Bayesian Inference Applied to Estimating the Uncertainty in Phase-3 and Phase-5 Log K Values

Martin B. Nemer[1], Paul Domski[2], Habib Najm[3]

[1]*Sandia National Laboratories, Albuquerque, NM*

email: mbnemer@sandia.gov

[2]*Sandia National Laboratories, Carlsbad, NM*

[3]*Sandia National Laboratories, Livermore, CA*

INTRODUCTION

Bayesian inference is a statistical technique that has gained popularity in recent years due to the availability of open-source code packages and computing power. Bayesian inference has been applied to various inverse problems including those in aqueous thermodynamics using the Pitzer EOS [1]. A notable quote from [1] reads "*A considerable amount of research has been reported on the forward propagation of assumed uncertainty in the chemical equilibrium models in geochemistry and geophysics; but the inverse problem of uncertainty quantification in the chemical equilibrium modeling has not received much attention*."

Bayesian inference is grounded in Bayes rule, a statement on conditional probability [2], and its use for statistical parameter estimation or model calibration given data. Written in terms of probability density functions (PDFs), Bayes rule in this context provides a relationship between the PDF of the data given model parameters and the PDF of model parameters given the data,

$$p(\theta \mid y)p(y) = p(y \mid \theta)p(\theta), \qquad (1)$$

where y is the experimental data that is being fit, $p(y|\theta)$ is the PDF of the data for given model parameters θ, $p(\theta|y)$ is the PDF of the model parameters for a given set of data y, and $p(y)$ and $p(\theta)$ are the marginalized PDFs of the data and parameters, respectively. Bayesian statistics [2] assigns a meaning to $p(\theta)$ as the prior PDF, which represents our prior knowledge on the model parameters θ, and $p(y)$ as the evidence for the model [2]. Then $p(\theta|y)$ is the posterior PDF, which represents our knowledge of the model parameters given the data y. The PDF $p(y|\theta)$ is the likelihood of the data y for given model parameters θ.

Once the posterior PDF $p(\theta|y)$ is determined, a predicted dataset y_{pred} can be calculated via the posterior predictive PDF $p(y_{pred}|y_{meas})$ and compared to the experimentally obtained dataset y_{meas}. This is obtained by integrating the product of the posterior and the likelihood over θ,

$$p(y_{pred} \mid y_{meas}) = \int p(\theta \mid y_{meas})\, p(y_{pred} \mid \theta)\, d\theta. \qquad (2)$$

A key element in this analysis is the development of the likelihood model. In aqueous thermodynamics, it has often been observed [3] that the difference between experimental data and an adequate equilibrium thermodynamic model is normally distributed on a log scale,

$$\log_{10}\left(m_{meas}\right) - \log_{10}\left(m_{\mathrm{mod}}(\xi,\theta)\right) \approx \varepsilon(\xi), \qquad (3)$$

$$\varepsilon(\xi) \sim N\left(\mu(\xi), \sigma^2(\xi)\right), \qquad (4)$$

where m_{meas} is the experimentally measured concentration (herein molality) of some ion in solution and $m_{mod}(\xi, \theta)$ is the modeled concentration of the same ion in solution obtained from an equilibrium thermodynamics code with model parameters θ and independent variables (chemical conditions) ξ. Neglecting $\mu(\xi)$, and presuming independence of $\varepsilon(\xi_i)$ for $i=1,\ldots,n$ separate experimental measurements with chemical conditions ξ_i, equation (5) provides the basis for an additive independent identically distributed (*i.i.d.*) Gaussian likelihood model in terms of the sum of square error between the model and the experiment

$$p(y_{meas} \mid \theta, \sigma) = \left(2\pi\sigma^2\right)^{-\frac{nk}{2}} \exp\left[-\frac{1}{2\sigma^2} S(\theta)\right], \qquad (5)$$

where $S(\theta)$ as the sum of square error between the two terms on the left hand side of equation (4), n is the number of experimental measurements, and k is the number of ions being included in the model.

DESCRIPTION OF THE WORK

This talk will give a demonstration of the above Bayesian inference methodology as applied to the problem of parameter estimation for determining the log K values for phase-3 and phase-5 in aqueous systems using the Pitzer equation of state. This abstract and presentation have been condensed from a full report [4].

RESULTS

The phase-3 experimental data used for this analysis came from two semi-independent sources [5, 6]. The experimental data used for phase-5 came from only one study [7]. The results of the analysis are shown in Fig. 1 and Fig. 2 for phase-3; for phase-5, the results are given in [4] and will be given in the presentation. The two datasets in [5, 6] were simultaneously fit by the method

described in [4] but measured using different techniques and target ions, hence the need to plot the model/experiment comparisons separately for each dataset. The results show that the posterior predictive distribution $p(y_{pred}|y_{meas})$ captures nearly all of the experimental data at the 95th % credible interval.

