



HAL
open science

Global sensitivity based estimability analysis for the parameter identification of Pitzer's thermodynamic model

Ilias Bouchkira, Abderrazak Latifi, Lhachmi Khamar, Saad Benjelloun

► **To cite this version:**

Ilias Bouchkira, Abderrazak Latifi, Lhachmi Khamar, Saad Benjelloun. Global sensitivity based estimability analysis for the parameter identification of Pitzer's thermodynamic model. Reliability Engineering and System Safety, 2021, 207, pp.107263. 10.1016/j.res.2020.107263 . hal-03201846

HAL Id: hal-03201846

<https://hal.science/hal-03201846>

Submitted on 15 Dec 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Global sensitivity based estimability analysis for the parameter identification of Pitzer's thermodynamic model

Ilias Bouchkira^{a,b}, Abderrazak M. Latifi^{a,b,*}, Lhachmi Khamar^a, Saad Benjelloun^a

^aMohammed VI Polytechnic University, Lot 990 Hay Moulay Rachid, Benguerir 43150, Morocco

^bLaboratoire Réactions et Génie des Procédés, CNRS-ENSIC, Université de Lorraine, 1 rue Grandville, BP20451, 54001, Nancy Cedex, France

Abstract

This paper deals with parameter estimability analysis and identification of the Pitzer model used in the prediction of thermodynamic properties and phase equilibria of electrolytic solutions. The estimability analysis method used is based on the orthogonalization of a sensitivity matrix to rank the unknown parameters from the most estimable to the least estimable. Although the obtained results are interesting, the algorithm shows its limits since it is based on the local sensitivities of the outputs with respect to the unknown parameters of the model and may significantly affect the reliability of the model. In this work, the algorithm is improved by computing the sensitivities by means of the method of global sensitivity analysis. For demonstration purposes, different sets of experimental measurements of sulfuric acid solutions are carried out at different temperatures and acid concentrations. They mainly consist of pH, density, conductivity and molality measurements. The improved algorithm is then applied to each of the experimental sets. The most estimable parameters are determined and identified using a branch-and-reduce optimization method and their accuracy is assessed by means of confidence intervals. Finally, the quality of the model is quantified by computing the Pearson product-moment coefficient, its high values show a very good agreement between the predictions and the measurements.

Keywords: Estimability analysis, Global sensitivity analysis, Thermodynamics, Pitzer model, Sulfuric acid solutions.

1. Introduction

In process engineering, different models are developed for prediction, design, optimization, and control of a single unit operation or a set of (inter) connected unit operations (i.e. plant). They are very often first principle models and consist of momentum, mass and energy balances along with equations describing thermodynamic, kinetic, and hydrodynamic phenomena. These models inevitably involve several unknown parameters that are often deduced from experimental measurements by means of a parameter identification method.

In the literature, most studies assume that the available experimental measurements contain the necessary information to accurately identify all the unknown parameters involved in the models. However, it is well recognized that it is not always the case. Indeed, due to limited data, noisy measurements, or sometimes correlated designs of experiments, the estimation of the unknown parameters may not be accurate [14] and the underlying question is to know which parameters are estimable from the available measurements and whether it is possible to design

appropriate experiments to determine the least estimable ones. To answer these questions, several research works have been devoted to the development of parametric estimability analysis methods [35-41]. They allow to rank the unknown parameters from the most estimable to the least estimable using the available experimental measurements.

One of the most interesting estimability analysis method was developed by Yao et al. [28]. It is a direct selection method based on the building of a sensitivity matrix of the model outputs with respect to the unknown parameters and its use within an orthogonalization algorithm. The method has proven to be effective provided that the estimability threshold value used is optimal. However, the method exhibits two major limitations. The first one deals with the choice of the estimability threshold value, which is actually made almost arbitrary. Therefore, the number of estimable parameters depends significantly on its chosen value. The second limitation concerns the sensitivity matrix whose formation is based on the computation of local sensitivities which depend on the initial values used for the unknown parameters. The resulting order of estimability of parameters may therefore change from one set of initial values to another.

*Corresponding author

Email address: abderrazak.latifi@univ-lorraine.fr
(Abderrazak M. Latifi)

Nomenclature

A_{Φ}	Debye-Hückel constant
C	Covariance matrix
$C.I.$	Confidence interval
c_{jj}	j^{th} diagonal element of the matrix
d	Number of model parameters
E	Mathematical expectation
$F(P)$	Objective function
G^{ex}	Free energy of Gibbs
I	Ionic strength (mol/kgw)
J	Jacobian matrix
K_i	Equilibrium constant at T_i
K_{i0}	Equilibrium constant at $25^{\circ}C$
L	Number of the most estimable parameters
m_i	Molale concentration of component i
N	Number of P samples
n	Number of data points
P	Vector of parameters
$P_{i,j}$	Pitzer interaction parameter
P^*	Vector of optimal parameters
\overline{P}_j	Average parameter j
pH	pH of the aqueous phase
r	Pearson product-moment coefficient
R	Gas constant $Jmol^{-1}K^{-1}$
$S_{i,j}$	Total Sobol sensitivity of Parameter j with respect to the measurement i
$S_{i,j}^*$	Normalized sensitivity index
$\overline{S}_{i,j}^e$	Average of experimental measurements
$\overline{S}_{i,j}^m$	Average of model predictions
T	Temperature (K)
V	Total variance of a model output
\hat{V}	Standard deviation of the model outputs
W	Weighing matrix
W_n	Mole fraction of water
\overline{y}_i	Average of the i^{th} model output
Z	Sensitivity matrix
z_i	Electrical charge of component i

Greek letters

γ_i	Activity coefficient of a component i
$\beta_{i,j}$	Pitzer binary interaction parameters
$\delta_{i,j}$	Number of elements j in component i
ΔH	Enthalpy of reaction (kJ)
$\theta(P)$	Residue vector
$\lambda_{i,j}$	Pitzer binary interaction parameters
$\psi_{i,j,k}$	Pitzer ternary interaction parameter

To overcome the arbitrary choice of the estimability threshold value, Wu et al. [14] and Eghtesadi and McAuley [15] have further developed the algorithm of Yao et al. [28] by proposing a forward selection methodology based on a mean-square error for simultaneous ranking and selection of parameters. Although the obtained results are very promising, the methodology is still dependent on how the elements of the sensitivity matrix are computed. In most cases, and particularly in some of our recent works [30-31], they are approximated by means of a centered finite difference method at a given set of initial values of the parameters (local sensitivities). A computation method of sensitivities which is independent of initial values of the parameters would significantly improve the estimability analysis method.

This is the objective of this paper which aims to develop an estimability analysis approach based on global sensitivities which are no longer influenced by the initial values of the model parameters. More specifically, the Sobol method [9] based on the decomposition of the total variance into a sum of different partial variances will be used. The case study deals with the prediction of physical properties and phase equilibria of concentrated sulfuric acid solutions using the Pitzer thermodynamic model. The latter is included as a thermodynamic property prediction method in several flowsheeting computation codes such as Aspen Hysys [34], Pro/II [17], PhreeQC [32] etc. Accurate estimation of its unknown parameters would definitely improve the design and operation of processes involving electrolytic solutions.

The paper is organized as follows. Section 2 describes the thermodynamic modeling of sulfuric acid solutions. The sensitivity and estimability analysis approaches are presented in Section 3, while the results are discussed in Section 4. Finally, a conclusion and perspectives of the work are provided in Section 5.

2. Pitzer's thermodynamic model

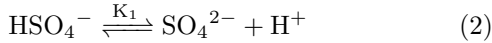
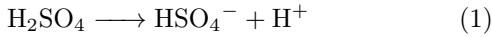
The Pitzer thermodynamic model [24] is widely used to describe the interactions between dissolved ions and solvent (water). Its use therefore depends on the chemical species present in solution (i.e. speciation) as well as their molalities. This section first presents the chemical speciation of sulfuric acid solutions, then the mass and charge balances, the Pitzer model equations as well as its unknown parameters, and finally the experimental measurements to identify them and to validate the model predictions.

Table 1: Matrix of components

j index		1	2	3	charge z_i
i index	components	S	H	O	
1	H^+	0	1	0	+1
2	HSO_4^-	1	1	4	-1
3	SO_4^{2-}	1	0	4	-2

2.1. Chemical speciation of sulfuric acid solutions

In water, sulfuric acid may dissociate into different ions depending on operating conditions such as temperature, pressure and total molality [1-6] as :



The first equation is considered complete, i.e. the sulfuric acid undergoes total speciation, whereas the second equation is in equilibrium. The chemical species present in the aqueous solution are therefore H^+ , HSO_4^- and SO_4^{2-} ions.

