Partial molar properties from molecular simulation using multiple linear regression

Tyler R. Josephson[a,b], Ramanish Singh[a,b], Mona S. Minkara[a], Evgenii Fetisov[a,c], and J. Ilja Siepmann[a,b]

[a] Department of Chemistry and Chemical Theory Center, [b] Department of Chemical Engineering and Materials Science, University of Minnesota, Minneapolis, MN, USA [c] Chemical Physics and Analysis, Pacific Northwest National Laboratory, Richland, Washington 99352, United States

ARTICLE HISTORY
Compiled July 20, 2019

ABSTRACT
Partial molar volumes, energies, and enthalpies can be computed from NpT-Gibbs ensemble simulations through a post-processing procedure that leverages fluctuations in composition, total volume, and total energy of a simulation box. By recording the instantaneous box volumes V and instantaneous number of molecules N_i of each of n species for M frames, a large $M \times n$ matrix $\mathbf{N}$ is constructed, as well as the $M \times 1$ vector $\mathbf{V}$. The $1 \times n$ vector of partial molar volumes $\bar{\mathbf{V}}$ may then be solved using $\mathbf{N} \cdot \bar{\mathbf{V}} = \mathbf{V}$. A similar construction permits calculation of partial molar energies using M instantaneous measurements of the total energy of the simulation box, and $\mathbf{N} \cdot \bar{\mathbf{U}} = \mathbf{U}$. Partial molar enthalpies may be derived from $\bar{\mathbf{U}}$, $\bar{\mathbf{V}}$, and pressure p. These properties may be used to calculate enthalpy and entropy of transfer (absorption, extraction, and adsorption) for species in complex mixtures. The method is demonstrated on three systems in the NpT-Gibbs ensemble: a highly compressible natural gas condensate of methane, n-butane, and n-decane, the liquid-phase adsorption of 1,5-pentanediol and ethanol onto the MFI zeolite, and a relatively incompressible mixture of ethanol, n-dodecane, and water at liquid-liquid equilibrium. Property predictions are compared to those from numerical differentiation of simulation data sequentially changing the composition and from equations of state. The method can also be extended to reaction equilibria in a closed system and is applied to a reactive first-principles Monte Carlo simulation of compressed nitrogen/oxygen.

KEYWORDS
Monte Carlo simulation, thermodynamics, phase equilibria, reaction equilibria

1. Introduction

Partial molar properties are thermodynamic quantities of mixtures that indicate how extensive properties vary with changes in molar composition when temperature and pressure are kept constant [1–3]. Partial molar properties can be leveraged to study the thermodynamics of multicomponent mixtures (especially liquid) and to design new chemical and biochemical processes [4–11]. Separation systems [12], physisorption and chemisorption processes, and reactive systems [10, 11, 13] can benefit from understanding partial molar properties.

CONTACT Tyler R. Josephson. Email: josep180@umn.edu

The partial molar property, $\bar{X}_i$ is a measure of the response of the extensive property X to changes in the number of molecules N_i of species i as $N_{j\neq i}$, T, and p are held constant (Eq. 1). Generally, $\bar{X}_i$ is a function of T and p and may also depend on the composition of the fluid, x_i.

$$\bar{X}_i = \left(\frac{\delta X}{\delta N_i}\right)_{T,p,N_{j\neq i}} \tag{1}$$

Experimentally, partial molar properties are measured using numerical differentiation. Small amounts of a pure component are added to a mixture and then a change in volume or enthalpy is measured with calorimetry [14]. Alternatively, an equation of state (EOS) is fit to experimental data, and partial molar properties are evaluated by solving EOS calculations [15].

Thermodynamic properties can also be predicted using molecular simulation, with tools such as molecular dynamics (MD) and Monte Carlo (MC) simulations. MD is most conveniently applied for closed systems, and a trajectory is generated according to the laws of Newtonian physics. In MC, the Markov chain trajectory is generated stochastically. Multiple statistical mechanical ensembles are available in MC simulations; the Gibbs ensemble (GEMC) [16–18] is particularly valuable for its ability to directly simulate coexisting phases in equilibrium. For example, to simulate vapor-liquid equilibrium (VLE) for a single-component system, a two-box NVT-GEMC simulation uses a low-density vapor box in equilibrium with a high-density liquid box. Within each box, MC moves sample the distribution of positions, orientations, and conformations of the molecules. Volume moves adjust the vapor/liquid volume fraction while keeping the total volume fixed, and molecules are swapped between boxes to reach mechanical and chemical equilibrium. These computational techniques can be predictive of experimental observations when accurate, transferrable force fields are used, such as TraPPE [19], OPLS [20], and CHARMM [21].

Several methods have been developed for predicting partial molar properties in molecular simulations. The experimental method discussed previously, numerical differentiation, can also be used to compute partial molar properties via simulation by numerically differentiating the total property while changing the number of molecules of the component of interest. For multicomponent systems, this method is not computationally efficient because of the number of independent simulations. Its accuracy depends strongly on the uncertainty of the total property, requiring long simulations [2, 22–26].

In the direct method, Frenkel et al. used the Widom test particle insertion method (WTPI) [27] alongside molecular simulation to directly compute partial molar volume and enthalpy in a single MC simulation in the NpT ensemble [24, 25]. However, this approach is inefficient for dense systems, and impossible when trying to insert large molecules [24, 28–32].

Alternatively, the difference method, also based on WTPI, can be used to avoid some of the sampling issues of the WTPI method by using identity changes between two molecule types [24, 25]. However, problems with this method occur if the molecules are very different in size or have very different interactions with the surrounding molecules.

