

ProtoFit Version 1.0

**A program for determining surface speciation constants from
titration data**

USER'S MANUAL

April 13, 2004

Benjamin F. Turner

Dept. of Civil Engineering and Geological Sciences
156 Fitzpatrick Hall
University of Notre Dame
Notre Dame, IN 46556

Copyright Benjamin F. Turner, 2004. All rights reserved.

This document may be freely copied and distributed.

Table of Contents

1. Introduction and Overview.....	1
2. Concepts and Approach.....	3
2.1 Use of titration data.....	3
2.1.1 Reduction of raw titration data.....	3
2.1.2 Calculation of the dataset derivative function y^*	4
2.1.3 Calculation of activity coefficients.....	5
2.1.4 Calculation of ionic strength.....	6
2.2 Model Calculation.....	6
2.2.1 Double Layer Model (DLM).....	7
2.2.2 Constant Capacitance Model (CCM).....	8
2.2.3 Donnan Shell Model (DSM).....	8
2.2.4 Non-electrostatic adsorption model (NEM).....	8
2.2.5 Calculation of the model derivative function.....	9
2.3. Optimization of model parameters.....	9
Addendum: Definition of symbols.....	9
3. Using ProtoFit.....	11
3.1 Constructing dataset files.....	11
3.2 Using ProtoFit-GUI.....	11
3.2.1 Installation.....	11
3.2.2 Invoking ProtoFit GUI.....	11
3.2.3 Working with ProtoFit GUI.....	12
3.2.4 Defining the model.....	12
3.2.5 Setting the optimization parameters.....	14
3.2.6 Running the optimization and viewing output.....	15
3.2.7 Running a simulation.....	16
3.3 Using ProtoFit from the command line.....	18
3.4 Tips and Tricks.....	18
3.5 Example Datasets.....	19
References.....	20
Appendix 1. Main Input File Format.....	21
Appendix 2. Compiling ProtoFit and ProtoFit GUI from source code.....	23

1. Introduction and Overview

ProtoFit is a computer tool for calculating surface protonation constants from titration data. ProtoFit can optimize constants for one of several surface complexation models with one to four discrete surface sites. ProtoFit is similar in function in many ways to FITEQL (Westall, 1982), but is easier to use and has some additional features. ProtoFit can optimize a model's constants to one or several datasets simultaneously. In addition to optimizing model parameters, ProtoFit also has the capability of simulating titrations using surface complexation models.

ProtoFit-GUI is an easy-to-use graphical user interface (GUI) to ProtoFit. The GUI offers a user friendly environment for conducting model optimizations and simulations. The GUI has built-in capability for visualization of results, including comparisons between observations and model predictions. Most users will want to run ProtoFit through ProtoFit-GUI, although ProtoFit can also be run from the command line for additional flexibility. ProtoFit and ProtoFit-GUI are designed for cross-platform portability, and should run on most popular operating systems including Linux, Microsoft Windows, and Macintosh OS X.

ProtoFit optimizes model parameters (e.g. equilibrium constants and surface site concentrations) by minimizing the sum of squares between a *dataset derivative function* and a *model derivative function*. The derivative function is essentially the pH-buffering capacity of the surface as a function of pH. The *dataset* derivative function is estimated from the raw titration data, and the *model* derivative function is calculated using the model. Use of the derivative function approach rather than a charge-balance approach bypasses the problematic “ T_{H0} ” issue inherent in conventional approaches to analyzing titrations of organic functional groups (See Westall et al., 1995 for more information on the “ T_{H0} ” issue).

ProtoFit works with models that involve surface reactions forming monovalent positively or negatively-charged surface species:



A functional group is termed “acidic” if it releases a proton to form a negatively-charged surface species (Equation 1.1a) and “basic” if it absorbs a proton to form a positively-charged surface species (Equation 1.1b). A functional group that exhibits both behaviors is termed “amphoteric.” In this type of surface site, neutral, positively-charged and negatively-charged surface species coexist. Functional groups on typical oxide minerals are commonly assumed to be amphoteric (i.e. protonation reactions for a given surface functional group can be expressed by both Equations 1.1a and 1.1b).

Organic surfaces tend to contain acidic functional groups, having only one surface protonation reaction



In this type of surface site, only neutral and negatively-charged surface species exist. Note that [Equation 1.2](#) is equivalent to [Equation 1.1a](#), with the exception of the notation differences reflecting how organic functional groups are conventionally expressed. A given surface may contain a number of surface functional groups, each with its own site concentration. For each functional group in a user-defined model, ProtoFit will optimize the site concentration and the equilibrium constant for each surface reaction.

The purpose of this manual is to provide the user with information regarding the approach and calculations used by ProtoFit, and practical information regarding the use of ProtoFit and ProtoFit-GUI. ProtoFit and ProtoFit-GUI are *free software* distributed under the terms of the Gnu General Public License (see <http://www.gnu.org/copyleft/gpl.html>). Additional information can be found at <http://protofit.sourceforge.net>.

2. Concepts and Approach

ProtoFit optimizes model parameters by minimizing the sum of squares between a *dataset derivative function* and a *model derivative function*. Measured titration data (including pH and volume of acid added) is reduced to the net protons added or removed from a surface ($y(i)$) for each titration step (i). The derivative of this function with respect to pH is the rate of proton exchange per pH unit, referred to here as the *dataset derivative function*, y^* . This function represents the proton-buffering capacity of the surface as a function of pH.

A simulated net proton exchange between a surface and solution ($f(i)$) is computed using a surface speciation model for the same titration steps (i). The derivative of this function, referred to as the *model derivative function* f^* , is compared to the dataset derivative function. The optimization proceeds by finding a set of surface speciation constants that provide a minimum weighted sum of squares between the two derivative functions. An optimized model parameter set is a set of model parameters (including reaction equilibrium constants and site concentrations) that results in a minimized weighted sum of squares.

