Isotherm ranking and selection using thirteen literature datasets involving hydrophobic organic compounds

CrossMark

L. Shawn Matott [a,*], Zhengzheng Jiang [b], Alan J. Rabideau [b], Richelle M. Allen-King [c]

[a] University at Buffalo, Center for Computational Research, Buffalo, NY, United States
[b] University at Buffalo, Department of Civil, Structural and Environmental Engineering, Buffalo, NY, United States
[c] University at Buffalo, Department of Geology, Buffalo, NY, United States

ARTICLE INFO

ABSTRACT

Numerous isotherm expressions have been developed for describing sorption of hydrophobic organic compounds (HOCs), including "dual-mode" approaches that combine nonlinear behavior with a linear partitioning component. Choosing among these alternative expressions for describing a given dataset is an important task that can significantly influence subsequent transport modeling and/or mechanistic interpretation. In this study, a series of numerical experiments were undertaken to identify "best-in-class" isotherms by refitting 10 alternative models to a suite of 13 previously published literature datasets. The corrected Akaike Information Criterion (AICc) was used for ranking these alternative fits and distinguishing between plausible and implausible isotherms for each dataset.

The occurrence of multiple plausible isotherms was inversely correlated with dataset "richness", such that datasets with fewer observations and/or a narrow range of aqueous concentrations resulted in a greater number of plausible isotherms. Overall, only the Polanyi-partition dual-mode isotherm was classified as "plausible" across all 13 of the considered datasets, indicating substantial statistical support consistent with current advances in sorption theory. However, these findings are predicated on the use of the AICc measure as an unbiased ranking metric and the adoption of a subjective, but defensible, threshold for separating plausible and implausible isotherms.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction and research objectives

Sorption to aquifer solids is an important process affecting groundwater transport of hydrophobic organic compounds (HOCs) and other contaminants (Ehlers and Kavanaugh, 2013; Kan and Tomson, 1990; Karapanagioti et al., 2012; NRC, 2012; Rakowska et al., 2012; Sharma et al., 2013). Sorption in such systems is often described using an isotherm expression relating equilibrium aqueous and sorbed concentrations. Describing experimental data using an isotherm requires estimating "best-fit" isotherm parameter values, a task that

has been termed isotherm fitting, parameter estimation, and calibration.

With the increasing availability of alternative isotherm models, robust procedures are needed to determine which expressions yield the most plausible explanation of sorption for each soil/solute combination. In this regard, previous studies have noted the difficulty of discriminating between fitted isotherms using just statistical "goodness-of-fit" analysis. For example, Xia and Ball (1999) explicitly noted that "we are reluctant to discriminate among these models based solely on goodness of fit criteria." Similarly, Xia and Pignatello (2001) noted that "a choice of model appropriateness on the basis of the MWSE (mean weighted squared error) cannot be made." To adequately discriminate between alternative isotherms,

* Corresponding author. Tel.: +1 716 881 7566; fax: +1 716 849 6656.
E-mail address: lsmatott@buffalo.edu (L.S. Matott).

these and other studies have generally augmented conventional (and biased) regression metrics (e.g. R^2, RMSE (root mean squared error), or MWSE) with additional considerations – for example, correspondence with maximum sorption capacity and other independently derived criteria.

The major objective of the current study was to explore how information-theoretic multi-model statistical analysis can complement existing non-statistical approaches to assessing alternative isotherms. To this end, an established ranking and selection criterion was applied across a wide variety of published sorption datasets. The resulting numerical experiments facilitated a detailed exploration of interrelated research questions, including: Does comparing various fitted isotherms across a multitude of datasets reveal any general trends? If so, is there a correlation between isotherm plausibility and specific organic compounds? How do "dual-mode" isotherms rank in comparison with traditional alternatives? What roles do various dataset characteristics (e.g. number of data points collected and their distribution along the sorbed vs. aqueous concentration curve) play in helping discriminate between alternative isotherms? Are information-theoretic measures of isotherm preference in agreement with disciplinary expertise about sorption mechanics?