Recent work by Rahbari, et al. showcases Continuous Fractional Component Monte Carlo (CFCMC), which is used to facilitate transfer of large molecules into dense systems [33]. In a CFCMC simulation [34–39], particle insertions take place gradually, which improves the efficiency of molecule exchanges. This method also allows

computation of chemical potentials without the use of test particle methods, and post-processing is used to compute the chemical potential or its derivatives [33].

An alternative paradigm is to use Kirkwood-Buff (KB) integrals [1, 22, 40–42]. Unlike other theories, KB integrals do not invoke assumptions about pairwise additivity (i.e., other theories assume that the total potential energy is only a sum of all two-body interactions within the system, neglecting the contribution of the many-body terms). However, KB integrals cannot be used for ternary closed systems in which two of the three species have the same mole fractions [1]. A unique advantage of KB integrals is that in some cases, they permit calculation of partial molar properties of closed systems [22], such as those sampled using molecular dynamics, by dividing the system into multiple sub-systems, tracking molecule populations and interactions in each subsystem, and correcting for finite-size effects.

Here we present an alternative method using multiple linear regression (MLR) to compute partial molar properties from GEMC simulation results. Benefits of this method include: 1) With just a few lines of code, partial molar properties of all species in a multicomponent mixture can be computed simultaneously. 2) This approach is generalizable to simulations in constant-p ensembles where the number of molecules fluctuate, including reactive closed systems. 3) No special setup, type of move, or accounting of specific interactions is required. 4) The property data (instantaneous values of V, U, and N_i) are available at no cost throughout the simulation, and the data size is small compared to particle coordinates needed for computation of radial distribution functions. 5) Legacy simulation data may be post-processed, so long as instantaneous macroscopic properties (V, U, N_i) have been recorded throughout the simulation trajectory. 6) No restrictions on the types of interactions are imposed. We demonstrate this technique for simulations using pair-wise additive force fields as well as using Kohn-Sham density functional theory; only the *total* volume and/or energy of each simulation box are required.

2. Partial molar properties from multiple linear regression

Instead of beginning from the derivative definition of the partial molar property (Eq. 1), we use the integral form that gives the total property as a composition-weighted sum of the partial molar properties (Eq. 2), such as the partial molar volume (Eq. 3). This formulation enables a system-of-equations approach: for a mixture with n species, use n equations to solve for n unknowns, as shown in Eq. 4 for partial molar volumes in a ternary mixture; measurements at only three compositions a, b, and c are sufficient to compute these properties, under the assumption that they are independent of composition over the set of N_i investigated.

$$\sum_i N_i \bar{X}_i = X \tag{2}$$

$$\sum_i N_i \bar{V}_i = V \tag{3}$$

$$\begin{aligned} N_1^a \bar{V}_1 + N_2^a \bar{V}_2 + N_3^a \bar{V}_3 &= V^a \\ N_1^b \bar{V}_1 + N_2^b \bar{V}_2 + N_3^b \bar{V}_3 &= V^b \\ N_1^c \bar{V}_1 + N_2^c \bar{V}_2 + N_3^c \bar{V}_3 &= V^c \end{aligned} \tag{4}$$

Consequently, for an n-component mixture, $\bar{V}_i$ can be computed by performing n measurements (whether from experiment or simulation) at n compositions to obtain n values each of V and N_i. However, if any two state points are particularly close, then the system of equations becomes ill-conditioned; if a pair of V, N, or x_i are equal, then two equations would become dependent. This creates a dilemma because partial molar properties are composition-dependent. Because Eq. 4 requires measurements at multiple state points, the corresponding measures of $\bar{V}_i$ correspond to some average across the domain of these state points, assuming that $\bar{V}_i$ is invariant with composition. Narrowing this domain necessarily makes equations 4 more ill-conditioned; selecting the "best" set of mixtures would not be trivial, especially in a high-dimensional space.

However, molecular simulations have an opportunity here not available with experiment. The *fluctuations* of V and N_i within a single simulation can be used to obtain $\bar{V}_i$ using Eq. 5, where $\boldsymbol{N}$ is the $M \times n$ matrix of instantaneous mole fractions for M frames and n species, $\bar{V}$ is the $1 \times n$ vector of partial molar volumes $\bar{V}_i$, and $\boldsymbol{V}$ is the $M \times 1$ vector of instantaneous V. Eq. 5 is an overdetermined system of M instances of Eq. 3 for n unknowns, solved using multiple linear regression.

$$\boldsymbol{N} \cdot \bar{\mathbf{V}} = \boldsymbol{V} \qquad\qquad \bar{\mathbf{V}} = (\boldsymbol{N}^T \boldsymbol{N})^{-1} \boldsymbol{N}^T \cdot \boldsymbol{V} \tag{5}$$

Replacing V with U in Eq. 5 gives partial molar internal energies $\bar{U}_i$ (Eq. 6). With $\bar{U}_i$ and $\bar{V}_i$ and the specified pressure p, partial molar enthalpies $\bar{H}_i$ can be computed using $\bar{H}_i = \bar{U}_i + p\bar{V}_i$ (Eq. 7). Alternatively, $\bar{H}_i$ (Eq. 8) can be computed directly using the instantaneous enthalpy of a given simulation frame $H = U + pV$, from instantaneous measures of U and V. A comparison of the instantaneous and ensemble average methods is discussed in Section 4.1.2, and also in recent work by Eggimann et al. [43].