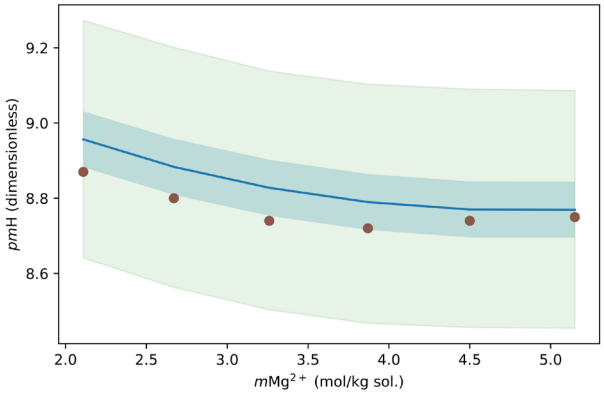


Fig. 1: Plot of data from [5] (brown dots), plot using the mean value of $\log_{10}K$ from the posterior distribution (solid blue line); the blue shaded region is the region that spans the results of the model within the 95th % credible interval on the marginal posterior distribution for $\log_{10}K$; green shaded area is the region that spans the 95th % credible interval on the posterior predictive distribution.

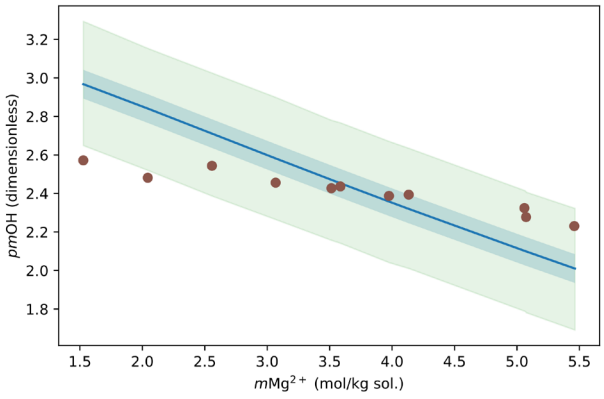


Fig. 2: Plot of the experimental data from [6] data (brown dots); plot of the model using the mean value of $\log_{10}K$ from the posterior distribution (solid blue line); the blue shaded region is the region that spans the results of the model within the 95th % credible interval on the marginal posterior distribution for $\log_{10}K$; the green shaded area is the region that spans the 95th % credible interval on the posterior predictive distribution.

CONCLUSION

It is the authors' opinion that the evidence presented supports the use of Bayesian inference for estimating uncertainty in thermodynamic geochemical model parameters and uncertainty in thermodynamic geochemical prediction.

When the model-form is uncertain and multiple potential models have been identified, Bayesian Model Averaging [8] can be utilized to produce a mean posterior prediction with uncertainty based on the underlying uncertainty in each model.

FUNDING STATEMENT

Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525. This research is funded by WIPP programs administered by the Office of Environmental Management (EM) of the U.S. Department of Energy.

REFERENCES

1. Sarkar, S., et al., "Bayesian calibration of thermodynamic parameters for geochemical speciation modeling of cementitious materials" *Cement and Concrete Research*, **42**, Page 889-902 (2012).

2. Silva, D.S., "Data Analysis; A Bayesian Tutorial". Second Edition ed. Norfolk, Great Britain: Oxford University Press. (2007).

3. Ahrens, L.H., "A Fundamental Law Of Geochemistry" *Nature*, **172**, Page 1148-1148 (1953).

4. Nemer, M.B., et al., "Bayesian Inference applied to the Solubility of Phase-3 and Phase-5", Sandia National Laboratories: Albuquerque, NM. SAND2023-01165R (2023).

5. Altmaier, M., et al., "Solid-liquid equilibria of Mg(OH)(2)(cr) and Mg-2(OH)(3)Cl center dot 4H(2)O(cr) in the system Mg-Na-H-OH-O-Cl-H2O at 25 degrees C" *Geochimica Et Cosmochimica Acta*, **67**, Page 3595-3601 (2003).

6. Pannach, M., et al., "Solubility Equilibria in the System $Mg(OH)_2$–$MgCl_2$–HO from 298 to 393 K" *Journal of Chemical & Engineering Data*, **62**, Page 1384-1396 (2017).

7. Xiong, Y.L., et al., "Experimental determination of the solubility constant for magnesium chloride hydroxide hydrate (Mg3Cl(OH)(5)center dot 4H(2)O, phase 5) at room temperature, and its importance to nuclear waste isolation in geological repositories in salt formations" *Geochimica Et Cosmochimica Acta*, **74**, Page 4605-4611 (2010).

8. Fragoso, T.M., et al., "Bayesian Model Averaging: A Systematic Review and Conceptual Classification" *International Statistical Review*, **86**, Page 1-28 (2018).