Marine Chemistry xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Marine Chemistry

journal homepage: www.elsevier.com/locate/marchem

Interpretation of complexometric titration data: An intercomparison of methods for estimating models of trace metal complexation by natural organic ligands

I. Pižeta [a], S.G. Sander [b], R.J.M. Hudson [c,*], D. Omanović [a], O. Baars [d], K.A. Barbeau [e], K.N. Buck [f,1], R.M. Bundy [e], G. Carrasco [g,h], P.L. Croot [i], C. Garnier [j], L.J.A. Gerringa [k], M. Gledhill [l,m], K. Hirose [n], Y. Kondo [o], L.M. Laglera [p], J. Nuester [q], M.J.A. Rijkenberg [k], S. Takeda [o], B.S. Twining [q], M. Wells [b,2]

[a] Division for Marine and Environmental Research, Ruđer Bošković Institute, 10000 Zagreb, Croatia
[b] NIWA/University of Otago Research Centre for Oceanography, Department of Chemistry, University of Otago, Dunedin 9054, New Zealand
[c] Department of Natural Resources and Environmental Sciences, University of Illinois, Urbana, IL 61801, United States
[d] Department of Geosciences, Princeton University, Princeton, NJ 08544, United States
[e] Geosciences Research Division, Scripps Institution of Oceanography, La Jolla, CA 92093, United States
[f] Bermuda Institute of Ocean Sciences, St. George's GE01, Bermuda
[g] Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, United States
[h] Department of Ocean, Earth and Atmospheric Sciences, Old Dominion University, Norfolk, VA 23529, United States
[i] Department of Earth and Ocean Sciences, National University of Ireland-Galway, Galway, Ireland
[j] RCMO-PROTEE Laboratory, Université de Toulon, 83957 La Garde, France
[k] Royal Netherlands Institute for Sea Research, 1790 AB Den Burg, Texel, The Netherlands
[l] School of Ocean and Earth Science, National Oceanography Centre, University of Southampton, Southampton SO14 3ZH, United Kingdom
[m] GEOMAR Helmholtz Centre for Ocean Research, D-24148 Kiel, Germany
[n] Department of Materials and Life Sciences, Sophia University, Chiyoda, Tokyo 101-8554, Japan
[o] Graduate School of Fisheries and Environmental Studies, Nagasaki University, Nagasaki 852-8521, Japan
[p] FI-TRACE, Department of Chemistry, University of the Balearic Islands (UIB), 01722, Spain
[q] Bigelow Laboratory for Ocean Sciences, East Boothbay ME 04544, United States

ARTICLE INFO

Article history:
Received 23 June 2014
Received in revised form 17 February 2015
Accepted 2 March 2015
Available online xxxx

Keywords:
Metal ions
Organic ligands
Speciation
Complexation
Equilibrium constant
Titration
Voltammetry
Multi-window titration
Data analysis

ABSTRACT

With the common goal of more accurately and consistently quantifying ambient concentrations of free metal ions and natural organic ligands in aquatic ecosystems, researchers from 15 laboratories that routinely analyze trace metal speciation participated in an intercomparison of statistical methods used to model their most common type of experimental dataset, the complexometric titration. All were asked to apply statistical techniques that they were familiar with to model synthetic titration data that are typical of those obtained by applying state-of-the-art electrochemical methods – anodic stripping voltammetry (ASV) and competitive ligand equilibration-adsorptive cathodic stripping voltammetry (CLE-ACSV) – to the analysis of natural waters. Herein, we compare their estimates for parameters describing the natural ligands, examine the accuracy of inferred ambient free metal ion concentrations ($[M_f]$), and evaluate the influence of the various methods and assumptions used on these results.

The ASV-type titrations were designed to test each participant's ability to correctly describe the natural ligands present in a sample when provided with data free of measurement error, i.e., random noise. For the three virtual samples containing just one natural ligand, all participants were able to correctly identify the number of ligand classes present and accurately estimate their parameters. For the four samples containing two or three ligand classes, a few participants detected too few or too many classes and consequently reported inaccurate 'measurements' of ambient $[M_f]$. Since the problematic results arose from human error rather than any specific method of analyzing the data, we recommend that analysts should make a practice of using one's parameter estimates to generate simulated (back-calculated) titration curves for comparison to the original data. The root–mean–squared relative error between the fitted observations and the simulated curves should be comparable to the expected precision of the analytical method and upon visual inspection the distribution of residuals should not be skewed.

* Corresponding author at: W-503 Turner Hall, 1102 S. Goodwin Ave., Urbana, IL 61801, United States.
E-mail addresses: pizeta@irb.hr (I. Pižeta), sylvia.sander@otago.ac.nz (S.G. Sander), rjhudson@illinois.edu (R.J.M. Hudson), omanovic@irb.hr (D. Omanović).
[1] Present address: College of Marine Science, University of South Florida, St. Petersburg FL 33701, United States.
[2] Present address: Department of Environmental Science, Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu 215123, China.

http://dx.doi.org/10.1016/j.marchem.2015.03.006
0304-4203/© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Please cite this article as: Pižeta, I., et al., Interpretation of complexometric titration data: An intercomparison of methods for estimating models of trace metal complexation by ..., Mar. Chem. (2015), http://dx.doi.org/10.1016/j.marchem.2015.03.006

Modeling the synthetic, CLE-ACSV-type titration dataset, which comprises 5 titration curves generated at different analytical windows or levels of competing ligand added to the virtual sample, proved to be more challenging due to the random measurement error that was incorporated. Comparison of the submitted results was complicated by the participants' differing interpretations of their task. Most adopted the provided 'true' instrumental sensitivity in modeling the CLE-ACSV curves, but several estimated sensitivities using internal calibration, exactly as is required for actual samples. Since most fitted sensitivities were biased low, systematic error in inferred ambient [M_f] and in estimated weak ligand (L_2) concentrations resulted.

The main distinction between the mathematical approaches taken by participants lies in the functional form of the speciation model equations, with their implicit definition of independent and dependent or manipulated variables. In 'direct modeling', the dependent variable is the measured [M_f] (or I_p) and the total metal concentration ($[M]_T$) is considered independent. In other, much more widely used methods of analyzing titration data – classical linearization, best known as van den Berg/Ružić, and isotherm fitting by nonlinear regression, best known as the Langmuir or Gerringa methods – [M_f] is defined as independent and the dependent variable calculated from both $[M]_T$ and [M_f]. Close inspection of the biases and variability in the estimates of ligand parameters and in predictions of ambient [M_f] revealed that the best results were obtained by the direct approach. Linear regression of transformed data yielded the largest bias and greatest variability, while non-linear isotherm fitting generated results with mean bias comparable to direct modeling, but also with greater variability.

Participants that performed a unified analysis of ACSV titration curves at multiple detection windows for a sample improved their results regardless of the basic mathematical approach taken. Overall, the three most accurate sets of results were obtained using direct modeling of the unified multiwindow dataset, while the single most accurate set of results also included simultaneous calibration. We therefore recommend that where sample volume and time permit, titration experiments for all natural water samples be designed to include two or more detection windows, especially for coastal and estuarine waters. It is vital that more practical experimental designs for multi-window titrations be developed.

Finally, while all mathematical approaches proved to be adequate for some datasets, matrix-based equilibrium models proved to be most naturally suited for the most challenging cases encountered in this work, i.e., experiments where the added ligand in ACSV became titrated. The ProMCC program (Omanović et al., this issue) as well as the Excel Add-in based KINETEQL Multiwindow Solver spreadsheet (Hudson, 2014) have this capability and have been made available for public use as a result of this intercomparison exercise.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

In order to predict the nutritive or toxic effects of a bioactive metal in a marine ecosystem, one must first determine its aqueous speciation (Lewis and Sunda, 1978; Moffett et al., 1997; Batley et al., 2004). The ideal speciation analysis would report both the identities and concentrations of each distinct aquo or complex species that contains the metal of interest. In reality, marine chemists typically perform two types of analysis. First, they measure the total concentration of the metal and of its principal chemical forms. Each 'form' can comprise a single chemical species, defined by oxidation state, charge, molecular structure, etc., but more often includes a group of species that i) contain the metal in a single oxidation state or organometallic compound and ii) are interrelated by reversible acid–base or complexation reactions. Second, they make measurements needed to model the equilibrium distribution between the different species that each 'form' comprises, i.e., between its free ion and its complexes with the multitude of inorganic and organic ligands present. Herein, we focus on this second type of speciation analysis.

For metal ions that are complexed mainly by inorganic ligands, it is often possible to develop a relatively complete speciation model from published stability constants and readily-measured inorganic ligand concentrations (Turner et al., 1981; Millero and Schreiber, 1982; van den Berg, 2001; Byrne, 2002; Gustafsson, 2014). For metals that are mainly bound by organic ligands, even identifying what these ligands are is a daunting if not impossible task, so the extent to which a metal is bound in organic complexes must be ascertained experimentally, i.e., by performing complexometric titrations. In this way, a quantitative chemical model for the relative abundance of free and complexed species of the metal in the analyzed water sample can be derived empirically.

The accuracy of such models, however, depends strongly on the data analysis methods used and the skill with which they are applied (Fig. 1). Among the community of researchers engaged in trace metal speciation analysis, several different approaches and custom programs are commonly used to model complexometric titration data. The fact that the 15 participants of this study applied 21 different approaches suggests that even a single researcher or laboratory may employ different tools on occasion. In addition, recent attempts to identify and resolve problems inherent in the most widely used methods of data analysis have not yet had much impact. Thus, it is hoped by all contributors that the intercalibration exercise reported herein will help reveal which methods perform the best and nudge the community toward using them so that the quality of our data analysis can more consistently match the investments in travel, clean sample collection and handling, and difficult chemical analyses that obtaining high quality trace metal speciation data requires.

2. Progress in complexometric titration modeling

When performing and interpreting experimental studies of metal speciation, the analyst must address 3 distinct methodological issues: methods of chemical analysis, experimental design, and methods of data analysis. In this report, we address methodological issues that arise in the analysis of voltammetric data from experiments in which incremental metal additions titrate the metal-binding ligands present in natural samples. To interpret such experiments, the analyst must choose an approach to data analysis or modeling. Each choice of approach is defined by i) the model structure employed for describing metal complexation equilibria, ii) the method of calibration adopted, iii) the mathematical transformations used, and iv) the software tools used for estimating the model parameters. These distinctions, together with some historically-significant references, are summarized in a proposed typology of approaches (Table 1).

Please cite this article as: Pižeta, I., et al., Interpretation of complexometric titration data: An intercomparison of methods for estimating models of trace metal complexation by ..., Mar. Chem. (2015), http://dx.doi.org/10.1016/j.marchem.2015.03.006

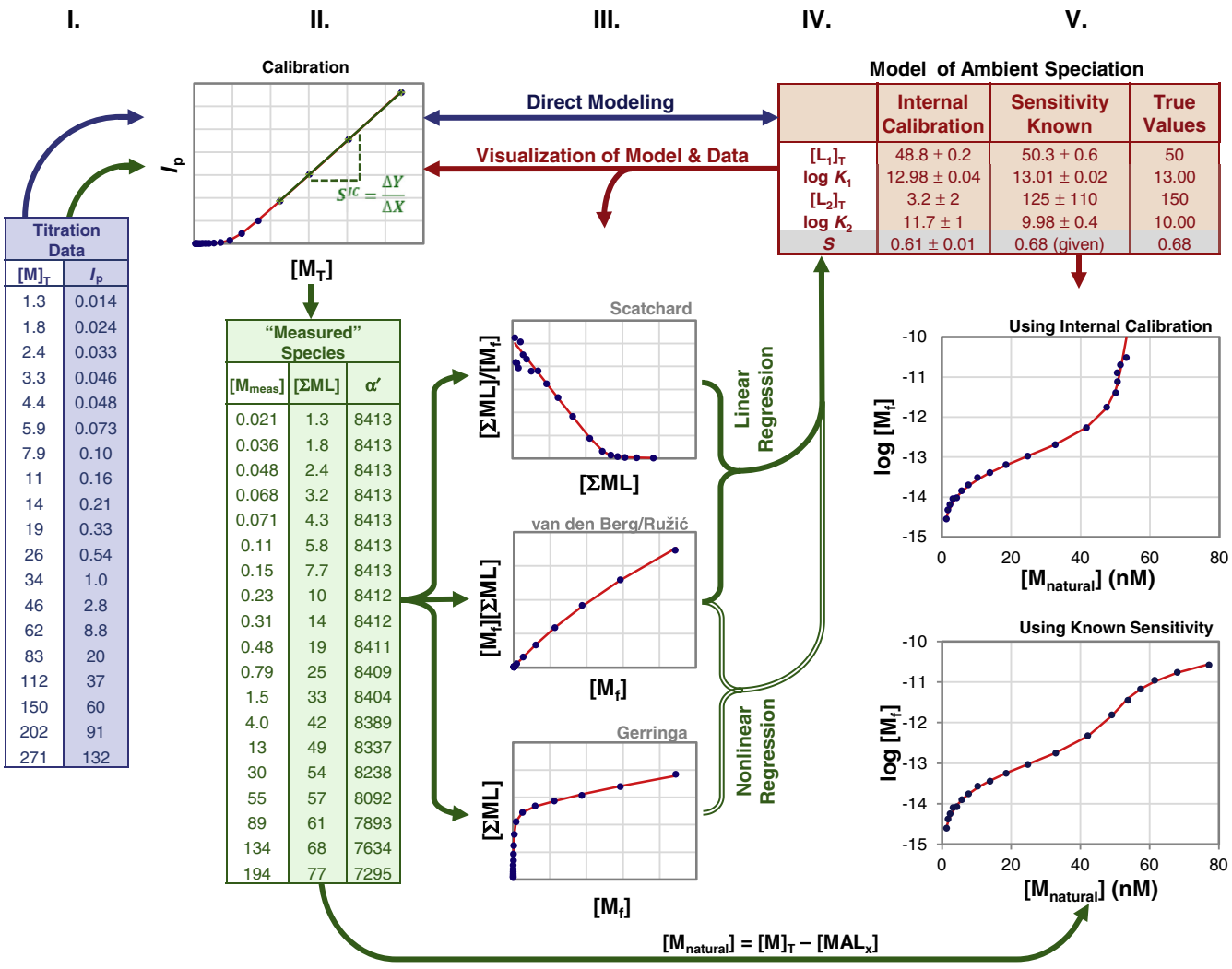

Fig. 1. Typical steps in retrieving parameters needed to model metal complexation by natural organic ligands from a single titration curve. (I) Measure voltammetric peak currents (I_p) in sample aliquots over a range of total metal concentrations $[M]_T$ sufficient to titrate natural ligands. Data shown are from window MW1-3. (II) Plot I_p on a linear graph and estimate the slope (S^{ic}) at high $[M]_T$. Assume $S = S^{ic}$ to calculate $[M_{meas}]$ from Eq. (16). Check for titration of AL and compute [AL′] (see Appendix 6). Compute exact α′, compute $[M_f]$, and (III) plot new Y and X variables according to the transformation taken, e.g., Ružić/van den Berg, Scatchard, or Gerringa (Table 1). (IV) Select an approach (linear regression, non-linear regression or direct modeling) to estimate natural ligand parameters ($[L_i]_T$ and K_i). Verify model accuracy by simulating the titration data using the parameter estimates (V) and comparing the simulated curves (in red) to the observations (in blue). Also make $[M_f]$–$[M_{natural}]$ plots to visualize in situ speciation as a function of different ambient $[M]_T$. When the model is appropriate and the parameters are recovered precisely, the red curves should fit the trends of the blue points in all plots and $[M_f]$–$[M_{natural}]$ plot should not curve backwards at high $[M_{natural}]$.

2.1. Models for natural ligands

The key step in interpretation of these complexometric titration data is, at the risk of introducing unfamiliar jargon, 'estimation of the model' that describes the equilibrium complexation of the metal by organic (or other strong) ligands, i.e., finding optimal values for the parameters describing metal–ligand interactions. Among workers in this field, the 'discrete ligand class model' has emerged as the most common model structure used to represent the diverse organic ligands present in natural waters. In this approach, the strongest class of organic ligands is called 'L_1' followed by L_2, … L_n to distinguish the progressively weaker ligand classes. Since the ligands present in humic acid, and presumably other common types of dissolved organic matter, form 1:1 complexes with the free metal ion (M_f) (Cabaniss and Shuman, 1988), one conventionally writes the equilibrium reaction for complexation by ligands of the *i*th class (L_i) as:

$$M_f + L_i \overset{K_i}{\leftrightarrow} ML_i \quad (1)$$

and the associated equilibrium mass law equation as:

$$K_i \equiv \frac{[ML_i]}{[M_f][L_i']} \quad (2)$$

where ML_i comprises all complexes of M with organic ligands of the *i*th class, L_i' denotes all ligands of this class not bound to M, and K_i is the average conditional stability constant of the ML_i. The complexation of M by the *i*th natural ligand class also depends on the total concentration of these ligands ($[L_i]_T$). The mass balance for each ligand class is simply:

$$[L_i]_T = [L_i'] + [ML_i]. \quad (3)$$

Herein, we refer to K_i and $[L_i]_T$ as 'natural ligand parameters,' or more succinctly 'ligand parameters,' and consider them to be specific to the mass of water from which the analyzed sample was taken.

ARTICLE IN PRESS

Table 1
Typology of approaches for modeling complexometric titration data. "Speciation models" are classified according to i) computational tools used (L, N, M), relationship between fitted Y/X and true dependent ([ΣML], $[M_x] = [M_{meas}]$ or $[M_f]$) and independent ($[M]_T$) variables, and by inclusion of fixed or variable α′ (see below). "Calibration" describes the coupling of this step to speciation modeling. Early/first workers to apply an approach are indicated, along with maximum number of ligands modeled (1L, 2L, or nL) and whether overload (O), unified multiwindow (MW) analysis, or reverse (R) titration was performed. Approaches taken by participants for modeling the MW1 dataset are shown in {bold, bracketed} letters. [Available programs shown in square brackets].