$$\boldsymbol{N} \cdot \bar{\mathbf{U}} = \boldsymbol{U} \qquad\qquad \bar{\boldsymbol{U}} = (\boldsymbol{N}^T \boldsymbol{N})^{-1} \boldsymbol{N}^T \cdot \boldsymbol{U} \tag{6}$$

$$\bar{\mathbf{H}} = \bar{\mathbf{U}} + p\bar{\mathbf{V}} \tag{7}$$

$$\boldsymbol{N} \cdot \bar{\mathbf{H}} = \boldsymbol{H} \qquad\qquad \bar{\mathbf{H}} = (\boldsymbol{N}^T \boldsymbol{N})^{-1} \boldsymbol{N}^T \cdot \boldsymbol{H} \tag{8}$$

We also propose a method to calculate enthalpies and entropies of transfer, $\Delta\bar{H}_{\mathrm{tr}}$ and $\Delta\bar{S}_{\mathrm{tr}}$, between two phases in equilibrium. The energy cost to remove one molecule of species i from phase I, holding T, p, and $N_{j\neq i}$ constant, is $-\bar{U}_i^{\mathrm{I}}$, and the energy gained by phase II upon addition of one molecule of species i from phase I, holding T, p, and $N_{j\neq i}$ constant, is $\bar{U}_i^{\mathrm{II}}$. Consequently, Eq. 9 indicates the energy change that

would occur upon transfer of this molecule between the two phases, $\Delta\bar{U}_{\mathrm{tr},i}$. This is valid so long as the perturbation is sufficiently small such that partial molar internal energies of all species do not change during the transfer (see extended derivation in supporting information).

$$\Delta\bar{U}_{\mathrm{tr},i} = \bar{U}_i^{\mathrm{II}} - \bar{U}_i^{\mathrm{I}} \tag{9}$$

$\Delta\bar{H}_{\mathrm{tr},i}$ is calculated in the same way, except now the pV work must be accounted for (Eq. 10). Although the pressure is set as thermodynamic constraint for both phases, the incremental change in volume due to the transfer of the molecule is different for each phase, and this will be captured in $\Delta\bar{V}_{\mathrm{tr},i}$ (Eq. 11). Special considerations must be taken when treating a constant-volume simulation box as encountered during simulations probing adsorption in an incompressible sorbent material (see extended derivation in Supporting Information and discussion in Section 4.3).

$$\Delta\bar{H}_{\mathrm{tr},i} = \Delta\bar{U}_{\mathrm{tr},i} + p\Delta\bar{V}_{\mathrm{tr},i} = \bar{H}_i^{\mathrm{II}} - \bar{H}_i^{\mathrm{I}} \tag{10}$$

$$\Delta\bar{V}_{\mathrm{tr},i} = \bar{V}_i^{\mathrm{II}} - \bar{V}_i^{\mathrm{I}} \tag{11}$$

$\Delta\bar{G}_{\mathrm{tr},i}$ is the incremental change in excess free energy due to transferring one molecule between the phases, holding T, p, and $N_{j\neq i}$ constant. This can be calculated using the ensemble averaged number densities ρ_i of species i in each phase, along with the gas constant R and T (Eq. 12) [44]. By using $\Delta\bar{G}_{\mathrm{tr},i} = \Delta\bar{H}_{\mathrm{tr},i} - T\Delta\bar{S}_{\mathrm{tr},i}$, the incremental change in excess entropy upon transferring a molecule of species i between the phases can also be computed.

$$\Delta\bar{G}_{\mathrm{tr},i} = -RT\ln\left(\frac{\rho_i^{\mathrm{II}}}{\rho_i^{\mathrm{I}}}\right) \tag{12}$$

For a single-component system at vapor-liquid equilibrium, these relationships simplify to the well-known relationships for free energy, enthalpy, and entropy of vaporization [2].

MLR may also be applied to a reactive closed system. As long as all species undergo reaction and N_i and V fluctuate, partial molar properties may be computed. Volumes, energies, and enthalpies of reaction may be computed from differences in these partial properties.

We introduce this method by calculating these properties in Monte Carlo simulations in the NpT-Gibbs ensemble for phase and adsorption equilibria (including a fixed-volume sorbent phase for the latter case) and the NpT ensemble for reaction equilibrium in a single phase. In addition, we also assess the MLR approach for NVT-Gibbs ensemble simulations.

3. Simulation Methods

3.1. VLE of natural gas condensate

For a three-component hydrocarbon mixture consisting of methane, n-butane, and n-decane, Monte Carlo simulations in the isochoric and isobaric versions of the Gibbs ensemble at $T = 333$ K [16–18] were performed using the in-house Monte Carlo for Complex Chemical Systems–MN software (MCCCS-MN) [45]. Monte Carlo (MC) move probabilities were distributed as follows: volume moves were attempted with a probability of $1/0.3/N_{\text{total}}$, and maximum displacements were adjusted to yield a target acceptance rate of 0.3, to obtain approximately 1 accepted volume move per MC cycle [46]. Coupled-decoupled [47] configurational-bias Monte Carlo [48–50] moves were attempted with a probability of about 30%, and 55% of moves were translations (proportionally distributed among molecule types), and 15% of moves were rotations (proportionally distributed between n-butane and n-decane). Interbox swap moves were attempted with a probability of around 0.5-1%, in order to achieve approximately 1 accepted swap move per MC cycle. Inter- and intramolecular interactions were described using the TraPPE force field for n-alkanes [19]. Simulations were initialised by randomly positioning molecules on a grid, followed by 2k MC cycles of melting at 3000 K, 3k MC cycles of cooling at the set temperature, and at least 50k MC cycles of equilibration. Production runs followed for up to 100k MC cycles for 64 independent simulations. The volume, energy, and composition of each box were output every MC cycle, while pressures were computed every 10 MC cycles.

