

# Manual for PSEQUAD

Update 5.31, Last revision: June 28, 2021

László Zékány

*University of Debrecen, Department of Inorganic and Analytical Chemistry  
Debrecen, Egyetem tér 1., Hungary H-4032*

István Nagypál<sup>†</sup> and Gábor Peintler

*University of Szeged, Department of Physical Chemistry  
Szeged, Rerrich Béla tér 1., Hungary H-6700*

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Brief History . . . . .	2
<b>2</b>	<b>Computational Bases</b>	<b>3</b>
2.1	Base Equations . . . . .	3
2.2	Process of Calculations . . . . .	6
2.2.1	Calculation of the Free Concentrations . . . . .	6
2.2.2	Refinement of the Formation Constants and/or the Molar Absorptivities	7
2.3	Model-selection . . . . .	8
2.4	Estimation of Parameter Correlation . . . . .	9
2.5	Program Limits . . . . .	9
<b>3</b>	<b>Creating Input Files</b>	<b>10</b>
3.1	Examples . . . . .	11
3.2	Type of Lines in Input Files . . . . .	13
3.3	Changing the Format of Lines . . . . .	14
3.4	Changing Default Parameter Values . . . . .	14
3.5	Input of Potentiometric and/or Photometric Data . . . . .	16
3.5.1	Simulation of Data . . . . .	19
3.6	Data Processing Options and Instructions . . . . .	21
3.7	Model-selection . . . . .	31
<b>4</b>	<b>Output of PSEQUAD</b>	<b>32</b>

# 1 Introduction

PSEQUAD is a comprehensive program for the evaluation of **P**otentiometric and/or **S**pectrophotometric **E**quilibrium data **U**sing **A**nalytical **D**erivatives. Potentiometry and spectrophotometry (and other methods which can be described by means of analogous mathematical formulas) are the most popular experimental techniques for equilibrium studies. This program can handle both types of experimental data either separately or simultaneously. There are more than 220 citations concerning to the use of PSEQUAD until 1999.

This manual explains the most important computational bases and contains reference information for creating PSEQUAD input files. The description of the previous mainframe version can be found in the book mentioned in PSEQUAD.HLP file.

## 1.1 Brief History

The first mainframe version of PSEQUAD was developed in 1983 by László Zékány and István Nagypál.

**Update 1:** Tape handling was added in April 1987.

**Update 2:** The program was modified for working under CMS in June 1987.

**Update 3:** The double precision version was introduced in December 1987. It was the last mainframe version.

**Update 4:** The first PC version was developed by Gábor Peintler in April 1990. Microsoft Fortran v5.1 compiler was used.

**Update 5:** The first revision of the PC version was carried out in April 2000 (subsequent revisions are detailed in the source file). The compiler was changed to GNU Fortran (versions 3.4.5–10.3.0 at the time of writing, depending on the operating system). More important improvements were introduced:

- The program limits have been increased significantly (*see: Section 2.5*).
- The length of input and output lines is not limited to 80 and 120 characters, respectively. The upper limit is now 16380 characters for both input and output files. The structure of any input line can be changed by the user through the new FORMATA and FORMATB keys. Moreover, a new form of output (FREEOUT key) is possible in which long lines are not broken beyond the 120<sup>th</sup> character.
- The carriage control characters are removed from the output files after interpreting them and modifying the output accordingly.

- The program does not longer suppose that the most integer values (both in input and output) are not larger than 99. This limit has been increased up to 999.
- Both 32 and 64 bit versions are available under Windows and Linux (only 32 bit version under DOS, of course). The 32 bit version of Windows executable is compatible even with Windows 98 second edition.

## 2 Computational Bases

Potentiometry and spectrophotometry<sup>1</sup> are the most frequently used experimental methods to study equilibrium systems in solutions; these methods may be used in many different types of experimental arrangements.

### 2.1 Base Equations

The equilibrium system can be described through the mass-balance equations as

$$C_t = \sum_{j=1}^n \alpha_{jt} [S_j] = \sum_{j=1}^n \alpha_{jt} \beta_j \prod_{i=1}^k [c_i]^{\alpha_{ji}} \quad (t = 1 \dots k) \quad (1)$$

where

$C_t$  is the total concentration of the  $t^{\text{th}}$  component,

$n$  is the number of species in the system, including the components,

$S_j$  is the  $j^{\text{th}}$  species present in the system,

$k$  is the number of components in the system,

$[c_i]$  is the equilibrium (free) concentration of the  $i^{\text{th}}$  component,

$\beta_j$  is the formation constant of the  $j^{\text{th}}$  species,<sup>2</sup> ( $\beta_j = [S_j] / \prod_{i=1}^k [c_i]^{\alpha_{ji}}$ ) and

$\alpha_{ji}$  are the stoichiometric numbers, giving the number of the  $i^{\text{th}}$  component in the  $j^{\text{th}}$  species.  $\alpha$  is called *composition matrix*.

The stoichiometric numbers are arranged in  $k$  columns and  $n$  rows forming the composition or  $\alpha$  matrix of the system. It is expedient to arrange the species in a special order, such that the first  $k$  rows of the composition matrix form a unit matrix (*i.e.*, the components occupy the first  $k$  rows). The stoichiometric numbers are usually positive integers, apart from those relating to

<sup>1</sup>Analogous methods (*e.g.*, conductometry) are also supported by PSEQUAD.

<sup>2</sup>According to the definition, the formation constants of the components are unity.

the component taking part in the self-dissociation of the solvent. The change of the solvent concentration is normally disregarded in solution equilibrium studies. Only one of the ions taking part in the self-dissociation process can be considered as a component. Therefore, the stoichiometric numbers of a component are negative integers for species containing the counter-ion or for the species having deficit in the selected component. This situation most frequently occurs in aqueous solutions, where, with the hydrogen ion taken as a component, the stoichiometric numbers belonging to the  $\text{OH}^-$  ion are  $-1$  for the hydrogen ion and zero for the other components. Thus it follows from this general definition that the formation constant of the  $\text{OH}^-$  ion is  $K_w = [\text{H}^+][\text{OH}^-]$ . The appropriate stoichiometric numbers for the species containing  $\text{OH}^-$  ion—or having fewer protons than that form of the ligand selected to be a component—are also negative integers for protons.

It follows from the above definition that the total concentration of the proton in aqueous solution may be negative; this special case has some consequences on the solution of the mass-balance equations, as we shall see later.

The experiments are frequently carried out as titrations, which means that the total concentration of one of the components is changing step by step over a relatively wide range, while the total concentrations of the other components are changing owing to dilution.<sup>3</sup> In this situation the volume of the titrant, or the total concentration of one of the components, may be regarded as an experimental datum, denoted by  $X_1^V$  in the following.

For potentiometry, the directly measured experimental data ( $X_l^P$ ) can be expressed by the free concentration of one or more of the components:

$$X_l^P = A_l + M_l \log[c_l] \quad (l = 2 \dots m) \quad (2)$$

where

$m - 1$  is the number of potentials measured in the system ( $m - 1 \leq k$ ),

$A_l$  is an additive term, for example,  $E_0$  if emf is measured, or an additive term to convert the directly measured pH into  $-\log[\text{H}^+]$  and

$M_l$  is multiplicative coefficient to be calculated from the Nernst-equation in case of emf measurements, or  $-1$  for pH measurements.

---

<sup>3</sup>Occasionally, there will be no total concentration change at all since their concentrations are the same in the two solutions being mixed. It can especially be advantageous in the case of photometric titrations.

For spectrophotometry, the following general relation is valid:

$$X_l^A = \sum_{j=1}^n \epsilon_{jl} \beta_j \prod_{i=1}^k [c_i]^{\alpha_{ji}} \quad (l = m + 1 \dots p) \quad (3)$$

where

$p - m$  is the number of wavelengths studied,

$X_l^A$  is the measured absorbance at the  $l^{\text{th}}$  wavelength and

$\epsilon_{jl}$  is the molar absorptivity of the  $j^{\text{th}}$  species at the  $l^{\text{th}}$  wavelength.

PSEQUAD solves Eq. (1) for the unknown free concentrations and obtains the unknown formation constants and/or molar absorbancies by minimizing the  $F$  function:

$$\begin{aligned} F &= \sum_{q=1}^{n_d} F_q = \\ &= \sum_{q=1}^{n_d} \sum_{i=1}^{r_q} \left( w_1 (\Delta X_1^V)^2 + \sum_{l=2}^m w_l (\Delta X_l^P)^2 + w_A \sum_{l=m+1}^p (\Delta X_l^A)^2 \right) \end{aligned} \quad (4)$$

where

$w_1$  is the weighting factor of the volume of the titrant or total concentrations,

$w_l$  is the weighting factor for the  $l^{\text{th}}$  potential measurements,

$w_A$  is the weighting factor for the absorbance measurements and

$F_q$  is any function that is “correctly composed” from the experimental data. The meaning of “correctly composed”, and the definition of  $n_d$  and  $r_q$  is detailed in the next paragraph.

