

# An analysis of distinguishing composite dissolved metal–ligand systems measurable by stripping voltammetry

Cédric Garnier, Ivanka Pižeta, Stéphane Mounier, Vlado Cuculi, Jean Yves Benaïm

► **To cite this version:**

Cédric Garnier, Ivanka Pižeta, Stéphane Mounier, Vlado Cuculi, Jean Yves Benaïm. An analysis of distinguishing composite dissolved metal–ligand systems measurable by stripping voltammetry. *Analytica Chimica Acta*, Elsevier Masson, 2005, 538, pp.263 - 271. <10.1016/j.aca.2005.02.043>. <hal-01096804>

**HAL Id: hal-01096804**

**<https://hal-univ-tln.archives-ouvertes.fr/hal-01096804>**

Submitted on 5 Jan 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# An analysis of distinguishing composite dissolved metal–ligand systems measurable by stripping voltammetry

Cédric Garnier<sup>a</sup>, Ivanka Pižeta<sup>b,\*</sup>, Stéphane Mounier<sup>a</sup>,  
Vlado Cuculić<sup>b</sup>, Jean Yves Benaim<sup>a</sup>

<sup>a</sup> *PROTEE Laboratory, University of Sud Toulon Var-BP 20132, 83957 La Garde, France*

<sup>b</sup> *Center for Marine and Environmental Research, Ruđer Bošković Institute, P.O. Box 180, Bijenicka c. 54, 10002 Zagreb, Croatia*

Received 7 October 2004; received in revised form 17 December 2004; accepted 16 February 2005  
Available online 14 March 2005

## Abstract

Voltammetry is a method able to distinguish in certain degree the speciation of dissolved metals. An analysis of its ability to discern composite and more complex dissolved metal–ligand systems has been carried out by simulating the experiments for determination of metal–ligand complexing parameters. Logarithmic equidistant addition of metal was presumed, covering 2.5 decades. The data obtained with the preset parameter values were subsequently fitted to the presumed models. Data points under the detection limit  $DL = 10^{-10} \text{ mol L}^{-1}$  were eliminated and random noise following a realistic shape was added to the points to approach them to the real experiment. Four models were applied for simulation and up to five models for fitting.

The analysis of the results shows that with the nowadays state-of-the-art measurement and data treatment techniques, in most of the cases it was possible to distinguish more complex and also more probable bi-ligand and mixed metal–ligand complexes from the simpler 1:1 metal–ligand systems. Statistical evidences to validate the right model were given. Its applicability has been confirmed by generating a similar data mining server (DMS) rule.

© 2005 Elsevier B.V. All rights reserved.

**Keywords:** Voltammetry; Metal–ligand complexing parameters; Multiligand systems; Mono- and bi-ligand complexes; Mixed-ligand(s) complexes; PROSECE; DMS

## 1. Introduction

As written in numerous papers, there has been a lot of work done and many efforts put in measuring metal–ligand interactions in model solutions and natural waters, while it is evident that the metals speciation is a good indicator of their bioavailability and hence, their toxicity [1–7]. Voltammetric methods with standard addition of metal ions of interest have shown to be one of the most non-destructive and subtle techniques, able to recognize metal–ligand complexation by directly measuring the free and labile fraction of

metal and distinguishing free metal and labile from inert metal complexes [8–12]. There are numerous models which deal with metal speciation with natural organic matter that have become sophisticated taking into account more and more components and parameters affecting the metal speciation and conditions when the bioavailable species could be present, as well as the models of natural organic matter description, which is heterogeneous and complex in its nature. They use large database knowledge about thermodynamic stability constants when discrete ligand approach is applied [13–18] or when continuous functions are associated with NOM properties and affinity to metal, major cations or proton [19–23], both needing a verification in experiments of various techniques. By the combination of such

\* Corresponding author. Tel.: +385 1 4561 190; fax: +385 1 4680 231.  
E-mail address: [pizeta@irb.hr](mailto:pizeta@irb.hr) (I. Pižeta).

model—experiment approach, one can advance in characterization of samples of interest and obtain a more global speciation conception. In concrete applications, i.e. experimental verifications, the complexity of the system imposed the need for simplification [24,25], a stepwise approach. Indeed, it has been shown that many situations and experimental data could be explained and covered by a simplified conception of 1:1 metal–ligand complexes with their numerical indications of capacity and conditional complex stability constants [26–33]; on the other hand for a more complex conception there were no means of enough sensitive measurements and calculations, regarding voltammetric method approach [34,35]. When facing the complex structure of the real sample matrix, we always have to bear in mind that from single technique measurements (voltammetric or any other) one can get information about the so called conditional stability constants, which does not diminish the significance of the analyses, but necessitates further interpretation.

Ligands with strong complexing abilities will tend towards making bi-ligand complexes with trace metals, especially in the situations of low concentrations of the metals available [18,24,34,36–46]. In natural water systems, which contain different ligands, mixed metal–ligand complexes are more probable to occur than the pure single metal–ligand ones [47]. Their quantitative characterization is not an easy task and is not often found in the literature. In contrast, theoretical calculations based on the knowledge of simpler systems, analyses of factors affecting the recognition of ternary complexes, and the reviews about particular ligand's behaviour in the environment (e.g. [48–52]) can be found. Mixed metal–ligand complexes were studied in order to improve the sensitivity of the voltammetry as electroanalytical method by multiple enhancement of metal–mixed ligands complexes adsorption onto mercury drop electrode and formation of metal–mixed ligands–surface complex, known as synergetic adsorption [53–55]. Also, this could be a possible type of process (among those well known) for metals removing from aqueous systems. Experimental results on synergetic adsorption phenomena of uranium and copper at the hanging mercury drop electrode have been published [53–55]. If we broaden the problematic to the formation of metal–ligand–surface complexes (surfaces found in natural waters and sediments and surface adsorptions acting as a ligand), there is a growing motivation for the study of ternary complexes [56].

In our previous paper, a detailed analysis of the state-of-the-art in metal complexing parameters determination in natural water systems for 1:1 metal–ligand systems has been given [57]. Starting from these results and using the optimal conditions described, possible to perform with our nowadays voltammetric experimental setup [58] (i.e. non-linear, logarithmic standard additions) we wanted to analyze in more detail the ability of our experimental conditions and analyzing tools to distinguish the situations different from 1:1 metal–ligand situations that might occur in natural water systems, such as one and two 1:2 metal–ligand

systems, as well as mixed ligand systems with the studied metal.

## 2. Simulation of experiment and formation of data sets for fitting

The simulation of experiment, i.e. numerical generation of data points that strictly imitate the distribution and the number of standard additions, that takes into account the existing sensitivity of the available instruments and noise distribution curve, is useful in predicting the recognition of information comprised in real, experimental data points, as shown previously [57,59]. According to the design of experimental standard additions, each point representing the sum of the free and the labile fraction of added metal was calculated by the program MINEQL [60].

