

Classification of metal ions according to their complexing properties: a data-driven approach

Igor V. Pletnev*, Vladimir V. Zernov

Department of Chemistry, Lomonosov Moscow State University, Vorobyovy Gory, 119899 Moscow, Russia

Received 1 February 2001; received in revised form 22 November 2001; accepted 28 November 2001

Abstract

Factor, cluster and self-organizing map analyses were performed for the stability constants of complexes of 24 metal ions and hydrogen with 3960 ligands (15606 values of $\log K_1$). Five factors reproduce 89% of data variability. Both direct clusterization and clusterization on the basis of factor analysis established the existence of six different classes of similar cations. The similarity series for metal ions and relative similarity of several ions are discussed and the Kohonen two-dimensional map, which visually represents the similarity, is presented. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Metal ion complexes; Stability constants; Chemometrics; Principal factor analysis; Hierarchical cluster analysis; Kohonen self-organizing map

1. Introduction

Analytical chemistry has a long-term continuing interest in rationalization and prediction of complexing ability of various metal ions. To achieve this goal, chemists established a number of correlation equations and qualitative rules (reviews [1–3] and monographs [4,5]). However, most of these rules are restricted to specific (though important) classes of metals/ligands.

The most general concepts, metal ions classification schemes related to complexing properties (Arhland–Chatt–Davies [6] and Pearson's HSAB principle [7]) are relatively old and were designed with the use of rather limited experimental data. This only emphasizes the significance of the work of those who pioneered the field, however, nowadays the array

of measured stability constants has significantly extended and one may attempt to classify metal cations in more detail.

We are particularly interested in a purely phenomenological classification which uses no a priori assumptions. From one side, such a *data-driven* approach may seem inherently limited as it ignores existing knowledge about the subject. From another side, it may appear extremely helpful in broadening our knowledge regarding metal ions nature and in eliminating misinterpretations/pitfalls that accompany even the most elegant and successful theoretical concepts (for peculiar examples, see Pearson's comments [8] on implementation of HSAB principle).

There exists a number of well-established statistical techniques for classification, namely, principal factor analysis (PFA) and cluster analysis [9–11]. *Factor analysis* tries to explain all the variability in complexation constants for m cations through the

* Corresponding author.

E-mail address: pletnev@analyt.chem.msu.ru (I.V. Pletnev).

linear combination of increments inherent to a small number k of factors, $k \ll m$. One may assume that the factors are related to archetypical metal ions or, better, to the archetypical features expressed, to different extent, in any metal's complexation behavior.

After the initial data set is reduced to k factors, it is possible to place the points corresponding to various metal ions in k -dimensional space and to study their distribution by means of *cluster analysis*. Cluster analysis may also be applied to the original data set itself, however, pre-factorization reduces the size of the problem, eliminates redundancy/noise and helps to treat missed data.

Finally, one may use a relatively new technique of Kohonen self-organizing maps (SOM) [13] for the analysis of similarity/dissimilarity in multidimensional space. The technique is based on projection of k -dimensional data onto a two-dimensional map thus providing easily analyzable patterns.

This paper reports on the results of application of factor, cluster and SOM analyses to a large array of metal ions stability constants (24 metal and hydrogen cations, 3960 ligands, 15,606 values of $\log K_1$).

2. Theoretical background

2.1. Data representation

The key concept in the following consideration is a metal ion complexation profile. The profile for an ion is an l -dimensional vector containing stability constant values for all the l ligands under study. It characterizes the relative affinity of the cation to different ligands.

The similarity in complexing behavior of two metal ions is measured for their profile vectors by means of Pearson correlation coefficient, r . Note that this measure refers not to similarity in absolute stability constant values but to similarity in trends, i.e. the similarity in how values change along the ligands series.

2.2. PFA [9,11]

Mathematically, PFA is a variation of a principal component analysis (PCA). The sample correlation matrix \mathbf{R} is constructed from normalized initial data

matrix $\mathbf{Z}(m \times n)$ as follows: $\mathbf{R} = (1/(n - 1))\mathbf{Z}'\mathbf{Z}$, where \mathbf{Z}' is transpose of matrix \mathbf{Z} (i.e. data matrix contains normalized stability constant values, the columns correspond to cations and the rows to ligands). The diagonal elements of $\mathbf{R}(m \times m)$, which is a unity in PCA, are replaced with the estimates of communality (here is the general difference between PFA and PCA). After that, factor analysis model is used for calculation of principal factors, $\mathbf{R} = \mathbf{A}\mathbf{A}' + \mathbf{U}$, where \mathbf{A} ($m \times k$, where k the number of factors) is the factor pattern. The matrix of characteristic vectors is associated with linear principal factors (columns of \mathbf{A} are *factor loadings*), \mathbf{A}' is \mathbf{A} transpose, \mathbf{U} the residual matrix. For easier interpretation, factors may be rotated according to various methods.

PFA, in our context, reduces a set of metal ions complexing profiles to a set of archetypical profiles, or factors (the estimate of the number of factors may not be self-evident). The set of factors, in terms of factor analysis, is the set of latent variables that adequately describes the variability of initial data, linear combination of several factors must accurately reproduce any ion profile. The number of factors should be less than that of metals since some metals are similar and, by design, factors are orthogonal (i.e. they represent "pure" behavioral types, not "mixed" ones). Mathematically, PFA/PCA is performed through eigendecomposition of correlation matrix or equivalently, through singular value decomposition of normalized data matrix.

2.3. Cluster analysis [10,11]

A common method to investigate how the samples group in multidimensional space is agglomerative cluster analysis. In this method, clusters are formed by grouping samples into bigger and bigger clusters until all samples become members of a single cluster. Before the analysis, each sample forms its own, separate cluster. At the first stage, two samples are combined in the single cluster, at the second, the third sample is added to the growing cluster, and so on. Graphically, this process is illustrated by agglomerative dendrogramm.

There are two important issues: the way of measuring the distance between samples (metrics) and the way of measuring the distance between samples

and cluster (linkage rule). In the present study, we applied the popular Euclidean, squared Euclidean and Manhattan (city-block) metrics in combination with complete linkage, Ward's linkage, and weighted/unweighted pair-group average linkage rules.

2.4. SOM analysis [13]

A powerful and popular method for visualization of multidimensional data on two-dimensional map is the so-called Kohonen SOM (they may be considered as a kind of unsupervised learning neural network).

The approach relies on the immersion of the two-dimensional grid into multidimensional space. Then, in the process of "training", the grid deforms in order to approximate the initial "data cloud" and to preserve data points neighborhood. In a final pattern, point representation is locally continuous, two points which are near on the two-dimensional grid are also near in the initial multidimensional space (the opposite is generally wrong as the two points which are closely located in the initial space may be located far on the grid, it is the price of dimension reduction). The resulting map is similar to a common topographical map, it is divided into regions which correspond to different objects (metal ions). The more closely located their regions are, the higher their similarity.

3. Results and discussion

3.1. Data selection and pre-treatment

We have used experimental data on the first stability constants, $\log K_1$, extracted from IUPAC stability constants database (SCDB, Academic Software, UK). Our licensed version of the database contains data up to 1993, inclusively.

