

Protofit: A program for determining surface protonation constants from titration data[☆]

Benjamin F. Turner^{*}, Jeremy B. Fein

Department of Civil Engineering and Geological Sciences, 156 Fitzpatrick Hall, University of Notre Dame, Notre Dame, IN 465561, USA

Received 27 April 2005; received in revised form 6 December 2005; accepted 15 December 2005

Abstract

Determining the surface protonation behavior of natural adsorbents is essential to understand how they interact with their environments. ProtoFit is a tool for analysis of acid–base titration data and optimization of surface protonation models. The program offers a number of useful features including: (1) enables visualization of adsorbent buffering behavior; (2) uses an optimization approach independent of starting titration conditions or initial surface charge; (3) does not require an initial surface charge to be defined or to be treated as an optimizable parameter; (4) includes an error analysis intrinsically as part of the computational methods; and (5) generates simulated titration curves for comparison with observation. ProtoFit will typically be run through ProtoFit-GUI, a graphical user interface providing user-friendly control of model optimization, simulation, and data visualization. ProtoFit calculates an adsorbent proton buffering value as a function of pH from raw titration data (including pH and volume of acid or base added). The data is reduced to a form where the protons required to change the pH of the solution are subtracted out, leaving protons exchanged between solution and surface per unit mass of adsorbent as a function of pH. The buffering intensity function Q_{ads}^* is calculated as the instantaneous slope of this reduced titration curve. Parameters for a surface complexation model are obtained by minimizing the sum of squares between the modeled (i.e. simulated) buffering intensity curve and the experimental data. The variance in the slope estimate, intrinsically produced as part of the Q_{ads}^* calculation, can be used to weight the sum of squares calculation between the measured buffering intensity and a simulated curve. Effects of analytical error on data visualization and model optimization are discussed. Examples are provided of using ProtoFit for data visualization, model optimization, and model evaluation.

© 2006 Elsevier Ltd. All rights reserved.

Keywords: Model; Acid; Base; Mineral; Data analysis

1. Introduction

The protonation behavior of surfaces affects a wide range of geochemical processes such as pH buffering, colloid stability, mineral dissolution and

precipitation behavior, electrostatic surface–surface interactions, metal binding, and organic solute binding to surfaces. ProtoFit is a software tool for analysis of potentiometric acid/base titration data, optimization of surface protonation models, and simulation of titrations. Although ProtoFit is primarily intended to be used for surfaces, it can also be used to model the protonation behavior of natural organic matter (NOM) and other solutes.

[☆] Program web site <http://profit.sourceforge.net>

^{*} Corresponding author. Tel.: +1 574 631 4370;

fax: +1 575 631 9236.

E-mail address: bturner@nd.edu (B.F. Turner).

ProtoFit was created to utilize a buffer value approach in parameterizing discrete-site protonation models utilizing simple acidic and basic proton exchange reactions, although the approach could be adapted to model continuous pK distributions. Currently, the program works with non-electrostatic surface complexation models, and electrostatic models including the diffuse layer, constant capacitance, and Donnan shell models. In addition, our approach provides a context for data visualization that has a number of advantages relative to more widely used approaches. A graphical user interface ProtoFit-GUI was developed to aid the user by providing efficient control of model optimization, simulation, and data visualization.

2. Background

2.1. Surface protonation modeling

By surface protonation, we refer to the mass transfer of protons (H^+ ions) between solution and binding sites on a solid surface or dissolved molecule. Although similar methods can be used for both surface proton binding and proton binding to dissolved molecules, the following discussion pertains to a surface reaction involving the exchange of protons between a solution and an adsorbent. Surface protonation reactions are typically envisioned as follows:



where “ $\equiv R$ ” represents either an inorganic crystallographically-bound or organic functional group on the surface (see Dzombak and Morel, 1990, for a review of the history of surface complexation modeling). In Eq. (1) the species $\equiv RH^0$ exhibits acidic behavior, releasing a proton to yield a negatively-charged surface species; in Eq. (2) the species $\equiv RH^0$ exhibits basic behavior, absorbing a proton to form a positively-charged surface species. If one binding site exhibits both behaviors, it is termed amphoteric, and complex surfaces or molecules can exhibit more than one type of binding site. The equilibrium mass action expressions for Eqs. (1) and (2) are

$$\frac{\{\equiv R^-\}a_{H^+}}{\{\equiv RH^0\}} = K_1 \exp\left(\frac{F\Psi}{RT}\right) \quad (3)$$

and

$$\frac{\{\equiv RH^0\}a_{H^+}}{\{\equiv RH_2^+\}} = K_2 \exp\left(\frac{F\Psi}{RT}\right) \quad (4)$$

constrained by the mass balance expression

$$\{\equiv RH\}_{tot} = \{\equiv RH^0\} + \{\equiv R^-\} + \{\equiv RH_2^+\} \quad (5)$$

where $\{\}$ denotes the concentrations of surface species in moles per L of solution, a_{H^+} is the thermodynamic activity of H^+ , K_1 and K_2 are the equilibrium constants, F is Faraday's constant, Ψ is the surface potential, R is the ideal gas constant, and T is absolute temperature (Stumm, 1992). The surface potential Ψ is a function of surface charge, σ , depending on the electrostatic surface complexation model, or equal to zero for a non-electrostatic model (NEM). The goal of determining the surface protonation constants would be to obtain values of K_1 and/or K_2 , and $\{\equiv RH\}_{tot}$ for each discrete surface site as required by a given model.

2.2. Optimizing models by the proton balance approach

The proton balance approach is perhaps the most widely used approach to optimizing protonation constants among geochemists. The approach does, however, present difficulties for some adsorbents which may be overcome using our approach (Section 3). A description of the proton balance approach follows. A titration designed for determining surface protonation constants usually involves a known amount of solid suspended in an electrolyte solution of known volume and concentration. Small aliquots of strong acid or base are added to the suspension, and pH is measured at each step. A system with NaCl as a background electrolyte and containing an adsorbent being titrated with HCl or NaOH titrant is subject to the following charge balance equation:

$$[Cl^-] + [OH^-] + \sum_j \{\equiv R^-\}_j = [Na^+] + [H^+] + \sum_j \{\equiv RH_2^+\}_j \quad (6)$$

where $[]$ denotes the concentration of an aqueous species per volume or mass of solution. As the titration proceeds, $[Cl^-]$ and $[Na^+]$ change

according to the concentration of acid or base added:

$$[\text{Cl}^-] = [\text{Cl}^-]_0 + C_A \quad (7)$$

$$[\text{Na}^+] = [\text{Na}^+]_0 + C_B \quad (8)$$

where $[\text{Cl}^-]_0$ and $[\text{Na}^+]_0$ are the initial concentrations of Cl^- and Na^+ , and C_A and C_B are the concentrations of acid and base, respectively, added per unit volume or mass of solution. If $[\text{Cl}^-]_0 = [\text{Na}^+]_0$, then substituting Eqs. (7) and (8) into (6) leads to

$$C_A - C_B = [\text{H}^+] - [\text{OH}^-] + \sum_j \{\equiv\text{RH}_2^+\}_j - \sum_j \{\equiv\text{R}^-\}_j \quad (9)$$

