



A semi-empirical equation modeling long-range contributions to the activity coefficients of individual ions at high ionic strengths



Kent M. Rapp, Michael J. Betenbaugh*, Marc D. Donohue

Department of Chemical and Biomolecular Engineering, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218, USA

ABSTRACT

Molecular thermodynamics is well developed with models like UNIQUAC, UNIFAC and SAFT. These models include robust codes and extensive databases for chemical and petroleum industries through ASPEN, DORTMUND, and others. However, important systems in the pharmaceutical and biotech industries now exist where these codes and their databases are not always sufficient. Cell culture media used to grow mammalian cells are multicomponent aqueous solutions with 50, 100, or more compounds including amino acids, salts, acids or bases, as well as sugars, fatty acids, etc. These components can exhibit both complexation and incomplete dissociation. The thermodynamics of these systems have proven challenging to model, as current solubility descriptions frequently assume complete dissociation in solutions containing a single solute in water at relatively low concentrations. However, not all compounds that are generally considered to be strong electrolytes are completely dissociated even at low concentrations. Further, mixtures of such components can show dramatic changes in dissociation and complexation with concentration, pH, and temperature. In this paper, we present a semi-empirical model for individual ions in aqueous solutions. We use data from 317 compounds to isolate the ionic contribution to a compound's activity coefficient in aqueous systems from confounding short-range effects. We compare eight robust regression M-estimators with a least-squares estimate and provide a two-parameter equation relating ionic strength, valence, and the activity coefficient not requiring arduous characterization of the solute of interest. We demonstrate that the Fair M-estimator generates an accurate model that can serve as an appropriate reference state for use with a model based on perturbation theory up to 29 molal.

1. Introduction

Solutions featuring multiple components, potentially with incomplete dissociation and at high concentrations, are frequently encountered in biological systems. For example, blood is a precisely buffered solution containing a rich mixture of organic molecules and inorganic ions some of which can be at high concentrations. Electrolytic solutions used in electrochemistry often require high salt concentrations for conductivity. Other natural systems such as oceans are a concentrated salt solution with many different ions, and their capacity to capture and store carbon dioxide via the formation of a weak acid is a key element in maintaining global CO₂ levels. One additional important current area is in the development of intensified media formulations for mammalian-cell biomanufacturing. Mammalian cell culture media often contains 50 to 100 different chemical compounds, including amino acids, sugars, fatty acids, and salts, and their development relies heavily on experimental optimization. Feeds, which are supplemented to the bioreactors during extended fed-batch or perfusion processes, may contain even higher concentrations of select media components. Because of the large space of possible media formulations, it is nearly impossible to efficiently find an optimal media formulation via experimentation alone.

The ability to accurately predict a compound's solubility in aqueous solution without having to rely solely on experiments could vastly accelerate media development. However, there is currently a dearth of both experimental data and computational tools that can model these complex environments, despite a compelling need for accurate predictions of their behavior.

Creating media formulations in which components remain in solution requires careful determination of the amounts of specific components that can dissolve in the media. This is especially challenging in intensified biomanufacturing processes with high-density cultures in which high concentrations of nutrients must be dissolved in the culture media. Part of this formulation process requires determining the solubilities of specific amino acids in solution. To conventionally accomplish this, laborious, time-consuming experiments have been needed to determine the activity coefficients and solubilities even for simple systems like a single amino acid in aqueous solution. Taking data for two or three amino acids in solution can be an even more daunting task.

An alternative strategy would be to computationally predict the solubilities of these systems using thermodynamic approaches. Being able to make *in silico* predictions of such behavior will aid in the development of new cell culture media formulations and will help

* Corresponding author.

E-mail address: beten@jhu.edu (M.J. Betenbaugh).

advance the development of biomanufacturing, especially for intensified (i.e., high cell concentration) processes.

However, since many media components, including amino acids and buffers, can either ionize partially or fully in solution, determining which compounds remain soluble and which precipitate is especially challenging. A number of predictive models already exist, but each has their limitations. Debye-Hückel [1] requires low ionic strengths, complete dissociation, and point charges, all of which are unsuitable for the systems discussed here. Guggenheim's work [2] as well as that of Glueckauf [3] present an alternative approach for modeling multicomponent electrolyte systems, but only capture interactions caused by ions of opposite charges. Similarly, the original Pitzer-Debye-Hückel formula [4] was developed for use only in binary systems. Its subsequent use to model the long-range portion of the activity coefficient in some electrolyte models, such as the multicomponent electrolyte non-random two-liquid model [5], overcame this limitation, but still required salt-specific fits that limit its use to specific sets of compounds. Indeed, this shortcoming was noted and recently addressed in an ion-based electrolyte non-random two liquid model [6]; however, the model has mixed results when predicting salt solubilities in multi-electrolyte systems. Other models using the Pitzer-Debye-Hückel equation can model weak electrolytes by incorporating hydration and equilibrium constants, but only model single electrolyte systems without complexation [7]. A fourth model, developed by Kusik, Meissner, and Field [8,9], is capable of predicting solubilities in multi-ion systems; nevertheless, it only is applicable to systems containing exclusively strong electrolytes, as it does not model the activity of individual ions for solving equilibrium equations. Finally, updated electrolyte versions of LIQUAC and LIFAC are only for strong electrolytes [10].

Predicting the solubilities of these compounds in aqueous solution requires a detailed knowledge of three main factors: ionization, complexation, and dissociation. Accurately describing the effect of each of these factors would provide improved solubility predictions for individual compounds. However, due to the complexity of the interactions in these systems, experimental validation still will be required. In other words, rigorous theoretical characterization is often tangential to the ultimate aim of developing concentrated media. Hence, a more general, coarse-grained estimate is a reasonable approach that often can supply sufficient information for initial media component concentrations that can subsequently be refined and tested through media development experiments.

Ionization is relatively straightforward to calculate if the chemical structure of a component is known: one must simply determine the valence of the atom or molecule in question. For example, the ionization of the CO_3^{2-} is simply 2. This feature can be used to determine the component's complexation with other components in solution, often by cross-referencing against a database of known compounds and their constituent ions. The extent to which ions (such as Na^+ and CO_3^{2-}) will form a complex (Na_2CO_3) in solution depends largely on the complex's dissociation constant. In contrast to ionization and complexation, which are discrete variables, the dissociation constant is continuous and can range from completely dissociated and ionized (only Na^+ and CO_3^{2-} in solution) to completely associated and complexed (only Na_2CO_3 in solution).

The dissociation constants of some media components, such as amino acids, have been characterized [11]. Nevertheless, solubility predictions using dissociation constants that rely on concentrations alone have low accuracy because they assume ideal mixtures; incorporating an activity model into dissociation calculations can address this shortcoming by accounting for non-ideality. For simple systems involving only a single solute, such as the strong electrolyte hydrochloric acid in water, the solute's activity can be modeled without differentiating between conjugate forms (HCl/Cl^- and $\text{H}_3\text{O}^+/\text{H}_2\text{O}$, as per Brønsted-Lowery acid-base theory). In contrast, multicomponent systems involving solute-solute interactions, such as a mixture of weak electrolytes like amino acids in water, necessitate a means to separately model the activity of conjugate

species, many of which are ionized.

Multiple factors affect a component's activity, including its charge, shape, and size. Modern activity models increasingly differentiate between these interaction effects by using complementary mathematical descriptions. For dissociating ionized species, such as amino acids, this can result in superior predictive capabilities. Contemporary media development relies heavily on experimental studies to develop highly concentrated, multicomponent products for their processes. Current computational approaches frequently do not take an individual ion approach to modeling, which is necessary for systems in which cross-species complexations and partial dissociation can occur. To this end, we have developed a new approach to modeling long-range electrostatic interactions that demonstrates improved performance over existing alternatives at high ionic strengths and does not require ion-specific parameters. By developing an activity model that uses individual ions as the modeling unit instead of complexes, our model serves as a spring-board for solubility predictions in a broader array of systems that include multiple components and partially dissociated species.

