JOURNAL OF CHEMOMETRICS
J. Chemometrics 2000; **14**: 197–211

From experimental design to images to particle size histograms to multiway analysis. An example of peat dewatering

Paul Geladi[1]*, Helén Bergner[2] and Lena Ringqvist[2]

[1]*Department of Chemistry, Umeå University, S-901 87 Umeå, Sweden*
[2]*Department of Agricultural Research for Northern Sweden, SLU Röbäcksdalen, Box 4097, S-90403 Umeå, Sweden*

SUMMARY

The efficiency of peat dewatering by filtering slurries is dependent on the sizes of fine and colloidal particles that clog the filter. A designed experiment was carried out to check the use of different treatments on particle coagulation. The resulting particle sizes were studied under the microscope by automated digital image analysis, leading to area histograms for 21 size classes. Seven treatments on five peat types give a two-way ANOVA in all-qualitative variables, but the 21 response variables are a bit too much for an ANOVA or MANOVA analysis. The data can also be arranged in a 5 (peat types) $\times$ 7 (treatments) $\times$ 21 (size classes) three-way array. This array is analyzed by PARAFAC and gives an effective three-way rank of 4. The three-way data have no obvious underlying trilinear structure, and curve resolution results are not expected. The three-way analysis gives a very parsimonious model that is easily interpreted as a function of the problem definition. The emphasis is on visualization of the results. Copyright © 2000 John Wiley & Sons, Ltd.

KEY WORDS: PARAFAC; three-way analysis; image analysis; peat dewatering; digital image analysis; ANOVA; pretreatment for coagulation

INTRODUCTION

Dried peat is mainly used as a growing medium and a biofuel, but it is also used for environmental protection and as a precursor of natural chemicals [1]. Peat is found in nature in peat bogs. These are formed continuously by dying vegetation on the surface. In a peat bog the water content in the material is approximately 95%. For industrial uses this water augments the cost of storage and transportation and it must be removed. Conventional peat production systems, using drying in open fields, have drawbacks such as short production seasons and weather dependence. For these reasons, research and development projects on new peat production systems involving mechanical dewatering were set up during the late 1980s and early 1990s.

Drying peat is very difficult, as pressing out the water changes the structure of the material, and hence only a minor fraction of all water is removed by pressure. A possible solution to the problem is to make a slurry by adding water. The slurry can be filtered and gives a reasonably dry residue on the filter, but fine particles clog the filters and slow down the filtration process, making it very impractical. These particles are mainly colloidal in nature, and properties such as negative surface

* Correspondence to: P. Geladi, Department of Chemistry, Umeå University, S-901 87 Umeå, Sweden.

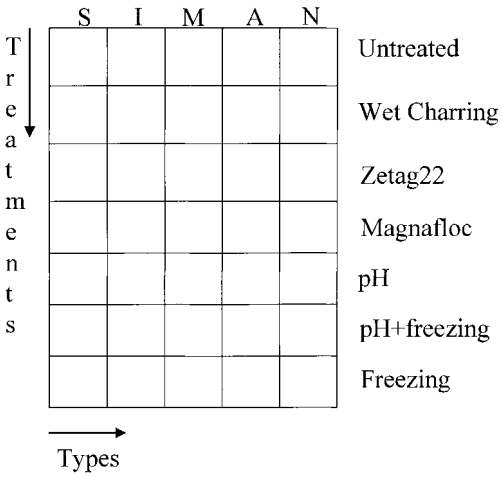

Figure 1. The particle coagulation experiment is a two-way ANOVA layout with five qualitative peat types and seven qualitative treatments. The response variables are 21 area classes, so ANOVA or MANOVA interpretations are not very easy to make.

charge are very important. The final decomposition products of peat are humic acids. Therefore the decomposition is called humification. Because of the humic acids, colloidal particles in a peat slurry carry a negative charge, so H^+, Al^{3+} or positively charged polyelectrolytes can be used to induce the aggregation of small particles into larger ones.

The data described here come from experiments in pretreatment of peat supposed to give minimal amounts of fine particles in the slurry. The ideal study would use an experimental design, but because of the qualitative nature of the factors used, this design becomes an ANOVA [2]. The experiment is a two-way ANOVA layout, with five peat types of industrial use and seven treatments as qualitative factors. The layout is given in Figure 1. The response is the particle size distribution in the slurry after removal of the coarse fraction. This particle size distribution is measured under an optical microscope. Automated image analysis of 1500 particles for each cell in the ANOVA layout gives a histogram of particle area in 21 size classes. Some histograms are given in Figure 2. For some cells in the ANOVA layout (Figure 1), filtration experiments were also carried out, but for reasons of time saving, the filtration data are not complete.

Although the experiment is an ANOVA two-way layout, it is not easy to analyse it in this way because of the many correlated responses in the particle size distribution. These could be reduced to a few histogram statistics which could be treated by MANOVA [3], but this does not seem to be the ideal way to go. The data can also be analyzed as a two-way 35 $(5 \times 7) \times 21$ data matrix. Principal component analysis (PCA) would be useful on such a data matrix. It was found that this does not give optimal information. The effective rank to be used is very hard to find, and the mixing of the types and treatments in the scores gives confusing plots.

The data can be arranged in a three-way array of $5 \times 7 \times 21$, i.e. 5 peat types $\times$ 7 treatments $\times$ 21 area intervals. This array is shown schematically in Figure 3. It was found that three-way analysis is best suited for this type of data. It is shown that three-way analysis by PARAFAC/CANDECOMP [4–6] (from here on simplified as PARAFAC) is an improvement over two-way data analysis techniques. The data are not really trilinear and there are no 'pure' particle area histograms, so no curve resolution results should be expected. The emphasis is not on equations or algorithms, but on the visual presentation of three-way results.