Speciation model				Calibration			
Math type	Y variable	X variable	α′[d]	Decoupled	Manual recursive	Simultaneous	Simultaneous multiwindow
Linear[a] (Ψ^L)	Analytical equations						
	$[\Sigma ML]/[M_x]$	$[\Sigma ML]$	Fix	1L: Scatchard (1949) 2L: Mantoura/Riley (1975) {**g**,**j**,**p**,**r**,**v**}		2L: Laglera (2001) {**i**}	
	$[M_x]/[\Sigma ML]$	$[M_x]$	Fix	1L: van den Berg/Kramer (1979) {**k**} 2L: Ružić;van den Berg (1982) {**o**, **r**, **u**} [VDB-XLS][ProMCC][G/R-NLR]	1L: van den Berg (2006) [VDB-XLS][ProMCC]	1L: Turoczy /Sherwood (1997) 2L: Laglera (2001) {**i**}	
	$1/[\Sigma ML]$	$1/[M_x]$	Fix	1L: Buffle (1977) 2L-R: Hirose (1982) [G/R-NLR]			
Nonlinear[b] (Ψ^N)	Analytical equations						
	$[M_x]/[\Sigma ML]$	$[M_x]$	Fix	nL: Pižeta/Branica (1997) {**e**} [ProMCC]			
	$[\Sigma ML]/[M_x]$	$[M_x]$	Fix	1L: Moffett (1995) nL: Croot/ Johansson (2000) {**h**}			
	$[\Sigma ML]$	$[M_x]$	Fix	2L: Gerringa (1995) {**a**,**b**,**l**,**m**,**s**,**x**} MW{**f**,**y**} nL: Cabaniss /Shuman (1988) [ProMCC][G/R-NLR]	[ProMCC]	[G/R-NLR]	
	$[M]_T$	$[M_x]$	Fix		2L: Wu/Jin (2009)	2L: Laglera (2013) {**n**} [Laglera-NLR][ProMCC]	
	Numerical model						
	$[M_x]$	$[M]_{natural}$	Fix	nL-MW,O: Voelker/ Kogut (2001) [FITEQL]			
Direct modeling[c] (Ψ^M)	Analytical equations						
	$[M_x]$	$[M]_T$	Fix	1L: Shuman/ Woodward (1973) 2L-MW: {**d**,**z**}		1L: Shuman/ Cromer (1979) 1L-R: Nuester / van den Berg (2006)	2L-MW: Hudson (2003)
				[Hudson-NLR]		[ProMCC]	[Hudson-NLR]
	Numerical model						
	$[M_x]$	$[M]_T$	Var	nL: McKnight-Westall (1983) nL-O+MW: Kogut/Voelker (2003) MW:{**c**} [FITEQL][PROSECE][ProMCC]			nL-MW: Sander/Wells (2011) {**t**} [KMS-XLS][Sander/Wells-NLR]

Available programs: XLS: Excel spreadsheets — VDB (van den Berg, 2014), KMS (Hudson, 2014); NLR: Non-linear regression codes written for various software packages — G/R (Gerringa et al., 2014), Hudson et al. (2003), Laglera et al. (2013), Sander et al.(2011); Stand-alone programs — FITEQL: Westall (1982); PROSECE: Garnier et al. (2004b); ProMCC: Omanović et al. (in this issue).

[a] Linear regression, incl. piecewise.
[b] Non-linear regression with some mixing of true Y (dependent) and X (independent) variables.
[c] Optimization with true X and Y maintained distinct.
[d] Fix means α′ is held constant for each data point, but can vary within a titration if calculated prior to regression modeling. Var means variation in α′ is computed explicitly within speciation model.

The ability of each L_i to compete for a metal is commonly expressed in terms of a 'side reaction coefficient' (α_{MLi}), which is the concentration ratio of complexed to free metal ion:

$$\alpha_{MLi} \equiv \frac{[ML_i]}{[M_f]} = K_i \cdot [L_i'] . \quad (4)$$

The maximum ability of a ligand class to compete for the metal occurs at the limit when M_f is too scarce for any ML_i to form, which leads to the following definition of the competition strength (KL_i) of the *i*th natural ligand class:

$$KL_i \equiv K_i \cdot [L_i]_T . \quad (5)$$

Several major inorganic anions naturally present in marine ecosystems – OH^-, Cl^-, F^-, CO_3^{2-}, and $B(OH)_4^-$ – also form complexes with metal cations (MX_{in}). As the anions' concentrations are generally much higher than those of trace metals, they are almost never titrated to a significant extent. Thus, a side reaction coefficient for each can be easily computed (Gerringa et al., 1995) from the product of i) its known metal complex stability constant(s) and ii) its ambient concentration, which can be measured directly or indirectly inferred. The aggregate effect of these ligands can be represented using the inorganic side reaction coefficient ($\alpha_{M'}$), which is defined by:

$$\alpha_{M'} \equiv [M']/[M_f] = ([MX_{in}] + [M_f])/[M_f] \quad (6)$$

where [M′] denotes the aggregate concentration of the aquo ion plus its complexes with major inorganic anions. Note that for most metals in marine waters, the value of $\alpha_{M'}$ is definable from measurements of *p*H, $[\Sigma CO_2]$, and salinity, and hence is specific to particular environmental conditions.

The complete mass balance for M in a natural water is then given by:

$$[M]_T = [\Sigma ML] + [M'] \quad (7)$$

where [ΣML] is the total concentration of metal bound by all classes of natural organic ligands. The basic equations – equilibrium mass laws and mass balances – underlying this speciation model are essentially the same across the numerous papers where they have been developed more fully, e.g., van den Berg (1982a), Ružić (1982), Cabaniss and Shuman (1988), Miller and Bruland (1997), and Wells et al. (2013).

Please cite this article as: Pižeta, I., et al., Interpretation of complexometric titration data: An intercomparison of methods for estimating models of trace metal complexation by ..., Mar. Chem. (2015), http://dx.doi.org/10.1016/j.marchem.2015.03.006

While the nomenclature these papers employ is also fairly consistent, the symbols used for key variables are not. Those used here (See Section 6) are drawn or adapted from these and other relevant papers.

2.2. Measurable metal species

Fundamentally, complexometric titrations depend on having an analytical method capable of quantifying the concentration of a chemically well-defined subset of the metal species present in a sample. While capable, non-electrochemical methods exist for most metals (van den Berg, 1982b; Hirose et al., 1982; Miller and Bruland, 1997) and remain methods of choice for certain important metals such as Hg (Black et al., 2007), anodic stripping voltammetry (ASV) and competitive ligand equilibration-adsorptive cathodic stripping voltammetry (ACSV) are currently the most widely used methods in studies of transition metal speciation. Both methods entail measuring the current (I_p) generated by the electrochemical stripping of metal accumulated during an antecedent deposition step. The fact that this metal comes solely from electroactive species whose concentrations are quantifiably linked to the [M_f] of the sample makes these methods useful for speciation analysis.

For each voltammetric method, the relationship of I_p to the solution speciation of a metal can be expressed in general form as:

$$I_p = \sum_{\ell=1}^{N_S} \left(s_\ell^* \cdot [M_\ell]\right). \tag{8}$$

Here, l is an index to distinguish each of the N_s metal species present and s_ℓ^* is the method's sensitivity to that species. Note that the s_ℓ^* are not necessarily constant between samples, as surface active components of seawater and possibly added synthetic ligands can diminish the accumulation of metals. Because the relative sensitivities of the individual metal species are usually not precisely known, it is common to calibrate an operational sensitivity parameter (S) that relates I_p to the aggregate concentration of well-defined species deemed to be 'measured' by the analytical method ([M_{meas}]):

$$I_p = S \cdot [M_{meas}]. \tag{9}$$

With ASV, metal complexes that are labile enough to dissociate within the electrode's diffusive boundary layer are electroactive, i.e., reduced during deposition and detected when the post-deposition oxidation current (I_p) is recorded. Thus, practitioners of ASV often adopt the hypothesis that only aquo metal ions and complexes with major inorganic anions are detected, leading them to define the concentration of ASV-labile or measurable M ([M^{asv}_{meas}]) as:

$$[M^{asv}_{meas}] \equiv [M']. \tag{10}$$

The application of ASV is, however, limited by the fact that it can only detect the amalgam-forming metals, such as Zn, Cd, Pb, and Cu.

In the case of ACSV, a fully-characterized ligand added to the sample, or 'added ligand' (AL), forms metal complexes that are preferentially deposited on hanging mercury drop electrodes (van den Berg and Nimmo, 1987). Thus, practitioners of ACSV typically assume that peak currents are proportional to [ΣMAL_x] and would define the concentration of metal species 'measured' by ACSV ([M^{acsv}_{meas}]) accordingly. Other significant ACSV workers, e.g., Laglera and van den Berg (2003), have defined S as the ratio of I_p and [M_{labile}], which includes M′ and corresponds to the total concentration of all well-defined M species present in an experimental aliquot. Taking [M_{labile}] to be equivalent to [M_{meas}], we thus define:

$$[M^{acsv}_{meas}] \equiv [M'] + [\Sigma MAL_x]. \tag{11}$$

Normally, [ΣMAL_x] is so much greater than [M′] that there is no practical difference between the values of S obtained using either definition.

CLE-ACSV practitioners match metal ions of interest with ligands that aid in analysis by forming strong complexes that adsorb onto mercury drop electrodes. Established combinations include Fe–NN (1-nitroso-2- naphthalene: Gledhill and van den Berg, 1994; van den Berg, 1995), Fe–SA (salicylaldoxime: Rue and Bruland, 1995; Buck et al., 2012a; Abualhaija and van den Berg, 2014), Fe–TAC (2-(2-thiazolylazo)-p-cresol: Croot and Johansson, 2000), Fe–DHN (2,3-dihydroxynaphthalene: van den Berg, 2006), Cu–SA (Campos and van den Berg, 1994), Cu–bzac (benzoylacetone: Moffett, 1995), Ni–DMG (dimethylglyoxime: van den Berg and Nimmo, 1987; Donat et al., 1994; Saito et al., 2004), Co–DMG (Saito et al., 2004; Baars and Croot, 2014), Co–nioxime (Ellwood and van den Berg, 2000; Baars and Croot, 2014), and Zn–APDC (ammonium pyrrolidine dithiocarbamate: van Den Berg and Nimmo, 1987; Ellwood and van den Berg, 2000; Lohan et al., 2005).

2.3. Titration experiments

In practice, a complexometric titration of a seawater sample comprises a series of voltammetric measurements made in aliquots to which increasing amounts of metal have been added and allowed to equilibrate with the L_i (Fig. 1, stage I). In principle, the analyst's goal is to add sufficient M to fully titrate all the ligands present in the sample. Additionally, while *p*H and major ion chemistry are kept constant across the aliquots, in ACSV one or more levels of synthetic ligand are added.

The combined competition strength of the natural inorganic and added organic ligands determine the range of KL_i of natural organic ligands that are detectable in any particular titration experiment (van den Berg and Donat, 1992). In recognition of this fact, analysts conventionally report a method-specific, detection window parameter (α'), which is an aggregate side reaction coefficient or 'competition strength' for all well-defined ligands known to be present in an experimental aliquot:

$$\alpha' \equiv \frac{[\Sigma MAL_x] + [M']}{[M_f]} = \alpha_{MAL_x} + \alpha_{M'} \tag{12}$$

where $\alpha_{M'}$ has been defined in Eq. (6). The side reaction coefficient of the added ligand (α_{MALx}) can be calculated using known conditional stability constants for the MAL_x and the concentration of AL not bound to M ([AL′]):

$$\alpha_{MAL_x} = K_{MAL} \cdot [AL'] + \beta_{M(AL)_2} \cdot [AL']^2 + \ldots + \beta_{M(AL)_n} \cdot [AL']^n. \tag{13}$$

Since $[AL]_T$ equals zero in ASV, titrations are only performed at a single detection window, i.e., $\alpha' = \alpha_{M'}$. On the other hand, for CLE-ACSV titrations at multiple $[AL]_T$ and therefore different α' are possible. Most often, ACSV titration curves are analyzed one at a time, in which case the only practical difference from analyzing ASV data is the magnitude of α'. Later, we will also encounter experimental designs in which data from curves of different α' are analyzed in a coherent manner. In order to model such cases, it is necessary to understand how the ACSV sensitivity varies with $[AL]_T$.

As discussed above and made explicit in Eq. (8), ACSV exhibits distinctly different sensitivities to different metal complexes. At present, the various s_ℓ^* are not precisely known. However, the available data indicate that S does change systematically along with the speciation of the metal. For example, the apparent ACSV sensitivity of Cu in NOM-free seawater containing SA depends on $[SA]_T$ (Fig. 2), as shown by the close agreement between works from three leading laboratories. While it has been argued that only $Cu(SA)_2$ is deposited in ACSV analysis (Campos and van den Berg, 1994), the observed reduction in sensitivity with decreasing $[AL]_T$ (symbols in Fig. 2) is much less than the decline in the relative abundance of $Cu(SA)_2$ predicted from published $CuSA_x$ stability constants for UV-oxidized seawater (dashed line in Fig. 2) and the

Please cite this article as: Pižeta, I., et al., Interpretation of complexometric titration data: An intercomparison of methods for estimating models of trace metal complexation by ..., Mar. Chem. (2015), http://dx.doi.org/10.1016/j.marchem.2015.03.006

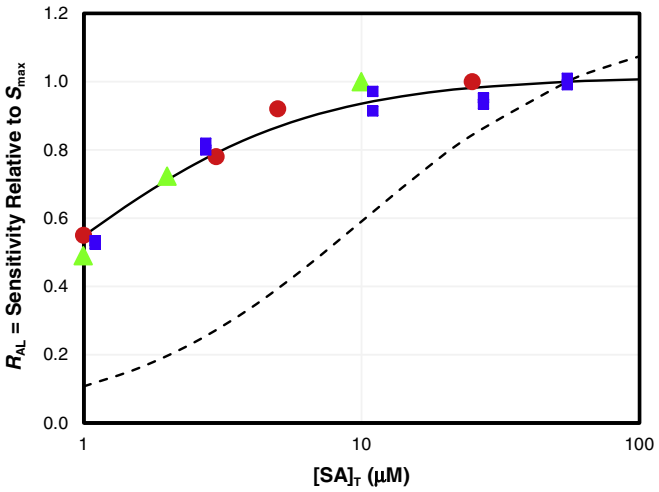

Fig. 2. Dependence of ACSV sensitivity ratio (R_{AL}) for Cu-SA on $[SA]_T$ in pH ~8 UV-oxidized seawater. Lines are modeled X_{AL} from Eq. (14) for published $K_{CuSA} = 10^{9.57}$ M^{-1} (- - - - -) and K_{CuSA} adjusted to $10^{8.5}$ M^{-1} (——) in order to fit data more closely. Measurements were made by Rue and Bruland (■), Kogut and Voelker (●), and Campos and van den Berg (▲) as reported in Hudson et al. (2003).

theoretical relationship (Nuester and van den Berg, 2005):

$$X_{AL} \equiv \frac{\alpha_{CuSA_2}}{\alpha_{Cu'} + \alpha_{CuSA_x}}. \quad (14)$$

Although an improved model may become available in the future, a simple, empirically-based relationship accounting for observed changes in S relative to an optimal value of S (S_{max}) while varying $[AL]_T$ but holding $[M]_T$ constant has proved useful:

$$S = S_{max} \cdot R_{AL}([AL]_T). \quad (15)$$

More empirical studies are needed before the best approach to predicting changes in S between windows in known. It may be better to define R_{AL} as a function of [AL′] and make it a non-linear function of M-AL speciation (Omanović et al., in this issue). In addition, the improvement in fit the X_{AL} after adjusting the stability constant for CuSA (solid line in Fig. 2) suggests that it may be worth reevaluating the published value.

Note that ACSV work with the Fe–SA system is subject to similar complications. Recent work suggests that at higher $[SA]_T$ the electroinactive $Fe(SA)_2$ complex becomes increasingly abundant relative to the electroactive FeSA species, resulting in a decrease in the signal at constant $[Fe]_T$ (Abualhaija and van den Berg, 2014) and hence in R_{AL}.

2.4. Calibration and data transformation

Once titration is completed, typically a set of model parameters, including S and one or more sets of K_i-$[L_i]_T$ pairs, is fitted by performing a series of data analysis/modeling steps. As a first step, the measured peak current (I_p) for each aliquot is plotted against the total metal concentration $[M]_T$ and a line fitted through the high $[M]_T$ part of the curve (Fig. 1, Stage II). When the weakest ligands in the sample are completely titrated, the slope of this line (S^{ic}) is equal to S, the method's true sensitivity to the measured species. Deriving S from the slope of the I_p-$[M]_T$ plot in this fashion is termed 'internal calibration' (Miller and Bruland, 1997).

Next, the value of S obtained by internal calibration is used to convert each observed I_p into a concentration of 'measured' M species:

$$[M_{meas}] \equiv I_p/S \quad (16)$$

which we define here to include M_f, MX_{in}, and MAL_x, but not the ML_i. Then, the overall mass balance for M, Eq. (7), is rearranged so that the total concentration of M complexed by the L_i ($[\Sigma ML]$) can be computed from the $[M_{meas}]$ and $[M]_T$ for each point in the titration curve (Fig. 1, II):

$$[\Sigma ML] = [M]_T - [M_{meas}]. \quad (17)$$

Generally, these measured concentrations are further transformed into new dependent (Y) and independent (X) variables that can be plotted and entered into statistical software in order to estimate speciation model parameters (Fig. 1, Stage III). The speciation model used in data analysis can be written generally in terms of X and Y as:

$$Y = f(X, \log K_i, [L_i]_T, \alpha') + \varepsilon \quad (18)$$

where both Y and X are functions of $[\Sigma ML]$ and either $[M_{meas}]$ or $[M_f]$ and ε is the fitting error. The different definitions of X and Y employed in the main approaches are summarized in Table 1.

At this point, we also introduce a variable, $[M_{natural}]$, that is intended to facilitate comparison of ACSV data to ambient $[M]_T$ levels (Moffett, 1995). $M_{natural}$ comprises all M species present either as aquo ions or in complexes with natural inorganic and organic ligands:

$$[M_{natural}] \equiv [M'] + [\Sigma ML] = [M]_T - [\Sigma MAL_x] \quad (19)$$

but excluding MAL_x complexes. Note that for ASV, $[M_{natural}]$ is identical to $[M]_T$ since $[\Sigma MAL_x]$ is always zero and that for ACSV, $[M_{natural}] \approx [\Sigma ML_i]$.

2.5. Parameter estimation

Historically, linearizing transformations (cf. Ψ^L, 'linear' approaches in Table 1), such as the van den Berg/Ružić (VDBR) plot (Ružić, 1982; van den Berg, 1982a), the Scatchard plot (Scatchard, 1949; Mantoura and Riley, 1975), and occasionally the double reciprocal plot (Buffle et al., 1977; Hirose et al., 1982), saw widespread use because they permitted easy estimation of parameters for a single natural ligand class using linear regression (Fig. 1, stage IV). As the widespread citation and continued use of linearization suggests, these early workers significantly advanced the science of modeling complexometric titration curves. However, it should be realized that all three of these linearizations are mathematically identical to the popular transformations of the Michaelis–Menten equation that biochemists have long employed (Woolf, 1932 as cited in Haldane, 1957) and carefully examined from a statistical perspective (Wilkinson, 1961).

While the use of simple linear regression to estimate ligand parameters is convenient, obtaining rigorous estimates of parameters for 2 ligand classes requires the application of non-linear regression (cf. Ψ^N non-linear approaches in Table 1). Data transformed using VBDR and reciprocal Langmuir transformations (Table 1, top 3 rows of Group Ψ^N) have been modeled using i) nonlinear regression (Pižeta and Branica, 1997; Moffett, 1995) as well as ii) iterative calibration/linearization after splitting the titration into 2 sections representative of those data where complexation is dominated by each class of ligands (Laglera-Baquer et al., 2001; van den Berg, 1984). However, the most widely-used approach to analyzing complexometric titration data using nonlinear regression software employs a different transformation entirely, one commonly known as the Langmuir isotherm or Gerringa plot and abbreviated here as the 'LG' model (Table 1, Group N). Note that simply adding the quantity $\alpha' \times [M_f]$ to Y of the LG model yields the mass balance equation for M with $Y = [M]_T$ and $X = [M_f]$, which

Please cite this article as: Pižeta, I., et al., Interpretation of complexometric titration data: An intercomparison of methods for estimating models of trace metal complexation by ..., Mar. Chem. (2015), http://dx.doi.org/10.1016/j.marchem.2015.03.006

is the model for a related approach to non-linear regression (Wu and Jin, 2009; Laglera et al., 2013). Significantly, this form of the speciation model variables almost exactly inverts the true independent ($[M]_T$) and dependent ($[M_f]$) variables of a titration, a trait shared by every linear and non-linear regression approach to varying extents (Table 1).