The concentration ranges, the number of additions of titrated metal as well as the logarithmic type of additions were taken from the previous results [57], i.e. for one and two ligand cases which were analyzed in this work, the concentration range of added metal from 1 to 457 nmol L<sup>-1</sup> is covered by 25 standard addition points with constant increments of  $pM_T$ , i.e.

$$M_{T,i} = 10^{-(pM_{T,0} + (pM_{T,0} - pM_{T,n}) \times (i/n))}$$

where  $n$  is 25 and  $M_{T,0}$ ,  $M_{T,n}$  and  $M_{T,i}$  are the initial, the final and  $i$ th total metal concentrations, respectively.

This design of titration, named logarithmic, has been the subject of the development of an automatic apparatus, with DPASV measurements of labile trace metal concentrations [61].

Considering the non-uniform repartition of the standard deviation of the measurements obtained on real logarithmic addition titration of cadmium [61], it seemed interesting to try reproducing that type of variation and apply it in this study. The noise added to the simulated data before fitting the best corresponds to:

$$e(M) = \frac{1}{\exp((M/M_\infty) \times 10) - ((e_0 - e_\infty - 1)/(e_0 - e_\infty)) + e_\infty}$$

where  $M$  and  $M_\infty$  are the concentrations of metal added during and at the end of the titration, respectively,  $e_0$  and  $e_\infty$  the noise limits at the beginning and end of the titration, 20 and 1%, respectively. So the added noise depends on metal addition and consists of adding random noise to the data points between  $-e(M)$  and  $+e(M)$ . This process of data modification by noise addition is presented in Fig. 1.

Taking into account that analytical techniques used for the characterization of natural ligand complexing properties, especially when real samples are concerned, can seldom detect the concentration values lower than 0.1 nmol L<sup>-1</sup>, this value was put as a limit to remove all the data points obtained by calculation which are lower than 0.1 nmol L<sup>-1</sup>, i.e. the detection

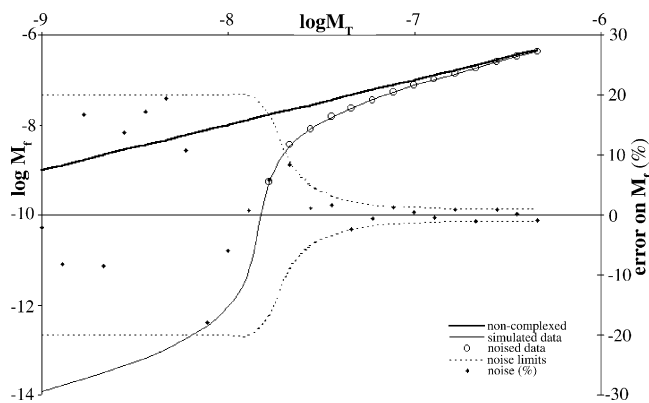


Fig. 1. An example of modification of data simulated by MINEQL after noise addition and removal of the points below  $1 \times 10^{-10} \text{ mol L}^{-1}$ .

limit (DL) =  $1 \times 10^{-10} \text{ mol L}^{-1}$ . So, the number of points of one data set exposed to fitting was not always 25, it came down even to 8 for the experiments where the complexation intensity was the strongest. In Fig. 1, 14 points are left after noise application and detection limit restriction.

The existence of two ligands (L) was simulated, a strong (X) and a weak (Y) of 30 and 300  $\text{nmol L}^{-1}$  concentrations, respectively. The position of the points selected for simulation in  $\log(\beta_1) \times \log(\beta_2)$  space, representing the first and the second stability constants between metal (M) and ligand (L = X or Y) ( $\beta_1 = [\text{ML}]/[\text{M}][\text{L}]$ ;  $\beta_2 = [\text{ML}_2]/[\text{M}][\text{L}]^2$ ), is presented in Fig. 2. These values have been chosen taking into account the results obtained in the literature (or on real systems) [27–29,31–33,57,39–41,62]. The stability constant values chosen for simulation are summarised in Table 1.

Regarding the formation constants of mixed complexes, i.e. MX<sub>2</sub>, it has been shown [50,51] that they should be related to the formation constants of the bi-ligand complexes by a relation of the following kind:  $\beta_{\text{MX}_2} = \omega \times (\beta_{2\text{X}} \times \beta_{2\text{Y}})^{1/2}$ , with  $\omega \geq 2$ . In our analysis, we used three different values of

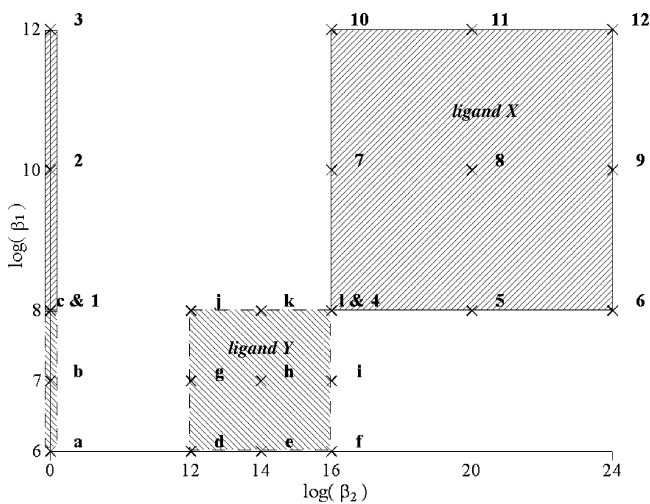


Fig. 2. The legend of distribution of metal first and second stability constants with stronger (X = 30  $\text{nmol L}^{-1}$ ) and weaker (Y = 300  $\text{nmol L}^{-1}$ ) ligand used to simulate the experiments.

Table 1

Values of stability constants of mono-, bi- and mixed complexes used for simulation in various combinations

$\log(\beta_{1\text{X}})$	$\log(\beta_{2\text{X}})$	$\log(\beta_{1\text{Y}})$	$\log(\beta_{2\text{Y}})$	$\log(\beta_{\text{XY}})$
8	16	6	12	14.3, 17.3 and 20.3
10	20	7	14	17.3, 20.3 and 23.3
12	24	8	16	20.3, 23.3 and 26.3

$\omega$  for the same combinations of formation constants as those used for simulation of single complexes, i.e.  $2$ ,  $2 \times 10^3$  and  $2 \times 10^6$  (Table 1).

Four different groups of models were simulated using MINEQL: S1 = ML (for X and Y); S2 = ML + ML<sub>2</sub> (for X and Y); S4 = MX + MX<sub>2</sub> + MY; and S5 = MX + MX<sub>2</sub> + MY + MY<sub>2</sub> + MX<sub>2</sub>. The preset values of complexing parameters used for simulation can be read out from Fig. 2 and from Tables 2–5, respectively. For example: data set 7 stands for  $\log \beta_1 = 10$ ,  $\log \beta_2 = 16$ ,  $L = X = 30 \text{ nmol L}^{-1}$ , the relevant results are found in Table 3.

