


Exploring uncertainty in thermodynamic modeling of the lead carbonate aqueous system

C. Kushnir and C. E. Robinson *

Department of Civil and Environmental Engineering, University of Western Ontario, 1151 Richmond St, London, ON N6A 3K7, Canada

*Corresponding author. E-mail: crobinson@uwo.ca

 CER, 0000-0002-6156-2328

ABSTRACT

Thermodynamic models of the lead carbonate aqueous system have been used in prior studies to evaluate the effect of water chemistry on lead solubility and corrosion scale composition in water distribution systems. A common challenge in these studies is uncertainty arising from the thermodynamic equilibrium constants ($\log K$ values) used to parameterize the models. The objective of this study is to evaluate the way in which uncertainty in $\log K$ values propagates through thermodynamic models of the lead carbonate system and provide guidance for future modeling efforts. This was done by conducting a full factorial statistical analysis implemented using a custom Python (v3) code coupled with a PHREEQC (v1.4.2) geochemical model, along with batch lead solubility experiments. Two lead carbonate solid phases (cerussite and hydrocerussite) and nine aqueous lead complexes are considered in the geochemical model with conditions simulated ranging from pH 4 to 11 and dissolved inorganic carbon (DIC) ranging from 0 to 250 mg C/L. Main effect analysis indicates that model uncertainty is predominately associated with $\log K$ values for five species (in order of decreasing effect): hydrocerussite, $\text{Pb}(\text{CO}_3)_2^{-2}$, cerussite, PbOH^+ , and PbCO_3^0 , with model uncertainty varying depending on the specific pH and DIC conditions simulated. Interaction effects show that $\log K$ values cannot be selected independently, as their influence on lead solubility is connected. Finally, by considering uncertainty in $\log K$ values it was possible for the thermodynamic model to match batch hydrocerussite solubility experimental data over a range of pH and DIC conditions. Six combinations of $\log K$ values that provided a good match between the simulated and experimental data were determined with the average difference between the simulated and experimental lead concentrations calculated to be 0.031 mg/L when the recommended $\log K$ value combination was used.

Key words: cerussite, drinking water, hydrocerussite, lead, thermodynamic model, uncertainty

HIGHLIGHTS

- Thermodynamic model uncertainty is predominately associated with $\log K$ values for five species (in order of decreasing effect): hydrocerussite, $\text{Pb}(\text{CO}_3)_2^{-2}$, cerussite, PbOH^+ , and PbCO_3^0 .
- $\log K$ values for species cannot be selected independently, as their influence on lead solubility is interconnected.
- Thermodynamic model uncertainty varies depending on specific pH and dissolved inorganic carbon conditions simulated.

INTRODUCTION

Lead is a pervasive toxic metal that can exist in the home, air, soil, or water. Industries focused on the removal or control of lead include wastewater treatment, battery recycling, and soil remediation. Furthermore, lead-bearing plumbing materials in drinking water distribution systems are common. The release of lead from corrosion scales that form inside these plumbing materials can cause elevated lead concentrations in drinking water at the tap. While many municipalities have programs for the replacement of lead-bearing plumbing materials, these programs are costly and progress slowly (Sandvig *et al.* 2008). As such corrosion control strategies are often implemented to reduce lead concentrations in drinking water. Water chemistry is generally the most accessible parameter for water utility operators to adjust to control corrosion and thus reduce lead concentrations. For instance, parameters such as alkalinity (or dissolved inorganic carbon, DIC) (Schock 1980; Marani *et al.* 1995; Tam & Elefsiniotis 2009; Wang *et al.* 2012); pH (Tam & Elefsiniotis 2009; Kim *et al.* 2011; Wang *et al.* 2012); additives such as orthophosphate (Gouider *et al.* 2009; Tam & Elefsiniotis 2009; Ng *et al.* 2012); and disinfectant type (i.e., free chlorine and chloramines) (Edwards & Dudi 2004; Vasquez 2005; Switzer *et al.* 2006; Lin & Valentine 2008; Lin & Valentine 2009) can be adjusted to stabilize

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY-NC-ND 4.0), which permits copying and redistribution for non-commercial purposes with no derivatives, provided the original work is properly cited (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

lead-bearing corrosion scales. As conditions vary between distribution systems, there is a need to be able to predict *a priori* the effect water chemistry adjustments may have on lead corrosion scale stability and thus lead concentrations.

The composition of the corrosion scale is important for determining the effectiveness of a corrosion control strategy (Docherty & Kariuki 2019). The main groups of phases in lead-bearing corrosion scales are lead (II) carbonate and lead (IV) oxide solids (Wang *et al.* 2012). The major lead (II) carbonates are hydrocerussite ($\text{Pb}_3(\text{CO}_3)_2(\text{OH})_2$) and cerussite (PbCO_3), while the major lead (IV) oxides are plattnerite (PbO_2) and scrutinyite (PbO_2) (Wang *et al.* 2012). The characterizations of experimental pipe loops and exhumed lead pipes have shown that lead (II) carbonates are often the dominant solid phase in distribution systems with higher DIC (Edwards *et al.* 1999; Liu *et al.* 2008; Noel *et al.* 2014).

Numerous studies over the last 50 years have used thermodynamic models to describe lead carbonate aqueous–solid phase interactions for varying water chemistry conditions. Details of models previously used are provided in Supplementary Material, Table S1. These thermodynamic models have provided valuable insight into the relationship between water chemistry, solid phases, and lead solubility (Hem & Durum 1973; Patterson *et al.* 1977; Schock 1980; Marani *et al.* 1995; Noel & Giammar 2008; Tam & Elefsiniotis 2009; Noel *et al.* 2014; Tully *et al.* 2019). A common theme among these studies is uncertainty in the lead solid phases included in the model (i.e., phases allowed to precipitate) and the equilibrium constants ($\log K$ values) used for the lead solid phases and aqueous complexes. There is a large variation in the values for equilibrium constants reported in the literature, and this can result in lead solubility and solid phase predictions that differ over several orders of magnitude between thermodynamic models, and between model predictions and experimental or field data (Noel & Giammar 2008; Noel *et al.* 2014). Equilibrium constants are typically determined by measuring species of interest under different experimental conditions (Vanbriesen *et al.* 2009). Equilibrium constant values reported may vary due to multiple reasons including the specific conditions of the experiment (temperature, pressure, ionic strength, and solution species present), the analytical methods used to measure the species concentrations (with methods improving over time), or variations in the morphology of the solid phases.