Several studies have already discussed the problems these methods have with specific types of data (Apte et al., 1988; Garnier et al., 2004b, 2005). For example, Gerringa et al. (1995) showed that the VDBR linearization is more susceptible to outliers than the LG non-linear regression approach. The Scatchard method is often used to fit more than one ligand class, but produces biased estimates of the ligand parameters (Sposito, 1982; Wu and Jin, 2009) because it overemphasizes the low $[M]_T$ points, which have the smallest current peaks and therefore are the least precise of any given titration (Miller and Bruland, 1997; Wells et al., 2013). Imprecise quantitation of peaks at low metal concentrations can cause spurious 'detection' of a non-existent stronger ligand (Omanović et al., 2010). On the other hand, non-linear fitting may not converge on the optimal parameter values if the analyst fails to provide adequate initial guesses for parameters. With these problems in mind, practitioners have modified these methods causing the number of variations in approach to grow almost as large as the number of research groups publishing metal complexation data.

One flaw is common to both the linear and nonlinear regression approaches: they conflate or even invert the actual independent ($[M]_T$) and dependent variables (I_p or $[M_{meas}]$) of the experiment. This violates the basic assumption of regression analysis: that the main error lies in measurements of the dependent variable. Note that linear transformations that do not make this error are possible for enzyme kinetics data (cf. Wilkinson, 1961), but not for voltammetric data. Furthermore, since neither class of methods minimize the model error with respect to the true dependent variable (I_p or $[M_f]$), by definition they cannot provide the closest fit to the untransformed observations.

Some workers have eliminated the problematic use of transformed variables in the speciation modeling step by directly solving for $[M_f]$ and $[M_{meas}]$ as a function of estimated ligand parameters and the true independent variables ($[M]_T$ and $[AL]_T$) (cf. Ψ^M, 'modeling' approach in Table 1). Shuman and Woodward (1973) took this approach using the analytical solution to a one-ligand model. McKnight et al. (1983) directly modeled ISE titrations using the FITEQL program (Westall, 1982), which numerically solves a complete speciation model. ACSV titrations have been directly modeled by Voelker and Kogut (2001) and Kogut and Voelker (2003) who also used FITEQL, and by Garnier et al. (2004a,b), who employed the PROSECE model. However, these works all retain the conceptual separation, or decoupling, of speciation modeling from calibration.

2.6. Coupling calibration and speciation modeling

The second common flaw lies in performing internal calibration uncoupled from speciation modeling, which introduces bias in model parameters (Voelker and Kogut, 2001) and fails to propagate uncertainty in S into that of ligand parameters (Hudson et al., 2003). Of course, workers in this area have long realized that S and the ligand parameters are interrelated and several have attempted to address the issue by coupling calibration and modeling.

Perhaps the most intuitive approach taken has been the use of recursion, i.e., first calibrating S and modeling the derived $[M_{meas}]$ and then recomputing S using speciation information from the back-calculated titration curve (Turoczy and Sherwood, 1997). Manual recursion has been employed with the linearized VDBR speciation model (van den Berg, 2006) and with non-linear regression based on the mass balance for M (Wu and Jin, 2009). Algorithms that automate this process have been devised, with linear and non-linear speciation modeling approaches nested within an outer loop where calibration is performed (Laglera-Baquer et al., 2001; Laglera et al., 2013). Such algorithms truly do seek simultaneous solutions for S and ligand parameters (Table 1).

However, the first application of fully coupled calibration/modeling considerably preceded these recursive approaches. Shuman and Cromer (1979) coupled calibration with a direct analytical solution to a one-ligand speciation model, but their method was not widely adopted. Perhaps the greater probability of encountering convergence problems with coupled calibration/modeling proved daunting, e.g., Laglera et al. (2013). Subsequently, Hudson et al. (2003) introduced an analytical solution to the two-ligand speciation model that made calibration with direct modeling possible for both ASV and ACSV. Coupling of calibration and a full matrix-based equilibrium model was first implemented by Sander et al. (2011). All of these workers are in effect estimating a model using Y equal to I_p and X equal to $[M]_T$ and $[AL]_T$. It must be emphatically stated, however, that additional natural ligand classes only make the problems of attaining convergence and of high parameter uncertainty worse for modeling single titration curves. Thus, progress in this area required an additional conceptual advance that grew out of further comparative experimental work.

2.7. Toward consistent parameter estimates

With the shared goal of building a common collection of mutually-consistent complexation model parameters, researchers in this field have cooperated in an inter-laboratory comparisons of methods for measuring organic complexation of Cu in estuarine (Bruland et al., 2000) and of Fe and Cu in oceanic waters (Buck et al., 2012a). In both studies, different groups used their normal analytical procedures, which differed in electrochemical instrumentation, solution chemistries, and data analysis methods, to analyze aliquots of a common large-volume sample. Reported parameters describing natural ligands were comparable between groups using similar conditions (some researchers performed multiple titrations at different detection windows), but the results of different approaches varied significantly. Bruland et al. (2000) assigned this variation to differences in analytical competition strengths (detection windows) of the methods used, noting that with increasing detection window $[L_i]_T$ values decreased, while K_i increased. Buck et al. (2012a) concluded that consensus between different labs could be attained when using the same analytical method at the same α', but inconsistencies were found for Fe between titrations with different added ligands. Recommendations arising from this inter-calibration were that: 1) a multi-detection window titrations would probably help better characterize the natural Cu-binding ligands in open ocean waters, and 2) all researchers should agree to use at least one 'standard' detection window to facilitate intercomparison of results between groups. In both cases, it became clear that along with refining analytical methods, achieving the goal of mutually-consistent complexation datasets requires a coherent strategy in choosing analytical detection windows, with multiple window titrations emerging as an important tool.

2.8. Multiwindow data analysis

The ability to analyze samples at multiple detection windows with ACSV simply by varying $[AL]_T$ raised an important question: how should the differences in parameter estimates obtained for each window be understood? A response heard from the earliest multiwindow analyses through to the inter-comparisons described above was to argue that different ligands were detected, which for a complex mixture of ligands is almost certainly true. However, the statistical significance of differences between ligand parameters obtained from titrations of different detection windows has rarely been rigorously examined. Nonetheless, Kogut and Voelker (2001) showed that data from multiple complexometric titrations of a single humic acid-rich solution performed at different $[AL]_T$ could be modeled using a single set of ligand parameters when transformed into $[M_f]$ vs. $[M_{natural}]$ space, i.e., using

Please cite this article as: Pižeta, I., et al., Interpretation of complexometric titration data: An intercomparison of methods for estimating models of trace metal complexation by ..., Mar. Chem. (2015), http://dx.doi.org/10.1016/j.marchem.2015.03.006

$Y = [M_f]$ and $X = [M_{natural}]$. Note that this approach would be considered 'direct' modeling were $[M_{natural}]$ not dependent on the measured $[M_f]$ (Table 1).

As noted above, a second consequence of decoupling 'calibration' from 'speciation modeling' is the systematic underestimation of the uncertainty in S and ligand parameters. Hudson et al. (2003) showed that S and parameters of weaker ligands can interact strongly, potentially causing very large increases in parameter uncertainty and preventing convergence when analyzing single titration curves. This is apparent from the increase in confidence limits for $[L_2]_T$ between the 'known S' and the simultaneous internal calibration cases in Fig. 1. This phenomenon can also be seen in real data sets of the highest quality, as seen in the reanalysis of Kozelka and Bruland (1998) data by Hudson et al. (2003). In that example, it is clear that 2 ligand classes were 'detected' in the normal statistical sense only if one accepts that S was known exactly.

Once one realizes that 2 or more parameters of a model can interact in ways that leave both poorly constrained – a problem known as collinearity – and that this problem affects *every* analysis of a single titration curve, the need for a means of properly calibrating S becomes glaring. Voelker and Kogut (2001) correctly realized that one solution was to obtain data in a window where the weak organic ligands interfere little or not at all. Their proposal to perform separate titrations at two detection windows was a major leap in solving the calibration problem. The 'overload' titration performed at very high $[AL]_T$ permits exact calibration to within measurement error, if the added ligand outcompetes all L_i present. The 'speciation' titration is performed at an $[AL]_T$ that permits accurate estimation of the ligand parameters of interest and R_{AL} used to correct for changes in S between windows. However, the decoupled analysis of data from the two windows maintains the artificial distinction between calibration and speciation modeling. As a result, practitioners of the overload approach do not make the fullest use of the information contained in the data generated (See Section 4.3).

The next step in addressing the problem was analyzing multiple titrations at different detection windows (Fig. 3) in a unified manner, while simultaneously estimating S_{max} and ligand parameters (Hudson et al., 2003). Since the points in a set of multiple ACSV titrations describe a surface with a single model underlying it (Fig. 3), it is in fact more appropriate to analyze multiwindow titrations as a single dataset than as individual curves. What made it possible to constrain S_{max} at the same time was the requirement that the same ligand parameters describe complexation in all detection windows. This validity of this 'criterion of coherence' was confirmed by improved fits and lower parameter uncertainties for models of the multiwindow datasets generated in the Bruland et al. (2000) intercomparison, especially Moffett's ACSV titrations using bzac. It also became clear that the large uncertainty inherent in analyzing any single curve meant that it was in no way justified to conclude that differences in ligand parameters obtained by fitting curves at different α' necessarily meant that different ligand classes were detected. In three different datasets, no more than 2 distinct classes could be fitted when simultaneously analyzing 5 distinct windows for each sample (Hudson et al., 2003). These insights were confirmed by implementing the unified multiwindow approach using a matrix-based equilibrium model (Sander et al., 2011).

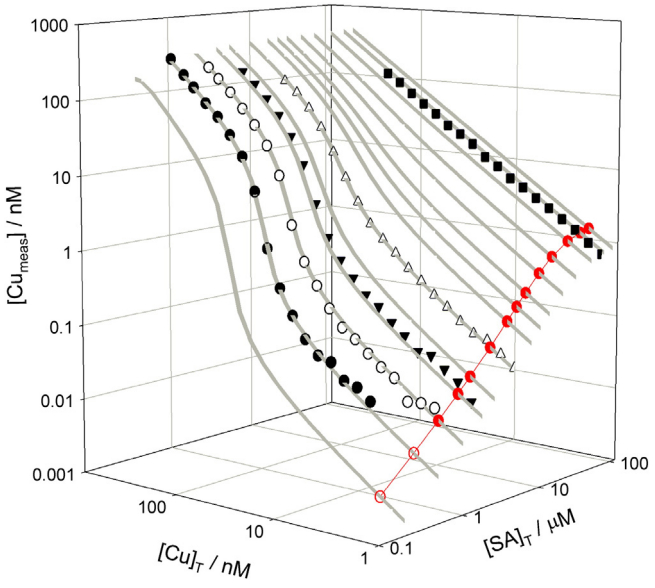

Fig. 3. Dependence of ACSV-measurable Cu ($[Cu_{meas}]$) on concentrations of total Cu ($[Cu]_T$) and added ligand ($[SA]_T$). Each set of points at constant $[SA]_T$ is a titration curve perfumed at a different 'detection window' from MW1 dataset. Gray lines are from variable α' model calculations. Red line connects data obtainable using reverse titration experimental design, i.e., $[SA]_T$ is varied while holding $[Cu]_T$ constant at ambient $[Cu]_T$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

A subsequent development in design of titration experiments derived from the unified multiwindow approach is the 'reverse titration,' which aims at quantifying very strong ligands present at concentrations lower than ambient $[M]_T$. Rather than titrating ligands by adding metals, as in a 'forward' titration, Nuester and van den Berg (2005) systematically increased $[AL]_T$ and analyzed the I_p data in a unified manner, despite the fact that each point differed in α_{CuSAx} (see also Hawkes et al., 2013). To accomplish this, they defined the ratio X_{AL} in Eq. (14), which plays a role identical to that of R_{AL} in Eq. (15), and calculated its values from the known stability constants of the Cu–SA species present.

The key similarity between reverse titration and the multiwindow approach is the unified analysis of data obtained at different α'. Since the model equations used in analyzing reverse titrations work with the ratios of I_p to the maximum value at high $[AL]_T$, S_{max} does not appear in the published equations. However, the effects of changes in $[AL]_T$ on sensitivity are included in the definition of X_{AL} in Eq. (14) in a manner that is analogous to R_{AL} (Fig. 2). Thus, in a reverse titration, calibration and speciation modeling are in effect performed simultaneously, albeit without necessarily using direct modeling. Note that by varying $[AL]_T$ and not $[M]_T$, this type of reverse titration only permits characterization of L_i present at concentrations near or below ambient $[M]_T$.

Finally, it should be noted that to date, the unified multiwindow approach has not been widely tested. So far, it seems to have been very successful in some cases, but not every set of multiwindow titrations is similarly coherent. Perhaps this is not surprising as the method relies on the underlying thermodynamic model of complexation by the added ligand being exact, as well as the assumption that equilibrium is actually attained at each point of the titrations. Thus, bias in stability constants and R_{AL} or differences in equilibration times between individual points or reaction kinetics of different natural ligands could make it difficult to 'unify' data from different curves. Similarly, interactions of DOM and AL on electrodes could make R_{AL} measured in UV-oxidized seawater inappropriate for samples containing DOM, e.g., if the presence of surface active substances affects S in a way that depends on $[AL]_T$. Clearly, experimentalists should devise tests of each assumption of the method, specifically for variables not related to the multiwindow approach: salinity, electrode size, potential deposition, etc. Nevertheless, the central insight of the approach is robust, even if its implementation requires that further adjustments to the model equations or even analytical procedures be made.

2.9. Uncertainty and parameter interactions

To understand the complexities of parameter estimation discussed above, it is helpful to also understand how uncertainty in estimates of model parameters is related to measurement error. Often, calculated $[M_f]$ values depend most directly on the parameters of the strong ligand, i.e., K_1 and $[L_1]_T$. This occurs when $[M]_T$ is much less than $[L_1]_T$ and M is almost entirely bound by L_1, permitting the approximation $[L_1'] \approx [L_1]_T$ to be combined with the mass law equation for ML_1 formation, Eq. (2),

to yield an inverse relationship between $[M_f]$ and the competition strength of L_1 (KL_1) at such $[M]_T$ values ($[M]_{T(Low)}$):

$$[M_f] = [ML_1]/(K_1 \cdot [L_1']) \approx [M]_{T(Low)}/KL_1. \quad (20)$$

Thus, *p*M must be linearly related to log KL_1 unless L_2 is strong enough to also exert an influence (▬ in Fig. 4). Since Eqs. (11), (12), and (16) together imply that I_p is proportional to $[M_f]$, experimental noise in I_p should introduce equal fluctuations in fitted KL_1 on a percentage basis,

The situation for KL_2 is more complicated. At realistic levels of natural ligands, L_2 buffers $[M_f]$ mainly at high $[M]_T$, but an analytical function analogous to Eq. (20) relating $[M_f]$ to KL_2 doesn't exist. However, it is simple to compute and graph a theoretical curve relating them for a concrete example (═ in Fig. 4). The curves show that $pM_{(High)}$ is much more sensitive to KL_2 than $pM_{(Low)}$ is to KL_1 and that it depends on L_1 parameters as well. Thus, a small fluctuation (experimental noise) in measured $[M_f]$ at low $[M]_T$ introduces an equal fluctuation in KL_1 on a percentage basis, but at moderate-to-high $[M]_T$, the random noise of the same relative magnitude causes a disproportionately large fluctuation in KL_2. Since the KL_2 curve also depends on the uncertain value of KL_1, much greater scatter in L_2 parameters result.

2.10. Summary

The analytical methods and experimental procedures used by trace metal biogeochemists have advanced to a point where, provided great care is taken and sufficiently sensitive instruments are employed, very high quality data can be generated. However, a considerable disparity remains in the sophistication of methods used by different workers in processing those data. As it would be a real benefit to the research community to identify which methods most consistently generate accurate parameters for modeling metal complexation by natural ligands, an intercomparison of data analysis methods was planned and accomplished as part of the terms of reference (ToR) of SCOR Working Group 139: Organic Ligands — A Key Control on Trace Metal Biogeochemistry in the Ocean (Sander et al., 2012; Buck et al., 2012b).

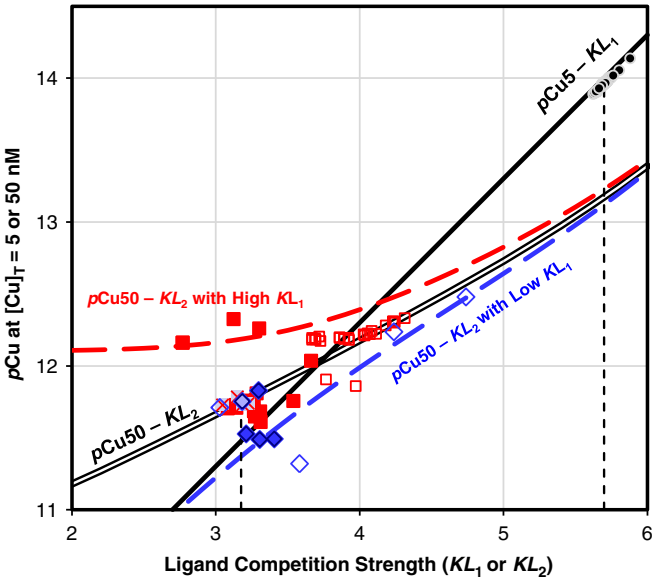

Fig. 4. Free copper concentrations (*p*Cu ≡ - log[Cu_f]) at $[Cu]_T$ of 5 nM (≡*p*Cu5) and 50 nM (≡*p*Cu50) plotted versus competition strength of either L_1 (KL_1) or L_2 (KL_2). (▬) *p*Cu5 vs. KL_1 computed from Eq. (20); (═) *p*Cu50 vs. KL_2 computed using 'alternate' ligand parameters and $[Cu]_T = 50$ nM in two-ligand model; (▬ ▬): *p*Cu50 vs. KL_2 for high-biased KL_1 ($K_1 = 10^{13.1}$ M^{-1}, $[L_1]_T = 55$ nM); (▬ · ▬): *p*Cu50 vs. KL_2 for low-biased KL_1 ($K_1 = 10^{12.9}$ M^{-1}, $[L_1]_T = 45$ nM). All *p*Cu were computed using ligand parameters from submitted responses for MW1 data set with L_2 detected and fixed α′. (●) KL_1 from submitted parameters and *p*Cu5 calculated from parameters for both L_1 and L_2. *p*Cu50 for responses with $S = R_{AL}$ (□■×) and with calibrated S (◇◆◆) using approaches Ψ^L, Ψ^N, Ψ^M respectively. Dashed lines indicate correct values of KL_1 (5.699) and KL_2 (3.176).

This intercomparison expands upon previous efforts to intercalibrate analytical methods by examining outcomes from participants' modeling of synthetic datasets that are similar to routinely generated ASV- and ACSV-type complexometric titrations. Participants from 15 different laboratories applied the data analysis tools that each routinely employ and submitted their interpretations of the data. Here, we report these results and compare them to the 'correct' values of ligand parameters and to $[M_f]$ predicted at both low and high ambient $[M]_T$.

3. Methods and models

3.1. Generation of synthetic datasets

Synthetic datasets representing complexometric titrations of organic, metal-binding ligands naturally-present in surface waters were generated by S.G.S. and I.P. using equilibrium speciation models plus the above equations relating I_p, $[M_{meas}]$, and R_{AL}. Each virtual sample is defined by a set of 'true' parameters (Table 2) describing the natural ligands present, i.e., their concentrations ($[L_i]_T$) and complex stability constants (K_i). The reagent additions made in each titration and the simulated instrumental response mimic complexometric analysis of real samples by either ASV or CLE-ACSV. All complete sets of synthetic data are available in Appendix 1.