For each group of simulated points, three forms of data sets were prepared, namely: (a) for all 25 originally calculated data points, (b) for data points after removal of the points lower than the selected determination limit DL, in our case  $\text{DL} = 1 \times 10^{-10} \text{ M}$  (number of data points diminished up to 8), and (c) for data points obtained after removal of data <DL and application of the random noise.

### 3. Fitting strategy

Also resulting from our previous paper [57], the program PROSECE and its complete procedure was applied in all the fittings. Besides its proven suitability shown in the previous analyses, its structure is much more flexible and able to adapt itself to various fitting models used in this study compared to the other programs that could have been applied instead [59]. Its drawback is, however, the lack of fitted parameters confidence limits estimate.

In order to limit the total number of fittings, which in such analyses easily becomes too big and the obtained results not easy to elaborate, the schedule of fitting generally comprised the models similar or simpler than those used to simulate the data sets. Altogether five different fitting models were applied to data sets, namely, F1 = ML; F2 = ML + ML<sub>2</sub>; F3 = MX + MY; F4 = MX + MX<sub>2</sub> + MY and F5 = MX + MX<sub>2</sub> + MY + MY<sub>2</sub> + MX<sub>2</sub>.

In the case of the “right” (matching) fitting model with the simulation model, one can discuss the errors of fitting, i.e. the differences between the preset and the obtained metal–ligand complexing parameters, but since in general we do not know which model is the right one, we should pay more attention to the other parameters of fitting, such as the amount and the shape of the residual of fitting. The parameter that also seemed to be interesting and drawing attention when analyzing the residuals was the number of zero crossings of the residual.

Table 2

Results obtained with PROSECE fitting of the six S1-type experiments using F1, F2 and F3 complexing models on c-type of data sets ( $n$ : number of simulated points after removal of points <DL,  $L_T$ : ligand concentrations in  $\text{nmol L}^{-1}$ ,  $\text{res}_n$ : normalized residuals of PROSECE fitting)

Ligand	S1 sim.	Preset values			F1* = ML			F2 = ML + ML <sub>2</sub>				F3 = MX + MY				
		$\log(\beta_1)$	$L_T$	$n$	$\log(\beta_1)$	$L_T$	$\text{res}_n$	$\log(\beta_1)$	$\log(\beta_2)$	$L_T$	$\text{res}_n$	$\log(\beta_{1X})$	$X_T$	$\log(\beta_{1Y})$	$Y_T$	$\text{res}_n$
X	1	8	30	25	8.0	30.0	0.951	8.0	12.7	30.2	0.952	8.0	30.0	4.2	245	0.949
	2	10	30	14	9.9	30.0	0.120	10.0	15.0	30.0	0.124	10.0	30.0	5.1	6.0	0.119
	3	12	30	11	12.1	30.0	0.173	11.5	17.6	30.0	0.173	12.7	30.0	-0.5	106	0.173
Y	a	6	300	25	6.0	288	0.600	5.5	11.7	550	0.598	6.2	19.7	6.0	271	0.600
	b	7	300	25	7.0	298	0.740	6.9	12.4	319	0.740	7.2	19.7	7.0	280	0.740
	c	8	300	20	8.0	304	0.994	8.0	9.4	304	0.994	8.0	18.5	8.0	285	0.994

Table 3

Results obtained with PROSECE fitting of the 18 S2-type experiments using F1, F2 and F3 complexing models on c-type of data sets; other parameters are the same as in Table 2

S2 sim.	Preset values				F1 = ML			F2* = ML + ML <sub>2</sub>				F3 = MX + MY				
	$\log(\beta_1)$	$\log(\beta_2)$	$L_T$	$n$	$\log(\beta_1)$	$L_T$	$\text{res}_n$	$\log(\beta_1)$	$\log(\beta_2)$	$L_T$	$\text{res}_n$	$\log(\beta_{1X})$	$X_T$	$\log(\beta_{1Y})$	$Y_T$	$\text{res}_n$
4	8	16	30	24	8.8	17	1.612	8.2	16.0	26.6	0.848	8.8	16.3	5.5	279	1.129
5	8	20	30	14	10.8	15	0.185	10.8	17.8	14.9	0.185	10.7	14.9	2.4	292	0.185
6	8	24	30	14	16.0	15	0.248	18.2	19.4	15.0	0.248	14.7	15.0	-63.0	1112	0.248
7	10	16	30	14	10.0	30	0.453	10.0	14.5	30.2	2.137	10.0	30.0	3.0	0	0.453
8	10	20	30	14	10.0	19	1.667	11.5	20.9	19.7	1.483	10.3	17.9	5.6	197	0.923
9	10	24	30	14	16.5	15	0.326	19.5	11.3	15.2	0.326	15.1	15.2	-14.4	808	0.326
10	12	16	30	11	10.7	30	0.157	10.6	9.1	29.8	0.156	11.1	29.7	4.3	123	0.156
11	12	20	30	14	16.4	30	0.123	20.1	8.9	29.9	0.123	11.2	29.9	5.8	11	0.109
12	12	24	30	14	10.0	20	1.631	11.2	20.7	20.9	1.394	10.1	18.7	5.6	200	0.908
d	6	12	300	25	6.3	216	0.351	6.2	11.4	227	0.351	6.7	22.1	4.2	202	0.353
e	6	14	300	24	7.9	110	1.678	6.8	14.2	238	1.114	8.4	23.4	5.9	203	1.171
f	6	16	300	10	9.5	140	3.918	9.5	16.5	142	1.029	8.2	26.5	5.7	206	1.470
g	7	12	300	25	7.1	282	0.837	6.9	12.7	322	0.784	7.6	18.9	5.1	199	0.777
h	7	14	300	24	7.9	156	1.868	7.4	14.1	220	1.124	8.6	7.9	6.1	188	1.513
I	7	16	300	10	9.6	132	3.444	9.6	16.1	135	2.387	7.8	37.3	5.3	217	0.944
j	8	12	300	20	8.0	300	1.360	8.0	9.6	300	1.360	8.0	24.0	5.5	204	1.361
k	8	14	300	19	8.1	276	1.623	2.8	13.8	755	1.277	8.8	21.0	6.3	201	1.222
l	8	16	300	9	9.4	158	5.398	9.2	16.4	184	2.373	6.6	20.1	4.1	200	4.886

Table 4

Results obtained with PROSECE fitting of the 15 S4-type experiments using the following fitting models: F1 = ML, F2 = ML + ML<sub>2</sub>, F3 = MX + MY and F4 = MX + MX<sub>2</sub> + MY; other parameters are the same as in Table 2