The size of the data set was defined as a result of a compromise between requirements to have compatible experimental conditions and the intention to obtain as large dataset as possible. We included metal ions for which data on stability constants, $\log K_1$, were measured for more than $N = 100$ ligands (the exception is potassium, $N = 99$). Stability constant values mea-

sured at ionic strength ≤ 1 and at temperature 20–30 °C were selected (we did not adjust data to a single ionic strength/temperature). If several $\log K_1$ values for the same complex were present, the median value was calculated and further used (however, if scattering exceeded three logarithmic units, the entry was excluded).

The final data set contained stability constants of complexes for 25 cations: Ag^+ , Al^{3+} , Ba^{2+} , Be^{2+} , Ca^{2+} , Cd^{2+} , Ce^{3+} , Co^{2+} , Cu^{2+} , Fe^{2+} , Fe^{3+} , H^+ , Hg^{2+} , K^+ , La^{3+} , Mg^{2+} , Mn^{2+} , Na^+ , Ni^{2+} , Pb^{2+} , Sr^{2+} , UO^{2+} , VO^{2+} , Y^{3+} , Zn^{2+} with 3960 different ligands. Naturally, the corresponding data matrix contains empty cells, the total number of valid entries being 15606.

A correlation matrix of size 25×25 was built for this data set (Fig. 1). Matrix element r_{ij} is a Pearson correlation coefficient between M_i and M_j cations. Each r_{ij} was calculated on the "common ligands base" (i.e. those ligands which form complexes with both metals i and j). In this relation, it would be more rigorous to call the obtained matrix a *pseudocorrelation* matrix, since "common ligands base" is different for different pairs of metals (interestingly, we have not met this approach to correlation matrix construction in literature). The size of "common ligands base" varies, for different ion pairs, from 16 to 1824, average number being 182 ligands.

3.2. Complexation profiles and their similarity

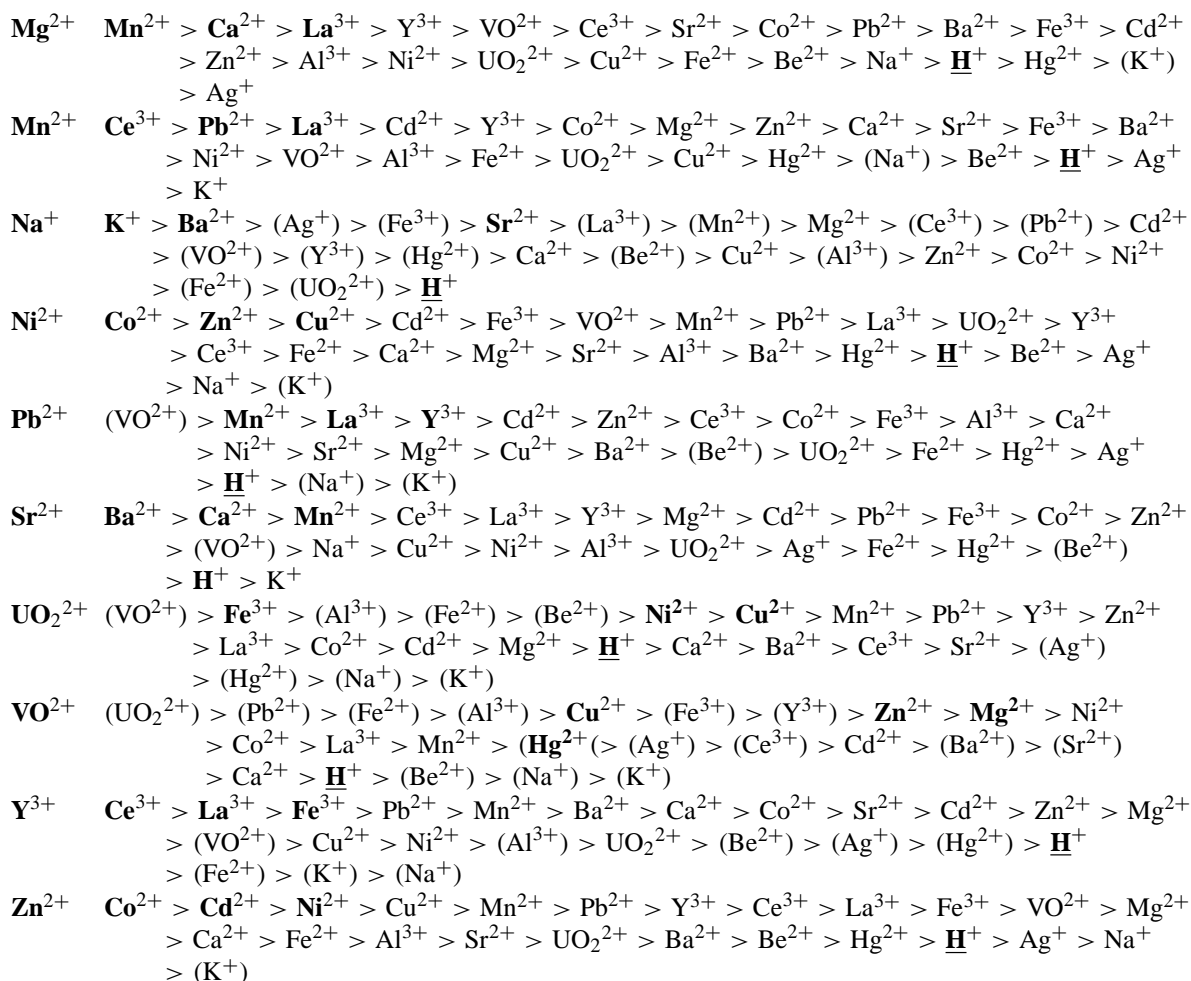
Each entry in a pseudocorrelation matrix defines a similarity between complexation profiles for two cations. One may also build a similarity series for any ion. In such series, the rest of the ions are placed in the decreasing order of the correlation coefficient. Per se, this series is a more informative and chemically perceptual way to represent the correlation matrix.

The similarity series for all the ions are presented below. The three nearest "neighbors" of each ion are shown in bold, in respective series. Ions with the "common ligands base" < 50 ligands are put in brackets (we can not consider them as trusted "neighbors"). Proton in all series is underlined (notably, the correlation with H^+ is poor for the majority of cations).

