Independent Component Analysis to Mass Spectra of Aluminium Sulphate

M. Heikkinen, A. Sarpola, H. Hellman, J. Rämö, and Y. Hiltunen

Abstract—Independent component analysis (ICA) is a computational method for finding underlying signals or components from multivariate statistical data. The ICA method has been successfully applied in many fields, e.g. in vision research, brain imaging, geological signals and telecommunications. In this paper, we apply the ICA method to an analysis of mass spectra of oligomeric species emerged from aluminium sulphate. Mass spectra are typically complex, because they are linear combinations of spectra from different types of oligomeric species. The results show that ICA can decomposite the spectral components for useful information. This information is essential in developing coagulation phases of water treatment processes.

Keywords—Independent component analysis, mass spectroscopy, water treatment, aluminium sulphate.

I. INTRODUCTION

BESIDES iron aluminium chemicals are widely used in water treatment technology in domestic as well as in industrial fields. A major application is coagulation in which, in addition to creating a positively charged hydroxide flock, charged oligomeric Al-species are generated in water aiming to neutralize surface charges of impurities like mineral particles, microbes and humic substance. This neutralization intentionally deteriorates repulsive electrical forces of the impurities aiming attractive van der Waals forces to take place and coagulation to occur. Formed precipitate is then removed by settling, flotation of filtration. In addition to the coagulation itself, the charged oligomers have a biological impact by stiftening cell membranes and cross-linking biomolecules.

Aluminium sulphate, alum, has been used for this purpose for decades. Later on, different kinds of polyaluminium

M. Heikkinen, is with the Department of Environmental Sciences, University of Kuopio, P.O. Box 1627, FIN–70211 Kuopio, Finland (telephone: +358 44 716 2334; e-mail: mikko.heikkinen@uku.fi).

A. Sarpola, is with the Water Resources and Environmental Engineering Laboratory, University of Oulu, P.O. Box 4000, FIN-90014 University of Oulu, Finland (phone: +358 44 378 4422; e-mail: arja.sarpola@oulu.fi).

H. Hellman, is with the Department of Chemistry, University of Oulu, P.O. Box 3000, FIN-90014 University of Oulu, Finland (phone: +358 44 066 0220, e-mail: heikki.hellman@oulu.fi).

J. Rämö, is with the Water Resources and Environmental Engineering Laboratory, University of Oulu, P.O. Box 4000, FIN-90014 University of Oulu, Finland (phone: +358 40 524 7462; e-mail: jaakko.ramo@oulu.fi).