3.1.1. ASV-type, single detection window datasets

The ASV-type, single-detection window (SW) datasets provided to the participants were all noise-free. 'Samples' A–C contained just a single ligand class and were titrated using linear-, logarithmic- and decadal- distributions of metal additions, resulting in 9 individual titration curves (Table 2). Samples D–F contained 2 ligand classes each and were titrated using the same 3 distributions of metal additions, resulting in 9 additional titration datasets. The final ASV-type titration was generated for a sample (G) with 3 ligand classes.

For samples A–F, participants were directed to choose a one- or two-ligand model, as they deemed appropriate. For dataset G, participants were advised to include between 2 and 4 ligand classes. The former primarily tested participants' ability to accurately estimate complexation model parameters, while dataset G tested their ability to detect the presence of more than 2 ligand classes in a sample.

3.1.2. CLE-ACSV-type, multi-detection window datasets

ACSV-type, multi-detection window (MW) titration datasets were generated for two different 'samples' containing two natural ligand classes each (Table 2). The first sample (dataset MW1) represented the more common case where $[L_1]_T < [L_2]_T$. The second sample (datasets MW2 and MW3) was designed to make L_2 difficult to detect since $[L_2]_T < [L_1]_T$. Titrations were generated for 5 different concentrations of added salicylaldoxime between 0.5 and 70 μM. Random measurement error (noise) was added to the synthetic I_p data in MW1 and MW3, while set MW2 was noise-free. Noise having both a proportional component of random error that averages 3% of the signal and a constant component was generated based on data from real Cu-SA titrations at the University of Otago (Sander et al., 2011). The detection limit was taken to be 0.01 nA; two out of 95 data points from the MW1 curves fell below the detection limit and were classified as 'nd.'

Participants were informed that the total side reaction coefficient for inorganic copper ($\alpha_{Cu'}$) was 13 and the overall value for Cu complexation by SA could be computed from Eq. (13) with $K_{CuSA} = 10^{9.57}$ M^{-1} and $\beta_{CuSA_2} = 10^{14.57}$ M^{-2} (Campos and van den Berg, 1994). Participants were further informed that the relative ACSV sensitivities (S) to SA-complexed Cu in each curve were defined by the empirical factor (R_{AL}) that accounts for the dependency of S on $[SA]_T$ (Fig. 2; Eq. 15). Since S at the highest $[SA]_T$, or S_{max}, was defined as unity, the lowest $[SA]_T$ window had a sensitivity of 0.36 (Table 3).

ARTICLE IN PRESS

Table 2
Correct complexation model parameters of synthetic samples.

Sample	$[L_1]$ (nM)	log K_1 (M^{-1})	$[L_2]$ (nM)	log K_2 (M^{-1})	$[L_3]$ (nM)	log K_3 (M^{-1})
Single detection window, ASV-type titrations (True model)						
A	10	10				
B	50	9				
C	90	8				
D	10	8	100	6		
E	10	10	100	6		
F	30	10	300	7		
G	10	16	25	13	100	10
Multiple detection window, ACSV-type titrations Variable α′ model (True)						
MW1	50	13	150	10		
MW2/3	50	15	15	12.5		
Fixed α′ model (Alternate)[a]						
MW1	50.02	13.001	268.1	9.756		

[a] Alternate parameters were derived by generating high-resolution (one point every 5 nM), noise-free true titration curves over the same range as each MW1 window and fitting them using weighted ($W = 1/I_p$) residuals ($N = 293$). RMS-RE = 0.15%; $R^2 = 1.000$.

3.2. Instructions to participants

Each participant received a package of data and instructions via email (See SCOR Working Group 139, 2014). For each simulated titration, participants were provided with experimental details, including the actual sensitivity (S for ASV and S_{max} for ACSV) used to generate the synthetic voltammetric data and all values of independent and dependent variables for each point in the titration curve (I_p, $[M]_T$, and $[AL]_T$ if present). Other information needed for complexation modeling, including conditional stability constants for MAL_x complexes, α'_{MAL_x} for each detection window, and α_M' were also provided.

Participants were instructed to analyze each SW or MW dataset as if it were from a real sample and report the parameters for each ligand class, along with the number of ligand classes detected. For the MW titrations, the directions also suggested that participants who lacked a modeling tool for analyzing the unified multiwindow dataset should fit the curves separately and decide whether the parameters were reliable or outside the detection window.

The participants were asked to describe their methods of data analysis according to the scheme of Fig. 1. In order to gain deeper insights into other attributes needed to evaluate and compare the results, the participants were asked whether they i) removed some data points before fitting; ii) plotted original data (linear–linear, log–log or both); iii) transformed data according to VDBR, Scatchard, or LG methods or examined the same curve with more than one transformation; iv) used linear, nonlinear or both fitting methods, and if nonlinear, how they obtained initial guess parameters; and v) simulated or 'back calculated' the titration curve using their fitted parameters ($[L_i]_T$, log K_i) in order to visually inspect the success of fitting. They were further asked about the criteria employed to define goodness of fit and to choose the number of ligands in the model fitted to the data. The full questionnaire provided to participants can be found in Appendix 5.

Table 3
Characteristics of the titration curves in the multi-window datasets.

	Window				
	1	2	3	4	5
Characteristics common to datasets MW1–MW3					
$[AL]_T$ (μM)	0.5	0.975	1.9	4.5	70
log α′	3.29	3.60	3.92	4.38	6.32
R_{AL}	0.36	0.52	0.68	0.84	1.00
Max. $[Cu]_T$ (nM)	477	346	271	202	150
Characteristics unique to dataset MW1					
Min. $[AL']/[AL]_T$	0.387	0.749	0.886	0.958	0.996
Sic[a]	0.286	0.411	0.593	0.722	0.850
Inferred S_{max}	0.795	0.791	0.872	0.859	0.850
Window-specific (alternate) parameters for fixed-α′ model.[b]					
log K_1 (M^{-1})	13.000	13.000	13.000	13.000	12.995
$[L_1]_T$ (nM)	50.03	50.02	50.03	50.03	50.55
log K_2 (M^{-1})	9.752	9.716	9.619	9.538	9.501
$[L_2]_T$ (nM)	265.0	288.5	360.6	440.0	448.8

[a] $S^{ic} = S$ determined from regression slope of final 4 points of titration curve, i.e., simple internal calibration.
[b] Estimated by fitting high-resolution (one point every 5 nM), noise-free true data over given range of each of curve from the MW1 dataset with inverse-square weighting of residuals ($N = 59$).

3.3. Participants' methods and results

In response to this intercomparison exercise, 15 participants reported 21 independent sets of results, which are referred to herein as the responses of participants *a* to *w* (Appendix 2), and described the methods they used to analyze the data (Appendix 3). All results have been categorized according to the methods of data analysis used (cf. Table 1).

Unexpectedly, 8 of the 21 responses reported that the sensitivity S had been fitted or adjusted for at least some curves, meaning that the provided correct S values had been overlooked. Since S is never known a priori for a real dataset, it could be said that these participants treated the synthetic data exactly as one must treat real data. The importance of this distinction between those that adjusted S and those that did not should be apparent from the discussion of calibration above. At the very least, having results of both types has provided us with an opportunity to examine the effects of S calibration on the accuracy of complexometric titration analysis. Further pertinent decisions made by the participants include the number of ligand classes in the speciation model, the choice of the initial guesses for parameters (if needed), and the criteria for convergence of iterations in nonlinear fitting.

Because the number of participants within each group was not sufficient to make valid statistical comparisons between methods and because it proved difficult to determine some important details of what each participant did, we 'invited' 3 pseudo-participants to submit results. These pseudo-participants followed prescribed procedures intended to resemble a standard means of applying an existing method (see Appendix 3), but since they knew the answers in advance their results are not blind tests. Pseudo-participant y' used the LG model using the given sensitivity ($S = R_{AL}$). Pseudo-participant x' also used the LG model, but adjusted slopes using internal calibration. Pseudo-participant z' used direct modeling to estimate the two-ligand model, both with and without adjusting slopes.

Please cite this article as: Pižeta, I., et al., Interpretation of complexometric titration data: An intercomparison of methods for estimating models of trace metal complexation by ..., Mar. Chem. (2015), http://dx.doi.org/10.1016/j.marchem.2015.03.006

3.4. Evaluation of results

Three basic metrics of the quality of participants' results are used herein. The first measures the quality of the fit to the data, while the second measures the accuracy of a parameter estimate. The third metric is the deviation of *p*M at particular $[M]_T$ of interest predicted from a particular parameter set from the value defined by the 'correct' ligand parameters.

To assess the quality of fit to the observed titration curve of any given parameter set, all data points of the relevant titration curve were first simulated (I_p^{calc}) for the reported parameter set and specified $[Cu]_T$ and $[AL]_T$ (see Fig. 1) using the program ProMCC (Omanović et al., in this issue). Two metrics of fit to the titration curve were then derived: *RMSEF* and *RMS-RE*. Note that these metrics may differ from the objective function optimized by participants.

Calculation of the RMS error function, or *RMSEF* (Sander et al., 2011), begins by computing a normalized residual, or relative percent difference (RPD_i), for each point in a fitted titration curve:

$$RPD_i \equiv \frac{2 \cdot \left(I_p^{calc} - I_p\right)}{\left(I_p^{calc} + I_p\right)} \tag{21}$$

The *RMSEF* for each parameter set is then calculated from the normalized residuals (RPD_i):

$$RMSEF = \sqrt{\frac{\sum_{i=1}^{n_m} RPD_i^2}{n_m - n_p}} \tag{22}$$

where n_m is the number of titration data points and n_p the number of estimated parameters. In some cases, *RMSEF* are multiplied by 100 and reported as a percentage. Note that the denominator in Eq. (22) normalizes the error function by the degrees of freedom of the model, i.e., the number of data points minus the number of parameters. This ensures that the *RMSEF* of a one-ligand model can be compared to that of a multiple-ligand model.

A second, closely-related metric of fit to titration curves is the RMS relative error, or *RMS-RE*. To compute this quantity, the relative error of each simulated data point (RE_i) was calculated using:

$$RE_i \equiv \frac{\left(I_p^{calc} - I_p^{obs}\right)}{I_p^{obs}} \tag{23}$$

Then the RMS relative error for a parameter set/titration curve combination is:

$$RMS\text{–}RE = \sqrt{\frac{\sum_{i=1}^{n_m} RE_i^2}{n_m}} \tag{24}$$

The bias in parameter estimates (E_P) was computed using log-transformed values of the parameters, as shown here for the generic parameter P:

$$E_P \equiv \log(P_{est}) - \log(P_{correct}) \tag{25}$$

For a group of *N* estimates of parameter *P*, it can prove useful to compute the RMS error of the estimates (*RMS-*E_P):

$$RMS\text{–}E_P = \sqrt{\frac{\sum_{i=1}^{n} E_{Pi}}{N}} \tag{26}$$

The RMS error in derived *p*Cu values is computed from Eq. (26) by setting log *P* in Eq. (25) equal to *p*Cu5 or *p*Cu50.

3.5. Sources of error in parameter estimates

An important complication in the MW1 dataset became apparent during the analysis of submitted results, namely that in the 3 MW1 titration curves with the lowest $[SA]_T$ added, α_{CuSAx} was reduced by 11–61% by the end of the titration (Table 3). Consequently, the assumption that α' remains fixed within a titration curve, which was made by all but 3 participants, was not consistent with the speciation model used to generate the synthetic data. Since less than 10% of data points are affected significantly and fits to the data were reasonably accurate with fixed α' assumed, the problem went largely unnoticed by participants. Real ACSV titrations performed at low $[AL]_T$ may also be affected by this problem (Kogut and Voelker, 2001), albeit generally less than in our window MW1-1.

The fact that most participants who followed common practice applied an incorrect speciation model, however, does raise the question of how to evaluate their results since their mis-specified model might prevent them from ever finding the 'true' parameter values. That is, since the difference between participant results and the 'true' parameters (E_P) must reflect a combination of systematic bias arising from model structure (Δ_{model}) and the method used in data analysis (Δ_{method}), plus random error in fitting (ε_{random}):

$$E_P = \Delta_{model}(P) + \Delta_{method}(P) + \varepsilon_{random}(P) \tag{27}$$

parameter estimates could appear inaccurate, with high E_P, in some cases and accurate in others, with low E_P, just by chance if the model and/or random errors added to or offset the method error. Since our objective is to assess Δ_{method} it follows that we should compare each parameter estimate to a 'correct' value that corresponds to the fixed or variable α' model (Tables 2 and 3) actually employed by the participant, rather than solely to the 'true' values.

In order to derive 'correct' parameters for the speciation model with fixed α', we fitted a set of ligand parameters to all five noise-free curves of MW1 both singly (Table 3) and simultaneously (Table 2). Herein, the simultaneously fitted set of parameters is referred to as the 'correct' parameters of the 'alternate' or 'fixed α'' model. Note that only the resultant L_2 parameters are noticeably different from the original model. Calculated $[Cu_f]$ at 5 and 50 nM differs by less than 1%. The residuals of the MW1 data points from the alternate model are distributed similarly to those of the original model. N.B. The alternate parameters were derived solely for the purpose of evaluating responses in this comparison. This approach should not be used as an alternate means of analyzing real data.

3.6. Models and programs

A variety of custom programs and commercial software packages were employed in this work. The synthetic data were generated using custom equilibrium speciation modeling software based on Morel Tablature (Sander et al., 2011). Much of the basic data manipulation was accomplished in Microsoft Excel, while specialized statistical analysis was performed using SAS v 9.4 (SAS Systems). ProMCC (Omanović et al., in this issue) was used to perform VDBR and LG fits of titration curves. Speciation calculations for graphs and evaluation of model error were performed using PROSECE (Programme d'Optimisation et de SpEciation Chimique dans l'Environnement) (Garnier et al., 2004a; Louis et al., 2009) and KINETEQL Multiwindow Solver (Hudson, 2014). Software used by the participants is described in Appendix 3.

4. Results and discussion

As in many inter-laboratory comparisons, this study has produced a collection of results that are most naturally organized into a two-dimensional, 'participant' by 'sample' matrix. Ideally, we would subject the results to a formal statistical analysis capable of resolving the inter-

participant differences into contributions from the methods or approaches taken (cf. Table 1) and individual factors, such as expertise. Now the submitted responses make clear that important inter-participant differences do exist, but enough variants of the main approaches to data analysis (Table 1) were used that the number of results reported using each is too low and too unbalanced to permit us to make a rigorous analysis of variance. Consequently, we must scale back our ambitions and rather explore these results with the goal of identifying the most important inter-method differences. It is our hope that the graphical comparisons and statistical analyses of these results, which were obtained from the actual implementation of diverse methods by practitioners, prove to be helpful and perhaps even offer compelling reasons to improve one's skills or adopt better methods as warranted.

To explain how differences in the quality of participants' results are linked to the various approaches to data analysis taken, a scheme for classifying these approaches is essential. These classes and our notation used to represent them (Ψ^{X}_{ASn}) must describe both the structure of the model applied, which is defined by the subscripts (A, S, n), and the mathematical approach to data transformation and speciation modeling, which is defined by the superscript (X) (cf. Ψ^{L}, Ψ^{N}, or Ψ^{M} in Table 1). The factors that distinguish the model structures include: A) whether fixed (F) or variable (V) α' was employed to account for titration of AL; S) whether the sensitivity used was the provided reference (R) value, one derived from simple internal calibration (C), or one obtained by simultaneous calibration + modeling (CM); and n) the number of natural ligand classes included in the speciation model.

In making this comparison, we employ several quantitative metrics. The main metrics of model fit to data are two different root–mean–squared residuals. These include the RMS error of fit, or *RMSEF* defined by Eq. (22), and the RMS relative error, or *RMS-RE* defined by Eq. (24). In addition, RMS errors in parameter estimates, or *RMS-E*$_P$ defined in Eq. (26) after log transformation are used to assess accuracy. Finally, a wholistic assessment of a set of parameter estimates is obtained from the difference between log [Cu$_f$] calculated using 'true' and fitted values of ligand parameters at [Cu]$_T$ of 5 and 50 nM. The complete set of calculated [Cu$_f$] is provided in Appendix 4.

4.1. ASV titration curves (noise-free data)

4.1.1. Descriptive statistics

Of the 21 participants, 12 submitted complete or partial sets of responses for the ASV-type, single detection window titration curves. Each submission comprises up to 19 responses in total, corresponding to the 3 different experimental designs for samples A–F and 1 design for sample G. For the entire set of 237 ASV-type responses, the geometric mean *RMSEF* is 0.16%, with 71% of responses having less error than this. Since an *RMSEF* of this magnitude can result from an error in log K_1 of only 0.002 for a sample such as F_1, the bulk of the responses should be considered to be precise fits of the data, which for the purposes of this discussion is defined as parameter sets with *RMSEF* between zero and 0.1%. Note also that 31% of the responses were exact fits, i.e., had *RMSEF* equal to zero (Fig. 5A). Since minor factors, such as the number of significant digits in the I_p data a participant entered into their software, could degrade the fit from 'exact' to 'precise', in this context the distinction is only important as a reminder that it is possible to achieve zero model error with noise-free data, cf. Eq. (27). With that in mind, we examine the ASV-type results more closely to see what caused the modest number of imprecise results that did arise. What was revealed by this relatively uncomplicated test foreshadows our findings from the ACSV-type data below: that underestimation of sensitivity and incorrect model structures (number of ligand classes) are primary causes of inaccurate fits to observational data.

Unsurprisingly, the best results were reported for the single-ligand samples (A–C), with only 7% of the responses yielding imprecise or poor fits (Fig. 5A). All 12 sets of responses reported the correct number

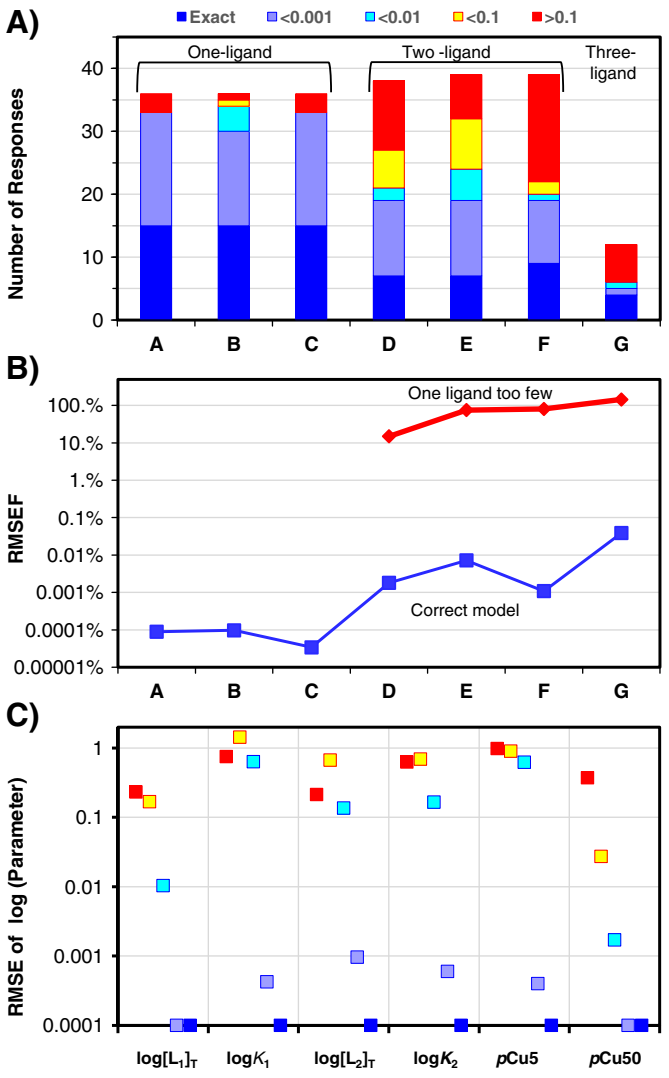

Fig. 5. Summary of results for ASV-type, single detection window titrations of samples A–G. A) Distribution of *RMSEF* values by sample (Reponses for different [M]$_T$ distributions aggregated by sample). Numbers of responses within following ranges of *RMSEF* values are shown: Exact, *RMSEF* = 0 (Dark blue); Precise, 0 < *RMSEF* ≤ 0.1% (Light blue); Good, 0.1% < *RMSEF* ≤ 1.0% (turquoise); Imprecise, 1% < *RMSEF* ≤ 10% (Yellow), and Poor, *RMSEF* > 10% (Red). B) RMSEF of different subsets of responses by sample. Average of results with too few ligands (Red), and median of results with correct model, i.e., S and number of ligands (Blue). C) RMS Error in estimates of log(ligand parameters) and predicted *p*Cu. Average of responses within indicated range of *RMSEF* (same color scheme as in A) . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of ligands and, with one exception, exact or precise fits to the data (Fig. 5A). All of the exact responses came from 5 participants, some of whom used linear regression (Ψ^{L}) and others non-linear (Ψ^{N}) approaches. Since direct modeling (Ψ^{M}) also yielded precise fits, each of the basic mathematical approaches of Table 1 passed this very simple test. The few poor fits, i.e., the 6.5% of responses with *RMSEF* greater than 10%, arose from one participant's use of *S* values derived using simple internal calibration (Ψ^{N}_{FC1}) rather than the provided, 'true' *S* of unity (see Sections 2.4 and 3.3).