| | Ag1 | Al3 | Ba2 | Be2 | Ca2 | Cd2 | Co2 | Cu2 | Fe2 | Fe3 | H | Hg2 | K1 | La3 | Mg2 | Mn2 | Na1 | Ni2 | Pb2 | Sr2 | UO2 2 | VO 2 | Y 3 | Zn2 | Ce3 |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Ag1 | 1 | 0.635 | 0.766 | 0.247 | 0.509 | 0.68 | 0.611 | 0.592 | 0.703 | 0.721 | 0.391 | 0.723 | 0.621 | 0.646 | 0.443 | 0.551 | 0.796 | 0.555 | 0.714 | 0.681 | 0.652 | 0.826 | 0.744 | 0.616 | 0.573 |
| Al3 | 0.635 | 1 | 0.734 | 0.830 | 0.703 | 0.826 | 0.813 | 0.865 | 0.783 | 0.892 | 0.812 | 0.735 | 0.451 | 0.848 | 0.807 | 0.844 | 0.582 | 0.743 | 0.888 | 0.733 | 0.890 | 0.914 | 0.843 | 0.822 | 0.832 |
| Ba2 | 0.766 | 0.734 | 1 | 0.674 | 0.915 | 0.836 | 0.801 | 0.743 | 0.663 | 0.850 | 0.601 | 0.720 | 0.693 | 0.913 | 0.857 | 0.906 | 0.842 | 0.741 | 0.836 | 0.964 | 0.732 | 0.795 | 0.919 | 0.792 | 0.934 |
| Be2 | 0.247 | 0.830 | 0.674 | 1 | 0.639 | 0.717 | 0.715 | 0.729 | 0.686 | 0.742 | 0.792 | 0.637 | 0.548 | 0.713 | 0.718 | 0.718 | 0.613 | 0.597 | 0.823 | 0.652 | 0.852 | 0.710 | 0.783 | 0.748 | 0.608 |
| Ca2 | 0.509 | 0.703 | 0.915 | 0.639 | 1 | 0.876 | 0.866 | 0.802 | 0.715 | 0.838 | 0.566 | 0.636 | 0.446 | 0.941 | 0.914 | 0.925 | 0.627 | 0.805 | 0.873 | 0.963 | 0.733 | 0.747 | 0.914 | 0.832 | 0.937 |
| Cd2 | 0.680 | 0.826 | 0.836 | 0.717 | 0.876 | 1 | 0.946 | 0.920 | 0.810 | 0.892 | 0.663 | 0.722 | 0.502 | 0.924 | 0.84 | 0.939 | 0.649 | 0.915 | 0.938 | 0.865 | 0.783 | 0.820 | 0.907 | 0.961 | 0.948 |
| Co2 | 0.611 | 0.813 | 0.801 | 0.715 | 0.866 | 0.946 | 1 | 0.931 | 0.834 | 0.872 | 0.622 | 0.657 | 0.319 | 0.902 | 0.861 | 0.933 | 0.530 | 0.975 | 0.898 | 0.830 | 0.786 | 0.872 | 0.912 | 0.969 | 0.922 |
| Cu2 | 0.592 | 0.865 | 0.743 | 0.729 | 0.802 | 0.920 | 0.931 | 1 | 0.770 | 0.905 | 0.679 | 0.789 | 0.443 | 0.861 | 0.769 | 0.833 | 0.613 | 0.946 | 0.845 | 0.774 | 0.838 | 0.910 | 0.857 | 0.945 | 0.869 |
| Fe2 | 0.703 | 0.783 | 0.663 | 0.686 | 0.715 | 0.810 | 0.834 | 0.770 | 1 | 0.897 | 0.51 | 0.769 | 0.322 | 0.75 | 0.741 | 0.836 | 0.386 | 0.839 | 0.805 | 0.680 | 0.875 | 0.932 | 0.704 | 0.832 | 0.746 |
| Fe3 | 0.721 | 0.892 | 0.850 | 0.742 | 0.838 | 0.892 | 0.872 | 0.905 | 0.897 | 1 | 0.629 | 0.855 | 0.662 | 0.906 | 0.843 | 0.914 | 0.785 | 0.876 | 0.889 | 0.853 | 0.933 | 0.892 | 0.945 | 0.885 | 0.927 |
| H | 0.391 | 0.812 | 0.601 | 0.792 | 0.566 | 0.663 | 0.622 | 0.679 | 0.51 | 0.629 | 1 | 0.443 | 0.336 | 0.663 | 0.636 | 0.600 | 0.378 | 0.615 | 0.699 | 0.594 | 0.771 | 0.743 | 0.709 | 0.679 | 0.663 |
| Hg2 | 0.723 | 0.735 | 0.720 | 0.637 | 0.636 | 0.722 | 0.657 | 0.789 | 0.769 | 0.855 | 0.443 | 1 | 0.510 | 0.787 | 0.544 | 0.753 | 0.632 | 0.652 | 0.724 | 0.655 | 0.627 | 0.856 | 0.717 | 0.682 | 0.764 |
| K1 | 0.621 | 0.451 | 0.693 | 0.548 | 0.446 | 0.502 | 0.319 | 0.443 | 0.322 | 0.662 | 0.336 | 0.51 | 1 | 0.514 | 0.445 | 0.522 | 0.848 | 0.293 | 0.542 | 0.543 | 0.260 | 0.469 | 0.684 | 0.342 | 0.554 |
| La3 | 0.646 | 0.848 | 0.913 | 0.713 | 0.941 | 0.924 | 0.902 | 0.861 | 0.75 | 0.906 | 0.663 | 0.787 | 0.514 | 1 | 0.900 | 0.939 | 0.76 | 0.860 | 0.941 | 0.914 | 0.804 | 0.869 | 0.981 | 0.895 | 0.992 |
| Mg2 | 0.443 | 0.807 | 0.857 | 0.718 | 0.914 | 0.840 | 0.861 | 0.769 | 0.741 | 0.843 | 0.636 | 0.544 | 0.445 | 0.900 | 1 | 0.930 | 0.687 | 0.799 | 0.858 | 0.877 | 0.778 | 0.881 | 0.896 | 0.837 | 0.878 |
| Mn2 | 0.551 | 0.844 | 0.906 | 0.718 | 0.925 | 0.939 | 0.933 | 0.833 | 0.836 | 0.914 | 0.600 | 0.753 | 0.522 | 0.939 | 0.930 | 1 | 0.725 | 0.869 | 0.946 | 0.917 | 0.834 | 0.864 | 0.934 | 0.926 | 0.949 |
| Na1 | 0.796 | 0.582 | 0.842 | 0.613 | 0.627 | 0.649 | 0.530 | 0.613 | 0.386 | 0.785 | 0.378 | 0.632 | 0.848 | 0.760 | 0.687 | 0.725 | 1 | 0.529 | 0.662 | 0.781 | 0.380 | 0.642 | 0.641 | 0.581 | 0.664 |
| Ni2 | 0.555 | 0.743 | 0.741 | 0.597 | 0.805 | 0.915 | 0.975 | 0.946 | 0.839 | 0.876 | 0.615 | 0.652 | 0.293 | 0.860 | 0.799 | 0.869 | 0.529 | 1 | 0.863 | 0.766 | 0.850 | 0.872 | 0.845 | 0.951 | 0.841 |
| Pb2 | 0.714 | 0.888 | 0.836 | 0.823 | 0.873 | 0.938 | 0.898 | 0.845 | 0.805 | 0.889 | 0.699 | 0.724 | 0.542 | 0.941 | 0.858 | 0.946 | 0.662 | 0.863 | 1 | 0.860 | 0.822 | 0.952 | 0.938 | 0.913 | 0.911 |
| Sr2 | 0.681 | 0.733 | 0.964 | 0.652 | 0.963 | 0.865 | 0.830 | 0.774 | 0.680 | 0.853 | 0.594 | 0.655 | 0.543 | 0.914 | 0.877 | 0.917 | 0.781 | 0.766 | 0.86 | 1 | 0.683 | 0.792 | 0.910 | 0.817 | 0.916 |
| UO2 2 | 0.652 | 0.890 | 0.732 | 0.852 | 0.733 | 0.783 | 0.786 | 0.838 | 0.875 | 0.933 | 0.771 | 0.627 | 0.260 | 0.804 | 0.778 | 0.834 | 0.380 | 0.850 | 0.822 | 0.683 | 1 | 0.962 | 0.809 | 0.807 | 0.722 |
| VO 2 | 0.826 | 0.914 | 0.795 | 0.710 | 0.747 | 0.820 | 0.872 | 0.910 | 0.932 | 0.892 | 0.743 | 0.856 | 0.469 | 0.869 | 0.881 | 0.864 | 0.642 | 0.872 | 0.952 | 0.792 | 0.962 | 1 | 0.890 | 0.881 | 0.834 |
| Y 3 | 0.744 | 0.843 | 0.919 | 0.783 | 0.914 | 0.907 | 0.912 | 0.857 | 0.704 | 0.945 | 0.709 | 0.717 | 0.684 | 0.981 | 0.896 | 0.934 | 0.641 | 0.845 | 0.938 | 0.910 | 0.809 | 0.890 | 1 | 0.897 | 0.984 |
| Zn2 | 0.616 | 0.822 | 0.792 | 0.748 | 0.832 | 0.961 | 0.969 | 0.945 | 0.832 | 0.885 | 0.679 | 0.682 | 0.342 | 0.895 | 0.837 | 0.926 | 0.581 | 0.951 | 0.913 | 0.817 | 0.807 | 0.881 | 0.897 | 1 | 0.895 |
| Ce3 | 0.573 | 0.832 | 0.934 | 0.608 | 0.937 | 0.948 | 0.922 | 0.869 | 0.746 | 0.927 | 0.663 | 0.764 | 0.554 | 0.992 | 0.878 | 0.949 | 0.664 | 0.841 | 0.911 | 0.916 | 0.722 | 0.834 | 0.984 | 0.895 | 1 |