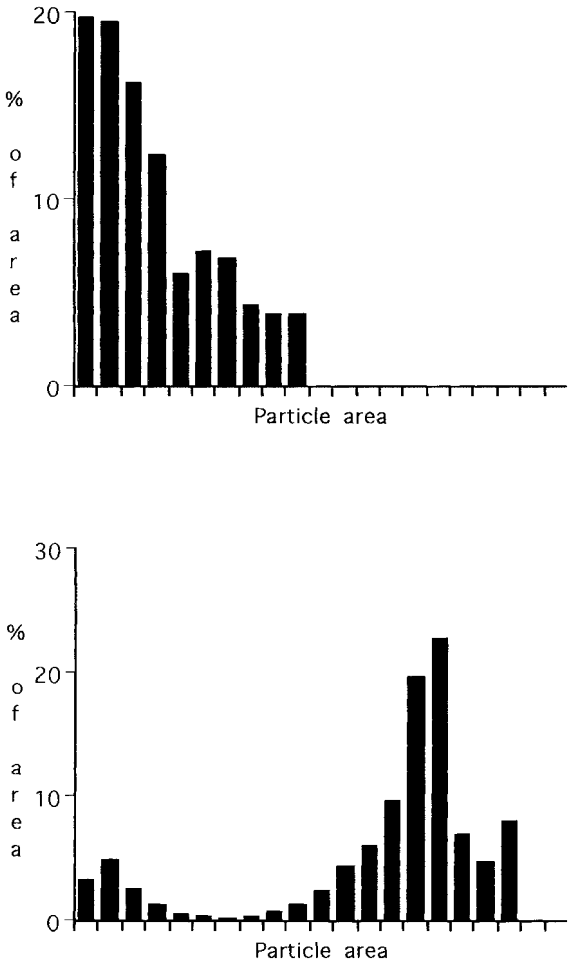

Figure 2. Particle area histograms of a slurry with many fine particles (top) and one with mainly medium-size particles (bottom).

EXPERIMENTAL

The five peat types form a small design with three Sphagnum peats and two Carex peats. Types A and S are of low humification and types M, I and N are of high humification. The description is given in Table I. The seven treatments are: U, untreated; C, mild wet carbonization or charring; Z22, Zetag 22 (a cationic polyelectrolyte); M+A, Magnafloc + $AlCl_3$ (Magnafloc is an anionic polyelectrolyte); PH, pH lowering to 2·5; PF, pH lowering combined with freezing; FR, freezing only.

Amounts of 25–50 g of peat were mixed with 300 ml of water and left on a shaker for a period of 16 h. The liquid with particles passing a 0·045 mm sieve was collected and treated. All seven treatments were applied to the sieved slurry [7]. Mild wet carbonization (charring) was done in an autoclave at 140 °C for 60 min. The cationic polyelectrolyte was Zetag 22 (polymethacrylate, charge 100%) from Allied Colloids (0·5% of the dry peat mass). The anionic polyelectrolyte was Magnafloc E10 (polyacrylate, charge 5%) from Allied Colloids (0·5% of the dry peat mass), used in combination

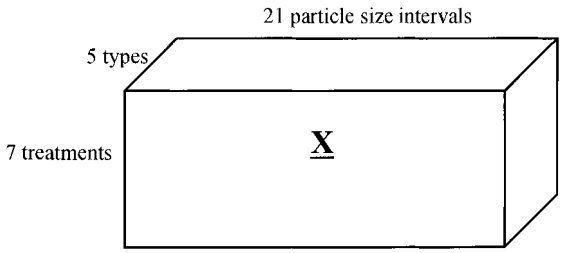

Figure 3. The three-way data array of $5 \times 7 \times 21$. The array is 5 peat types $\times$ 7 treatments $\times$ 21 area intervals.

with $AlCl_3$ (2% of the dry peat mass). Hydrochloric acid (1 M) was used to decrease the pH to 2·5. Freezing was done by keeping the slurry at -20 °C overnight.

The particle sizes were measured by drying a few drops of the liquid containing the colloidal and suspended particles on a glass microscope plate. The microscope used was a Zeiss Universal with 3·2× and 10× objectives and xenon arc lamp illumination filtered to give blue light. The microscope had a computer-programmable scanning stage. The image measurement was done with a Grundig SN76 black-and-white tube TV camera and a Kontron MIAP digitizer in images of 512×512 pixels and 256 grey levels. Automated image analysis was done using the program IBAS 2·0 from Kontron working in DOS. Some particles are shown in Figure 4. The description of the image analysis sequence is given in Figure 5. The digitized image in 256 grey levels is background-corrected to remove uneven illumination, and the contours of the particles are enhanced. Then a binary image is created with a black background and white particles. From this image the scrap (too fine particles) is removed and the objects are cleaned by binary operations such as erosion, dilation and hole filling. Then the binary objects are detected and their area is measured and collected in the histogram.

The background image compensates for unevenness in optics and illumination. It was saved in a file before each measurement series and used for background correction. This can be done by taking an image without a specimen present and blurring it by spatial averaging. The threshold for segmentation was set manually before each measurement series. The histograms were not constructed from one frame, but were the results from many frames in a rectangular scanning pattern. Images were not saved. They were discarded as soon as their contribution to the accumulated histogram had been calculated. This means that images such as those in Figure 4 cannot be directly compared to histograms such as those in Figure 2.