Equation of state (EOS) calculations were performed using AspenPlus V8.6 using the Peng-Robinson equation of state [51] both with and without binary interaction parameters, taken from Knapp et al. [52]. First, a pT curve for the mixture at the base liquid composition was generated for finding an initial estimate of the bubble point pressure at $T = 333$ K. Then, at that pressure and temperature, the vapor composition was calculated by simulating a flash process. Once the temperature, pressure, and vapor composition were determined using the above steps, partial molar properties were calculated by simulating a single stream separately for the two phases. For each phase, the molar flow rate of each component was individually perturbed to calculate the partial molar properties. The partial molar properties for both phases were found be numerically stable for perturbations of 0.01% and 0.001% of the molar flow rate.

3.2. Liquid-liquid equilibrium

Harwood et al. [53] carried out Gibbs ensemble Monte Carlo simulations for the liquid-liquid equilibrium of ethanol, water, and n-dodecane at $T = 333$ K and $p = 100$ kPa. The simulated systems contained 600 ethanol, 300 n-dodecane, 6 n-hexane (to facilitate n-dodecane transfers through identity switch moves [54]), and either 100, 300, or 500 water molecules. Compete simulation details are provided in Ref. [53].

3.3. Liquid-phase adsorption

Sun et al. [55] carried out Gibbs ensemble Monte Carlo simulations for solution-phase adsorption of 1,5-pentanediol and ethanol into all-Si MFI zeolite at $T = 323$ K and $p = 100$ kPa. Two state points from this work are analysed here: a mixture with 452 ethanol and 48 diol molecules and a mixture with 208 ethanol and 198 diol molecules, where the sorbent phase consisted of 12 MFI unit cells for both mixtures. Complete

simulation details are provided in Ref. [55].

3.4. Reaction equilibria

Fetisov et al. [56] performed reactive first principles Monte Carlo simulations in the isothermal-isobaric ensemble, for a system of 32 N and 64 O atoms at $T = 3000$ K and $p = 30$ GPa, using the BLYP density functional [57, 58] with the third-generation Grimme dispersion correction (D3) [59]. In this simulation, molecules are treated as aggregates of atoms, and chemical equilibrium is sampled using MC moves such as atom and molecule translations, molecule rotations, bond length changes, and atom identity exchanges. While the total number of N and O atoms are conserved, the population of molecules fluctuates as configuration space is sampled. After equilibration, 32 independent simulations were performed for ∼4000 MC steps each. Complete simulation details are provided in Ref. [56].

4. Results

4.1. Natural Gas Condensate

We introduce and validate this method using a model natural gas condensate characterised experimentally by Urlic et al. [60]. Urlic et al. measured bubble and dew points for a model natural gas condensate with $x_{C1} = 0.6$, $x_{C4} = 0.31$ and $x_{C10} = 0.09$. At $T = 333$ K, they measured a bubble point pressure of 17.4 MPa. After performing a few iterations of VLE simulations in the NVT-Gibbs ensemble, we identified simulation settings that achieved a similar liquid composition, so as to compare our bubble point properties with those of Urlic et al.

The reference case (enabling calculation of the bubble point pressure) for this study usesan NVT-Gibbs ensemble simulation of cubic vapor and liquid boxes, each with initial box lengths of 55 Å. After equilibration, the actual liquid composition was found to be slightly enriched in n-decane relative to the target composition by Urlic et al. [60], with $x_{\mathrm{C1}} = 0.570 \pm 0.004$, $x_{\mathrm{C4}} = 0.310 \pm 0.002$, and $x_{\mathrm{C10}} = 0.120 \pm 0.002$.Using the pressure measured in the vapor box of the NVT-Gibbs simulation, $p = 16220 \pm 40$ MPa, simulations in the NpT-Gibbs and NpT ensembles were also run for 100k MC cycles. Partial molar properties from the GEMC simulations are compared to those calculated using numerical differentiation of single-box simulations in the NpT ensemble, with incremental changes in N_i holding $N_{j \neq i}$ constant (Eq. 1, and Fig. S1 in the Supporting Information).

MLR presents a challenge in assigning uncertainty, because the simulation data is highly correlated. To obtain 95% confidence intervals for these properties, we separately compute partial molar properties for each of 64 independent simulations, then compute the mean and standard deviation of these independent estimates. The standard deviation is multiplied by $2/\sqrt{N-1}$ to give the 95% confidence intervals; error propagation is not used.

For finite-sized systems, correlations may influence the ensemble averages of derived properties such as the compressibility factor, Z [43], or in this case, H. To explore this closely, we compute H and $\bar{H}_i$ two ways. $\bar{H}_{i,\mathrm{inst}}$ are computed using Eq. 8, after computing the mechanical observable H_{inst} using instantaneous measures of U and V for each frame, and p is taken as either the instantaneous value for NVT-Gibbs simulation or as the specified thermodynamic constraint for NpT-Gibbs simulations.

Alternatively, $\bar{H}_{i,\text{ave}}$ assumes no correlation between U and V, and is computed using Eq. 7 after $\bar{V}_i$, $\bar{U}_i$, and p have been computed separately. Parity plots and histograms of the MLR in the NpT-Gibbs ensemble are provided in Figure 1, and thermodynamic properties are compiled in Table 1. Correlations among V, U, and p are shown in Figure S2 in Supporting Information. Energies and enthalpies of transfer are reported in Table 2. System size effects were also examined, and are presented in Section 2.4 of the Supporting Information.

