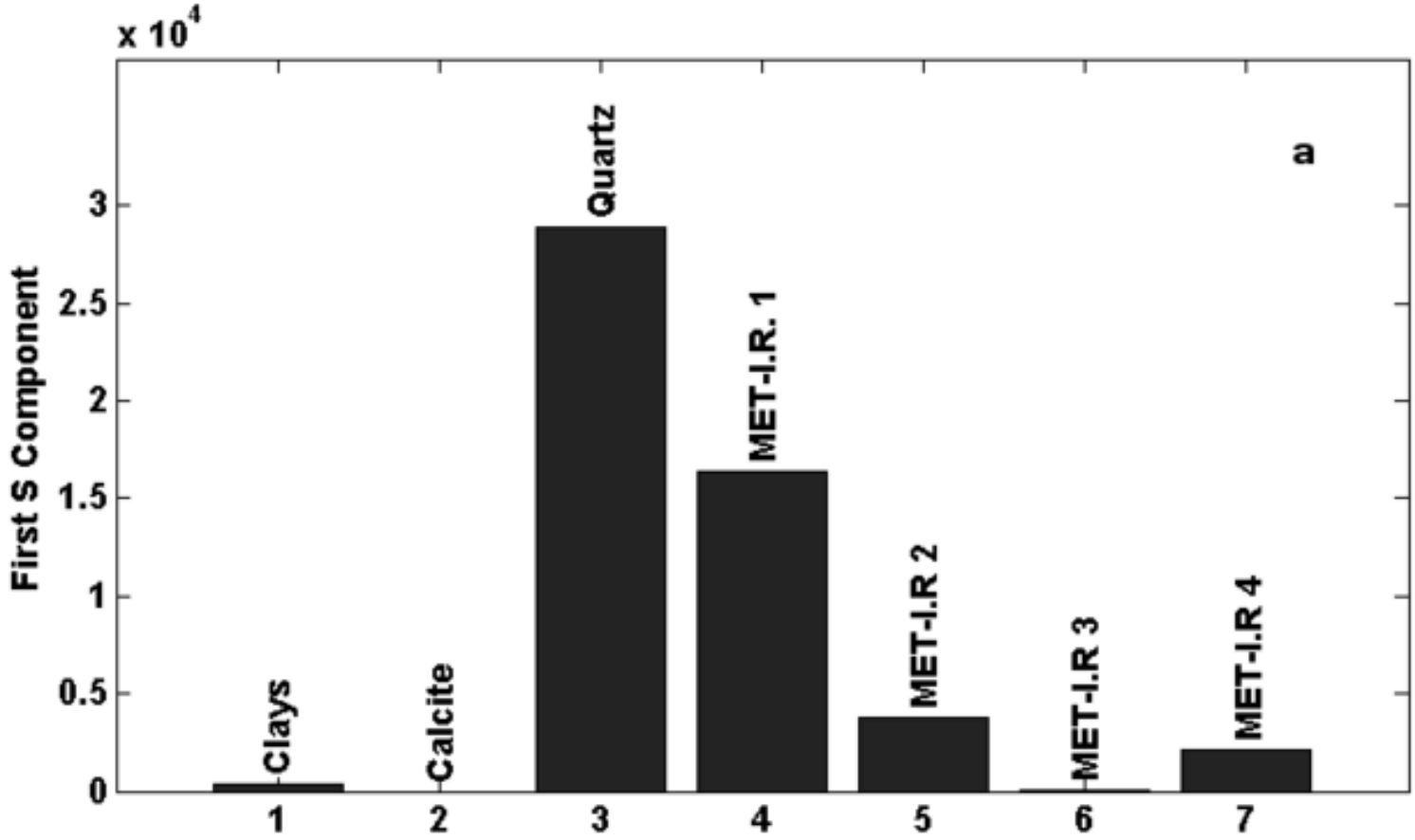

Figure 7b
Click here to download high resolution image