Particle size ranges in μm^2 are given in Table II. Method limitations are more frequently noticed at both ends of the particle size range. The smallest particles (size interval 1) suffer most from artefacts in the digital treatment. The largest particles have large counting statistics errors. Intervals 2–16 are the most reliable ones in Table II. Most histograms used are averages of two or more individual

Table I. Peat types

Sample code	Type	Humification
S	Sphagnum	Low
I	Sphagnum	High
M	Sphagnum	High
A	Carex	Low
N	Carex	High

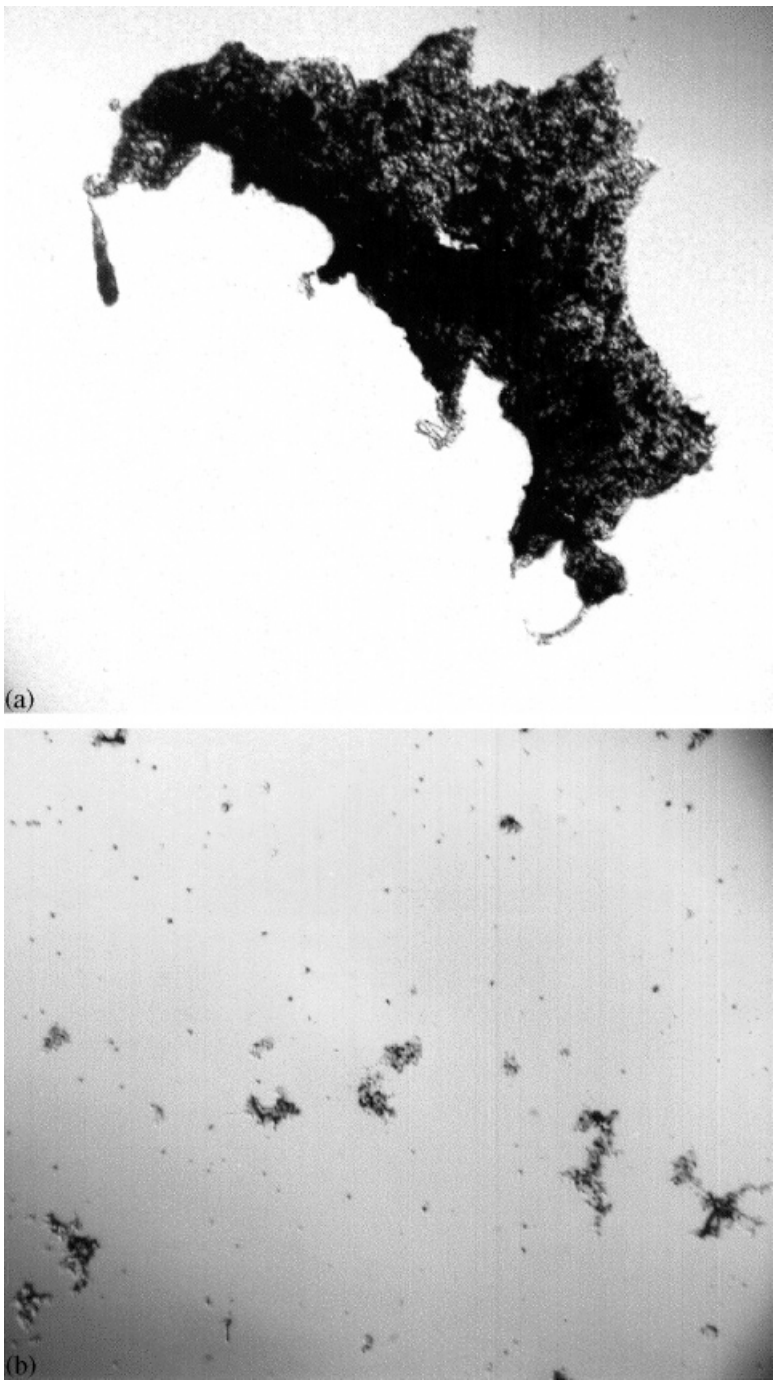

Figure 4. Some of the particles studied. The histograms were made by studying many frames like these. (a) A large coagulated particle. (b) Small particles.

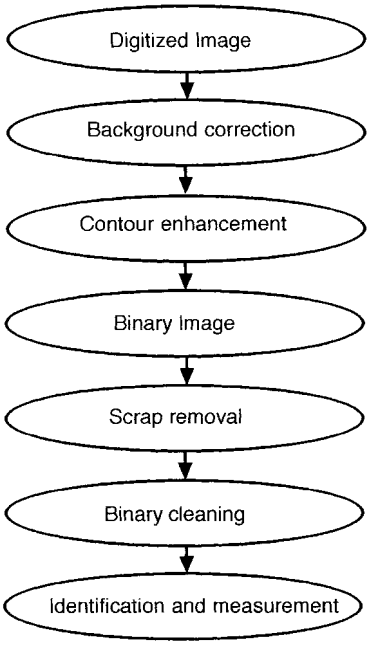

Figure 5. The operations sequence of the image analysis.

histograms, with some obvious outliers removed. The data were used as they were obtained, without detailed outlier detection or filtering. The histograms each contain the results for 1500 particles.

The PARAFAC calculations were done with the PLS toolkit in MATLAB from Eigenvector Research [8]. All other calculations were done in MATLAB. Some plots were made using Clarisworks.

THREE-WAY DATA ARRAYS AND ANALYSIS

The data form a three-way array of $5 \times 7 \times 21$, called $\underline{\mathbf{X}}$. The array is peat type $\times$ treatment $\times$ particle area. A graphical representation of the three-way array is given in Figure 3. The data are used without centering or scaling. A three-way array can be scaled and centered in many ways, but if no good reason based on a physical model exists, errors and confusion often result. Mean centering was found not to improve the interpretation. The scales of the histograms (particle size in μm^2) are all in the same units, and therefore scaling by the variable-wise standard deviation was not considered necessary.