The Arrhenius-type equation is used to compute the equilibrium constants at different temperatures as [22]:

$$\log(K_i) = \log(K_{i0}) + \frac{\Delta H}{R} \left(\frac{1}{T_i} - \frac{1}{T_0} \right) \quad (3)$$

where K_{i0} and ΔH are the reference equilibrium constant and enthalpy respectively. Their values are taken from the *Minteq* thermodynamic database [23].

2.2. Mass and electroneutrality balances and ionic strength

The development of mass and electroneutrality balances is based on the matrix of chemical components present in the aqueous medium. Table 1 lists the number of chemical elements (i.e. sulfur (S), hydrogen (H) and oxygen (O)) contained in each component involved in the reactions of sulfuric acid dissociation (Eqs.1 and 2). The charge z_i of each component i is also reported.

Mass and electroneutrality balances are given by the following equations :

$$(S)_{Total} = \sum_{i=1}^{i=3} \delta_{i,1} m_i \quad (4)$$

$$(H)_{Total} = \sum_{i=1}^{i=3} \delta_{i,2} m_i \quad (5)$$

$$(O)_{Total} = \sum_{i=1}^{i=3} \delta_{i,3} m_i \quad (6)$$

$$\sum_{i=1}^{i=3} q_i z_i m_i = \sum_{i=1}^{i=3} n_i z_i m_i \quad (7)$$

where m_i is the molality of the component i , $\delta_{i,j}$ is the number of the chemical element j in the component i . q_i is equal to 1 if the i^{th} component is charged + and to 0 otherwise. Likewise, n_i is equal to 0 if the i^{th} component is charged + and to 1 otherwise. The ionic strength, denoted I , is expressed as:

$$I = \frac{1}{2} \sum_{i=1}^{i=3} m_i z_i^2 \quad (8)$$

In order to solve the resulting model equations which consist of equilibrium constants, material and electroneutrality balances and ionic strength, it is important to compute the activity coefficients of the ions present in the aqueous phase. It is noteworthy that for electrolytic systems where the ionic strength is very low, some empirical expressions can accurately approximate these activity coefficients such as the Debye-Huckel equation [7]. However, when the ionic strength is high, it is necessary to use a suitable thermodynamic model.

In this work, we consider the Pitzer model [24, 27] as one of the most used, flexible and accurate models that can be used to compute the activity coefficients of the ions.

2.3. Pitzer model equations

The Pitzer model equations [24] allow computing important properties such as osmotic and activity coefficients of ions based on the basic virial development of excess free enthalpy as:

$$\frac{G^{ex}}{W_n RT} = f(I) + \sum_{i,j} m_i m_j \lambda_{i,j} + \sum_{i,j,k} m_i m_j \psi_{i,j,k} + \dots \quad (9)$$

The derivative of this expression of Gibbs free energy with respect to the mole fraction of an ion allows to compute its activity coefficient in the aqueous phase as [24]:

$$\ln(\gamma_i) = f^\gamma(I) + 2 \sum_j m_j \lambda_{i,j}(I) + z_i^2 \sum_j \sum_k m_j m_k \lambda'_{i,j}(I) + 3 \sum_j \sum_k m_j m_k \psi_{i,j,k} + \dots \quad (10)$$

where $f^\gamma(I)$ is the derivative of the Debye-Huckel term. $\lambda_{i,j}(I)$ and $\lambda'_{i,j}(I)$ are the binary interaction parameters and their derivatives with respect to ionic strength respectively, and expressed as :

$$\lambda_{i,j}(I) = \beta_{i,j}^{(0)} + \beta_{i,j}^{(1)} \frac{2}{1.2} \left(1 - (1 + \alpha\sqrt{I}) e^{-\alpha\sqrt{I}} \right) \quad (11)$$

$$\lambda'_{i,j}(I) = \frac{\beta_{i,j}^{(1)}}{\alpha^2 I} \left(-1 + (1 + \alpha\sqrt{I} + \alpha I) e^{-\alpha\sqrt{I}} \right) \quad (12)$$

The resulting model involves many unknown parameters that should be identified from experimental measurements. In our case, the Pitzer model involves binary interaction parameters $\beta_{i,j}^{(0)}$ and $\beta_{i,j}^{(1)}$, and a ternary interaction parameter $\psi_{i,j,k}$ between H^+ , HSO_4^- and SO_4^{2-} ions. The binary interaction parameters $\beta_{i,j}^{(0)}$ and $\beta_{i,j}^{(1)}$ between these different ions are involved in (Eqs 11 and 12). It is assumed that $\beta_{i,j} = \beta_{j,i}$ and the binary interaction parameters between the same ions are equal to zero [24].

The number of interaction parameters is therefore 7, i.e. 6 binary and 1 ternary parameters (Table 2). Since they are temperature dependent, all the parameters are expressed as functions of temperature as [20]:

$$\beta_{i,j}^{(0)}(T) = a_{i,j}^{(0)} + b_{i,j}^{(0)} \cdot T + \frac{c_{i,j}^{(0)}}{T} \quad (13)$$

$$\beta_{i,j}^{(1)}(T) = a_{i,j}^{(1)} + b_{i,j}^{(1)} \cdot T + \frac{c_{i,j}^{(1)}}{T} \quad (14)$$

$$\psi_{i,j,k}(T) = a_{i,j,k}^{(2)} + b_{i,j,k}^{(2)} \cdot T + \frac{c_{i,j,k}^{(2)}}{T} \quad (15)$$

As a result, the total number of 21 parameters (listed in Table 2) are to be determined from the experimental measurements. However, it is essential to have access to a database of experimental measurement in order to identify these unknown model parameters.

2.4. Experimental methods and measurements

For demonstration purposes, different sets of experimental measurements of sulfuric acid solutions are carried out. They consist of pH, density, electrical conductivity and total concentrations measurements obtained as follows : (i) the pH measurements are carried out by means of a glass electrode pH meter which was calibrated using a *KCl* solution at different temperatures, (ii) a pycnometer of known volume and mass is used to measure the density of the samples, (iii) the total molal concentrations of the ions present in the acid solutions are calculated using a specific conductivity based method developed by McClesley et al. [33].

For all experiments, a pure (98w%) sulfuric acid purchased from Sigma Aldrich is used to prepare different samples of 100 ml. The temperature and concentration of the samples vary from 298 K to 353 K and from infinite dilution to 4 moles/kgw respectively.

The question that arises is to know whether the available experimental data contain the necessary information to identify all the unknown parameters or only some of them. To answer this question, an estimability analysis method is developed and described in the next section.

Table 2: Unknown parameters of the investigated model

$\beta_{i,j}^{(0)}$					
1	$a_{HSO_4^-, SO_4^{2-}}^{(0)}$	8	$b_{HSO_4^-, SO_4^{2-}}^{(0)}$	15	$c_{HSO_4^-, SO_4^{2-}}^{(0)}$
2	$a_{HSO_4^-, H^+}^{(0)}$	9	$b_{HSO_4^-, H^+}^{(0)}$	16	$c_{HSO_4^-, H^+}^{(0)}$
3	$a_{SO_4^{2-}, H^+}^{(0)}$	10	$b_{SO_4^{2-}, H^+}^{(0)}$	17	$c_{SO_4^{2-}, H^+}^{(0)}$
$\beta_{i,j}^{(1)}$					
4	$a_{HSO_4^-, SO_4^{2-}}^{(1)}$	11	$b_{HSO_4^-, SO_4^{2-}}^{(1)}$	18	$c_{HSO_4^-, SO_4^{2-}}^{(1)}$
5	$a_{HSO_4^-, H^+}^{(1)}$	12	$b_{HSO_4^-, H^+}^{(1)}$	19	$c_{HSO_4^-, H^+}^{(1)}$
6	$a_{SO_4^{2-}, H^+}^{(1)}$	13	$b_{SO_4^{2-}, H^+}^{(1)}$	20	$c_{SO_4^{2-}, H^+}^{(1)}$
$\psi_{i,j,k}$					
7	$a_{HSO_4^-, SO_4^{2-}, H^+}^{(2)}$	14	$b_{HSO_4^-, SO_4^{2-}, H^+}^{(2)}$	21	$c_{HSO_4^-, SO_4^{2-}, H^+}^{(2)}$

3. Estimability analysis method

The estimability analysis method used in this work is based on the computation of sensitivities of the measured outputs with respect to the unknown parameters of the model and implemented within an orthogonalization algorithm described below.