(Stumm, 1992). Eq. (9) is reasonable if the amount of strong acid or base present at the beginning of the titration experiment can be expected to be zero. In other words, Eq. (9) applies if it can be assumed that there is no imbalance between positive and negative counterions prior to initiation of the titration. Otherwise, a T_H^0 term should be added to the left side of Eq. (9) (see below). Since C_A and C_B are known, and $[\text{OH}^-]$ and $[\text{H}^+]$ are calculated from pH measurements, the value of $\sum \{\equiv\text{RH}_2^+\}_j - \sum \{\equiv\text{R}^-\}_j$ can be calculated for every pH value covered over the course of the titration by Eq. (9). Using a proton balance algorithm for optimizing model parameters such as FITEQL 4.0 (Herbelin and Westall, 1999), the problem of determining constants for the above model would be solved by finding values of K_1 , K_2 , and $\{\equiv\text{RH}\}_{\text{tot}}$ that would result in a minimum weighted sum of squares difference between the left and right sides of Eq. (9). For natural organic surfaces or natural dissolved ligands, the difficulty with the approach inherent in Eq. (9) is that an imbalance in electrolyte counterions may exist prior to initiation of the titrations, causing the mass balance to fail. Preparation of natural organic matter or biological surfaces (such as bacteria) for titration is generally more complex than preparation of mineral surfaces, decreasing the likelihood that $[\text{Cl}^-]_0 = [\text{Na}^+]_0$. For example, preparation of Leonardite humic acid by Westall et al. (1995) by dialysis designed to promote complete dissolution and eliminate the size fraction readily passing through the dialysis membrane necessarily altered the acid–base balance of the system. As another example, preparation of biological surfaces (such as bacteria) for titration involves introducing to

suspension wet biomass, for which the initial acid–base balance is unknown (Fein et al., 2005).

The following is an explanation of how the difficulty described above is dealt with by the proton balance approach. Assuming no counterion complexation with functional groups, the proton mass balance for a system where an adsorbent is being titrated is

$$T_H^{\text{calc}} = [\text{H}^+] - [\text{OH}^-] + \sum_j \{\equiv\text{RH}_2^+\}_j - \sum_j \{\equiv\text{R}^-\}_j \quad (10)$$

where T_H^{calc} is the total concentration of component H^+ calculated from species in contact with solution (Westall et al., 1995). Implicit in the understanding of Eq. (10) is that $\{\equiv\text{RH}^0\}$ is the reference state, or zero proton condition, for ligand j (i.e. protons in $\{\equiv\text{RH}^0\}_j$ do not contribute to T_H^{calc}). In an experimental system, $\sum \{\equiv\text{RH}_2^+\}_j - \sum \{\equiv\text{R}^-\}_j$ is unknown, thus the experimental value of T_H , T_H^{exp} , is defined as

$$T_H^{\text{exp}} = C_A - C_B + T_H^0 \quad (11)$$

where T_H^0 represents the concentration of strong acid or strong base initially present (Westall et al., 1995). For a system containing only NaCl as a background electrolyte, we can think of T_H^0 as equaling the difference $[\text{Cl}^-]_0 - [\text{Na}^+]_0$. This difference may be a significant part of the proton balance while insignificant in terms of the total NaCl concentration. Under such conditions, Eq. (9) may be rewritten as

$$C_A - C_B + T_H^0 = [\text{H}^+] - [\text{OH}^-] + \sum_j \{\equiv\text{RH}_2^+\}_j - \sum_j \{\equiv\text{R}^-\}_j \quad (12)$$

Since experimental determination of T_H^0 is likely to be impractical in most cases, Westall et al. (1995) suggested that T_H^0 be treated as an adjustable parameter. Sets of equilibrium constants and site concentrations, plus T_H^0 , are optimized by minimizing the weighted sum of squares between T_H^{exp} and T_H^{calc} . This approach has the disadvantage in that the concept of T_H^0 is confusing, and that it requires the use of one additional adjustable parameter beyond the protonation constants sought. As discussed in Section 3, ProtoFit avoids this problem by using a different approach to the optimization problem.

2.3. Use of buffer value in the analysis of titration data

When conducting a potentiometric (acid–base) titration, the goal is to constrain the pH buffering behavior of either the solution, a dissolved component in solution, or a suspended adsorbent. A convenient way interpret titration data is through a buffer value function for the solution or adsorbent. This function represents the capacity of a solution or adsorbent to resist changes in system pH as a function of pH; this is the behavior of interest to the experimenter. van Slyke (1922) was the first to use the concept of buffer value in the interpretation of titration data. The buffer value β , of a solution as defined by van Slyke is the instantaneous rate of change in the amount of strong base added to the solution (B) per pH unit:

$$\beta = \frac{dB}{dpH} \quad (13)$$

van Slyke (1922), later investigators such as Grunwald (1951) and Kilpi (1952), and recent investigators such as Chen et al. (1996) used derivative-based approaches to determine protonation constants for dissolved acids. Villieras et al. (1992, 1997) used a derivative-based approach to interpret measurement of gases adsorbed to solid surfaces. Prelot et al. (2002) adapted the approach of Villieras et al. to interpret the protonation behavior of heterogeneous oxide surfaces. ProtoFit, as described in Section 3, provides a new algorithm for calculating a buffering value function for an adsorbent and optimizing surface complexation models to fit the observed buffering behavior. Since derivative functions such as van Slyke's represent the behavior of interest to the experimenter, they are useful for data visualization. Villieras et al. (1997) noted that visualization of a derivative-based adsorption function, rather than undifferentiated adsorption data, made it much easier to draw conclusions regarding the adsorption behavior of interest. However, they also warned that problems with experimental noise could be exacerbated by using a derivative approach. ProtoFit enables the analyst to easily examine the experimental buffering value function and compare it to model predictions (see Section 4). The algorithm used by ProtoFit automatically estimates the standard error in the experimental buffering value function, which can be used to mitigate problems with experimental noise in the optimization problem (see Sections 3.2

and 3.3). Further discussion of the effects of noise on data visualization and model optimization is provided in Section 5.

3. Problem solving approach used in ProtoFit

3.1. Calculation of the buffer value

The basis for the ProtoFit approach is the following proton mass balance equation valid for a system with zero solution alkalinity:

$$\Delta n_{H^+,total,i} = \Delta n_{H^+,wat,i} + \Delta n_{H^+,ads,i} \quad (14)$$

where Δn_{H^+} refers to the amount (i.e. moles) of proton added to the system as a whole (total), water, and adsorbent from the beginning of the titration to step i . Calculation of $\Delta n_{H^+,total}$, representing the amount of strong acid added (or removed) from the system, comes directly from the titration data

$$\Delta n_{H^+,total,i} = V_{titr,i} N_{acid} \quad (15)$$

where $V_{titr,i}$ is the total volume of acid or base added at step i , and N_{acid} is the normality of the acid (negative value if base). The value of $\Delta n_{H^+,wat}$, representing the amount of protons exchanged with water over the course of the titration, is found by speciating the solution using measured pH values:

$$\Delta n_{H^+,wat} = (V_0 + V_{titr,i}) \left(\frac{10^{-pH_i}}{\gamma_{i,H^+}} - \frac{10^{-14+pH_i}}{\gamma_{i,OH^-}} \right) - V_0 \left(\frac{10^{-pH_0}}{\gamma_{0,H^+}} - \frac{10^{-14+pH_0}}{\gamma_{0,OH^-}} \right) \quad (16)$$

where γ is the activity coefficient, and V_0 refers to the solution volume at the beginning of the titration. ProtoFit can calculate activity coefficients by a number of methods, including the Davies, Debye–Huckel, and Truesdell–Jones equations, or can assume activity coefficients equal unity. The value of $\Delta n_{H^+,ads}$ is obtained after substituting Eqs. (15) and (16) into (14). Hence the amount of protons exchanged with the adsorbent, $\Delta n_{H^+,ads}$, is calculated directly from experimental measurements.