The PARAFAC three-way decomposition [4–6 9 10] is given by

$$x_{ijk} = a_{i1}b_{j1}c_{k1} + a_{i2}b_{j2}c_{k2} + \ldots + e_{ijk} \tag{1}$$

or alternatively by

$$x_{ijk} = l_1 a_{i1}b_{j1}c_{k1} + l_2 a_{i2}b_{j2}c_{k2} + \ldots + e_{ijk} \tag{2}$$

where

Table II. Particle size ranges

Number	Size range (μm^2)	Remarks
1	5–8	Method artefacts
2	8–13	
3	13–22	
4	22–37	
5	37–62	
6	62–100	
7	100–170	
8	170–290	
9	290–490	
10	490–810	
11	810–1400	
12	1400–2300	
13	2300–3800	
14	3800–6400	
15	6400–10700	
16	10700–18000	
17	18000–30000	Counting
18	30000–50000	statistics gives
19	50000–84000	noise for
20	84000–140000	17–20
21	140000 and up	Accumulator for large

x_{ijk} an element of the three-way array $\underline{\mathbf{X}}$
i the index of the A loadings
j the index of the B loadings
k the index of the C loadings
a_{i1}, a_{i2} elements of the first and second A loadings
b_{j1}, b_{j2} elements of the first and second B loadings
c_{k1}, c_{k2} elements of the first and second C loadings
l_1, l_2 component 'sizes' when the A, B and C loadings are normalized to length 1
e_{ijk} the residual

The A loadings are peat type, the B loadings are treatment and the C loadings are particle size. The number of loading vectors in each mode is the effective rank of the model. This effective rank is chosen by consideration of noise and instability of the solution. The decomposition is chosen such that the loadings contain meaningful structure and the residual contains noise. Principal component analysis (PCA) of the 35×21 two-way array was done on raw data and on mean-centered and scaled data. The eigenvalues, or singular values, did not really give a sharp cut-off, but 75%–85% of the sum of squares explained with four or five components is reasonable for this type of data.

EFFECTIVE THREE-WAY RANK

The results of the PARAFAC decomposition are shown in Figures 6 and 7. Figure 6 shows the percentage explained by a PARAFAC model as the complexity of the model (effective rank) increases. For the PARAFAC analysis there is no additivity as for the eigenvalues of the PCA. From Figure 6 it is very diffcult to see what effective rank should be used. The percentage explained increases with rank, but each model may have totally different components. The only observation to be made is how much the residual decreases as a function of rank. With the specific background

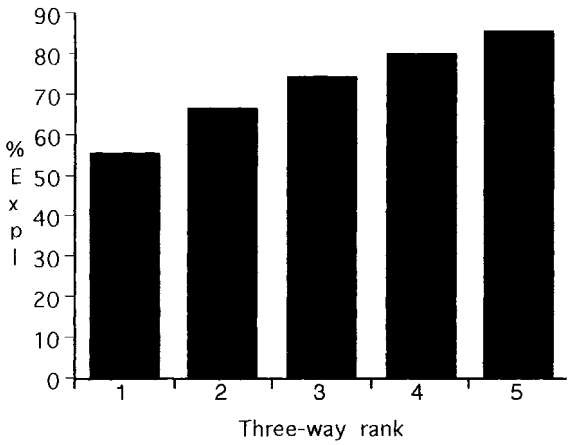

Figure 6. Percentage of the total sum of squares explained for PARAFAC models of increasing effective rank.

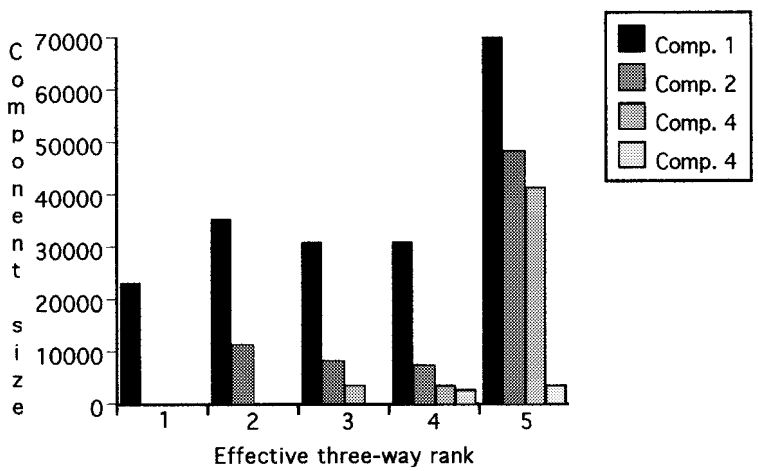

Figure 7. The bar plot shows the stability of the components as the effective rank increases. The rank 5 model is inconsistent with the models of lower rank. Only up to four components are shown.

information of the expected noise in the data, one should be pleased with a model that explains 75%–85% of the total sum of squares. Figure 7 shows the 'sizes' of the PARAFAC components for models of effective rank 1–5. The rank 1 model is just included for completeness. It was known in advance to be insufficient. The rank 2–4 models seem to agree quite well, and then the rank 5 model is quite deviating. The rank 4 model is selected for further use. With the effective rank selected, loadings can be studied.