3.1. Orthogonalization-based methods

Among the parameter selection methods developed in the literature, the orthogonalization based methods have proven to be particularly relevant to rank the parameters from the most to the least estimable. The method used in this work is based on the use of a sensitivity coefficient matrix Z that is computed from the individual sensitivity coefficients S_{ij} as [16, 18, 28] :

$$Z = \begin{pmatrix} S_{1,1} |_{m_1, T_1} & \cdots & S_{1,d} |_{m_1, T_1} \\ \vdots & \ddots & \vdots \\ S_{n,1} |_{m_n, T_n} & \cdots & S_{n,d} |_{m_n, T_n} \end{pmatrix}$$

where T_i is the operating temperature, m_i is the total molality of H_2SO_4 , d is the number of unknown parameters and n is the number of experimental measurements. In the Z matrix, the individual sensitivity coefficients $S_{i,j}$ are usually approximated by means of a finite difference method as:

$$S_{i,j} \approx \frac{\Delta Y_i}{\Delta P_j}, \quad i \in [1; n], \quad j \in [1; d] \quad (16)$$

where Y_i is the i^{th} model output (e.g., ion concentration or water activity) and P_j is the j^{th} unknown parameter.

The components of the vector P of the unknown parameters are defined by the coefficients listed in Table 2. The elements $S_{i,j}$ are then multiplied by scaling factors, i.e. \overline{P}_j and \overline{Y}_i , in order to ensure dimensional consistency and guarantee the same order of magnitude of the elements:

$$S_{i,j}^* = \frac{\overline{P}_j}{\overline{Y}_i} S_{i,j} \quad (17)$$

The Z matrix is then implemented within the estimability algorithm developed by Yao et al. [28] in order to rank the parameters. Moreover, the implementation of this method requires the definition of an estimability threshold whose value is chosen almost arbitrarily. The number of the most estimable parameters therefore depends on the chosen value of the threshold and thus makes the method less robust.

To overcome this arbitrary choice, Wu et al. [14] proposed a method based on the computation of the mean-square errors. The algorithm of their method is presented below.

- Rank the parameters from the most to the least estimable using the orthogonalization algorithm of Yao et al. [28] as detailed in (Appendix A):
- Identify the first estimable parameter by minimizing the square errors J^{mse} between the experimental measurements and the model predictions. Next, identify the first two estimable parameters, then the first three and so on until all the estimable parameters are identified. The least estimable parameters are set constant. The mean-square errors J^{mse} are computed as:

$$\min J^{mse} = \sum_{i=1}^n \sum_{j=1}^k \left(\frac{y_{i,j}^{exp} - y_{i,j}^{mod}}{\omega_i} \right)^2$$

where $y_{i,j}^{exp}$ and $y_{i,j}^{mod}$ are the experimental measurements and the model outputs respectively, and ω_i are weighing factors.

- Note the value of the objective function for the identification of the L most estimable parameters, while setting the remaining $d - L$ parameters constant.
- Compute the critical ratio $r_{c,L}$ for $L = 1, \dots, d$ as :

$$r_{c,L} = \frac{J_L^{mse} - J_d^{mse}}{d - L}$$

where J_L^{mse} defines the objective function for the identification of the L most estimable parameters. This critical ratio represents the ratio of the squared bias over the variance when L parameters are estimated rather than all d parameters.

- Compute the corrected critical ratio $r_{cc,L}$ as :

$$r_{cc,L} = \frac{d - L}{nk} (r_{cKub,L} - 1)$$

where:

$$r_{cKub,L} = \max \left(r_{c,L} - 1, \frac{2}{d - L + 2} r_{c,L} \right)$$

$r_{cc,L}$ is used to compare several simplified models with different numbers of parameters.

- Deduce the value of L which minimizes the $r_{cc,L}$. L corresponds to the number of estimable parameters to be identified from the available measurements.

Although the algorithm allows to avoid the arbitrary choice of the estimability threshold value, the parameter ranking is still dependent on the calculation method of the Z matrix. The latter changes with the initial values of the parameters used to compute its elements, i.e. local sensitivities. To overcome this problem, we propose to fill the Z matrix with global sensitivities computed using the Sobol approach.

3.2. Global sensitivity computation method

In sensitivity analysis, Sobol indices [9] are used to quantify the influence of an X_i model input on the variance of a model output Y . A popular variance-based method to calculate the total sensitivity index ST_i was developed by Homma and Saltelli [13]. The ST_i index accounts for the total contribution to the output variation due to the input X_i , i.e. its first-order effect plus all higher-order effects due to interactions. It is expressed as :

$$ST_i = 1 - \frac{\mathbf{E}(\mathbf{V}(Y|X_{\sim i}))}{\mathbf{V}(Y)} \quad (18)$$

where \mathbf{V} and \mathbf{E} are respectively the variance and the mathematical expectation of the output. $X_{\sim i}$ denotes all the model inputs except X_i . However, almost all the practitioners use the following algorithm to estimate total Sobol sensitivity indices :

$$\hat{ST}_i = 1 - \frac{U_{\sim i} - \hat{f}_0^2}{\hat{V}} \quad (19)$$

where:

$$U_{\sim i} = \frac{1}{N} \sum_{k=1}^N Y_{A,k} \cdot Y_{C_i,k} \quad (20)$$

$$\hat{f}_0 = \frac{1}{N} \sum_{k=1}^N Y_{A,k} \quad (21)$$

$$\hat{V} = \frac{1}{N} \sum_{k=1}^N Y_{A,k}^2 - \hat{f}_0^2 \quad (22)$$

and:

$$Y_{A,k} = f(X_{k,1}^{(A)}, \dots, X_{k,i}^{(A)}, \dots, X_{k,d}^{(A)}) \quad (23)$$

$$Y_{C_i,k} = f(X_{k,1}^{(A)}, \dots, X_{k,i}^{(B)}, \dots, X_{k,d}^{(A)}) \quad (24)$$

The input vectors $X_{i,j}^{(A)}$ and $X_{i,j}^{(B)}$ are generated by means of a Monte-Carlo sampling method and used as the elements of the following matrices A and B as :

$$A = \begin{pmatrix} X_{1,1}^{(A)} & \dots & X_{1,i}^{(A)} & \dots & X_{1,d}^{(A)} \\ \vdots & & \ddots & & \vdots \\ X_{N,1}^{(A)} & \dots & X_{N,i}^{(A)} & \dots & X_{N,d}^{(A)} \end{pmatrix}$$

$$B = \begin{pmatrix} X_{1,1}^{(B)} & \dots & X_{1,i}^{(B)} & \dots & X_{1,d}^{(B)} \\ \vdots & & \ddots & & \vdots \\ X_{N,1}^{(B)} & \dots & X_{N,i}^{(B)} & \dots & X_{N,d}^{(B)} \end{pmatrix}$$

These matrices are used in turn to generate the C_i matrices which are specific to each unknown parameter i as:

$$C_i = \begin{pmatrix} X_{1,1}^{(A)} & \dots & X_{1,i}^{(B)} & \dots & X_{1,d}^{(A)} \\ \vdots & & \ddots & & \vdots \\ X_{N,1}^{(A)} & \dots & X_{N,i}^{(B)} & \dots & X_{N,d}^{(A)} \end{pmatrix}$$

where N is the number of samples and d is the number unknown parameters. The method is further explained by Iooss et al. [10] and Saltelli et al. [11-13]. The Z sensitivity matrix is then filled using total sensitivity indices as in Rodriguez-Fernandez et al. [46] who used first the sensitivity indices to fill the Fisher Information matrix in order to develop an optimal design of experiments.

4. Results and discussions

4.1. Global sensitivity analysis

The results of the sensitivity analysis are investigated for two main purposes. The first objective is to ensure that all interactions between the ions considered in the aqueous medium are important. This is achieved by computing the global sensitivities of the 21 interaction parameters and comparing them to a cutoff criterion proposed by Zhang et al. [29] who assumed that a parameter with a global sensitivity greater than 0.05 must be necessarily taken into account. The second objective is to use these computed global sensitivity indices to fill the Z estimability matrix.

Figure 1 shows the Sobol global sensitivity indices of the measured model outputs, i.e. water activity and ions molalities. The sensitivities are computed using a Monte-Carlo sequence with random sampling to generate 10,000 uniformly distributed samples of size 21 (number of parameters) each. Taking into account the sensitivity computation algorithm described above, 230,000 simulations are carried out to calculate the global sensitivity indices. The CPU time needed to perform all the simulations is estimated to 64 hours using a Dell Precision T7810 Bi-Xeon 12 x Core 64GB work station.

It can be seen that the model outputs are very sensitive to the variation of the model parameters (Fig. 1). The elimination of one or more of these parameters would produce large uncertainties on the model outputs. The estimability analysis will therefore allow us to know which parameters are estimable from the available data and those whose values should be fixed either from the literature or from previous works.