Y. Hiltunen, is with the Department of Environmental Sciences, University of Kuopio, P.O. Box 1627, FIN–70211 Kuopio, Finland (phone: +358 40 732 0355; e-mail: yrjo.hiltunen@uku.fi).
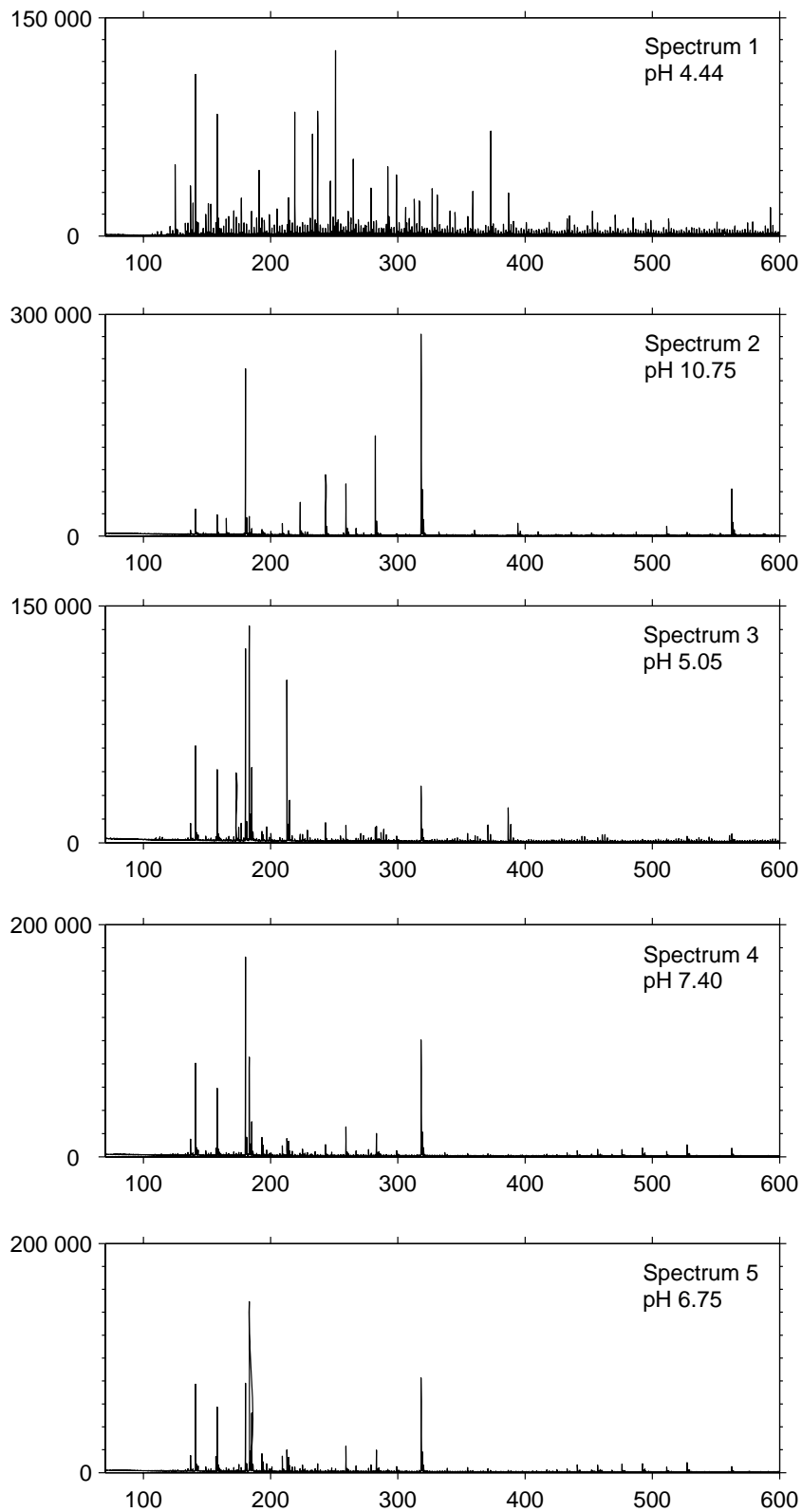
chemicals have occupied the market, but the alum is still overwhelming the most applied chemical due to its availability and performance. The dosage of the chemicals is still generally based on intuition or short scale testing rather than on scientific information. When thinking about immensity of global water treatment and material flows involved in it, it is clear that breeding oligomeric species and thus material efficiency of the aluminium chemicals must be managed much better than today. In addition to the waste of these chemicals, overdoses generally cause restabilization of the impurity particles driving them back to the solution.

The mass spectrometric (MS) method is a novel application for investigation of oligomeric species emerged from aluminium sulphate, where species are distinguished on the basis of their molecular mass to charge (m/z) ratios [1-2]. However, these kinds of mass spectra are typically complex, because they are linear combinations of spectra/signals originating from different types of oligomeric aluminium hydrolysis products, pH adjusting chemical and potential impurities. Therefore, it would be useful to decompose the spectral components and then analyze these composite spectra. However, both the spectral components and their respective amounts are unknown, resulting in a mathematically difficult blind source problem.