LOADINGS AS LINE AND SCORE PLOTS

The loadings are shown as bar plots in Figures 8 and 9. It may be observed that loading 1 is mainly a small- and intermediate-particle loading. Loading 2 is a contrast between small and intermediate

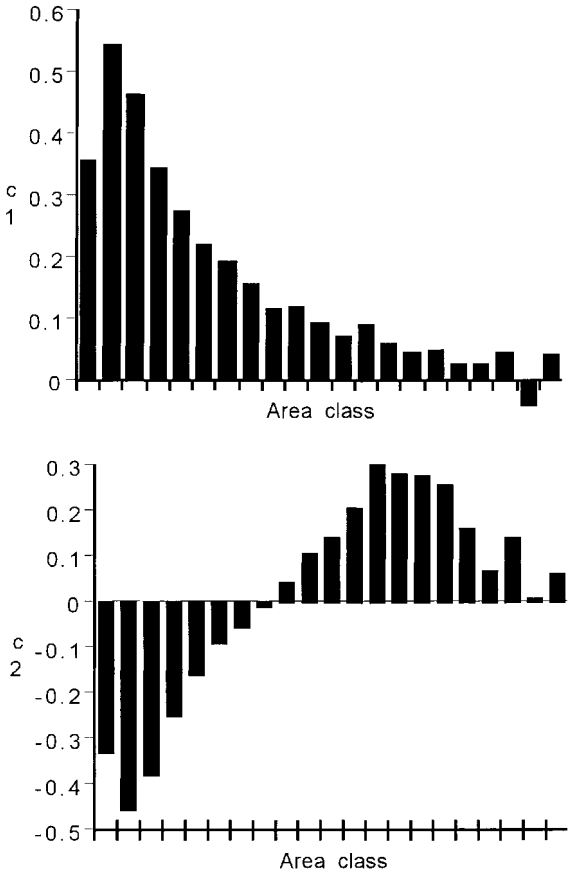

Figure 8. The first and second C loadings as bar plots.

particles. Loadings 3 and 4 are mainly related to large particles. Loading 3 is mainly influenced by class 21. Loading 4 has a high negative value for class 20. Because of the noise in the large-particle classes, there is no point in giving a too detailed interpretation.

Scatter plots are used to plot a number of loadings against each other. They can be made in 2D or 3D. When more than three loadings have to be shown, more plots can be arranged in one figure. In two-way analysis, orthogonal loadings (or scores) can be plotted very efficiently in this manner to give an overview of the reduced data space. Multiway analysis by PARAFAC does not provide orthogonal loadings by definition, but still the principle of scatter plots can be made useful. The loading plots are given in Figures 10–12. The loadings are arranged in pairs 1–2 and 3–4.

With scatter plots one has to be careful in the interpretation. This is already true in two-way analysis and becomes even more complicated in three-way analysis. The plots (score plots, loading plots) are usually made to fill a square or rectangle no matter what the size of the used scores or loadings is. It is very important that the percentage of the total sum of squares (%SS) is displayed to avoid misinterpretations. Another complication with loading plots is that loadings from three-way analysis are not always orthogonal and can even be very correlated. The angles between loading vectors are given in Table III. The angle between $\mathbf{a}_1$ and $\mathbf{a}_2$ is only 7 °, so these components are almost parallel. For other vector pairs the angles are closer to orthogonal ($\mathbf{b}_3$, $\mathbf{b}_4$ and $\mathbf{c}_2$, $\mathbf{c}_4$), but in most cases something intermediate is obtained for the angle.

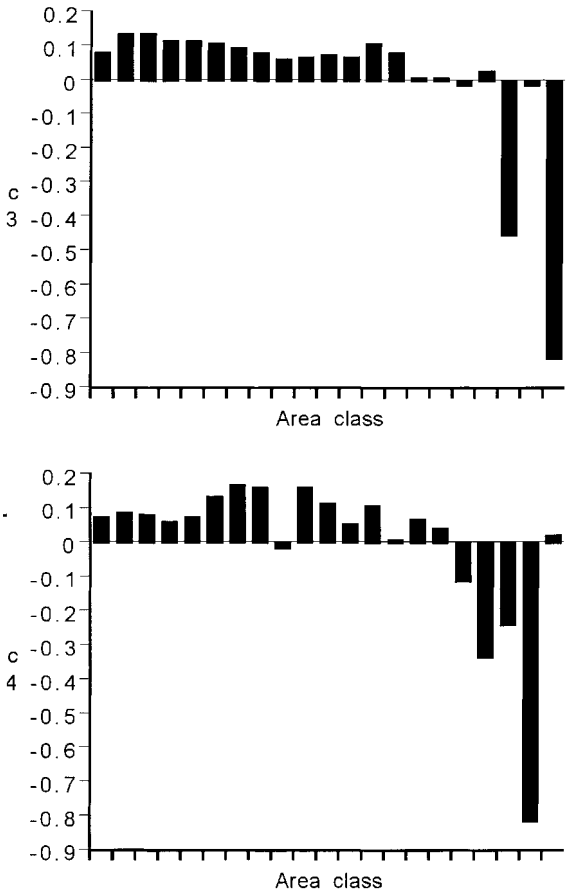

Figure 9. The third and fourth C loadings as bar plots.

An important point is to check whether the origin is in the plot or not. In the $\mathbf{a}_1$–$\mathbf{a}_2$ plot (in Figure 10) the origin is very far outside the plot, and this explains how the angle between the vectors can be only 7 °. There is also a sign indeterminacy that gives mirror images leading to false interpretations. This can be shown by repeating Equation (1):

$$x_{ijk} = a_{i1}(-b_{j1})(-c_{k1}) + a_{i2}b_{j2}c_{k2} + \ldots + e_{ijk} \tag{3}$$

This shows that any pair of loadings can be reflected around the origin (mirrored) without changing the model or fit. The only solution to this type of indeterminacy is having background knowledge and using it wisely to introduce some meaningful mirrorings. It was decided that mirroring the third and fourth A and C loadings would give a better match with the third and fourth B loadings.