Figure 1 goes here.

For both the liquid and vapor phases, MLR does an excellent job matching the regressed volume and internal energy to the simulated volume, as shown in the parity plots and histograms of Fig. 1. The instantaneous enthalpy is not described well in the liquid phase; although MLR necessarily reproduces the correct mean value, the histogram is significantly narrower for the regressed distribution, suggesting that MLR struggles to capture the low- and high-enthalpy frames (Fig. 1g and Fig. 1o). This is a consequence of major fluctuations in the instantaneous pressure which is used for calculating the instantaneous enthalpy, which lead to a large difference between the simulation and regression results. MLR performs much better for predicting simulation enthalpies calculated frame-by-frame using the ensemble-average p and the instantaneous V and U (Fig. 1h and 1p), in which p is set precisely and assumed to be constant. MLR still misses the low- and high-enthalpy tails of the distribution of H_{ave} (MLR shifts the low-enthalpy tail $H_{\text{ave}} < -0.6 \times 10^6$ K into enthalpies of around -0.6×10^6 K (Fig. 1h)), but not as severely as H_{inst}, suggesting that damping out the extreme fluctuations in p improves the performance of the MLR.

Table 1 goes here.

The partial molar properties are reported in Table 1. Several non-intuitive trends are worth noting. The partial molar volumes and energies are not proportional to the number of LJ sites in each molecule; they are not even monotonic. For example, in the liquid phase, n-butane has a smaller partial molar volume compared to methane. In the vapor phase, both n-butane and n-decane have smaller partial molar volumes compared to methane. This has been noted before - large molecules in supercritical fluids can even give negative partial molar volumes [61], and the majority component here, methane, is far above its critical point.

4.1.1. Validation with numerical differentiation

To validate the MLR methodology, partial molar properties for the natural gas condensate system were also calculated using numerical differentiation (ND). The properties were calculated in the NpT ensemble using the MCCCS-MN software [45].
First, the coexistence pressure and compositions were determined for each phase from the NVT-Gibbs ensemble simulations. Then, in the NpT ensemble, molecules of one species N_i were removed or added to the simulation box while $N_{j\neq i}$, T, and p were held constant. $\langle V \rangle$, $\langle U \rangle$, and $\langle H \rangle = \langle U \rangle + \langle p \rangle \cdot \langle V \rangle$ were measured at 5 or 6 different compositions, and a linear equation was fit to this data to measure the slope of the thermodynamic property with respect to number of molecules of that species to calculated partial molar properties. Regression results are shown in Figure S1.

The ND results are in excellent agreement with MLR for NpT-Gibbs, and reveal the inadequacy of the NVT-Gibbs ensemble for predicting partial molar properties (Table 1). Uncertainties for the MLR method are smaller for partial molar properties of methane and larger for that of butane and decane compared to uncertainties from ND. MLR provides computational savings compared to ND. For a ternary system,

independent simulations must be performed at 7 different compositions (the mean composition, plus positive and negative perturbations in N_i for all three species) for finding the partial molar properties of the three components in one phase. We performed simulations at 14 different compositions for more robust statistics (5, 5, and 6 incremental compositions for methane, n-butane, and n-decane, respectively). After performing the NVT-Gibbs simulation, either NpT-Gibbs or NpT simulations could be used to compute partial molar properties. For this system, single-box NpT simulations are faster but require more parallel simulations than two-box NpT-Gibbs simulations; overall, we find that the MLR method is 4.5 and 9 times faster than ND with 7 or 14 NpT simulations, respectively (with the same number of MC cycles). If the reference simulation is in the NpT-Gibbs ensemble, then partial molar properties may be calculated at no additional cost.

4.1.2. Uncertainty analysis

The first assessment of the quality of these regressions is the ability of the partial molar properties to describe the total properties. Table S1 shows the contributions of each component to the total V, U, and H for both the NVT-Gibbs and NpT-Gibbs data sets.

All partial molar properties are consistent with the total properties in both phases for both ensembles, within statistical uncertainty. In the liquid phase, all three species contribute to the total volume, but in the vapor phase, both x_{C10} and $\bar{V}_{\mathrm{C10}}$ are very small, so the contribution of n-decane is negligible; even smaller than the uncertainty in V_{box}. the 95% confidence interval for $\bar{V}_{\mathrm{C10}}$ is much smaller than that for V_{box}; a broader range of values could close the volume balance in the ensemble average of V; when MLR minimises the error between $\sum \bar{V}_i N_i$ and V_{box} across all frames of the simulation, a more precise estimate of V_{box} is found than that constrained by the ensemble average V.

While NVT-Gibbs and NpT-Gibbs have similar V, the relative contributions from the components are significantly different, especially in the vapor phase. Both data sets are self-consistent, but only the NpT-Gibbs simulation agrees with the ND validation data. This also holds for U, H_{inst}, and H_{ave}.

4.1.3. Comparison to equations of state

Thermodynamic properties calculated using the Peng-Robinson equation of state [51] are provided in Table 2. Experimentally, the bubble point pressure is 17400 kPa, which is greater than all predicted results: PR with k_{ij} gives $p = 16900$ kPa, TraPPE gives $p = 16220 \pm 40$ kPa and PR without k_{ij} gives $p = 15800$ kPa. The PR EOS is improved by adding binary interaction parameters, but at the cost of requiring experimental data in multicomponent systems. Experimental measures of liquid and vapor density, and vapor composition, are not available at this state point.