The independent component analysis (ICA) is a newly developed statistical method that has proven to be a successful technique in signal processing [3]. The ICA method has been applied in many fields, e.g. brain imaging [4, 5], vision research [6, 7], telecommunications [8] and financial research [9, 10]. ICA is a method for finding underlying factors or components from multivariate (multidimensional) statistical data. What distinguishes ICA from other methods is that it looks for components that are both statistically independent, and non-Gaussian [3].

The complexity of spectral information can be approached by assuming that spectra from oligomeric species are statistically independent. The principal component analysis (PCA) is a standard approach to analyze spectral data [3]. PCA uses second-order statistics, which is appropriate for dealing with Gaussian distributed data. However, MS data can hold interesting information with non-Gaussian distribution that can be analyzed with ICA.

In the present study, the ICA have been applied to an analysis of mass spectra of oligomeric species emerged from aluminium sulphate.

Fig. 1 Mass spectra of aluminium sulphate. X-axis represent molecular mass to charge (m/z) ratios and Y-axis is relative intensity of signals

II. Materials and Methods

A. Chemical Compounds

Aluminium sulphate octadecahydrate $Al_2(SO_4)_3 \cdot 18H_2O$ (Merck, purum) 0.03g was dissolved in 1000 ml deionized, distilled water (0.1×10^{-3} mol dm^{-3} Al^{3+}). The solution was divided into five parts. The first part of the solution was left at its initial pH 4.44 and the other parts of solution were adjusted to pH 10.75, 5.05, 7.40, and 6.75 by a drop wise addition of concentrated tetramethyl ammonium hydroxide pentahydrate, $(CH_3)_4NOH \cdot 5H_2O$, (TMA, Fluka, reinst).

The pH was determined with a Mettler Toledo MP 220 pH meter using a Mettler Toledo InLab 410 Ag/AgCl-electrode calibrated with FF-Chemicals buffer solutions pH 4.00 and 7.00. Samples were allowed to equilibrate prior to recording for 1 h.

B. Mass Spectroscopy

The mass spectrometric method is based on detection of mass to charge ratio, m/z value of the studied chemical components. The mass spectra were recorded by a Micromass LCT time-of-flight mass spectrometer equipped with a Z-spray electrospray interface. The solutions were introduced into the spectrometer by a Harvard Apparatus Model 11 syringe pump at flow-rates of 10μL/min. The operating conditions used were tested by the experimental design calculated by MODDE 3.0- computer program. Signals from all the pH values were collected for 1 minute and combined to a sum spectrum. All ESI mass spectral data in the positive and negative ion modes were acquired using Masslynx NT software (version 3.4) and transformed to Microsoft ® Excel 2002 for calculations.

III. Independent Component Analysis

It is here assumed that there are n observed spectra, OS_1, OS_2, . . ., OS_n, which are linear combinations of m independent components, IC_1, IC_2, . . ., IC_m. The equation for IC_i can be written as:

$$OS_i = a_{i1}IC_1 + a_{i2}IC_2 + ... + a_{im}IC_m = \sum_{j=1}^{m} a_{ij}IC_j \quad (1)$$

where i=1, 2, . . ., n and the a_{ij} are real coefficients (contributions of ICs). The independent components, IC_j, and also the corresponding coefficients, a_{ij}, are unknown. The statistical model in Eq. (1) is called the independent component analysis model. The ICA model is a generative model that describes how the observed data are generated by a process of mixing the components IC_i. Both IC_i and a_{ij} need to be estimated using the observed spectra. The starting point for ICA is the assumption that the components IC_i are statistically independent, which can be concluded from non- Gaussianity [3]. Here, a fixed-point algorithm (Fast-ICA) was used as an implementation of ICA [3]. The analysis was performed using

the Fast-ICA toolbox under the Matlab software platform (Mathworks, Natick, MA, USA).