The A loadings show a grouping in peats of high (I, M, N) and low (A, S) humification in the first and third components. There is no grouping in Carex and Sphagnum peats. The $\mathbf{a}_1$–$\mathbf{a}_2$ loading plot is special. The angle between the vectors is only 7 ° (see Table III) and also the origin is very far outside the plot. Two-way analysis showed little importance for the differences between peat types, and a Tucker decomposition needed fewer loadings in the A mode. Also this indicates that peat type is not a very important factor.

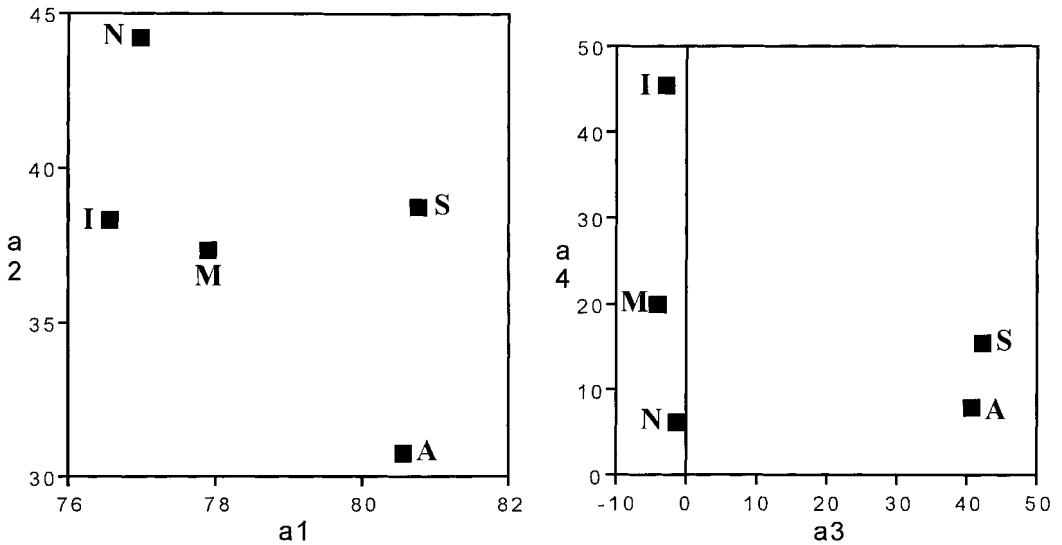

Figure 10. Loading plots for the rank 4 PARAFAC: A loadings 1–2 and 3–4. The 3–4 loading plot was mirrored after calculation.

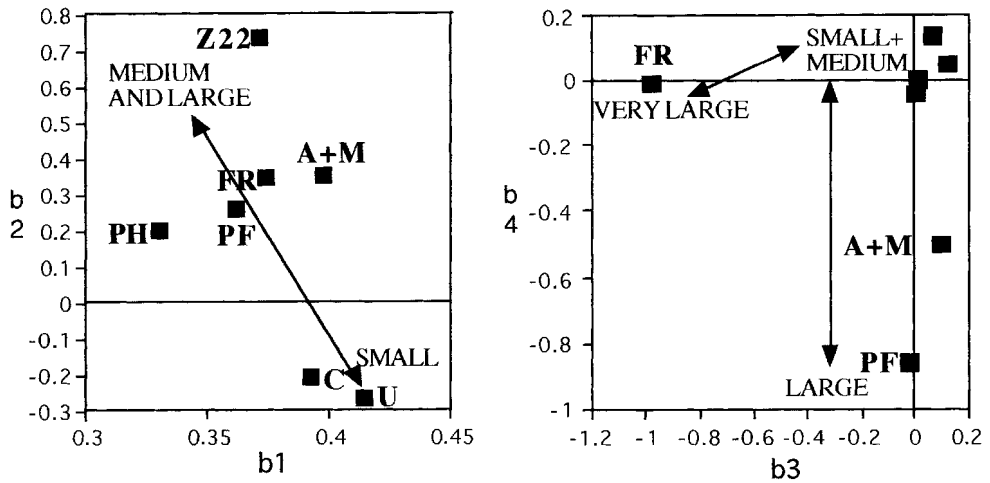

Figure 11. Loading plots for the rank 4 PARAFAC: B loadings 1–2 and 3–4.

The B loadings show important differences between the treatments, and these are related to the particle size classes in the C loadings. A full interpretation of the B and C loadings is given below.

THE DENDROGRAM

No matter how useful scatter plots of loadings or scores are, when the effective rank of the model becomes high, there are quite a number of them to be studied. In these cases it may be more useful to