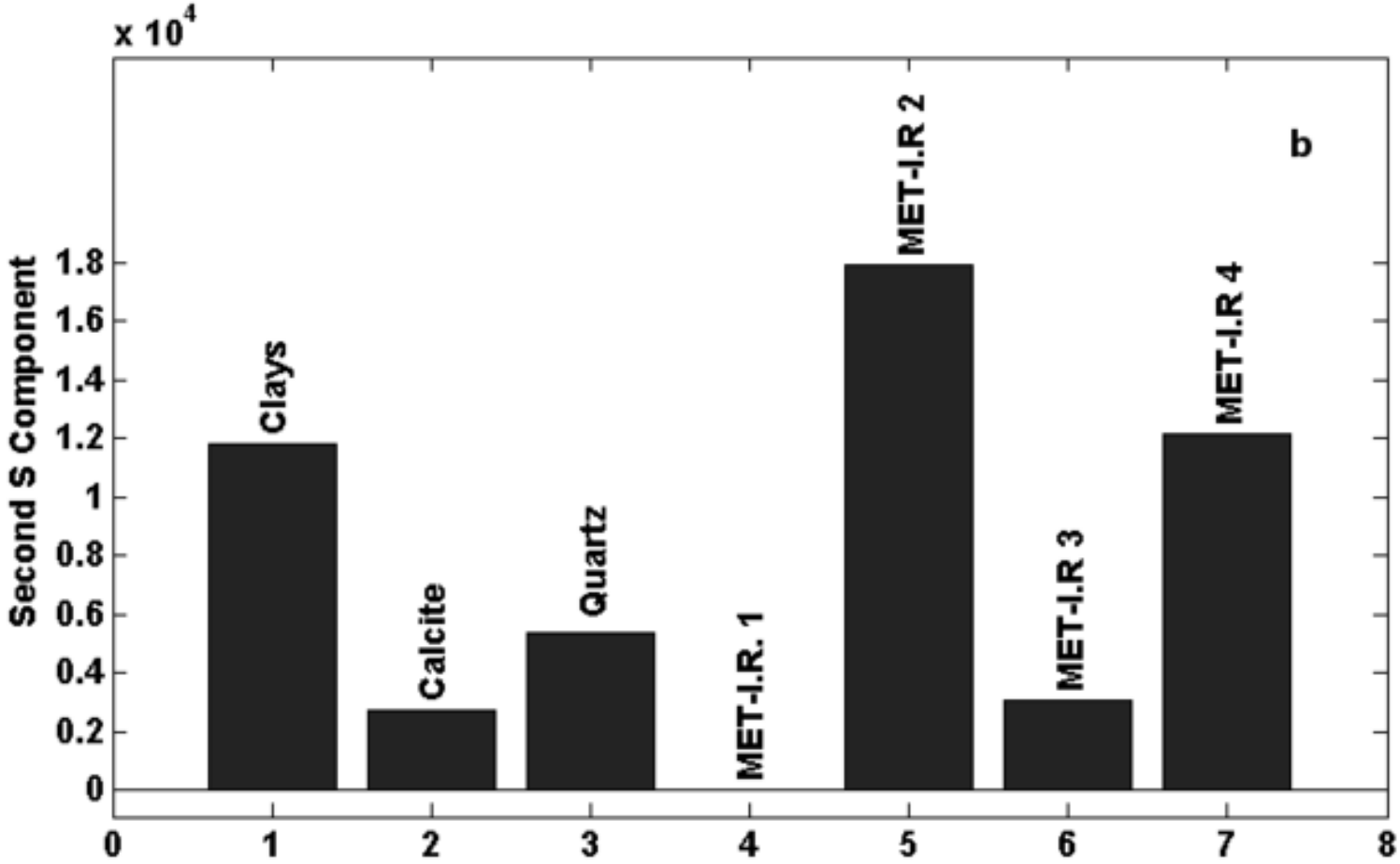

Figure 7c
Click here to download high resolution image