IV. Results and Discussion

The five mass spectra of oligomeric species were analyzed without a priori knowledge of the ICs and corresponding mixing coefficients. A key issue in using the ICA method is to decide the number of ICs to be estimated. For the spectral set, only the physically meaningful components were chosen, i.e. the signals of components are positive as in the original spectra. The number of physically meaningful ICs was three. These components are shown in Fig. 2. The values of the coefficients aij for each sample are in Table II.

TABLE I

THE COEFFICIENTS A$_{IJ}$ OF THREE INDEPENDENT COMPONENTS FOR EACH SAMPLE

Spectrum	IC 1	IC 2	IC 3
1	-6,2843	111,78	4819,1
2	6195,1	113,8	271,13
3	1609,1	3254,5	483,06
4	2751,9	2119,1	624,68
5	1933,3	2825,3	689,89

The distribution of the oligomeric species may be dominated by factors like pH, reaction time and concentration in addition to the type of aluminium chemicals. Therefore, the pHs of samples were plot as a function of the coefficients a_{ij} (Fig. 3). The correlation coefficients between the coefficients aij of the independent components and pHs of the samples were 0.98, 0.58 and 0.19, respectively.

The first component clearly consists of signals assigned to pH adjusting chemical. For instance the x-value 318 could be assigned to $[TMA_3(SO_4)]^+$ and the value 183 $[TMA_2(OH)(H_2O)]^+$. The signal m/z 562, which could have several assignments including aluminium sulphates, can be said to originate from $[TMA_5(SO_4)_2]^+$.

The second component, IC 2, seems to consist of signals, which could merely be assigned to very soluble trimeric aluminium oxo hydroxide, which, when born after first pH elevation, is independent of pH. It vanishes from the spectrum when the pH is over 10, but the reason could be ion suppression due to the domination of TMA and its clusters in the measurement.