The program has the capacity to evaluate simultaneously different types of measurements carried out in solutions having varying compositions, or even in groups of solutions in which the number of the components differ from one subset to another. For example, let us consider a four-component system including the proton (H), two ligands (A and B) and a metal ion (M). Examples for valid subsets are H/A, H/B, H/A/M, H/B/M and H/A/B/M. Titration curves may be evaluated simultaneously for the  $pK_a$ 's of the ligands and for the formation constants of  $M_xA_yH_z$ ,  $M_xB_wH_z$  and  $M_xA_yB_wH_z$  complexes. In this way the effect of error accumulation, which may occur if the titrations are evaluated step by step, can be avoided. If, for example, these

measurements are supplemented with spectrophotometric measurements in the same or separate samples, these may also be evaluated simultaneously. This possibility is implied by Eq. (4), where

$n_d$  is the number of sets of measurements derived from different types of primary experimental data and

$r_q$  is the number of experimental points in the  $q^{\text{th}}$  set of measurements.

## 2.2 Process of Calculations

There are two main steps in the calculations: the solution of Eq. (1) for the unknown free concentrations and the refinement of the formation constants and/or molar absorptivities.

### 2.2.1 Calculation of the Free Concentrations

The calculation of the unknown free concentration is based on the standard Newton-Raphson procedure, by solving the equations

$$(C_i^{\text{calc}} - C_i^{\text{exp}}) = 0 \quad (5)$$

for  $C_i^{\text{calc}}$ .

The free concentrations are calculated on a  $\log[c_i]$  scale, thereby preventing the occurrence of negative concentrations during the course of the iterative procedure. The starting  $[c_i]$  values are  $C_i/2$  if  $C_i \geq C_{\min}^0$  or  $C_{\min}^0$  if  $C_i \leq C_{\min}^0$ .<sup>4</sup> This restriction is necessary because, as we have seen, the total concentration for the protons may be negative. There is a step-size control in the algorithm that limits a change in  $[c_i]$  to two orders of magnitude;<sup>5</sup> the step sizes of the other  $[c_i]$  values are proportionally decreased on a logarithmic scale. The iteration terminates when

$$\left| \frac{C_i^{\text{exp}} - C_i^{\text{calc}}}{C_i^{\text{exp}}} \right| \leq \text{FNRMAX}$$

for all of the concentrations.<sup>6</sup> If the standard Newton-Raphson method fails to converge, then the equation

$$\ln \frac{C_i^{\text{calc}}}{C_i^{\text{exp}}} = 0 \quad (6)$$

<sup>4</sup>The default value for  $C_{\min}^0$  is  $10^{-6}$  but it can be changed by field f3 of 91 SUPERVIS line (page 15).

<sup>5</sup>This maximum value for the change can be modified by field f3 of 92 SUPERVIS line (page 15).

<sup>6</sup>The default value for FNRMAX is  $5 \times 10^{-4}$  but it can be changed by field f1 of 92 SUPERVIS line (page 15).

is solved for  $-\log[c_i]$  by the Newton-Raphson method. Four orders of magnitude of change for  $[c_i]$  are allowed in this iteration.<sup>7</sup>

As we have seen, the total concentration for the hydrogen ion may be negative; thus Eq. (6) is modified for the hydrogen ion as follows:

$$\ln \left( \frac{C_{+H}^{\text{calc}} + \frac{\text{sign}(C_H^{\text{exp}}) - 1}{2} C_H^{\text{exp}}}{C_{-H}^{\text{calc}} + \frac{\text{sign}(C_H^{\text{exp}}) + 1}{2} C_H^{\text{exp}}} \right) = 0 \quad (7)$$

where  $C_{-H}^{\text{calc}}$  and  $C_{+H}^{\text{calc}}$  are the calculated total hydrogen concentration with negative and positive stoichiometric numbers, respectively. Iteration terminates when the absolute values of the left sides of Eqs. (6,7) are less than FHMAX.<sup>8</sup>

The above procedure is used only for the first iterative cycle for the formation constants and/or molar absorptivities. In the second and subsequent cycles, the previous  $[c_i]$  values are used as  $[c_i]_0$  for the current cycle. For the evaluation of titration curves, the  $[c_i]_0$  values defined above are used only at the first titration point. For the second and subsequent titration points, they are calculated by extrapolation from the previous point using the method of analytical derivatives of the implicit function systems (*continuation method*).

### 2.2.2 Refinement of the Formation Constants and/or the Molar Absorptivities

The Gauss-Newton method is used for the refinement of the parameters by minimizing any function correctly composed from Eq. (4). However, since some of the primary data included in Eq. (4) are interdependent, inconclusive results may follow. Allowed combinations of the  $X_l$  ( $l = 1 \dots p$ ) values are as follows.

- (a) If  $X_1^V$  (volume of the titrant or the total concentration of the first component) is not included in the experimental data to be fitted, then all possible combinations of the remaining  $X_l$  ( $l = 2 \dots p$ ) values are allowed and all the  $C_1 \dots C_k$  values are fixed data.
- (b) If some of the measured potentials are fixed values, then all potential values must be arranged so that the fixed potential(s) are the first  $X_l^P$  ( $l = 2 \dots m_1$ ) values, and the potential(s) to be fitted are the remaining  $X_l^P$  ( $l = m_1 + 1 \dots m$ ) values. In this instance the volume of the titrant

<sup>7</sup>This maximum value for the change can be modified by field f4 of 92 SUPERVIS line (page 15).

<sup>8</sup>The default value for FHMAX is  $5 \times 10^{-4}$  but it can be changed by field f2 of 92 SUPERVIS line (page 15).



containing the first component, or the  $C_1$  values, as well as the other  $X_l$  values ( $l = m_1 + 1 \dots p$ ) may be optionally combined in Eq. (4). The  $C_2 \dots C_{m_1}$  data are then not used either as fixed parameters, or as variables to be fitted. This enables the processing of photometric measurements at fixed pH values.

- (c) It may be expedient for pH-titrations (or potentiometric one, of course), to fit the parameters by a combined minimization of the volume–pH data (orthogonal regression<sup>9</sup>). The ratio of the variance of the volume and pH will be required for this situation.

The unknown formation constants are nonlinear, whereas the molar absorptivities are linear parameters; thus the equations relating to the molar absorptivities (Beer's law) are solved first, and the  $\beta$  values are refined subsequently.

The molar absorptivities are calculated directly, no restriction being imposed on their values. There is, however, a step-size control for the refinement of the formation constants. The allowed maximum step is one logarithmic unit for one cycle.<sup>10</sup>

Refinement is complete when the change in the error square sum is either less than 0.005%<sup>11</sup> or the maximum number of iterations is exceeded. If the error square sum in a subsequent iteration cycle is found to be increasing, then half the calculated step sizes is used for the shift of the formation constants in that cycle.

## 2.3 Model-selection

The model-selection is an algorithm to change the used chemical model by adding and/or removing species systematically into the subsequent calculations. Originally, model-selection was included in the mainframe version of PSEQUAD. It made possible to try many models in a single run. Due to the short computing time nowadays, the use of this algorithm is not really useful with the PC-version. Although model-selection is possible with each version of PSEQUAD, it is no longer supported and only briefly explained (see: Subsection 3.7).<sup>12</sup>

<sup>9</sup>Yu. V. Linnik, *Method of Least Squares and Principles of the Theory of observations*, Pergamon Press, Oxford (1961).

<sup>10</sup>This default value can be changed by the f2 field of 91 SUPERVIS line (page 15).

<sup>11</sup>This default value can be changed by the f1 field of 91 SUPERVIS line (page 15).

<sup>12</sup>A fairly detailed description of model-selection can be found in the book *Computational Methods for the Determination of Formation Constants*, Ed.: David J. Leggett, Plenum Press, New York, 1985.



## 2.4 Estimation of Parameter Correlation

Several numbers of statistical parameters are calculated by the program. Since most of the formulas are given in the output, they are not described here. However, the meaning of the partial, multiple, and total correlation coefficients will be discussed here. These are calculated from the elements of the matrix

$$\mathcal{B} = \mathcal{J}^T \cdot \mathcal{W} \cdot \mathcal{J} \quad (8)$$

where  $\mathcal{J}$  is the Jacobian-,  $\mathcal{J}^T$  is its transpose, and  $\mathcal{W}$  is the diagonal weighting matrix used in the Gauss-Newton method. The *partial correlation coefficients*,  $r_{ij}$ , give the measure of interdependence between two constants  $\beta_i$  and  $\beta_j$  assuming that the other constants have fixed values:

$$r_{ij} = \frac{-\mathcal{B}_{ij}}{\sqrt{\mathcal{B}_{ii}\mathcal{B}_{jj}}} \quad (9)$$

The *total correlation coefficients*,  $S_{ij}$  also provide a measure of interdependence between two constants, the other constants being regarded as fitted parameters:

$$S_{ij} = \frac{-C_{ij}}{\sqrt{C_{ii}C_{jj}}} \quad (10)$$

The *multiple correlation coefficients*,  $R_i$  give the measure of the independence of a given constant from that of all the others:

$$R_i = \sqrt{1 - \frac{1}{\mathcal{B}_{ii}C_{ii}}} \quad (11)$$

The appropriate elements of  $C = \mathcal{B}^{-1}$  are denoted by  $C_{ij}$  in Eqs. (10, 11).