Fig. 1. Correlation matrix for stability constants of 25 cations. Each Pearson correlation coefficient calculated for the same set of ligands “common ligands base” for two ions. The “bases” are different for different pairs of ions.

| | |
|------------------------|---|
| Ag⁺ | (VO ²⁺) > (Na ⁺) > Ba²⁺ > (Fe ³⁺) > (Y ³⁺) > Hg²⁺ > Pb²⁺ > Fe ²⁺ > Sr ²⁺ > Cd ²⁺ > (UO ₂ ²⁺) > (La ³⁺) > (Al ³⁺) > (K ⁺) > Zn ²⁺ > Co ²⁺ > Cu ²⁺ > (Ce ³⁺) > Mn ²⁺ > Ni ²⁺ > Ca ²⁺ > Mg ²⁺ > H⁺ > (Be ²⁺) |
| Al³⁺ | (VO ²⁺) > Fe³⁺ > (UO ₂ ²⁺) > Pb²⁺ > Cu²⁺ > La ³⁺ > Mn ²⁺ > (Y ³⁺) > (Ce ³⁺) > (Be ²⁺) > Cd ²⁺ > Zn ²⁺ > Co ²⁺ > H⁺ > Mg ²⁺ > (Fe ²⁺) > Ni ²⁺ > (Hg ²⁺) > Ba ²⁺ > Sr ²⁺ > Ca ²⁺ > (Ag ⁺) > (Na ⁺) > (K ⁺) |
| Ba²⁺ | Sr²⁺ > Ce³⁺ > Y³⁺ > Ca ²⁺ > La ³⁺ > Mn ²⁺ > Mg ²⁺ > Fe ³⁺ > Na ⁺ > Pb²⁺ > Cd ²⁺ > Co ²⁺ > (VO ²⁺) > Zn ²⁺ > Ag ⁺ > Cu ²⁺ > Ni ²⁺ > Al ³⁺ > UO ₂ ²⁺ > Hg ²⁺ > (K ⁺) > (Be ²⁺) > Fe ²⁺ > H⁺ |
| Be²⁺ | (UO ₂ ²⁺) > (Al ³⁺) > (Pb ²⁺) > H⁺ > (Y ³⁺) > Zn²⁺ > (Fe ³⁺) > Cu²⁺ > Mg ²⁺ > Mn ²⁺ > (Cd ²⁺) > Co ²⁺ > (La ³⁺) > (VO ²⁺) > (Fe ²⁺) > (Ba ²⁺) > (Sr ²⁺) > Ca ²⁺ > (Hg ²⁺) > (Na ⁺) > (Ce ³⁺) > Ni ²⁺ > (K ⁺) > (Ag ⁺) |
| Ca²⁺ | Sr²⁺ > La³⁺ > Ce³⁺ > Mn ²⁺ > Ba ²⁺ > Y ³⁺ > Mg ²⁺ > Cd ²⁺ > Pb²⁺ > Co ²⁺ > Fe ³⁺ > Zn ²⁺ > Ni ²⁺ > Cu ²⁺ > VO ²⁺ > UO ₂ ²⁺ > Fe ²⁺ > Al ³⁺ > Be ²⁺ > Hg ²⁺ > Na ⁺ > H⁺ > Ag ⁺ > K ⁺ |
| Cd²⁺ | Zn²⁺ > Ce³⁺ > Co²⁺ > Mn ²⁺ > Pb²⁺ > La ³⁺ > Cu ²⁺ > Ni ²⁺ > Y ³⁺ > Fe ³⁺ > Ca ²⁺ > Sr ²⁺ > Mg ²⁺ > Ba ²⁺ > Al ³⁺ > VO ²⁺ > Fe ²⁺ > UO ₂ ²⁺ > Hg ²⁺ > (Be ²⁺) > Ag ⁺ > H⁺ > Na ⁺ > (K ⁺) |
| Ce³⁺ | La³⁺ > Y³⁺ > Mn²⁺ > Cd ²⁺ > Ca ²⁺ > Ba ²⁺ > Fe ³⁺ > Co ²⁺ > Sr ²⁺ > Pb²⁺ > Zn ²⁺ > Mg ²⁺ > Cu ²⁺ > Ni ²⁺ > (VO ²⁺) > (Al ³⁺) > Hg ²⁺ > Fe ²⁺ > UO ₂ ²⁺ > H⁺ > (Na ⁺) > (Be ²⁺) > (Ag ⁺) > (K ⁺) |
| Co²⁺ | Ni²⁺ > Zn²⁺ > Cd²⁺ > Mn ²⁺ > Cu ²⁺ > Ce³⁺ > Y ³⁺ > La ³⁺ > Pb²⁺ > Fe ³⁺ > VO ²⁺ > Ca ²⁺ > Mg ²⁺ > Fe ²⁺ > Sr ²⁺ > Al ³⁺ > Ba ²⁺ > UO ₂ ²⁺ > Be ²⁺ > Hg ²⁺ > H⁺ > Ag ⁺ > Na ⁺ > (K ⁺) |
| Cu²⁺ | Ni²⁺ > Zn²⁺ > Co²⁺ > Cd ²⁺ > VO ²⁺ > Fe ³⁺ > Ce³⁺ > Al ³⁺ > La ³⁺ > Y ³⁺ > Pb²⁺ > UO ₂ ²⁺ > Mn ²⁺ > Ca ²⁺ > Hg ²⁺ > Sr ²⁺ > Fe ²⁺ > Mg ²⁺ > Ba ²⁺ > Be ²⁺ > H⁺ > Na ⁺ > Ag ⁺ > (K ⁺) |
| Fe²⁺ | (VO ²⁺) > Fe³⁺ > (UO ₂ ²⁺) > Ni²⁺ > Mn²⁺ > Co ²⁺ > Zn ²⁺ > Cd ²⁺ > Pb²⁺ > (Al ³⁺) > Hg ²⁺ > Cu ²⁺ > La ³⁺ > Ce³⁺ > Mg ²⁺ > Ca ²⁺ > Ag ⁺ > (Y ³⁺) > (Be ²⁺) > Sr ²⁺ > Ba ²⁺ > H⁺ > (Na ⁺) > (K ⁺) |
| Fe³⁺ | Y³⁺ > UO₂²⁺ > ³⁺ Ce³⁺ > Mn ²⁺ > La ³⁺ > Cu ²⁺ > (VO ²⁺) > Fe ²⁺ > Cd ²⁺ > Pb²⁺ > Al ³⁺ > Zn ²⁺ > Ni ²⁺ > Co ²⁺ > Sr ²⁺ > Hg ²⁺ > Ba ²⁺ > Mg ²⁺ > Ca ²⁺ > (Na ⁺) > (Ag ⁺) > (Be ²⁺) > (K ⁺) > H⁺ |
| H⁺ | Al³⁺ > Be²⁺ > UO₂²⁺ > VO ²⁺ > Y ³⁺ > Pb²⁺ > Cu ²⁺ > Zn ²⁺ > Ce³⁺ > La ³⁺ > Cd ²⁺ > Mg ²⁺ > Fe ³⁺ > Co ²⁺ > Ni ²⁺ > Ba ²⁺ > Mn ²⁺ > Sr ²⁺ > Ca ²⁺ > Fe ²⁺ > Hg ²⁺ > Ag ⁺ > Na ⁺ > K ⁺ |
| Hg²⁺ | (VO ²⁺) > Fe³⁺ > Cu²⁺ > La³⁺ > Fe ²⁺ > Ce³⁺ > Mn ²⁺ > (Al ³⁺) > Ag ⁺ > Pb²⁺ > Cd ²⁺ > Ba ²⁺ > (Y ³⁺) > Zn ²⁺ > Co ²⁺ > Sr ²⁺ > Ni ²⁺ > (Be ²⁺) > Ca ²⁺ > (Na ⁺) > (UO ₂ ²⁺) > Mg ²⁺ > (K ⁺) > H⁺ |
| K⁺ | Na⁺ > (Ba ²⁺) > Y³⁺ > Fe³⁺ > Ag ⁺ > (Ce ³⁺) > Be ²⁺ > Sr ²⁺ > (Pb ²⁺) > (Mn ²⁺) > (La ³⁺) > (Hg ²⁺) > (Cd ²⁺) > (VO ²⁺) > (Al ³⁺) > Ca ²⁺ > (Mg ²⁺) > (Cu ²⁺) > (Zn ²⁺) > H⁺ > (Fe ²⁺) > (Co ²⁺) > (Ni ²⁺) > (UO ₂ ²⁺) |
| La³⁺ | Ce³⁺ > Y³⁺ > Pb²⁺ > Ca ²⁺ > Mn ²⁺ > Cd ²⁺ > Sr ²⁺ > Ba ²⁺ > Fe ³⁺ > Co ²⁺ > Mg ²⁺ > Zn ²⁺ > VO ²⁺ > Cu ²⁺ > Ni ²⁺ > Al ³⁺ > UO ₂ ²⁺ > Hg ²⁺ > (Na ⁺) > Fe ²⁺ > (Be ²⁺) > H⁺ > (Ag ⁺) > (K ⁺) |