Based on prior thermodynamic models and laboratory experiments (see Supplementary Material, Tables S1 and S2), there are two dominant solid phases (hydrocerussite and cerussite) and up to nine lead aqueous species ($\text{Pb}(\text{CO}_3)_2^{-2}$, $\text{Pb}(\text{OH})_2$, $\text{Pb}(\text{OH})_3^-$, $\text{Pb}_2\text{OH}^{+3}$, $\text{Pb}_3(\text{OH})_4^{+2}$, $\text{Pb}_4(\text{OH})_4^{+4}$, PbCO_3^0 , PbHCO_3^\pm , and PbOH^+) in the basic lead carbonate system. Based on thermodynamic constraints, hydrocerussite is the most stable lead (II) carbonate when the pH is above 7–7.5, whereas cerussite is stable under more acidic conditions (Schock 1980; Schock *et al.* 1980). A range in equilibrium constants has been reported for hydrocerussite and cerussite. For instance, the hydrocerussite equilibrium constant ($\log K$) has been reported to range from –16.25 to –19.04 (Supplementary Material, Tables S1 and S2). While Noel *et al.* (2014) and Tully *et al.* (2019) recently illustrated the sensitivity of lead solubility predictions to the hydrocerussite equilibrium constant, predictions also depend on the equilibrium constants for lead aqueous species. Equilibrium constants for these species also have a range of reported values. There is a need to understand the way in which uncertainty in equilibrium constants, including those for aqueous species, propagates through the thermodynamic model to guide the selection of equilibrium constant values, understand uncertainty in predictions, and thus to develop more robust thermodynamic lead solubility models.

Uncertainty in thermodynamic models caused by variation in equilibrium constant values has been explored previously for other chemical systems. For instance, Vanbriesen *et al.* (2009) used a Markov chain Monte Carlo to regress 27 thermodynamic constants for an EDTA system simultaneously. More recently, Guo *et al.* (2019) used a Monte Carlo with Latin Hypercube sampling to propagate equilibrium constant uncertainty through thermodynamic simulations of the mercury system. Both studies found that slight variations in equilibrium constants can result in significant changes to the output species concentrations spanning multiple orders of magnitude. Furthermore, Guo *et al.* (2019) highlighted that there is a significant lack of thermodynamic parameter uncertainty quantification for environmentally important metals such as As, Pb, Cd, and Hg.

While Monte Carlo methods are popular, they require probability distributions for input parameters (Novoselov *et al.* 2015). If probability distributions for input parameter values are not available, as is the case for the equilibrium constants for the lead carbonate system, options to evaluate uncertainty are more limited. Modifying one factor (i.e., one equilibrium constant value) at a time can be cumbersome and miss critical interaction effects between factors. The factorial analysis is a sensitivity analysis method useful for assessing many factors without requiring large computational resources (Liu *et al.* 2020). A full factorial analysis involves simulating the permutation of every combination of factors at every level (high or low value) investigated. One of the most important aspects of factorial analysis is the ability to analyze interactions between the factors (Mee 2009). For example, Liu *et al.* (2020) showed important interaction effects between factors (e.g., between

snowfall temperature and threshold temperature for snowmelt) in a model used to simulate average streamflow. In a thermodynamic model, these factors would be the equilibrium constants for the various species included in a model for simulating total dissolved lead.

The main objective of this study is to evaluate how uncertainties in equilibrium constants for lead solid phases and aqueous complexes propagate through a thermodynamic model used to simulate the lead carbonate aqueous system. In particular, the study aims to (i) identify species' equilibrium constants that contribute most to model uncertainty; (ii) determine how uncertainty contributed by each species equilibrium constant varies with changing pH and DIC conditions; and (iii) provide recommendations for future thermodynamic modeling of this system. The lead carbonate system considering two solid phases (cerussite and hydrocerussite) and nine aqueous lead complexes is simulated using a full factorial statistical approach that considers the range of equilibrium constant values previously reported. Batch laboratory lead solubility experiments are also conducted with results compared with output from the factorial analysis. While the lead carbonate system simulated in this study is simplified relative to the water chemistry and solid corrosion scales found in drinking water distribution systems, analyzing this simpler system provides important fundamental insight into the uncertainty inherent in thermodynamic simulations of the aqueous lead system. The study findings can be applied to guide the future application of thermodynamic models including consideration of uncertainty inherent in these models.

METHODS

Factorial methods

A full factorial statistical analysis was performed to analyze the way in which uncertainty in equilibrium constants ($\log K$ values) for all lead species in the lead carbonate system propagates through the thermodynamic model. A custom Python (v.3) wrapper code was used to iterate through factorial combinations of $\log K$ values that were used as parameter inputs to a thermodynamic model implemented in PHREEQC (v1.4.2). This code has been uploaded to a public repository (see Data Availability Statement for URL). The PHREEQC model was a batch simulation with the lead carbonate solid-phase hydrocerussite initially present in excess in a buffered solution with ionic strength 0.01 M that was closed to the atmosphere at room temperature (20 °C). A second lead carbonate, cerussite, was not initially present but was allowed to precipitate if the saturation index is >0 . Nine aqueous lead species were considered: $\text{Pb}(\text{CO}_3)_2^{-2}$, $\text{Pb}(\text{OH})_2$, $\text{Pb}(\text{OH})_3^-$, $\text{Pb}_2\text{OH}^{+3}$, $\text{Pb}_3(\text{OH})_4^{+2}$, $\text{Pb}_4(\text{OH})_4^{+4}$, PbCO_3^0 , PbHCO_3^+ , and PbOH^+ . These were chosen based on prior modeling and experimental studies (Schock 1980; Marani *et al.* 1995; Xie *et al.* 2010a, 2010b; Noel *et al.* 2014). Dissolution and formation equations for the lead aqueous species and solid phases are provided in Table 1. The MINTEQ.V4 database (Allison *et al.* 1991) was used to specify the equilibrium constants for non-lead aqueous and solid species (e.g., H_2CO_3 , HCO_3^- , and CO_3^{2-}) that were considered in the model. Simulations were conducted with pH ranging from 4 to 11 and DIC ranging from 0 to 250 mg C/L. These pH and DIC conditions were coded into the PHREEQC model via the python wrapper.

Table 1 | Species, equations, and $\log K$ levels used for 2^{11} factorial analysis