2. Materials and methods

2.1. Curve fitting

Curve fitting was done using the "fitnlm" function in MATLAB R2021b. Ten thousand linearly spaced points were generated between zero and the maximum molality of the 1:1 ions (29 molal). These points were converted into ionic strengths, and data for all compounds was generated such that they would have estimates at the same ionic strengths. Thus, a compound with a maximum concentration of 1 molal would have half as many points as a compound with a maximum concentration of 2 molal. Specifically, despite different maximum concentrations, each compound would have calculated data points at the same ionic strengths for the regions they overlap (e.g., at 0, 0.0029, 0.0058, ..., 0.29, ..., 1 molal). To facilitate fitting, the data was scaled and centered by using the natural logarithm of the activity coefficient, $\ln(\gamma_{\pm})$.

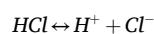
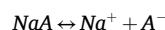
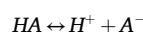
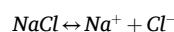
3. Theory/calculation

3.1. Multicomponent systems

The behavior of strong electrolytes in solution is among the most widely studied systems in thermodynamics due to its well-defined physical attributes. In complex media, however, the components are less well-behaved, exhibiting only partial dissociation and complexation with other media components. Here, the necessity of individual ion activity models becomes apparent. The dissociation constant K for a compound is defined as:

$$K = \prod_j a_j^{\nu_j} = \prod_j \frac{m_j \gamma_j}{m^\ominus} \quad (1)$$

where a_j is the activity of a component j in solution, ν_j is the stoichiometric number of the component, m_j is the component's molality, m^\ominus is the standard state molality, and γ_j is the component's activity coefficient [12]. For a dissociation reaction with complexation, such as that which occurs when sodium chloride and an amino acid both are present in solution, the system can be described by:



Each of these dissociation reactions can be modeled using their corresponding dissociation constants and activities. Despite the necessity to model these systems for complex solutions, the activities of individual ions in solution are not well-described in the literature.

3.2. Prior work

Some previous efforts have been successful in predicting complexation with reasonable accuracy for multicomponent systems using individual ion activities. Of note are MINTEQA2 [13] (which forms the basis of Visual MINTEQ [14]) and WATEQ [15]. Both of these approaches, however, have limitations.

MINTEQA2 uses either a modified Debye-Hückel equation [1] or the Davies equation [16] to estimate the activity coefficient. The developers do not recommend use in solutions with an ionic strength above 0.5 molal. This limits the usefulness of MINTEQA2 to dilute systems, precluding its application to broader systems like seawater and complex culture media.

WATEQ relies on the MacInnes hypotheses [17]. In short, the hypotheses state: (1) that a given ion's activity is independent of the ions associated with it and is therefore dependent only on the ionic strength and temperature of the solution and (2) that the K^+ cation and Cl^- anion have the same activity coefficients. While these assumptions have allowed the development of useful software for solubility predictions of complex solutions, there are specific issues undermining their functionality.

The first hypothesis tacitly requires that the ions are completely dissociated, which is doubtful even for many systems that are considered strong electrolytes [5,18] and false for many ionic species that are not strong electrolytes. Indeed, models have been able to explain the non-ideality of strong electrolytes as a function of incomplete dissociation and solvation [18,19]. Moreover, experimental studies have suggested that strong electrolytes do form ion pairs [20], with an estimated 60 % of KCl ions pairing near saturation [21]; similarly, surface force measurements of concentrated electrolyte solutions have shown the formation of ion layers [22]. Comparisons with molecular dynamics simulations have yielded similar results [23–29]. For other ionizable species, such as carbonic acid, dissociation is incomplete, invalidating the use of the MacInnes assumptions for accurate estimates of the activities for such species.

The second hypothesis was made based on the fact that the ions K^+ and Cl^- have similar weights and molar conductivities. However, this hypothesis is also questionable. While these two ions may have some similar characteristics, they differ in many more, such as ionic radius [30], hydration number [31], and energies of formation [10]. Additionally, it is impossible to rely on this assumption above the maximum solubility of KCl (approximately 5 molal [32]) despite many salts having significantly higher solubilities, limiting its explanatory power.

In summary, it is clear that a more complete model is needed to capture the behavior of ionic species in multicomponent systems containing incompletely dissociating compounds at moderate to high concentrations. Existing approaches either are only descriptive models of salts with existing data, do not account for multiple components that form new molecular compounds in solution, or are unsuitable at high ionic strengths. Here we develop a model that requires only the charge of the ion and the solution's ionic strength, thereby eliminating the need for time-consuming parameterization necessitated by other models. This will facilitate future modeling efforts by allowing for the evaluation of equilibrium equations using activity coefficients for each ion and molecule involved in the reaction across several interaction reactions in solution (see Section 4.1). Though this model is general and not completely explanatory for the behavior of any specific ion in solution, it captures typical behavior for ions with a valence up to 5 and, more importantly, gives reasonable predictions up to approximately 30 molal. This latter point is particularly relevant because all other current

species-agnostic activity coefficient models, such as the Davies equation, give dramatically unrealistic activities at high ionic strengths.

3.3. Individual ion activity model

Activity coefficients capture a range of non-ideal behaviors spanning ionic, solute–solute, and solute–solvent interactions. Distinct models describing these interactions are increasingly being developed and combined through models characterizing factors such as hydration [33], partial dissociation [18], group interactions [34], screening length [35], and the dielectric constant [36]. These factors operate across different length scales. Thus, to model these separate effects, the overall activity often is written as:

$$\gamma_{\pm} = \gamma_{LR}\gamma_{SR} \quad (2)$$

where γ_{LR} considers long-range ionic contributions to the activity coefficient and γ_{SR} accounts for short-range contributions to the activity coefficient. This is similar to the approach taken by LIQUAC [37], LIFAC [38], and updated versions thereof [10]. LIQUAC and LIFAC have demonstrated prowess in property predictions involving electrolytes in mixed solvent solutions [10]; still, they are not built to account for species where partial dissociation and complexation are major factors influencing solution properties, necessitating an alternative approach incorporating equilibrium constants as described by Equation (1).

In contrast to LIQUAC and LIFAC models, we omit a separate mid-range contribution term and assume that interactions at this length scale are sufficiently captured by the long-range and short-range activity coefficient terms. This reduces the overhead necessary to describe a given system, as the mid-range term requires interaction parameters for every species pair in solution; future work could explore the benefits of including a mid-range term.

In our approach, we use a long-range activity coefficient model that is species-agnostic, as interactions happen at length scales significantly larger than the size of the species in solution. In contrast, short-range models are species-specific, and interactions happen at length scales at or below the size of the species in solution. Still, short-range models often can estimate interaction effects for novel components by predicting them from existing group interactions, minimizing the necessity for experimental measurements. A number of different approaches can be used to model the short-range contributions, such as UNIFAC [34] or PC-SAFT [39]. The focus of the model developed in this paper, however, is on estimating the long-range, electrostatic contributions for charged components, and we will restrict our discussion accordingly. This is similar to the semi-empirical approach taken in Samson, Lemaire, Marchand, and Beaudoin [40], which has been successfully used in modeling complex cement and concrete systems [41–43].

We should also note that the approach of factoring the partition function or activity coefficient used here (and in nearly all modern activity models) traces back to the concept of perturbation theory. It has been shown by Nezbeda [44,45] and Aranovich, Donohue, and Donohue [46] that the appropriate choice of reference state is necessary for any factoring of the partition function or activity coefficient for any perturbation theory to work. In this case, because of the strength of electrostatic forces present in solutions containing ionized compounds, long-range interactions are the only appropriate reference state.

While this model can provide a suitable reference state at high ionic strengths, any model only accounting for long-range interactions should be used with caution as ionic strengths increase. At high ionic strengths, short-range interactions become increasingly important to a compound's activity; a long-range model alone may be insufficient in this scenario. Incorporating a short-range model can help capture the variation in behavior between different salts, given a reasonably accurate estimate of the activity coefficient. The work presented here provides such an estimate, which will increase the predictive capability of future models built to capture compound-specific activity coefficients.