2.1 Use of titration data

2.1.1 Reduction of raw titration data

During a titration, some amount of acid or base is added to the system at each titration step, resulting in a change in solution pH. In a system with zero solution alkalinity, the proton mass balance can be written as

$$\Delta n_{H^+, total, i} = \Delta n_{H^+, solution, i} + \Delta n_{H^+, surface, i} \quad (2.1.1)$$

where Δn_{H^+} refers to the amount of proton added to the system as a whole (total), solution, and surface from the beginning of the titration to step i . Calculation of $\Delta n_{H^+, total}$ comes directly from the titration data:

$$\Delta n_{H^+, total, i} = \Delta V_i \cdot N_{acid} \quad (2.1.2)$$

where ΔV_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 $\Delta n_{H^+, solution}$ is found by speciating the solution:

$$\Delta n_{H^+, solution, i} = V_i \left(\frac{10^{-pH_i}}{\gamma_{i, H^+}} - \frac{10^{14+pH_i}}{\gamma_{i, OH^-}} \right) - V_0 \left(\frac{10^{-pH_0}}{\gamma_{0, H^+}} - \frac{10^{14+pH_0}}{\gamma_{0, OH^-}} \right) \quad (2.1.3)$$

where γ is the activity coefficient (see [Section 2.1.3 Calculation of activity coefficients](#)), V refers to the total solution volume, and subscripts I and i refer to the initial state of the system and the i^{th} titration step, respectively. The value of $\Delta n_{H^+, surface}$ is found after substituting [Equations \(2.1.2\) and \(2.1.3\)](#) into [\(2.1.1\)](#). The value of $y(i)$ (normalized to adsorbent mass) is then found by

$$y(i) = \frac{\Delta n_{H^+, surface, i}}{M_{ads}} \quad (2.1.4)$$

for $i > I$ where M_{ads} is the mass of adsorbent, and $y(I) = 0$. A value for y is calculated for each titration point. The variable y represents the cumulative amount of proton added or subtracted

from the surface per adsorbent mass at each point in the titration.

2.1.2 Calculation of the dataset derivative function y^*

Examining the derivative of y with respect to pH or $\log a_{H^+}$ has a number of advantages relative to examining y itself:

- The pH-dependence of surface protonation becomes more strongly apparent. The “signal” that the titration is intended to measure is amplified.
- While the value of y is strongly dependent on the initial pH of the titration, the derivative is completely independent of the starting pH.
- All values of the derivative will be like-signed, while the sign of y will be positive or negative depending on whether a titration is performed with an acid or a base.

A disadvantage of examining the derivative is that analytical error is magnified just as signal is magnified. Fortunately, the method of calculating the derivative below allows for analytical error to be estimated. The analytical error is used to weight the sum of squares calculation so that poorly constrained data points are given less weight than well-behaved data points.

The derivative of y with respect to pH (i.e. $-\log a_{H^+}$) for data point i is calculated by performing a second-order polynomial regression on a subset of y . The data subset used are the five data points closest to i . The regression is designed to estimate the parameters b_0 , b_1 , and b_2 for the polynomial

$$y = b_0 + b_1 x + b_2 x^2 \quad (2.1.5)$$

where x is the pH value. The slope of this function at point i is the derivative of the above equation:

$$y'_i = b_1 + 2 b_2 x_i \quad (2.1.6)$$

The polynomial regression is performed using a series of matrix operations suitable for a multiple regression (Neter et al., 1985). The matrix X containing the independent variables and Y containing the dependent variables are defined as

$$X = \begin{bmatrix} 1 & x_{i-2} & x_{i-2}^2 \\ 1 & x_{i-1} & x_{i-1}^2 \\ 1 & x_i & x_i^2 \\ 1 & x_{i+1} & x_{i+1}^2 \\ 1 & x_{i+2} & x_{i+2}^2 \end{bmatrix} \quad \text{and} \quad Y = \begin{bmatrix} y_{i-2} \\ y_{i-1} \\ y_i \\ y_{i+1} \\ y_{i+2} \end{bmatrix} \quad (2.1.7)$$

The matrix b containing the regression parameters is calculated by

$$b = (X' X)^{-1} X' Y \quad (2.1.8)$$

where $b_0 = b[1,1]$, $b_1 = b[2,2]$, and $b_2 = b[3,3]$. The variance-covariance matrix $s^2(b)$ is calculated by

$$s^2(b) = MSE (X' X)^{-1} \quad (2.1.9)$$

where

$$MSE = \frac{(\mathbf{Y} - \mathbf{Xb})'(\mathbf{Y} - \mathbf{Xb})}{k - p} \quad (2.1.10)$$

and k is the number of data points used in the regression (i.e. five), and p is the number of parameters in the regression model (i.e. three). Note that $k-p$ is the number of degrees of freedom.

Treating y' as a linear combination of b_1 and b_2 , the variance in the slope estimate is calculated by

$$V(y') = V(b_1) + 4x_i \text{Cov}(b_1, b_2) + 4x_i^2 V(b_2) \quad (2.1.11)$$

where the variance $V()$ and covariance $\text{Cov}()$ values are obtained from $s^2(\mathbf{b})$.

The derivative y' (i.e. dy/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

$$y^* = \frac{dy}{d \log a_{H^+}} = \frac{-dy}{d pH} = -y' ; \quad V(y^*) = V(y') \quad (2.1.12)$$

Values of $V(y^*)$ are used to weight the sum of squares calculation (See [Section 2.3](#)).

For a titration data set of m points, a value of y_i^* is calculated by the above method for points 3 through $(m-2)$. For points 1 and 2, or $m-1$ and m , regressions are performed at $i=3$ or $i=(m-2)$, respectively, and values of $x_{i-2}, x_{i-1}, x_{i+1}, x_{i+2}$ are substituted as appropriate into [Equations 2.1.6 and 2.1.11](#) to obtain values of y' and $V(y')$, respectively. Hence values of y^* and $V(y^*)$ are calculated for every point in the titration. However, values of y^* calculated at $x_{i-2}, x_{i-1}, x_{i+1}, x_{i+2}$ (as done only for the beginning or ending two points in the titration) are naturally less well-constrained than values calculated at x_i ; this is reflected in a tendency for higher $V(y^*)$ values at the beginning and end of the titration.