ProtoFit uses the values of $\Delta n_{H^+,ads}$ to calculate an adsorbent proton buffering function (Q_{ads}^*) as a function of pH, representing the buffer values of the adsorbent. This quantity expresses what the titration is intended to measure: the pH-dependence of the surface's ability to buffer solution pH. The function Q_{ads}^* is the derivative of $\Delta n_{H^+,ads}$ normalized to adsorbent mass, calculated by performing a

polynomial regression on a closely-grouped set of data points. The polynomial regression is performed using a series of matrix operations suitable for a multiple regression (see Turner, 2005 for details). For a calculation of Q_{ads}^* at data point i , the group of points involved in the regression includes point i plus the two points before and after (i.e. $i - 2, i - 1, i, i + 1, i + 2$). The regression estimates the parameters b_0 , b_1 , and b_2 for the polynomial

$$Q_{ads} = \frac{\Delta n_{H^+,ads,i}}{M_{ads}} = b_0 + b_1x + b_2x^2 \quad (17)$$

where M_{ads} is the mass of the adsorbent, and x is the pH value. The slope Q'_{ads} of this function at point i is the derivative of the above equation:

$$Q'_{ads,i} = b_1 + b_2x_i \quad (18)$$

Calculation of Q'_{ads} at the first two and last two points in the titration requires a slight modification to the above method (see Turner, 2005 for details). The derivative Q'_{ads} (i.e. dQ_{ads}/dpH) yields negative values; since it is more convenient to visually compare positive values, the derivative is converted to one yielding a positive value by

$$Q_{ads}^* = \frac{dQ_{ads}}{d \log a_{H^+}} = \frac{dQ_{ads}}{dpH} = -Q'_{ads} \quad (19)$$

3.2. Error analysis

One advantage of our algorithm for calculating Q_{ads}^* is that an error analysis is intrinsically part of the computation. Treating Q_{ads}^* as a linear combination of b_1 and b_2 , the variance in the slope estimate is calculated by

$$V(Q_{ads}^*) = V(Q'_{ads}) = V(b_1) + 4x_i \text{Cov}(b_1, b_2) + 4x_i^2 V(b_2) \quad (20)$$

where the variance $V()$ and covariance $\text{Cov}()$ values of the coefficients are obtained from the associated variance-covariance matrix (see Turner, 2005 for details). Values of $V(Q_{ads}^*)$ reflect the uncertainty in Q_{ads}^* due to experimental noise. The error values $V(Q_{ads}^*)$ permit the optimization routine to give more weight to the highest quality portions of the titration data, while giving less weight to noisy or poorly-constrained portions of the data (see Section 3.3). Also, the estimate of Q_{ads}^* tends to be the least well-constrained at the very beginning and end of the titration curve, and this is reflected in $V(Q_{ads}^*)$ (see Turner, 2005 for details). In addition to

weighting of the optimization, visualization of $V(Q_{ads}^*)$ provides the user with a gauge of the quality of the dataset.

3.3. Parameter optimization

ProtoFit, similarly to FITEQL, uses a weighted least sum of squares approach to determine surface protonation constants. However, ProtoFit avoids the use of the T_H^0 parameter often necessary for using the proton balance approach by following a fundamentally different way of defining the problem. ProtoFit optimizes model constants by minimizing the weighted sum of squares difference between Q_{ads}^* (derived from experimental data) and a model proton buffering function F_{ads}^* (calculated using a surface protonation model). The optimization proceeds by adjusting the model constants until the weighted sum of squares difference between Q_{ads}^* and F_{ads}^* arrives at a minimum. The simulated derivative function F_{ads}^* is obtained by finding the slope of a function F_{ads} defined as

$$F_{ads}(i) = \frac{(\sigma_i - \sigma_1)SSA}{F_d} \quad (21)$$

for $i > 1$ where σ is surface charge (e.g. Coulombs m^{-2}) calculated using the surface protonation model, F_d is Faraday's constant (e.g. 96490 C mol^{-1}), and SSA is the specific surface area (e.g. $m^2 kg^{-1}$) of the surface. The value of σ is defined as

$$\sigma = \frac{F_d}{SSA} \sum_j (\{ \equiv RH_2^+ \}_j - \{ \equiv R^- \}_j) \quad (22)$$

where the subscript j refers to a given surface site. When using the non-electrostatic model (NEM), where calculation of surface potential Ψ from σ is not necessary, the choice of SSA value can be arbitrary, provided that $SSA > 0$. The function F_{ads} has the same dimensions as Q , and F_{ads}^* is obtained from F_{ads} in a way similar to how Q_{ads}^* is obtained from Q_{ads} , except with zero analytical error (see Turner, 2005 for details on finding the derivative of F_{ads}). The weighted sum of squares objective function SS^* minimized by ProtoFit is

$$SS^* = \frac{\sum_i (F_{ads}^* - Q_{ads}^*)^2}{\sum_i V(Q_{ads}^*)^{0.5w}} \quad (23)$$

where w is the error weighting parameter. This function is designed so that points with greater uncertainty in Q_{ads}^* are given less weight in the

optimization depending on the value of w used. An analysis of the effect of experimental noise on data analysis and model optimization is reported in Section 5. When more than one dataset is used in the optimization, each dataset is given equal weight in calculating a total SS^* . ProtoFit adjusts the values of the model parameters until SS^* reaches a minimum. The algorithm used for minimizing SS^* is adapted from the “pattern search” algorithm of Hook and Jeeves (1960).

Since least squares optimization techniques can often yield non-unique solutions or solutions representing local rather than global minima, the solution arrived at is often dependent on the starting guess. ProtoFit addresses this problem by automatically iterating over a number of starting guesses according to search parameters specified by the user. The algorithm then tabulates a set of best-fitting solutions and reports them to the user.