Table 2 goes here.

The liquid density from TraPPE is 5% greater than that of the PR EOS (k_{ij} does not affect the liquid phase density, nor is its density affected by changes in pressure). The liquid phase $\bar{V}_i$ from MC are similar to PR EOS with and without k_{ij}, with relative deviations of 1% for C1 (which has the largest contribution to the total volume), 17% for C4, and 7% for C10. More substantial differences are seen in the vapor phase as the vapor composition and pressure are quite different.

The vapor density varies by up to 10% among all three cases, following the trend

in vapor pressures, with the lowest ρ and p for PR EOS without k_{ij} and the highest for PR EOS with k_{ij}. The vapor phase density predicted by TraPPE is in between that of PR EOS with and without k_{ij}. $\bar{V}_i$ of C1 is within 10% for all three cases. However, for C4 and C10, $\bar{V}_i$ vary significantly between the two EOS models and the MC simulation. In the vapor phase, partial molar properties are significantly more sensitive than bulk properties. Even within the PR EOS, addition of k_{ij} parameters changes the vapor density by just 8%, but $\bar{V}_i$ and $\bar{H}_i$ change by a factor of two in some cases. Major changes in $\bar{V}_i$ of C4 and C10 do not dramatically affect V because $\bar{V}_{\mathrm{C4}}$ and x_{C10} are close to zero, so large changes in density can dramatically change $\bar{V}_i$.

The absolute values of $\bar{U}_i$, and $\bar{H}_i$, are quite different between TraPPE and the PR EOS because they use different reference states. However, the differences between the vapor and liquid properties can be compared directly (Table 2). For these properties, TraPPE qualitatively agrees with PR, giving similar magnitudes and trends in $\Delta\bar{U}_{\mathrm{tr}}$ and $\Delta\bar{H}_{\mathrm{tr}}$, with the exceptions of $\Delta\bar{U}_{\mathrm{tr,C10}}$ which is equal to -4.9 ± 0.5 kJ/mol but the highly negative $\bar{V}_i$ of $n-$decane raises the contribution of pV energy and leads to a positive $\Delta\bar{H}_{\mathrm{tr,C10}}$ (1.42 kJ/mol). Interestingly, while $\Delta\bar{U}_{\mathrm{tr},i}$ shows that transferring n-butane or n-decane into the liquid is energetically favorable, with $\Delta\bar{U}_{\mathrm{tr,C4}} = -3.1 \pm 0.2$ kJ/mol and $\Delta\bar{U}_{\mathrm{tr,C10}} = -4.9 \pm 0.5$ kJ/mol, methane is unusual; it experiences an energetic penalty of 0.45 ± 0.05 kJ/mol when it transfers from the vapor to the liquid, even though the liquid phase has a much higher density of interaction sites. Moreover, since both butane and n-decane have negative $\bar{V}_i$ in the vapor phase, the enthalpies of transfer these two species are higher than the respective energies of transfer and vice-versa for methane.

4.2. Liquid-liquid equilibrium

To demonstrate this technique in a liquid-liquid equilibrium (LLE) system, we considered the ternary ethanol/n-dodecane/water mixture, as reported in Harwood, et al. [53], with $x^{\mathrm{total}}_{\mathrm{H2O}} = 0.1$, $x^{\mathrm{total}}_{\mathrm{H2O}} = 0.25$, and $x^{\mathrm{total}}_{\mathrm{H2O}} = 0.36$. Mixtures partitioned into ethanol-rich and n-dodecane-rich phases, reaching average compositions as shown in Table S3 in Supporting Information.

MLR (Eq. 5 - Eq. 8) is used to compute the partial molar properties. The quality of the regression is shown in Fig. 2. The bimodal distributions indicate that the phase space was not evenly sampled; nonetheless, MLR accurately fits V and U. For this liquid, the pV contribution to H is negligible, so $U \approx H$. Partial molar volumes and internal energies for the three overall compositions are reported in Table 3.

Figure 2 goes here.

Table 3 goes here.

For these mixtures, the partial molar volumes are relatively insensitive to the composition, but partial molar enthalpies are strongly composition dependent. $\bar{V}_{\mathrm{EtOH}}$ ranges from 0.0611 to 0.0648 L/mol (6% change), $\bar{V}_{\mathrm{C12}}$ ranges from 0.23329 to 0.245 L/mol (5% change), and $\bar{V}_{\mathrm{H2O}}$ ranges from 0.018 to 0.023 (22% change). $\bar{U}_{\mathrm{EtOH}}$ and $\bar{H}_{\mathrm{EtOH}}$ range from -33.8 kJ/mol to -30.3 (10% change), while $\bar{U}_{\mathrm{C12}}$ and $\bar{H}_{\mathrm{C12}}$ range from -6.1 kJ/mol to -16.53 as C12 concentration increases. $\bar{U}_{\mathrm{H2O}}$ and $\bar{H}_{\mathrm{H2O}}$ range from -39.1 to -32.3 kJ/mol as H_2O concentration increases (Table 3).

The enthalpy of transfer can also be calculated for each species as it transfers from the C12-rich liquid into the EtOH-rich liquid (Table 3). For all species, $\Delta U_{\mathrm{tr},i} \approx \Delta H_{\mathrm{tr},i}$ because the pV contribution is negligible. Transferring an EtOH molecule from the C12-rich phase to the EtOH-rich phase is enthalpically favorable ($\Delta H_{\mathrm{tr,ave,EtOH}}$ varies

from -3.1 ± 1.1 to -2.3 ± 0.7 kJ/mol). Likewise, transferring an n-dodecane molecule is enthalpically unfavorable (and more unfavorable in systems with more water content). While the trace species H_2O and n-hexane have larger uncertainties, they follow the same trends as EtOH and C12, respectively.