2.1.3 Calculation of activity coefficients

Activity coefficients define the relationship between an ion's concentration and its thermodynamic activity. The quantity pH is usually defined as the $-\log$ of the H^+ activity. In order to convert measured pH values to concentrations of H^+ and OH^- , activity coefficients must be used.

One commonly used expression for calculating the activity coefficient γ_i for a given ion i is the Extended Debye-Huckel (EDH) equation:

$$\log \gamma_i = \frac{-A z^2 \sqrt{I}}{1 + B a \sqrt{I}} \quad (2.1.13)$$

where A and B are constants equal to 0.5092 and 0.3283, respectively, at 25°C, z is the charge of the ion for which the coefficient is being calculated for (± 1 in the case of H^+ , OH^-), I is ionic strength, and a is an ion-specific parameter (Langmuir, 1997). Values of a for H^+ and OH^- are 9.0 and 3.5, respectively. The EDH equation is considered to be the most accurate for ionic strengths up to about 0.1 M (Langmuir, 1997).

Another commonly used expression, considered to be accurate up to ionic strengths of about 0.7 M is the Davies equation

$$\log \gamma_i = -A z^2 \left(\frac{\sqrt{I}}{1 + \sqrt{I}} - 0.3 I \right) \quad (2.1.14)$$

(Langmuir, 1997). The Truesdell-Jones (TJ) equation (sometimes referred to as the modified Extended Debye-Huckel equation)

$$\log \gamma_i = \frac{-A z^2 \sqrt{I}}{1 + B a \sqrt{I}} + b I \quad (2.1.15)$$

where b_i is an ion-specific parameter. Values of a and b used in the TJ equation are respectively 4.78 and 0.24 for H^+ and 10.65 and 0.21 for OH^- . The TJ equation is considered to be accurate to ionic strengths of ~ 2.3 M (Langmuir, 1997). General-purpose geochemical speciation models such as PHREEQC and MINTEQA2 by default use the EDH equation for H^+ and OH^- , while FITEQL uses the Davies equation.

In some cases, such as when pH is defined as the $-\log$ of the H^+ concentration (as for some low-pH electrode calibration techniques), it is useful to assume that activity coefficients equal one in determining the dataset derivative function. Thus, ProtoFit provides the user with this option.

2.1.4 Calculation of ionic strength

The quantity *ionic strength* is important both for the calculation of activity coefficients, and for the calculation of surface potentials for some electrostatic surface complexation models. ProtoFit keeps track of the changes in ionic strength over the course of the titration. The ionic strength at the beginning of the titration, I_0 , is equal to the C_s value supplied in the dataset input file (see [Section 3.1](#)). As the titration proceeds, I is updated according to the amount of acid or base added, and the changes in H^+ and OH^- concentrations:

$$I(i) = \frac{I_0 V_1}{V_i} + \frac{0.5 \cdot |N_{acid}| \cdot \Delta V_i}{V_i} + \frac{10^{-pH_i}}{\gamma_{i,H^+}} + \frac{10^{14+pH_i}}{\gamma_{i,OH^-}} \quad (2.1.16)$$

for each titration point i .

2.2 Model Calculation

For comparison with the dataset derivative function y^* , ProtoFit calculates a model derivative function f^* . This involves calculating a simulated derivative value for pH measurements corresponding to each titration data point. Each pH value is passed to a speciation solver which solves the surface speciation according to a given model. Because the solver uses pH values, and not H^+ or OH^- concentration values, the choice of activity coefficient model does not affect the model derivative function.

Surface speciation models used by ProtoFit include the Double Layer Model (DLM), Constant Capacitance Model (CCM), Donnan Shell Model (DSM), and a non-electrostatic adsorption model (NEM).

2.2.1 Double Layer Model (DLM)

The simulated titration curve is calculated by assuming that pH is fixed at each titration step, and that the surface becomes equilibrated with respect to the solution at each step. Some number of surface deprotonation reactions of the form of [Equations 1.1 or 1.2](#) are assumed to occur at the surface, each with a corresponding equilibrium constant and site concentrations. For the DLM, the mass action expressions for such reactions are

$$\frac{\{SO^-\}[H^+]}{\{SOH^0\}} = K_1 \exp\left(\frac{-\Delta Z \cdot F \Psi}{RT}\right) \quad (2.2.1a)$$

and

$$\frac{\{SOH^0\}[H^+]}{\{SOH_2^+\}} = K_2 \exp\left(\frac{-\Delta Z \cdot F \Psi}{RT}\right) \quad (2.2.1b)$$

for [Equations 1.1a and 1.1b](#), and

$$\frac{\{R-L_i^-\}[H^+]}{\{R-L_iH^0\}} = K_i \exp\left(\frac{-\Delta Z \cdot F \Psi}{RT}\right) \quad (2.2.2)$$

for [Equation 1.2](#), respectively, where $\{\}$ denotes the concentrations of surface species in moles per kg of adsorbent, $[\]$ denotes activity of a solution species, K is the equilibrium constant (optimized by ProtoFit), ΔZ is the change in charge of the surface species for the reaction (i.e. -1 in the case of [Equations 1.1 and 1.2](#)), F is Faraday's constant, Ψ is the surface potential, R is the ideal gas constant, and T is absolute temperature (Stumm, 1992). The mass action expression is subject to the mass balance constraint

$$\{SOH\}_{tot} = \{SOH^0\} + \{SO^-\} + \{SOH_2^+\} \quad (2.2.3)$$

for [Equations 1.1a and 1.1b](#), and

$$\{R-L_iH\}_{tot} = \{R-L_iH^0\} + \{R-L_i^-\} \quad (2.2.4)$$

for [Equation 1.2](#). The total concentration of a given functional group is one of the model parameters optimized by ProtoFit. The surface charge, σ , is calculated by

$$\sigma = F \sum_j \left(\frac{-\{SO^-\}_j + \{SOH_2^+\}_j}{SSA} \right) \quad (2.2.5)$$

and

$$\sigma = F \sum_j \left(\frac{-\{R-L^-\}_j}{SSA} \right) \quad (2.2.6)$$

for [Equations 1.1 and 1.2](#), respectively, where j refers to a specific functional group. The value of Ψ is then calculated using the Double Layer Model (DLM) by

$$\Psi = \frac{2RT}{zF} \sinh^{-1} \left(\frac{\sigma}{\sqrt{8RT\epsilon\epsilon_0c}} \right) \quad (2.2.7)$$

where z is the counterion valence, c is the counterion concentration, ϵ is the dielectric constant of water, and ϵ_0 is the permittivity of free space (Stumm, 1992). Given a set of surface speciation constants (i.e. K and total functional group concentrations), [Equations 2.2.1-2.2.7](#) are solved

simultaneously by iteration for the pH value corresponding to each titration step. The net proton exchange function f (i.e. protons added or removed from the surface per mass of adsorbent) is then calculated by

$$f(i) = \frac{(\sigma_i - \sigma_1) \cdot SSA}{F} \quad (2.2.8)$$

for $i > 1$.