Fitting the samples containing 2 ligands (D–F) proved to be more challenging. Of the complete responses submitted by 13 participants, 20% are exact fits, 29% precise, and another 7% good (Fig. 5A). Two participants submitted only exact responses while 4 more submitted a mixture of exact and precise results; all 6 used either non-linear regression or direct modeling. The 44% of responses that are imprecise or poor fits

ARTICLE IN PRESS

appear to have been from participants that encountered problems of one kind or another while using both linear and non-linear regression. The participant (*g*) that adopted internal calibration (S^{ic}) in A–C did the same for D–F and obtained imprecise fits in 4 of 9 titrations and poor fits in the remaining 5. Two participants (*k* and *l*) applied a one-ligand model to every titration in the intercomparison. Consequently, the quality of their responses is mostly 'poor' and rose to the level of 'imprecise' only for sample D_1, which had the lowest difference between K_1 and K_2.

For the sample containing 3 ligands (G), 11 participants submitted a response, of which 3 are exact, 1 precise, and 1 good fits, while the remaining 6 describe the data poorly (Fig. 5A). All exact to good results came from participants using either non-linear regression or direct modeling; poor results were derived using both linear and non-linear regression. Parameters for the correct three-ligand model were reported by 7 participants, with parameters for two- and four-ligand models reported by 3 and 1 participants respectively. Two of the responses with correct models fit very poorly, apparently because the linearizations used to derive them sufficed for fitting L_1 and L_3 at the ends of the titration curve but not in the middle. As for A–F, simple internal calibration yielded poor *RMSEF*.

4.1.2. Assessing error and causes of poor fit

To discern what poor *RMSEF* values mean, we need to consider the two main steps in the process of data analysis — calibration and modeling — separately. Just as in analyzing real titration data, *RMSEF* values calculated here reflect differences between measured and back-calculated I_p, or equivalently $[M_{meas}]$. An incorrect calibration, or biased *S*, leads to incorrect $[M_{meas}]$ calculated from Eq. (16), which subsequently causes bias in estimated ligand parameters. However, a complexation model that closely fits incorrect $[M_{meas}]$ will also fit the original I_p closely, since the initial effect of incorrect *S* is exactly cancelled out when back-calculating I_p. Thus, *RMSEF* cannot diagnose calibration bias, but it can indicate the presence of other deficiencies in the fitted complexation model.

Now, all but 2 participants used the provided *S* of unity to analyze the ASV-type titrations. Participant *n* employed simultaneous calibration+modeling (Ψ^{N}_{FCM2}) of these data and obtained exact fits, along with correct slopes and parameter estimates. The one participant, *g*, that used simple internal calibration submitted responses with poor *RMSEF*. Since these were uncorrelated with the extent of bias in *S* and the number of ligands was correct in most cases, it appears that the ligand parameters were fitted incorrectly.

The well-known importance of matching the number of ligand classes in the speciation model with that of the sample is also confirmed here. Perhaps with the intent of providing examples demonstrating that too-simple models are inadequate, two participants chose to never include more than one ligand in the speciation models they applied to analyze the titrations from all ASV-type samples. Those responses yielded very poor fits (Fig. 5B) as did those from 2 additional participants that included either one too few or one too many ligands in analyzing sample G. In all of these cases, the participant could have realized that the parameter estimates were of low quality simply by simulating the titration data and computing *RMSEF*.

Since all approaches to data analysis involve optimizing model parameters in order to fit observations, it may seem surprising that any response should have a poor *RMSEF*. By virtue of the wide range in *RMSEF* values, clearly back calculation and computation of *RMSEF* could have at least identified the lower quality responses and helped the participant avoid the biased estimates of ligand parameters and predicted environmental *p*Cu that occur only in the poorly-fitting responses (Fig. 5C). Note that while all parameter estimates from poor fits exhibit an average bias of 0.1 log units or more, some like K_1, K_2, and $[L_2]_T$, are much more sensitive, with a bias of 0.1 occurring when *RMSEF* exceeds just 0.1%. Perhaps the explanation lies in the fact that a modeler using the Ψ^{L} or Ψ^{N} approaches does not automatically optimize *RMSEF* during the fitting process, while those using Ψ^{M} do. Thus, the former group should be especially careful to simulate the fitted titration curve after analysis and compute *RMSEF*, rather than rely solely on optimizing the fit to the transformed data.

4.1.3. Visual inspection of data and simulation (back calculation) of titration curves

It is well known among statisticians that when assessing model quality, conventional statistical metrics should be supplemented by visually inspecting the fit of model curves to observational data, particularly when fitting transformed data. A graphical comparison of the observations for sample F_1 with simulated (back calculated) titration curves provides a case in point (Fig. 6A–F). The curves of participants *e*, *f*, and *j* all depict two-ligand models, but are from the high *RMSEF* group discussed above. Participant *k* fitted a one-ligand model to the data. On the linear $[M_{meas}]$–$[M]_T$ plot (Fig. 6A), the poor fit of curve *j* is apparent. The others appear to fit the raw data reasonably well despite the fact that *k* includes too few ligands. The poor fit of curves *f*, *j*, and *k* become apparent at low $[M_{meas}]$ on the log $[M_{meas}]$–linear $[M]_T$ plot (Fig. 6B). Similarly, poor fits of model curves at either end of a titration may be diagnosed in different plots of transformed data, with the log-LG (Fig. 6D), VDBR (Fig. 6E), and Scatchard (Fig. 6F) plots being best at low $[M_{meas}]$ and the linear-LG (Fig. 6C) at high $[M_{meas}]$. The very poor fit of curve *j* in all plots except the Scatchard plot (Fig. 6F) strongly suggests that the participant used this transformation to fit the data, but neglected to create any alternative visualizations. Thus, we highly recommend examining the quality both of one's parameter estimates using multiple visualizations, a step easily performed using ProMCC, and of the *RMSEF* when assessing one's interpretation of this type of data.

Of course, taking steps to correct poor model fits is crucial. Since non-linear fitting depends on initial guesses, it may suffice to choose an alternate initial guess and re-estimate ligand parameters. Removal of data points, which quite a few participants reported doing for the MW1 dataset, should not be undertaken just to improve model fit since doing so indulges confirmation bias. Statistical tests proving that a given point is truly an outlier are available and should be applied if data are suspect (Miller, 1993).

4.1.4. Detection of ligand classes

One aspect of titration curve interpretation that has received little attention in the trace metal speciation literature is the process of deciding how many ligands to report. Visualization can help show when too few ligands are included in a model, such as curve *k* in Fig. 6, and many workers report testing for the presence of more than one ligand class by creating a Scatchard plot. However, workers in this field do not commonly report the use of statistical criteria to decide whether or not a class was truly detected. Objective means of deciding which model to employ include i) comparing parity-adjusted RMS errors (Sander et al., 2011), ii) applying the F-test for regressions, and iii) simply restricting the model to ligands whose parameters are statistically different from zero (Hudson et al., 2003). Statistical methods designed for selecting between multiple models should be given serious consideration, such as the Akaike Information Criterion (Aho et al., 2014).

4.2. Analysis of individual ACSV titration curves (noised data)

The ACSV-type titrations of sample MW1 are the most realistic of the 3 multiwindow datasets and therefore it is the participants' responses to these data that we analyze herein. While acknowledging the limited nature of what can be proven by any analysis, we seek to discern whether or not methodological differences affected the quality of the responses submitted.

Please cite this article as: Pižeta, I., et al., Interpretation of complexometric titration data: An intercomparison of methods for estimating models of trace metal complexation by ..., Mar. Chem. (2015), http://dx.doi.org/10.1016/j.marchem.2015.03.006

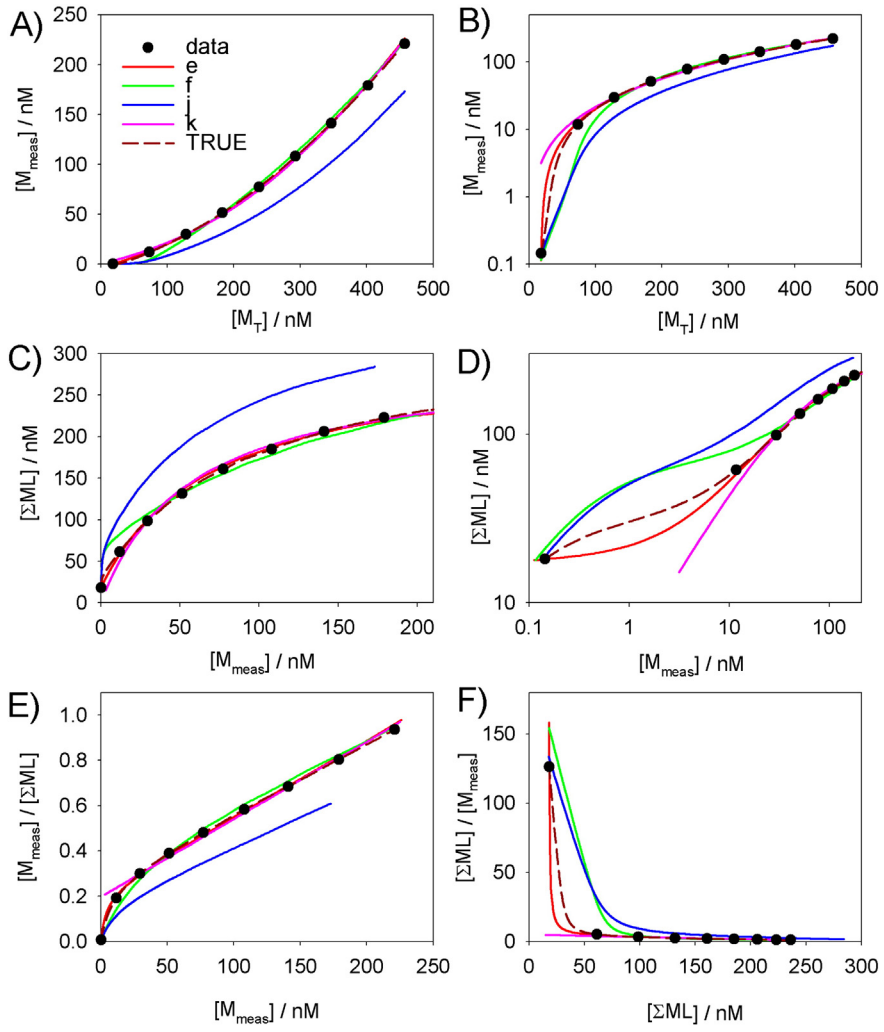

Fig. 6. Use of transformations to visually compare ASV-type data from titration F_1 (●) to simulated curves generated from parameter estimates of participants (*e*, *f*, *j*, *k*) and from the 'true' parameters. A) and B) $[M_{meas}]$–$[M]_T$ plots in lin–lin and log–lin formats respectively; C) and D) Langmuir isotherm/Gerringa plots in lin–lin and log–lin formats, respectively; E) van den Berg/Ružić plot; F) Scatchard plot. Parameter values (L_1 (nM), log K_1, L_2 (nM) and logK_2) submitted by participant *e*) 18, 11.44, 280, 7.15; *f*) 66.1, 9.5, 315.9, 6.71; *j*) 63.2, 9.45, 322.2, 7.10; *k*) 291.6, 7.24 and 'true' values: 30, 10, 300, 7) (see Appendix 2 to examine results in more detail). S = 1 assumed in all cases. *RMSEF* values are 0.11, 0.65, 0.80, and 0.22 for *e*, *f*, *j*, and *k* respectively.

4.2.1. Overview of participants' responses

Accurately modeling the five ACSV-type titrations of the MW1 sample (Fig. 3) is much more challenging than the ASV-type data because of i) the mismatch between α′ and ligand competition strength in some ACSV titrations and ii) the random measurement error included in the ACSV data. Note that all five titrations were performed on the same virtual seawater 'sample' but at different levels of $[AL]_T$ or detection windows (α′) (Table 3). Thus, although the I_p-$[Cu]_T$ curves differ markedly in appearance (Fig. 3), they are not independent and should in principle yield identical parameter estimates – S_{max}, number of ligand classes, and ligand parameters – and predicted *p*Cu. Based on their analyses of these 5 curves, 21 participants submitted up to 6 responses for MW1, each comprising ligand parameters (K_1, K_2, $[L_1]_T$, $[L_2]_T$) plus assumed or fitted *S*.

The *p*Cu5 values (N = 82) are for the most part clustered near the correct value (Fig. 7A, Table 4), although the spread in the central part of the *p*Cu5 distribution is somewhat broader than expected from the 3% measurement error inherent in the data. There are also five anomalous *p*Cu5 values that are biased low by >3 log units. All come from high α′ windows that have strongly low-biased $[L_1]_T$ (Fig. 7C) and where L_2 was not detected (Fig. 7D). In contrast to the *p*Cu5 predictions, the *p*Cu50 values (Fig. 7B) are generally more scattered, with the most biased values of *p*Cu50 coming from fits of the highest α′ windows (curves 3–5) in which L_2 was not detected (Fig. 7D). Overall, for results that include only L_1, the mean bias in *p*Cu5, or average E_{pCu5}, is −0.6 log units and the *RMS*-E_{pCu5} is 1.5. Among results that include both ligands, the mean bias in *p*Cu5 is only 0.0006 and *RMS*-E_{pCu5} only 0.049.

An obvious question to ask is what factor(s) determined whether participants included one or two ligand classes in their models? While the problem of window-ligand mismatch is well known and clearly affected attempts to model the higher $[AL]_T$ curves, another influential factor proves to be whether, for each curve, *S* was set equal to R_{AL} (S_{max} = 1) or calibrated. Of the 21 participants, 14 adopted the given $S = R_{AL}$ (N = 50 responses) and 7 used calibrated *S* values (N = 32 responses). All responses using $S = R_{AL}$ included parameters for two ligands in windows 1 and 2, but the L_2 detection rate declined monotonically from 87 to 14% between windows 3 and 5 (Fig. 8). One response reported that no ligand was detectable in window 5. It is hardly surprising that ligands are difficult to detect when competition strength and α′ are mismatched (van den Berg and Donat, 1992), as is especially the case for the weaker ligand class (L_2) at the highest α′ (window 5).

The 7 participants that employed internal calibration of *S* included L_2 parameters less frequently. In 3 cases, L_2 was not reported in any

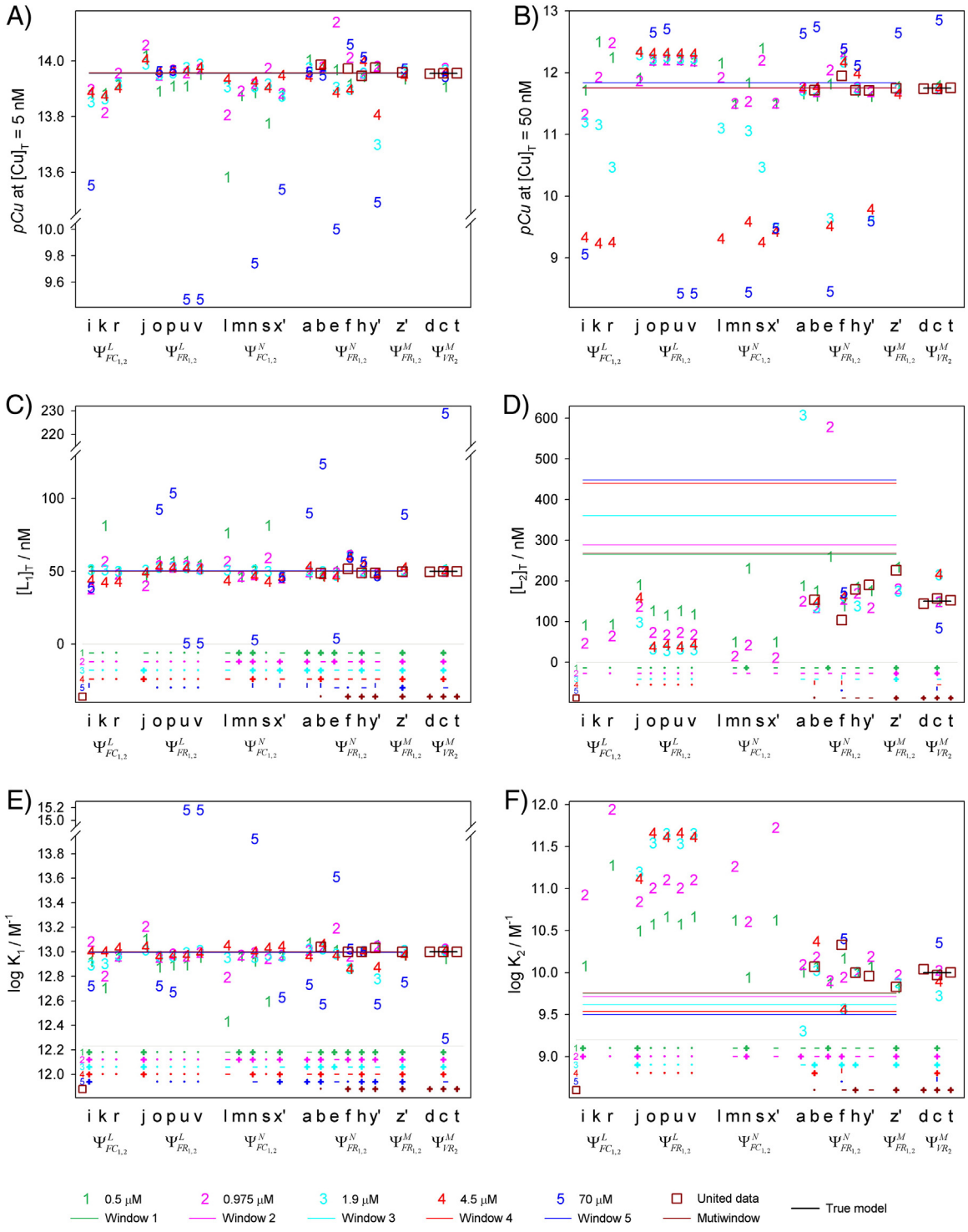

Fig. 7. Results from all participants' responses for MW1 dataset: A) *p*Cu at 5 nM; B) *p*Cu at 50 nM; C) $[L_1]_T$; D) $[L_2]_T$; E) log K_1; F) log K_2. Participants *i*, *k* and *r* (Ψ^{L}_{FCn}); *j*, *o*, *p*, *u* and *v* (Ψ^{L}_{FRn}); *l*, *m*, *n*, *s* and *x*′ (Ψ^{N}_{FCn}); *a*, *b*, *e*, *f*, *h* and *y*′ (Ψ^{N}_{FRn}); *z*′ (Ψ^{M}_{FRn}); *d*, *c* and *t* (Ψ^{M}_{FV2}). Results for different windows are represented by corresponding numbers, results for unified solutions by squares, 'correct' parameters for all windows for the alternate model are represented by lines with the same color code as detection window numbers (see Table 3). (See text for explanation of Ψ indexing.) Test of whether 'correct' value lies within reported confidence limits: (+) Yes, (−) No, (•) CL not reported, (|) CL very large, (blank) parameter not reported. Color of the numeral or line corresponds to color code of corresponding window. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

window, reflecting the participants' consistent application of a one-ligand model across all windows, rather than a failure to detect L_2. In the remaining 4 responses, both ligands were detected in windows 1 and 2, but L_2 was not reported in any parameter sets from windows 3–5. Since internal calibration effectively masks weak, untitrated ligands by including their effects in the adjusted *S* (Voelker and Kogut, 2001), it is not surprising to find a lower detection rate for L_2 in these responses. Note that these responses were obtained using both linear and nonlinear regression approaches.