S4 sim.	Preset values						F1, $\text{res}_n$	F2, $\text{res}_n$	F3, $\text{res}_n$	F4*					
	$\log(\beta_{1X})$	$\log(\beta_{2X})$	$X_T$	$\log(\beta_{1Y})$	$Y_T$	$n$				$\text{res}_n$	$\log(\beta_{1X})$	$\log(\beta_{2X})$	$X_T$	$\log(\beta_{1Y})$	$Y_T$
a+4	8	16	30	6	300	24	3.716	2.233	0.914	0.732	8.5	15.9	19.8	6.1	256
a+6	8	24	30	6	300	14	4.059	2.954	0.170	1.166	14.0	27.5	21.6	5.7	475
a+8	10	20	30	6	300	14	5.877	5.833	0.419	0.428	10.2	16.5	18.0	6.4	186
a+10	12	16	30	6	300	11	3.366	3.351	0.233	0.235	11.4	18.1	30.4	6.0	302
a+12	12	24	30	6	300	14	5.797	5.758	0.410	0.419	10.3	17.1	18.3	6.4	185
b+4	8	16	30	7	300	23	5.956	5.955	1.158	0.911	8.7	15.9	16.8	7.0	317
b+6	8	24	30	7	300	14	6.910	6.910	0.337	0.844	13.7	27.4	27.2	6.9	297
b+8	10	20	30	7	300	14	8.875	7.764	0.608	0.593	11.4	13.3	15.9	7.0	302
b+10	12	16	30	7	300	11	5.767	5.767	0.917	0.516	11.2	11.3	30.8	7.0	306
b+12	12	24	30	7	300	14	9.279	9.279	0.348	0.348	10.7	12.8	17.6	7.0	317
c+4	8	16	30	8	300	19	1.231	1.231	1.062	1.126	8.0	16.6	12.6	8.0	318
c+6	8	24	30	8	300	13	4.594	4.594	3.958	0.931	11.6	9.4	16.0	8.0	304
c+8	10	20	30	8	300	13	4.592	4.592	4.313	0.743	15.1	7.9	14.9	8.0	309
c+10	12	16	30	8	300	11	6.771	6.771	6.566	0.761	13.8	-2.1	30.4	8.0	299
c+12	12	24	30	8	300	13	4.993	4.993	4.693	1.061	10.4	11.6	19.3	8.0	307

Table 5

Results obtained with PROSECE fitting of the 15 S5-type experiments using the following fitting models: F1 = ML, F2 = ML + ML<sub>2</sub>, F3 = MX + MY, F4 = MX + MX<sub>2</sub> + MY and F5 = MX + MX<sub>2</sub> + MY + MY<sub>2</sub> + MXY; other parameters are the same as in Table 2

S5 sim.	$\omega$	$n$	F1, res <sub>n</sub>	F2, res <sub>n</sub>	F3, res <sub>n</sub>	F4, res <sub>n</sub>	F5*							
							res <sub>n</sub>	log( $\beta_{1X}$ )	log( $\beta_{2X}$ )	X <sub>T</sub>	log( $\beta_{1Y}$ )	log( $\beta_{2Y}$ )	log( $\beta_{XY}$ )	Y <sub>T</sub>
d+4	2	23	4.356	2.708	1.100	0.930	0.908	8.5	16.1	23.9	5.8	12.1	12.2	368
d+4	2 × 10 <sup>3</sup>	12	4.501	4.501	0.247	0.406	0.924	9.0	18.0	27.4	6.3	12.9	20.1	188
d+4	2 × 10 <sup>6</sup>	11	3.464	3.422	0.211	0.191	0.236	9.3	24.6	30.4	6.3	11.2	29.5	211
f+6	2	8	2.984	2.984	0.694	1.612	0.576	7.3	14.7	19.7	9.6	16.6	23.5	156
f+6	2 × 10 <sup>3</sup>	8	3.547	3.547	2.411	2.732	0.795	7.2	15.3	18.2	9.7	16.8	19.6	155
f+6	2 × 10 <sup>6</sup>	8	2.656	2.656	1.393	0.595	0.765	5.9	18.7	16.8	9.6	16.5	18.3	152
h+8	2	13	9.316	9.316	1.085	1.305	1.001	8.8	16.1	19.2	7.6	14.3	19.2	220
h+8	2 × 10 <sup>3</sup>	11	6.013	6.013	1.008	3.567	0.475	9.6	16.3	29.8	7.4	14.0	18.9	237
h+8	2 × 10 <sup>6</sup>	11	6.504	6.504	3.604	1.344	0.395	8.8	12.0	29.6	7.2	14.1	17.9	264
j+10	2	11	7.084	7.084	6.883	0.545	0.560	4.4	4.6	31.0	8.0	12.3	26.4	332
j+10	2 × 10 <sup>3</sup>	11	7.015	7.015	6.834	0.748	0.917	7.1	15.2	29.4	8.0	7.5	19.9	331
j+10	2 × 10 <sup>6</sup>	11	7.037	7.037	6.709	0.472	0.443	8.6	18.0	30.9	8.0	9.0	19.1	303
l+12	2	8	8.021	7.900	1.957	8.088	1.969	8.3	15.1	13.8	9.1	16.6	18.6	217
l+12	2 × 10 <sup>3</sup>	8	6.697	6.697	5.035	5.609	1.660	7.3	16.2	16.9	9.2	16.7	17.1	211
l+12	2 × 10 <sup>6</sup>	8	7.051	7.051	5.634	6.879	1.721	7.5	14.4	22.1	9.2	16.8	17.0	207

#### 4. Results and discussion

Before we start the fitting procedure that should comprise several models, we should try to globally estimate the results by watching the shape of the data expressed in term of  $M_f$  versus  $M_T$ ,  $pM_f$  versus  $pM_T$  and linearization modes of Scatchard [63] and/or Ružić–Van den Berg [35,64–66], where  $M_f$  stands for concentration of metal (free and labile) measured by voltammetry and  $M_T$  for total metal added in standard addition procedure. That could give us the idea of the kind of fitting model to be used, or at least to discard some of the models. In the first presentation ( $M_f$  versus  $M_T$ ), one can try to estimate the total ligand capacity by noticing where the tangent to the data points intercepts X-axis, suiting the Chau method of linearization [67]. In the second presentation ( $pM_f$  versus  $pM_T$ ), if there is more than one inflection, and in the third presentation (linearization) if there is an indication of curvature of the data points, we can be sure that the structure is more complicated than 1:1 metal–ligand complexation.