4.2. Global estimability analysis results

The orthogonalization method proposed in [28] requires to fill the estimability matrix with the sensitivities of the unknown parameters. In several works [16, 18, 19, 42], the authors computed the elements of the Z matrix by means of a finite difference approximation which makes the method less powerful, and the reliability of the results depends in particular on the initial values of the parameters. Here, the global sensitivities computed previously are used to fill the Z-matrix.

Fig. 2A presents the estimability results of the 21 unknown parameters based on 56 water activity measured values at four different temperatures and 14 total sulfuric acid concentrations. In this case, the Z sensitivity matrix consists of 21 columns (parameters) and 56 data points (experiments). It is noteworthy that the magnitude of 13 parameters is very low, which may be due to the fact that the water activity alone does not contain enough information to estimate them. Moreover, most of the parameters that have a large magnitude value are temperature-independent. This is consistent with many literature works such as [25] where the authors used only water activity as experimental measurements and considered only the temperature-independent parameters in the modeling of similar binary electrolytic systems.

In Fig. 2B, the molalities of the different components are considered, i.e. H^+ , HSO_4^- , and SO_4 . They are measured in the same operating conditions as in the case of water activity with a total number of 168 measurements. In this case, the Z sensitivity matrix consists of 21 columns and 168 data points. The analysis of Fig.2B shows that the magnitudes of the unknown parameters

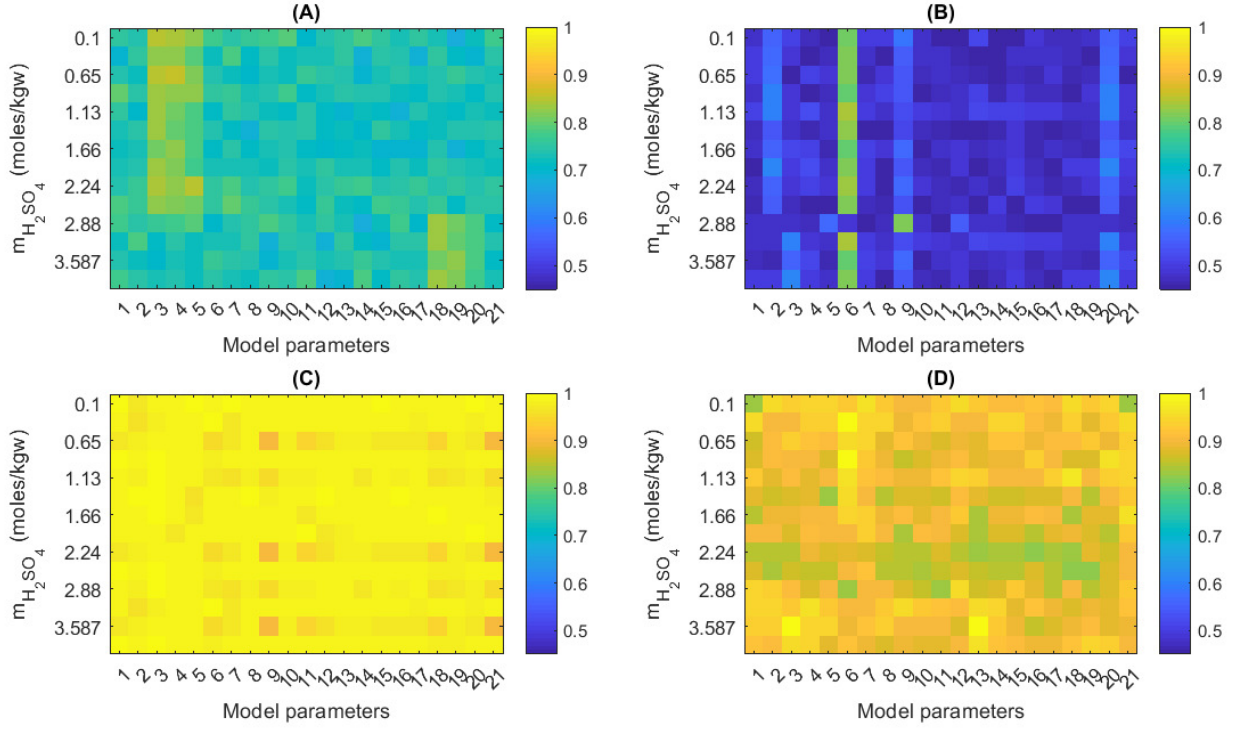


Figure 1: Sobolj's global sensitivity indices of the 21 model parameters computed using a set of 10,000 samples. (A), (B), (C) and (D) are the global sensitivities with respect to HSO_4^- molality, SO_4^{2-} molality, water activity and H^+ molality respectively.

are higher compared to the results of the first case, thus showing that this set of measurements contains more information to identify more unknown parameters.

Fig. 2C presents the case where the measurements of both water activity and components molalities are simultaneously taken into account. In this case, the Z sensitivity matrix consists of 21 columns and 224 data points. In Fig. 2C, all the magnitudes of the parameters are higher compared to the previous cases. This is consistent with the fact that by adding more measurements, the number of estimable parameters should in principle increase due to the increase of the level of information they contain.

4.3. Optimal value of the estimability threshold

Based on the previous results obtained for each of the three considered cases, sets of mean-square errors (mse) are defined and consist of the following elements:

$$J^{mse} = \sum_k \sum_T \sum_m \left(\frac{k_{m,T}^{exp} - k_{m,T}^{mod}}{\omega_i} \right)^2 \quad (25)$$

where k corresponds to the measured variable, i.e. water activity and component molalities, ω_i are weighing factors, T and m refer to the temperature and total concentration of sulfuric acid respectively.

The first mse set contains only the most estimable parameter as an unknown parameter, and all the other parameters are fixed at their nominal values. The second mse set contains the first and second most estimable parameters and so on until the last mse set where all the unknown parameters are estimated. Each mse set is minimized using GAMS optimization environment. More specifically, Baron optimizer based on a branch-and-reduce approach [21] is used to solve the identification problems to global optimality.

Figure 3 presents the effect of the number of estimated parameters on the value of the objective function, i.e. J^{mse} . It can be seen that as expected, the inclusion of more parameters in the identification process leads to a decrease in the value of the objective function. However, the main objective is to know which parameters are estimable in each of the considered cases. An optimal estimability threshold value for each case should therefore be determined. For this purpose, the corrected critical ratio $J_{cc,L}$ is computed and presented in Fig. 4 versus the objective function J^{mse} for each of the three sets of experimental measurements. The coordinates of the minimum of the curve correspond to the number of the most estimable parameters and to the optimal value of the estimability threshold respectively.

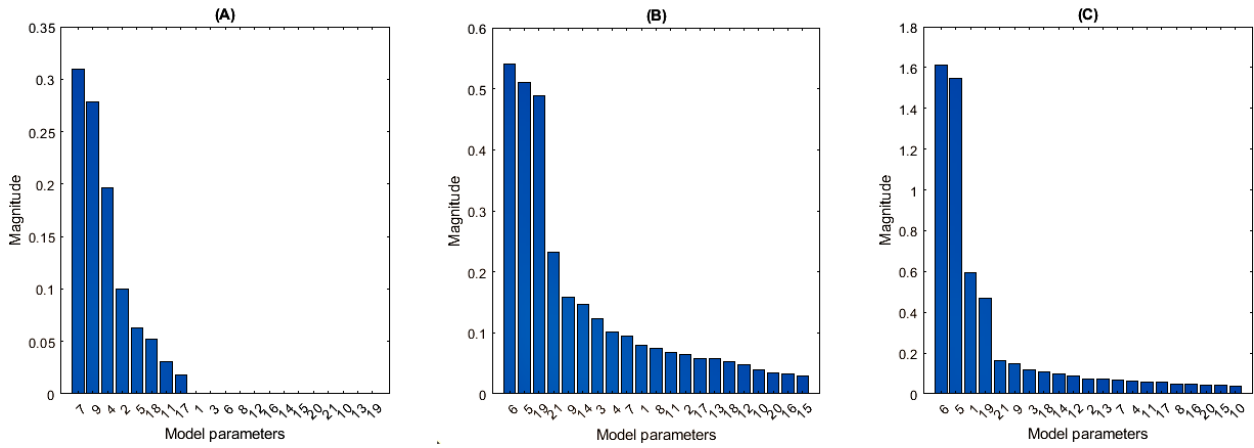


Figure 2: Ranking of the 21 unknown parameters using the orthogonalization method based on total sensitivity indices of Sobol. (A): case of water activity, (B): case of ions molalities, (C) case of water activity and ions molalities.