4. Interpreting the Q_{ads}^* function

ProtoFit operates by calculating from the raw titration data an adsorbent proton buffering function Q_{ads}^* which represents the behavior that the titration is intended to measure. Since Q_{ads}^* is independent of the starting pH of the titration, it provides for a meaningful direct comparison between datasets that would not be possible for raw titration data unless starting pH values were exactly the same. Furthermore, the shapes of the Q_{ads}^* curves have physically meaningful interpretations. In a discrete-site surface protonation model, each dissociation reaction for each site will produce a “peak” in the Q_{ads}^* function. For example, titration of acetic acid, a monoprotic organic acid, yields a single peak in Q_{ads}^* (Fig. 1). The width and position of the peak will depend on the surface protonation model used: peaks generated by a nonelectrostatic model (as is applicable to the acetic acid example) will be relatively narrow and centered at the pK for the reaction; peaks generated by an electrostatic model will be broader and shifted either down-pH (if the reaction involves a positively-charged surface species) or up-pH (if the reaction involves a negatively-charged surface species). If electrostatics play an important role in surface speciation, the peaks will shift position as ionic strength changes.

The use of the Q_{ads}^* function amplifies the “signal” from the titration data, and also may be useful for bringing the limitations of the titration data to light. For instance, compared to raw titration data or

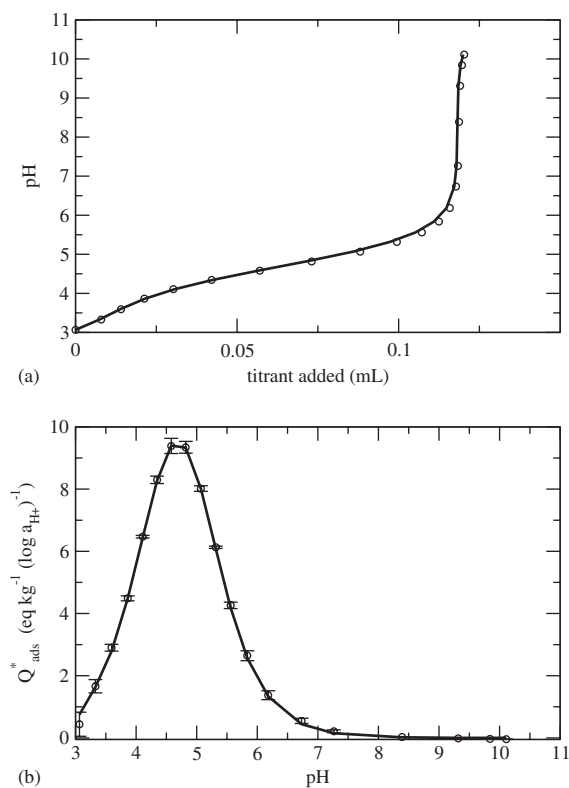


Fig. 1. Comparison of raw titration data to function Q_{ads}^* used by ProtoFit: (a) raw titration data for acetic acid; (b) function Q_{ads}^* (“proton buffering” capacity as a function of pH). Solid lines in (a) and (b) represent one-site NEM optimized by ProtoFit: $\log K = -4.69$, $\log C = 1.23 \log(\text{mol kg}^{-1})$. Error bars indicate standard error of Q_{ads}^* (i.e. square root of $V(Q_{ads}^*)$).

“ $C_A - C_B$ ” vs. pH plots, visualizing the Q_{ads}^* function may make it easier to determine whether the titration adequately captures the surface protonation behavior of concern. Also, comparing a simulated titration to Q_{ads}^* may make it easier to see the limitations of a model that may otherwise appear reasonable compared to titration data plotted in a conventional sense. For example, the one-site double layer model (DLM) for corundum (Al_2O_3) parameterized using FITEQL by Hayes et al. (1991) appears reasonable when compared to the raw titration data (Fig. 2), but its limitations become strongly apparent when compared to Q_{ads}^* . A one-site DLM is clearly not suitable for this surface; a more complex model would be required to explain the surface’s protonation behavior. The reason that visualization of raw titration data obscures the poor fit is that much of the base added during the titration was used to increase the pH of

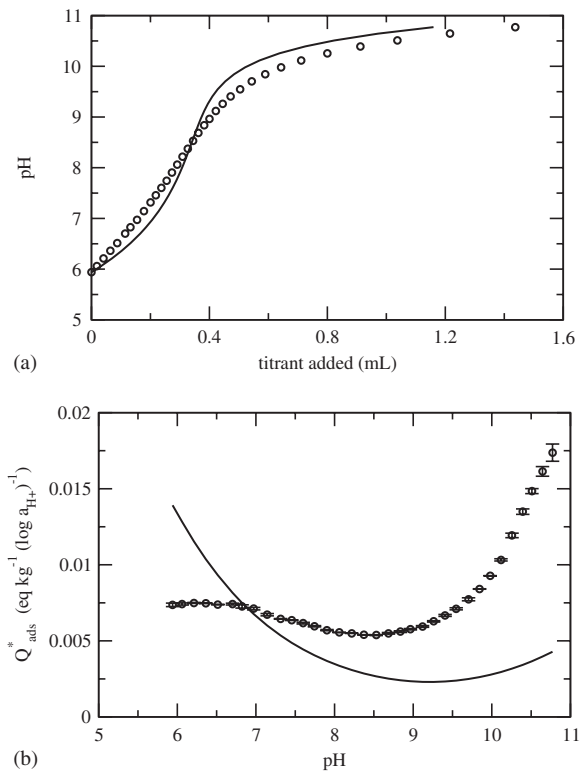


Fig. 2. Comparison of a prediction by a published one-site DLM for corundum (Al_2O_3) to a titration dataset used to develop model: (a) model simulation (solid line) appears to reasonably fit raw titration data (open circles); (b) comparison between Q_{ads}^* (open circles) and model prediction (solid line). Error bars indicate standard error of Q_{ads}^* . Data and model from Hayes et al. (1991); base volumes back calculated assuming a normality of 1 N and an initial solution volume of 1 M.

the solution, such that the base consumed in changing the surface protonation was overshadowed.

5. Effects of experimental noise

Experimental noise has important effects on the analysis of titration data, affecting both the quality of data interpretation and model optimization. There are two potential sources of experimental noise: (1) error in the pH measurement; and (2) error in titrant volume. When calculating Q_{ads}^* , an additional source of error is the uncertainty inherent in estimating the activity coefficients for H^+ and OH^- , which contributes to error in estimating the buffering capacity of water. For these reasons, error in Q_{ads}^* will tend to be greatest at low and high pH

where the buffering capacity of water is high compared to the buffering capacity of the adsorbent, and lowest at near-neutral pH where the buffering capacity of water is least.

While the Q_{ads}^* function amplifies the “signal” sought from the titration, it also amplifies the noise. Thus, noise will be much more readily apparent in the Q_{ads}^* function than in the raw titration data, for instance. This amplification of errors can be useful to the experimentalist in a number of ways; for example, experimental problems producing unacceptable error may be more readily identifiable than otherwise. Also, error estimates can be utilized to weight the sum of squares calculation such that the optimization is not heavily skewed by poorly-constrained data points.