The third component, IC 3 resembles the original spectrum 1, which has been measured with out pH handling. Hence, it could be said that with this operation, the signals that are originated from the pH handling could be separated and omitted from the interpretation.
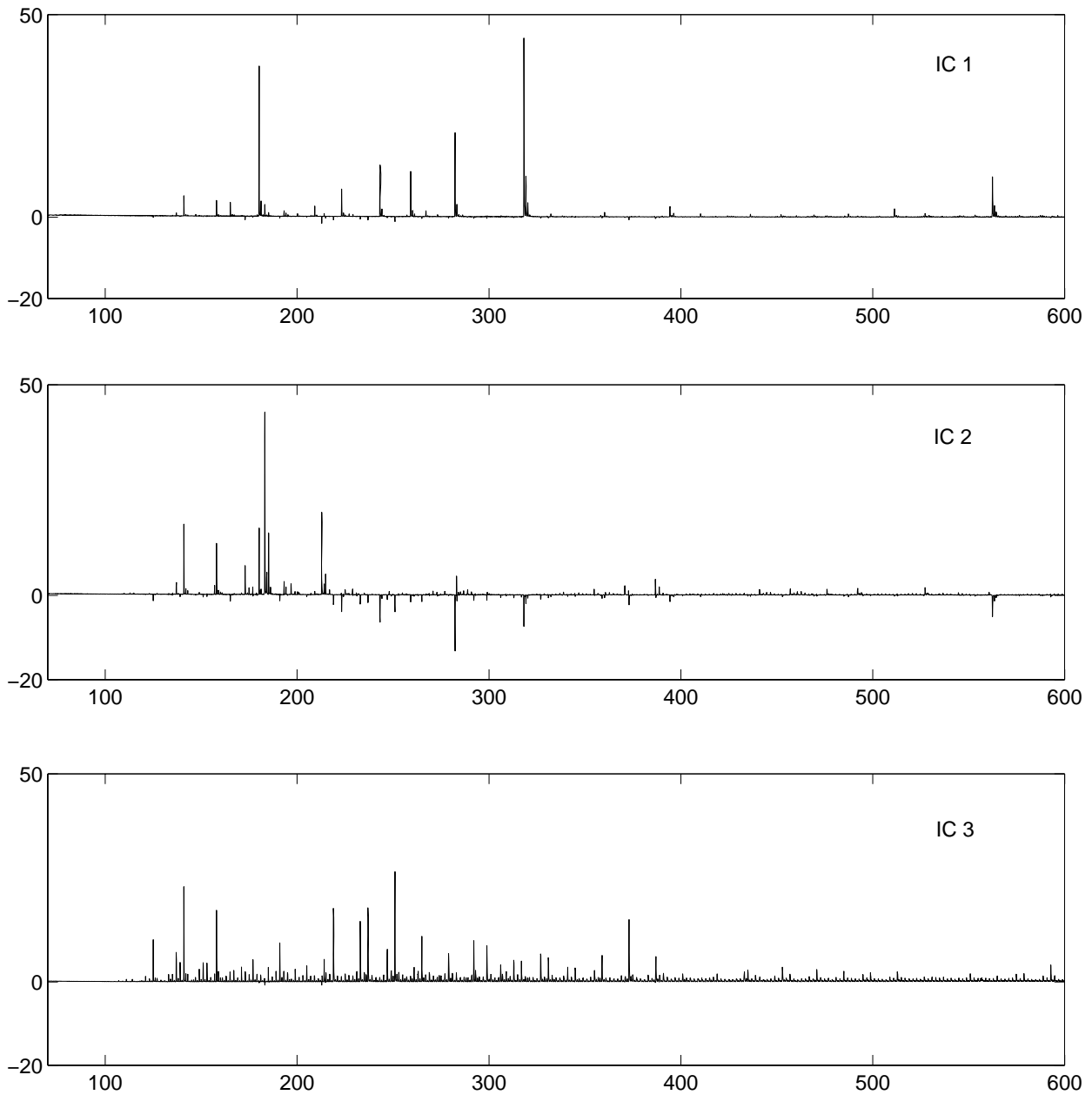
Fig. 2 The independent components of oligomeric species. The first component (IC 1) consists of signals assigned to pH adjusting chemical. The second component (IC 2) to very soluble trimeric aluminium oxo hydroxide and the third component (IC 3) to the original spectrum1.
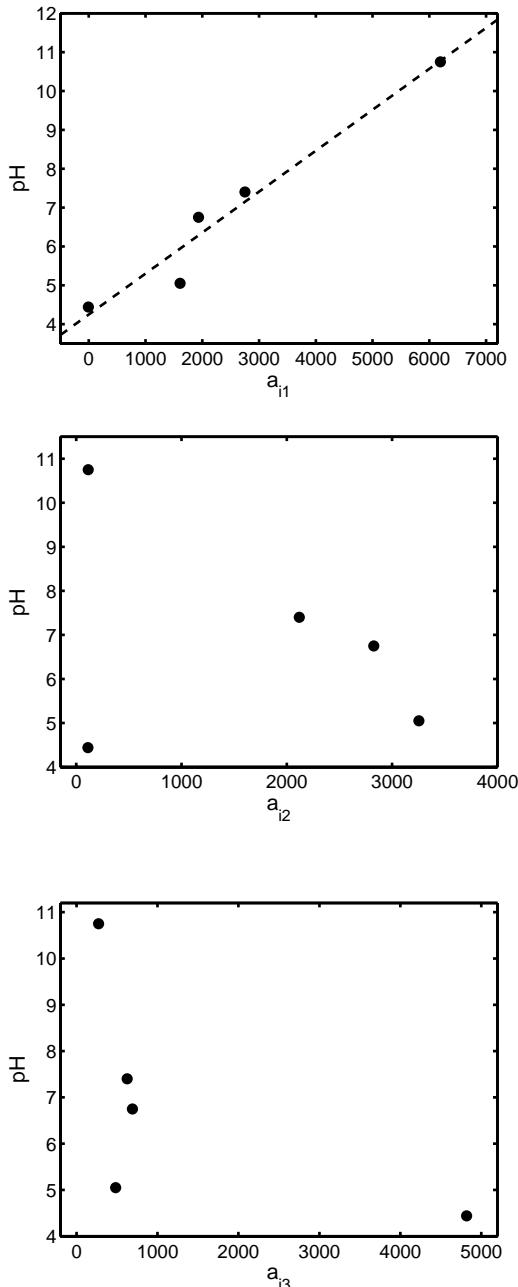
Fig. 3 The pHs of samples as a function of the coefficients a_{ij}