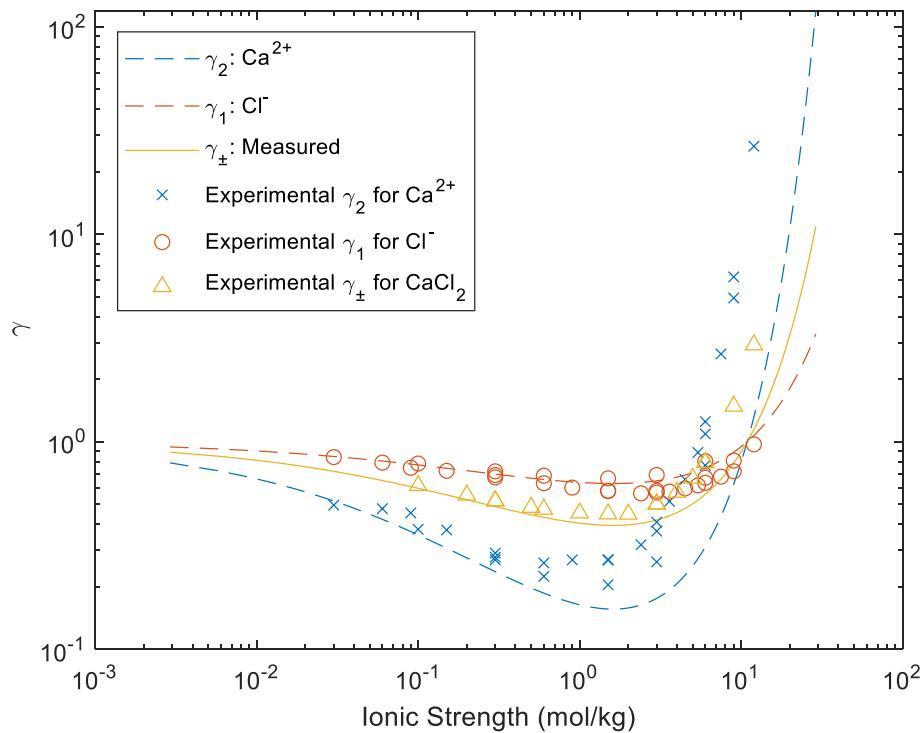


Fig. 1. Relationship between individual ion ionic activity coefficient values (dashed lines) and experimentally measured ionic activity coefficient values (solid line) for the bi-univalent compound CaCl_2 . The individual ions' activity coefficient cannot easily be experimentally determined, and what is often measured and reported instead is the mean activity coefficient for a given compound. Experimental data sources for the individual ions and mean activity coefficient of CaCl_2 are listed in Table S3.

In the special case of components that completely or almost completely dissociate, we assume $\gamma_{SR} \approx 1$, such that the mean activity coefficient for a compound M_pN_q is only a function of γ_{LR} and can be defined in terms of its constitutive ions as

$$\gamma_{\pm} = (\gamma_+^p \gamma_-^q)^{\frac{1}{p+q}} \quad (3)$$

where γ_+ is the activity coefficient of the cation and γ_- is the activity coefficient of the anion [47]. It is important to note that γ_{\pm} is the measurable value of the activity coefficient for an ionic compound, representing the geometric mean of the individual activity coefficients of all ions in solution; it is not the individual activity coefficient of the ions themselves. Much discussion has surrounded whether it is possible to measure individual ion activity coefficients and if they meaningfully exist at all [48,49]. Here we surmise that, regardless of their existence, the use of an activity coefficient model for individual ions is nevertheless useful for the prediction of complex systems.

Assuming that the long-range interactions are a function solely of the ionic strength I of the solution, the valence z of the ions j and k , and temperature T , for a salt with two univalent ions we can write

$$\gamma_{\pm} \equiv \gamma_{\pm}(I, z_j, z_k, T) \quad (4)$$

whereby

$$\gamma_{\pm} = \gamma_+ = \gamma_- = \gamma_1 \quad (5)$$

Because the ionic strength, valence, and temperature are identical for the univalent cation and univalent anion, there is no parameter to distinguish between them and the activity coefficient γ_1 can be used to more generally represent the activity of an ion with a valence of one. The variation observed in experiments among ions of the same charge is attributed to short-range interactions.

Given γ_{\pm} for a compound containing a univalent anion and γ_1 , it becomes trivial to determine the activity coefficient for other valences

via simple rearrangement of Equation (3):

$$\gamma_z = \left(\frac{\gamma_{\pm}^{p+q}}{\gamma_1^q} \right)^{\frac{1}{p}} \quad (6)$$

or, in the case where the cation has a valence of one,

$$\gamma_z = \left(\frac{\gamma_{\pm}^{p+q}}{\gamma_1^p} \right)^{\frac{1}{q}} \quad (7)$$

where z represents the counterion's valence in a compound containing a univalent ion. This process can be replicated for other valences and ions, given that sufficient data for them exists. An empirical equation then can be fit to measured γ_{\pm} data using these relationships to develop a simple function that only depends on ionic strength, charge, and temperature. Assuming that at least one of the ions present in the compound is univalent, we can use z to simplify Equation (3) into the form

$$\ln(\gamma_{\pm}) = \left(\frac{1}{1+z} \right) (\ln(\gamma_z) + z \ln(\gamma_1)) \quad (8)$$

This equation allows for simultaneous fitting of any number of compounds, so long as at least one constituent ion is univalent. An idealized visual representation of the relationship between γ_{\pm} and long-range ionic effects is presented in Fig. 1 for calcium chloride, a bi-univalent salt, assuming no short-range interactions.

The semi-empirical activity model developed here takes a similar approach to other models [16,50–52] by adding empirical fitting terms to the Debye-Hückel activity model and consists of three main components, which can be decomposed into independent functions describing the impact of ionic strength $f(I)$, the impact of valence $g(z)$, and the impact of temperature $h(T)$. The model used here is

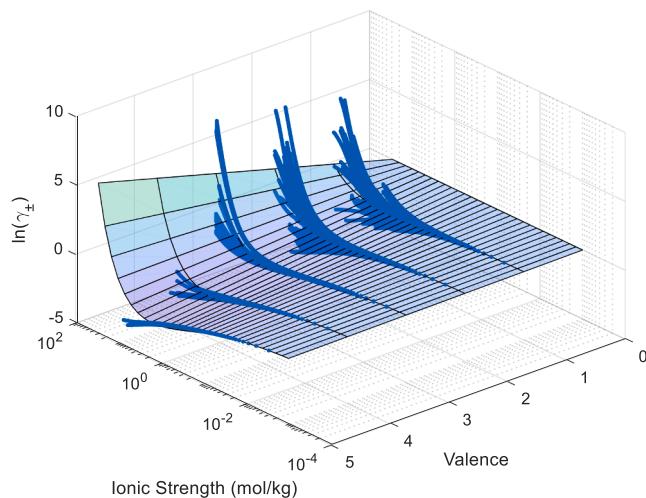


Fig. 2. Fitted model prediction (surface) and experimental data (blue points).

$$\ln(\gamma_z) = f(I)g(z)h(T) = \left(\frac{-A\sqrt{I}}{1 + c_1\sqrt{I}} + c_2I \right) z^2 \left(\frac{298.15}{T} \right)^{\frac{3}{2}} \quad (9)$$

where $A = 1.172 \left(\frac{\text{kg}}{\text{mol}} \right)^{\frac{1}{2}}$ is the temperature-dependent Debye-Hückel constant of water at 25 °C, and c_1 and c_2 are empirically determined fitting constants. Note that, as the Debye-Hückel constant used here is for water, this model applies only to aqueous systems. With a different constant, it could likely be used for additional solvents; with additional modeling and testing, it could potentially be used for mixed solvent systems, but that is beyond the scope of the work presented here. In this paper, we deal exclusively with aqueous solutions containing multiple solutes. Additionally, as is conventional for electrolyte solutions, this

model uses asymmetric activity coefficients, such that $\lim_{I \rightarrow 0} \gamma_z = 1$.

After the individual ion activity coefficients are determined using Equation (9), Equation (3) can be used to determine the mean activity coefficient. Given a sufficiently large dataset, this relationship can be used to estimate the long-range component of individual ion activities.

It is worth noting that choosing an adequate empirical model is subjective. In general, $f(I)$ resembles the Debye-Hückel equation [1] without the linear term c_2I included here. The Davies equation does have such a linear term included but is missing the fitting constant c_1 in the denominator. We find that two fitting constants provide a better description of the behavior at low and high ionic strengths by minimizing the root mean square error (RMSE) of the data. For completeness, the authors have examined several forms of $f(I)$ and provided statistical metrics (Table S2), but a fitting constant in the denominator combined with a single linear term seems to best balance functionality with simplicity. The functions $g(z)$ and $h(T)$ are expected to remain as-is: a more complex $g(z)$ or $h(T)$ necessitates more fitting constants, impacting the Debye-Hückel constant and resulting in problematic behavior at low ionic strengths. Additionally, $h(T)$ comes from the temperature component of the Debye-Hückel constant, as it is often desirable to model these systems at a variety of temperatures. For biological media, the system of focus for this work, these temperatures typically range from 4 °C to 37 °C.