It is intuitive that error in Q_{ads}^* will affect the optimization of model parameters, but it is not clear without further analysis how and to what extent. To address this issue, we have conducted a set of simulations designed to investigate the effect of error in pH on visualization of Q_{ads}^* and the optimization of model parameters. The study is based on two simulated titrations calculated using PHREEQC (Parkhurst and Appelo, 1999), one with 104 data points, and one with 26 data points. Error in the pH measurements were simulated by adding a random normal deviate to each pH value according to

$$\text{pH}_{sim,i} = \text{pH}_i + \sigma_{err} Z_i \quad (24)$$

where σ_{err} is the standard error in the simulated pH measurement and Z_i is the random normal deviate (see Table 1 for a description of the simulations). The values of σ_{err} used were 0.003 and 0.03 pH units. In order to investigate the effect of errors on Q_{ads}^* and the accuracy of the model optimization,

Table 1
Model parameters used to simulate titration using PHREEQC

| Parameter | Value |
|----------------------------|---------------|
| Number of sites | 3 |
| Site type | Acidic |
| pK values | 4.0, 6.0, 8.0 |
| Site concentration | 1 mol/kg |
| Surface complexation model | NEM |
| Background electrolyte | 0.01 M NaCl |
| Titrant | 1.0 N NaOH |
| Solution volume | 1 L |
| Adsorbent mass | 1 g |

each simulated curve was modified by Eq. (24) using 10 different sets of random normal deviates.

The error weighting scheme has an effect on the parts of the titration curve that receive the most weight during the optimization process. When no error weighting is used, the portions of the curve with the greatest density of data points are given the most weight in the optimization (in this case, at high and low pH). However, since these portions tend to have the most error associated with them (due in part to uncertainties in estimating the proton buffering of water), they are weighted less in the optimization if error weighting is utilized, yielding a more balanced distribution of weights.

The number of data points in the titrations has a significant effect on the appearance of the Q_{ads}^* function. Specifically, more closely-spaced data points lead to greater uncertainty in individual Q_{ads}^* values (Fig. 3). At a σ_{err} value of 0.03, the Q_{ads}^* function with 104 data points is very noisy, while the function with 24 data points is relatively smooth. The reason for this is that the more closely-spaced the points, the larger the uncertainty in the position of the points is relative to the distance between points; hence the greater uncertainty in the slope estimate. An implication for this is that the user may choose to eliminate data points in the data set that are too close together in order to obtain a less noisy Q_{ads}^* signal. In order to test the effect of this on optimization, another set of simulated data was generated by “weeding” the 104-point datasets so that no data points are separated from adjacent points by less than 0.05 pH units. Since closely-spaced points lead to uncertainty in the slope estimate, the weeding of the dataset yields less apparent noise in Q_{ads}^* .

The effects of number of titration data points, error amount, error weighting, and data point weeding on accuracy of the model optimization were tested, with results reported in Table 2. The most striking trend is that the larger value of σ_{err} consistently yielded greater uncertainty in the optimization results. At the lower value of σ_{err} , it is not clear whether error weighting or weeding improve the accuracy of the optimization results. At the higher value of σ_{err} , weeding of the datasets yielded more accurate results, while it is not clear if error weighting improved accuracy in the 104-point datasets. However, error weighting did improve the accuracy of results in the 26-point datasets at a σ_{err} value of 0.03 pH units. Overall, the most accurate results were obtained from the weeded 104-point

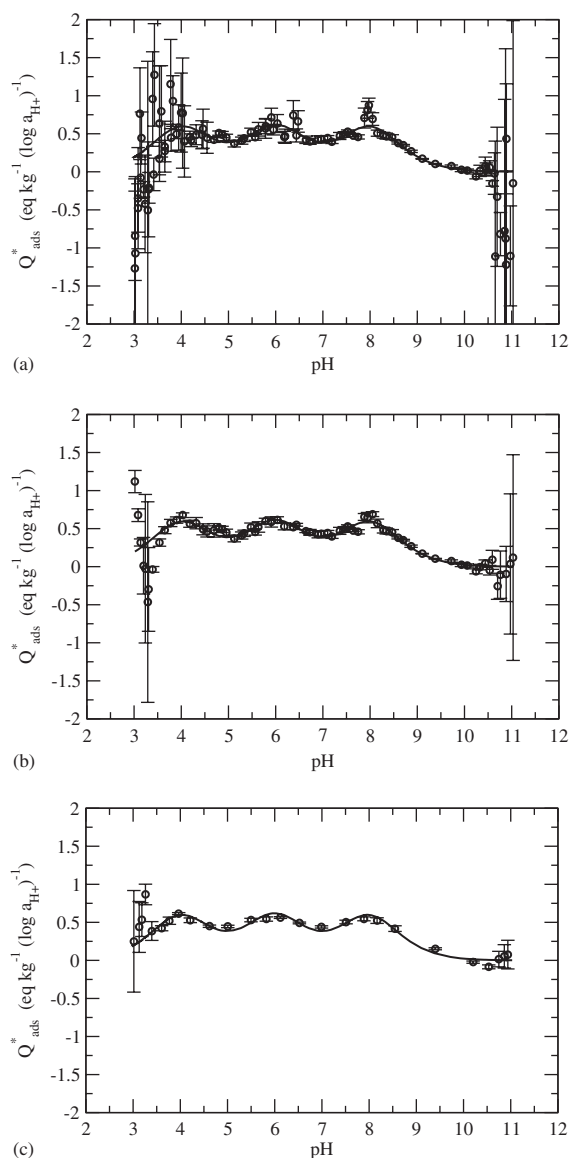


Fig. 3. Effect on appearance of Q_{ads}^* of number and spacing of data points in a simulated titration with artificially introduced error: (a) titration data with 104 data points; (b) 104-point data set “weeded” to remove closely-spaced points (so that no points are closer than 0.05 pH units); (c) titration data with 26 data points. Value of σ_{err} in each case is 0.03. Solid lines indicate F_{ads}^* . Error bars indicate standard error of Q_{ads}^* .

datasets with a σ_{err} value of 0.003 pH units and error weighting enabled, and the least accurate results were obtained from the 26-point datasets with a σ_{err} value of 0.03 pH units and without error weighting.

This investigation of the effect of simulated errors on data analysis and model optimization processes yields insights that are not otherwise obvious. For

Table 2

Results of optimization exercise involving simulated dataset. Values reported are average of optimized values with standard deviations below