Each of these correlation coefficients may take values between zero and  $\pm 1$ . Zero implies the total independence of the species,  $+1$  or  $-1$  means a complete correlation, and consequently the two species in question should not be refined simultaneously. The correlation coefficients are arranged in matrix form in the output. The diagonal elements contain the multiple correlation coefficients, the upper triangle contains the partial, while the lower triangle contains the total correlation coefficients.

## 2.5 Program Limits

There are only a few limitations to the program:

- (a) A maximum of five different types of potentials can simultaneously be measured.
- (b) Most of the calculations are carried through in two common blocks. Their sizes are 30000 for the integers and 200000 for the real variables, in the present form of the program. These sizes may easily be changed as shown in the source code.
- (c) Several parameters (working precision, convergence criteria for the Newton-Raphson or the Gauss-Newton procedure, step sizes, etc.), given in the source code as definite numbers, may be changed via the input data. Moreover, the program may be used to simulate experimental data, or to calculate the concentration distribution of species in a known system. For this latter option the maximum number of Gauss-Newton iterations and the allowed step size of the formation constants are set to zero.
- (d) As can be seen, the program is very flexible and capable of handling almost all potentiometric and/or spectrophotometric types of measurements. As we have seen, however, no numerical constraints are applied to the molar absorptivities; thus negative values may be calculated. This also increases the run time, but it is still considerably lower than numerical differentiation would be employed.

### 3 Creating Input Files

PSEQUAD is a very flexible program that can handle pH-metric, potentiometric, and photometric data separately or in any combination; refines all or part of the equilibrium model; and performs data conversion and various types of data weighting and statistical analysis. As a consequence of it, the structure of an input file can be complicated. The most important fact the user must know that the program is written in standard FORTRAN 66 language, therefore, the input/output processes are exclusively based on the “field” conception of the FORTRAN’s FORMAT command. In spite of the modern programming languages—in which almost only *the sequence of the input data* is important—the field conception means that each individual input data must have exact position and length in the input file. This manual does not explain the specifications of the FORMAT statement. The user should become acquainted with them from any FORTRAN related textbook or from the official description of the standard FORTRAN 77 language.<sup>13</sup>

---

<sup>13</sup>It can also be reached online from <http://www.fortran.com/>.

### 3.1 Examples

To understand the structure of the input file, ten data files of eight examples (EX\*.DAT) are included into the PSEQUAD distribution. They illustrate almost all important aspects of the input files.

**Example 1 (EX1.DAT)** shows how to change the format of the primary experimental data and how to evaluate photometric measurements made at different pathlengths. It should be mentioned that the current version of PSEQUAD also offers more efficient ways to change the format specifications.

**Example 2 (EX2.DAT)** is the most important example because it was worked out especially for illustrating the use of PSEQUAD. It presents how to evaluate pH-metric, potentiometric and absorbance measurements (titrations). Moreover, data pertaining to  $pK_w$ ,  $pK_a$ , and  $\log \beta$  calculations may be processed within a single run of PSEQUAD. Throughout this chapter, EX2.DAT file is used as a reference example.<sup>14</sup>

EX2.DAT comprises two major sections. The first section contains the experimental data: pH, potential and absorbance data, solution compositions. The second section contains program directives and the total equilibrium model description.

The input protocol, to be described, will use as an example data derived from a study of the  $\text{Cu}^{2+}$ -NTA-proton equilibrium system. The separate experiments are as follows:

- A. Determination of  $pK_w$ , one titration, pH measured.
- B. Determination of NTA  $pK_a$ 's, one titration, pH measured.
- C. Titration of copper(II) with NTA, two titration curves, pH and pCu (potential) measured for each.
- D. Titration of copper(II) and NTA with KOH, two titrations, pH and pCu (potential) measured for each.
- E. Photometric titration of copper(II) and NTA (1:2) with KOH, one titration, 11 absorbances per spectrum, 15 spectra.
- F. Photometric titration of copper(II) with NTA (1:1.25) with KOH, one titration, 11 absorbances per spectrum, 12 spectra.

The potentiometric data have been collected by a titrimetric technique. The pH measurements were made using a Radiometer PHM-52 pH-meter and a GK-2322C combination electrode. The electrode system

<sup>14</sup>These references are typeset by serif fonts in narrower paragraphs.

was calibrated following Irving *et al.*<sup>15</sup>, the calibration constant,  $I_d$ , being found to be 0.05 for the equation  $\text{pH}_{\text{obs.}} = I_d - \log[\text{H}^+]$ .

Copper ion measurements were performed using a Radelkis OP-208 pH-meter equipped with a Radelkis OP-8303 calomel electrode and a Radiometer F3002 Cu Selectrode. The copper ion selective electrode was calibrated from data obtained from copper(II)-glycine titrations, measuring pH and emf simultaneously. The titrations were evaluated using the pH measurements and the copper electrode subsequently calibrated from emf. *vs.*  $-\log[\text{Cu}^{2+}]$  plots which provided  $E^0$  (313.8 mV) and Nernst-slope (28.935 mV).

Absorbance measurements were obtained using a Beckman ACTA M4 spectrophotometer. Solution from the titration vessel was pumped through a 1 cm pathlength flow-cell using a peristaltic pump. Absorbance values were digitized at 20 nm intervals starting at 800 nm.

The pH, pCu titrations were performed as follows: initial solutions of  $\text{Cu}(\text{NO}_3)_2$  (its value is  $5.551 \times 10^{-3}$  M, curve 1;  $4.683 \times 10^{-3}$  M, curve 2; 0.10 M  $\text{KNO}_3$ ) were titrated with a solution of  $\text{H}_3\text{NTA}$  ( $9.92 \times 10^{-3}$  M for both curves; 0.10 M  $\text{KNO}_3$ ). The data from each titration comprise Set 3 (*see: key=01 SET later*). After an appropriate amount of  $\text{H}_3\text{NTA}$  had been added the titrations were completed using 0.2324 M KOH, comprising Set 4. The data were collected in the order Set 3, Curve 1 followed by Set 4, Curve 1 followed by Set 3, Curve 2, Set 4 Curve 2. However, PSEQUAD processes the curves for Sets 3 and 4 by set.

**Examples 3 and 7** demonstrate the calculation of concentration distribution in a dozen-component system (EX3.DAT), and in a more realistic three-component system (EX7.DAT).

**Example 4 (EX4.DAT)** is a real example to show the evaluation of pH-metric experiments in the Ni(II)-aspartic acid-glycine system.

**Example 5 (EX5.DAT)** is another complicated example for the evaluation of different experimental data together and for the simulation with the calculated formation constants.

**Example 6 (EX6.DAT)** demonstrates the important aspects of the curve simulation, including pseudo-random-error generation.

**Example 8 (EX8ORG.DAT, EX8LONG.DAT and EX8SHORT.DAT)** illustrates different possibilities for arranging the input data resulting the same output file.

---

<sup>15</sup>H.M. Irving, M.G. Miles, and L.D. Pettit, A Study of some Problems in Determining the Stoichiometric Proton Dissociation Constants of Complexes By Potentiometric Titrations Using a Glass Electrode, *Anal. Chim. Acta* **38**, 475–488 (1967).

## 3.2 Type of Lines in Input Files

Only three types (formats) of data line are employed by PSEQUAD. They are as follows:

**Type A** default FORMAT: (I2, 2A4, 10I2, 5F10.0)  
 lines are used to indicate specific options or instructions for data processing. This line has four sections:

- (a) *Key* is an integer that determines the role of line and, in many instances, one or more lines that follow. The default format specifier is I2.
- (b) *Keyword* is an eight-character label, used to identify the purpose of the particular line. The keywords used in this manual and in the examples are of our own making and may be changed to suit the user's purposes. PSEQUAD uses the integer value of the key rather than the *keyword*. The default format specifier is 2A4.
- (c) *Ten Integers* (i1 i2 i3 ... i10) used for a variety of purposes. Originally, they are positioned from column 11 through column 30 and their default format specifier is 10I2. In the following examples designations such as

- i. i1 i2 i3 i4 i5 i6 i7 i8 i9 i10    or    i1 i2 ... i10
- ii. i1 i2 i3
- iii. i1    i3    i5    i7

mean that (i.) a value must be supplied for each integer; (ii.) a value is needed only for the first three integers; (iii.) a value is required only for the first, third, fifth and seventh integers.

- (d) *Five Floating point numbers* (f1 f2 f3 f4 f5) used for a variety of purposes, that are positioned from column 31 through column 80 originally (default format specifier is 5F10.0). The designation f1 f3 f5 or f1 f2 has the same implication as for the *integers* above.