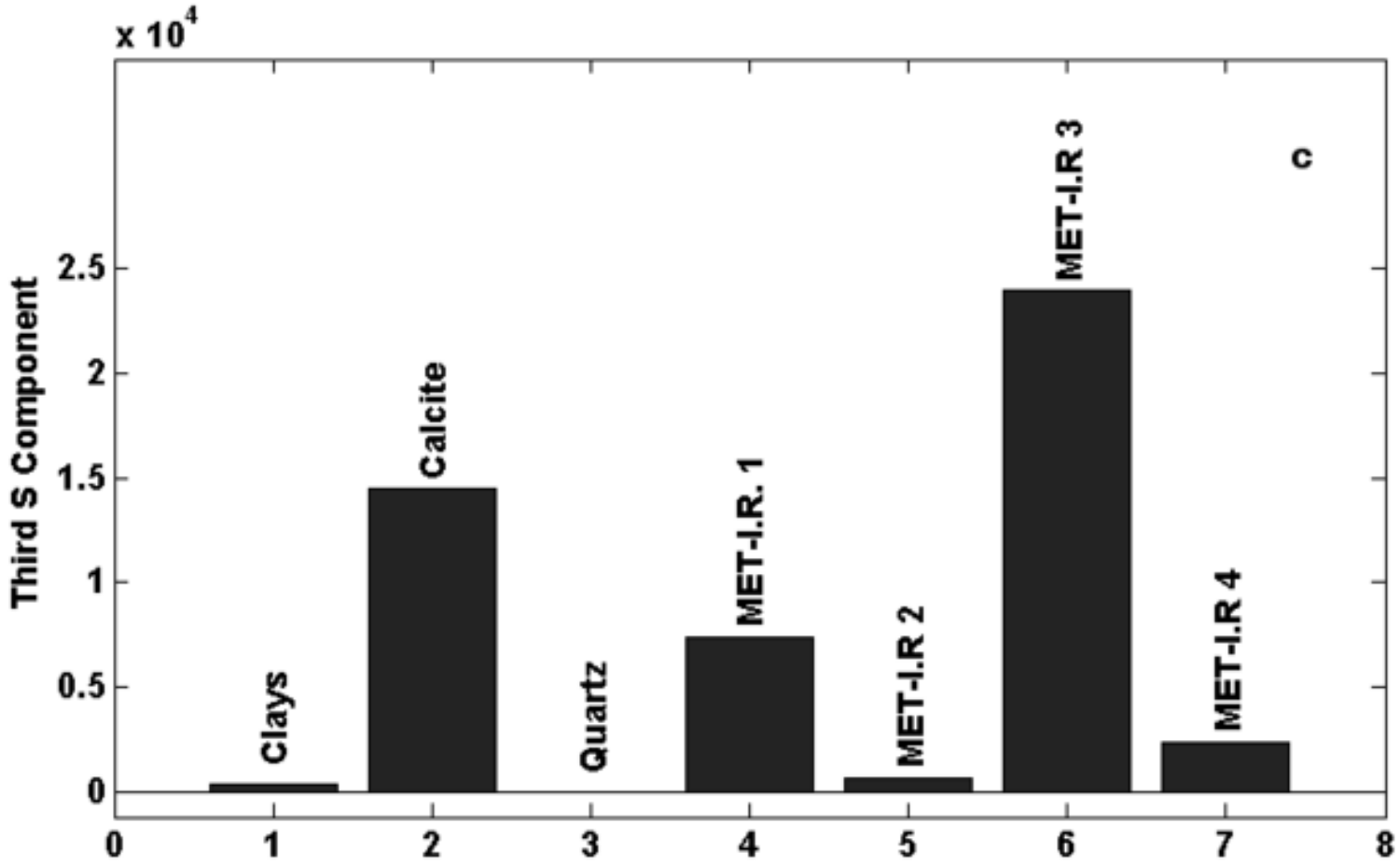

Figure 7d
Click here to download high resolution image