The comparison of the three forms of data set results shows that difference is bigger between (a) and (b) than between (b) and (c). This is promising since the error function applied to the simulated data points is arbitrary in any case. Also, the detection limit of  $10^{-10}$  mol L<sup>-1</sup>, used to remove the data points was rather severe; it could be expected to be lowered up to two orders of magnitude, depending on the analytical technique used. It has to be noticed that noise insertion was not that important for the final results concerning the goodness of fit.

As only the (c) case represents possible simulation of the experiment and as usually we do not know which model for fitting the experimental data is the right one, the most valuable information about the fit we have, besides logical parameter values, i.e. positive numbers for concentration and stability constants values, is the value of the residuals

(res = sum(abs(ln( $C_{calc}$ ) – ln( $C_{sim}$ )))), which is the sum of absolute values of differences between the natural logarithms of calculated and simulated data points. The values of residuals are normalised to the number of data points in order to allow the comparison of the results for data sets of various lengths. It is interesting to compare consequent values obtained for the same set of simulated data by application of different fitting models.

The normalized residuals of all fittings for the data sets (c) (data points obtained after removal of data <DL and application of the random noise) of the four groups of simulated S1, S2, S4 and S5 models fitted by up to five F1–F5 fitting models, are shown in Fig. 3 and will be discussed separately. In general, at first sight, for the S1 and S2 cases the residuals are similar for all fittings, in which case, with the help of statistics we decide whether to accept a more complicated model or stay to the simpler one. For the S4 and S5 cases, we can notice that the application of the right model can in average be distinguished from the wrong ones by comparing the normalized residual values. In the tables the matching model is marked with an asterisk (\*).

##### 4.1. One-ligand systems (S1 and S2 simulations)

One-ligand system was presented with both ligands, L = X and Y, and their corresponding stability constants for mono-ligand as well as bi-ligand cases. Each of the simulated data sets was fitted by three different models, F1 = ML, F2 = ML + ML<sub>2</sub> and F3 = MX + MY.

###### 4.1.1. S1 simulation

Six different data sets were considered, three for stronger (X) and three for weaker (Y) ligands; the preset values of complexing parameters, capacity and stability constants are given in Table 2 and can also be read out from Fig. 2. The results of (a) and (b) fittings with ML model were rather



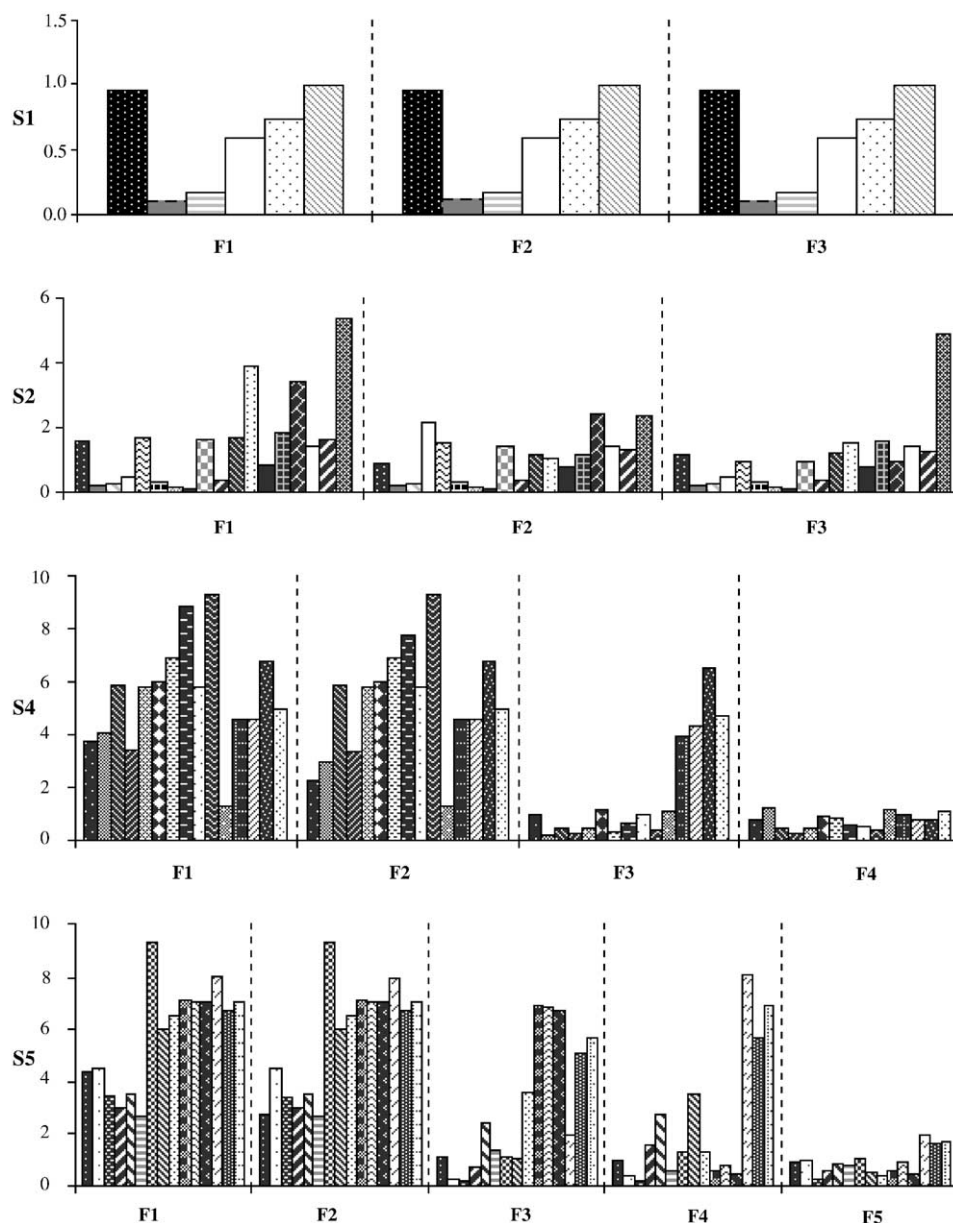


Fig. 3. Residuals of fitting for data sets obtained after removal of data below the DL and application of random noise. S for simulation and F for fitting models: S1 = F1 = ML; S2 = F2 = ML + ML<sub>2</sub>; F3 = MX + MY; S4 = F4 = MX + MX<sub>2</sub> + MY; S5 = F5 = MX + MX<sub>2</sub> + MY + MY<sub>2</sub> + MYX. Legends correspond to tables notation and are named from left to right for each F: S1: 1, 2, 3, a, b and c; S2: 4, 5, 6, 7, 8, 9, 10, 11, 12, d, e, f, g, h, i, j, k and l; S4: a + 4, a + 6, a + 8, a + 10, a + 12, b + 4, b + 6, b + 8, b + 10, b + 12, c + 4, c + 6, c + 8, c + 10 and c + 12; S5: d + 4 + ω = 2, d + 4 + ω = 2 × 10<sup>3</sup>, d + 4 + ω = 2 × 10<sup>6</sup>, f + 6 + ω = 2, f + 6 + ω = 2 × 10<sup>3</sup>, f + 6 + ω = 2 × 10<sup>6</sup>, h + 8 + ω = 2, h + 8 + ω = 2 × 10<sup>3</sup>, h + 8 + ω = 2 × 10<sup>6</sup>, j + 10 + ω = 2, j + 10 + ω = 2 × 10<sup>3</sup>, j + 10 + ω = 2 × 10<sup>6</sup>, l + 12 + ω = 2, l + 12 + ω = 2 × 10<sup>3</sup> and l + 12 + ω = 2 × 10<sup>6</sup>.