4. Results and discussion

In order to generate an aggregate description of the activity coefficient for ions with a valence of unity, we must fit a curve to empirical data. Fortunately, a large collection of data for compounds containing a univalent ion has already been compiled and fit by various sources [50–62]. In total, we have compiled activity data for 317 compounds containing a univalent ion and used this to develop an accurate model to estimate effects stemming from long-range interactions (Fig. 2). Much reported data is not raw values but fitted trends. Thus, we opted to use

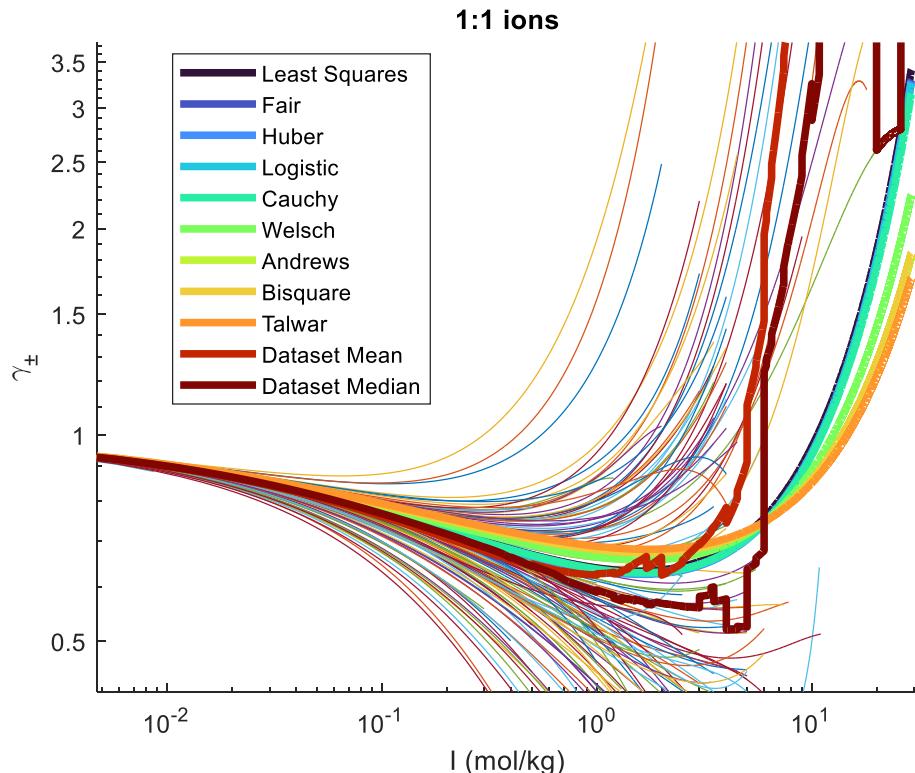


Fig. 3. Comparison of best fit M-estimator results to the dataset mean and median. Thin lines represent expected activity values for each of 140 different 1:1 salts in aqueous solution, while the wider lines represent either the results of fitting a given M-estimator, the dataset mean, or the dataset median.

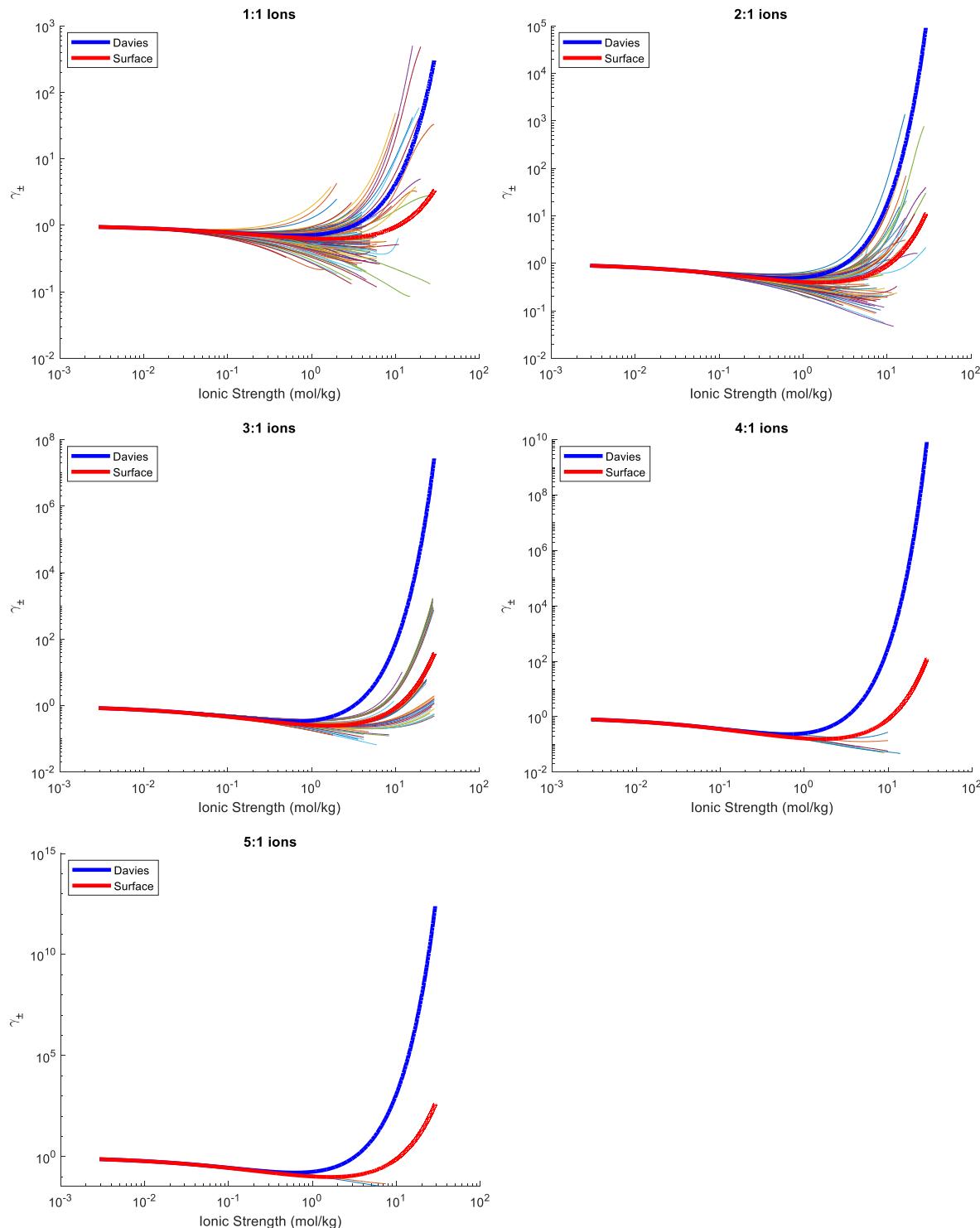


Fig. 4. Comparison of Davies equation predictions with the proposed model. Thin lines represent the mean activity coefficient, γ_{\pm} , of a single electrolyte in water.

these fits for all compounds and accepted them at face value. Additionally, the compiled data exclusively represents the mean activity coefficient, γ_{\pm} , for compounds that have at least one univalent ion. Because of this, we can substitute the semi-empirical Equation (9) into Equation (8) to generate an equation that will fit this dataset. Because the equation is being fit to a large body of electrolyte activity coefficient data rather than that of a specific salt, the resulting fitting constants will represent a typical trend for any given salt, which ultimately reflects the underlying long-range contributions to each salt's activity coefficient. If this approach were to be used for an individual electrolyte, it would be

impossible to disentangle long-range and short-range interactions.