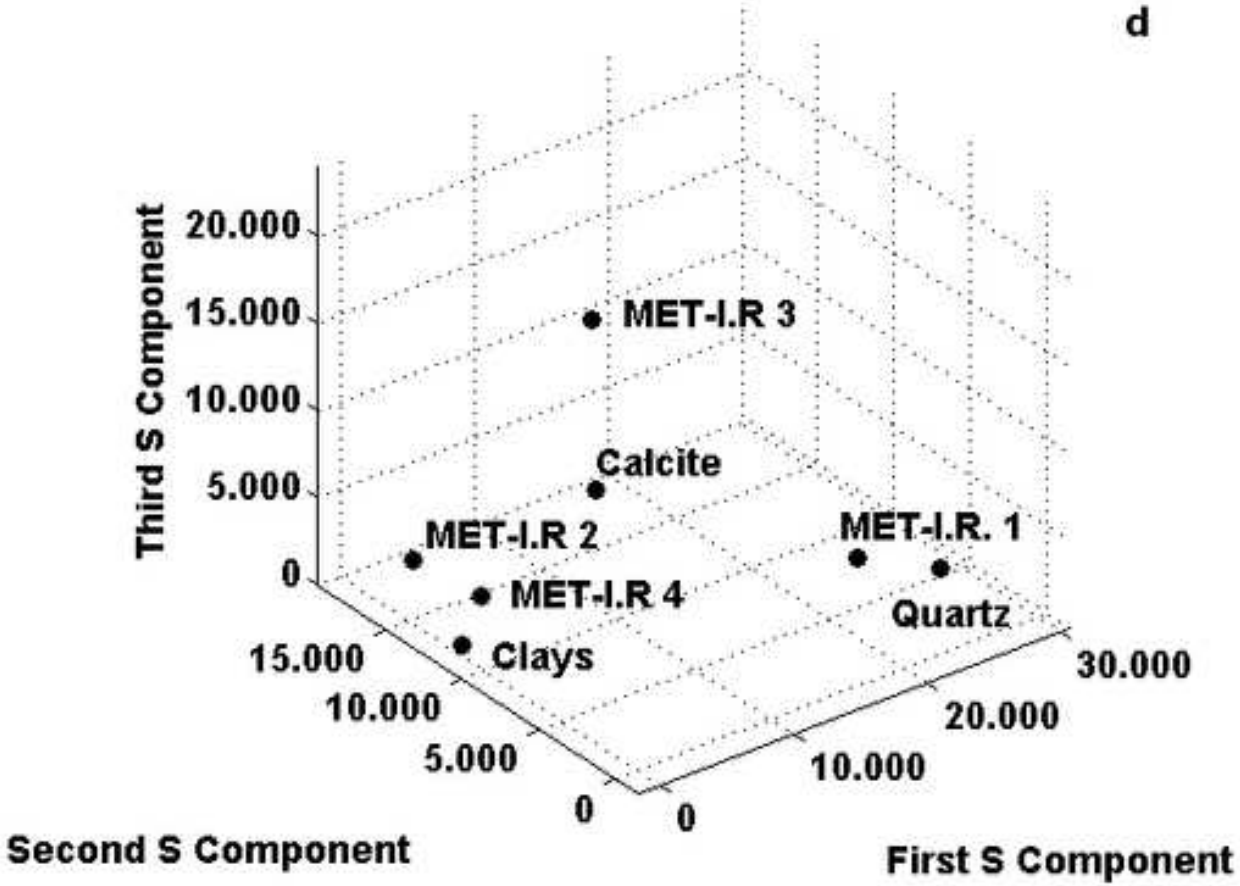

Figure 8a
Click here to download high resolution image

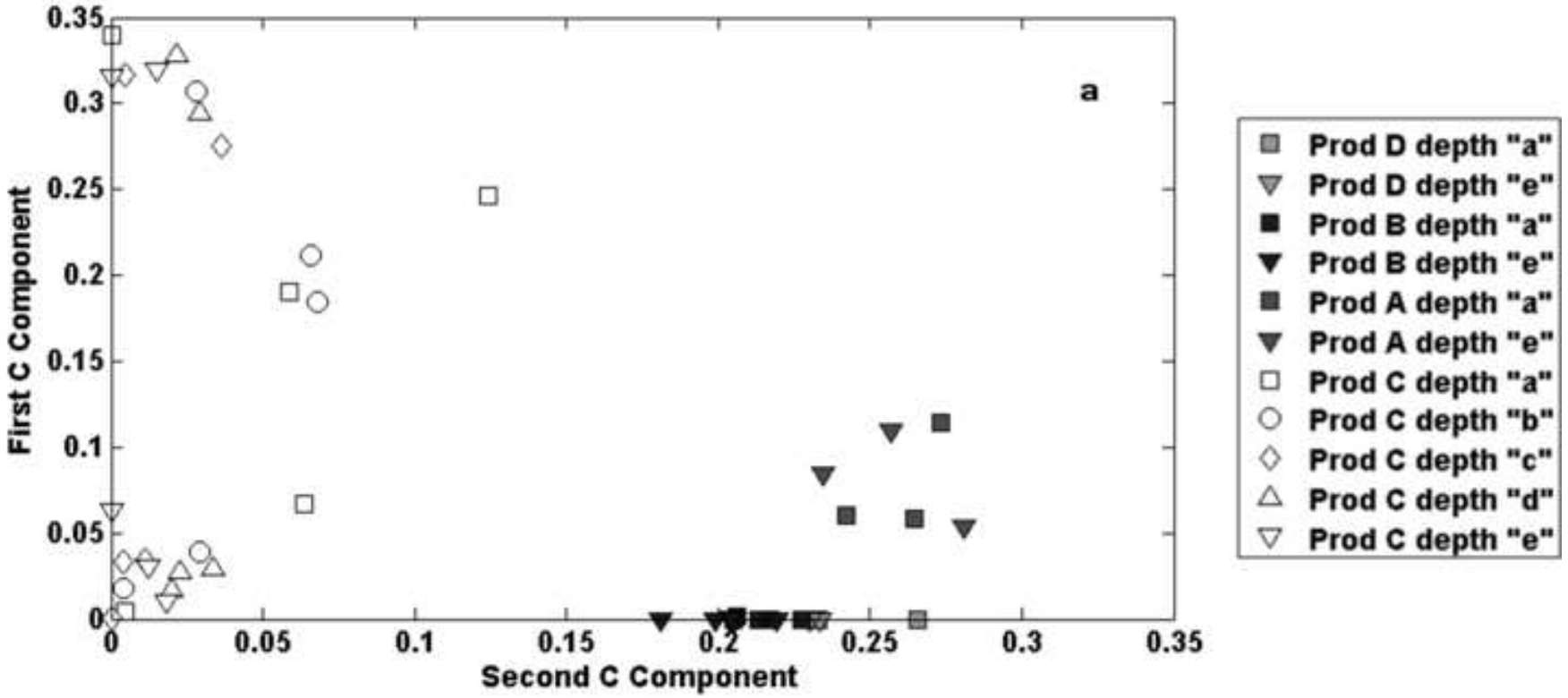