straightforward, almost perfectly matching the preset values, and are not surprising and to be discussed. The results of the simulated points fitting after removal of <DL points and noise insertion with three different models are shown in Table 2. Concerning normalized residuals (residuals divided by number of points), all the three fittings have similar residuals. In that case the simplest model should be accepted, which is the right decision according to the *F*-test [68]. It is interesting that for F2 model, the second constant  $\beta_2$  has just been figured out by the fitting program, with almost no influence on

the result of total ligand (except for the Sim. a case). The F3 model applied on strong ligands gives unacceptable results, but when applied on weaker ligands it just divides the total ligand in two groups of almost equal constants, which points that the decision of accepting the F1 model as the right one would be the correct choice.

#### 4.1.2. S2 simulation

Eighteen different data sets were simulated, 9 for stronger (X) and 9 for weaker (Y) ligands. The results of data fitting

from the (c) cluster with the same three models as in the S1 case are shown in Table 3. Looking only to the residuals provides no enough evidence to determine which model for fitting is the right one. On the contrary, for the set of stronger (X) ligands, lower residuals were obtained with a model offering two, rather than one ligand with two constants for mono- and bi-ligand complexes. In the group of S2 simulations, there is expressed the strongest impact of passing from fitting of all 25 data points (a), to fitting of data points without those eliminated by the DL (b) and fitting of noisy data (c). By analyzing only the (c) cluster, one can hardly decide about the right model and consequently about the values of the complexing parameters. In such a case the answer lies in redesigning the experiments either by measuring more replicates or refining and increasing the number of standard additions in order to increase the detection limit and lower the noise.

#### 4.2. Two-ligand systems (S4 and S5 simulations)

Two-ligand systems were presented with the mixtures of ligands X and Y with the metal M and stability constants for mono-, bi-ligand cases as well as mixed metal–ligand complexes. Each of the simulated data sets was fitted by four or five various models, F1 = ML, F2 = ML + ML<sub>2</sub>, F3 = MX + MY, F4 = MX + MX<sub>2</sub> + MY and F5 = MX + MX<sub>2</sub> + MY + MY<sub>2</sub> + MX<sub>2</sub>Y.

##### 4.2.1. S4 simulation

Fifteen different data sets were simulated following the S4 model, i.e. there is one stronger (X) ligand that forms mono- and bi-ligand–metal complexes and one weaker (Y) ligand that forms mono complex. According to their simulation names given in Table 4 and Fig. 2, one can read out their ligand concentrations and stability constants used. In Table 4, the obtained fitting parameters for the matching model as well as the values of the residuals for all the models applied are given together with the number of points of each data set for the (c) cluster.

In 9 out of the 15 cases the normalized residual was the smallest for the matching model. The weaker ligand concentration was retrieved with better score than the stronger one, with less than 10% of error, 11 out of the 15 and 3 out of the 15, respectively. Stability constant of the weaker ligand was well retrieved in all the 15 cases, while for the stronger ligand this happened in only 3 out of the 15 cases. Those positive cases were characterized at the same time with the number of points considerably high, i.e. the constants were the lowest of the proposed constant sets.

##### 4.2.2. S5 simulation

Fifteen different data sets were simulated following the S5 model, i.e. there is one stronger ligand that forms mono- and bi-ligand–metal complexes, one weaker ligand that forms mono- and bi-ligand–metal complexes, and there is one mixed-ligand complex. According to their simulation names

given in Table 5, Fig. 2 and Table 1, one can read out their ligand concentrations and stability constants used. In Table 5 the obtained fitting parameters for the matching model as well as the values of the residuals for all the models applied are given together with the number of points of each data set for the (c) cluster. In 10 out of the 15 cases the normalized residual was the smallest for the matching model, which is a better score than in the case of S4 simulation. However, the values of complexing parameters were estimated with higher tolerances. Ligand concentrations were underestimated, weaker ligands in higher extent than the stronger ones, in more than half of the cases. From five of the simulated stability constants the best retrieved constants after noising procedure were  $\beta_{1Y}$  and  $\beta_{2Y}$ , the constants for weaker mono- and bi-ligand–metal complexes. It is evident that for such rather complicated model, broader range of values of simulated parameters should be selected and the simulated data consequently treated in order to survey a specific rule and/or a range of more certain recognition of preset parameters.

Starting from these results, two characteristic cases were studied in more detail, i.e. the case d + 4,  $\omega = 2$  and the case f + 6,  $\omega = 2 \times 10^6$ . Three-fold repetition of the experiments was simulated in order to test its influence on the correct retrieving of the parameters. The results given in Table 6 brought to the conclusion that measurement repeating in great extent eliminates the influence of the noise (seen from the comparison of the corresponding (b) and (c) set of results), but cannot compensate for the missing of the data points <DL (comparison of a-type of data with b- and c-types in Table 6), which in fact carry the major part of information about the strong complex parameter values.

#### 4.3. Data mining server rules

Handling with a big quantity of calculated data, we tried to possibly find a general rule that could help us recognize the right model of fitting, in our case simulated data, but in general, measured data of some unknown samples.

To that purpose we have prepared our data according to the rules of the data mining server (DMS) [69], where data analysis is performed based on knowledge induction by the inductive learning by logic minimization (ILLM) system. Using known data sets and their classification this program searches for classification rules that could be applied to the new data sets of the same kind. An attempt of mixing (a), (b) and (c) sets of data did not give any meaningful result, so we restricted ourselves to the (c) set. A table with 239 example rows (containing all the mentioned sets of simulated data fitted to all the mentioned models) was constructed, 70 of them was found in the positive class, which in our case was a matching simulation and fitting model. Defined attributes were all preset parameter values of metal complexing capacity and their mutual differences and products, the number of data points in a particular set, the results of fitting including the amount of residual of fitting, and the residual normalized



Table 6

Results obtained with PROSECE fitting of two S5-type experiments using the fitting model F5; type: data type—see text; av: average of the three repetitions of the same simulated point with different random noise; other parameters are the same as in Table 2