Taken together these results suggest that the success of calibration, which in real samples dictates that the analyst estimate a value of *S* from the titration data, is fundamental to the success of the modeling

Table 4
Accuracy of MW1 parameter estimates. Average parameter error, $\bar{E}_P$, and RMS-E_P (in parentheses) calculated using Eqs. (25) and (26). [a] Both are computed from log-transformed variables, i.e., log $[L_i]_T$, log K_i, *p*Cu5, log KL_1, etc.

	Correct value	Responses included in mean		
		All	L_2 in model	L_2 not in model
*p*Cu5	13.956	−0.233 (0.95)	0.001 (0.05)	−**0.599** (1.52)
*p*Cu50	11.749	−0.248 (1.18)	**0.254** (0.40)	−**1.032** (1.82)
$[L_1]_T$	50	−0.045 (0.37)	0.020 (0.10)	−0.146 (0.59)
K_1	13.00	0.010 (0.39)	−0.026 (0.12)	0.067 (0.61)
KL_1	5.699	−0.035 (0.11)	−0.006 (0.05)	−**0.079** (0.17)
$[L_2]_T$	150/268[b]	−0.396 (0.53)	−**0.396** (0.53)	–
K_2	10.00/9.76[b]	0.804 (1.07)	**0.804** (1.07)	–
KL_2	3.176	0.415 (0.62)	**0.415** (0.62)	–
N		82	50	32

[a] Bold number indicates $\bar{E}_P$ different from zero at 0.05 level by *t*-test.
[b] True/alternate model values (see Table 2). Errors in L_2 parameters are computed from correct values in Table 3 depending on the use of variable/fixed α′ model in each parameter set.

effort. While the goal of internal calibration is to obtain an accurate sensitivity that accounts for surfactants present in the sample matrix, to the extent that any slope obtained by internal calibration is biased by untitrated weak ligands, the chances of detecting L_2 will also be reduced. This in turn increases the error in both *p*Cu5 and *p*Cu50 and propagates error into estimates of the L_2 parameters. Thus, the responses with calibrated slope provide us with a scenario closer to real-life, i.e., where *S* is truly unknown and its estimate is biased low. Responses based on adopting the known $S = R_{AL}$ provide us with a less realistic scenario, but a better dataset for comparing methods of estimating parameter values. Both sets of results are valuable, but need to be considered somewhat differently.

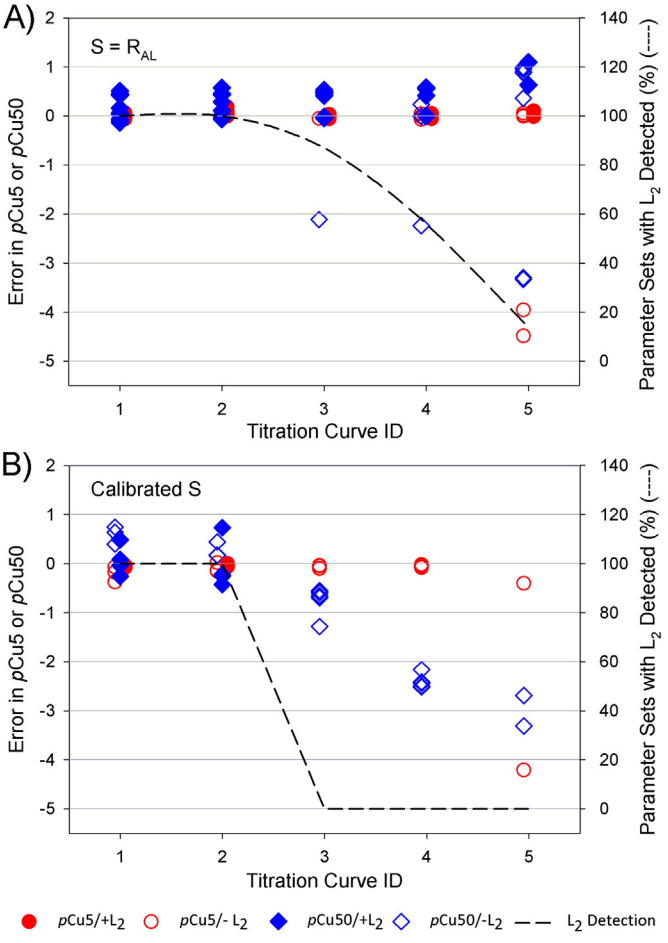

Fig. 8. Error in *p*Cu 5 nM and 50 nM by titration curves (windows) of MW1. A) results of responses with reference slope; B) results of responses with calibrated (internal) slope. Both L_1 and L_2 detected (blue diamonds), L_2 not detected (red dots). Right axis of each plot corresponds to percentage of parameters sets with L_2 detected. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.2.2. Results with true sensitivity adopted

Among the responses with $S = R_{AL}$ adopted and L_2 detected (Table 5), there are sufficient results to compare the three basic approaches to analyzing the data (Ψ^L_{FR2}, Ψ^N_{FR2} and Ψ^M_{VR2}), but not to discriminate at any deeper level of detail, e.g., type of linear transformation. It is apparent (Fig. 7) that the *p*Cu5 values for these responses are clustered tightly around the correct value, with mean errors of +0.05, +0.04, and +0.03 for the three approaches respectively (Table 5). Neither the biases in *p*Cu5 nor the method-specific mean errors of the L_1 parameters are significantly different from zero. Note that the mean bias in L_1 competition strength (KL_1) for each method is also non-significant. As expected, *p*Cu5 and log (KL_1) for the individual responses (Fig. 4) are linearly correlated ($R^2 = 0.98$; slope = 0.94).

In contrast to L_1, most estimates of L_2 parameters are biased. The Ψ^L_{FR2} and Ψ^N_{FR2} estimates of $[L_2]_T$ were biased 54 to 70% lower than the correct $[L_2]_T$ of 268 nM and values of log K_2 were 1.3 to 0.4 log units higher than the correct value of $10^{9.76}$ M^{-1}. The KL_2 results of the Ψ^L_{FR2} approach had a very high bias of $10^{+0.755}$ and its parameter-specific RMSE was also largest, reflecting a combination of high bias and high variability. On the *p*Cu50-logKL_2 plot (Fig. 4), these results fall along the ideal line at log KL_2 near 4.0 and *p*Cu50 near 12.2, illustrating how a narrow range in *p*Cu50 (±0.1) can yield a wide range in KL_2 (±0.4) and ultimately in K_2 (Table 4). In contrast, the log KL_2 of the Ψ^N_{FR2} approach had a non-significant mean bias and mostly fell close to the ideal *p*Cu50-logKL_2 line on Fig. 4 at log KL_2 near 3.2. Apart from these clusters, the scatter in remaining points of the $S = R_{AL}$ group mostly is caused by the fixed-α′ parameter sets with high-biased $[L_1]_T$.

The poor quality of the $[L_2]_T$ estimates for both Ψ^L_{FR2} and Ψ^N_{FR2} methods is somewhat puzzling. In fact, only three parameter sets based on $S = R_{AL}$, fixed-α′ models had $[L_2]_T$ in excess of 200 nM. The responses of pseudo-participant *y*′ had $[L_2]_T$ from 133 to 175 nM and even the three parameter sets submitted by *z*′ (Ψ^M_{FR2}) only spanned a range in $[L_2]_T$ from 173 to 233 nM. Upon close examination of the data using ProMCC software, we found that the systematic low bias resulted in large part from the peculiarities of these noised data on the fixed α′ models.

Finally, we conclude our discussion of approaches that used $S = R_{AL}$ by noting that the lowest bias was exhibited by the two responses applying direct modeling (Ψ^M). The mean bias for response c) (Ψ^M_{VR2}) was non-significant for all parameters, with essentially exact means for $[L_2]_T$ of 153 nM and K_2 of $10^{9.91}$ M^{-1}. Since it employed the fixed α′ model, response z′ was not quite as accurate, but it did yield tightly clustered parameters and predictions (Fig. 7).

4.2.3. Results with sensitivity calibrated

Three participants submitted 7 responses with calibrated sensitivities specified for each MW1 window. Two of them derived simple internal calibration slopes for each window; the third submitted 2 responses derived using simultaneous calibration with either linearizations (*n*) or non-linear fitting (*i*). With one exception, the slopes were all near the internal calibration slopes (Table 3), which correspond to $S_{max} = 0.79$ to 0.87 or log $S_{max} = -0.06$ to −0.10. The exception corresponds to the optimal S_{max} for window MW1-1 (see discussion of Fig. 9 below).

Consideration of the equations relating *S* to $[Cu_f]$ and $[Cu_f]$ to KL_1, Eqs. (16) and (20), reveals that *p*Cu5 and log KL_1 should be proportional to the value of log *S* adopted in transforming the data. In fact, within the population of all parameter sets for MW1-1 to MW1-4, *p*Cu5 and KL_1 are linearly correlated with log *S* ($P < 0.0001$; slope near unity). This

Table 5
Analysis of bias in MW1 parameter estimates with L_1 and L_2 detected (window 5 excluded). Average error, $\bar{E}_P$, and RMS-E_P (in parentheses) calculated using Eqs. (25) and (26).[a] Both are computed from log-transformed variables, i.e., log $[L_i]_T$, log K_i, *p*Cu5, log KL_1, etc.

	Correct value[b]	Responses adopting $S = R_{AL}$			Responses with calibrated S	
		Linear (Ψ^L_{FR2})	Nonlinear (Ψ^N_{FR2})	Direct Model (Ψ^M_{VR2})	Linear (Ψ^L_{FC2})	Nonlinear (Ψ^N_{FC2})
*p*Cu5	13.956	0.017 (0.05)	0.002 (0.03)	−0.005 (0.03)	−0.037 (0.04)	**−0.070** (0.07)
*p*Cu50	11.749	**0.427** (0.45)	**0.105** (0.26)	−0.003 (0.02)	**0.270** (0.46)	−0.246 (0.28)
$[L_1]_T$	50	0.013 (0.03)	0.014 (0.05)	0.001 (0.01)	−0.012 (0.02)	**−0.047** (0.07)
K_1	13.00	−0.007 (0.08)	−0.014 (0.07)	−0.007 (0.03)	−0.044 (0.05)	−0.018 (0.06)
KL_1	5.699	0.006 (0.05)	0.000 (0.04)	−0.005 (0.03)	−0.056 (0.06)	**−0.065** (0.07)
$[L_2]_T$	150/268	**−0.531** (0.63)	**−0.196** (0.25)	0.075 (0.11)	**−0.493** (0.57)	**−0.817** (0.87)
K_2	10.00/9.76	**1.288** (1.40)	0.191 (0.33)	−0.099 (0.15)	**1.180** (1.40)	**0.962** (1.06)
KL_2	3.176	**0.766** (0.81)	0.003 (0.20)	−0.024 (0.07)	**0.695** (0.95)	0.154 (0.25)
N		22	14	4	4	4

[a] Bold text indicates mean error of parameter is different from zero at $P < 0.05$ level by *t*-test.
[b] True/alternate model values (cf. Table 2). Error in L_2 parameters computed from correct values in Table 2 depending on the use of variable/fixed α′ model in each parameter set.

confirms the subtle, but deterministic bias in L_1 parameters introduced by the incorrect value of S from internal calibration. Although in this exercise this bias is only 10–20%, even small biases of this magnitude should be avoided if reasonable effort permits.

As discussed above, adjusting S also has indirect effects on predictions and parameters due to the indirect effects it has on $[L_2]_T$. As has been shown previously (Voelker and Kogut, 2001), low-biased S values lead to low-biased estimates of $[L_2]_T$. This is readily apparent from a comparison of $[L_2]_T$ reported for the Ψ^N_{FC2} and Ψ^N_{FR2} approaches (Fig. 7; Table 5). For windows MW1-1 and MW1-2, $[L_2]_T$ obtained using calibration are all <50 nM while using S equal to R_{AL} gives values above 150 nM. Since KL_2 is constrained by observations in the mid-$[Cu]_T$ range, bias in $[L_2]_T$ propagates into K_2 as well, making it higher where calibration was used. These inferences are supported by comparing the responses of pseudo-participants x' (Ψ^N_{FCn}) and y' (Ψ^N_{FRn}). Both employed non-linear regression approaches, but the former obtained very low values of $[L_2]_T$ and much more biased values of K_2 (Fig. 7).

4.2.4. Effects of adjusting S on estimates of ligand parameters

The calibration of S, while widely known to be problematic, has effects that remain poorly understood. For a two-ligand model, the error of fit weakly depends on S (Fig. 9). For example, for MW1-1, the noisy data causes fits to have a barely perceptible minimum at $S_{max} = 1.083$, while the fit to window MW1-3 has a minimum at $S_{max} = 0.89$. Thus, realistic confidence limits (CL) on S when fitted to individual curves must be at least ±0.1 or so for the MW1 data. When combined with the already substantial CL generated by the fitting software for each parameter at fixed S, the uncertainty in S leads to even larger CL for the parameters. The difficulty that nonlinear regression software can have in converging for a two-ligand model applied to curves like the MW1 titrations reflects interactions of S and ligand parameters. The very weak dependence on S, particularly at higher α′, leads to convergence problems when attempting to model S and ligand parameters simultaneously.

4.2.5. Visual inspection of data and simulation (back calculation) of titration curves

While to this point, the focus has been on effects of calibration, we now consider an important aspect of the art of analyzing titration curves using non-linear regression tools, namely deriving initial guesses and recursively improving fits. To this end, we examine the data of the intermediate MW1-3 window ($[SA] = 1.9$ μM) and the characteristic parameter estimates of six different participants (Fig. 10A to D). In addition to plotting the back-calculated titration curves and given data (Fig. 10A), we used the reported parameters to generate points that were then transformed into LG, VDBR, and Scatchard plots using ProMCC software (Omanović et al., in this issue). Looking at the curves in Fig. 10A, all but one result seems to have successfully fitted the data points. A less satisfactory picture is found in Fig. 10B–D, where different transformations accentuate different segments of data points, and show a significant deviation of some model curves from the data points. Clearly, by using only transformation and visualization of the model fit to data some of the points at either end of the curve may be overlooked and the correct model missed. An example is the result of y', who fitted a single ligand to the curve and as a result has a poor fit over part of it. ProMCC is designed to make convenient the process of generating multiple visualizations, adjusting initial guesses, and re-estimating model parameters.

Sensitivity, which unintentionally became a factor in this study, is a very important parameter and if not properly determined influences fitted ligand parameters as well. A mis-calibrated slope can sometimes be detected by visual inspection of the aforementioned transformations (VDBR, Scatchard, and/or LG), i.e., VDBR-transformed curves become concave instead of convex, LG-transformed points start to decrease toward higher $[M_{meas}]$ values instead of increasing toward total capacity, and Scatchard-transformed data can appear to bend backwards when S is too low. Unfortunately, a slope that is too high is hard to detect in this way.

4.3. Unified analysis of multiple window datasets

When fitting a single titration curve, neither simple internal calibration (Table 3) nor fully coupled calibration/direct modeling (Fig. 9) can guarantee accurate estimates of S. However, by considering even two curves with distinct α′ simultaneously, the composite error function develops a distinct minimum near the correct slope (Fig. 9A). With more windows, the minimum becomes even better defined and the uncertainties in model parameters greatly diminished (Hudson, et al., 2003; Sander et al., 2011). Using the 5 windows of the MW1 dataset, participant t applied simultaneous calibration/modeling to obtain an S_{max} that is correct to within 0.01%, estimates of all ligand parameters that are effectively exact, and a mean fitting error (*RMS-RE* = 3.12%) nearly as low as the purely experimental error of 3.11% (Table 6).

An excellent way to visualize the virtues of unified multi-window analysis is to plot the data and models for all windows on the same *p*Cu-[$Cu_{natural}$] plot (Fig. 11). As we saw previously (Fig. 7), analysis of single curves at different α′ can yield disparate estimates of ligand parameters, even when the correct slope is known a priori. Such discrepancies are worst when a ligand class is not detected, as often occurred for windows MW1-4 and MW1-5 (Fig. 11A). Participants f, h, and y' all used non-linear regression to model all 5 MW1 curves simultaneously. The high quality of their results (Fig. 7, h in Fig. 11A, Table 6) confirms that the criterion of coherence between windows is robust and powerful. The very good predictions of *p*Cu made by participants h and y' are also noteworthy. Participant b also showed the value of this criterion even when performed manually. By averaging the best-fitting single window results, b obtained a very good overall fit to the complete dataset (*RMS-RE* = 4.0%).