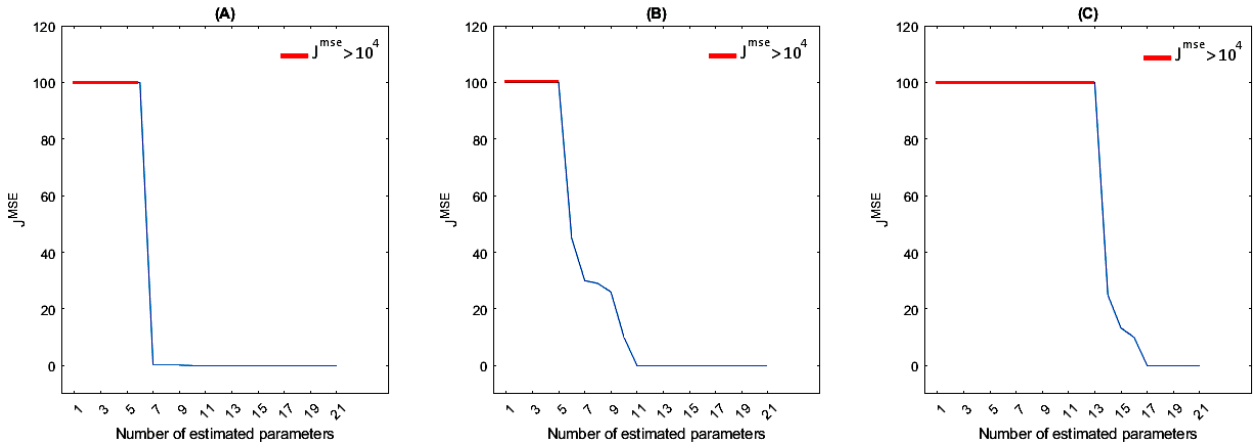


Figure 3: Value of J^{mse} for each mse case. (A): water activity, (B): molalities of components, (C) water activity and molalities of components.

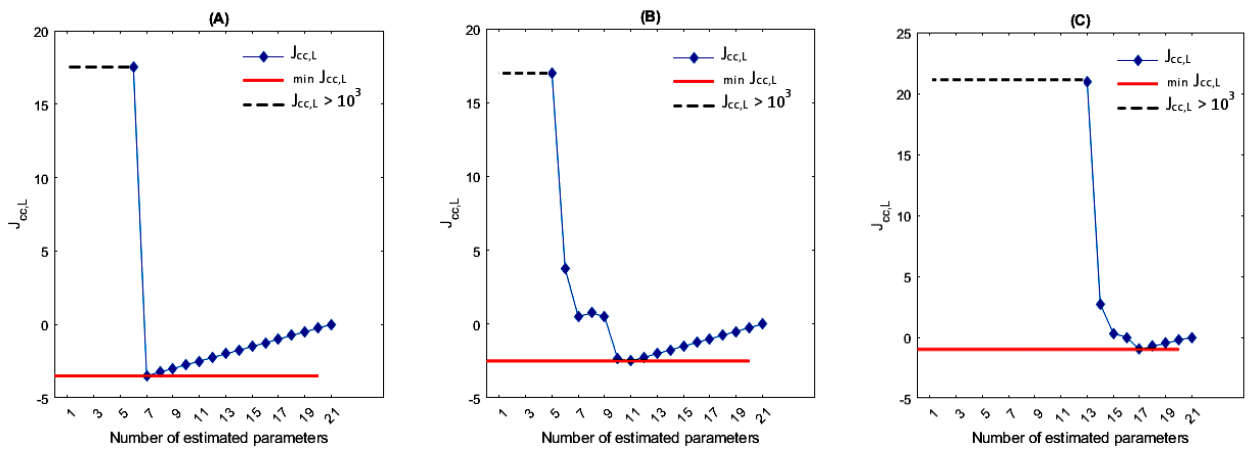


Figure 4: Value of $J_{cc,L}$ for each mse case. (A): water activity, (B): molalities of components, (C) water activity and molalities of components.

Fig. 4A shows that 7 parameters are estimable when only water activity measurements are used. The corresponding estimability threshold is about 0.03 (the ordinate of parameter 7 in Fig. 2A) meaning that for a parameter to be estimable, a variation of 10% of its value should cause at least a variation of 1.75% of the model outputs.

When only the measurements of components molalities are used, the number of the most estimable parameters as well as the corresponding value of the estimability threshold are higher than in the previous case and are equal to 11 and 0.07 respectively (Fig. 4B). This means that the components molalities contain more information than the water activity and the minimum variations of the model outputs caused by a 10% of an estimable parameter are about 2.65%.

Finally, as expected, when the two sets of measurements, i.e. water activity and components molalities, are simultaneously used, the number of the most estimable parameters is even higher and equals 17 (Fig. 4C). The corresponding estimability threshold value lies between the values of the two previous cases and is about 0.05. Likewise, the minimum variation of the model outputs caused by a 10% variation of an estimable parameter is between the values of the two previous cases and is equal to 2.24%.

It is noteworthy that unlike the algorithm of Yao et al. [28] which arbitrarily fixes the value of the estimability threshold at 0.04, the method of Wu et al. [14] allows us to determine the optimal threshold value and consequently the number of the most estimable parameters for each set of measurements. This is consistent with the fact that the level of information contained in the sets of measurements differs from one set to another.

In our case, the threshold values are respectively 0.03, 0.07 and 0.05 for the measurements of water activity, molalities of components, and water activity and molalities of components, and their corresponding number of estimable parameters are 7, 11 and 17 respectively. With the algorithm of Yao et al. [28], the number of estimable parameters is underestimated to 6 in the case of water activity measurements, and is overestimated to 18 and 21 in the two sets of measurements of molalities of components, and of water activity and molalities of components respectively. Although the number of estimable parameters is not optimal in the two cases, an overestimated number is more critical to the accuracy of parameter identification.

4.4. Identification of the most estimable parameters

The most estimable unknown parameters previously determined are then identified using the three sets of available measurements, and their accuracy is assessed by

means of confidence intervals computed below [26]. In parametric identification, the objective function can be expressed as:

$$F(P) = \theta(P)^T \cdot W \cdot \theta(P) \quad (26)$$

where $\theta(P)$ is the residue vector, i.e. the difference between the predictions of the model and the measured values of the outputs, W is a weighing matrix and P corresponds to the vector of the unknown parameters. Assuming that the measurement errors are normally distributed and independent, the covariance matrix C of the least squares problem is approximated as:

$$C \approx \frac{F(P^*)}{d-n} (J^T W J)^{-1} J^T W^2 J (J^T W J)^{-1} \quad (27)$$

where P^* is the vector of parameters which minimize the objective function $F(P)$, J is the Jacobian matrix of the $\theta(P)$ vector. This approximation is more precise when non-linearities are not strong. In this work, all the measurements have the same weight equal to 1 in the objective function, therefore W is equal to the identity matrix. The Jacobian matrix is then given by:

$$J = \begin{pmatrix} \frac{\partial \theta_1}{\partial P_1} & \cdots & \frac{\partial \theta_1}{\partial P_j} \\ \vdots & \ddots & \vdots \\ \frac{\partial \theta_i}{\partial P_1} & \cdots & \frac{\partial \theta_i}{\partial P_j} \end{pmatrix}$$

Subscripts i and j correspond to the number of outputs and the number of unknown parameters respectively. The Jacobian matrix is computed using a local one-at-a-time (OAT) method. It consists in disturbing the value of each parameter P_j^* by 10% forward and backward, then, the centred finite difference method is used to approximate the elements of J .

This requires 9,408 simulations in the case where the three sets of experimental measurements are simultaneously used (21 parameters x 224 measurements x 2), 7,056 simulations when only the molalities of components are used (21 parameters x 168 measurements x 2), and 2,352 simulations in the case where only water activity is used (21 parameters x 56 experience x 2).

The uncertainty on a parameter j is calculated as :

$$\epsilon_{P_j} = \pm \frac{\sqrt{c_{jj}} t_{1-\alpha/2, \nu}}{P_j^*} \cdot 100\% \quad (28)$$

where c_{jj} is the j^{th} diagonal element of the C matrix, $t_{1-\alpha/2, \nu}$ is deduced from the Student t-distribution with ν degrees of freedom and corresponds to the probability of $1 - \alpha/2$ that the true value of the parameter lies within the confidence interval given as:

$$P_j \in [P_j^* - \sqrt{c_{jj}} t_{1-\alpha/2, \nu}; P_j^* + \sqrt{c_{jj}} t_{1-\alpha/2, \nu}] \quad (29)$$

In this work the probability $1 - \alpha/2$ is taken equal to 95%.