Species	Equation	Min ('-') LogK	Max ('+') LogK	Type
Cerussite	$\text{PbCO}_{3(s)} = \text{Pb}^{+2} + \text{CO}_3^{-2}$	-13.64	-10.91	Dissolution
Hydrocerussite	$\text{Pb}_3(\text{OH})_2(\text{CO}_3)_{2(s)} + 2\text{H}^+ = 3\text{Pb}^{+2} + 2\text{H}_2\text{O} + 2\text{CO}_3^{-2}$	-19.04	-16.25	Dissolution
$\text{Pb}(\text{CO}_3)_2^{-2}$	$\text{Pb}^{+2} + 2\text{CO}_3^{-2} = \text{Pb}(\text{CO}_3)_2^{-2}$	8.2	12.29	Formation
$\text{Pb}(\text{OH})_2$	$\text{Pb}^{+2} + 2\text{H}_2\text{O} = \text{Pb}(\text{OH})_2 + 2\text{H}^+$	-17.85	-16.42	Formation
$\text{Pb}(\text{OH})_3^-$	$\text{Pb}^{+2} + 3\text{H}_2\text{O} = \text{Pb}(\text{OH})_3^- + 3\text{H}^+$	-29.24	-27.99	Formation
$\text{Pb}_2\text{OH}^{+3}$	$2\text{Pb}^{+2} + \text{H}_2\text{O} = \text{Pb}_2\text{OH}^{+3} + \text{H}^+$	-7.28	-6.24	Formation
$\text{Pb}_3(\text{OH})_4^{+2}$	$3\text{Pb}^{+2} + 4\text{H}_2\text{O} = \text{Pb}_3(\text{OH})_4^{+2} + 4\text{H}^+$	-23.91	-22.48	Formation
$\text{Pb}_4(\text{OH})_4^{+4}$	$4\text{Pb}^{+2} + 4\text{H}_2\text{O} = \text{Pb}_4(\text{OH})_4^{+4} + 4\text{H}^+$	-21.02	-18.98	Formation
PbCO_3^0	$\text{Pb}^{+2} + \text{CO}_3^{-2} = \text{PbCO}_{3(aq)}^0$	5.4	7.4	Formation
PbHCO_3^+	$\text{Pb}^{+2} + \text{CO}_3^{-2} + \text{H}^+ = \text{PbHCO}_3^+$	12.96	13.2	Formation
PbOH^+	$\text{Pb}^{+2} + \text{H}_2\text{O} = \text{PbOH}^+ + \text{H}^+$	-8.72	-6.18	Formation

The minimum ('-') and maximum ('+') values determined from the literature (see Supplementary Material, Tables S1 and S2 for more details).

The custom python wrapper substituted logK values into the PHREEQC batch simulation for each solid and aqueous lead species. This substitution was based on a target logK table (Table 1 and Supplementary Material, Table S3) which identified the levels (e.g., minimum and maximum) for each factor to be included in the factorial analysis. The levels used for each logK value were based on an extensive literature review including Lothenbach *et al.* (1999) who previously compiled relevant logK values for the lead system that were reported in studies published between 1922 and 1998. Primary references for logK values obtained from Lothenbach *et al.* (1999) are provided in Supplementary Material, Tables S1 and S2. Once all combinations of logK values were then run through the PHREEQC simulation, the python code collated the model output for analysis. An initial factorial analysis was run with 11 factors 2 levels (2^{11}), meaning that each of the 11 factors (aqueous or solid phase logK values) investigated were varied across 2 levels: a high level and a low level. For this factorial analysis, a level was either the minimum or maximum logK value previously reported in the literature for that factor (Table 1).

A second factorial analysis was run with 5 factors 12 levels (12^5). This factorial analysis only included the five factors that were identified in the 2^{11} analysis to be associated with the greatest uncertainty in the model output. Additional levels that fell between the minimum and maximum values were included in this factorial analysis to further understand the sensitivity of each factor (Supplementary Material, Table S3). In this factorial analysis, the logK values in the MINTEQ. V4 database (Allison *et al.* 1991) were also set as a level for each factor. Finally, the results of the 12^5 factorial analysis were compared with batch lead solubility experimental results to evaluate whether the experimental data could be simulated when considering all logK combinations for these five factors that contribute to the greatest uncertainty in the model output. In doing this, recommendations are provided for selecting logK values for these five factors and for considering uncertainty in future thermodynamic simulations of the lead carbonate system.

The factorial results were analyzed by main effect analysis which determines the response of the model given a change in one factor (logK value for a given species), allowing for all other factors to be averaged. The main effect of varying factor x was calculated as follows:

$$\text{Main Effect}_x = \frac{\sum Y_{x+}}{n_{x+}} - \frac{\sum Y_{x-}}{n_{x-}} \quad (1)$$

where Y is the measured response of the model (in this case total dissolved lead), n is the number of model runs, ‘-’ indicates the factor at a low level, and ‘+’ indicates the factor at a high level. The indicators of ‘-’ and ‘+’ are standard factorial notation in statistical analysis. Using this equation, the main effect of x is calculated using the average of the response (total dissolved lead), where x is at a high level minus the average of the response where x is at a low level. Main effects were calculated considering the complete range of pH and DIC conditions simulated, as well as for each pH and DIC condition simulated.

The model output for the 2^{11} factorial was also analyzed to examine interaction effects. It is important to understand interaction effects because they show whether the selection of one factor (logK value for one species) may change the uncertainty contribution of another factor (logK value for a different species). If interaction effects are high between different factors, this indicates that it is not possible to select the logK value for an individual species without considering the way in which the value depends on the logK value selected for another species. This would mean that the selection of appropriate logK values for all species in an aqueous system may need to be done simultaneously.

Batch laboratory experiments

Batch hydrocerussite dissolution experiments were conducted across drinking water conditions (pH 8–10; DIC 20, 100, and 200 mg C/L) to determine whether the thermodynamic model was able to simulate total lead solubility from hydrocerussite dissolution when considering uncertainty in the logK values. The matching of the experimental data with the 12^5 factorial model output was used to determine appropriate logK values (or combinations of values) for the aqueous and solid-phase species to be used in the model. It should be noted that while the logK values selected based on the matching are appropriate for the solid phases used in our experiments, it is possible that logK values for solid phases may vary with crystallite size and other aging phenomenon.

Experimental materials

Hydrocerussite ($\text{Pb}_3(\text{CO}_3)_2(\text{OH})_2$; Sigma-Aldrich) was used for all experiments. XRD analysis and Rietveld refinement indicated that the solid phase was dominant hydrocerussite. Sodium bicarbonate (NaHCO_3 ; Sigma-Aldrich, >99.5%) and sodium nitrate (NaNO_3 ; Sigma-Aldrich, >99.0%) were combined with mega-pure water to prepare a stock solution at the target DIC and at 0.01 M ionic strength, respectively. The target pH was maintained with organic buffers: EPPS (Alfa Aesar, 0.2 M, pH

8), CHES (Alfa Aesar, 0.5 M, pH 9), CAPSO (Alfa Aesar, 0.2 M, pH 9.5), and MOPSO (Alfa Aesar, 0.2 M, pH 7). Sodium hydroxide (NaOH; Sigma-Aldrich) and concentrated nitric acid (HNO₃; EMD Milipore, Omnitrace, 67–70% v/v condensed) were used to adjust the stock solution to the target pH (± 0.05 pH units) immediately before each experiment.

Experimental methods

All experiments were performed as closed experiments at room temperature (20 °C). The influence of the initial amount of hydrocerussite present (solid loading) on the time to equilibrium was first evaluated by conducting experiments with varying solid loadings (10–1,000 mg/L; Supplementary Material, Figure S1) at pH 8 and DIC 20 mg C/L. Equilibrium was determined to have been reached when the measured dissolved lead concentrations were within 10% of each other for at least three samples collected at consecutive times. Based on these experiments, a solid loading of 175 mg/L was used for all subsequent experiments with a minimum of 4 days required until equilibrium was reached.