Sim.	Fitting model	<i>n</i> Preset values	Type	<i>n</i> of meas.	res <sub>n</sub>	log( $\beta_{1X}$ ) 8.0	log( $\beta_{2X}$ ) 16.0	log( $\beta_{1Y}$ ) 6.0	log( $\beta_{2Y}$ ) 12.0	log( $\beta_{XY}$ ) 14.3	X <sub>T</sub> 30.0	Y <sub>T</sub> 300
d + 4 $\omega$ = 2	F5*	25	a	1	0.194	8.7	16.1	6.0	12.7	14.4	17.4	277
	F5*	23	b	1	0.056	8.4	16.1	6.0	12.5	14.6	21.6	286
	F5*	23	c	1	0.908	8.5	16.1	5.8	12.1	12.2	23.9	368
	F5*	69	c	3	1.066	8.4	16.1	5.9	12.2	14.2	24.0	341
	F5*	23 (av)	c	3	0.623	8.5	16.1	5.7	12.2	8.4	23.9	383
Sim.	Fitting model	<i>n</i> Preset values	Type	<i>n</i> of meas.	res <sub>n</sub>	log( $\beta_{1X}$ ) 8.0	log( $\beta_{2X}$ ) 24.0	log( $\beta_{1Y}$ ) 6.0	log( $\beta_{2Y}$ ) 16.0	log( $\beta_{XY}$ ) 26.3	X <sub>T</sub> 30.0	Y <sub>T</sub> 300
f + 6 $\omega$ = 2 × 10 <sup>6</sup>	F5*	25	a	1	3.071	−3.6	25.2	5.4	16.1	26.3	31.2	264
	F5*	8	b	1	0.583	6.7	16.6	9.6	16.7	24.2	17.5	156
	F5*	8	c	1	0.765	5.9	18.7	9.6	16.5	18.3	16.8	152
	F5*	24	c	3	1.326	7.1	15.9	9.6	16.7	20.9	19.9	157
	F5*	8 (av)	c	3	0.663	7.2	16.1	9.6	16.7	21.6	19.2	157

to the number of data points (res<sub>n</sub>) as well as the number of zero crossings of the residual.

A reasonable result is a rule obtained by the generalization parameter  $g = 100$ , which claims: fitting is good if res<sub>n</sub> < 2.53. This rule has a sensitivity of 100% and a specificity of 53.9%. Such a result was a confirmation of classical statistical approach to the fitting problem where the analysis of residuals is provided.

Another table (with 75 example rows, 15 out of which were positive, i.e. matching simulation and fitting model) was prepared with the examples of only S5, fitted with all the F1–F5 fitting models. A reasonable result is a rule obtained by the generalization parameter  $g = 20$ , which claims fitting is good if res<sub>n</sub> < 2.19 (sensitivity of 100% and specificity of 72.4%). When compared to the first table of simulated and fitting models all together, the second table with only one simulated model (the most complex in our analysis), shows more specificity for less generalization, which was expectable. Furthermore, the second table is a good model of a real experiment data treatment. By applying this rule, we could count on selecting all (100%) the matching models and 27.6% of the wrong models. Further combining with other disposable evidences such as critical logical inspection of the parameters values and error distribution (e.g. number of zero crossing of the residuals) could eventually bring us higher specificity.

## 5. Conclusion

The analyzed web of simulated experiments presuming and modelling the existence of more complex and probable bi-ligand and mixed metal–ligand complexes in aquatic systems, showed that with the nowadays state-of-the-art in measurement and data treatment techniques, it is possible to distinguish them from the simpler 1:1 metal–ligand systems in most of the analyzed cases.

This study showed that by means of voltammetric instrumentation supplied with automatic burettes and sophisticated data treatment that enables quick scanning through different fitting models, it is possible to conclude more precisely about the metal–ligand structure in a measured sample. However, for more accurate values of the complexing parameters, a more detailed study should be undertaken including repeating of experiments, and/or putting denser metal additions (in cases of strong ligands when larger number of first additions are so complexed to leave the free metal below the detection limit), considering that rather a subtle structure of dissolved metal–ligand complexes is tried to be distinguished.

Analyses by the method of data mining server confirmed the statistical methods of residuals analysis and have shown to be applicable in such simulation—fitting modelling systems as a prediction tool and help as one of the criteria in data interpretation. A more specific rule is to be expected with a denser web of modelled parameters, which could be the aim of some future work.

## Acknowledgement

The authors wish to thank the Ministry of Science, Education and Sport of the Republic of Croatia and the Ministry of National Education, Teaching and Research of France. Also we thank the Ph.D. school of the University of Sud Toulon Var for financial support of collaboration between France and Croatia and D. Gamberger for his help in explanation and application of DMS rules. We would like to thank two anonymous reviewers for their comments and suggestions.

## References

- [1] K.W. Bruland, E.L. Rue, J.R. Donat, S.A. Skrabal, J.W. Moffett, Anal. Chim. Acta 405 (2000) 99.