With three different compositions studied in the LLE system, Eq. 4 may be used to solve for the partial molar properties using a system-of-equations approach (Table S4). This approach is not as effective as MLR; details are provided in the Supporting Information.

4.3. Partial molar properties of adsorption equilibria

The MLR construction given in Eq. 5 does not require that all N_i and V are fluctuating; the MLR is necessarily invalid only when two or more are invariant. Consequently, if one species is not permitted to swap, as in the osmotic ensemble [17], in which N_{water} is fixed, while V and all other N_i may vary, then MLR may still be used to measure $\bar{V}_i$ for all species. In a sense, $\bar{V}_i$ are measured for all swapping species, and $\bar{V}_i$ for the fixed species is calculated through the difference from the total volume. However, if two or more species are not permitted to swap, then the relative contributions of these will be impossible to define; the matrix $\mathbf{N}$ becomes rank-deficient, and the MLR will fail.

For simulations with a simulation box of constant volume, such as adsorption in the NpT-Gibbs ensemble with a fixed adsorbent box, the internal energy fluctuates, and $\bar{U}_i$ and $\Delta U_{\text{tr},i}$ can be calculated using Eq. 6 and 9. Consequently, the enthalpy of adsorption may be calculated using Eq. 10, $\Delta H_{\text{tr},j} = \Delta U_{\text{tr},j} + p\Delta V_{\text{tr},j}$. However, because the adsorption box volume is fixed, $\Delta V_{\text{tr},j}$ only depends on $\bar{V}_i$ in the fluid phase; the contribution from estimating $\bar{V}_i$ in the adsorbed phase is zero (see Supporting Information). This is also the case for grand canonical Monte Carlo (GCMC) simulations, which fix the chemical potential μ, T, and V. Additional care must be taken if the GCMC simulation is sampling VLE; if the same box explores both liquid and vapor phases at coexistence, these two populations would need to be separated prior to applying MLR, or $\bar{U}_i$ would represent an average internal energy of both phases.

We apply MLR to the liquid-phase adsorption of 1,5-pentanediol and ethanol into the MFI zeolite at 323 K and 1 bar [55]. Table 4 and Figure 3 show the MLR results. The major difference between using MLR in a fixed simulation box as compared to a flexible one, is that only one volume is sampled in the adsorbed phase. As the number of molecules fluctuates frame-by-frame, the regressed V from the instantaneous N_i and $\bar{V}_i$ is distributed around the fixed V of the simulation box, with mean values in agreement, as enforced by the MLR. Estimates of $\bar{V}_i$ in the adsorbed phase represent the optimal coefficients that solve the MLR problem, but do not share the same meaning of $\bar{V}_i$ in fluctuating-volume systems, where the derivative formulation of partial molar properties (Eq. 1) is valid and $\bar{V}_i$ corresponds to the incremental change of volume that would occur due to a change in N_i.

U in the adsorbed phase can take multiple values. In contrast to the liquid phase (Fig. 3b-c), in which the regressed V, U, and H_{ave} give similar distributions to the simulated distribution, in the adsorbed phase, the regressed U is much narrower than the simulated distribution (Fig. 3a); the low- and high-energy tails of the distribution are not well-described by a linear combination of partial molar properties. Nonetheless, the average U from regression closely matches that simulated, as enforced by the MLR.

In the liquid, $\bar{V}_{\text{Diol}}$ is nearly twice $\bar{V}_{\text{EtOH}}$, and $\bar{U}_{\text{Diol}}$ is 60% greater than $\bar{U}_{\text{EtOH}}$ in the liquid phase, and 86% greater in the zeolite. For this system, the partial molar

properties are relatively consistent across composition (see Table 4). Internal energies and enthalpies of adsorption can also be calculated using Eq. 9 and 10, which are identical because pV is negligible. Both adsorbates experience an enthalpic gain upon adsorption: -12 ± 3 kJ/mol for EtOH and -31 ± 3 kJ/mol for the diol. After calculating the excess free energy of transfer $\Delta\bar{G}_{\mathrm{tr},i}$ using Eq. 12, we can obtain the excess entropies of transfer from $\Delta\bar{G}_{\mathrm{tr},i} = \Delta\bar{H}_{\mathrm{tr},i} - T\Delta\bar{S}_{\mathrm{tr},i}$. As expected, entropy significantly decreases for both species upon adsorption (Table 4); the zeolite is significantly more confined than the solution phase.

Table 4 goes here.

Figure 3 goes here.

4.4. Reaction equilibria of compressed nitrogen and oxygen

Multiple linear regression may also be applied to closed systems if the molecule populations fluctuate due to chemical reactions. This is demonstrated in a reactive first-principles MC simulation of a compressed nitrogen/oxygen mixture consisting of 16 N_2 and 32 O_2 molecules at $T = 3000$ K and $p = 30$ GPa, as reported in Ref. [56]. By monitoring the molecule populations, volume, and total energy of the system, partial molar properties may be calculated using Eq. 5-8. With $\bar{V}_i$ and $\bar{U}_i$ for each species, changes in volume and energy upon reaction may be calculated using Eq. 13.