Experiments were run at conditions typical for drinking water systems and for which hydrocerussite is the dominant solid phase (pH 8–10) and at varying DIC levels (20, 100, and 200 mg C/L; Supplementary Material, Table S4). Complications due to possible cerussite formation can occur at pH lower than 7.5 (Schock 1980). Therefore, while some drinking water systems operate at lower pH, all experiments were run at pH 8 or higher to avoid complications of cerussite/hydrocerussite dominance within the experiments. Organic buffers were used to maintain target pH values, since a positive pH-driven feedback loop can occur in unbuffered solutions as hydrocerussite dissolution increases as pH increases. The buffers used have been used in previous hydrocerussite dissolution experiments (Noel *et al.* 2014). Additional experiments (not presented) were conducted which indicated that organic buffers do not alter the dissolved equilibrium lead concentrations at the range of pH values studied.

For all experiments, powdered hydrocerussite was first weighed into 150 mL polypropylene bottles. Each bottle was then half-filled with a pre-prepared stock solution and agitated to evenly disperse the hydrocerussite powder. After agitation, the bottles were filled completely and placed on a shaker table. A minimum of five sample bottles were run for each experimental data point (Supplementary Material, Table S4). While the experiments were run under closed conditions, the stock solution was not prepared under closed conditions and so alkalinity was measured to verify target DIC values. Interference from atmospheric CO₂ was not observed with all DIC measurements at expected target concentrations.

All samples were taken from individual bottles that were sacrificed after sampling. Samples were immediately filtered through a 0.2 μ m polyethersulfone filter. It is possible that the selection of this filter size (0.2 versus 0.45 μ m) may have influenced the dissolved equilibrium concentrations and thus the factorial analyses matching results, but the influence of filter selection was not examined in this study. Samples for total dissolved lead analysis were acidified immediately using 2% HNO₃ and analyzed using graphite furnace atomic absorption spectroscopy (GFAAS, Agilent Technologies; limit of detection for lead = 0.001 mg/L). Standard lead drying, ashing, and volatilization procedures for lead were followed without a matrix modifier. Experimental blanks and laboratory blanks and spikes were included in analysis for quality control and assurance.

Matching methods

The total dissolved lead concentrations for the different experimental conditions were compared with the 12⁵ factorial model output to determine set(s) of logK values that could be used to simulate the experimental data within a given tolerance. The tolerance was set to five times the standard deviation for each experimental data point (Supplementary Material, Table S4). When a factorial model output of total dissolved lead was within the tolerance of an experimental point, the grouping of logK values used for that factorial PHREEQC simulation were considered a ‘match’. Common sets of logK values that ‘matched’ all nine experimental data points were identified and assembled into groups for evaluation (Table 2).

RESULTS AND DISCUSSION

Main effects

The main effect of a factor (species logK value) indicates the impact of this factor on the dependent variable (total dissolved lead). A large main effect indicates that changing the factor from a low value to a high value has a large effect on the simulated total dissolved lead concentration. The model uncertainty associated with a factor with a large main effect is high. The calculated average main effects for all factors considering the entire condition space simulated (pH 4–11, DIC 0–250 mg C/L) are shown in Figure 1(a) with the data provided in Supplementary Material, Table S5A. Hydrocerussite (22.3 mg Pb/L), Pb(CO₃)₂⁻² (12.4 mg Pb/L), PbOH⁺ (7.6 mg Pb/L), cerussite (4.1 mg Pb/L), and PbCO₃ (1.1 mg Pb/L) were found to have

Table 2 | Sets of logK values from 5¹² factorial analysis that match with experimental data

Group	Hydrocerussite LogK	Cerussite LogK	Pb(CO ₃) ₂ ⁻² LogK	PbCO ₃ ⁰ LogK	PbOH ⁺ LogK	Average difference (mg/L)	Maximum difference (mg/L)
Set 1a	-17.77	-13.13	10.06	6.31	-8.72 to -8.03	-0.031	0.104
Set 1b	-17.77	-13.13	10.06	6.13	-7.8 to -7.597	-0.031	0.104
Set 2	-17.77	-13.13	9.938	6.31	-8.72 to -8.03	-0.032	0.101
Set 3	-17.52	-13.13	9.938	6.31	-8.72	0.021	0.312
Set 4	-18.28	-13.39	10.06	6.67	-8.72 to -8.26	-0.043	-0.432
Set 5	-18.03	-13.13	10.06	5.95	-7.1	0.050	-0.402
Set 6	-19.04	-13.64	10.43	6.85	-8.72 to -7.597	-0.069	-0.195

These have been ordered from the 'best' fit (1a/b) to the 'worst' fit (6). The average difference and the maximum difference are calculated as the experimental lead concentration minus the simulated lead concentration.

highest average (across the entire condition space) main effects. The total sum of average main effects considering all factors is 48.7 mg Pb/L – this value can be helpful in determining the relative importance of each factor on the model output (total dissolved lead). In comparing this main effect to the magnitude of experimental lead concentrations (<1 mg/L), the importance of logK value selection is evident, as the main effect of a selected logK can have a large impact on the simulation results when compared to experimental concentrations.

Hydrocerussite accounts for 46% of the average main effects, indicating that almost half of the uncertainty of the model arises from uncertainty in its logK value. Pb(CO₃)₂⁻² represents 25% of the average main effects, with PbOH⁺ (16%), cerussite (8%), and PbCO₃⁰ (2%) accounting for the remaining approximately 25% of the average main effects. All other factors had average main effects of less than 1 mg Pb/L and account for only approximately 3% of the total average main effects, indicating that uncertainty associated with these factors is minimal. Based on average main effect analysis, there are two primary factors: hydrocerussite and Pb(CO₃)₂⁻² and three secondary factors: PbOH⁺, cerussite, and PbCO₃⁰ that contribute to uncertainty in the total dissolved lead predictions. These five factors (hydrocerussite, Pb(CO₃)₂⁻², PbOH⁺, cerussite, and PbCO₃⁰) were further investigated by running a 12⁵ factorial analysis.