Incorporating ICA into the spectroscopic data analysis allows for decomposition of spectral data from oligomeric species emerged from aluminium chloride into analytical useful components and new clues. This new information can be used for guideline in further study and may lead to better chemical solutions. The present study shows that this kind of approach is a fruitful way of analyzing complex mass spectra.

REFERENCES

[1] A. Sarpola, H. Hellman, V. Hietapelto, J. Jalonen, J. Jokela, J. Rämö, J. Saukkoriipi, "Hydrolysis products of water treatment chemical aluminum sulfate octadecahydrate by electrospray ionization mass spectrometry; connection with mineral compositions", *Polyhedron*, accepted.

[2] A. Sarpola, J. Saukkoriipi, V. Hietapelto, J. Jalonen, J. Jokela, K. Laasonen, J. Rämö, "Identification of hydrolysis products of AlCl$_3$•6H$_2$O with presence of sulphate by electrospray ionization time of flight mass spectrometry and computational methods", *Physical Chemistry Chemical Physics*, 2007, vol. 9, pp. 377–388. DOI: 10.1039/b614814j.

[3] A. Hyvärinen , J. Karhunen, E. Oja, *Independent component analysis*. John Wiley & Sons, 2001

[4] J. Pulkkinen, A. Häkkinen, N. Lundbom, A. Paetau, R. Kauppinen, Y. Hiltunen, "Independent component analysis to proton spectroscopic imaging data of human brain tumours," in *European Journal of Radiology*, 2005, vol. 56, issue 2, pp. 160-164.

[5] V. D. Calhoun, T. Adali, G. D. Pearlson, P. C. M. van Zijl, and J. J. Pekar, ``Independent component analysis of fMRI data in the complex domain,'' in *Magnetic Resonance in Medicine*, 2002, vol. 48, pp. 180-192.

[6] L. Zhang, J. Mei, "Shaping up simple cell's receptive field of animal vision by ICA and its application in navigation system", in *Neural Networks*, 2003, vol. 16, issues 5-6, pp. 609–615.

[7] M. Ameen and H. Szu, "Early vision image analyses using ICA in unsupervised learning ANN", in *Neural Networks*, 1999, IJCNN '99

[8] T. Ristaniemi, and J. Joutsensalo, "On the performance of blind source separation in CDMA downlink", in *Proc. Int. Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, 1999, Aussois, France, pp. 437–441.

[9] K. Kiviluoto and E. Oja, "Independent component analysis for parallel financial time series", In *Proc. Int.Conf. on Neural Information Processing (ICONIP'98)*, 1998, vol. 2, Tokyo, Japan, pp. 895–898.

[10] A. D. Back and A. S. Weigend, "A first application of independent component analysis to extracting structure from stock returns", 1997, *Int. J. on Neural Systems*, vol. 8, issue 4, pp. 473–484.

V. CONCLUSION

The current process optimization of coagulating flocculating units base predominantly on single experiments, not on scientific information. Variations in the composition of raw water cause problems, which can be managed much better if the most efficiently functioning species and particles are known. Improved understanding of the impact of process parameters on purification results will provide a key to enhance material efficiency of processes. This can be achieved with increased understanding of the chemistry behind the process.