A few features are noteworthy. First, for an aqueous component in solution with no charge, the long-range electrostatic component for the activity coefficient is unity, as expected. Thus, any variation in solubility between two solubilized molecules with a valence of zero would be accounted for solely by a short-range activity coefficient model. Second, there is significant variation in the activity coefficient of different counterions as concentrations begin to exceed 0.5 M (Figure S1); we attribute this variation to short-range interactions which could be accounted for better through the inclusion of a short-range activity

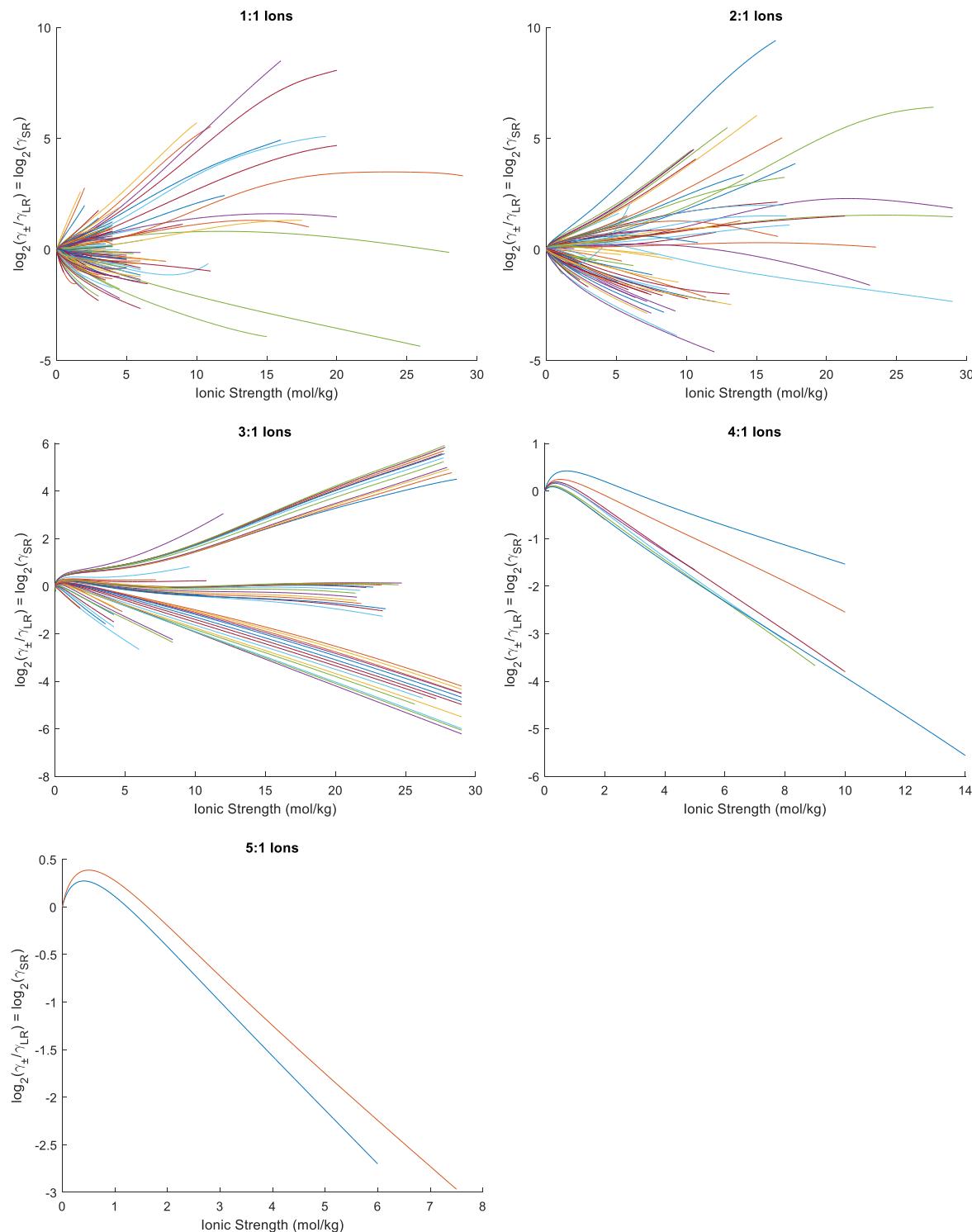


Fig. 5. Plots of the expected γ_{SR} values that could be fit using a variety of different models. Each line represents the mean activity coefficient of a single electrolyte in water.

coefficient model (Equation (2)). Finally, as the ionic strength approaches infinite dilution, the activity coefficient approaches unity, which is the expected result from Debye-Hückel theory.

The data used comes from several different sources for a number of salts, and it seems unlikely that all the data is of comparable quality. Additionally, because short-range effects are increasingly prominent at higher ionic strengths due to the closer proximity of ions to each other, the unexplained variance of the datasets is expected to increase at higher ionic strengths. Altogether, this poses two challenges to accurately

estimating the contribution of long-range, ionic effects to the activity coefficient: the potential presence of outliers and spurious fits plus heteroskedasticity.

To mitigate these effects, we used robust regression approaches and compared several different M-estimators, which have frequently been employed to achieve more accurate estimates in the face of outliers and violations of normality assumptions [63]. Fit statistics for these various methods are in Table S1. Of these, we have opted to use the Fair M-Estimator due to it resulting in the highest adjusted R^2 , lowest RMSE,

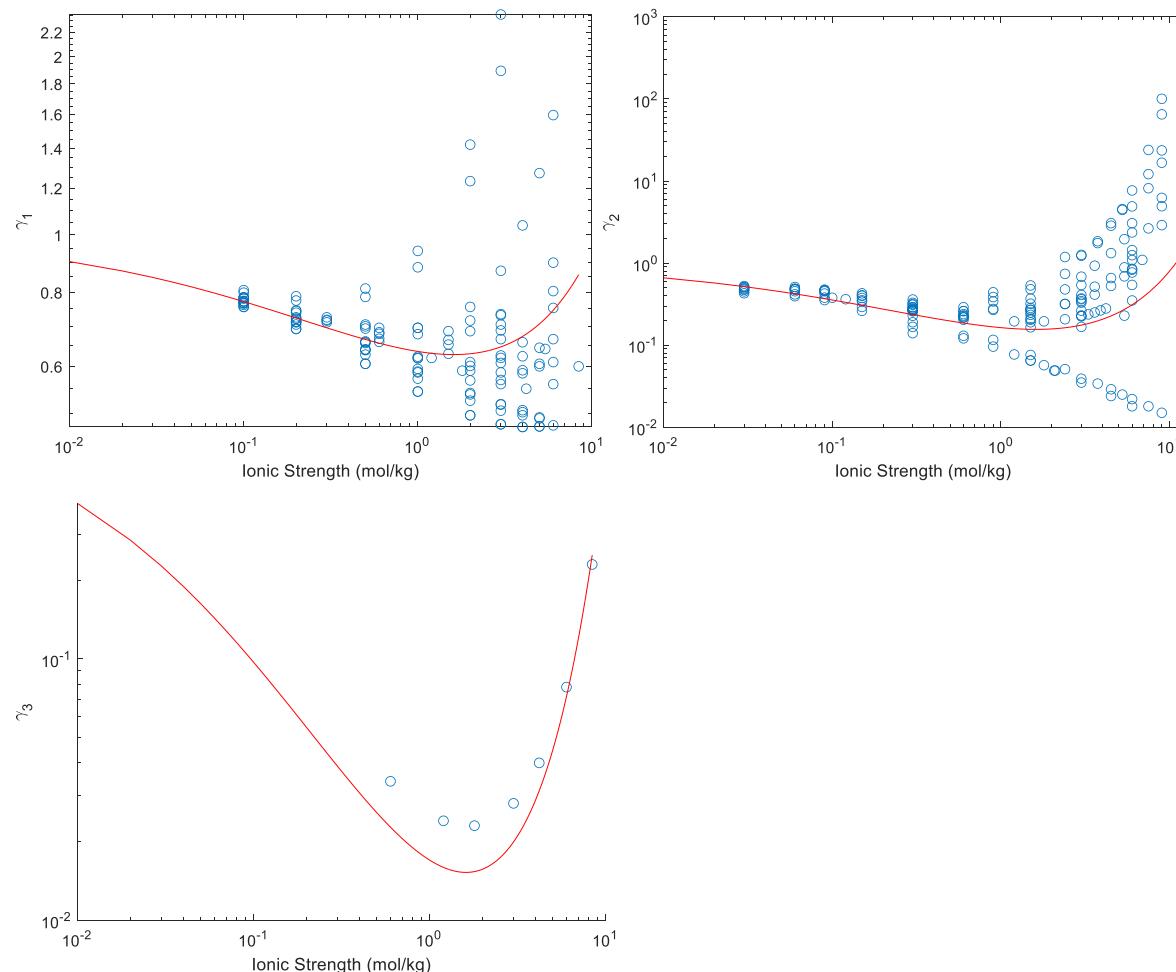


Fig. 6. Comparison of surface model with experimental individual ion activity coefficient data of valences one, two, and three (γ_1 , γ_2 , and γ_3 respectively). The model tends to track the center of the aggregated datasets, representing the long-range portion of the activity coefficient for individual ions. Sources for the experimental data plotted here are provided in Table S3.

and lowest sum of squared errors (SSE) of the M-estimators tested. Moreover, close inspection of the fitting parameters shows that there are only minimal differences between many of the M-estimators examined; generally speaking, monotonous M-estimators produce similar results, but redescending M-estimators produce somewhat different fits (Fig. 3, Figure S2).