While average main effects indicate the influence of changing one factor across the entire condition space on the dependent variable (total dissolved lead), it is also helpful to examine the maximum main effects for each factor and at what conditions they occur (Figure 1(b), Supplementary Material, Table S5B). The maximum main effect is the highest main effect for each factor considering each condition (pH 4–11, DIC 0–250 mg C/L). The largest maximum main effects were for Pb(CO₃)₂⁻² (146 mg Pb/L, pH 9.5, DIC 250 mg C/L), hydrocerussite (95.5 mg Pb/L, pH 9.5, DIC 250 mg C/L), cerussite (86.7 mg Pb/L, pH 4, DIC 250 mg C/L), and PbOH⁺ (28.8 mg Pb/L, pH 4, DIC 0 mg C/L). This indicates that model uncertainty due

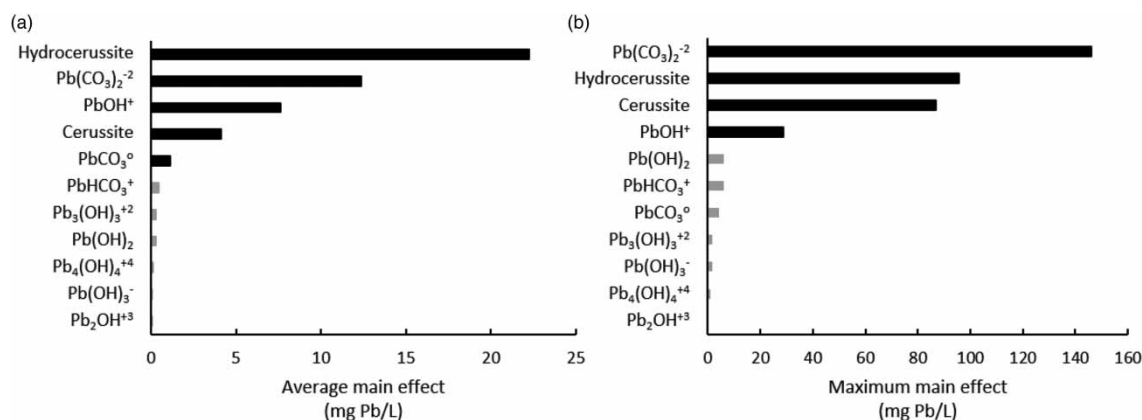


Figure 1 | (a) Average and (b) maximum main effects for all solid and aqueous species in the lead carbonate system. Data are provided in Supplementary Material, Table S5. Black bars are used for species that have the most effect on model uncertainty.

to the $\text{Pb}(\text{CO}_3)_2^{-2}$ logK value was the largest and can change the simulated total dissolved lead by up to 146 mg Pb/L at a pH value of 9.5 and a DIC value of 250 mg C/L. The maximum main effects all occur near the limits considered in the study, typically at either pH 4 or pH 10.5. It is therefore recommended that greater care be taken in selecting equilibrium values for all factors when simulating these extreme conditions. All other factors studied had maximum main effects less than 6 mg Pb/L with these maximum main effects again occurring at extreme pH values (4 or 10.5).

The main effects across the condition space for the five factors identified to have the largest average main effects are shown in Figure 2 (see Supplementary Material, Figure S2 for other factors). Note that the heat maps have an uneven contouring scale where green is less than 10 mg Pb/L. Typical drinking water conditions (between 20 and 200 mg C/L and pH 7 and 10) are also indicated. From Figure 2(a), it can be seen that uncertainty in the hydrocerussite logK results in three main areas of high model uncertainty: low pH (<5.5); mid-high pH (8–10) and high DIC (100–250 mg C/L); and extremely high pH (11). At these conditions changing the hydrocerussite logK value has a large impact on the total dissolved lead predictions. Furthermore, the results suggest that any model simulating drinking water conditions will have high uncertainty associated with the selection of the hydrocerussite logK value. Cerussite has a similar main effect heat map to hydrocerussite with three regions of high uncertainty: low pH (<5) with mid to high DIC (>50 mg C/L); mid-high pH (9–10) with mid to high DIC (>100 mg C/L); and extremely high pH (11; Figure 2(b)). For typical drinking water conditions (pH>7 and DIC >0 mg C/L), and there is moderate uncertainty associated with the selection of the cerussite logK value. $\text{Pb}(\text{CO}_3)_2^{-2}$ has one area that has high main effects at mid to high pH and mid to high DIC – this overlaps with typical drinking water conditions so care needs to be taken in selecting this logK value (Figure 2(c)). Finally, PbOH^+ and PbCO_3^0 have low main effects when considering typical drinking water conditions with both species having higher main effects at lower pH (pH <8) or at extremely high pH (11; Figure 2(d) and 2(e)).

Interaction effects

Interaction effects indicate whether a factor is dependent upon another factor. This is important to consider because if there is a large interaction between factors, both factors need to be simultaneously considered/adjusted when selecting their values as

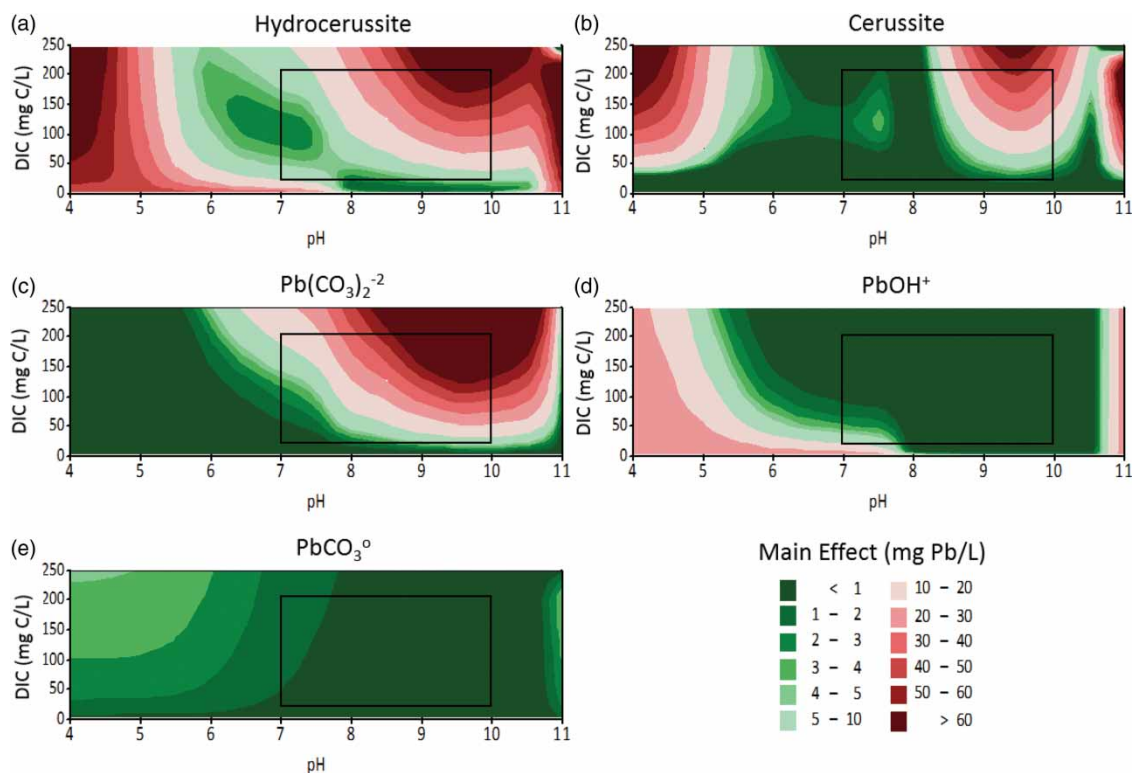


Figure 2 | Heat map of main effects across pH 4–11 and DIC 0–250 mg C/L for the five species with the largest effect on model uncertainty: (a) hydrocerussite, (b) cerussite, (c) $\text{Pb}(\text{CO}_3)_2^{-2}$, (d) PbOH^+ , and (e) PbCO_3^0 . The black box indicates the region of typical drinking water conditions.

their impact on the model output is interconnected. Prior thermodynamic modeling of the lead carbonate system has only examined the effects of modifying a single logK at a time to determine its impact on the model output, thereby neglecting interaction effects (e.g., Noel *et al.* 2014; Tully *et al.* 2019).