**Type B** default FORMAT: (8F10.0)  
 lines are used for giving experimental data.

**Type C** default FORMAT: (20A4)  
 lines are used for descriptive titles. These lines do not influence the calculations.

The next few sections detail the possible keys, keywords and structures in input files. These sections follow that sequence of input lines which is adequate for the most problems. Alternative possibilities are indicated at the appropriate places.

### 3.3 Changing the Format of Lines

The next three keys redefine the format specifications for the lines in input/output files. Their usage are illustrated in Example 8.

#### 99 FORMATA

new format

The format concerning to the type A lines can be changed by the help of these two lines. For example, if an integer number less than  $-9$  is necessary, the use of these lines is compulsory. The `new format` represents a sequence of FORMAT specification codes within parentheses. These codes correspond to the standard FORTRAN language. The 99 line may be included before either the 10 DATAIN or 11 TASK lines (*see: later*) but the best place is the top of the input file. *It is necessary that the new specification reserves places for one integer number, one string containing 8 characters, ten integer numbers and 5 real numbers, in this order strictly!*

#### 98 FORMATB

new format

The purpose of this key is the same as that of the *key=99* but it concerns to the type B lines. The new specification must reserve field(s) for one or more real number(s).

#### 97 FREEOUT

new format

In the previous versions of PSEQUAD, the length of output lines could not be longer than 120 characters. Giving *key=97* in the top of an input file overrides this limitation. The large data structures (*e.g.*, concentration matrix) are written into the output file in two forms: (1) in the classical form by forming data up to 120-character long lines and (2) without breaking the rows of data structures. Moreover, (a) the matrices of the calculated data and the residuals (*see: page 27*) are separated, (b) the transposed matrices are also written into the output file and (c) the distribution diagram is indicated with numbers, too.

### 3.4 Changing Default Parameter Values

As indicated earlier, the values of a number of parameters—mainly related to the various least-squares algorithms employed by PSEQUAD—may be reset. This is achieved using type A lines with *key=91* or *92*. These lines may be

placed before a 10 DATAIN or 11 TASK line. They are supervisor lines, hence the keyword is SUPERVIS. The format of the lines is as follows:

91 SUPERVIS i1 i2 f1 f2 f3

- i1: The minimum number of subsequent iterations for which the convergence criterion (*see*: f1 below) should be fulfilled before terminating the Gauss-Newton (GN) algorithm. Usually i1=1 but may be set to 2, 3, etc. for ill-conditioned systems.
- i2: The maximum number of half-steps permitted for the GN algorithm in case of increasing fitting parameter.
- f1: Convergence criterion for GN algorithm.
- f2: Maximum step size in the fitted  $\log \beta$  values between two successive iterations.
- f3: The lowest total concentration value at which the starting value for  $[c_i]_0$  is  $C_i^0$  (*see*: Subsection 2.2.1).

92 SUPERVIS i1 i2 f1 f2 f3 f4 f5

- i1: The maximum number of iterations in Eq. (5) allowed for the standard Newton-Raphson (NR) algorithm (MXITNR).
- i2: The maximum number of iterations allowed for the NR algorithm minimizing on Eqs. (6,7) in Subsection 2.2.1 (MXITH).
- f1: Convergence criterion for standard NR (FNRMAX).
- f2: Convergence criterion for NR algorithm minimizing on Eqs. (6,7) (FNMAX).
- f3: Maximum step size for  $\log[c_i]$  in standard NR applied for Eq. (5) (XMAX).
- f4: Maximum step size for  $\log[c_i]$  in NR for Eqs. (6,7) (XHMAX).
- f5: Maximum step size for the continuation method (*see*: last paragraph of Subsection 2.2.1) (XCMAx).

PSEQUAD uses the following default values for each of the above parameters:

Key	i1	i2	f1	f2	f3	f4	f5
91	1	1	$5.0 \times 10^{-5}$	1.0	$1.0 \times 10^{-6}$	–	–
92	25	75	$5.0 \times 10^{-4}$	$5.0 \times 10^{-4}$	2.0	4.0	3.0

Note that even if only one parameter on 91 and 92 need be altered, all of the other parameters must be “reset” to the default values (*see*: Example 2). If all default values are acceptable then the 91 and/or 92 lines can be omitted.



93 KRAND i1 i2 i3 i4

This key initializes the pseudo-random-number generator. i1, i2, i3 and i4 must have an integer value between 0 and 99. Examples 5, 6 and field i6 of *key*=10 ENDTASK show the usefulness of this key.

### 3.5 Input of Potentiometric and/or Photometric Data

Having established the convergence criteria, explicitly or implicitly, the experimental data are now read in. Each experiment is preceded and followed by type A lines that describe the type of data to be expected.

10 DATAIN

*Key*=10 indicates that the input file reading routines are activated. It means that (1) either experimental data set(s) begin(s) (keyword DATAIN), (2) all information relating to a particular experiment has been entered (keyword ENDSET), (3) entering experimental data has been finished (keyword ENDDATA) or (4) entering model description has been finished (keyword ENDTASK). The f5 field must be empty in the first case. The other fields are not relevant.

01 SET i1 i2

*Key*=01 indicates that data-type description lines follow. For titrations, integer i1 is used to indicate the number of titration curves (pH, potentiometric, or photometric) of the same type that follow. For pointwise measurements, one data set per solution, i1 indicates the number of solutions having different total concentrations.

If i2>0, the format specification of type B lines can be changed temporarily for only the local set. In this case, a FORMAT specification must be given with a type C line after the next 10 ENDSET line (or after the description line if it follows the 10 ENDSET line, *see*: below). This temporary format specification will be valid for the first i2 experimental point(s) of the actual set. Therefore, the titration curves should consist of at least i2 points. The initial volume and/or the initial total concentrations, however, must be given applying the original format specification.

Example 1 illustrates the usage of this possibility. The *key*=98 FORMATB, however, provides a better way to change the structure of type B lines.

02 TOTAL i1 i3 i5 i7 i9 f1 f2 f3 f4 f5

*Key*=02 indicates that this is a total concentration(s) line, for titrant. The integers i1, i3, i5, i7 and i9 are column numbers in the  $\alpha$  matrix (*see*:

*Eq. (1)*). The floating point numbers  $f1 \dots f5$  are the total concentrations of the components  $i1, i3, i5, i7$  and  $i9$ , respectively, in the burette. A negative value is used for total hydrogen ion concentration when the titrant is a strong base. For more than five components, a second 02 line is used.

In Example 2, in the first set (titled by DETERMINATION OF PKW),  $i1=1$  (hydrogen ion is titrant) and  $f1=-0.2324$  (strong base, 0.2324 M).

### 03 TITR

*Key=03* implies that titration-type data are to be evaluated, *i.e.*, several data points relate to a single solution. Pointwise measurement (one datum per solution) can be processed by PSEQUAD by omitting this line.

### 04 POT $i1 \ i3 \ i5 \ i7 \ i9$

*Key=04* signifies that potentiometric measurements have been made. Integers  $i1, i3, i5, i7$  and  $i9$  have the same meaning as for *key=02* line. Thus for pH measurements (hydrogen ion—which equals to component 1—measured)  $i1=1$  and  $i3-i9$  are set to zero. If no potential measurements have been made then this line is omitted.

The sequence of column numbers is not independent of each others in *key=02* and *04*. Those component numbers must be written in the first positions ( $i1, i3, \dots$ ) of 02 line, the concentrations of which are measured potentiometrically. These components must be given in the line having *key=04*, too.

### 05 ABS $i1 \ i2$

*Key=05* indicates that absorbance data will follow. Integer  $i1$  is the serial number of the first, while  $i2$  is the serial number of the last wavelength for the absorbance data that are read in. In this way, the wavelength number increases strictly by one, *i.e.*, real wavelength values cannot be used, only their serial numbers.

### 10 ENDSET $i1$

*Key=10*, in this context, signals the completion of this particular set of type A lines. Thus for every *key=01* line there must be a *key=10* line. PSEQUAD now expects a title line (type C) if  $i1=1$ , or a type B line if  $i1=0$ .

The *key=10* line can be used in the following form too:

```
10 ENDSET  $i1 \ i3 \ i4 \ i5 \ i6 \ i7 \ i8 \ i9 \ i10$            $f2 \ f3 \ f4 \ f5$ 
```

The meaning of  $i1$  is unchanged.

If  $i3=n$  ( $n \neq 0$ ) then the  $n^{\text{th}}$  measured data in a titration curve (or in a pointwise measurement) is changed in the following way:

$$X'_n = (X_n + f2) \times f3 \quad (12)$$

where  $X_n$  is the  $n^{\text{th}}$  measured data and  $X'_n$  is the modified data. If  $i4=m$  ( $m \neq 0$ ) then Eq. (12) is applied to all data between the  $n^{\text{th}}$  and the  $m^{\text{th}}$  measured points in every type B line. The meanings of  $i5$  and  $i6$  correspond to  $i3$  and  $i4$  so two ranges can be given for modifying the experimental data before calculation.