Please cite this article as: Pižeta, I., et al., Interpretation of complexometric titration data: An intercomparison of methods for estimating models of trace metal complexation by ..., Mar. Chem. (2015), http://dx.doi.org/10.1016/j.marchem.2015.03.006

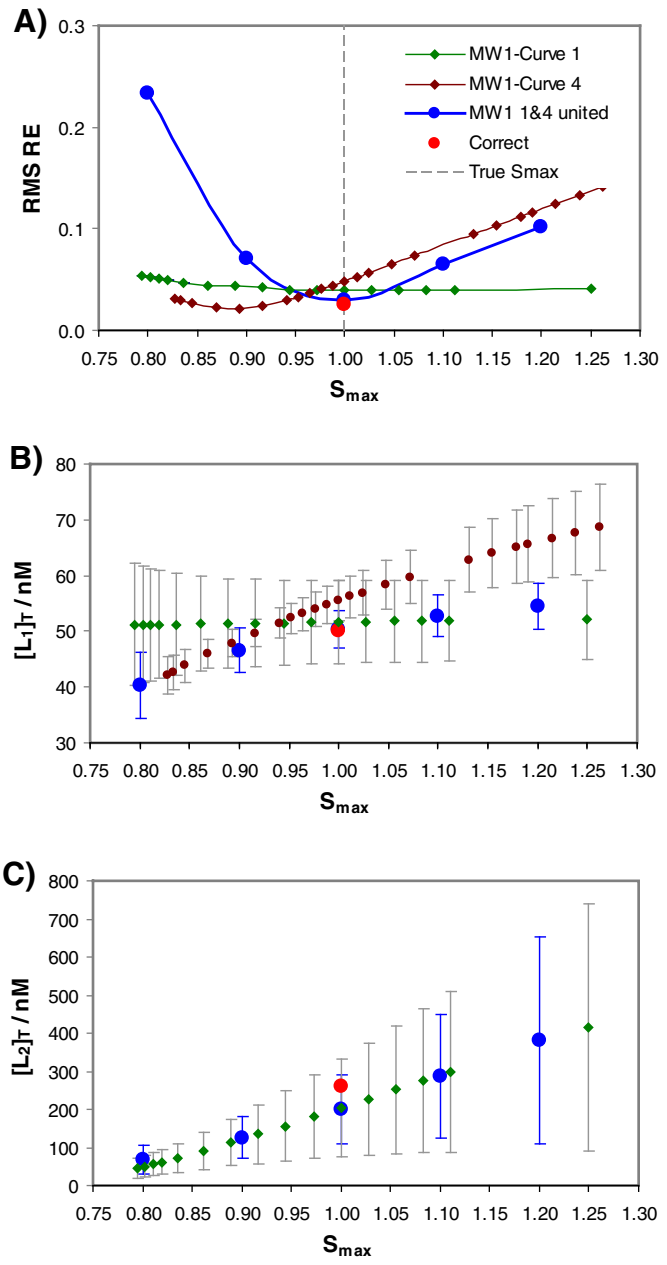

Fig. 9. Effects of adjusting sensitivity (S) on A) RMS-RE of fit and B) estimated $[L_1]_T$ and C) $[L_2]_T$. Parameter estimates and uncertainties fitted at various fixed S_{max} values using Ψ^{N}_{FR2} approach with ProMCC. Results with error bars include: MW1 titration curves 1 (●) and 4 (●) fitted individually and united fit of the same two curves (●). Optimal S_{max} for each curve or combination of curves corresponds to the minimum in the RMS-RE function. Correct 'alternate' model parameter values (●). True S_{max} indicated by vertical line.

Analyzing single titration curves when S_{max} is not known is more challenging, however. Biased calibrations can be identified as suspect by visual inspection of speciation data on $[Cu_f]$–$[Cu_{natural}]$ plots (Fig. 11B). These transformations of MW1 data made using biased S^{ic} values (Table 3) obviously lack coherence between windows, especially at the upper end of each curve. The estimated ligand parameters also yield widely disparate curves for $[Cu_f]$ as a function of $[Cu_{natural}]$. None of the participants that modeled single curves with calibrated S reported unified parameters for MW1.

As discussed above, participant *t* was able to successfully calibrate S_{max} by directly modeling the MW1 data in a unified fashion. This approach yields the simplest possible model that explains all of the data in a coherent manner, as is evident from the $[Cu_f]$–$[Cu_{natural}]$ plot (Fig. 11C). Two other participants, *c* and *d*, also employed direct modeling, but without simultaneous calibration, instead accepting the given S_{max} of 1. The accuracy of their ligand parameter estimates and the very close fit to the titration data (*RMS-RE* = 3.14 and 3.16% respectively) confirm that the criterion of coherence between analytical windows is powerful and that direct modeling of multiple windows is potentially very accurate.

In summary, only 6 of the 15 participating laboratories, plus two pseudo-participants, reported results for a unified analysis of the five MW1 titrations. In all cases, they yielded very good to excellent predictions of *p*Cu5 and *p*Cu50 and ligand parameters (Fig. 7; Table 6), despite the inclusion of the problematical data from window MW1-5. Consider the example of participant *c*, whose *p*Cu50 is more than a log unit high when calculated from window 5 alone (Fig. 7), but within 0.01 log units of the correct value based on the unified analysis performed on windows 1–5. Participants *b*, *h*, *y′*, and *z′* also had poor predictions for window 5, but quite good predictions for the unified datasets. The implied robustness of the unified multiwindow approach is heartening, since it suggests that even including single titration curves that are very difficult to interpret can be useful. While one would not intentionally design experiments to generate such data, it is helpful to know that it is not necessary to exclude such portions of a dataset and thereby run the risk of introducing confirmation bias.

Simultaneous analysis of multiple window titrations is not commonly performed, except by a few practitioners who use direct modeling. However, two participants in this intercomparison did in fact use a non-linear regression approach in simultaneous multiwindow analysis. To do so, $[M_f]$ was calculated from $[M_{meas}]$ and fixed α', $[\Sigma ML]$ was calculated as indicated by Eq. (17), and all points analyzed using non-linear regression software (Appendix 3). Since $[M_{meas}]$ must be calculated prior to inputting to the regression software, participants using non-linear regression were forced to use the given $S_{max} = 1$. The resultant *RMS-RE* were good (*f*) to very good (*h*). Simultaneous calibration/speciation modeling could also be performed using MW1 data using a recursive LG algorithm, such as developed by Laglera et al. (2013).

The use of information from multiple windows as a means of more accurately calibrating voltammetric data was pioneered by Kogut and Voelker (2001), who proposed the 'overload titration.' To test their experimental design, we analyzed data from windows MW1-1 and MW1-5. Despite the very high $[AL]_T$ in window 5, the internal calibration slope remained low-biased, $S_{max} = 0.85$, due to the high $[L_1]_T$ in this sample. As a consequence, the $[L_2]_T$ derived from modeling window 1 is also low-biased and the *RMS-RE* of this fit (w1 + OV in Table 6) exceeds the experimental error of 3% by a factor of 4 or so. Interestingly, simultaneous calibration/direct modeling of the two curves together yields much improved fits to the data (w1 + w5 in Table 6).

Finally, it should be noted that excellent results were attained using both variable and fixed α' speciation models, with the same minimum RMS relative error of fit (0.0311), and that their estimates of ligand parameters are close to their respective 'correct' values. Note also that the fixed and variable α' models both yield excellent fits to the $[Cu_f]$–$[Cu_{natural}]$ plots (Fig. 11A, C).

5. Conclusions and recommendations

1. Analysis of complexometric titration data requires three key steps: i) calibration, ii) choice of mathematical transformation of the raw data, and iii) definition of the speciation model. Based on the responses submitted by participants for both noise-free and noised data, it is clear that properly performing all steps is essential for obtaining quality results.
2. Calibration, the first step, is more fraught than many realize. The common approach of estimating sensitivity (S) from the final points of a titration curve–formally 'internal calibration'– can cause underestimation of sensitivity when weak natural ligands are not

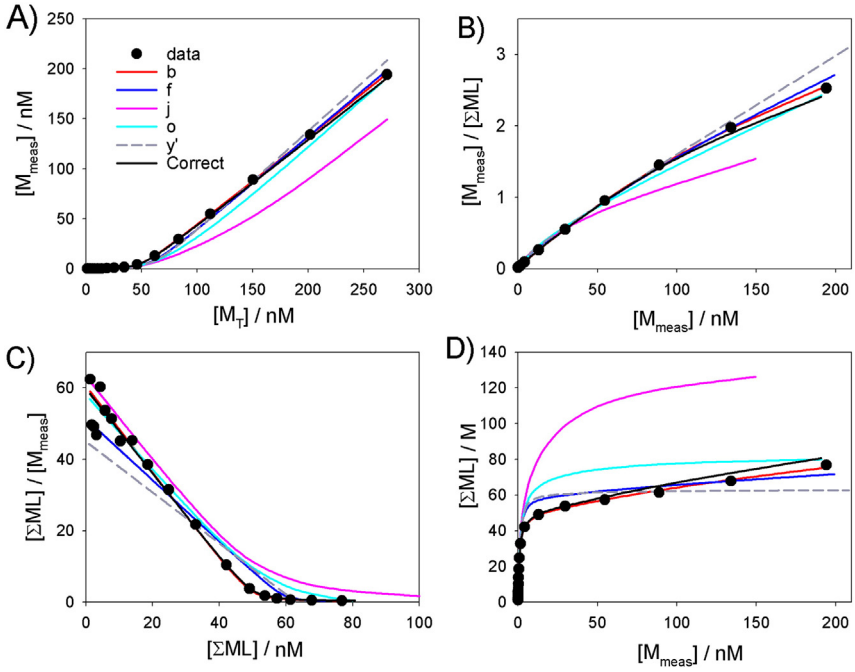

Fig. 10. Use of transformations to visually compare ACSV-type titration data from MW1-3 (●) to simulated curves back-calculated from parameter estimates submitted by participants *b*, *f*, *j*, *o*, *y′* and from 'Correct' alternate model parameters (see Table 2). A) linear $[M_f]$–$[M]_T$ plot; B) van den Berg/Ružić plot; C) Scatchard plot; and D) Langmuir isotherm/Gerringa plot. Parameter estimates (L_1 (nM), log K_1, L_2 (nM) and logK_2) from participant *b*) 49.65, 13.02, 133.91, 10.02; *f*) 59.57, 12.86, 161.52, 9.56; *j*) 50.60, 13.01, 97.17, 11.19; *o*) 52.46, 12.96, 31.2, 11.54; *y′*) 63.1, 12.78, L_2 not reported. All 'data' derived from $[M]_T$ and I_p using Eqs. (16) and (17) with S = 0.68.

completely titrated. Here, participants that used internal calibration had *S* values that were biased low by ~15% and predicted free metal ion concentrations that were proportionately high.

3. Participants' approaches to transformation spanned the full range of mathematical approaches available. High quality responses (sets of parameter estimates) derived using each of the major approaches were submitted, but the quality of submissions by those using the traditional linear transformation methods alone was considerably more biased and variable.
4. As for every titration of an environmental sample, participants had to decide the appropriate number of ligand classes to include in the speciation model used to fit the data. Instances of both under- and over-parameterizations, i.e., too few or too many ligand classes, were observed and contributed to a) poor fits to data, b) significant bias in parameter estimates, and c) inaccurate predictions of ambient metal speciation.
5. Likely out of habit, very few participants were aware that the added ligand became titrated within some of the ACSV titration curves analyzed here. This caused α′ to decline for high $[M]_T$–low $[AL]_T$ conditions and parameter estimates to be biased.
6. This problem is most likely to occur in samples with high DOM content, where it could seem to the analyst that metal additions of the same magnitude as the AL concentrations should be made. While matrix-based, direct modeling methods can compensate in such situations, the analyst should consider diluting the sample with UV-oxidized water of similar major ion composition to that of the sample in order to keep proper analytical conditions (Laglera et al., 2013). Addition of metals from stocks containing equimolar $[M]_T$ and $[AL]_T$ may also solve the problem.
7. Simulation (back-calculation) of the original titration data using one's fitted parameter estimates is a simple and effective step in evaluating their quality. If the number, distribution, or quality of the original data

Table 6
Parameter estimates for various unified multi-window data analyses of sample MW1 using a two-ligand model. Values shown are parameter estimates ± 95% confidence limits. Approach described in last 4 columns.

Participant	RMS- RE[a]	$[L_1]_T$ (nM)	log K_1 (M^{-1})	$[L_2]_T$ (nM)	log K_2 (M^{-1})	*p*Cu50	Calibrate	Transform	Math	Unification
Variable α′ Model										
c	0.0314	50.0 ± 1.6	13.00 ± 0.20	156.8 ± 14.4	9.97 ± 0.85	11.741	R	Direct	Matrix	Simultaneous fit
d	0.0316	49.8 ± 0.7	13.00 ± 0.02	144.0 ± 20.0	10.04 ± 0.09	11.738	R	Direct	Analytical	Simultaneous fit
t	0.0312	50.0 ± 0.4	13.00 ± 0.02	151.9 ± 15.7	10.00 ± 0.17	11.752	MW	Direct	Matrix	Simultaneous fit
w1 + w5	0.0365	50.2 ± 0.5	12.98 ± 0.02	133 ± 12.5	10.05 ± 0.06	11.753	MW	Direct	Analytical	Simultaneous fit
w1 + OV	0.122	50.1 ± 0.5	12.91 ± 0.02	73.3 ± 12.5	10.14 ± 0.12	11.623	OV	Direct	Analytical	Uncoupled
True	0.0311	50	13	150	10	11.748				
Fixed α′ model										
b	0.0397	48.8 ± 0.0	13.04 ± 0.00	153.7 ± 0.0	10.07 ± 0.00	11.716	R	RL	Inverted	Selective average
f	0.0699	51.7 ± 2.1	13.00 ± 0.05	103.5 ± 7.6	10.33 ± 0.09	11.947	R	LG	Inverted	
h	0.0380	49.0 ± 1.0	13.00 ± 0.07	179.0 ± 5.6	10.00 ± 0.08	11.714	R	RL	Inverted	
y′	0.0349	48.9 ± 0.0	13.03 ± 0.00	190.0 ± 0.0	9.96 ± 0.00	11.711	R	LG	Inverted	Simultaneous fit
z′	0.0316	49.9 ± 0.0	13.00 ± 0.00	225.8 ± 0.0	9.83 ± 0.00	11.746	R	Direct	Analytical	Simultaneous fit
Alternate	0.0311	50	13	268.1	9.756	11.756	R			

[a] RMS-Relative Error computed as in Eq. (24).

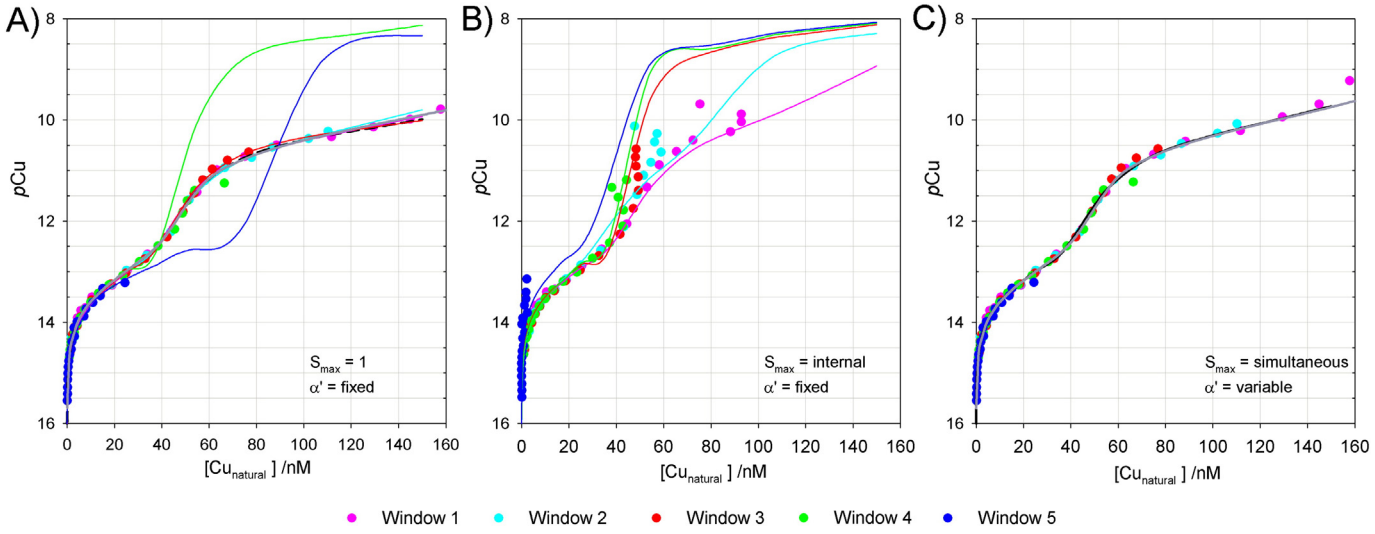

Fig. 11. $[Cu^{2+}]$–$[Cu_{Natural}]$ plots for all five MW1 titrations. A) I_p data transformed using $S_{max} = 1$ and fixed α′. Dashed black line is for correct alternate model; gray line is unified analysis of response *h*; colored lines represent single window results of response *a* (Ψ^{N}_{FRn}); B) I_p data transformed using S^{ic} for each curve and fixed α′. Colored lines represent single window results of response *i* (class Ψ^{L}_{FCn}); C) I_p data transformed using $S = 1.000762$ from unified multiwindow calibration/direct modeling with variable α′ as in response *t* (class Ψ^{M}_{FCM2}). Gray line is calculated from parameter estimates of *t*. Black line represents 'true' model.

does not match the back-calculated curve in all plot styles – linear or log axis scaling – it indicates that the parameters are inaccurate or the wrong number of ligands is included in the model. In such cases, the parameter estimation should be repeated with fewer or more ligands in the model and/or different initial guesses for parameters. Many of the inaccurate results seen in this intercomparison would have become obvious had the participants followed this simple rule. The program ProMCC is a suitable tool for the task and available in the public domain.

8. Along with visual inspection, the RMS relative error of the model fit –

Table 7
Definitions of variables and parameters.

Variable	Description of variable	Equation Number	Alternate variable names [a]
M	The metal in the oxidation state or compound under consideration.		
M_f	Free (aquo) metal ion	(1)	M^{z+}
MX_{in}	Complexes with all major inorganic anions	(6)	M_{IN}
M′	Aquo metal ion plus its inorganic complexes.	(6)	
ΣML	All 'unknown' metal complexes, i.e. with natural organic ligands.	(7)	
MAL_x	All metal-added ligand complex species	(11)	
M_{meas}	All well-defined, measurable metal species, specific to the analytical method used.	(9)	
$[M]_T$	Total concentration of M at each point in a titration curve.	(7)	C_M
AL	Added ligand in CLE-ACSV titrations		
AL′	All species of added ligand not complexing M	(13)	AL_f
$[AL]_T$	Total concentration of added ligand in all species including both complexes and unbound.	(15)	C_{AL}
X_{in}	Major inorganic ligands present in sample (OH^-, Cl^-, CO_3^{2-}, $B(OH)_4^-$, F^-, etc.).		X_i
L_i	Natural organic ligands of the *i*th class	(1)	
$[L_{i'}]$	Concentration of all species of *i*th ligand class not in a complex with M.	(2)	
$[L_i]_T$	Total concentration of *i*th natural organic ligand	(3)	
K_{ML_i}	Conditional stability constant of ML_i complex.	(4)	
K_{MAL}	Conditional stability constant of 1:1 MAL complex.	(13)	
β_{MAL_2}	Conditional stability constant of 1:2 MAL_2 complex.	(13)	
$\alpha' = \alpha_{MAL_x} + \alpha_{M'}$	Aggregate side reaction coefficient of all well-defined M species, i.e., M_{meas}	(12)	
α_{MAL_x}	Aggregate side reaction coefficient of all M-AL complexes.	(13)	
$\alpha_{M'}$	Side reaction coefficient for MX_{in} plus M_f.	(6)	α_M
S	Sensitivity (aggregate) for an arbitrary medium composition and $[AL]_T$.	(9, 15)	S
S_{max}	Maximum sensitivity (aggregate) possible under particular conditions, i.e., high $[AL]_T$.	(15)	
s^*_l	Sensitivity of individual species l	(8)	
R_{AL}	Ratio of sensitivity at a given $[AL]_T$ to sensitivity at highest analytical window (all in UVSW).	(15)	X
*p*Cu5	Negative logarithm of free cupric ion concentration at $[Cu]_T = 5$ nM; calculated from ligand parameters.		
*p*Cu50	Negative logarithm of free cupric ion concentration at $[Cu]_T = 50$ nM; calculated from ligand parameters.		
P_{est}	Estimated value of parameter.	(25)	
$P_{correct}$	'Correct' value of parameter *P*.	(25)	
E_P	Error of parameter estimate.	(25, 27)	
RPD	Relative percent difference between observed and simulated I_p value normalized by the average of the two.	(21)	
RMSEF	Root–mean–square of Error Function, i.e., square root of sum of squared RPD for all points in a dataset divided by degrees of freedom.	(22)	
RE	Relative error of model fit to a single point in titration curve,	(23)	
RMS-RE	Root–mean–square of Relative Error, i.e., square root of sum of squared RE for all points in a dataset divided by the number of points	(24)	
ε	Random error of model fit to measurement.	(18)	
Ψ	Data modeling approach: includes combination of mathematical method and speciation model used to analyze data.		