Large interaction effects observed from the factorial analyses are shown in Figure 3. In interpreting these figures, parallel lines indicate no interaction between factors, while divergent slopes indicate high interaction (results for all interaction effects are provided in Supplementary Material, Table S6). As can be seen in Figure 3(a)–3(c), cerussite has three major factors that it interacts with: $\text{Pb}(\text{CO}_3)_2^{-2}$, PbOH^+ , and PbCO_3^0 . As such the cerussite logK value needs to be considered in assessing the effect of the logK values for these other species. For instance, as the logK for cerussite increases, the model uncertainty associated with the logK values of $\text{Pb}(\text{CO}_3)_2^{-2}$ (Figure 3(a)) and PbCO_3^0 (Figure 3(c)) increases. Alternatively, as the logK for cerussite increases, the model uncertainty associated with the PbOH^+ logK value decreases (Figure 3(b)). These interaction effects mean that it is important to determine the logK values for cerussite, $\text{Pb}(\text{CO}_3)_2^{-2}$, PbOH^+ , and PbCO_3^0 simultaneously. Cerussite only has minor interactions with other species including $\text{Pb}(\text{OH})_2$, $\text{Pb}_3(\text{OH})_4^{+2}$, $\text{Pb}_4(\text{OH})_4^{+4}$, and PbHCO_3^- (Supplementary Material, Table S6). Finally, hydrocerussite has two major interactions: $\text{Pb}(\text{CO}_3)_2^{-2}$ (Figure 3(d)) and PbOH^+ (Figure 3(e)) with the uncertainty associated with both these species logK values increasing as the hydrocerussite logK increases. A minor interaction between hydrocerussite and PbCO_3^0 is shown in Figure 3(f) for comparison. This indicates that the logK values used for hydrocerussite, $\text{Pb}(\text{CO}_3)_2^{-2}$, and PbOH^+ should be determined simultaneously.

Experimental and matching results

The results for the lead solubility experiments at various pH values (8–10) and DIC values (20, 100, and 200 mg C/L) are shown in Figure 4. All experimental points at pH 8 have a similar total dissolved lead concentration (~ 0.05 mg/L) regardless of the DIC with the lead concentrations diverging as pH increases. For the conditions examined, the total dissolved lead is the highest for pH 10 and DIC 200 mg C/L (0.62 mg/L).

Before comparing the experimental data to the 2^{12} factorial results, the contribution of each of the factors considered in the 2^{11} factorial analysis to the total main effects was examined to confirm that the five factors included in the 2^{12} factorial analysis accounted for the majority of the model uncertainty at the experimental conditions. Of the factors considered in the 2^{11} factorial analysis but not included in the 2^{12} factorial analysis, only $\text{Pb}(\text{OH})_2$ was found to contribute to uncertainty ($< 4\%$) at low DICs (see Supplementary Material, Figure S3). Twenty-one combinations of logK values out of the 248,832 combinations

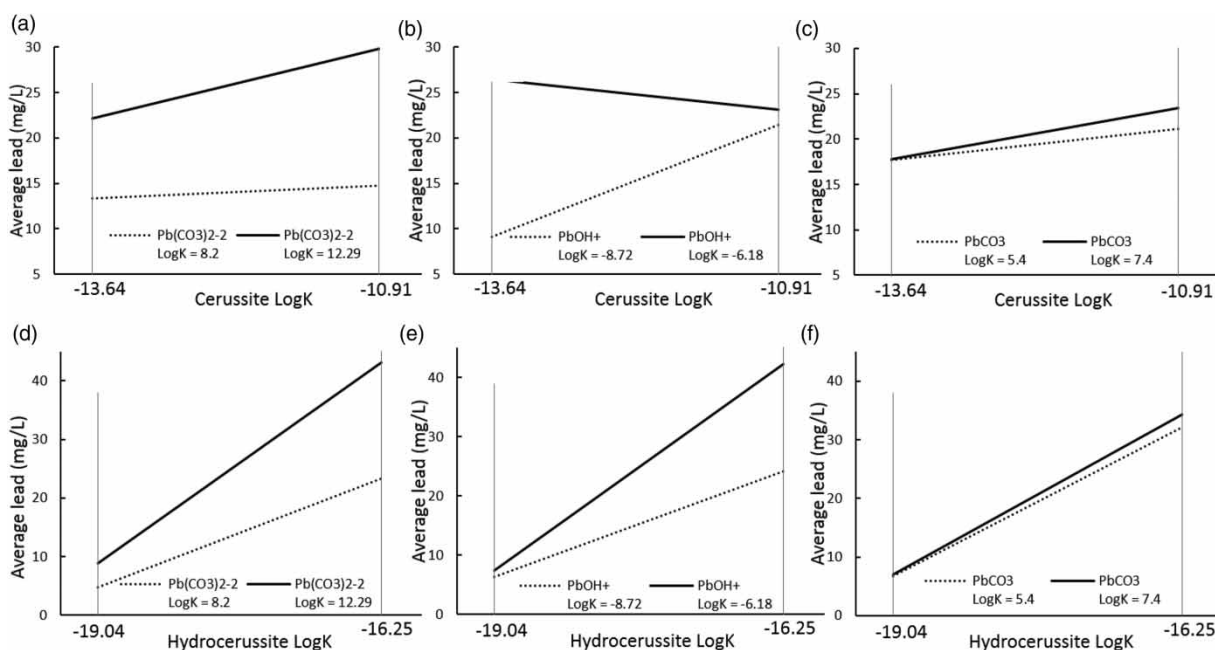


Figure 3 | Key interaction effects for the lead carbonate solids hydrocerussite and cerussite. Cerussite interacts with (a) $\text{Pb}(\text{CO}_3)_2^{-2}$, (b) PbOH^+ , and (c) PbCO_3^0 . Hydrocerussite interacts with (d) $\text{Pb}(\text{CO}_3)_2^{-2}$ and (e) PbOH^+ and has a minor interaction with (f) PbCO_3^0 . A summary of all interaction effects is provided in Supplementary Material, Table S6.