3.3. Factor and cluster analyses

PFA established that the minimal number of factors describing the whole dataset equals to four. This result was obtained from comparison of eigenvalues of pseudomatrix and eigenvalues of random (normal distribution) matrix. However, the other popular criteria recommends to retain all factors describing 90% of data variance, this results in selection of five factors which reproduce 89%. Characteristics of several first factors compared with random values are presented in Fig. 2. In Table 1, the factors are shown in more details.

In PFA experiment, we estimated initial communalities, i.e. diagonal elements of correlation matrix, as squared multiple correlation coefficients for each

original variable (with all the others). The improvement of factor structure has been performed through varimax rotation.

As is shown on Table 1, the total variance for the first two factors is high, ~80%. It should be logical to try

Table 1
Characteristics of the first five factors

| Number of factors | Eigenvalue | Total variance (%) | Cumulative total variance (%) |
|-------------------|------------|--------------------|-------------------------------|
| 1 | 18.574 | 74.297 | 74.297 |
| 2 | 1.519 | 6.076 | 80.373 |
| 3 | 0.937 | 3.748 | 84.121 |
| 4 | 0.816 | 3.265 | 87.386 |
| 5 | 0.373 | 1.494 | 88.880 |

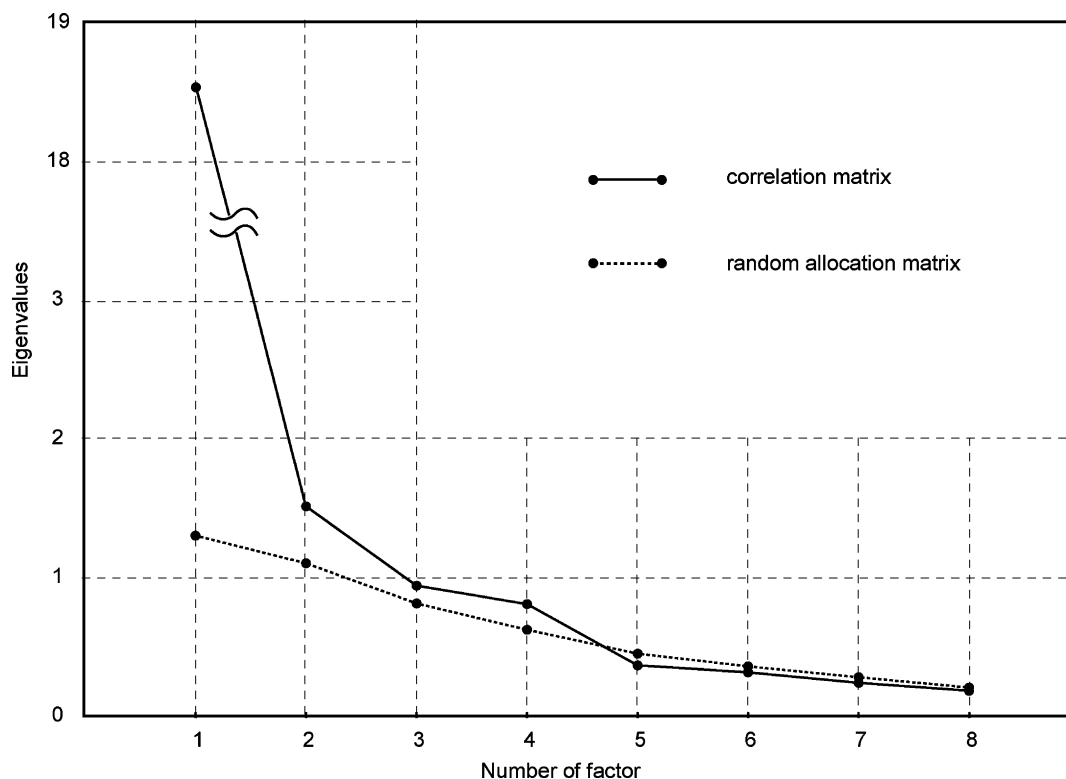


Fig. 2. Comparison of the eigenvalues calculated for the correlation matrix under study and the random allocation matrix.

to cluster cations in the simple two-dimensional space of these factors, however, such analysis (performed also for other factor pairs) did not appear informative.