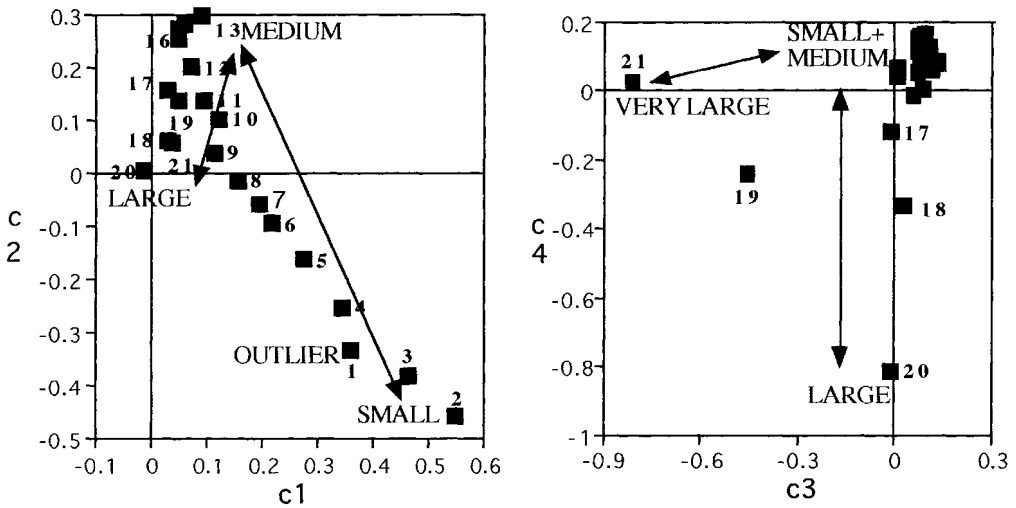

Figure 12. Loading plots for the rank 4 PARAFAC: C loadings 1–2 and 3–4. The 3–4 loading plot was mirrored after calculation.

use dendrograms to get an overview. Dendrograms are based on some distance-based classification in multivariate space. They give a good overview of the structure of clusters and outliers in the many dimensions used. It is a good idea to make the dendrogram on loadings (scores) instead of on the original variables. This saves calculation time and separates out the residual that is assumed to contain mainly noise. One complication with clustering is that there are many clustering methods and ways of preprocessing the data that may supply more confusion than help.

For the four B loadings of the peat example the distance-based classification was carried out after mean centering and using the nearest-neighbour method [11,12]. The results are shown in Figure 13. The figure shows clearly that FR is very different from the other treatments and that U and C are clustered close together. The loading plots (Figures 10–12) may make more sense when compared to the dendrogram. The use of *K*-means clustering gave very similar results.

INTERPRETATION OF THE LOADING PLOTS

The interpretation of the A loadings was given earlier. Figures 11 and 12 show the B and C loadings for the four components as scatter plots, with interpretation. The first component is larger than the second one (sizes 31 000 and 7000). Components 3 and 4 are of about equal size (3400 and 2800). In the C loadings it can be seen that particle sizes are autocorrelated and that there is a gradient from small to large particles. Particle size class 1 does not follow the behavior of the others. This class was

Table III. Angles (deg) between loading vectors, calculated as $\theta = \cos^{-1}[\mathbf{a}_1{}^{\mathrm{T}}\mathbf{a}_2\,(\|\mathbf{a}_1\|\,\|\mathbf{a}_2\|)^{-1}]$

	$\mathbf{a}_2$	$\mathbf{a}_3$	$\mathbf{a}_4$		$\mathbf{b}_2$	$\mathbf{b}_3$	$\mathbf{b}_4$		$\mathbf{c}_2$	$\mathbf{c}_3$	$\mathbf{c}_4$
$\mathbf{a}_1$	7	125	143	$\mathbf{b}_1$	59	74	62	$\mathbf{c}_1$	125	74	74
$\mathbf{a}_2$		120	142	$\mathbf{b}_2$		77	71	$\mathbf{c}_2$		103	97
$\mathbf{a}_3$			76	$\mathbf{b}_3$			90	$\mathbf{c}_3$			78

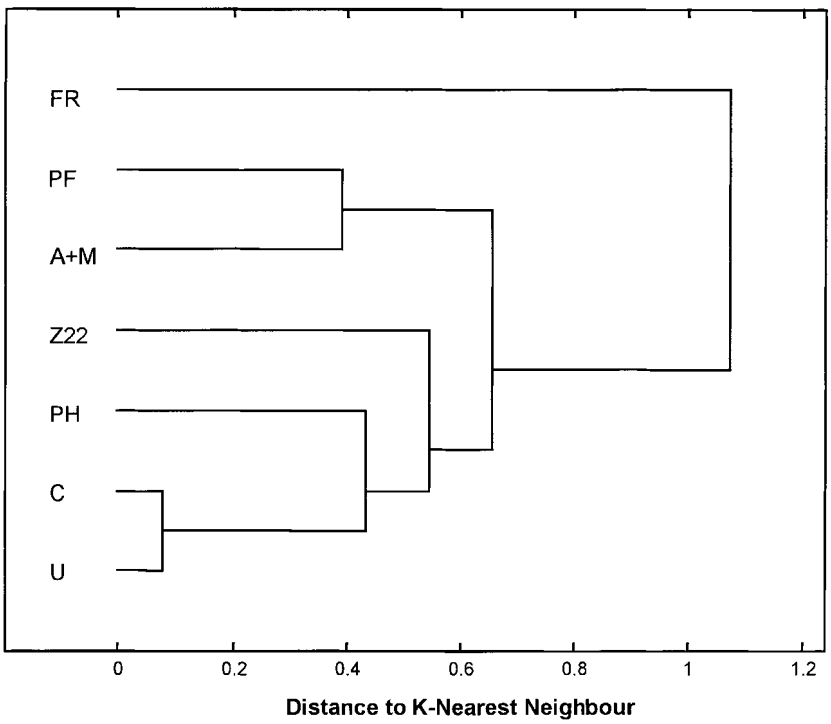

Figure 13. The dendrogram for clustering based on four B loadings. The similarities and dissimilarities of the treatments are clearly shown.

difficult to measure and was influenced most by the binary operations. Also the counting noise in the larger-particle classes 17–21 can be noticed in the C loadings. The B and C loadings considered together show that the two treatments U and C give mainly small particles. The C loadings show a grouping in small + intermediate particles and large particles (bins 17–21) as outliers. The B loadings $\mathbf{b}_3$ and $\mathbf{b}_4$ show a grouping of treatments U, C, PH and Z22. The group of treatments corresponds to small and intermediate particles. The main conclusion is that freezing gives the largest particle aggregates, and obviously freezing combined with pH lowering also gives large aggregates. The $AlCl_3$ + Magnafloc treatment seems to be the next best treatment. The Zetag 22 treatment gives medium-size aggregates.