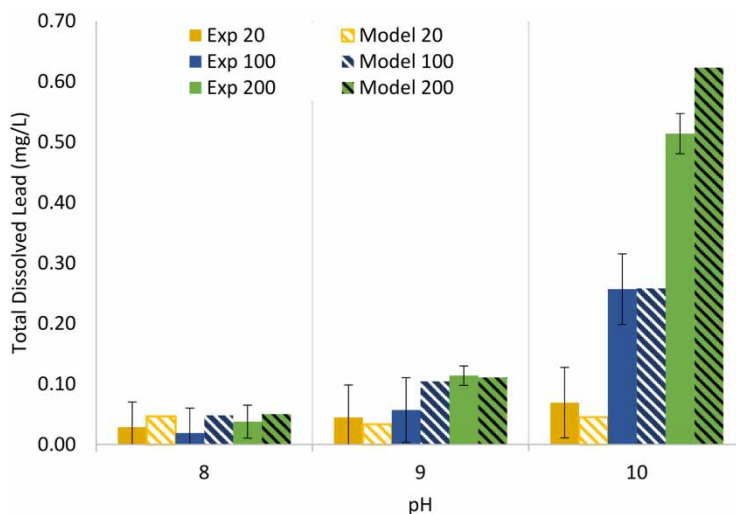


Figure 4 | Experimental (solid bars) and simulated (hatched bars) total dissolved lead concentrations for experimental conditions examined (pH 8–10; DIC 20, 100, 200 mg C/L). Values in the legend indicate DIC concentrations. The simulated results use logK values from Sets 1a and 1b (see Table 2). Error bars indicate the tolerance calculated from the standard deviation of replicate samples.

considered in the 12^5 factorial analysis were found to fit all experimental data points with a tolerance of five times the standard deviation between replicate experimental samples. These 21 combinations of logK values that match experimental data can be grouped according to the logK values of $\text{Pb}(\text{CO}_3)_2^{-2}$, hydrocerussite, and cerussite to form six general groups (Table 2). The uncertainties in the logK values for PbOH^+ and PbCO_3^0 have less effect on the model output relative to the logK values for hydrocerussite, cerussite, and $\text{Pb}(\text{CO}_3)_2^{-2}$. As the logK value for PbOH^+ did not considerably affect the dissolved lead concentrations, a range of values could be used.

Two parameters, the average difference and the maximum difference between the experimental values and the simulated values, were calculated to quantitatively compare the goodness of the fit of the model for each of the general sets of logK values (Table 2). The average difference considering all the experimental conditions ranged from 0.021 mg/L (Set 3) to -0.0686 mg/L (Set 6), whereas the maximum difference ranged from 0.101 mg/L (Set 2) to -0.432 mg/L (Set 4). The logK values in Sets 1a and 1b provided the closest match with the experimental data considering the calculated average and maximum differences (Table 2). While the matching suggests that any of the combinations of logK values shown in Table 2 could be used in future simulations, in general, it is recommended that the logK values associated with Set 1 (Table 2; Figure 4) are used as it was the most common result from matching. In contrast to prior studies that have been unable to simulate experimental results within the tolerance used in this current study (e.g. Noel *et al.* 2014), here this was possible by understanding the uncertainty and the interactions between the logK values for the multiple solid and aqueous phase species in the lead carbonate system. This highlights the need to consider the uncertainty of all logK values and the interactions between these values when using thermodynamic models to simulate the lead carbonate aqueous system.

CONCLUSIONS

This study has shown how uncertainty in equilibrium constant values for solid and aqueous phase species propagates through thermodynamic models of the lead carbonate aqueous system. It was determined that the logK values for the following species contributed most to uncertainty: hydrocerussite, $\text{Pb}(\text{CO}_3)_2^{-2}$, cerussite, PbOH^+ , and PbCO_3^0 . The model output was shown to be highly sensitive to the hydrocerussite logK value with average main effects, indicating that this value accounts for 46% of the uncertainty in the thermodynamic lead solubility predictions for pH 4–11 and DIC 0–250 mg C/L. When simulating typical drinking water conditions (between pH 7 and 10 and DIC 20 and 200 mg C/L), care must also be taken in selecting the equilibrium constant for $\text{Pb}(\text{CO}_3)_2^{-2}$, as it showed the most effect under these conditions. The $\text{Pb}(\text{CO}_3)_2^{-2}$ logK value was associated with the second-highest average uncertainty (up to 25%) across the pH and DIC condition space considered, and also had the highest maximum main effect. Overall, maximum main effects indicated that model uncertainty due to logK selection is typically greatest at extreme pH values (pH <5 and >10). Significant interactions effects were

also observed between the logK values for cerussite, $\text{Pb}(\text{CO}_3)_2^{-2}$, PbOH^+ , and PbCO_3^0 and between the logK values for hydrocerussite and $\text{Pb}(\text{CO}_3)_2^{-2}$ and PbOH^+ . This indicates that logK values used for these species should be selected simultaneously for thermodynamic modeling of the lead carbonate system. Combinations of logK values that provided a good match with experimental data were determined with the average difference between the simulated and experimental lead concentrations calculated to be 0.031 mg/L for the best logK value combination (Set 1a/1b).

Overall, this study provides valuable insight into the uncertainty present in thermodynamic models of the lead carbonate system. While the simulated lead carbonate system is simplified relative to the water chemistry in drinking water distribution systems, analyzing this simpler system provides important insight into the uncertainty inherent in thermodynamic simulations. Lead chemistry is more complex in drinking water distribution systems due to reaction kinetics, temperature changes, and the presence of species including chlorine, natural organic matter, and calcium. It is recommended that this uncertainty analysis be extended to other species including lead phosphate (inorganic) species and that the effects of temperature are explored. Additionally, it is recommended that lead solubility experiments that use well-characterized lead corrosion scales from real distributions systems be combined with similar thermodynamic model uncertainty analysis to further evaluate the sets of recommended logK values derived in this study.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information. A compilation of logK values from literature sources, tabulated experimental results, additional simulation results, and a summary of all interaction effects are provided in the Supplementary Materials. The python code is used to implement the factorial analysis and matching process, and all main and interaction effects plots are available in a data repository: <https://github.com/UWO-RESTORE/Kushnir-2021>.