2.2.2 Constant Capacitance Model (CCM)

The Constant Capacitance Model (CCM) is very similar to the DLM, except for the relationship between Ψ and σ :

$$\Psi = \frac{\sigma}{C} \quad (2.2.9)$$

where C is the capacitance. While in some ways simpler than the DLM, the CCM inherently assumes no ionic strength dependence on Ψ .

2.2.3 Donnan Shell Model (DSM)

While the DLM and CCM assume that surface charge exists in a plane, the DSM assumes the charge is distributed throughout a shell-like volume enveloping the surface. The charge per unit shell volume σ_s is calculated by

$$\sigma_s = \sum_j \left(\frac{(-\{SO^-\}_j + \{SOH_2^+\}_j) \cdot F}{SSV} \right) \quad (2.2.10)$$

and

$$\sigma_s = \sum_j \left(\frac{-\{R-L^-\}_j \cdot F}{SSV} \right) \quad (2.2.11)$$

for [Equations 1.1 and 1.2](#), respectively, where SSV is the specific shell volume, the volume of shell per unit mass of adsorbent. The Donnan Shell potential Ψ_{DON} is calculated by

$$\Psi = \frac{RT}{zF} \sinh^{-1} \left(\frac{\sigma_s}{2zFc} \right) \quad (2.2.12)$$

(Wonders et al., 1997). The value of f is obtained by

$$f(i) = \frac{(\sigma_{s,i} - \sigma_{s,1}) \cdot SSV}{F} \quad (2.2.13)$$

for $i > 1$. The DSM is otherwise similar to the DLM.

2.2.4 Non-electrostatic adsorption model (NEM)

In this model, Ψ is assumed to be zero. Hence, no electrostatic effects affect adsorption, and thus [Equation 2.2.1](#) becomes

$$\frac{\{SO^-\}[H^+]}{\{SOH^0\}} = K_1 \quad (2.2.14a)$$

and

$$\frac{\{SOH^0\}[H^+]}{\{SOH_2^+\}} = K_2 \quad , \quad (2.2.14b)$$

and Equation 2.2.2 becomes

$$\frac{\{R-L_i\}[H^+]}{\{R-L_iH^0\}} = K_i \quad . \quad (2.2.15)$$

The model is otherwise similar to the DLM.

2.2.5 Calculation of the model derivative function

The model derivative function f^* is calculated from f for each measured pH value in a similar way to how y^* is calculated from y (Section 2.1.2). However, since f is not affected by experimental error, it is permissible for the polynomial regression to involve zero degrees of freedom (i.e. it uses three data points rather than five) for faster computation.

2.3. Optimization of model parameters

ProtoFit optimizes surface speciation constants by minimizing the weighted sum of squares between the dataset derivative function y^* and the model derivative function f^* . The program can optimize a set of model parameters to one or more datasets simultaneously. For each value of y^* , a corresponding value of f^* is calculated. The weighted sum of squares SS^* is then calculated by

$$SS^* = \frac{\sum_i (f_i^* - y_i^*)^2}{\sum_i V(y_i^*)} \quad . \quad (2.3.1)$$

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 optimization algorithm used is adapted from the “pattern search” algorithm of Hook and Jeeves (1960).

Since optimized parameter sets for complex models tend to be non-unique, ProtoFit will typically iterate over a number of starting guesses to arrive at a number of different optimized parameter sets. The user then has the option of choosing from among the best-fitting parameter sets.

Addendum: Definition of symbols

Dimensions

L	length
M	mass

N	number (e.g. moles)
T	time
Θ	temperature

Variables

a	ion-specific parameter (dimensionless)
A	constant in Davies, EDH, and TJ expressions ($N^{-0.5}L^{1.5}$)
b	ion-specific parameter ($N^{-1}L^3$)
b_0, b_1, b_2	fitting parameters in multiple regression
B	constant in EDH and TJ expressions ($N^{-0.5}L^{1.5}$)
c	counterion concentration ($N L^{-3}$)
C	capacitance ($N^2T^2M^{-1}L^{-2}$)
f	simulated reduced titration curve ($N M^{-1}$)
f^*	model derivate function ($N M^{-1} \log N^{-1}L^3$)
F	Faraday's constant (dimensionless)
i	titration step number (N)
j	surface group number (N)
k	number of data points in used in polynomial regression (N)
I	ionic strength ($N L^{-3}$)
K	equilibrium constant (dimensionless)
m	number of titration data points (N)
M_{ads}	mass of adsorbent (M)
n	number of protons (N)
N_{acid}	normality of acid ($N L^{-3}$)
p	number of parameters in regression model (N)
R	ideal gas constant ($M L^2 T^{-2} N^{-1} \Theta^{-1}$)
SSA	specific surface area of adsorbent ($L^2 M^{-1}$)
SSV	specific shell volume of adsorbent ($L^3 M^{-1}$)
T	absolute temperature (Θ)
V	volume of solution (L^3)
V()	variance statistic (dimensions of the object of the statistic)
x	pH value ($\log N^{-1}L^3$)
y	reduced titration curve ($N M^{-1}$)
y^*	dataset derivate function ($N M^{-1} \log N^{-1}L^3$)
z	charge of ion (N)
Z	charge of surface species (N)
γ	activity coefficient (dimensionless)
ϵ	dielectric constant of water (dimensionless)
ϵ_0	permittivity of free space ($N^2 T^2 L^{-3} M^{-1}$)
σ	surface charge ($N L^{-2}$)
σ_s	charge per unit shell volume ($N L^{-3}$)
Ψ	surface potential ($M L^2 T^{-1} N^{-1}$)

3. Using ProtoFit

ProtoFit can be run in one of two modes: in a graphical user interface (GUI) or from the command line. Most users will run ProtoFit using the GUI, which provides the user with a user-friendly environment for controlling the program and for visualizing data and model simulations. The command-line program provides for greater flexibility in how ProtoFit may be used, allowing ProtoFit to be run from fast and efficient scripts, etc.