The same rules can be applied to  $i7$ – $i10$ ,  $f4$  and  $f5$  too. In this case  $i7$ – $i10$ ,  $f4$  and  $f5$  correspond to  $i3$ – $i6$ ,  $f2$  and  $f3$ . Example 1 shows how to use this possibility in order to evaluate photometric data simultaneously but measured at different pathlengths.

At this point, data for the particular experiment will follow using the type B lines (preceded, optionally by a title line, which is good practice). The arrangement of data on the type B lines will be determined by the presence (or absence) of lines 02, 03, 04 and/or 05.

**Titration data.** The first line contains the initial volume followed by the initial total concentrations of components *in the solution to be titrated*. The order of the total concentrations is determined by the order established on the *key=02* line. The following lines will contain the volume of titrant for that point and then the measured potentials (or pH), followed by the absorbances, if the potentials and absorbances are measured in the same solutions (this is not the case in Example 2). Consequently, either potentials or absorbances may be absent. Again, the ordering of the potential (or pH) measurements follows the order specified by the *key=04* line while the ordering of the absorbances follows the order specified by the *key=05* line. If more than eight data (*i.e.*, floating point numbers) are to be entered a new line is used. However, data for a new point always must begin on a new line.

**Pointwise measurements.** The type B lines are set up as follows: total concentrations ordered according to the *key=02* line, measured potentials in the *key=04* line order and measured absorbances in the *key=05* line order.

The end of the data for a titration curve is indicated by  $-1.0$  in the field of the first real number (by default, columns 1–10). It should be noted that if more than one line is needed for each point, then—when terminating with

–1.0,—add as many blank lines as necessary after the termination line as continuation lines (*see*: set titled PHOTOMETRIC TITRATION... in Example 2).

If for the 01 line,  $i_1$  is greater than unity, and there is a 03 line, then the 2<sup>nd</sup>, 3<sup>rd</sup> ...  $i_1$ <sup>th</sup> sets of titration data will follow, terminated by –1.0.

For pointwise measurements, the line containing –1.0 should not be given because  $i_1$  of *key*=01 SET determines the number of points.

### 3.5.1 Simulation of Data

The “experimental” data can be simulated, too (*see*: Examples 6,7). In this case, the meanings of *keys*=02–05 remain unchanged but *key*=01 SET and 10 ENDSET have another form.

01 SET  $i_1$   $i_2$

$i_1$  is the same as earlier while  $i_2=-2$  (it means the simulation).

The *key*=10 ENDSET can have two forms for simulation:

*For pointwise measurements:*

10 ENDSET  $i_1$   $i_3$

where  $i_1$  is the same as earlier and  $i_3$  is the number of points to be simulated.  $i_3$  must be larger than one.

Two type B lines follow this line. The first one contains the total concentrations of the first point to be simulated. The order of total concentrations is determined by *key*=02 line. The second line contains the differences between the total concentrations of data to be simulated successively. PSEQUAD increases the data of the last simulated point with the differences and repeats this calculation ( $i_3-1$ ) times. In this way the program simulates  $i_3$  “experimental” points.

A slightly more complicated way can also be used:

10 ENDSET  $i_1$   $i_3$   $i_4$   $i_5$   $i_6$   $i_7$   $i_8$   $i_9$   $i_{10}$

where  $i_1$  is the same as earlier and  $i^*$  ( $=3...10$ ) are the number of points to be simulated successively. The total number of the simulated points will be  $\sum_{j=3}^{10} i_j$ .  $i^*$  ( $=3...10$ ) must be larger than one or must equal to zero. If a  $i^?$  equals to zero then all  $i^*$  having larger index must also be zero.

If the first  $n$  elements of  $i^*$  ( $=3...10$ ) are larger than zero,  $2n$  type B lines follow. The syntax is the same as it was detailed above and  $i_j$  belongs to the  $2(j-3)+1$ <sup>th</sup> and the  $2(j-3)+2$ <sup>th</sup> lines.

*For titration-type curves:*

10 ENDSET i1 i3 i4 i5 i6 i7 i8 i9 i10

where i1 is the same as earlier and i\* (\* = 3...10) is the number of the points to be simulated for the (\*-2)<sup>th</sup> titration curve.

Three type B lines follow this line. The first one contains the initial volume and total concentrations of the components in the solution to be titrated. The second one contains the data of the first point to be simulated. It can consist of the volume of titrant, the “measured” potential(s) (or pH) and the “measured” absorbance(s). The third one contains the differences between two successive points within one curve. The data of the first simulated point are increased with the data of the third line (i\*-1) times.

If the experimental points are simulated, a line containing “-1.0” must not be placed between two titration curves!

Once all the data for the particular experiment have been entered then either

01 SET i1 or

10 ENDDATA i1 i2

should appear. The first possibility indicates the beginning of a new experiment and *keys*=01, 02, 03, 04, 05 and 10 lines are used as indicated above. The second possibility means that all the data pertaining to all the experiments have been entered. For this situation i1 and i2 have the following significance:

i1=0 no title line,

i1=1 a title line follows,

i2=0 no experimental or simulated data to be printed,

i2=1 print out solution compositions and other relevant input data and

i2=2 print out all input and simulated data.

In this context the 10 ENDDATA line pairs with the 10 DATAIN line.

There are—probably less important—two additional keys influencing the input data. Both of them must be given before *key*=10 ENDDATA!<sup>16</sup>

-1 DELETE

This key deletes all the previously entered input data with the exception of the titles.

---

<sup>16</sup>Further keys existed in the previous versions of PSEQUAD: -2 INTAPE and -3 OUTTAPE but they are obsoleted. Also, an earlier structure of input data (called KAPA) cannot be used any more.

-4 POTCHNG i1 i2 i3 i4

This key exchanges the sequence of the  $i3^{\text{th}}$  and  $i4^{\text{th}}$  potentials. If  $i2$  equals to zero, this exchange concerns only the  $i1^{\text{th}}$  set (*i.e.*, data after the  $i1^{\text{th}}$  *key=01 SET*). If  $i2 > 0$  then  $i2$  must be greater than (or equal to)  $i1$ . In this case, the change concerns all sets between the  $i1^{\text{th}}$  and  $i2^{\text{th}}$  ones. For example:

```

          i1 i2 i3 i4
04 POT      04      02
-4 POTCHNG 01      04 02

```

results the same effect than the

```

04 POT      02      04

```

single line.

### 3.6 Data Processing Options and Instructions

The second section of the input file comprises the instructions to PSEQUAD for the several data processing options relating to the various experiments, including model-selection. More sets of calculations can be defined in a single input file. Each set of calculations begins with a *key=11* line (default keyword is TASK) and ends with a *key=10* line (default keyword is ENDDATA). We shall refer to these calculation sets as tasks.

In EX2.DAT file, immediately following 10 ENDDATA the overall title for the second section for the data part (*i.e.*, PH-METRIC, POTENTIOMETRIC...) is found. The first task to be performed is the calculation of the  $pK_w$ . The second task will be the refinement of the  $pK_a$ 's for NTA, and subsequent tasks will be the refinement of all or some of the formation constants using the model-selection option. Notice that the tasks are performed in a logical sequence so that the value of the  $pK_w$  from the first task will be incorporated into the equilibrium model for the second task.

The following keys can be used for data processing:

11 TASK i1 i2 i3 i4 i7 i8 i9 i10

*Key=11 TASK* indicates the beginning of a task (*i.e.*, calculation). The integers  $i1 \dots i4$  and  $i7 \dots i10$  are used to indicate the number of iterations, the number of colored species, etc. The omission of certain values, *i.e.*,  $i^*$  ( $=1 \dots 4, 7 \dots 10$ ) set to zero, also provides PSEQUAD with information concerning the type of experimental data to be included in this task. If no data given before a 11 TASK line is processed, PSEQUAD terminates with a "Missing data" error message. The significances of the integer values are as follows:

- i1: The number of components, *i.e.*, the number of columns in  $\alpha$  matrix (see: Eq. (1)).
- i2: The number of complexes *plus* components, *i.e.*, the number of rows in  $\alpha$  matrix (see: Eq. (1)).
- i3: The maximum number of iterations in GN algorithm for fitting the parameters to be refined.
- i4: The maximum number of the additional calculations in the model-selection procedure. If i4=0 there is no model-selection (i4=0 is suggested with PC-version).
- i7: The number of species with unknown molar absorptivities.
- i8: The total number of absorbing species.
- i9: The serial number of the first wavelength used in the calculation.
- i10: The serial number of the last wavelength used in the calculation.

The last two integers, i9 and i10, are not necessarily the same as i1 and i2 for the *key*=05 ABS line. Any continuous part of the absorbance data that has been read in may be selected for calculation. This is the purpose of i9 and i10. (This option was not used in Example 2.)