COMPARISON TO FILTRATION EXPERIMENTS

For some of the treatments, slurry filtration experiments were also carried out [7]. The response of these experiments is the filtration capacity (ml of slurry per second) in a standard filtration set-up. They agree very well with the findings from the B loadings. The most efficient pretreatment method obtained for peat dewatering was a combination of pH lowering and freezing, whereby an extremly high filtration capacity was achieved [7] (see PF in Figure 11). This method flocculated the fine-particle fraction to very large aggregates. Compared to freezing, addition of the cationic polyelectrolyte gave a lower filtration capacity. The aggregates obtained were also smaller. Addition of a cationic polyelectrolyte had a stronger positive influence on the dewatering properties compared to an anionic polyelectrolyte [13]. The flocculated particles differed in size: the anionic

polyelectrolyte gave both small and large aggregates, while the cationic polyelectrolyte gave medium-size aggregates (see Z22 and A+M in Figure 11). Lowering the pH in the peat samples caused an aggregation of particles, but the aggregates were small, leading to an increased filtration capacity compared to an untreated sample but rather low compared to polyelectrolyte addition [7]. Untreated and mild wet carbonized samples contained the finest particles, and a low filtration capacity was achieved [7]. The effect of carbonization is obtained when high pressure is applied on the samples. The filtration results seem to corroborate the particle size findings in the PARAFAC loadings.

CONCLUSION

The data are a 5×7 two-way ANOVA layout with 21 responses in each cell. This forms a $5 \times 7 \times 21$ three-way array, but underlying trilinear structure or curve resolution aspects are not present or are unknown. Three-way analysis gives a four-component PARAFAC model. The determination of effective rank is easier than for two-way analysis, and the model is more parsimonious. The histograms fit very well to a PARAFAC model as they are. The uncertainty of whether or not to scale or center the data, which occurs frequently in PCA modeling, did not occur here. The A loadings, representing the peat types, show a separation in degree of humification but not in origin (Sphagnum or Carex). Otherwise the peat type used is of less importance than the treatments, and this is confirmed by a Tucker3 analysis requiring fewer A loadings. The B and C loadings together give information on the similarity or dissimilarity of the treatments and on the relation of the treatments to particle sizes obtained. This is confirmed by a dendrogram giving combined information for the four B loadings at once. Having the treatments and peat types separately in their own loadings is a huge advantage over two-way methods. Freezing alone or combined with pH lowering is best at giving large particles. Also the polyelectrolytes give reasonable results. Wet charring in an autoclave gives almost no advantage over the untreated material. The findings agree very well with incomplete filtration experiments. Some of the treatments used in the lab cannot be used in an industrial situation because of the high cost of the polyelectrolytes (Zetag 22, Magnafloc), the long duration of the treatment (freezing) or environmental considerations (pH lowering).

The data could have been analyzed by three separate PCA models of two-way matrices with three different reorganizations: 21×35, 7×105 and 5×147. This would have been almost the same as a Tucker3 model. A simple Tucker3 model presented the same problem of choice of components as the choice of effective rank in a PCA model and gave a confusing core array. It was therefore abandoned. The PARAFAC results used here gave an easy calculation, an easy determination of effective rank and a useful interpretation. With some effort a rotated Tucker3 model would have given similar results.

REFERENCES

1. Lappalainen E (ed.). *Global Peat Resources*. International Peat Society: Jyväskylä, 1996.
2. Cochran W, Cox G. *Experimental Designs* (2nd edn). Wiley: New York, 1957.
3. Mardia K, Kent J, Bibby J. *Multivariate Analysis*. Academic Press: London, 1979.
4. Geladi P. Analysis of multi-way (multi-mode) data. *Chemometrics Intell. Lab. Syst.* 1989; **7**: 11–30.
5. Smilde A. Three-way analyses. Problems and prospects. *Chemometrics Intell. Lab. Syst.* 1992; **15**: 143–157.
6. Bro R. PARAFAC: tutorial and applications. *Chemometrics Intell. Lab. Syst.* 1997; **38**: 149–171.
7. Ringqvist L, Igsell P. Four pretreatment methods for mechanical dewatering of peat. *Int. Peat J.* Special Edition 1992; 213–224.
8. Wise B, Gallagher N. *PLS Toolbox 2·0 for Use with MATLAB*. Eigenvector Research: Manson, WA, 1998.

9. Law H, Snyder C, Hattie J, McDonald R (eds). *Research Methods for Multimode Data Analysis*. Praeger: New York, 1984.
10. Coppi R, Bolasco S (eds). *Multiway Data Analysis*. North-Holland: Amsterdam, 1989.
11. Sharaf M, Illman D, Kowalski B. *Chemometrics*. Wiley: New York, 1986.
12. Henrion R, Henrion G. *Multivariate Datenanalyse. Methodik und Anwendung in der Chemie und Verwandten Gebieten*. Springer: Berlin, 1995.
13. Ringqvist L, Igsell P. Dual polymer system in peat dewatering. *Energy Fuels* 1994; **8**: 953–959.