Both the GUI and command line approaches involve passing instructions to ProtoFit, reading dataset input files, and producing output. Dataset files are not created by the GUI, and must be created by the user.

3.1 Constructing dataset files

One dataset file must be created by the user for each dataset. The files include parameters specific to a given experiment, including paired pH and volume (mL) of acid/base added. The files are plain text files that can be created in a text editor or spreadsheet. The files must have the format as described in [Table 3.1](#). A few things to note include:

- The exact precision and positioning of the numbers on each line is unimportant.
- Lines can be annotated after the numerical values (e.g. 140.0 ! m2/g).
- pH values of adjacent data points cannot be identical. ProtoFit will produce an error message indicating where identical pH values are detected.
- Values in exponential notation are accepted (e.g. 1.40E+2).

Also note that values of Z and T should not be changed from 1.0 and 298.15, respectively.

3.2 Using ProtoFit-GUI

3.2.1 Installation

Installation instructions are found in the distribution media, typically in a file named *install.txt*. For Microsoft Windows, installation is performed simply by running the setup program and following the on-screen instructions. For instructions on how to compile ProtoFit from source code, see the Appendix “Compiling ProtoFit.”

3.2.2 Invoking ProtoFit GUI

The method of invoking ProtoFit GUI depends on the operating system used. In the Microsoft Windows version, *ProtoFit GUI* is typically started by double-clicking on the ProtoFit icon.

On Unix-like systems (e.g. Linux, Mac OS X), ProtoFit GUI is invoked by 1) cd to the ProtoFit directory; then 2) invoke the shell script from the command line (“\$” indicates command prompt):

```
$/run_protokit_gui.sh
```

Table 3.1. Dataset Input File Format.

<i>line</i>	<i>variables</i>	<i>description</i>
1	none	title or description for dataset
2	nob (integer)	number of data points in dataset
3	init_vol (real)	initial solution volume (L)
4	Cs (real)	background electrolyte concentration (mol/L)
5	Z (real)	valence of background electrolyte ions (note: the program is currently not designed to handle values other than one)
6	T (real)	absolute temperature (K) (note: the program is currently not designed to handle values other than 298.15)
7	adsorbent_mass (real)	mass of adsorbent (g)
8	SSA (real)	specific surface area of adsorbent (m ² /g)
9	Nacid (real)	normality of acid (eq/L) (note: use a negative value if using a base)
10 to (9+nob)	titr_pH, titr_mL (real, real)	solution pH and volume of acid or base (mL) added at each titration step. The first value of “titr_mL” must equal zero.

3.2.3 Working with ProtoFit GUI

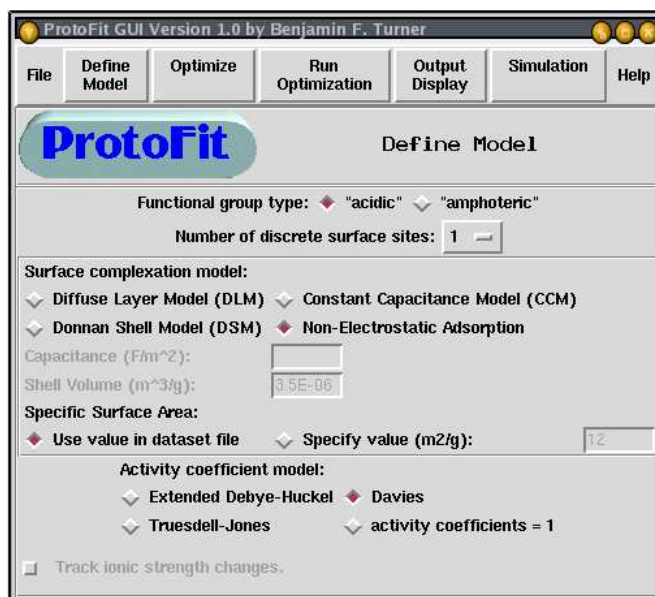
The first screen to appear is the “Define model” screen. Settings saved from a previous session (including all model, optimization and simulation parameters) will automatically be loaded into ProtoFit GUI. To save your settings so that they will be automatically loaded next time you launch ProtoFit, select *File: Save Settings* from the menu bar.

Also note that the optimization output and model comparison chart (see following sections) will not automatically update until you run the optimization or simulation routines, respectively.

3.2.4 Defining the model

The first screen to appear in the ProtoFit GUI is the “Define Model” screen. This screen can be accessed at any time by clicking the “Define Model” button on the menu bar (see [Figure 3.2.1](#)). This screen allows you to define the model that ProtoFit will either optimize or simulate.

Figure 3.2.1. “Define Model” window.



Surface type Select “amphoteric” if your surface functional groups obey [Equation 1.1](#), and “acidic” if they obey [Equation 1.2](#).

Number of discrete surface sites Select the number of surface sites for the model.

Surface complexation model Choose a model from among the choices provided. If you choose “Constant Capacitance Model (CCM),” then you will be given the opportunity to specify the capacitance.

Capacitance Enter value if CCM is used.

Shell Volume Enter value if DSM is used.

Specific Surface Area You are given the choice of using the value provided in the dataset input file (see [Section 3.1](#)), or specifying a value.