[a] Equivalent symbols used elsewhere.

ARTICLE IN PRESS

RMS-RE obtained from the deviation of fitted model from observed data – should be compared to the known measurement precision of one's experimental system in order to guide model selection, i.e., choosing the number of ligands 'detected' in the sample. Estimates of measurement precision can be developed from repeated measurements or averaging model error over multiple titration curves. Note that the magnitude of the error often depends on the signal measured.

9. Participants that performed a unified analysis of ACSV titration curves at multiple detection windows for a sample improved their results regardless of the basic mathematical approach taken. Overall, the top three most accurate sets of results were obtained using automatic unified analysis while the single most accurate set of results combined simultaneous calibration and parameter estimation. We therefore recommend that where sample volume and time permit, titration experiments be designed to include more than one detection window for all natural water samples, especially for coastal and estuarine waters. Note that even two windows can help substantially. It is vital that practical experimental designs for multi-window titrations be developed.
10. Everybody working in this field of research should verify their methods of data analysis, whether they use a simple spreadsheet model or custom software, by analyzing simulated data before applying it to data from real samples. We recommend using the simulated datasets presented here in Appendices 1 and 2. If data with little or no noise cannot be modeled with good precision and accuracy, it cannot be expected that real data obtained from real samples by ASV or ACSV will yield correct parameter estimates for metal binding ligands.
11. Although titrations using metals other than copper or samples from other types of aquatic ecosystems undoubtedly will require some differences in the design of titration studies and in some specifics of the data analysis process, most of the insights gained here are transferable to other ionic metals. For example, for samples from open ocean waters, which normally have lower ligand concentrations, lower amounts of metals would be added during the course of a titration, but the complications inherent in calibration discussed herein still need to be considered. Any differences in modeling approaches that are required for particular metals are worth documenting and reporting.

6. Definitions of symbols used

All symbols for variables used in the body of this paper are defined in Table 7 above.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.marchem.2015.03.006.

Acknowledgments

Logistical and partial financial support for Working Group 139, 'Organic Ligands — A Key Control on Trace Metal Biogeochemistry in the Ocean', was provided by the Scientific Committee on Oceanic Research (SCOR), from grant OCE-1243377 from the U.S. National Science Foundation, and from national SCOR committees. The following funding agencies supported this work through grants or PhD scholarships. IP and DO were supported by the Ministry of Science, Education and Sport of the Republic of Croatia, within the project 'Interactions of Trace Metals in Aquatic Environment'(098-0982934-2720). SGS and MW were supported by the New Zealand Ministry of Business, Innovation and Employment (CO1X1005) via a subcontract from the National Institute of Water and Atmosphere Research, and a University of Otago Research Grant. RMJH was supported by the USDA National Institute of Food and Agriculture, Hatch project 875-913. OB was supported by US National Science Foundation grant OCE-1315200 to FMM Morel. RMB was supported by US National Science Foundation grant OCE-1233733 to KAB. KNB was supported by institutional funding from the Ray Moore Endowment Fund and the Walwyn Hughes Fund for Innovation at the Bermuda Institute of Ocean Sciences. GC was supported by US National Science Foundation grants OCE-0926204 to EA Boyle, OCE-0136977 to JR Donat, and US-GEOTRACES and by Kuwait Foundation for the Advancement of Science grant P401900 to EA Boyle. CG was supported by ANR MARSECO and MERMEX-WP3-C3A projects. LJAG and MJAR were supported by the Dutch Program for Ocean and Coastal Research (ZKO) grant 839.08.410 of the Netherlands Organisation for Scientific Research (NWO, www.nwo.nl/en). MG was supported by the UK Natural Environment Research Council grant NE/E013546/1. LML was supported by the Campus de Excelencia Internacional Program, Spanish Ministerio de Educación, Cultura y Deportes. JN and BST were supported by US NSF grant 1061545. This manuscript is BIOS contribution number 2033.

References

Abualhaija, M.M., van den Berg, C.M.G., 2014. Chemical speciation of iron in seawater using catalytic CSV with ligand competition against salicylaldoxime. Mar. Chem. 164, 60–74.
Aho, K., Derryberry, D., Peterson, T., 2014. Model selection for ecologists: the worldviews of AIC and BIC. Ecology 95, 631–636.
Apte, S.C., Gardner, M.J., Ravenscroft, J.E., 1988. An evaluation of voltammetric titration procedures for the determination of trace metal complexation in natural waters by use of computer simulation. Anal. Chim. Acta 212, 1–21.
Baars, O., Croot, P.L., 2014. Dissolved cobalt speciation and reactivity in the eastern tropical North Atlantic. Mar. Chem. http://dx.doi.org/10.1016/j.marchem.2014.10.006.
Batley, G.E., Apte, S.C., Staube, J.L., 2004. Speciation and bioavailability of trace metals in water: progress since 1982. Aust. J. Chem. 57, 903–919.
Black, F.J., Bruland, K.W., Flegal, A.R., 2007. Competing ligand exchange-solid phase extraction method for the determination of the complexation of dissolved inorganic mercury (II) in natural waters. Anal. Chim. Acta. 598, 318–333.
Bruland, K.W., Donat, S.A., Moffett, J.W., Rue, E.L., Scrabal, S.A., 2000. Intercomparison of voltammetric techniques to determine the chemical speciation of dissolved copper in a coastal seawater sample. Anal. Chim. Acta 405, 99–113.
Buck, K.N., Moffett, J.W., Barbeau, K.A., Bundy, R.M., Kondo, Y., Wu, C., 2012a. The organic complexation of iron and copper: an intercomparison of competitive ligand exchange-adsorptive cathodic stripping voltammetry (CLE-ACSV) techniques. Limnol. Oceanogr. Methods 10, 496–515.
Buck, K.N., Lohan, M.C., Sander, S.G., SCOR Working Group 139, 2012b. Organic ligands — a key control on trace metal biogeochemistry. IUPAC Chem. Int. 34, p. 23
Buffle, J., Greter, F.L., Haerdi, W., 1977. Measurement of complexation properties of humic and fulvic acids in natural waters with lead and copper ion-selective electrodes. Anal. Chem. 49, 216–222.
Byrne, R.H., 2002. Inorganic speciation of dissolved elements in seawater: the influence of pH on concentration ratios. Geochem. Trans. 3, 11–16.
Cabaniss, S.E., Shuman, M.S., 1988. Copper binding by dissolved organic matter: I. Suwannee River fulvic acid equilibria. Geochim. Cosmochim. Acta 52, 185–193.
Campos, M.L.A.M., van den Berg, C.M.G., 1994. Determination of copper complexation in sea water by cathodic stripping voltammetry and ligand competition with salicylaldoxime. Anal. Chim. Acta 284, 481–496.
Croot, P.L., Johansson, M., 2000. Determination of iron speciation by cathodic stripping voltammetry in seawater using the competing ligand 2-(2-thiazolylazo)-p-cresol (TAC). Electroanalysis 12, 565–576.
Donat, J.R., Lao, K.A., Bruland, K.W., 1994. Speciation of dissolved copper and nickel in South San Francisco Bay: a multi-method approach. Anal. Chim. Acta 284, 547–571.
Ellwood, M.J., van den Berg, C.M.G., 2000. Organic complexation of zinc in the north eastern Atlantic. Mar. Chem. 68, 295–306.
Garnier, C., Mounier, S., Benaïm, J.Y., 2004a. Influence of dissolved organic carbon content on modelling natural organic matter acid–base properties. Water Res. 38, 3685–3692.
Garnier, C., Pižeta, I., Mounier, S., Benaïm, J.Y., Branica, M., 2004b. Influence of the type of titration and of data treatment methods on metal complexing parameters determination of single and multi-ligand systems measured by stripping voltammetry. Anal. Chim. Acta 505, 263–265.
Garnier, C., Pižeta, I., Mounier, S., Cuculić, V., Benaïm, J.Y., 2005. An analysis of distinguishing composite dissolved metal–ligand systems measurable by stripping voltammetry. Anal. Chim. Acta 538, 263–271.
Gerringa, L.J.A., Herman, P.M.J., Poortvliet, T.C.W., 1995. Comparison of the linear van den Berg/Ružić transformation and a non-linear fit of the Langmuir isotherm applied to Cu speciation data in the estuarine environment. Mar. Chem. 48, 131–142.
Gerringa, L.J.A., Rijkenberg, M.J.A., Thuróczy, C.-E., Maas, L.R.M., 2014. A critical look at the calculation of the binding characteristics and concentration of iron complexing ligands in seawater with suggested improvements. Environ. Chem. 11, 114–136.
Gledhill, M., van den Berg, C.M.G., 1994. Determination of complexation of iron(III) with natural organic complexing ligands in seawater using cathodic stripping voltammetry. Mar. Chem. 47, 41–54.
Haldane, J.B.S., 1957. Graphical methods in enzyme chemistry. Nature 179, 832.
Hawkes, J.A., Gledhill, M., Connelly, D.P., Achterberg, E.P., 2013. Characterisation of iron binding ligands in seawater by reverse titration. Anal. Chim. Acta 766, 53–60.

ARTICLE IN PRESS

Hirose, K., Dokiya, Y., Sugimura, Y., 1982. Determination of conditional stability constants of organic copper and zinc complexes dissolved in seawater using ligand exchange method with EDTA. Mar. Chem. 11, 343–354.
Hudson, R.J.M., Rue, E.L., Bruland, K.W., 2003. Modeling complexometric titrations of natural water samples. Environ. Sci. Technol. 37, 1553–1562.
Kogut, M.B., Voelker, B.M., 2001. Strong copper-binding behavior of terrestrial humic substances in seawater. Environ. Sci. Technol. 35, 1149–1156.
Kogut, M.B., Voelker, B.M., 2003. Kinetically inert Cu in coastal waters. Environ. Sci. Technol. 37, 509–518.
Kozelka, P.B., Bruland, K.W., 1998. Chemical speciation of dissolved Cu, Zn, Cd, Pb in Narragansett Bay, Rhode Island. Mar. Chem. 60, 267–282.
Laglera, L.M., van den Berg, C.M.G., 2003. Copper complexation by thiol compounds in estuarine waters. Mar. Chem. 82, 71–89.
Laglera, L.M., Downes, J., Santos-Echeandia, J., 2013. Comparison and combined use of linear and non-linear fitting for the estimation of complexing parameters from metal titrations of estuarine samples by CLE/AdCSV. Mar. Chem. 155, 102–112.
Laglera-Baquer, L.M., González-Dávila, M., Santana-Casiano, J.M., 2001. Determination of metallic complexing capacities of the dissolved organic material in seawater. Sci. Mar. 65, 33–40 (SUPPLEMENT).
Lewis, J.A.M., Sunda, W.G., 1978. Effect of Complexation by natural organic ligands on the toxicity of copper to a unicellular alga, *Monochrysis lutheri*. Limnol. Oceanogr. 23, 870–876.
Lohan, M.C., Crawford, D.W., Purdie, D.A., Statham, P.J., 2005. Iron and zinc enrichments in the northeastern subarctic Pacific: ligand production and zinc availability in response to phytoplankton growth. Limnol. Oceanogr. 50, 1427–1437.
Louis, Y., Garnier, C., Lenoble, V., Omanović, D., Mounier, S., Pižeta, I., 2009. Characterisation and modelling of marine dissolved organic matter interactions with major and trace cations. Mar. Environ. Res. 67, 100–107.
Mantoura, R.F.C., Riley, J.P., 1975. Analytical concentration of humic substances from natural-waters. Anal. Chim. Acta 76, 97–106.
McKnight, D.M., Feder, G.L., Thurman, E.M., Wershaw, R.L., Westall, J.C., 1983. Complexation of copper by aquatic humic substances from different environments. Sci. Total Environ. 28, 65–76.
Miller, J.M., 1993. Outliers in experimental data and their treatment. Analyst 118, 455–461.
Miller, L.A., Bruland, K.W., 1997. Competitive equilibration techniques for determining transition metal speciation in natural waters: evaluation using model data. Anal. Chim. Acta 343, 161–181.
Millero, F.J., Schreiber, D.R., 1982. Use of the ion pairing model to estimate activity coefficients of the ionic components of natural waters. Am. J. Sci. 282, 1508–1540.
Moffett, J.W., 1995. Temporal and spatial variability of Cu complexation by strong chelators in the Sargasso Sea. Deep-Sea Res. I 42, 1273–1295.
Moffett, J.W., Brand, L.E., Croot, P.L., Barbeau, K.A., 1997. Cu speciation and cyanobacterial distribution in harbors subject to anthropogenic Cu inputs. Limnol. Oceanogr. 42, 789–799.
Nuester, J., van den Berg, C.M.G., 2005. Determination of metal speciation by reverse titrations. Anal. Chem. 77, 11–19.
Omanović, D., Garnier, C., Louis, Y., Lenoble, V., Mounier, S., Pižeta, I., 2010. Significance of data treatment and experimental setup on the determination of copper complexing parameters by anodic stripping voltammetry. Anal. Chim. Acta 664, 136–143.
Omanović, D., Garnier, C., Pižeta, I., 2014. ProMCC: an all-in-one tool for trace metal complexation studies. Mar. Chem. http://dx.doi.org/10.1016/j.marchem.2014.10.011 (in this issue).
Pižeta, I., Branica, M., 1997. Simulation and fitting of anodic stripping voltammetric data for determination of the metal complexing capacity. Anal. Chim. Acta 351, 73–82.
Rue, E.L., Bruland, K.W., 1995. Complexation of iron(III) by natural organic ligands in the central North Pacific as determined by a new competitive ligand equilibration adsorptive cathodic stripping voltammetric method. Mar. Chem. 50, 117–138.
Ružić, I., 1982. Theoretical aspects of the direct titration of natural waters and its information yield for trace metal speciation. Anal. Chim. Acta 140, 99–113.
Saito, A.M., Moffett, J.W., DiTullio, G.R., 2004. Cobalt and nickel in the Peru upwelling region: a major flux of labile cobalt utilized as a micronutrient. Glob. Biogeochem. Cycles 18, GB4030.
Sander, S.G., Hunter, K.A., Harms, H., Wells, M., 2011. Numerical approach to speciation and estimation of parameters used in modeling trace metal bioavailability. Environ. Sci. Technol. 45, 6388–6395.
Sander, S.G., Buck, K.N., Lohan, M., 2012. Improving understanding of organic metal-binding ligands in the ocean. Eos 93, 244.
Scatchard, G., 1949. The attractions of proteins for small molecules and ions. Ann. N. Y. Acad. Sci. 51, 660–672.
Shuman, M.S., Cromer, J.L., 1979. Copper association with aquatic fulvic and humic acids. Estimation of conditional formation constants with a titrimetric anodic stripping voltammetry procedure. Environ. Sci. Technol. 13, 543–545.
Shuman, M.S., Woodward Jr., G.P., 1973. Chemical constants of metal complexes from a complexometric titration followed with anodic stripping voltammetry. Anal. Chem. 45, 2032–2035.
Sposito, G., 1982. On the use of the Langmuir equation in the interpretation of 'adsorption' phenomena: II. The 'two-surface' Langmuir equation. Soil Sci. Soc. Am. J. 46, 1147–1152.
Turner, D.R., Whitfield, M., Dickson, A.G., 1981. The equilibrium speciation of dissolved components in freshwater and seawater at 25 °C and 1 atm pressure. Geochim. Cosmochim. Acta 45, 855–881.
Turoczy, N.J., Sherwood, J.E., 1997. Modification of the van den Berg/Ružić method for the investigation of complexation parameters of natural waters. Anal. Chim. Acta 354, 15–21.
van den Berg, C.M.G., 1982a. Determination of copper complexation with natural organic ligands in seawater by equilibration with MnO_2. Part 1: Theory. Mar. Chem. 11, 307–322.
van den Berg, C.M.G., 1982b. Determination of copper complexation with natural organic ligands in seawater by equilibration with MnO_2. Part 2: experimental procedures and application to surface seawater. Mar. Chem. 11, 323–342.
van den Berg, C.M.G., 1984. Determination of the complexing capacity and conditional stability constants of complexes of copper (II) with natural organic ligands in seawater by cathodic stripping voltammetry of copper-catechol complex ions. Mar. Chem. 15, 1–18.
van den Berg, C.M.G., 1995. Evidence for organic complexation of iron in seawater. Mar. Chem. 50, 139–157.
van den Berg, C.M.G., 2006. Chemical speciation of iron in seawater by cathodic stripping voltammetry with dihydroxynaphthalene. Anal. Chem. 78, 156–163.
van den Berg, C.M.G., Donat, J.R., 1992. Determination and data evaluation of copper complexation by organic ligands in sea water using cathodic stripping voltammetry at varying detection windows. Anal. Chim. Acta 257, 281–291.
van den Berg, C.M.G., Kramer, J.R., 1979. Determination of complexing capacities and conditional stability constants for copper in natural waters using MnO_2. Anal. Chim. Acta 106, 113–120.
van den Berg, C.M.G., Nimmo, M., 1987. Determinations of interactions of nickel with dissolved organic material in seawater using cathodic stripping voltammetry. Sci. Total Environ. 60, 185–195.
Voelker, B.M., Kogut, M.B., 2001. Interpretation of metal speciation data in coastal waters: the effects of humic substances on copper binding as a test case. Mar. Chem. 74, 303–318.
Wells, M., Buck, K.N., Sander, S.G., 2013. New approach to analysis of voltammetric ligand titration data improves understanding of metal speciation in natural waters. Limnol. Oceanogr. Methods 11, 450–465.
Westall, J.C., 1982. FITEQL: a computer program for determination of chemical equilibrium constants from experimental data. Department of Chemistry, Oregon State University (101 pp.).
Wilkinson, G.N., 1961. Statistical estimations in enzyme kinetics. Biochem. J. 80, 324–332.
Wu, J., Jin, M., 2009. Competitive ligand exchange voltammetric determination of iron organic complexation in seawater: examination of accuracy using computer simulation and elimination of artifacts in ideal two-ligand case using iterative non-linear multiple regression. Mar. Chem. 114, 1–10.

Web references (accessed November 28, 2014)

Gustafsson, J.P., 2014. Software: visual MINTEQ version 3.0, 2011. http://www2.lwr.kth.se/English/OurSoftware/vminteq/.
Hudson, R.J.M., 2014. Software: KINETEQL Multiwindow Solver (KMS), Version 1.0. https://sites.google.com/site/kineteql/.
Omannović, D., 2014. ProMCC: an all-in-one tool for trace metal complexation studies. https://sites.google.com/site/mccprosece/.
SCOR Working Group 139, 2014. Simulated complexometric titration data. http://neon.otago.ac.nz/research/scor/achievements.html.
van den Berg, C.M.G., 2014. Excel spreadsheet: vdB complexing capacity calculation for Cu. http://www.liv.ac.uk/~sn35/Documents/Useful_links.html.

Please cite this article as: Pižeta, I., et al., Interpretation of complexometric titration data: An intercomparison of methods for estimating models of trace metal complexation by ..., Mar. Chem. (2015), http://dx.doi.org/10.1016/j.marchem.2015.03.006