1.1. Isotherms considered in this study

Popular isotherms used in groundwater modeling practice include the linear, Freundlich and Langmuir models (Abulaban and Nieber, 2000; Abulaban et al., 1998; Bosma and van der Zee, 1995; Bosma et al., 1996; Tompson, 1993). Although used less frequently, the Polanyi, Toth, Dual-Langmuir, and Generalized-Langmuir–Freundlich (GLF) isotherms are also suitable. Furthermore, theoretical work and laboratory studies of sorption mechanics have led to the development of more complex isotherms for HOCs (Allen-King et al., 2002). For example, many researchers now advocate treating sorption as a combination of linear partitioning to "soft carbon" and nonlinear adsorption to "hard carbon" (Accardi-Dey and Gschwend, 2002; Luthy et al., 1997; Ran et al., 2003; Tiwari and Guha, 2012; Wang et al., 2005; Weber et al., 1992; Yang et al., 2009; Zhang and He, 2009). Consequently, researchers have proposed augmenting traditional nonlinear isotherms with a linear partitioning term so as to make them more consistent with evolving theory (Allen-King et al., 2002; Cornelissen, 2005; Huang et al., 2003; Lohmann, 2003; Weber et al., 1992; Xia and Ball, 1999; Xia and Pignatello, 2001; Xing and Pignatello, 1997 among many others). Such isotherms are referred to as "dual-mode" or "distributed reactivity" models. While the present study emphasizes hard/soft terminology, some have proposed an alternative glassy/rubbery character-ization of dual-mode heterogeneity, with a "phase transition" occurring with either temperature or solvent concentration (Huang et al., 1997; LeBoeuf and Weber, 1997; Xing and Pignatello, 1997). In practice, few measurements have attempted to distinguish between the hard/soft and glassy/rubbery paradigms and both have demonstrated relevance. Furthermore, there can be significant overlap in the isotherm shapes that characterize these theoretical models. Consequent-ly, this paper is not concerned with evaluating whether hard/soft or glassy/rubbery is the correct paradigm, but with the

processes of parameter estimation and model selection for dual-mode and other complex isotherms.

Table 1 lists 10 isotherm equations suitable for describing pollutant sorption that were considered in this study (Ball and Roberts, 1991; Kinniburgh, 1986; Kumar et al., 2010; Moazed and Viraraghavan, 2005; Weber et al., 1992; Xia and Ball, 1999). These range from a simple one-parameter linear isotherm to complex multi-parameter dual-mode formulations (e.g. the Polanyi-, Freundlich-, and Langmuir-with-partitioning isotherms). Of particular interest for HOC sorption is the Polanyi-partition (PP) isotherm, which assumes a pore-filling mechanism for adsorption and incorporates additional param-eters relative to conventional isotherms (Xia and Pignatello, 2001).

1.2. Procedures for isotherm fitting

A variety of approaches for isotherm fitting have been advocated and adopted, and these are typically differentiated by the selected fitness measure and associated error structure and data weighting strategy (Ball and Roberts, 1991; Bolster, 2008; Bolster and Hornberger, 2007; Bolster and Tellinghuisen, 2010; Kinniburgh, 1986; Tellinghuisen and Bolster, 2010; Xia and Ball, 1999; Xia and Pignatello, 2001). In this study a traditional weighted sum of squared residuals (WSSR) fitness function (Φ) was utilized (Ball and Roberts, 1991; Bolster, 2008; Kinniburgh, 1986):

$$\Phi = \sum_{i=1}^{n} \left(w_{qi} \left[q_i - q_i^* \right] \right)^2 \tag{1}$$

where q_i is the measured or inferred sorbed concentration, $q_i^* = f(C_i, \boldsymbol{\theta})$ is the simulated sorbed concentration calculated via an isotherm equation with parameter set $\boldsymbol{\theta}$ (see Table 1), w_{qi} is the weight associated with each q_i, and Φ is the sum of weighted squared differences between the measured and simulated sorbed concentrations.

Based on maximum likelihood theory, the ideal data weight (w_i) for a given residual ($q_i - q_i^*$) is the inverse of the standard deviation of the corresponding measurement (Hill and Tiedeman, 2007). However, only one of the datasets considered in this study contained estimates of the standard deviation of q. For the remaining datasets, and following Kinniburgh (1986) and Ball and Roberts (1991), this study adopted a *relative* weighting approach in which weights were assumed to be inversely proportional to observed sorbed concentrations. Such an assumption would be appropriate, for example, if analytical measurement errors are the dominant source of error and are proportional to concentration. However, the methods used in this work are generalizable to any theoretical or experimentally-derived weighting scheme.

1.3. Procedures for isotherm comparison and selection

As evidenced by Table 1, practitioners have many options when considering candidate isotherms for describing a given dataset. The chosen expression can significantly influence the results of subsequent transport modeling (Hinz, 2001; Hinz et al., 1994; Matott et al., 2009; Rabideau and Miller, 1994), making it desirable to consider and compare multiple