On the other hand, the quality of the model predictions compared to the experimental measurements is quantified by means of the Pearson product-moment coefficient computed as [26]:

$$r = \frac{\sum_{i,j} (S_{ij}^{mod} - \overline{S_{ij}^{mod}}) (S_{ij}^{exp} - \overline{S_{ij}^{exp}})}{\sqrt{\left(\sum_{i,j} (S_{ij}^{mod} - \overline{S_{ij}^{mod}})^2\right) \left(\sum_{i,j} (S_{ij}^{exp} - \overline{S_{ij}^{exp}})^2\right)}} \quad (30)$$

where S_{ij}^{mod} is the model prediction, S_{ij}^{exp} is the corresponding experimental value and $\overline{S_{ij}}$ is the average value with respect to the total number of available measurements carried out at different molalities and temperatures. Subscripts i and j refer to the molalities and temperatures respectively.

The results of the identification of the most estimable parameters as well as their confidence intervals are presented in tables 3A-C. Table 3.A shows the results of the case where all the experimental measurements are considered. These results are used to compare the model outputs to the experimental measurements (Figs. 5A-E). The Pearson product-moment coefficient r is higher than 0.98 showing the good quality of the model predictions as well as the accuracy of the identification of parameters. Table 3.B shows the identified parameters along with their confidence intervals for the case where the molalities are used. The high value of the Pearson product-moment coefficient r points out the quality of the model predictions (Figs. 5G-J). Finally, the results of the third and last case where only water activity measurements are used are presented in Table 3.C. The model predictions are compared to the experimental measurements on Fig. 5F with a high value of r . Here again the model predictions are in very good agreement with the measurements.

It is noteworthy that although the confidence intervals are an approximation of the accuracy of the identified parameters, particularly when strong non-linearities are involved in the model, they show that the parameters are determined with good accuracy.

5. Conclusions

The estimability analysis approach developed in this work for the identification of the unknown parameters of Pitzer's model is based on the orthogonalization method developed by Yao et al. [28] and modified by Wu et al. [14]. The modification was necessary in order to determine the optimal value of the estimability threshold and therefore the number of the most estimable unknown parameters. In this paper, the modified algorithm was further improved by computing the elements of the sensitivity matrix Z by means of Sobol global sensitivity method. With this improvement, the initial values of the unknown parameters used in the calculation of local sensitivities no longer have any influence on the calculation of the elements of the sensitivity matrix.

The improved algorithm was implemented and applied to the thermodynamic modeling of sulfuric acid solutions. More specifically, three sets of experimental measurements of water activity and components molalities carried out at different values of temperature and sulfuric acid concentration were tested.

Although the results of the ranking of parameters are very promising, it should be noted that the efficiency of the approach developed in this paper can still be improved, in particular through some interesting issues which will be addressed in future works. Among them, the development of surrogate models as suggested by Sudret [8] using the Polynomial Chaos extensions, and the shift from random sampling to quasi-Monte Carlo sampling based on Sobol sequences, could considerably reduce the computation time which remains very high [44, 45].

Another important improvement is to use the Bayesian calibration method [47] which is fundamentally different from conventional calibration methods where the difference between the observed data and the model outputs is minimized. Bayesian calibration is an iterative process where the uncertainty distributions of the model parameters are updated consistently with the observed data. The main advantages of this method are: (i) it accounts for all sources of uncertainty, i.e. parameter uncertainty, data uncertainty, and uncertainty in the structure of the model, (ii) and the posterior probabilities are a natural measure of uncertainty. In addition, a practical by-product of Bayesian calibration is an estimate of the best fit to experimental data.

Table 3: Unknown interaction parameters (298 K to 353K) and their confidence intervals: parameters without confidence intervals are not estimable, their values are fixed from previous identifications or from literature [32]. a , b and c are the coefficients of Eqs. (13-15).

(A)						
	a	C.I	b	C.I	c	C.I
$\beta_{HSO_4^-, SO_4^{2-}}^{(0)}$	-2.471	[-3.00;-1.934]	0.007	[0.004;0.010]	0.01	-
$\beta_{HSO_4^-, H^+}^{(0)}$	1.046	[0.518;1.573]	-0.002	[-0.003;-0.001]	0.01	-
$\beta_{SO_4^{2-}, H^+}^{(0)}$	-8.001	[-9.665;-6.336]	0.01	-	1684.0	[1221.1;2147.04]
$\beta_{HSO_4^-, SO_4^{2-}}^{(1)}$	-254.8	[-312.8;-196.8]	0.405	[0.341;0.468]	$3.9 \cdot 10^4$	$[3.1 \cdot 10^4; 4.8 \cdot 10^4]$
$\beta_{HSO_4^-, H^+}^{(1)}$	-296.9	[-489.1;-104.8]	0.453	[0.341;0.564]	$4.8 \cdot 10^4$	$[2.7 \cdot 10^4; 7.0 \cdot 10^4]$
$\beta_{SO_4^{2-}, H^+}^{(1)}$	8.039	[4.879;11.19]	-0.023	[-0.032;-0.013]	0.01	-
$\psi_{HSO_4^-, SO_4^{2-}, H^+}$	1.696	[1.142;2.249]	-0.003	[-0.004;-0.001]	-277.209	[-335.9;-218.4]
(B)						
	a	C.I	b	C.I	c	C.I
$\beta_{HSO_4^-, SO_4^{2-}}^{(0)}$	-4.449	[-5.716;-3.181]	0.011	[0.009;0.013]	0.01	-
$\beta_{HSO_4^-, H^+}^{(0)}$	1.046	-	-0.002	[-0.003;-0.001]	0.01	-
$\beta_{SO_4^{2-}, H^+}^{(0)}$	-7.533	[-9.881;-5.184]	0.01	-	1684.0	-
$\beta_{HSO_4^-, SO_4^{2-}}^{(1)}$	-254.7	[-323.1;-186.2]	0.405	-	$3.9 \cdot 10^4$	-
$\beta_{HSO_4^-, H^+}^{(1)}$	-288.3	[-383.8;-192.8]	0.453	-	$4.5 \cdot 10^4$	$[2.7 \cdot 10^4; 6.4 \cdot 10^4]$
$\beta_{SO_4^{2-}, H^+}^{(1)}$	1.053	[0.654;1.451]	-0.023	-	0.01	-
$\psi_{HSO_4^-, SO_4^{2-}, H^+}$	1.155	[0.812;1.497]	-0.002	[-0.003;-0.001]	-173.5	[-228.3;-118.8]
(C)						
	a	C.I	b	C.I	c	C.I
$\beta_{HSO_4^-, SO_4^{2-}}^{(0)}$	-2.47	-	0.007	-	0.01	-
$\beta_{HSO_4^-, H^+}^{(0)}$	1.445	[0.828;2.062]	$-3 \cdot 10^{-3}$	$[-4.24 \cdot 10^{-3}; -1.76 \cdot 10^{-3}]$	0.01	-
$\beta_{SO_4^{2-}, H^+}^{(0)}$	-8.001	-	0.01	-	1684.0	-
$\beta_{HSO_4^-, SO_4^{2-}}^{(1)}$	-1.028	[-1.309;-0.747]	$4.44 \cdot 10^{-5}$	$[2.03 \cdot 10^{-5}; 6.83 \cdot 10^{-5}]$	4.414	[1.190;7.637]
$\beta_{HSO_4^-, H^+}^{(1)}$	-297.5	[-372.9;-222.2]	0.453	-	$4.87 \cdot 10^4$	-
$\beta_{SO_4^{2-}, H^+}^{(1)}$	8.039	-	-0.023	-	0.01	-
$\psi_{HSO_4^-, SO_4^{2-}, H^+}$	1.827	[1.027;2.627]	-0.003	-	-277.2	-