Because the data used to produce the model was generated from previous curve-fitting procedures, it is possible to directly compute an expected center for any ionic strength, facilitating our ability to determine the best overall model. Hence, we computed the expected mean and median for the dataset and visually examined them with the results of fitting various M-estimators (Fig. 3). The Least Squares, Fair, Huber, Logistic, and Cauchy models all produce very similar fits; in contrast, the Welsch, Andrews, Bisquare, and Talwar models begin to differ more significantly. The use of the Fair M-estimator is again reinforced here, as it most closely tracks the dataset center throughout.

The Davies equation—perhaps the most commonly used activity model in software due to its simplicity—is plotted for reference and tends to overestimate the experimental activity coefficient curves, especially at ionic strengths exceeding 0.5 molal (Fig. 4). As the ionic strength increases, the fitted surface becomes exponentially different from the predictions resulting from the use of the Davies equation; further, as the valence of the counterion increases, the rate of this change increases as well (Figure S3). This is not surprising. The Davies equation is generally not recommended for use at ionic strengths exceeding 0.5 molal, but it is nevertheless used in multicomponent

system modeling software (such as WATEQ [15] and MINTEQA2 [13]) because few alternatives exist. The work presented in this paper provides a reference state for the long-range activity coefficient that can be built upon. Indeed, compared to the Davies equation, the model described by Equation (9) provides a reasonable (i.e., within the range of existing experimental data) estimate, particularly at high ionic strengths and valences, extending as high as 29 molal. Additionally, the residuals resulting from this fit have a physical meaning: they represent the short-range interactions between the ions. Rearrangement of Equation (2) allows for the straightforward generation of the short-range interaction data (Fig. 5) that is conducive to fitting with UNIFAC [34] or another short-range activity coefficient model.

By using individual ions instead of compounds as the modeling unit, one can quantify the long-range component of the activity coefficient for individual ions. Indeed, previous work has experimentally determined the activity coefficient for a variety of electrolytes, and recent reviews have compiled these sources [64]. Upon comparing the results of this work with aggregated experimental individual ion activity coefficient data (see Table S3 for ions used and relevant sources), the predicted values follow the approximate center of the dataset (Fig. 6), with a RMSE of 11.7. Just as the variance of mean activity coefficient data increases with ionic strength, so too does the variance of individual ion activity coefficient data increase with ionic strength. At high ionic strengths, long-range interactions are still quite strong, but short-range interactions become increasingly prevalent. This model successfully provides a satisfactory baseline activity coefficient for ions in solution.

The model presented here can accurately model long-range contributions to the activity coefficients of ionic species in solution. By incorporating hundreds of datasets for ionizing compounds, it is possible to generate an improved activity coefficient model that does not rely on *a priori* assumptions and is functional over a wide range of ionic strengths. Furthermore, this activity model benefits significantly from its simplicity and can therefore be used for systems involving large numbers of interactions without having to laboriously characterize each individual species and their interactions with other components in solution.

Still, this approach does not fully capture the complex interactions of aqueous species. For systems where a high degree of accuracy is desired, models that include additional terms describing interactions at shorter length scales could prove beneficial. Indeed, the model developed in this paper can be used to quantitatively describe short-range activity coefficient contributions for well-characterized systems (Fig. 5), which can subsequently be described by a model such as UNIFAC, as has been done elsewhere [10]. Finally, we reiterate the limitations of this model alone in high ionic strengths, especially given the impact of short-range interactions on the activity coefficient in these conditions. For example, in a solution with an ionic strength of 1 mol/kg, the semi-empirical equation developed in this work estimates the mean activity coefficient of a 1:1 compound to be approximately 0.64, but the activity coefficient can span the range from approximately 0.22 to 1.9, depending on the salt (Fig. 4). Other semi-empirical models are already used in this regime [40] due to their simplicity; however, the model presented here provides a more mathematically rigorous approach than other models by incorporating a large body of experimental data. Nevertheless, we caution use of this model alone to accurately predict a specific compound's activity at ionic strengths exceeding 1 mol/kg. Instead, we recommend using this model to provide an improved reference state for more comprehensive models incorporating short-range activity coefficient behavior based on perturbation theory [44,45].

A systematic comparison and assessment of various M-estimators also has been conducted. The data used does not appear to suffer greatly from gross outliers—a case where redescending M-estimators tend to excel—and so the choice of M-estimator depends more heavily on its ability to handle heteroskedasticity. As has been described previously, the Fair M-estimator performs exceptionally well in this case [65]. Given that activity coefficient data is highly heteroskedastic, the high performance of the Fair M-estimator compared to other M-estimators is expected.

5. Conclusion

Real-world solutions often contain three main obstacles that make modeling challenging: incomplete dissociation, multiple components with cross-species complexation, and high ionic strengths. Current prediction software often assumes that compounds are completely associated or dissociated, are simple solute-solvent systems, or low (<0.5 molal) ionic strengths. An accurate estimate of the electrostatic contribution to the activity coefficient can facilitate the study of complex systems without the need for detailed parameterization by incorporating well-known and easily determined equilibrium constant data. Here, we provide a model to estimate individual ion activities that can serve as an improved reference state over existing long-range models for perturbation theory. Most importantly, this model will enable solubility predictions for solutions containing mixtures of weak electrolytes at high concentrations. This paves the way for future efforts modeling complex systems in myriad areas including media and feed formulation for intensified bioprocessing and perhaps extending to mathematical descriptions of blood, marine, and concentrated electrolysis solutions.

CRediT authorship contribution statement

Kent M. Rapp: Methodology, Software, Writing – original draft.
Michael J. Betenbaugh: Writing – review & editing, Supervision,

Project administration. **Marc D. Donohue:** Writing – review & editing, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. 2139757. Additionally, this work was funded and supported by the Advanced Mammalian Biomanufacturing Innovation Center (AMBIc) through the Industry-University Cooperative Research Center Program under the US National Science Foundation (Grant numbers 1624684 and 2100800). We would like to express our gratitude to all AMBIc member companies for their mentorship and financial support.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.molliq.2024.125144>.