The hierarchical cluster analysis (we experimented with Euclidean and Manhattan distances and with several linkage rules) was carried out in five factors loadings space. Alternatively, cluster analysis was applied to the initial data, where pseudocorrelation matrix was considered as a matrix of distances in 25-dimensional space (metrics was $1 - \text{Pearson } r_{ij}$). The dendrogram of the last version of clusterization is presented on Fig. 3.

3.4. Metal ions classification

All the analysis versions performed arrived to the similar picture. At the next to simple ternary classification (like [6,7]) level, there appear six classes. These are presented below. Cations that may migrate from class to class in different versions of clusterization are

placed in brackets (note that, for all migrations, involved cation may migrate only between two clusters).

- I. $\text{K}^+ \text{Na}^+$
- II. $\text{Ag}^+ \text{Hg}^{2+}$
- III. $\text{H}^+ \text{Be}^{2+} (\text{Al}^{3+})$
- IV. $\text{UO}_2^{2+} \text{VO}^{2+} \text{Fe}^{2+} (\text{Fe}^{3+}) (\text{Al}^{3+})$
- V. $\text{Co}^{2+} \text{Cu}^{2+} \text{Ni}^{2+} \text{Zn}^{2+} (\text{Cd}^{2+})$
- VI. $\text{Ba}^{2+} \text{Sr}^{2+} \text{Ca}^{2+} \text{Mg}^{2+} \text{Mn}^{2+} \text{Pb}^{2+} \text{La}^{3+} \text{Y}^{3+}$
 $\text{Ce}^{3+} (\text{Fe}^{3+}) (\text{Cd}^{2+})$

Some notes seem appropriate here.

Hg^{2+} and Ag^+ , formally falling into a single class, form, in fact, two classes, which are located closer to each other than to the rest of clusters. Be^{2+} and H^+ form a tight group, notably, these cations have the smallest ionic radii. Similarity of alkali and alkaline earth metal ions is self-explanatory.

Borderline (according to Pearson) metal ions do group well. However, the position of manganese is interesting. Mn^{2+} ion is located nearer to alkaline earth

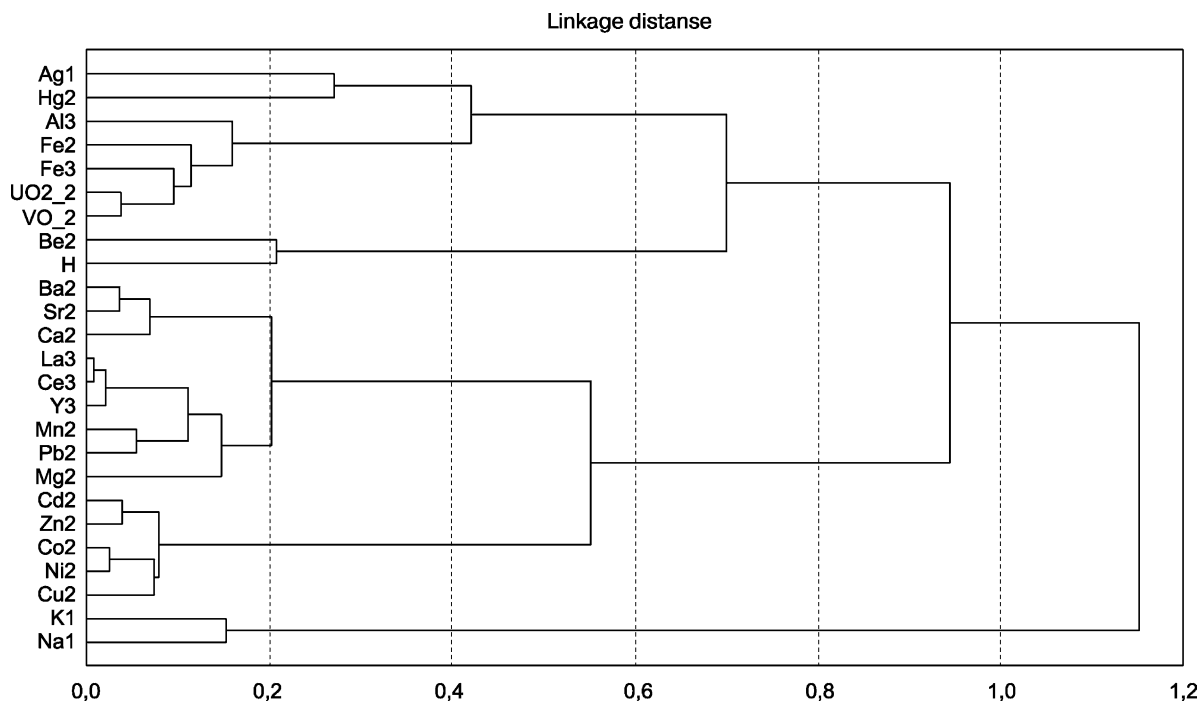


Fig. 3. Example clusterization diagram for 25 ions (Wards linkage rule, Pearson metrics).

and rare earth elements, as well as to Pb^{2+} (this may be attributed to d^5 -electron configuration which does not supply electron stabilization by ligand field).

The obtained classification does not contradict the well-known schemes by Arhland–Chatt–Davies/Pearson but complements them. It allows to examine the complexing properties in more detail.

Let us consider a typical practical task. For some metal ions, the number of measured stability constants is small, due to “poor” chemistry in aqueous solutions or simply as a result of the researchers’ neglect. Estimation of missed stability constants would be of interest. One may naturally turn to similar metal ions and try to obtain the suitable (linear) dependence of their stability constants to those for metal of interest. After calibration, the dependence may be used to predict stability constants in missed cases. Naturally, the more similar cation we choose, the more chances exist for valid predictions. The classification helps to select “nearest neighbor” for building a trusted relationship.

Let us consider two examples.

1. The divalent lead ion. According to Pearson, Pb^{2+} is at borderline. It seems natural to correlate $\log K_1$ values for Pb^{2+} with those for ions in the same group, e.g., Ni^{2+} . However, this correlation is far from perfect, as is illustrated by Fig. 4a.

According to our classification, lead belongs to cluster VI. The similarity series shows that its nearest neighbor in cluster is Mn^{2+} (VO_2^{2+} is even closely located but this conclusion is based on too small a “common ligands base”).

The result is rather unexpected. In different variants of HSAB formulation, Mn^{2+} is either missing or classified as “hard”, depending on the author (Pearson himself did not classify Mn^{2+} [8]). Most authors place the Pb^{2+} into borderline group. Anyway, the similarity between Mn^{2+} and Pb^{2+} is not what all the chemists were to expect.