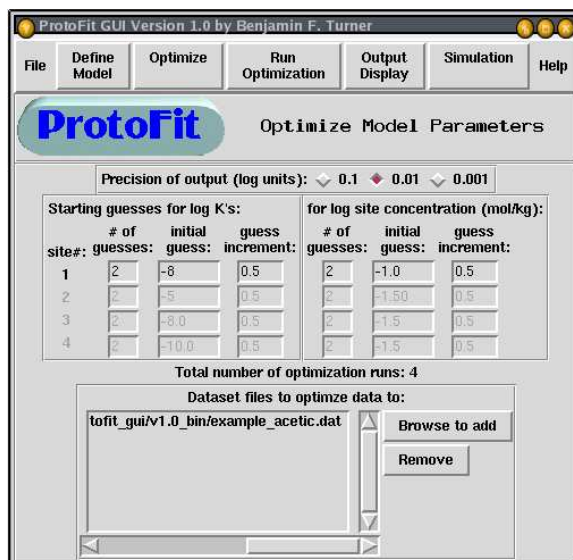
Activity coefficient model Choose a model from among the choices.

Track ionic strength changes Check this box if you want ProtoFit to track changes in ionic strength as the titration progresses (recommended). Do not check if you want to assume ionic strength is constant.

3.2.5 Setting the optimization parameters

Before running ProtoFit in optimization mode, it is necessary to set the optimization parameters in the “Optimize” screen (Figure 3.2.2). The “Optimize” screen is accessed by clicking the “Optimize” button on the menu bar.

Figure 3.2.2. “Optimize” window.



Precision of output Output from the optimization routine is expressed in log units. Select the level of precision in the output values. Higher precision values will take longer to compute.

Starting guesses Since optimized parameter sets for complex models tend to be non-unique, ProtoFit is usually set to iterate over several sets of initial guesses. In this control, you are asked to provide parameters controlling how ProtoFit will iterate over initial guesses for *log K* and *log site concentration* values. If your site type is “amphoteric,” you must provide values for both *log K₁* and *log K₂*. You are asked to provide the following three values for each of the optimized model parameters.

- *# of guesses* This is the total number of different initial guesses for the given parameter.
- *initial guess* This is the first starting guess in the series of starting guesses the routine will iterate over.
- *guess increment* This is the increment in the starting guess that the routine will use in its iterations. This value is added to “initial guess” at each increment.

For example, if for a given *log K* value, values of “# of guesses,” “initial guess,” and “guess increment” are 3, -4.5, and 1.0, respectively, then the routine will iterate over three starting

guesses for the $\log K$ value: -4.5, -3.5, and -2.5.

The total number of optimization runs will equal the product of all the “# of guesses” values. For example, if there is a 2-site model, and all of the “# of guesses” values are equal to 3, then there would be $3*3*3*3 = 3^4 = 81$ optimization runs.

Dataset files to optimize data to: ProtoFit requires one or more dataset files in order to run an optimization. The box on the left contains a list of files that ProtoFit will simultaneously optimize the model to. The entry box and buttons on the right are tools that allow you to edit the list.

To remove file names from the list, select one or more file names and click the “Remove” button.

To add a file name to the list, click the “Browse to Add” button. A dialog window will appear allowing you to browse for the file.

3.2.6 Running the optimization and viewing output

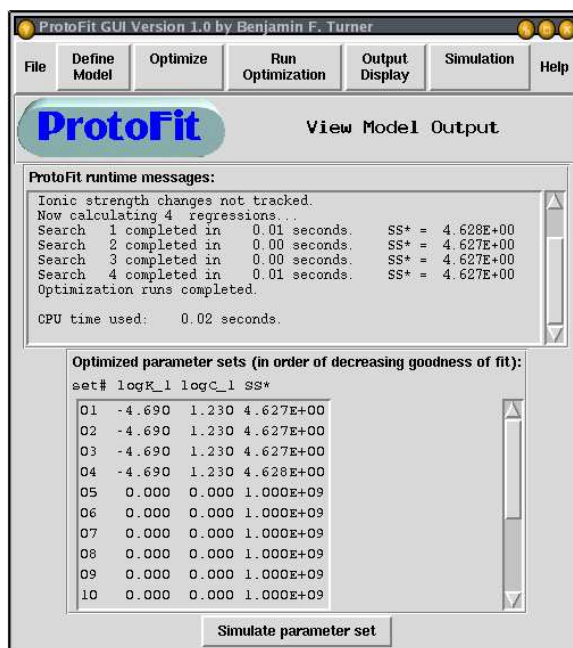
Once the model definition and optimization parameters are set to your satisfaction, the optimization routine is initiated by clicking the “Run Optimization” button. This button takes you to the “Output Display” screen (Figure 3.2.3) and starts the routine. The contents of this screen can be reviewed at a later time by clicking on the “Output Display” button.

ProtoFit runtime messages As the optimization routine runs, messages generated by the routine will scroll by in this window. This will keep you in touch with the progress of the optimization, and will be a place where error messages generated by the routine will appear.

Optimized parameter sets After the optimization routine finishes, this box will be refreshed with the latest optimization results. As the optimization routine was running, it kept track of the parameter sets that best fit the dataset files optimized to. The list that appears in this box is sorting in order of decreasing goodness of fit. The goodness of fit is reflected in the SS^* value (smaller is better). The optimized values of $\log K$ and $\log C$ for each site are display according to the header.

Simulate parameter set You will likely want to visualize how your optimized model compares to your data. If you select a parameter set from the above list and press the “Simulate parameter set” button, the parameter set will be automatically entered into the “Simulation” screen.

Figure 3.2.3. “Output Display” window.