- [2] J.P. Gustafsson, P. Pechová, D. Berggren, *Environ. Sci. Technol.* 37 (2003) 2767.
- [3] B. Koukal, C. Guéguen, M. Pardos, J. Dominik, *Chemosphere* 53 (2003) 953.
- [4] J.I. Lorenzo, O. Nieto, R. Beiras, *Aquat. Toxicol.* 58 (2002) 27.
- [5] S.E. Mylon, B.S. Twining, N.S. Fisher, G. Benoit, *Environ. Sci. Technol.* 37 (2003) 1261.
- [6] V.I. Slaveykova, K.J. Wilkinson, A. Ceresa, E. Pretsch, *Environ. Sci. Technol.* 37 (2003) 1114.
- [7] M.T.S.D. Vasconcelos, M.F.C. Leal, *Mar. Chem.* 75 (2001) 123.
- [8] J. Heyrovský, J. Kůta, *Principles of Polarography*, Academic Press, New York, 1966.
- [9] Z. Galus, *Fundamentals of Electrochemical Analysis*, Ellis Horwood, Chichester, 1976.
- [10] A.M. Bond, *Modern Polarographic Methods in Analytical Chemistry*, Marcel Dekker, New York, 1980.
- [11] A.J. Bard, L.R. Faulkner, *Electrochemical Methods. Fundamentals and Applications*, 2nd ed., Wiley, New York, 2000.
- [12] R.M. Town, H. Emons, J. Buffle, in: R. Cornelis, et al. (Eds.), *Speciation Analysis by Electrochemical Methods, Handbook of Elemental Speciation: Techniques and Methodology*, Wiley, 2003.
- [13] Y. Lu, H.E. Allen, *Water Res.* 36 (2002) 5083.
- [14] J. Masini, G. Abate, E. Lima, L. Hahn, M. Nakamura, J. Lichtig, H. Nagatomy, *Anal. Chim. Acta* 364 (1998) 223.
- [15] G. Sposito, *Environ. Sci. Technol.* 15 (1981) 396.
- [16] G. Sposito, *Chemical Equilibria and Kinetics in Soils*, Oxford University Press, 1994, 268 pp.
- [17] E. Tipping, M.A. Hurley, *Geochim. Cosmochim. Acta* 56 (1992) 3627.
- [18] E. Tipping, *Colloids Surf. A* 73 (1993) 117.
- [19] M.F. Benedetti, W.H. Van Riemsdijk, L.K. Koopal, D.G. Kinniburgh, D.C. Goody, C.J. Milne, *Geochim. Cosmochim. Acta* 60 (1996) 2503.
- [20] J.B. Christensen, T.H. Christensen, *Water Res.* 34 (2000) 3743.
- [21] D.G. Kinniburgh, W.H. Van Riemsdijk, L.K. Koopal, M. Borkovec, M.F. Benedetti, M.J. Avena, *Colloids Surf.* 151 (1999) 147.
- [22] J. Milne, D.G. Kinniburgh, E. Tipping, *Environ. Sci. Technol.* 35 (2001) 2049.
- [23] J.D. Ritchie, J.M. Perdue, *Geochim. Cosmochim. Acta* 67 (2003) 85.
- [24] J. Buffle, *Analytical Chemistry*, Ellis Horwood, 1988, 692 pp.
- [25] D. Dzombak, W. Fish, F.L.L. Morel, *Environ. Sci. Technol.* 20 (1986) 669.
- [26] B.T. Hart, *Environ. Technol. Lett.* 3 (1981) 95.
- [27] P.B. Kozelka, K.W. Bruland, *Mar. Chem.* 60 (1998) 267.
- [28] F.L.L. Muller, *Mar. Chem.* 52 (1996) 245.
- [29] F.L.L. Muller, *Mar. Chem.* 67 (1999) 43.
- [30] T. Rozan, B. Gaboury, H. Marsh, Y. Chin, *Environ. Sci. Technol.* 33 (1999) 1766.
- [31] R.M. Town, M. Filella, *Limnol. Ocean.* 45 (2000) 1341.
- [32] M.L. Wells, P.B. Kozelka, K.W. Bruland, *Mar. Chem.* 62 (1998) 203.
- [33] H. Xue, L. Sigg, *Anal. Chim. Acta* 363 (1998) 249.
- [34] R.H. Byrne, *Mar. Chem.* 12 (1983) 15.
- [35] I. Ružić, *Anal. Chim. Acta* 140 (1982) 99.
- [36] R.H. Byrne, *Mar. Chem.* 9 (1980) 75.
- [37] V. Cuculić, I. Pižeta, M. Branica, *Electroanalysis*, in press.
- [38] D.S. Smith, J.R. Kramer, *Anal. Chim. Acta* 416 (2000) 211.
- [39] E. Tipping, *Geochim. Cosmochim. Acta* 66 (2002) 3211.
- [40] Y. Dudal, F. Gérard, *Earth-Sci. Rev.* 66 (2004) 199.
- [41] D.S. Smith, R.A. Bell, J.R. Kramer, *Comp. Biochem. Physiol. C: Pharmacol. Toxicol.* 133 (2002) 65.
- [42] D.G. Kinniburgh, C.J. Milne, M.F. Benedetti, J.P. Pinheiro, J. Filius, L.K. Koopal, W.H. Van Riemsdijk, *Environ. Sci. Technol.* 30 (1996) 1687.
- [43] J.B. Christensen, J.J. Botma, T.H. Christensen, *Water Res.* 33 (1999) 3231.
- [44] K. Hirose, *Anal. Chim. Acta* 284 (1994) 621.
- [45] C. Huber, M. Filella, R.M. Town, *Comput. Geosci.* 28 (2002) 587.
- [46] H. Xue, L. Sigg, *Aquat. Geochem.* 5 (1999) 313.
- [47] R.H. Byrne, W.L. Miller, in: M.L. Sohn (Ed.), *Organic Marine Geochemistry*, ACS Symposium Series, vol. 305, 1986, 358 pp.
- [48] R. Djogić, M. Branica, *Chem. Speciation Bioavailability* 5 (1993) 101.
- [49] B. Nowack, *Environ. Sci. Technol.* 36 (2002) 4009.
- [50] M. Zelić, *Anal. Chim. Acta* 271 (1993) 275.
- [51] M. Zelić, *Anal. Chim. Acta* 281 (1993) 435.
- [52] M. Zelić, *Electroanalysis* 7 (1995) 350.
- [53] V. Cuculić, M. Mlakar, M. Branica, *Anal. Chim. Acta* 339 (1997) 181.
- [54] M. Mlakar, M. Branica, *J. Electroanal. Chem.* 256 (1988) 269.
- [55] M. Mlakar, M. Branica, *Mar. Chem.* 46 (1994) 61.
- [56] A.P. Davis, *Adsorption of Metal Complexes at Oxide and Related Surfaces*, *Encyclopedia of Surface and Colloid Science*, Marcel Dekker, 2002, pp. 440–449.
- [57] C. Garnier, I. Pižeta, S. Mounier, J.Y. Benaïm, M. Branica, *Anal. Chim. Acta* 505 (2004) 263.
- [58] C. Garnier, S. Mounier, J.Y. Benaïm, *Environ. Technol.* 25 (2004) 589.
- [59] I. Pižeta, M. Branica, *Anal. Chim. Acta* 351 (1997) 73.
- [60] J.C. Westall, J.L. Zachary, F.M.M. Morel, *MINEQL: a computer program for the calculation of chemical equilibrium composition of aqueous systems*, Technical Note No. 18, Ralph M. Parsons Laboratory, Department of Civil Engineering, MIT, Cambridge, MA, 1976.
- [61] C. Garnier, Ph.D. Thesis, University of Sud Toulon Var, France, 2004, 200 pp.
- [62] R.M. Town, M. Filella, *Sci. Total Environ.* 300 (2002) 143.
- [63] G. Scatchard, *Anal. NY Acad. Sci.* 57 (1949) 660.
- [64] C.M.G. Van den Berg, *Mar. Chem.* 11 (1982) 307.
- [65] C.M.G. Van den Berg, *Mar. Chem.* 11 (1982) 323.
- [66] C.M.G. Van den Berg, *Mar. Chem.* 15 (1984) 1.
- [67] Y.K. Chau, K. Lum-Shue-Chan, *Water Res.* 8 (1974) 383.
- [68] J.N. Miller, J.C. Miller, *Statistics and Chemometrics for Analytical Chemistry*, Pearson Education, Essex, England, 2000, 52 pp.
- [69] D. Gamberger, T. Šmuc, *Data Mining Server* [<http://dms.irb.hr/>], Rudjer Boskovic Institute, Laboratory for Information Systems, Zagreb, Croatia, 2001.