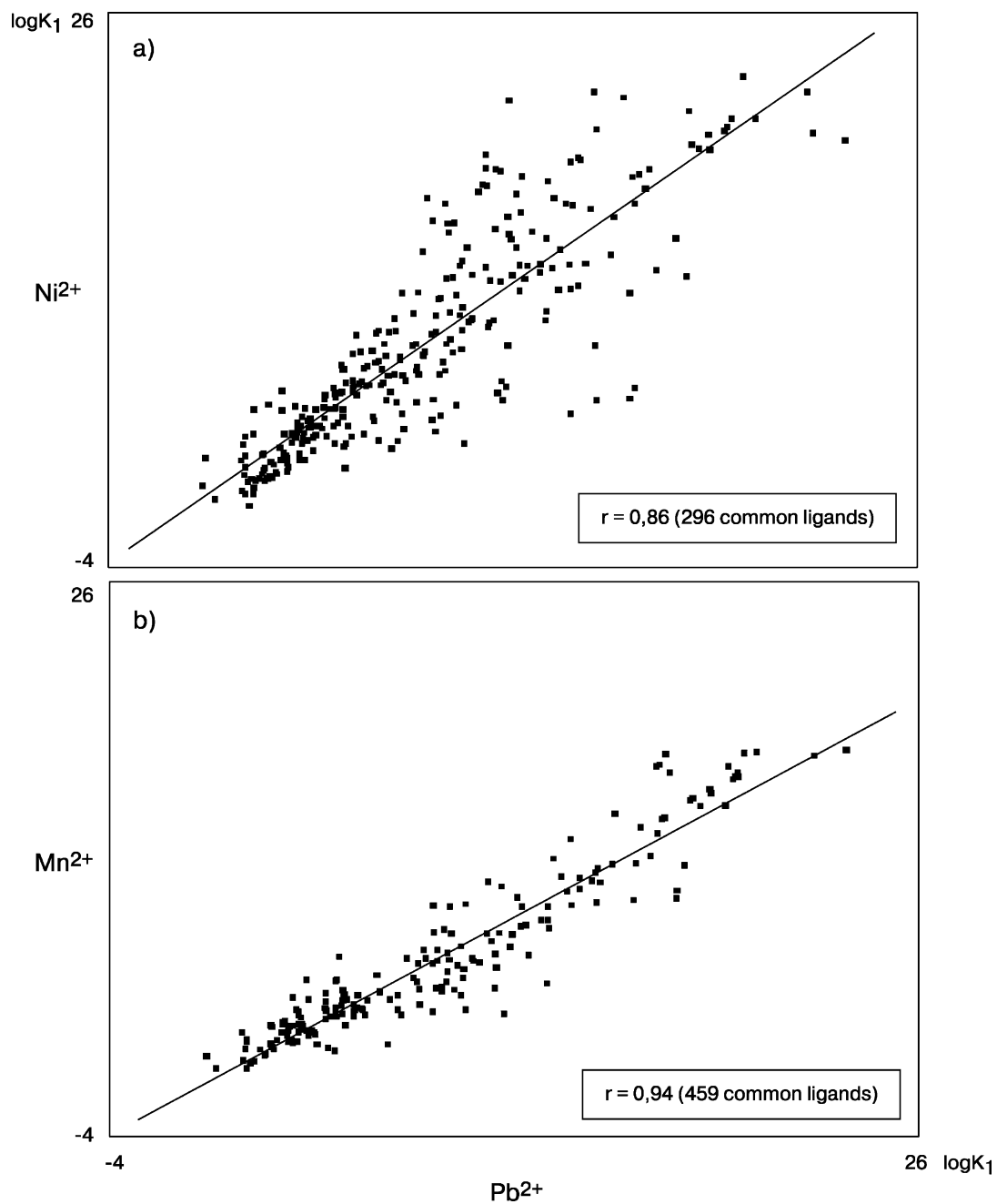


Fig. 4. Correlation between the stability constants for complexes of Pb^{2+} vs. Ni^{2+} and Mn^{2+} .

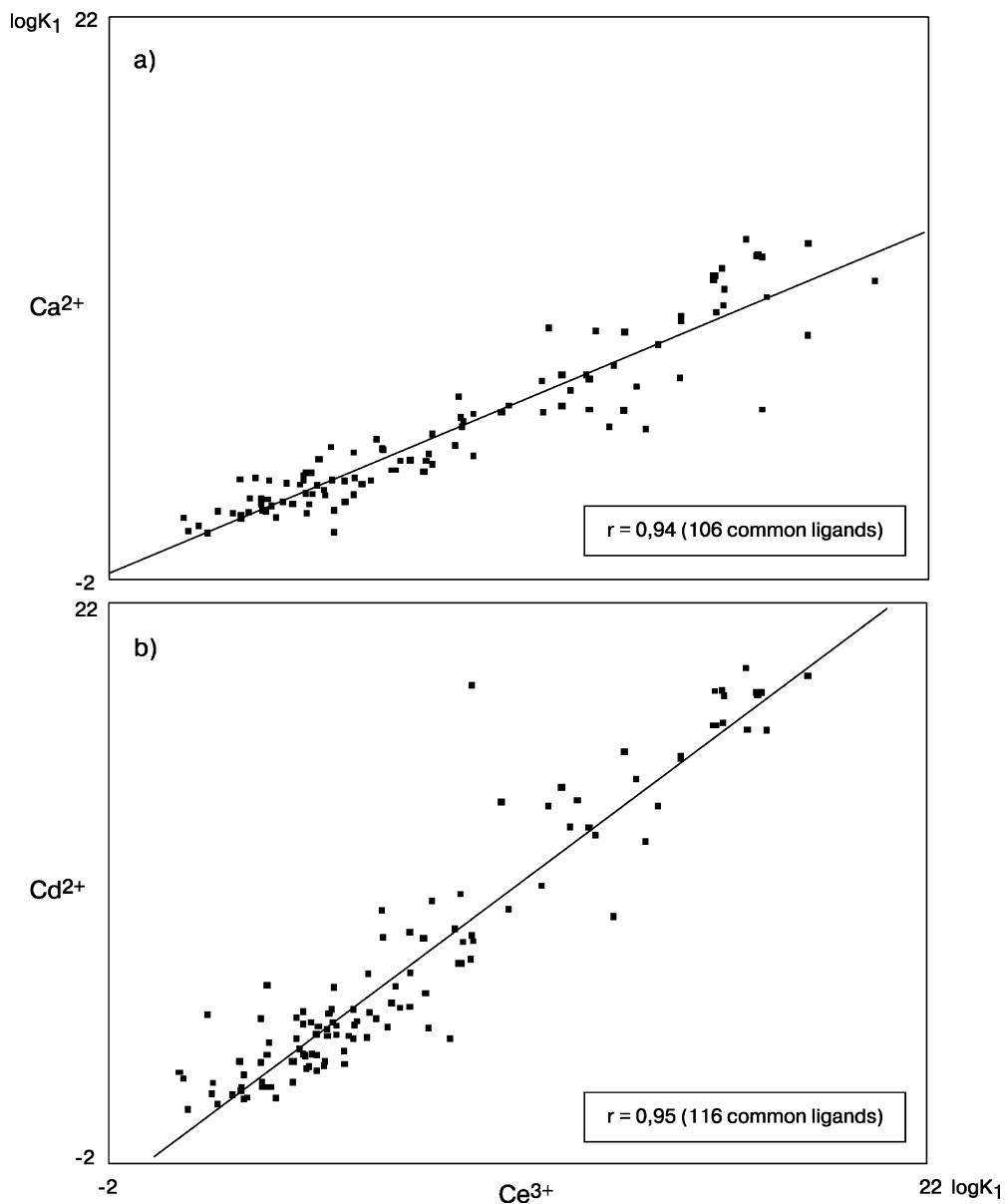


Fig. 5. Correlation between the stability constants for complexes of Ce^{3+} vs. Ca^{2+} and Cd^{2+} .