| | $\log K_1$ | $\log K_2$ | $\log K_3$ | $\log C_1$ | $\log C_2$ | $\log C_3$ |
|---|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Simulated model | −4.000 | −6.000 | −8.000 | 0.000 | 0.000 | 0.000 |
| 104 data points, $\sigma_{err} = 0.003$ | | | | | | |
| Unweighted, unweeded | −4.039 0.030 | −6.011 0.010 | −8.004 0.004 | −0.004 0.002 | −0.003 0.003 | 0.000 0.001 |
| Unweighted, weeded | −4.034 0.019 | −6.009 0.006 | −8.003 0.003 | −0.006 0.003 | −0.002 0.002 | 0.000 0.001 |
| Weighted, unweeded | −4.017 0.012 | −6.005 0.007 | −8.004 0.006 | −0.001 0.003 | 0.001 0.003 | 0.000 0.003 |
| Weighted, weeded | −4.025 0.009 | −6.006 0.007 | −8.005 0.006 | −0.004 0.002 | 0.001 0.002 | 0.000 0.002 |
| 104 data points, $\sigma_{err} = 0.03$ | | | | | | |
| Unweighted, unweeded | −3.805 1.235 | −6.007 0.182 | −8.000 0.059 | 0.234 0.783 | 0.004 0.062 | 0.014 0.014 |
| Unweighted, weeded | −4.271 0.132 | −6.078 0.030 | −8.018 0.022 | −0.033 0.046 | −0.013 0.019 | 0.001 0.004 |
| Weighted, unweeded | −4.300 0.516 | −6.204 0.568 | −8.123 0.252 | −0.030 0.052 | −0.010 0.037 | −0.088 0.243 |
| Weighted, weeded | −4.080 0.082 | −6.016 0.061 | −8.032 0.042 | 0.010 0.033 | −0.004 0.028 | −0.002 0.011 |
| 26 data points, $\sigma_{err} = 0.003$ | | | | | | |
| Unweighted | −4.034 0.033 | −6.005 0.018 | −7.952 0.009 | −0.006 0.002 | −0.013 0.003 | −0.022 0.002 |
| Weighted | −4.001 0.024 | −5.986 0.015 | −7.959 0.021 | −0.012 0.003 | −0.006 0.003 | −0.012 0.005 |
| 26 data points, $\sigma_{err} = 0.03$ | | | | | | |
| Unweighted | −3.617 1.004 | −5.824 0.475 | −7.858 0.252 | 0.174 0.369 | 0.000 0.040 | −0.006 0.052 |
| Weighted | −4.109 0.111 | −6.047 0.101 | −7.955 0.049 | −0.001 0.030 | −0.018 0.016 | −0.026 0.021 |

Statistics are based on 10 random instances of each dataset. Value of weighting parameter used is 1.0.

the purpose of model optimization, these results indicate that a larger number of data points will tend to yield more accurate results, and accuracy may be improved in some cases by using error weighting in the optimization and/or weeding out closely-spaced data points. For data visualization purposes, weeding out closely-spaced data points yields a Q_{ads}^* function with less noise.

6. Using ProtoFit

Prior to using ProtoFit, titration data must be recorded in an input file in the format specified in the User's Manual (Turner, 2005). Most users will create the file using a text editor. Required parameters include the pH and volume of titrant for each titration data point, normality of acid or base, adsorbent mass, and the initial solution

volume and ionic strength. Adsorbent specific surface area is also required if electrostatic surface complexation models are to be used. Titrant volume is specified relative to the beginning of the dataset, rather than relative to the pH of immersion.

ProtoFit runs in two modes: (1) optimization mode, and (2) simulation mode. In optimization mode, the parameters for a surface protonation model are optimized to fit the function Q_{ads}^* (see Section 3.3). ProtoFit will typically iterate over several sets of starting guesses for the parameters, converging on a number of parameter sets that may or may not be significantly different (depending largely on the complexity of the model). The user will typically choose the best-fitting parameter set for the model. In simulation mode, a simulated titration is calculated given a parameterized surface protonation model. The simulated titration is usually used for

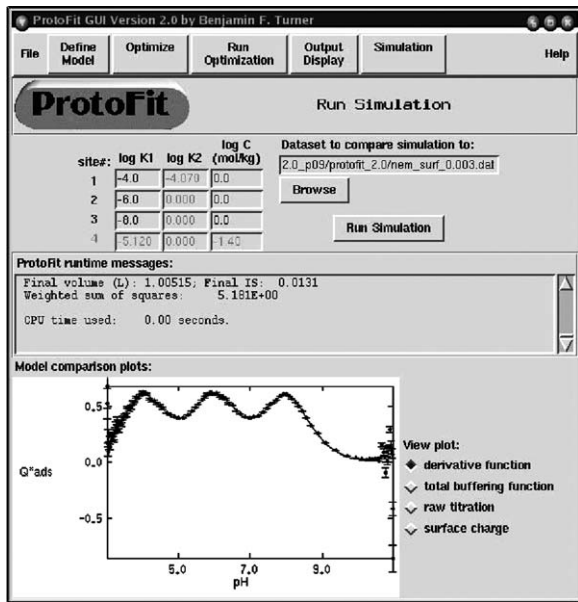


Fig. 4. Screenshot of simulation window in ProtoFit-GUI. Interface provides for rapid visualization of data, including comparisons of data to model simulations of Q_{ads}^* and raw titration. Error bars indicate standard error of Q_{ads}^* .

comparison with an observed titration in order to judge the reasonableness of the model.

ProtoFit will typically be run through the graphical user interface ProtoFit-GUI (Fig. 4). In ProtoFit-GUI, the user is able to (1) define a surface protonation model; (2) specify how the optimization should be run, including the dataset(s) to which the model is to be optimized; (3) run ProtoFit in optimization mode and peruse optimized parameter sets; and (4) run ProtoFit in simulation mode, and graphically compare the simulated titration to the observed raw titration data and Q_{ads}^* function. Individual optimized parameter sets can be selected and automatically input into the titration simulator. The simulation component of ProtoFit-GUI automatically produces plots of observed and model calculated adsorbent buffer value (Q_{ads}^*), system buffer value (water plus adsorbent), and raw titration data for comparison with the model simulation.

7. Examples

7.1. Example 1: Modeling ionic strength effects by optimizing to multiple datasets simultaneously

An important advantage of electrostatic surface complexation models such as the double layer

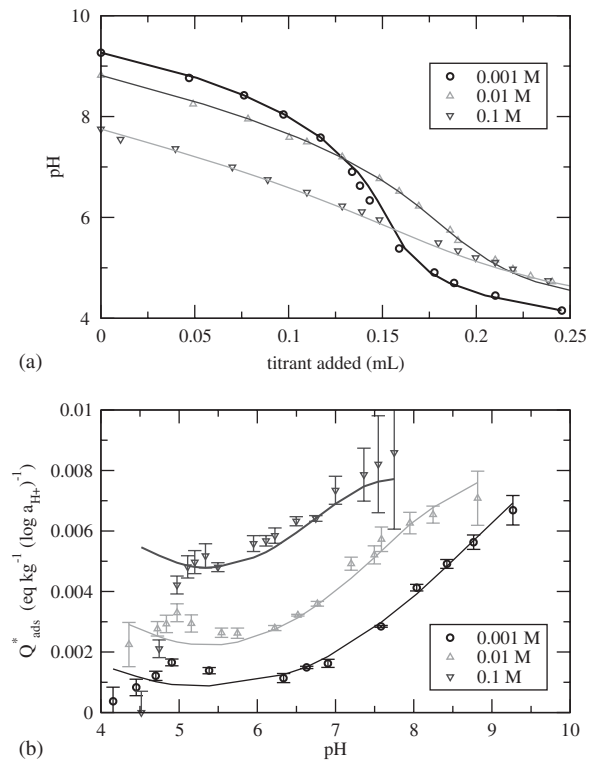


Fig. 5. A one-site DLM optimized for rutile (TiO_2) at three different ionic strengths: (a) raw titration data compared with model simulations; (b) Q_{ads}^* function compared with model predictions. Error bars indicate standard error of Q_{ads}^* . Model parameters: $\log K_1 = -6.30$, $\log K_2 = -4.46$, $\log C = -1.63 \log(\text{mol kg}^{-1})$ (i.e. 0.70 sites nm^{-2}). Data from Hayes et al. (1991); mL values back calculated assuming solution volume of 1.0 L and acid normality of 1.0 N. Error bars indicate standard error of Q_{ads}^* .

model (DLM) is the ability to predict the effect of ionic strength on surface speciation. Ideally, a single model can predict surface protonation behavior of a given surface over a range of ionic strength conditions. Hence, it is advantageous to parameterize a model by optimizing to a number of datasets simultaneously (with datasets spanning a range of ionic strength values), rather than to individual datasets.