References

- [1] P. Debye, E. Hückel, Zur Theorie der Elektrolyte, *Phys. Zeitschrift.* (n.d.).
- [2] E.A. Guggenheim, J.C. Turgeon, Specific interaction of ions, *Trans. Faraday Soc.* 51 (1955) 747–761, <https://doi.org/10.1039/tf9555100747>.
- [3] E. Glueckauf, Activity Coefficients in Concentrated Solutions Containing Several Electrolytes, *Nature* 163 (1949) 414–415, <https://doi.org/10.1038/163414a0>.
- [4] K.S. Pitzer, Electrolytes. From Dilute Solutions to Fused Salts, *J. Am. Chem. Soc.* 102 (1980) 2902–2906, <https://doi.org/10.1021/ja00529a006>.
- [5] C.-C. Chen, L.B. Evans, A local composition model for the excess Gibbs energy of aqueous electrolyte systems, *AIChE J.* 32 (1986) 444–454, <https://doi.org/10.1002/aic.690320311>.
- [6] J. Wang, J. Li, W. Cao, L. Zhu, R. Lakerveld, An ion-specific electrolyte non-random two-liquid segment activity coefficient model with improved predictive capabilities for aqueous electrolyte solutions, *Fluid Phase Equilib.* 517 (2020) 112605, <https://doi.org/10.1016/j.fluid.2020.112605>.
- [7] L. Ben Gáida J.B. Gros C.G. Dussap Activity coefficients of concentrated strong and weak electrolytes by a hydration equilibrium and group contribution model *Fluid Phase Equilib.* 289 (2010) 40–48, [10.1016/j.fluid.2009.10.013](https://doi.org/10.1016/j.fluid.2009.10.013).
- [8] C.L. Kusik, H.P. Meissner, E.L. Field, Estimation of phase diagrams and solubilities for aqueous multi-ion systems, *AIChE J.* 25 (1979) 759–762, <https://doi.org/10.1002/aic.690250503>.
- [9] C.L. Kusik, H.P. Meissner, *Electrolyte Activity Coefficients in Inorganic Processing, AIChE Symp. Ser.* 74 (1978) 14–20.
- [10] A. Mohs, J. Gmehling, A revised LIQUAC and LIFAC model (LIQUAC*/LIFAC*) for the prediction of properties of electrolyte containing solutions, *Fluid Phase Equilib.* 337 (2013) 311–322, <https://doi.org/10.1016/j.fluid.2012.09.023>.
- [11] J.R. Rumble, ed., *Properties of Amino Acids*, in: CRC Handb. Chem. Phys., 103rd ed., CRC Press/Taylor & Francis, Boca Raton, FL, 2022.
- [12] M.B. Ewing, T.H. Lilley, G.M. Olofsson, M.T. Rätzsch, G. Somsen, Standard quantities in chemical thermodynamics: Fugacities, activities, and equilibrium constants for pure and mixed phases, *J. Chem. Thermodyn.* 27 (1995) 1–16, <https://doi.org/10.1006/jcht.1995.0001>.
- [13] J.D. Allison, D.S. Brown, K.J. Novo-Gradac, MINTEQA2/PRODEFA2, A Geochemical Assessment Model for Environmental Systems, Version 3.0 User's Manual (1991). <https://www.osti.gov/biblio/5673069>.
- [14] J.P. Gustafsson, Visual MINTEQ 3.1 user guide, Dep. L. Water Resources, Stock. Sweden. (2011) 1–73.
- [15] A.H. Truesdell, B.F. Jones, Wateq, a Computer Program for Calculating Chemical Equilibria of Natural Waters, *J. Res. u.s. Geol. Surv.* 2 (1974) 233–248. <http://pubs.er.usgs.gov/publication/70007400>.
- [16] C.W. Davies, The Extent of Dissociation of Salts in Water. Part VIII. An Equation for the Mean Ionic Activity Coefficient of an Electrolyte in Water, and a Revision of the Dissociation Constants of Some Sulphates, *J. Chem. Soc.* (1938) 2093–2098, <https://doi.org/10.1039/JR9380002093>.
- [17] D.A. MacInnes, The Activities of the Ions of Strong Electrolytes, *J. Am. Chem. Soc.* 1640 (1919) 1086–1092, <https://doi.org/10.1021/ja02228a006>.

[18] R. Heyrovská, Physical Electrochemistry of Strong Electrolytes Based on Partial Dissociation and Hydration: Quantitative Interpretation of the Thermodynamic Properties of NaCl(aq) from “Zero to Saturation”, *J. Electrochem. Soc.* 143 (1996) 1789–1793, <https://doi.org/10.1149/1.1837144>.

[19] M. Landstorfer, On the dissociation degree of ionic solutions considering solvation effects, *Electrochim. Commun.* 92 (2018) 56–59, <https://doi.org/10.1016/j.elecom.2018.05.011>.

[20] H. Bian, X. Wen, J. Li, H. Chen, S. Han, X. Sun, J. Song, W. Zhuang, J. Zheng, Ion clustering in aqueous solutions probed with vibrational energy transfer, *Proc. Natl. Acad. Sci. u. s. a.* 108 (2011) 4737–4742, <https://doi.org/10.1073/pnas.1019565108>.

[21] H. Ohtaki, N. Fukushima, A structural study of saturated aqueous solutions of some alkali halides by X-ray diffraction, *J. Solution Chem.* 21 (1992) 23–38, <https://doi.org/10.1007/BF00648978>.

[22] T.S. Groves, C.S. Perez-Martinez, R. Lhermerout, S. Perkin, Surface Forces and Structure in a Water-in-Salt Electrolyte, *J. Phys. Chem. Lett.* 12 (2021) 1702–1707, <https://doi.org/10.1021/acs.jpclett.0c03718>.

[23] J. Gujt, M. Bešter-Rogač, B. Hribar-Lee, An investigation of ion-pairing of alkali metal halides in aqueous solutions using the electrical conductivity and the Monte Carlo computer simulation methods, *J. Mol. Liq.* 190 (2014) 34–41, <https://doi.org/10.1016/j.molliq.2013.09.025>.

[24] A.A. Chen, R.V. Pappu, Quantitative characterization of ion pairing and cluster formation in strong 1:1 Electrolytes, *J. Phys. Chem. b.* 111 (2007) 6469–6478, <https://doi.org/10.1021/jp0708547>.

[25] D.J. Kozuch, K. Ristroph, R.K. Prud'Homme, P.G. Debenedetti, Insights into Hydrophobic Ion Pairing from Molecular Simulation and Experiment *ACS Nano* 14 (2020) 6097–6106, [10.1021/acsnano.0c01835](https://doi.org/10.1021/acsnano.0c01835).

[26] N.F.A. Van Der Vegt, K. Haldrup, S. Roke, J. Zheng, M. Lund, H.J. Bakker, Water-Mediated Ion Pairing: Occurrence and Relevance, *Chem. Rev.* 116 (2016) 7626–7641, <https://doi.org/10.1021/acs.chemrev.5b00742>.

[27] Y. Luo, W. Jiang, H. Yu, A.D. MacKerell, B. Roux, Simulation study of ion pairing in concentrated aqueous salt solutions with a polarizable force field, *Faraday Discuss.* 160 (2013) 135–149, <https://doi.org/10.1039/c2fd20068f>.

[28] C.J. Fennell, A. Bizjak, V. Vlachy, K.A. Dill, Ion pairing in molecular simulations of aqueous alkali halide solutions, *J. Phys. Chem. b.* 113 (2009) 6782–6791, <https://doi.org/10.1021/jp908484v>.

[29] J. Qian, W.A. Henderson, W. Xu, P. Bhattacharya, M. Engelhard, O. Borodin, J. Zhang, High Rate and Stable Cycling of Lithium Metal Anode (2015), <https://doi.org/10.1038/ncomms7362>.

[30] Y. Marcus, Ionic Radii in Aqueous Solutions, *Chem. Rev.* 88 (1988) 1475–1498, <https://doi.org/10.1021/cr00090a003>.

[31] H. Chen, E. Ruckenstein, Hydrated Ions: From Individual Ions to Ion Pairs to Ion Clusters, *J. Phys. Chem. b.* 119 (2015) 12671–12676, <https://doi.org/10.1021/acs.jpcb.5b06837>.

[32] S.P. Pinho, E.A. Macedo, Solubility of NaCl, NaBr, and KCl in water, methanol, ethanol, and their mixed solvents, *J. Chem. Eng. Data.* 50 (2005) 29–32, <https://doi.org/10.1021/je049922y>.

[33] R.G. Bates, B.R. Staples, R.A. Robinson, Ionic Hydration and Single Ion Activities in Unassociated Chlorides at High Ionic Strengths, *Anal. Chem.* 42 (1970) 867–871, <https://doi.org/10.1021/ac60290a006>.

[34] A. Fredenslund, R.L. Jones, J.M. Prausnitz, Group-contribution estimation of activity coefficients in nonideal liquid mixtures, *AIChE J.* 21 (1975) 1086–1099, <https://doi.org/10.1002/aic.690210607>.

[35] A.A. Lee, C.S. Perez-Martinez, A.M. Smith, S. Perkin, Underscreening in concentrated electrolytes, *Faraday Discuss.* 199 (2017) 239–259, <https://doi.org/10.1039/c6fd00250a>.

[36] A.M. Smith, A.A. Lee, S. Perkin, The Electrostatic Screening Length in Concentrated Electrolytes Increases with Concentration, *J. Phys. Chem. Lett.* 7 (2016) 2157–2163, <https://doi.org/10.1021/acs.jpclett.6b00867>.