00 TEXT i1 i2 i3 i4 i5 i6 i7 i8 i9 i10 f1

*Key*=00 implies initial equilibrium model specification for a species. Every species (except the components) has to be described by a line having *key*=00, therefore, i2–i1 (from line having *key*=11 TASK) lines beginning with 00 must be in the model description. At this point in the data set, the total equilibrium model is defined, including probable and not so probable species. TEXT (eight characters maximum) may be used for the formula of the species, for example HL, ML2, H or OH<sup>-</sup>. The integers i1–i10 have the following meanings:

- i1: Serial number for the species. This should be started at (number of components+1) or at 1 if the components are also indicated (see: below).
- i2: A value assigned to each species according to the model selection rules. This value is called *selection number*. If the model-selection is not used, i2 can have the following values:
  - 1 means that the formation constant is to be fitted (*i.e.*, the  $\log \beta$  given in f1 field is used and changed).
  - 0 means that the formation constant is to be fixed (*i.e.*, the  $\log \beta$  given in f1 field is used but it is not changed).



-1 means that the species in question is to be omitted for the current calculation. If  $i_4$  of  $key=11$  TASK is 0, any negative integer results the omission of the species in question.

$i^*$ : ( $=3 \dots 10$ ) are the elements of the  $\alpha$  matrix of Eq. (1), one row per line.

$f_1$ : Value for  $\log \beta$ . If  $i_2=0$ , this value is fixed during the next calculation. If  $i_2>0$ , this value considered to be refined, so it is an initial estimation.

If there are more than eight components, only  $i_1$ ,  $i_2$  and  $f_1$  are entered on this line and the elements of the  $\alpha$  matrix should be given on consecutive lines (see: Example 3).

Note that the model definition lines need appear only once.

In EX2.DAT file, they occur in the first task, calculation of  $pK_w$ . All  $i_2$  values, except for species 4, are set to -1, meaning that they will not be included in the next calculation.

If names should also be given for the components, they must be indicated in the input file, too. In this case, every  $key=00$  line concerning to the components has to be placed immediately after the  $key=11$  line (see: Example 8 and the example below). The selection number of the components must be zero and their  $\log \beta$  values must be fixed to 0.0, by definition.

For example, a pH-metric titration is given in a citric acid solution. The model for the calculation of  $\log \beta_i$ 's ( $i = 1, 2, 3$ ) can be described in the following way<sup>17</sup> (the unknown keys are detailed later):

I2A4__A4__i1i2i3i4_____f1	← Positions
<hr/>	
11_TASK_____2_612	
00_H_____1_0_1_0_____0.0	
00_A_____2_0_0_1_____0.0	
00_HA_____3_1_1_1_____5.4	
00_H2A_____4_1_2_1_____10.1	
00_H3A_____5_1_3_1_____13.2	
00_OH-_____6_0-1_0_____ -13.79	
15_STDDEV__1_____1.0	
17_ADD_____1_____0.032	
10_ENDTASK_0_2	

The species to be included for refinement in the subsequent tasks are specified through the 13 or 14 lines.

<sup>17</sup>The ' ' sign means a space in the followings.

---

```
13 SPECIES i1 i2 i3 i4 i5 i6 i7 i8 i9 i10      f1 f2 f3 f4 f5
```

i1: The serial number for the species in the  $\alpha$  matrix in Eq. (1). It has already been defined in i1 field of *key*=00 lines.

i2: The same as i2 in *key*=00 lines.

f1: Value for  $\log \beta$  for the i1<sup>th</sup> species.

(i3, i4, f2), (i5, i6, f3), (i7, i8, f4) and (i9, i10, f5) groups carry the same meaning as (i1, i2, f1) but for the i3<sup>th</sup>, i5<sup>th</sup>, i7<sup>th</sup> and i9<sup>th</sup> species, respectively.

```
14 SPECIES i1 i2 i3 i4 i5 i6 i7 i8 i9 i10
```

The *key*=14 line is the same as *key*=13 one but the  $\log \beta$ 's do not get a new value before the new calculation so the initially given or previously calculated value remains valid.

The *key*=13 and *key*=14 lines are to be used in conjunction with the *key*=00 lines. Their purpose is to reset the i2 value of the *key*=00 lines defined earlier (with *key*=14 lines), or to reset the i2 and f1 values of the *key*=00 lines (with *key*=13 lines), in subsequent tasks.

In Example 2, the calculation of  $pK_w$  is required initially, followed by the calculation of  $pK_a$ 's for NTA. This is accomplished by using a *key*=14 line in the second task having the values from the previous task:

```
14 SPECIES 4 0 5 1 6 1 7 1
```

which implies that the constant for species 4 ( $pK_w$ ), obtained in the first task, will be used as a fixed value, i2=0, replacing the value assigned by the original *key*=00 line for species 4. Additionally, the formation constants of species 5, 6, and 7 have to be refined (i4, i6, i8=1). It can also be seen from EX2.DAT that the third task involves the refinement of the full equilibrium model with the constants for species 4–7 held fixed at values obtained from the previous tasks. Note that a second *key*=14 line is necessary in this third task.

After all the experimental data have been printed out together with processing options required, the various parameters are refined together with the estimation of  $pK_w$ . The evaluation is continued with the refinement of the  $pK_a$ 's of NTA using set 2, minimizing pH. Note that the first task provided the value of  $pK_w$  (13.7745) which is now used in the second and subsequent tasks. The next task is the calculation of the formation constants using sets 3–6. Only MA, MA<sub>2</sub>, and MAH<sub>-1</sub>, are considered at this point in the calculations. The species MAH and MAH<sub>2</sub> are included only in subsequent tasks (–2 and –3 are assigned to these species on the 14 line). The absorbance data are included in these calculations. by using appropriate values for i6–i10 on *key*=11 TASK. This task is entitled "Evaluation of all experiments with model-selection."

The remaining type A lines relate to various data preprocessing options. Two lines, *keys*=15 and 16, are used to provide statistical weighting information to be applied to the potentiometric and/or absorbance data. These

lines determine the *fitting factors* for the different sorts of measurements. Fitting factors (FF) are the reciprocal values of the square root of the weighting factors defined in Eq. (4).

```
15 STDDEV i1 i3 i5 i7 i9 f1 f2 f3 f4 f5
```

*i*\*: (\*=1, 3, 5, 7, 9) indices relating to the experimental sets of measurements that use the same weighting procedure. For example, if the experimental data belonging to the third *key*=01 SET are to be included in the fitting procedure, one of the *i*\* must have a value 3.

*f*1: Fitting factor for the volume of the titrant (in the case of titrations) or for the first total concentration (in the case of pointwise measurement).

*f*2: The ratio of the fitting factors for the volume and for the  $-\log[c_1]$  data for orthogonal regression (only for potentiometry or pH-metry!). Two arrangements can be given:

- If the experimental data measured potentiometrically are to be fitted (*i.e.*, *f*3≠0.0, see below) then the fitting factors are the followings:

$$\text{FF}(\text{potential}) = \text{FF}(\text{orthogonal}) \text{ and}$$

$$\text{FF}(\text{volume}) = f2 \times \text{FF}(\text{orthogonal})$$

- If the volume data are to be fitted (*i.e.*, *f*1≠0.0) then the following equations are valid:

$$\text{FF}(\text{potential}) = \text{FF}(\text{orthogonal}) / f2 \text{ and}$$

$$\text{FF}(\text{volume}) = \text{FF}(\text{orthogonal})$$

*f*3: Fitting factor of the first measured potential.

*f*4: Fitting factor of the second measured potential.

*f*5: Fitting factor of the absorbance measurements.

One of the *f*1, *f*3, *f*4 and *f*5 should be 1.0 for a simple fitting.

If *f*3, *f*4 or *f*5 equals to −1.0, the appropriate experimental data are not taken into consideration during the next refinement. For example, if there are both potentiometric and absorbance measurements and only the absorbance data are to be fitted then *f*3 must be −1.0 and *f*5>0.0!

If the values of experimental data are modified by the help of a line having *key*=10 ENDSET then the fitting factor corresponds to the modified data (see: Example 1). For another example, let the primary data measured in a cuvette having 0.2 cm pathlength. These data must be multiplied by five to transform them to 1.0 cm length (10 ENDSET . . . *f*3=5.0).

In this case, '15 STDDEV...f5=1.0' regards to the modified data while '15 STDDEV...f5=5.0' would regard to the primary measured data.

If more than two potentials are measured, then a line having *key*=16 is necessary:

```
16 STDDEV i1 i3 i5 i7 i9      f3 f4 f5
```

*i1...i9* have the same meaning as in the *key*=15 line and *f3*, *f4* and *f5* are the fitting factors of the subsequent potentials.

If more than five sets of measurements are evaluated in the same manner, then a second *key*=15 line (and *key*=16 line) is necessary. If the fitting factor of the particular type of data is positive then these data are included in Eq. (4); if zero then they are fixed (*i.e.*, they are considered to be accurate without random error); and if negative then the given data are disregarded in the calculation. Again, the weighting factors are the reciprocal values of the square of the fitting factors.