Figure 8b
Click here to download high resolution image

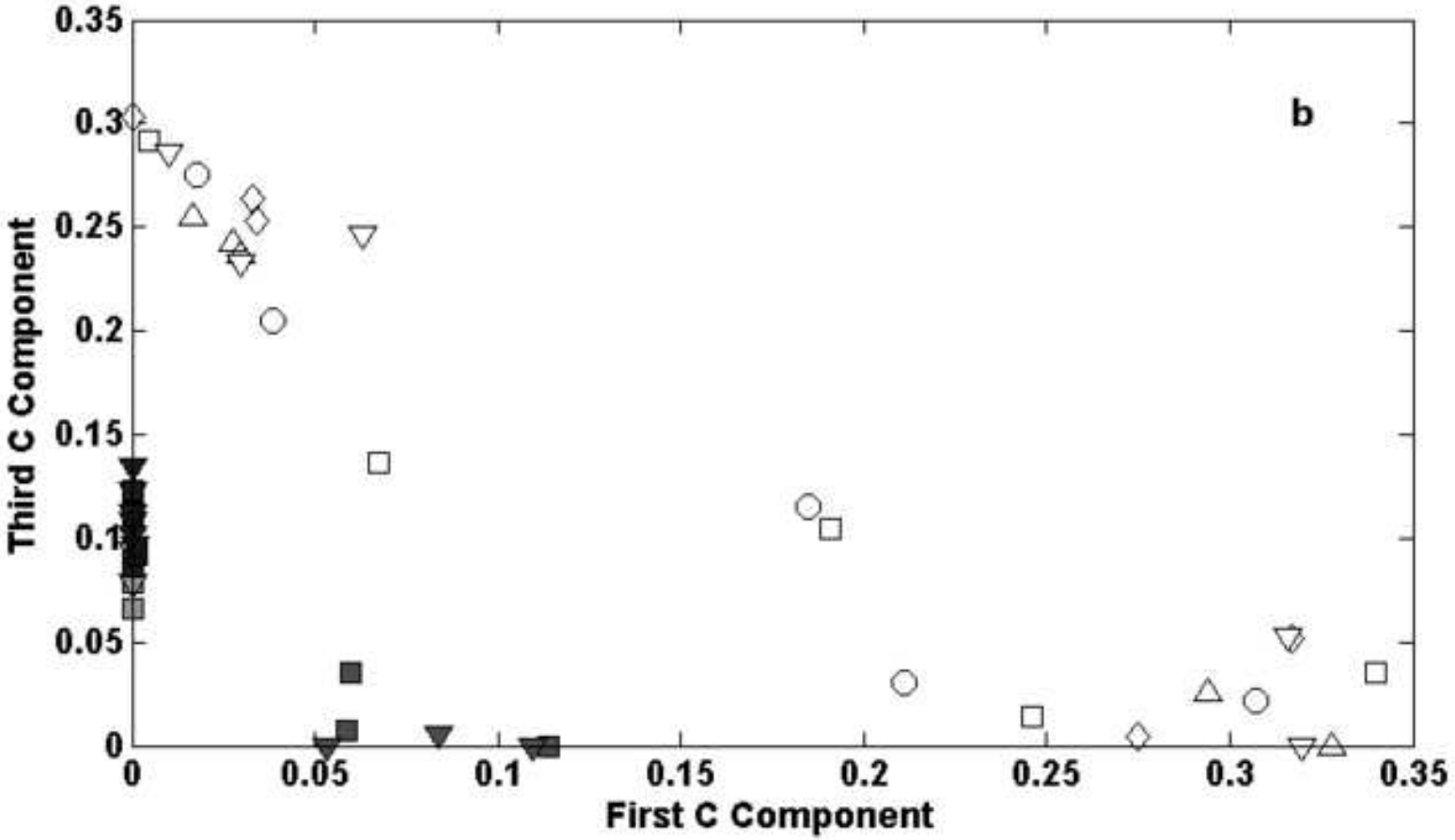