Nevertheless, the observed correlation is evidently good (Fig. 4b). Linear regression makes it possible to estimate $\log K_1$ for Pb^{2+} through $\log K_1$ for Mn^{2+} (and vice versa) in a broad range of stability constant values, from 0 to 23 logarithmic units.

2. The trivalent cerium ion. It is, undoubtedly, “hard” cation in Pearson’s classification. According to similarity series, Mn^{2+} and Cd^{2+} are similar to it. If Mn^{2+} is generally considered as “hard”, Cd^{2+} is commonly claimed to be “soft”. However, Ce^{3+} and Cd^{2+} is similar,

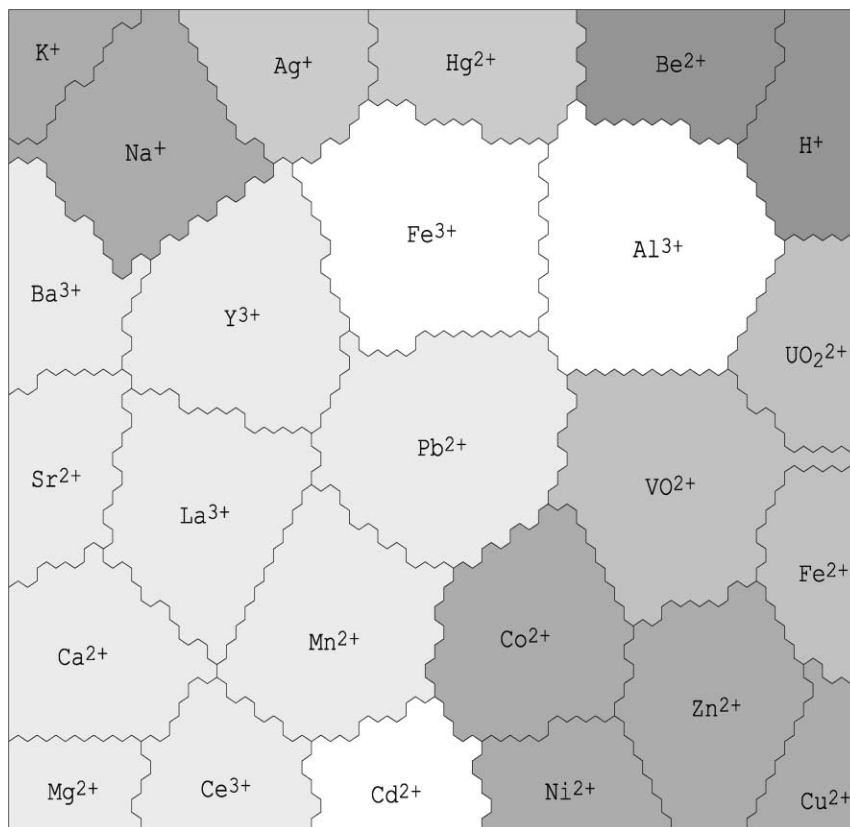


Fig. 6. The Kohonen self-organizing map representing distribution of metal ions by their complexing properties. The different gradations of gray indicate different clusters (see text); cations that may migrate from cluster to cluster are marked with white.

and, actually, more similar than Ce^{3+} and Ca^{2+} (Fig. 5). This observation seems even more unexpected if we consider ionic radii: 1.01 Å for Ce^{3+} and 0.85 Å for Cd^{2+} against practically identical values for Ce^{3+} and Ca^{2+} , 1.01 and 1.00 Å [12].

3.5. Two-dimensional map of metal ions similarity

To obtain visual patterns of similarity/dissimilarity in metal ions complexing behavior, we built a Kohonen map, Fig. 6.

The map is built for the data in factor space where the points are positioned precisely, note that the direct mapping of initial complexation data is impossible due to empty cells in data matrix. The factor loadings of the five most significant factors were used. The map

was built with use of Viscovery SOMine program [14]. The map has hexagonal structure, grid size: 45×42 nodes “neurons”, gaussian neighborhood function was used.

The map is a powerful visualization tool, which complements similarity series and classification scheme. For example, the similarity between lead and manganese ions becomes evident at a first glance, also is evident that cadmium is located close to cerium, as is shown in the second example above.

Multi-colored shading was used to help visualize the clusters established by hierarchical clusterization and to check the mapping for artifacts existing (two closely located points lying within the initial space may be located far on the grid). It appeared that the map reflects cations distribution in factor loading space with good precision.

3.6. Limitations

We should emphasize that the reported analysis is based on a purely data-driven approach, with all its inherent drawbacks. Herein, we intentionally did not interpret the obtained factors and classes.

The analysis missed some relatively important points. In particular, the stability constants were not adjusted to a common ionic strength and common temperature. Another problem is that measuring the inter-cation similarity was limited to existing “common ligands base”, which may be insufficient. For example, calculated similarity of K^+ – Hg^{2+} or Na^+ – Ag^+ is, undoubtedly, overestimated since the stability constants for complexes of alkaline metals with sulfur-containing ligands are practically not known and, naturally, are missed. Possibly, it would be better to assign them a formally negligible value but such an approach could not be considered general.

Nevertheless, the presented classification reflects the numerous important patterns in metal ions complexation behavior, we hope that it will be helpful for analytical and coordination chemists.

Acknowledgements

The authors thank Dr. O. Obrezkov, Prof. Yu. Zolotov (Moscow University) and anonymous referees for

their interest and useful comments. We are also grateful to Eudaptics gmbh (Wien, Austria) who generously made their SOM analysis software (limited version) available from the Web site.

References

- [1] R.D. Hancock, *Analyst* 122 (1997) 51R.
- [2] P.W. Dimmock, P. Warwick, R.A. Robbins, *Analyst* 120 (1995) 2159.
- [3] R.D. Hancock, A.E. Martell, *Chem. Rev.* 89 (1989) 1875.
- [4] M. Bek, I. Nadpal, *Chemistry of complex equilibria*, Mir, Moscow, 1989 (in Russian).
- [5] V.N. Kumok, *Principles of coordination complexes stability in solution*, TGU-Publishing, Tomsk, 1977 (in Russian).
- [6] S. Ahrland, J. Chatt, N.R. Davies, *Quart. Rev.* 12 (1958) 265.
- [7] R.G. Pearson, *J. Am. Chem. Soc.* 85 (1963) 3533.
- [8] R.G. Pearson, *Inorg. Chim. Acta* 240 (1995) 93.
- [9] E.R. Malinovski, *Factor Analysis in Chemistry*, 2nd Edition, Wiley, New York, 1991.
- [10] J.A. Hartigan, *Clustering algorithms*, Wiley, New York, 1975.
- [11] M.A. Sharaf, D.L. Illman, B.R. Kowalski, *Chemometrics*, Wiley, New York, 1986.
- [12] R.D. Shannon, *Acta Crystallogr. Sect. A* A32 (1976) 761.
- [13] T. Kohonen, *Self-Organizing Map*, 2nd Edition, Springer, Berlin, 1997, 426 pp.
- [14] Viscovery SOMine 3.0, copyright© by 1999 Eudaptics software gmbh; free limited version available from <http://www.eudaptics.com>.