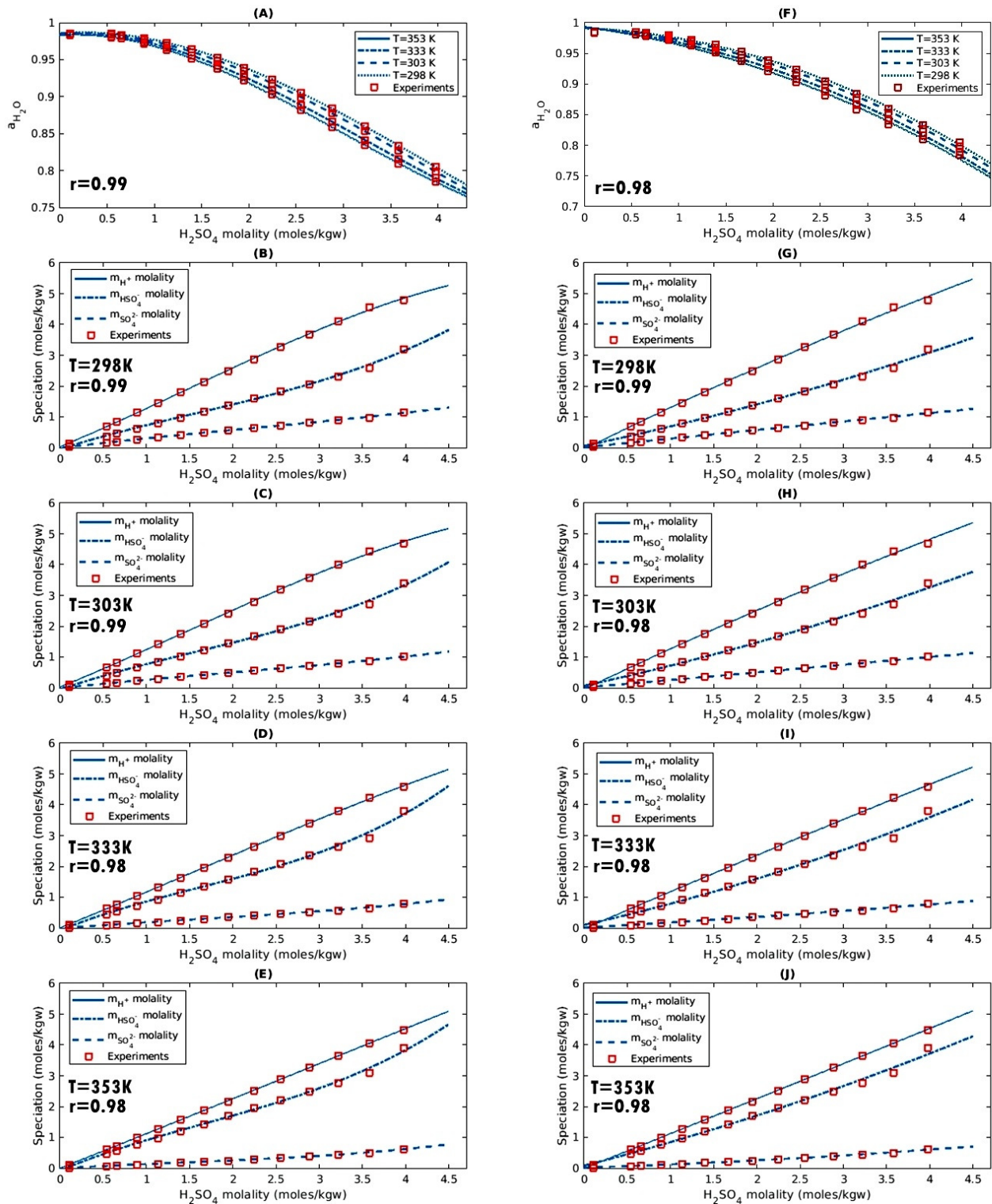


Figure 5: Model predictions compared to the experimental measurements, (A-E): case of water activity and molalities of components, (F): case of water activity, (G-J): case of molalities of components

Appendix A

The orthogonalization algorithm developed by Yao et al. [28] is based on the sensitivity matrix Z , and is presented below:

1. Calculate the magnitude of each column of Z .
2. Select the parameter whose column in Z has the largest magnitude (the sum of squares of the elements) as the first estimable parameter.
3. Mark the corresponding column as X_L ($L = 1$ for the first iteration).
4. Calculate Z^L the prediction of the full sensitivity matrix Z , using the subset of columns X_L :

$$Z^L = X_L(X_L^T X_L)^{-1} X_L^T Z \quad (\text{A.1})$$

5. Compute the residual matrix R_L as:

$$R_L = Z - Z^L \quad (\text{A.2})$$

6. Calculate the sum of squares of the residuals in each column of R_L . The column with the largest magnitude corresponds to the next estimable parameter.
7. Select the corresponding column in Z , and augment the matrix X_L by including the new column. Denote the augmented matrix as X_{L+1} .
8. Advance the iteration counter by one and repeat steps 4 to 7 until the column of largest magnitude in the residual matrix is smaller than a prescribed cut-off value.

Acknowledgments

The authors are grateful to OCP Group for its financial support.

Declaration of interests

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

References

- [1] Pitzer, K. S., Roy, R. N., & Silvester, L. F. (1977). Thermodynamics of electrolytes. 7. Sulfuric acid. Journal of the American Chemical Society, 99(15), 4930-4936. <https://doi.org/10.1021/ja00457a008>
- [2] Clegg, S. L., Rard, J. A., & Pitzer, K. S. (1994). Thermodynamic properties of 0-6 mol kg⁻¹ aqueous sulfuric acid from 273.15 to 328.15 K. Journal of the Chemical Society, Faraday Transactions, 90(13), 1875-1894. <https://doi.org/10.1039/FT9949001875>
- [3] Zeleznik, F. J. (1991). Thermodynamic properties of the aqueous sulfuric acid system to 350 K. Journal of Physical and Chemical Reference Data, 20(6), 1157-1200. <https://doi.org/10.1063/1.555899>
- [4] Que, H., Song, Y., & Chen, C. C. (2011). Thermodynamic modeling of the sulfuric acid water sulfur trioxide system with the symmetric electrolyte NRTL model. Journal of Chemical & Engineering Data, 56(4), 963-977. <https://doi.org/10.1021/je100930y>
- [5] Sippola, H. (2012). Thermodynamic modelling of concentrated sulfuric acid solutions. Calphad, 38, 168-176. <https://doi.org/10.1016/j.calphad.2012.06.008>
- [6] Sippola, H., & Taskinen, P. (2014). Thermodynamic properties of aqueous sulfuric acid. Journal of Chemical & Engineering Data, 59(8), 2389-2407. <https://doi.org/10.1021/je4011147>
- [7] Helgeson, H. C., & Kirkham, D. H. (1974). Theoretical prediction of the thermodynamic behavior of aqueous electrolytes at high pressures and temperatures; II, Debye-Huckel parameters for activity coefficients and relative partial molal properties. American Journal of Science, 274(10), 1199-1261. <https://doi.org/10.2475/ajs.274.10.1199>
- [8] Sudret, B. (2008). Global sensitivity analysis using polynomial chaos expansions. Reliability Engineering & System Safety, 93(7), 964-979. <https://doi.org/10.1016/j.res.2007.04.002>
- [9] Sobol, I. M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Mathematics and Computers in Simulation, 55(1-3), 271-280. [https://doi.org/10.1016/S0378-4754\(00\)00270-6](https://doi.org/10.1016/S0378-4754(00)00270-6)
- [10] Iooss, B., & Lemaître, P. (2015). A review on global sensitivity analysis methods. In Uncertainty management in simulation-optimization of complex systems (pp. 101-122). Springer, Boston, MA. <https://doi.org/10.1007/978-1-4899-7547-85>
- [11] Saltelli, A. (2002). Making best use of model evaluations to compute sensitivity indices. Computer physics communications, 145(2), 280-297. [https://doi.org/10.1016/S0010-4655\(02\)00280-1](https://doi.org/10.1016/S0010-4655(02)00280-1)
- [12] Saltelli, A., Tarantola, S., Campolongo, F., & Ratto, M. (2004). Sensitivity analysis in practice: a guide to assessing scientific models. New York: Wiley. <https://doi.org/10.1111/j.1467-985X.2005.35813.x>
- [13] Homma, T., & Saltelli, A. (1996). Importance measures in global sensitivity analysis of nonlinear models. Reliability Engineering & System Safety, 52(1), 1-17. [https://doi.org/10.1016/0951-8320\(96\)00002-6](https://doi.org/10.1016/0951-8320(96)00002-6)
- [14] Wu, S., McLean, K. A., Harris, T. J., & McAuley, K. B. (2011). Selection of optimal parameter set using estimability analysis and MSE-based model-selection criterion. International Journal of Advanced Mechatronic Systems, 3(3), 188-197. <https://doi.org/10.1504/IJAMECHS.2011.042615>
- [15] Eghtesadi & McAuley, K. B. (2014). Mean square error based method for parameter ranking and selection to obtain accurate predictions at specified operating conditions. Industrial & Engineering Chemistry Research, 53, 60336046. <https://doi.org/10.1021/ie5002444>
- [16] Lei, M., Vallières, C., Grévillet, G., & Latifi, M. A. (2013). Thermal swing adsorption process for carbon dioxide capture and recovery: modeling, simulation, parameters estimability, and identification. Industrial & Engineering Chemistry Research, 52(22), 7526-7533. <https://doi.org/10.1021/ie3029152>
- [17] sw.aveva.com/engineer-procure-construct/process-engineering-and-simulation/pro-ii-process-engineering.
- [18] Ngo, V. V., Michel, J., Gujisaite, V., Latifi, A., & Simonnot, M. O. (2014). Parameters describing nonequilibrium transport of polycyclic aromatic hydrocarbons through contaminated soil columns: Estimability analysis, correlation, and optimization. Journal of Contaminant Hydrology, 158, 93-109. <https://doi.org/10.1016/j.jconhyd.2014.01.005>
- [19] Ngo, V. V., Latifi, A., & Simonnot, M. O. (2013). Soil hydraulic parameters characterizing preferential water flow:

- estimability analysis and identification. *Journal of Hydrologic Engineering*, 19(10), 04014017.
[https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000953](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000953)
- [20] Simoes, M. C., Hughes, K. J., Ingham, D. B., Ma, L., & Pourkashanian, M. (2017). Temperature Dependence of the Parameters in the Pitzer Equations. *Journal of Chemical & Engineering Data*, 62(7), 2000-2013.
<https://doi.org/10.1021/acs.jced.7b00022>
- [21] Sahinidis, N. V. (1996). BARON: A general purpose global optimization software package. *Journal of Global Optimization*, 8(2), 201-205.
<https://doi.org/10.1007/BF00138693>
- [22] Peleg, M., Normand, M. D., & Corradini, M. G. (2012). The Arrhenius equation revisited. *Critical Reviews in Food Science and Nutrition*, 52(9), 830-851.
<https://doi.org/10.1080/10408398.2012.667460>
- [23] Gustafsson, J. P. (2011). Visual MINTEQ 3.0 user guide. KTH, Department of Land and Water Resources, Stockholm, Sweden.
- [24] Pitzer, K. S. (2018). *Activity Coefficients in Electrolyte Solutions: 0*. CRC press.
- [25] Jiang, C. (1996). Thermodynamics of aqueous phosphoric acid solution at 25 C. *Chemical Engineering Science*, 51(5), 689-693.
[https://doi.org/10.1016/0009-2509\(95\)00284-7](https://doi.org/10.1016/0009-2509(95)00284-7)
- [26] Keith, T. Z. (2014). *Multiple regression and beyond: An introduction to multiple regression and structural equation modeling*. Routledge.
<https://doi.org/10.1111/jedm.12108>
- [27] Pitzer, K. S., & Mayorga, G. (1993). Thermodynamics of Electrolytes.: II. Activity and Osmotic Coefficients for Strong Electrolytes with One or Both Ions Univalent. In *Molecular Structure And Statistical Thermodynamics: Selected Papers of Kenneth S Pitzer* (pp. 396-404).
<https://doi.org/10.1142/97898127959600058>
- [28] Yao, K. Z., Shaw, B. M., Kou, B., McAuley, K. B., & Bacon, D. W. (2003). Modeling ethylene/butene copolymerization with multi-site catalysts: parameter estimability and experimental design. *Polymer Reaction Engineering*, 11(3), 563-588.
<https://doi.org/10.1081/PRE-120024426>
- [29] Zhang, X. Y., Trame, M. N., Lesko, L. J., & Schmidt, S. (2015). Sobol sensitivity analysis: a tool to guide the development and evaluation of systems pharmacology models. *CPT: Pharmacometrics & Systems Pharmacology*, 4(2), 69-79.
<https://doi.org/10.1002/psp4.6>
- [30] Benyahia, B., Latifi, M. A., Fonteix, C., & Pla, F. (2013). Emulsion copolymerization of styrene and butyl acrylate in the presence of a chain transfer agent. Part 2: Parameters estimability and confidence regions. *Chemical Engineering Science*, 90, 110-118.
<https://doi.org/10.1016/j.ces.2012.12.013>
- [31] Lei, M., Lesage, F., Latifi, M. A., & Tretjak, S. (2015). Modeling and parameter estimation of coke combustion kinetics in a glycerol catalytic conversion reactor. In *Computer Aided Chemical Engineering 37*, 335-340. Elsevier.
<https://doi.org/10.1016/B978-0-444-63578-5.50051-7>
- [32] Parkhurst, D. L. (1995). User's guide to PHREEQC: A computer program for speciation, reaction-path, advective-transport, and inverse geochemical calculations (Vol. 95, No. 4227). US Department of the Interior, US Geological Survey.
- [33] McCleskey, R. B., Nordstrom, D. K., Ryan, J. N., & Ball, J. W. (2012). A new method of calculating electrical conductivity with applications to natural waters. *Geochimica et Cosmochimica Acta*, 77, 369-382.
<https://doi.org/10.1016/j.gca.2011.10.031>
- [34] www.aspentech.com/en/products/engineering/aspens-hysys.
- [35] Velez-Reyes, M., & Verghese, G. C. (1995, September). Subset selection in identification, and application to speed and parameter estimation for induction machines. In *Proceedings of International Conference on Control Applications* (pp. 991-997). IEEE.
<https://doi.org/10.1109/CCA.1995.555890>
- [36] Weijers, S. R., & Vanrolleghem, P. A. (1997). A procedure for selecting best identifiable parameters in calibrating activated sludge model no. 1 to full-scale plant data. *Water Science and Technology*, 36(5), 69-79.
[https://doi.org/10.1016/S0273-1223\(97\)00463-0](https://doi.org/10.1016/S0273-1223(97)00463-0)
- [37] Brun, R., Kühni, M., Siegrist, H., Gujer, W., & Reichert, P. (2002). Practical identifiability of ASM2d parameters—systematic selection and tuning of parameter subsets. *Water Research*, 36(16), 4113-4127.
[https://doi.org/10.1016/S0043-1354\(02\)00104-5](https://doi.org/10.1016/S0043-1354(02)00104-5)
- [38] Degenring, D., Froemel, C., Dikta, G., & Takors, R. (2004). Sensitivity analysis for the reduction of complex metabolism models. *Journal of Process Control*, 14(7), 729-745.
<https://doi.org/10.1016/j.jprocont.2003.12.008/>
- [39] Chu, Y., & Hahn, J. (2007). Parameter set selection for estimation of nonlinear dynamic systems. *American Institute of Chemical Engineers journal*, 53(11), 2858-2870.
<https://doi.org/10.1002/aic.11295>
- [40] Chu, Y., & Hahn, J. (2009). Parameter set selection via clustering of parameters into pairwise indistinguishable groups of parameters. *Industrial & Engineering Chemistry Research*, 48(13), 6000-6009.
<https://doi.org/10.1021/ie800432s>
- [41] Chu, Y., Huang, Z., & Hahn, J. (2009). Improving prediction capabilities of complex dynamic models via parameter selection and estimation. *Chemical Engineering Science*, 64(19), 4178-4185.
<https://doi.org/10.1016/j.ces.2009.06.057>
- [42] Bedel, S., Vallières, C., & Latifi, M. A. (2017). Parameters estimability analysis and identification for adsorption equilibrium models of carbon dioxide. *Adsorption*, 23(2-3), 373-380.
<https://doi.org/10.1007/s10450-017-9864-7>
- [43] Saltelli, A., Tarantola, S., Campolongo, F., & Ratto, M. (2004). *Sensitivity analysis in practice: a guide to assessing scientific models* (Vol. 1). New York: Wiley.
https://doi.org/10.1111/j.1467-985X.2005.358_13.x
- [44] Kucherenko, S., Feil, B., Shah, N., & Mauntz, W. (2011). The identification of model effective dimensions using global sensitivity analysis. *Reliability Engineering & System Safety*, 96(4), 440-449.
<https://doi.org/10.1016/j.res.2010.11.003>
- [45] Kucherenko, S., Albrecht, D., & Saltelli, A. (2015). Exploring multi-dimensional spaces: A comparison of Latin hypercube and quasi-Monte Carlo sampling techniques. *arXiv preprint arXiv:1505.02350*.
- [46] Rodriguez-Fernandez, M., Kucherenko, S., Pantelides, C., Shah, N. (2007). Optimal experimental design based on global sensitivity analysis. In *Computer Aided Chemical Engineering* (Vol. 24, pp. 63-68). Elsevier.
[https://doi.org/10.1016/S1570-7946\(07\)80034-4](https://doi.org/10.1016/S1570-7946(07)80034-4)
- [47] Smith, R. C. (2013). *Uncertainty quantification: theory, implementation, and applications* (Vol. 12). Siam.