ProtoFit allows for a model parameter set to be optimized to several titration datasets simultaneously to obtain a best fit between the model and the entire set of titrations. ProtoFit accomplishes this by calculating SS^* for each titration dataset by Eq. (23), and optimizing the parameter set by minimizing the sum of SS^* values.

In the example given in Fig. 5, titration datasets of rutile at ionic strength values of 0.001, 0.01, and 0.1 M

were used to obtain an optimized 1-site DLM assuming surface reactions as written in Eqs. (1) and (2). A rough idea of the reasonableness of the model can be obtained by examining how well the model predicts the raw titration data. From Fig. 5a it appears that the model predicts the raw titration behavior very well. More detailed information regarding the reasonableness of the model is obtained by examining the fit of the model to the Q_{ads}^* function. It is apparent from Fig. 5b that the model reasonably predicts the ionic strength dependence of Q_{ads}^* . However, a closer examination reveals that the titration data are perhaps not as well constrained as it might appear at first glance. According to the model, the pH range covered in the titrations corresponds to a “trough” in the Q_{ads}^* function between two peaks corresponding to the pK’s for Eqs. (1) and (2). In this light, it is apparent that neither peak is fully captured by the titrations, with the lower pH peak captured to a lesser extent than the higher pH peak. While the model reasonably predicts the ionic strength dependence of the surface’s buffering capacity in the pH range covered by the experiments, it is unclear whether the model will make reasonable predictions outside of this range or whether the model is a physically reasonable representation of the surface.

7.2. Example 2: Modeling protonation behavior of a biological surface

The determination of the surface protonation characteristics of biological surfaces is especially problematic since the initial degree of surface protonation is usually not known. Since ProtoFit does not require this as a constraint, it has a distinct advantage over the proton balance approach (such as that used by FITEQL), which must treat the initial hydrogen ion component concentration (T_H^0) as an optimizable parameter.

In this example, a set of bacterial titration data for *B. subtilis* (Fein et al., 2005) is used to obtain a surface protonation model. Over the pH range typical for titrations of biological surfaces, it is reasonable to assume that each surface site is subject to an acidic protonation reaction as expressed by Eq. (1). Since biological surfaces are more complex than mineral surfaces, and unlike mineral surfaces the charge on biological surfaces is not concentrated in a single plane, it is reasonable to avoid the electrostatic surface complexation models used for mineral surfaces and use a NEM. A four-site NEM

fitted to the data by ProtoFit provides a reasonable fit to both Q_{ads}^* and the raw titration data (Fig. 6).

Alternatively, this model may be optimized by the proton balance approach (Section 2.2), which requires input of a data series comprised of paired $C_A - C_B$ and pH values. In this example, the total exchange of protons with the bacterial surface prior to titration (e.g. through bacterial growth, rinsing, and acid washing procedures) is unquantified. Hence the quantity $C_A - C_B$, rather than reflecting the addition of acid or base relative to a zero proton condition (i.e. where $T_H = 0$), reflects the addition of acid or base relative to a reference point where T_H is unknown. The unknown value of T_H at the

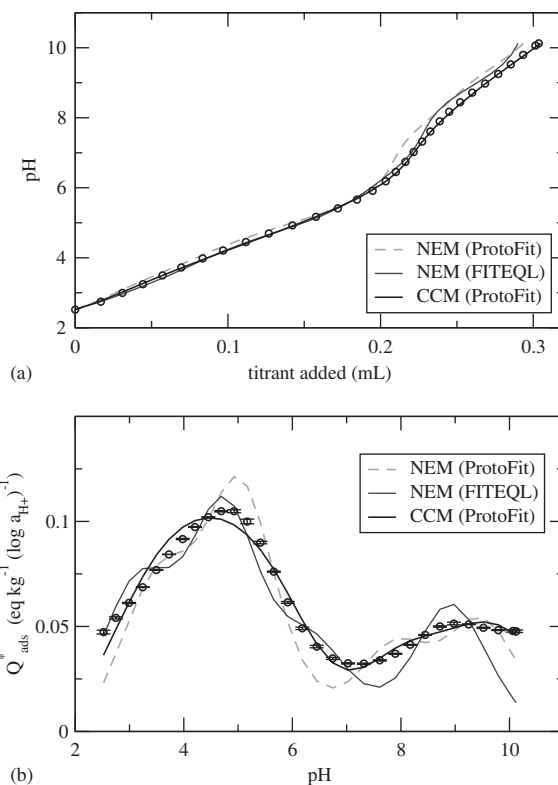


Fig. 6. Fitting of four-site NEM and three-site CCM to bacterial titration data: (a) raw titration data compared with model simulations; (b) Q_{ads}^* function compared with model predictions. Error bars indicate standard error of Q_{ads}^* . Model parameters for ProtoFit NEM: $\log K$ values: $-3.44, -5.07, -7.54, -9.48$; $\log C$ values ($\log(\text{mol kg}^{-1})$): $-0.87, -0.73, -1.17, -0.98$. Model parameters for FITEQL NEM: $\log K$ values: $-3.13, -4.75, -6.40, -8.92$; $\log C$ values ($\log(\text{mol kg}^{-1})$): $-0.93, -0.74, -1.17, -0.97$. Model parameters for ProtoFit CCM: $\log K$ values: $-3.63, -6.23, -7.45$; $\log C$ values ($\log(\text{mol kg}^{-1})$): $-0.46, -1.17, -0.98$; capacitance = 2.4. Titration data is of *B. subtilis* in 0.1 M NaClO₄ from Fein et al. (2005). Error bars indicate standard error of Q_{ads}^* . Error weighting option was not used in parameter optimization by ProtoFit in this example.

reference point is treated by the algorithm as an optimizable parameter (T_H^0). To illustrate the use of this approach, we optimize the NEM using FITEQL and defining the reference point where $C_A - C_B = 0$ to be the start of the base titration dataset (pH 2.521). The optimized model provides a reasonable fit to the data similar to the model optimized by ProtoFit (Fig. 6). Since the speciation of the bacterial surface at the reference point can be predicted using the model in a speciation calculator such as PHREEQC (Parkhurst and Appelo, 1999), the model-predicted value of T_H^{calc} (1.27×10^{-3} eq L⁻¹) at the point can be calculated by evaluating Eq. (10). However, comparing this with the optimized value of T_H^0 (8.12×10^{-4} eq L⁻¹) reveals that the two values are not the same. Since the approach minimizes the sum of squares difference between T_H^{calc} and T_H^{exp} over the dataset as a whole (not just at the reference point), T_H^{calc} is not identical to T_H^{exp} when $C_A - C_B = 0$. Therefore, the optimized value of T_H^0 does not necessarily match what one would predict using the model.