[37] J. Li, H.M. Polka, J. Gmehling, A gE model for single and mixed solvent electrolyte systems. 1. Model and results for strong electrolytes, *Fluid Phase Equilib.* 94 (1994) 89–114, [https://doi.org/10.1016/0378-3812\(94\)87052-7](https://doi.org/10.1016/0378-3812(94)87052-7).

[38] W. Yan, M. Topphoff, C. Rose, J. Gmehling, Prediction of vapor-liquid equilibria in mixed-solvent electrolyte systems using the group contribution concept, *Fluid Phase Equilib.* 162 (1999) 97–113, [https://doi.org/10.1016/S0378-3812\(99\)00201-0](https://doi.org/10.1016/S0378-3812(99)00201-0).

[39] J. Gross, G. Sadowski, Perturbed-Chain SAFT: An Equation of State Based on a Perturbation Theory for Chain Molecules, *Ind. Eng. Chem. Res.* 40 (2001) 1244–1260, <https://doi.org/10.1021/acs.iecr.9b01515>.

[40] E. Samson, G. Lemaire, J. Marchand, J.J. Beaudoin, Modeling chemical activity effects in strong ionic solutions, *Comput. Mater. Sci.* 15 (1999) 285–294, [https://doi.org/10.1016/s0927-0256\(99\)00017-8](https://doi.org/10.1016/s0927-0256(99)00017-8).

[41] F.P. Glasser, J. Marchand, E. Samson, Durability of concrete - Degradation phenomena involving detrimental chemical reactions, *Cem. Concr. Res.* 38 (2008) 226–246, <https://doi.org/10.1016/j.cemconres.2007.09.015>.

[42] S. Qin, D. Zou, T. Liu, A. Jivkov, A chemo-transport-damage model for concrete under external sulfate attack, *Cem. Concr. Res.* 132 (2020) 106048, <https://doi.org/10.1016/j.cemconres.2020.106048>.

[43] P. Wang, R. Mo, S. Li, J. Xu, Z. Jin, T. Zhao, D. Wang, A chemo-damage-transport model for chloride ions diffusion in cement-based materials: Combined effects of sulfate attack and temperature, *Constr. Build. Mater.* 288 (2021) 123121, <https://doi.org/10.1016/j.conbuildmat.2021.123121>.

[44] I. Nezbeda, Molecular-thermodynamic reference equations of state, *Fluid Phase Equilib.* 87 (1993) 237–253, [https://doi.org/10.1016/0378-3812\(93\)85029-L](https://doi.org/10.1016/0378-3812(93)85029-L).

[45] I. Nezbeda, On Molecular-Based Equations of State: Perturbation Theories, Simple Models, and SAFT Modeling, *Front. Phys.* 8 (2020) 1–17, <https://doi.org/10.3389/fphy.2020.00287>.

[46] G. Aranovich, P. Donohue, M. Donohue, A lattice model for fluids with directional interactions, *J. Chem. Phys.* 111 (1999) 2050–2059, <https://doi.org/10.1063/1.479473>.

[47] P. Atkins, J. de Paula, *Physical Chemistry*, Oxford University Press, New York, Ninth, 2010.

[48] D.P. Zarubin, The nature of single-ion activity coefficients calculated from potentiometric measurements on cells with liquid junctions, *J. Chem. Thermodyn.* 43 (2011) 1135–1152, <https://doi.org/10.1016/j.jct.2011.02.022>.

[49] J.H. Vera G. Wilczek-Vera Comment on “the nature of single-ion activity coefficients calculated from potentiometric measurements on cells with liquid junctions” by Dmitri P. Zarubin, *J. Chem. Thermodyn.* 43, 1135–1152 J. Chem. Thermodyn. 47 (2012) 2011 442 444 10.1016/j.jct.2011.11.011.

[50] W.J. Hamer, Y.-C. Wu, Osmotic Coefficients and Mean Activity Coefficients of Univalent Electrolytes in Water at 25°C, *J. Phys. Chem. Ref. Data.* 1 (1972) 1047–1100, <https://doi.org/10.1063/1.3253108>.

[51] R.N. Goldberg, R.L. Nuttall, Evaluated activity and osmotic coefficients for aqueous solutions: The alkaline earth metal halides, *J. Phys. Chem. Ref. Data.* 7 (1978) 263–310, <https://doi.org/10.1063/1.555569>.

[52] J.A. Rard, H.O. Weber, F.H. Spedding, Isopiestic Determination of the Activity Coefficients of Some Aqueous Rare Earth Electrolyte Solutions at 25 °C. 2. The Rare Earth Perchlorates, *J. Chem. Eng. Data.* 22 (1977) 187–201, <https://doi.org/10.1021/je60073a015>.

[53] J.A. Rard, F.H. Spedding, Isopiestic Determination of the Activity Coefficients of Some Aqueous Rare Earth Electrolyte Solutions at 25 °C. 5. Dy(No3)3, Ho(No3)3, and Lu(No3)3, *J. Chem. Eng. Data.* 26 (1981) 391–395.

[54] J.A. Rard, F.H. Spedding, Isopiestic Determination of the Activity Coefficients of Some Aqueous Rare Earth Electrolyte Solutions at 25 °C. 6. Eu(No3)3, Y(No3)3, and YCl3, *J. Chem. Eng. Data.* 27 (1982) 454–461. DOI: 10.1021/je60073a015.

[55] K.S. Pitzer, G. Mayorga, Thermodynamics of electrolytes. II. Activity and osmotic coefficients for strong electrolytes with one or both ions univalent, *J. Phys. Chem.* 77 (1973) 2300–2308, <https://doi.org/10.1021/j100638a009>.

[56] R.N. Goldberg, Evaluated activity and osmotic coefficients for aqueous solutions: thirty-six uni-bivalent electrolytes, *J. Phys. Chem. Ref. Data.* (1981) 671, <https://doi.org/10.1063/1.555646>.

[57] R.N. Goldberg, Evaluated activity and osmotic coefficients for aqueous solutions: Bi-univalent compounds of lead, copper, manganese, and uranium, 1979. DOI: 10.1063/1.555611.

[58] R.N. Goldberg, Evaluated activity and osmotic coefficients for aqueous solutions: Bi-univalent compounds of zinc, cadmium, and ethylene bis(trimethylammonium), Chloride and Iodide (1981), <https://doi.org/10.1063/1.555639>.

[59] R.N. Goldberg, R.L. Nuttall, B.R. Staples, Evaluated Activity and Osmotic Coefficients for Aqueous Solutions: Iron Chloride and the Bi-Univalent Compounds of Nickel and Cobalt (1979), <https://doi.org/10.1063/1.555615>.

[60] F.H. Spedding, H.O. Weber, V.W. Saeger, H.H. Petheram, J.A. Rard, A. Habenschuss, Isopiestic Determination of the Activity Coefficients of Some Aqueous Rare Earth Electrolyte Solutions at 25 °C. 1. The Rare Earth Chlorides, *J. Chem. Eng. Data.* 21 (1976) 3, <https://doi.org/10.1021/je60073a015>.

[61] J.A. Rard, L.E. Shiers, D.J. Heiser, F.H. Spedding, Isopiestic Determination of the Activity Coefficients of Some Aqueous Rare Earth Electrolyte Solutions at 25 °C. 3. The Rare Earth Nitrates, *J. Chem. Eng. Data.* 22 (1977) 337–347, <https://doi.org/10.1021/je60073a015>.

[62] J.A. Rard, D.G. Miller, F.H. Spedding, Isopiestic Determination of the Activity Coefficients of Some Aqueous Rare Earth Electrolyte Solutions at 25 °C. 4. La(No3)3, Pr(No3)3, and Nd(No3)3, *J. Chem. Eng. Data.* 24 (1979) 348–354.

[63] D.Q.F. de Menezes, D.M. Prata, A.R. Secchi, J.C. Pinto, A review on robust M-estimators for regression analysis, *Comput. Chem. Eng.* 147 (2021) 107254, <https://doi.org/10.1016/j.compchemeng.2021.107254>.

[64] L. Sun X. Liang N. Von Solms G.M. Kontogeorgis Analysis of Some Electrolyte Models including Their Ability to Predict the Activity Coefficients of Individual Ions 2020 10.1021/acs.iecr.0c00980.

[65] W.J.J. Rey Introduction to Robust and Quasi-Robust Statistical Methods 1st ed., 1983 Springer Berlin Heidelberg 10.1007/978-3-642-69389-2.