3.2.7 Running a simulation

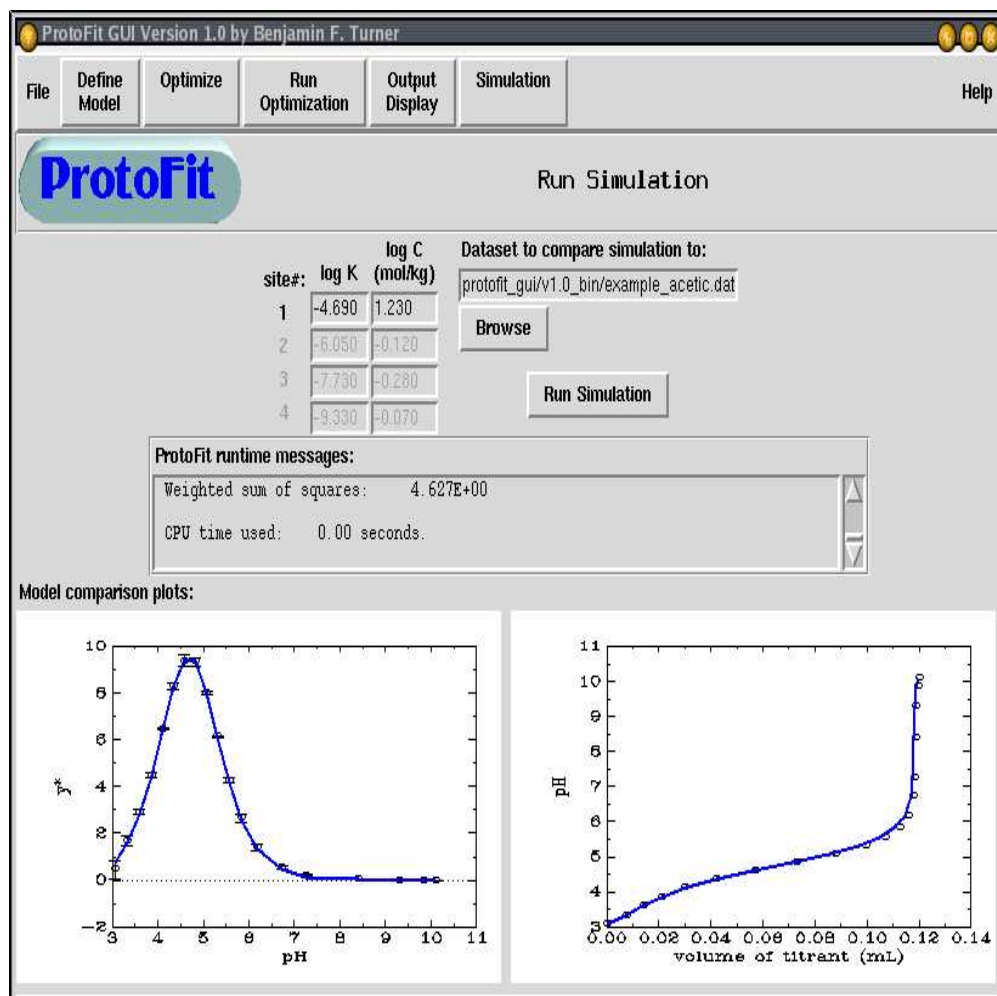
The ProtoFit GUI provides tools for simulating and visualizing your model. The “Simulation” screen (Figure 3.2.4) is accessed by clicking the “Simulation” button on the menu bar, or by clicking the “Simulate parameter set” button in the “Output Display” screen.

Model Parameters This is an area where you enter the $\log K$ (or $\log K_1$ and $\log K_2$) and $\log C$ values of the model you wish to simulate. If you entered the “Simulation” screen by clicking the “Simulate parameter set” button in the “Output Display” screen, the values from the parameter set you selected will be automatically entered.

Dataset to compare model to Here is where you enter the filename of the dataset to compare the model to. The file name can be entered in the entry box by one of two ways: 1) by typing in the name directly; or 2) by browsing for the file after clicking the “Browse” button.

Run Simulation Clicking this button initiates the simulation routine and causes the graph below to be updated after the routine is completed.

Figure 3.2.4. “Simulation” window.



ProtoFit runtime messages As the simulation routine runs, messages generated by the routine will scroll by in this window. This will keep you in touch with the progress of the routine, and will be a place where error messages generated by the routine will appear.

Model comparison plots One plot on the left compares the dataset derivative function (black with error bars) with the model derivative function (blue). The error bars are the square root of the variance of y^* , $V(y^*)$ (see Equation 2.1.12). The plot on the right compares the raw titration data to a simulated model. The plots are only updated after the “Run Simulation” button is clicked and the simulation routine is completed.

3.3 Using ProtoFit from the command line

There may be instances where it is advantageous to use ProtoFit from the command line rather than the GUI. The program file run from the command line (“`protofit.exe`”) is the same program called by the GUI for the optimization and simulation routines. As with the GUI, the user must create a dataset input file (see [Section 3.1](#)) for each dataset to be used.

In addition, the user must create an input file containing the controls for the program. The GUI automatically creates such a file, named “`protofit_main.inp`”. This file may be edited using a text editor. The format of the main input file is documented in [Appendix 1](#). As with the dataset files, the exact precision and positioning of the numbers on each line is unimportant.

Once the input files are ready, ProtoFit is invoked by the following command (assuming the current working directory contains the file “`protofit.exe`”):

Unix, Linux, Mac OS X, etc. (“\$” indicates command prompt):

```
$/protofit.exe < protofit_main.inp
```

MS Windows (“>” indicates command prompt):

```
>protofit.exe < protofit_main.inp
```

Output from the program is written to the file “`protofit_results.txt`”.

3.4 Tips and Tricks

Calculating speciation constants for organic acids

ProtoFit is primarily designed for calculating speciation constants for surfaces. However, it may also be used to calculate speciation constants for organic acids. Create the dataset file as you would for a surface, but use any non-zero value for the SSA. Choose “Non-electrostatic model” for the surface complexation model. Site concentrations generated by ProtoFit will be in units of mol per kg of acid.

Inserting the comparison plots into documents

The model comparison plots are intended for “quick and dirty” visualization, but may be imported into any document that supports the PNG graphics format. The file names for the plot graphics are “`plot1.png`” and “`plot2.png`”.

Using data generated by ProtoFit

In optimization mode, ProtoFit writes the optimized parameter sets to a file called *protofit_results.txt*.

In simulation mode, ProtoFit writes two output files: *comp_curves.csv*, which contains the values of y^* , the square root of $V(y^*)$, f^* , σ and Ψ ; and *raw_curves.csv*, which contains the

data for comparing the raw titration to the simulated titration. Both files are designed to be readable by spreadsheet applications.

3.5 Example Datasets

Three sets of example data are distributed with ProtoFit: an acetic acid titration, a bacteria titration, and a set of rutile titrations (at three different ionic strengths).