In Example 2, it was our experience that the pCu measurements are about half as precise as the pH data, and absorbance measurements are about three times as precise as the pH data. This information is relayed to PSEQUAD using the *key*=15 STDDEV line so that the function, *F* defined in Eq. (4), to be minimized becomes

$$F = (\Delta\text{pH})^2 + \frac{1}{57.87^2}(\Delta\text{emf})^2 + \frac{1}{0.3^2}(\Delta A)^2$$

where

$$\frac{1}{57.87^2}(\Delta\text{emf})^2 = \frac{1}{2^2}(\Delta\text{pCu})^2$$

This function is defined by the second and third *key*=15 lines. It will be seen from the output that the inclusion MAH decreases *F* significantly from 0.0353 to 0.0213. However, the inclusion of MAH<sub>-2</sub> leads to only a slight decrease in the sum of squares from 0.0213 to 0.021. The standard deviation for this latter constant is an order of magnitude higher compared to the other constants.

Finally all six sets of data are included in the refinement and *F* is further decreased to 0.0203. It should be noted that all the refined values of the constants are changed slightly in this final task as there is now no error accumulation from one task to the next.

Since PSEQUAD accepts emf data, a mechanism is required whereby the potentials may be directly converted into log[*c<sub>i</sub>*] values. The Nernst-equation, in the form given by Eq. (2), serves this purpose. A *key*=17 line provides the appropriate additive term *A<sub>l</sub>* and a *key*=18 line gives the multiplicative factor *M<sub>l</sub>* for the *l<sup>th</sup>* component:

```
17 ADDITIV i1 i3 i5 i7 i9      f1 f2 f3 f4 f5
```

*i\**: (\*=1, 3, 5, 7, 9) have the same meaning as in the *key*=15 line.

$f^*$ : ( $=1 \dots 5$ ) are the additive terms to convert potentials into  $\log[c_i]$  values according to Eq. (2). They are ordered in the same manner as the potentials on the  $key=04$  line.

Different addition terms for different sets can be given on more subsequent  $key=17$  lines.

```
18 MULTIPL i1 i3 i5 i7 i9      f1 f2 f3 f4 f5
```

This line contains the multiplicative coefficients in the same sense as for the  $key=17$  line.

These lines are necessary only if the additive terms differ from zero and/or the multiplicative terms differ from  $-1$ . The values used in the  $key=17$  and 18 lines remain in force from one task to next, until (1) new  $key=17$  and/or 18 lines are encountered or (2) a new  $key=10$  ENDDATA line is given.

The end of the task is indicated by a  $key=10$  line. The values of the integers  $i1-i10$  on this line give rise to a variety of further processes.

```
10 ENDTASK i1 i2 i3 i5 i6 i7 i8 i9 i10      f1 f2
```

$i1=0$ : There is no following title line.

$=1$ : The next line will be a type C.

$i2=0$ : The program lists only the calculated parameters together with their standard deviations. No other calculated quantities are listed.

$=1$ : The program additionally lists the back-calculated data and the deviation between the measured and calculated data (residuals).

$=2$ : The program gives a complete output, including the concentration distribution of each species.

$=3$ : The program provides only the concentration distribution in the output besides the calculated parameters.

The required output (determined by  $i2$ ) can be modified by  $i3$ .

$i3=1$ : In the case of model-selection (*i.e.*  $i4>0$  in  $key=11$  TASK line), the program lists the required output only if the model-selection has been finished and the fitting parameter has a minimum.

$=3$ : In the case of model-selection, the program does not give output list before changing the size of the composition matrix (*i.e.*, before adding or eliminating a species).

$=4$ : In the case of model-selection, the program does not give output list before eliminating a species).

- i5=0: The continuation method (*see*: page 7) is used only for titration curves.
- =1: The continuation method (*see*: page 7) is used both for titration curves and for pointwise measurements.
- i6≠0: This field helps to generate random error for the primary (usually simulated) data (*see*: Example 5). If i6≠0 and i2=1 or 2, the primary measured data ( $D_M$ ) are overwritten. The new “measured” data ( $D_R$ ) and the residuals ( $R_R$ ) are calculated according to the following rules:

$$\begin{aligned} i6=-2: & \quad D_R=C+FF \times \xi, & R_R=D_R-C=FF \times \xi \\ i6=-1: & \quad D_R=C, & R_R=0 \\ i6=+1: & \quad D_R=D_M, & R_R=D_M-C \\ i6=+2: & \quad D_R=D_M+FF \times \xi, & R_R=D_M-C+FF \times \xi \end{aligned}$$

where  $C$  is the calculated data,  $FF$  is the fitting factor (defined on page 25 and  $\xi$  is a random number with standard normal distribution ( $[-\infty < \xi < \infty]$ )).

The i6=1 essentially corresponds to the case i6=0. The only difference is that when the residuals are plotted for a potentiometric (or pH-metric) measurements, the unit is the calculated standard deviation for i6=0 and it is the fitting factor for i6=1. When orthogonal regression is used, i6 must be zero! Random number simulation is possible for either the volume of the titrant or the measured potential but not for both!

- i7=0: The experimental volumes are used for all the iterations when the total concentrations are calculated for a titration curve:

$$C_i^{\text{calc}} = \frac{V_0 C_i^0 + V_{\text{expr}} C_i^{\text{burette}}}{V_0 + V_{\text{expr}}} \quad (i = 1 \dots k) \quad (13)$$

- =1: The previously calculated volumes are used for the second and subsequent iteration:

$$C_i^{\text{calc}} = \frac{V_0 C_i^0 + V_{\text{calc}} C_i^{\text{burette}}}{V_0 + V_{\text{calc}}} \quad (i = 1 \dots k) \quad (14)$$

In *Eqs. (13, 14)*  $C_i^{\text{calc}}$  means the calculated total concentration of the  $i^{\text{th}}$  component,  $V_{\text{expr}}$  and  $V_{\text{calc}}$  mean the experimental and calculated volume of the titrant in the actual point,  $V_0$  means the initial volume of the solution to be titrated,  $C_i^0$  and  $C_i^{\text{burette}}$  mean the total concentrations of the  $i^{\text{th}}$  component in the the solution before the titration and in the burette and  $k$  is the number of the components.

i8=n: If  $n \neq 0$  then the back-calculation procedure is carried out at least i8 times.

When fitting titration curves, it is expedient to use the i7=1 code. If i7=1 and i8>0 then the procedure is carried out i8 times, using previously calculated volumes to obtain the total concentrations. This is the recommended procedure for the back-calculation of titration curves when there is not any fitted formation constant.

i9=0: Distribution diagram is not created.

=g: If  $g > 0$  then the distribution diagram concerning to the  $g^{\text{th}}$  component will be plotted. The diagram shows the following ratios:

$$r_j = \frac{\alpha_{jg}[c_j]}{\sum_{i=1}^n \alpha_{ig}[c_i]} \quad (j = 1 \dots n) \quad (15)$$

where  $n$ ,  $c_i$  and  $\alpha_{ij}$  have already defined in Eq. (1).

The distribution diagram cannot be complete if:

- more than 20 species contain the  $g^{\text{th}}$  component.
- any  $\alpha_{ij}$  is negative (e.g., if the  $g^{\text{th}}$  component is  $\text{H}^+$  and  $\text{OH}^-$  is defined among the species).

It is also possible to create a distribution diagram concerning to a species which is not a component. In this case, the composition matrix must contain a fictitious species and i9 is its serial number. Let consider the following *key=00* lines:

I2A4__A4__i1_i2_i3_i4_i5	← Positions
00_H_____1__0__1__0__0__...	
00_ML_____2__0__0__1__0__...	
00_ML2_____3__0__0__0__1__...	
00_L_____4__1__0__-1__1__...	
00_ML2OH_____5__1__-1__0__1__...	
00_M_____6__1__0__2__-1__...	
00_FICTSP_____7__-1__0__1__1__...	

The seventh species is fictitious. Its  $\alpha_{jk}$  (see: Eq. (1)) values mean that how many M occurs in a given component. In this case, PSEQUAD calculates for every species (including the components, too) the  $g_i$  weighting factor:

$$g_i = \sum_{k=1}^3 \alpha_{ik} \alpha_{fk} \quad (i = 1 \dots n)$$



where  $\alpha_{ik}$  is the stoichiometric number of the  $k^{\text{th}}$  component in the  $i^{\text{th}}$  species and  $\alpha_{fk}$  is the same number for the fictitious species.

If  $i9=7$ , PSEQUAD plots the

$$r_j = \frac{g_j[c_j]}{\sum_{i=1}^n g_i[c_i]} \quad (j = 1 \dots n)$$

ratio which is essentially the concentration distribution concerning M.<sup>18</sup>

$i10=0$ : The calculation is continued even if the previous task has failed.

$=1$ : The calculation is continued only if the previous task was successful.

$=2$ : The calculation is continued only if the minimum of error square sum was found in the previous task.

$f1=h$ : If  $i9 \neq 0$ ,  $f1$  defines the width of the diagram.  $h$  can be 0.0, 21.0, 31.0, ... 101.0. If  $f1$  equals to 0.0, the width of the diagram equals to 61 characters.