Although the NEM describes the data reasonably, analysis of Q_{ads}^* (Fig. 6) reveals “peaks” that are broader than what would be expected for a surface with a small number of discrete sites governed by the NEM. One approach, frequently used in the modeling the protonation behavior of humic acids, is to assume a continuous distribution of pK values (e.g. Perdue et al., 1984). Another possible approach is to use a simple electrostatic model to add breadth to the peaks through simulated electric field effects. For comparison with the NEM, a three-site constant capacitance model (CCM) was used to fit the observed surface protonation behavior. It is apparent from Fig. 6 that the CCM provides a superior fit to the surface’s protonation behavior. However, although this use of the CCM provides a good empirical fit to the data, the electrostatic effects implied in the model are likely not mechanistically or quantitatively realistic for a bacterial surface.

7.3. Example 3: Using ProtoFit with molecular acids

Although ProtoFit is primarily designed for determining speciation constants for surfaces, it can just as easily be used to determine speciation constants for some molecular acids such as humic acids. This can be accomplished using the non-electrostatic adsorption model just as one would for a surface. Since specific surface area is not

meaningful in this case, the SSA value supplied to the program would be any arbitrary value (provided $SSA > 0$). An example of using ProtoFit for determination of speciation constants for an organic acid was provided in Fig. 1. A similar approach is possible for modeling the more complicated protonation behavior of humic and fulvic acids. However, the NEM is unable to predict the subtle ionic strength effects affecting protonation of humic and fulvic acids.

8. Summary and conclusions

The protonation behavior of natural surfaces including mineral and biological substrates is important to a wide range of geochemical processes. ProtoFit is a software package for data visualization, model simulation, and optimization of surface protonation models using acid–base titration data that characterize the buffering capacity of these surfaces. The optimization approach used by ProtoFit is to calculate an adsorbent proton buffering function (Q_{ads}^*) from raw titration data, and to minimize the weighted sum of squares between this function and the model prediction. The approach offers important features including:

- ProtoFit does not require that the initial surface charge be known, nor does it require that initial surface charge be effectively treated as an optimizable parameter.
- Automatically computed error estimates can be used to weight the sum of squares calculation, such that well-constrained data are given greater weight than poorly-constrained data.
- The adsorbent proton buffering function Q_{ads}^* represents the behavior that the titration is intended to measure. Visual examination of Q_{ads}^* provides greater insight into data and model limitations than visualization of raw titration data.

The approach used in ProtoFit is particularly suitable for organic/biological adsorbents for which the initial surface charge cannot be measured.

Our stochastic analysis of the effects of noise on model optimization indicates that a larger number of data points yield more accurate model optimization results. Accuracy can also be improved by error weighting the sum of squares function and weeding out closely-spaced data points.

The software package includes ProtoFit-GUI, a graphical user interface providing user-friendly control of model optimization, simulation, and data visualization. The software is designed for cross-platform portability, and should run on Microsoft Windows and most popular Unix-like operating systems. ProtoFit is written in Fortran-95 and ProtoFit-GUI is written in Tcl/Tk. ProtoFit and ProtoFit-GUI are Free Software licensed under the Gnu General Public License. The ProtoFit software package and documentation can be downloaded at <http://protofit.sourceforge.net>

Acknowledgments

Funding for this work was provided by a NSF Environmental Molecular Science Institute grant (NSF-EAR-0221966). Comments from Mark Jensen, Johnson Haas and three anonymous reviewers led to improvements to the manuscript.

References

- Chen, J.F., Xia, Y.X., Choppin, G.R., 1996. Derivative analysis of potentiometric titration data to obtain protonation constants. *Analytical Chemistry* 68, 3973–3978.
- Dzombak, D.A., Morel, F.M.M., 1990. *Surface Complexation Modeling: Hydrous Ferric Oxide*. Wiley-Interscience, New York, 393pp.
- Fein, J.B., Boily, J., Yee, N., Gorman-Lewis, D., Turner, B.F., 2005. Potentiometric titrations of *Bacillus subtilis* cells and a comparison of modeling approaches. *Geochimica et Cosmochimica Acta* 69, 1123–1132.
- Grunwald, E., 1951. A differential method of measuring acid and base dissociation constants. *Journal of the American Chemical Society* 73, 4934–4938.
- Hayes, K.F., Redden, G., Ela, W., Leckie, J.O., 1991. Surface complexation models: an evaluation of model parameter estimation using FITEQL and oxide mineral titration data. *Journal of Colloid Interface Science* 142, 448–469.
- Herbelin, A.L., Westall, J.C., 1999. FITEQL 4.0: a computer program for determination of chemical equilibrium constants from experimental data. Report 99-01. Department of Chemistry, Oregon State University, Corvallis.
- Hook, R., Jeeves, T.A., 1960. “Direct search” solution of numerical and statistical problems. *Journal of the Association for Computing Machinery* 8, 212–229.
- Kilpi, S., 1952. A differential method of measuring acid and base dissociation constants. *Journal of the American Chemical Society* 74, 5296–5298.
- Parkhurst, D.L., Appelo, C.A.J., 1999. User’s guide to PHREEQC (version 2)—a computer program for speciation, batch-reaction, one-dimensional transport, and inverse geochemical calculations. U.S. Geological Survey Water Investigations Report 99-4259, 312pp.
- Perdue, E.M., Reuter, J.H., Parish, R.S., 1984. A statistical model of proton binding by humus. *Geochimica et Cosmochimica Acta* 48, 1257–1263.
- Prelot, B., Charmas, R., Zarzycki, P., Thomas, F., Villieras, F., Piasecki, W., Rudzinski, W., 2002. Application of the theoretical 1-pK approach to analyzing proton adsorption isotherm derivatives on heterogeneous oxide surfaces. *Journal of Physical Chemistry B* 106, 13280–13286.
- Stumm, W., 1992. *Chemistry of the Solid–Water Interface*. Wiley-Interscience, New York, 428pp.
- Turner, B.F., 2005. ProtoFit Version 2.0, a program for determining surface speciation constants from titration data: user’s manual. Department of Civil Engineering and Geological Sciences, University of Notre Dame, Notre Dame, 27pp.
- van Slyke, D.D., 1922. On the measurement of buffer values and on the relationship of buffer value to the dissociation constant of the buffer and the concentration and reaction of the buffer solution. *Journal of Biological Chemistry* 52, 525–570.
- Villieras, F., Cases, J.M., Francois, M., Michot, L.J., Thomas, F., 1992. Texture and surface energetic heterogeneity of solids from modeling of low pressure gas adsorption isotherms. *Langmuir* 8, 1789–1795.
- Villieras, F., Michot, L.J., Bardot, F., Cases, J.M., Francois, M., Rudzinski, W., 1997. An improved derivative isotherm summation method to study surface heterogeneity of clay minerals. *Langmuir* 13, 1104–1117.
- Westall, J.C., Jones, J.D., Turner, G.D., Zachara, J.M., 1995. Models for association of metal ions with heterogeneous environmental sorbents. 1. Complexation of Co(II) by Leonardite Humic Acid as a function of pH and NaClO₄ concentration. *Environmental Science & Technology* 29, 951–959.