REFERENCES

- Allison, J. D., Brown, D. S. & Novo-Gradac, K. J. 1991 *MINTEQA2/PRODEFA2 Version 4 - A Geochemical Assessment Model for Environmental Systems*. United States Environmental Protection Agency, Athens, Georgia.
- Docherty, B. & Kariuki, M. 2019 [Creating an effective corrosion control program to eliminate lead in drinking water in Hamilton, Ontario](#). *Journal American Water Works Association* **111** (10), 28–38.
- Edwards, M. & Dudi, A. 2004 [Role of chlorine and chloramine in corrosion of lead-bearing plumbing materials](#). *AWWA* **96** (10), 69–81.
- Edwards, M., Jacobs, S. & Dodrill, D. 1999 [Desktop guidance for mitigating Pb and Cu corrosion by-products](#). *AWWA* **91** (5), 66–77.
- Gouider, M., Bouzid, J., Sayadi, S. & Montiel, A. 2009 [Impact of orthophosphate addition on biofilm development in drinking water distribution systems](#). *Journal of Hazardous Materials* **167** (1), 1198–1202.
- Guo, L., Painter, S. L., Brooks, S. C., Parks, J. M. & Smith, J. C. 2019 [A probabilistic perspective on thermodynamic parameter uncertainties: understanding aqueous speciation of mercury](#). *Geochimica et Cosmochimica Acta* **263**, 108–121.
- Hem, J. & Durum, W. 1973 [Solubility and occurrence of lead in surface water](#). *AWWA* **65** (8), 562–568.
- Kim, E. J., Herrera, J. E., Huggins, D., Braam, J. & Koshowski, S. 2011 [Effect of pH on the concentrations of lead and trace contaminants in drinking water: a combined batch, pipe loop and sentinel home study](#). *Water Research* **45** (9), 2763–2774.
- Lin, Y.-P. & Valentine, R. L. 2008 [Release of Pb \(II\) from monochloramine-mediated reduction of lead oxide \(PbO₂\)](#). *Environmental Science & Technology* **42** (24), 9137–9143.
- Lin, Y.-P. & Valentine, R. L. 2009 [Reduction of lead oxide \(PbO₂\) and release of Pb \(II\) in mixtures of natural organic matter, free chlorine and monochloramine](#). *Environmental Science & Technology* **43** (10), 3872–3877.
- Liu, H., Korshin, G. V. & Ferguson, J. F. 2008 [Investigation of the kinetics and mechanisms of the oxidation of cerussite and hydrocerussite by chlorine](#). *Environmental Science & Technology* **42** (9), 3241–3247.
- Liu, Y. R., Li, Y. P. & Sun, J. 2020 [A two-stage fuzzy-stochastic factorial analysis method for characterizing effects of uncertainties in hydrological modelling](#). *Hydrological Sciences Journal* **65** (12), 2057–2071.
- Lothenbach, B., Ochs, M., Wanner, H. & Yui, M. 1999 *Thermodynamic Data for the Speciation and Solubility of Pd, Pb, Sn, Sb, Nb and Bi in Aqueous Solution*. Japan Nuclear Cycle Development Inst., Tokai, Ibaraki, Japan.
- Marani, D., Macchi, G. & Pagano, M. 1995 [Lead precipitation in the presence of sulphate and carbonate: testing of thermodynamic predictions](#). *Water Research* **29** (4), 1085–1092.
- Mee, R. 2009 *A Comprehensive Guide to Factorial Two-Level Experimentation*. Springer-Verlag, New York.
- Ng, D., Strathmann, T. J. & Lin, Y.-P. 2012 [Role of orthophosphate as a corrosion inhibitor in chloraminated solutions containing tetravalent lead corrosion product PbO₂](#). *Environmental Science & Technology* **46** (20), 11062–11069.
- Noel, J. & Giammar, D. 2008 *The Influence of Water Chemistry on Dissolution Rates of Lead (II) Carbonate Solids Found in Water Distribution Systems*. Water Quality Technology Conference Proceedings, Cincinnati, OH.
- Noel, J. D., Wang, Y. & Giammar, D. E. 2014 [Effect of water chemistry on the dissolution rate of the lead corrosion product hydrocerussite](#). *Water Research* **54**, 237–246.

- Novoselov, A. A., Popov, S. & Filho, C. R. d. S. 2015 Evaluation of uncertainties in solid–aqueous–gas chemical equilibrium calculations. *Computers & Geosciences* **79**, 118–128.
- Patterson, J. W., Allen, H. E. & Scala, J. J. 1977 Carbonate precipitation for heavy metals pollutants. *Journal of the Water Pollution Control Federation* **49** (12), 2397–2410.
- Sandvig, A., Kwan, P., Kirmeyer, G., Maynard, B., Mast, D., Trussell, R. R., Trussell, S., Cantor, A. & Prescott, A. 2008 *Contribution of Service Line and Plumbing Fixtures to Lead and Copper Rule Compliance Issues*. AWWA Research Foundation, Denver, CO.
- Schock, M. R. 1980 Response of lead solubility to dissolved carbonate in drinking water. *AWWA* **72** (12), 695–704.
- Schock, M. R., Mueller, W. & Buelow, R. W. 1980 *Laboratory Technique for Measurement of pH for Corrosion Control Studies and Water Not in Equilibrium with the Atmosphere*.
- Switzer, J. A., Rajasekharan, V. V., Boonsalee, S., Kulp, E. A. & Bohannon, E. W. 2006 Evidence that monochloramine disinfectant could lead to elevated Pb levels in drinking water. *Environmental Science & Technology* **40** (10), 3384–3387.
- Tam, Y. & Elefsiniotis, P. 2009 Corrosion control in water supply systems: effect of pH, alkalinity, and orthophosphate on lead and copper leaching from brass plumbing. *Journal of Environment Science and Health, Part A Environmental Science* **44** (12), 1251–1260.
- Tully, J., DeSantis, M. K. & Schock, M. R. 2019 Water quality-pipe deposit relationships in Midwestern lead pipes. *JAWWA Water Science* **1** (2), 1–18.
- Vanbriesen, J. M., Small, M., Weber, C. & Wilson, J. 2009 Modelling chemical speciation: thermodynamics, kinetics and uncertainty, Chapter 4. In: *Modelling of Pollutants in Complex Environmental Systems, Volume II. Part II Aquatic Modelling and Uncertainty*. ILM Publications, Plymouth, UK, pp. 133–149.
- Vasquez, F. A. 2005 *The Effect of Free Chlorine and Chloramines on Lead Release in a Distribution System*. M.Sc., University of Central Florida Orlando, Florida.
- Wang, Y., Xie, Y. & Giammar, D. E. 2012 *Lead (IV) Oxide Formation and Stability in Drinking Water Distribution Systems*. Water Research Foundation, St Louis.
- Xie, Y., Wang, Y. & Giammar, D. E. 2010a Impact of chlorine disinfectants on dissolution of the lead corrosion product PbO_2 . *Environmental Science & Technology* **44** (18), 7082–7088.
- Xie, Y., Wang, Y., Singhal, V. & Giammar, D. E. 2010b Effects of pH and carbonate concentration on dissolution rates of the lead corrosion product PbO_2 . *Environmental Science & Technology* **44** (3), 1093–1099.

First received 11 May 2021; accepted in revised form 4 October 2021. Available online 19 October 2021