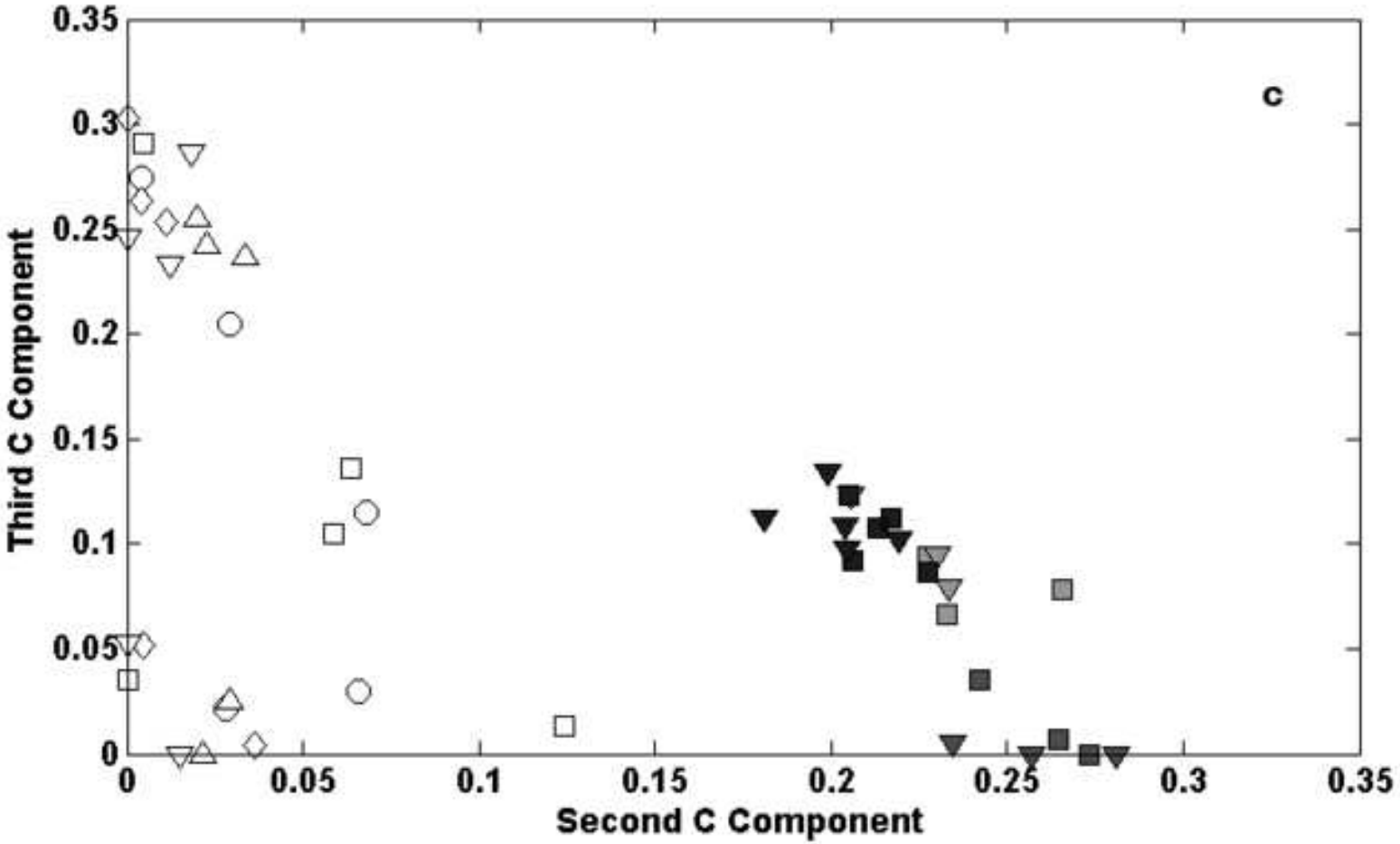
Figure 8c
Click here to download high resolution image