$f2 < 0$ : The correlation matrix disappears from the output.

When photometric titrations are evaluated in a given task, some additional lines should be given. After the occasional type C line (*i.e.*, title line), a type A line must follow from which the serial numbers of the colored species must be listed in the  $i1 \dots i10$  positions. The species with unknown molar absorptivities must occupy the first positions. More than ten integers can be given in further type A line(s).

The type A line(s) must then be followed by type B lines. The first field of the first line (F10.0, by default) contains the concentration of a species with measured molar absorptivities. The subsequent fields are for the absorbances measured in the wavelength order determined by  $i9$  and  $i10$  on the  $key=11$  line.<sup>19</sup> If more than seven wavelengths are used then the absorbances are continued on subsequent lines (columns 1–80, by default). Data for each new species must begin on a new line. If 1.0 is written into the first field of the first line for a species, then the molar absorptivities should be given,

<sup>18</sup>If the seventh fictitious species were defined as 07 -1 00 01 02 in fields  $i1 \dots i5$ , the distribution diagram would concern L instead of M.

<sup>19</sup>Warning! It also means that if  $i9$  and  $i10$  on the  $key=11$  line give only a subrange of the input data, the known absorbances can be given only for this subrange. If either  $i9$  or  $i10$  is changed, the given absorbances must also be changed!

otherwise, the program calculates the molar absorptivities from the given concentrations and measured absorbances.

This completes the description of the second section of the input files. One or more additional lines may be present:

## 82 STOP

This line indicates the end of the input file.

## 11 TASK i1 i2 ... i10

A new task may be initiated as described above. In the second and subsequent tasks, however, the i1 and i2 positions can be set to new values for defining a new composition matrix or they can also be set to zero. In the latter case, the last defined model remains valid (*i.e.*, the composition matrix remains unchanged) and the parameter values established in previous tasks may be left unchanged or reset within the new task. Changing the properties of the species is possible by applying the *key=13 SPECIES* and/or the *key=14 SPECIES* lines.

There is an (practically unimportant) exception. If the size of the newly defined composition matrix equals to that of the previously used matrix, the defined species and their  $\log \beta$  values remain unchanged but their names are deleted and their selection numbers (*Subsection 3.6* on page 22) are set to -1. To avoid this problem, set i1 and i2 to zero if the previously defined component matrix is to be used.

## 10 DATAIN

This line indicates that additional data are to be given as described above. After the additional data set(s) a *key=10 ENDDATA* line must be given. *This line deletes data given in previous key=17 and key=18 lines so the necessary additive and multiplicative factors must be redefined.*

### 3.7 Model-selection

The aim of model-selection is explained in *Subsection 2.3*. Model-selection is carried out if i4>0 in the *key=11 TASK* line. Species can be divided into five groups for model-selection, depending on their selection number (*see: Subsection 3.6*):

1. Species to be included into the next calculation with fixed formation constant. Their selection numbers are 0.
2. Species to be included into the next calculation with refinable formation constant. Their selection numbers are +1.

3. Species which may be omitted from the next calculation with refinable formation constant. Their selection numbers are positive integers.
4. Species which may be included into the next calculation with refinable formation constant. Their selection numbers are negative integers less than  $-1$ .
5. Species to be omitted from the next calculation. Their selection numbers are  $-1$ .

The algorithm of the model-selection is the following:

- Species having zero or positive selection number are taken into the first calculation.
- The species denoted by the largest positive integer is omitted from the next calculation if
  - a minimum has been reached and the relative change of the fitting parameter is less than  $f1$  on *key=91 SUPERVIS* (0.005 %, by default) *or* the maximum number of iterations ( $i1$  on *key=92 SUPERVIS* (25, by default)) has been reached

*and*

- there is at least one species having selection number larger than  $+1$  *and*
- the logarithm of the formation constant of this species has changed more than one unit.

If the circumstances detailed above are fulfilled, a new calculation begins by omitting the species with the largest positive selection number. This procedure is repeated until the appropriate fitting is reached (*i.e.*, at least one of the conditions above is not true).

- Additional calculations begin with including species denoted by negative selection numbers. Species having larger selection numbers ( $-2 > -3!$ ) are included sooner.

## **4 Output of PSEQUAD**

The output of PSEQUAD is designed as a self-explaining text so the best way to learn the use of PSEQUAD is to study the output files created from *EX\*.DAT* files. The output files include all primary data (on request), the calculated parameters, their deviations, the calculated data, concentration distributions, residuals, distribution diagrams, etc. The defining formulas for the statistical parameters are also included for the sake of unambiguous interpretation.

# Index

-1 (DELETE); 20  
-2 (INTAPE); 20  
-3 (OUTTAPE); 20  
-4 (POTCHNG); 21  
00 (TEXT); 22  
01 (SET); 16, 19  
02 (TOTAL); 16  
03 (TITR); 17  
04 (POT); 17  
05 (ABS); 17  
10 (DATAIN); 16, 31  
10 (ENDDATA); 20  
10 (ENDSET); 17, 19  
10 (ENDTASK); 16, 27  
11 (TASK); 21, 31  
13 (SPECIES); 24  
14 (SPECIES); 24  
15 (STDDEV); 25  
16 (STDDEV); 26  
17 (ADDITIV); 26, 31  
18 (MULTIPL); 27, 31  
82 (STOP); 31  
91 (SUPERVIS); 6, 8, 15  
92 (SUPERVIS); 7, 15  
93 (KRAND); 16  
97 (FREEOUT); 14  
98 (FORMATB); 14  
99 (FORMATA); 14

**A**  
ABS (05); 17  
ADDITIV (17); 26, 31  
 $\alpha$  (composition matrix); 3, 31

**C**  
component; 3  
composition matrix ( $\alpha$ ); 3, 31  
continuation method; 7  
correlation coefficients; 9  
counter-ion; 4

**D**  
DATAIN (10); 16, 31  
DELETE -1(); 20  
default parameter values; 15

**E**  
ENDDATA (10); 20

ENDSET (10); 17, 19  
ENDTASK (10); 16, 27  
equilibrium concentration; 3

**F**  
field conception; 10  
FF (fitting factor); 25  
fitting factor (FF); 25  
FORMATA (99); 14  
FORMATB (98); 14  
free concentration; 3  
FREEOUT (97); 14

**I**  
INTAPE (-2); 20

**K**  
key

-1 (DELETE); 20  
-2 (INTAPE); 20  
-3 (OUTTAPE); 20  
-4 (POTCHNG); 21  
00 (TEXT); 22  
01 (SET); 16, 19  
02 (TOTAL); 16  
03 (TITR); 17  
04 (POT); 17  
05 (ABS); 17  
10 (DATAIN); 16, 31  
10 (ENDDATA); 20  
10 (ENDSET); 17, 19  
10 (ENDTASK); 16, 27  
11 (TASK); 21, 31  
13 (SPECIES); 24  
14 (SPECIES); 24  
15 (STDDEV); 25  
16 (STDDEV); 25  
17 (ADDITIV); 26, 31  
18 (MULTIPL); 27, 31  
82 (STOP); 31  
91 (SUPERVIS); 6, 8, 15  
92 (SUPERVIS); 7, 15  
93 (KRAND); 16  
97 (FREEOUT); 14  
98 (FORMATB); 14  
99 (FORMATA); 14

keyword  
ABS (05); 17

ADDITIV (17); 26, 31  
DATAIN (10); 16, 31  
DELETE -1(); 20  
ENDDATA (10); 20  
ENDSET (10); 17, 19  
ENDTASK (10); 16, 27  
FORMATA (99); 14  
FORMATB (98); 14  
FREEOUT (97); 14  
INTAPE (-2); 20  
KRAND (93); 16  
MULTIPL (18); 27, 31  
OUTTAPE (-3); 20  
POT (04); 17  
POTCHNG (-4); 21  
SET (01); 16, 19  
SPECIES (13); 24  
SPECIES (14); 24  
STDDEV (15); 25  
STDDEV (16); 25  
STOP (82); 31  
SUPERVIS (91); 6, 8, 15  
SUPERVIS (92); 7, 15  
TASK (11); 21, 31  
TEXT (00); 22  
TITR (03); 17  
TOTAL (02); 16

KRAND (93); 16

## M

model-selection number;

*see*: selection number

molar absorbances; 30

MULTIPL (18); 27, 31

multiple correlation coefficients; 9

## O

OUTTAPE (-3); 20

## P

partial correlation coefficients; 9

photometric data; 30

pointwise measurement; 18

POT (04); 17

POTCHNG (-4); 21

## R

reference example; 11

residual; 27

## S

selection number; 22

SET (01); 16, 19

SPECIES (13); 24

SPECIES (14); 24

STDDEV (15); 25

STDDEV (16); 25

STOP (82); 31

SUPERVIS (91); 6, 8, 15

SUPERVIS (92); 7, 15

## T

TASK (11); 21, 31

TEXT (00); 22

TITR (03); 17

titration data; 4, 18

TOTAL (02); 16

total concentration; 3

total correlation coefficients; 9