The acetic acid dataset is described well by a one-site “acidic” NEM. The bacteria data set exhibits more complex behavior, but can be empirically described reasonably well by either a three-site “acidic” CCM (with capacitance value of about 1.8) or a four-site “acidic” NEM.

The rutile data set can be described reasonably by a one-site “amphoteric” DLM. The three rutile titrations can be used together to simultaneously optimize a model that reasonably describes the ionic strength dependence of the surface's protonation behavior.

References

Hook, R., and Jeeves, T. A., 1960. "Direct search" solution of numerical and statistical problems. *J. Assoc. Comput. Mach.*, 8: 212-229.

Langmuir, D. L., 1997. *Aqueous Environmental Geochemistry*. Prentice Hall, Upper Saddle River, New Jersey, 600 p.

Neter, J., Wasserman, W., and Kutner, M. H., 1985. *Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Designs*. Irwin, Homewood, Illinois, 1127 p.

Stumm, W., 1992. *Chemistry of the Solid-Water Interface*. Wiley-Interscience, New York, 428 p.

Westall, J. C., FITEQL. A Computer Program for Determination of Chemical Equilibrium Constants, Version 2.1; Report 82-02; Department of Chemistry, Oregon State University: Corvallis, OR, (1982).

Westall, J. C., Jones, J. D., Turner, G. D., and Zachara, J. M., 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 (1995).

Appendix 1. Main Input File Format

<i>line</i>	<i>variables</i>	<i>description</i>
1	np (int)	Number of parameters to optimize
2 to 14	imax(i), inc(i), init(i) (int, real, real)	<p><i>imax(i)</i>: the number of starting guesses for parameter <i>i</i>; <i>inc(i)</i>: the increment in the starting guess for parameter <i>i</i>; <i>init(i)</i>: the initial starting guess for parameter <i>i</i>.</p> <p>Example: 3 1.0 -4.6 Three starting guesses for <i>i</i> will be used: -4.6, -3.6 and -2.6. The total number of starting guesses will be $\prod imax(i)$, for <i>i</i>=1 to <i>np</i> Lines 2, 3, and 4 are for the parameters K1, K2 and log C for the first surface site. Lines 5, 6, and 7 are for the second surface site, etc. The last line in this set (line 14) is reserved for the adjustable surface parameter (see <i>isurfparam</i> below)</p>
15	nsites (int)	number of surface functional groups
16	itype (int)	Functional group type: 0 = "acidic", 1 = "amphoteric"
17	imod (int)	Adsorption model to be used: 1 = Guoy-Chapman Double Layer Model (DLM) 2 = Constant Capacitance Model (CCM) 3 = Donnan Shell Model (DSM) 4 = Non-electrostatic adsorption
18	capacitance (real)	capacitance value for CCM (Farad / m ²)
19	iact (int)	Specifies model for calculating activity coefficients 1 = Extended Debye-Huckel 2 = Davies 3 = Truesdell-Jones 4 = activity coefficients = 1
20	isurfparam	Specifies whether a surface parameter (e.g. specific surface area, specific shell volume) should be optimized 0 = do not optimize surface parameter 1 = optimize surface parameter 2 = specify a value (use value for <i>init</i> on line 10)

<i>line</i>	<i>variables</i>	<i>description</i>
21	iIS (int)	Specifies whether or not to track ionic strength changes 0 = do not 1 = do
22	nbest (int)	Number of best-fitting parameter sets to sort and print upon completion of program
23	nwex (int)	Specifies amount of output displayed to screen by parameter optimization subroutine: 0 = no output (recommended) 1 = writes copious output indicating progress of optimizing routine (entertaining but not recommended)
24	icompare, isetcompare (int, int)	<i>icompare</i> specifies whether or not program will be run in “compare mode:” 0 = “optimization-mode” 1 = “simulation-mode:” program will run a simulation of the dataset specified by <i>isetcompare</i> using the parameter values specified in the variable <i>init</i> above. <i>isetcompare</i> specifies the number of the dataset (in list below line 27) to be used in the simulation
25	step1 (real)	Specifies the “coarse adjust” step size for adjusting parameters. This value should be about an order of magnitude larger than <i>step2</i> .
26	step2 (real)	Specifies the “fine adjust” step size for adjusting parameters. This value determines the precision of your optimized parameters.
27	nsets (int)	Specifies the number of datasets to read. In optimization-mode, the parameters will be optimized simultaneously to all datasets read by the program.
28 to (28+nsets)	dat_filenm(i) (char)	The filenames of each dataset to be read

Appendix 2. Compiling ProtoFit and ProtoFit GUI from source code

Most users will install ProtoFit from an executable (i.e. ready to run) distribution and will not want to modify the source code. This appendix is for skilled users who want to 1) set up ProtoFit without an executable distribution; or 2) modify the source code for either ProtoFit or ProtoFit GUI. The source file for ProtoFit is located in the *src* subdirectory.

ProtoFit

Compiling the ProtoFit program (used for both the GUI and command line methods) requires a *Fortran 90/95* or *F* compiler. A typical command for compiling the program under a Unix-like environment (assuming source file is named “*proffit1.0.f90*”) would be

```
$ f90 -O proffit1.0.f90 -o proffit.exe
```

The executable should then be copied to the main ProtoFit directory.

ProtoFit GUI

The source file for the ProtoFit GUI has the extension *tcl*. ProtoFit GUI is written in the language Tcl/Tk. The source file does not need to be compiled after modification, since Tcl/Tk is an interpreted language.

Setting up the Tcl/Tk interpreter and Gnu graph application

If an executable distribution of ProtoFit was not obtained, the supporting software for ProtoFit GUI must be set up. Specifically, this includes:

- Tcl/Tk
- The BLT extension for Tcl/Tk
- The Gnu graph application (part of the Gnu plotutils)
- Gnu graph requires zlib and libpng

The Linux executable distribution contains the Tcl/Tk interpreter with the BLT extension, but requires Gnu graph to be installed by the user. All the supporting software is *Open Source* software and freely available on the Internet. Follow the instructions supplied with each of the above software packages.