$$\begin{aligned} \Delta\bar{V}_{N_2+O_2\rightarrow 2NO} &= 2\bar{V}_{\mathrm{NO}} - \bar{V}_{\mathrm{N_2}} - \bar{V}_{\mathrm{O_2}} \\ \Delta\bar{U}_{N_2+O_2\rightarrow 2NO} &= 2\bar{U}_{\mathrm{NO}} - \bar{U}_{\mathrm{N_2}} - \bar{U}_{\mathrm{O_2}} \end{aligned} \tag{13}$$

The computational expense of this technique limited the available statistics: after equilibration, 32 independent simulations were performed for ∼4000 MC steps each. These short, highly-correlated trajectories could not be analyzed using standard multiple regression, which occasionally yielded extreme and unphysical partial molar properties (ca. $\bar{V}$ or $\bar{U} = \pm10^9$ L/mol or kJ/mol). Ridge regression (Eq. 14), which is a variation on multiple linear regression that allows control over the magnitude of the fitted parameters, performed much better than MLR – a small bias $\alpha = 10^{-5}$ was used to make the regression well-behaved without affecting the final values of the coefficients. Shifting V and U by the average V and U was also found to be beneficial; the large atomization energies (but including pseudopotentials for the core electrons) and energies of formation result in large negative energies with only relatively small fluctuations (for example, ranging from -1432.98268 to -1432.37738 Hartrees) and led to erratic fitting behavior. (For the natural gas condensate, shifting U and V by the molar U and V multiplied by N prior to performing MLR did not influence the quality of the fit; shifting is therefore advantageous in data-poor simulations and becomes less important as the amount of data increases.) The first-principles trajectories did not have sufficient statistics to compute partial molar properties of trace species (NO_2, N_2O, and O), so frames including these species were excluded, and frames with only the major species, NO, O_2, and N_2, were analyzed. Each independent trajectory was analyzed with ridge regression to compute averages and 95% confidence intervals. Table 5 and Figure 4 give the MLR results.

$$\begin{aligned} &\text{Linear regression:} \quad \text{minimize} \quad (\boldsymbol{V} - \boldsymbol{N} \cdot \bar{\mathbf{V}})^2 \\ &\text{Ridge regression:} \quad \text{minimize} \quad (\boldsymbol{V} - \boldsymbol{N} \cdot \bar{\mathbf{V}})^2 + \alpha * \sum \bar{V}_i^2 \end{aligned} \tag{14}$$

Table 5 goes here. Fig. 4 goes here.

For this system, ridge regression produced very precise estimates of the partial molar volume for all species, indicating consistency among the independent simulations, presumably aided by the very low compressibility at the extreme pressure. However, the volume distribution was highly skewed toward lower volumes, and the linear model did not capture this behavior. The small number of molecules also produced striping in the regressed values, which may have also contributed to the poor fit. As expected for an incompressible mixture, the volume of reaction is zero within statistical uncertainties. However, the energy distribution was not skewed, and ridge regression produced partial molar energies in a narrower distribution, similar to that observed in the zeolite. These estimates were much less certain - improved statistics from longer runs or more efficient sampling would produce more precise partial molar energies. Nonetheless, Eq. 13 could be used to measure the reaction energy, which was found to be endothermic by 81 ± 32 kJ/mol and of a reasonable order of magnitude when compared to equation-of-state approaches [62–64].

5. Conclusions

Here we demonstrate the use of multiple linear regression (MLR) for calculating the partial molar properties of multicomponent mixtures in constant-pressure simulations that allow for fluctuations in the number of molecules either through phase transfer or reactive moves. This method uses frame-by-frame measurements of total N_i, V, and U for an individual simulation box; no special Monte Carlo moves or accounting of interactions is required, so long as the population of molecules fluctuates in a constant-pressure ensemble. We introduce this method by calculating the partial molar properties of a model natural gas condensate in the NVT-Gibbs, NpT-Gibbs, and NpT ensembles. MLR quantitatively reproduces partial molar properties as calculated with numerical differentiation (ND) in the NpT-Gibbs ensemble, but the coupled volume fluctuations in the NVT-Gibbs ensemble lead to incorrect partial molar properties, especially in the vapor phase. MLR with NpT-Gibbs provides significant cost savings relative to ND multiple simulations in the NpT ensemble. Through simple post-processing of statistics from previously published simulations, partial molar properties are easily obtained. We demonstrate this by calculating partial molar properties for the liquid-phase adsorption of 1,5-pentanediol and ethanol into the MFI zeolite [55], liquid-liquid equilibrium in ethanol/n-dodecane/water mixtures [53], and in the reactive NpT ensemble for compressed nitrogen and oxygen [56]. Notably, we find MLR to be useful in reactive closed systems, as well, and that for systems with poor statistics and an ill-conditioned composition matrix, ridge regression can also be used to fit reasonable partial molar properties. The numerical stability of MLR and the resulting distributions can be improved by shifting by the negative average value of a given property, so that instantaneous values fluctuate around zero.

Acknowledgements

This research was primarily supported by the Department of Energy, Office of Basic Energy Sciences, Division of Chemical Sciences, Geosciences, and Biosciences under Award DE-FG02-17ER16362 (development and testing of the MLR approach and simulations of adsorption equilibria). This research was also supported by the Abu Dhabi Petroleum Institute Research Center, Project Code LTR14009 (simulations of ethanol/*n*-dodecane/water and natural gas mixtures). Simulations of reactive equilibria of compressed NO were supported by the National Science Foundation (CHE-1265849) and used resources of the Argonne Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC02-06CH11357. A portion of the computer resources were provided by the Minnesota Supercomputing Institute at the University of Minnesota. The University of Minnesota Disability Resource Center also supported this work through providing access assistants to MSM. We thank Andrew Yangzesheng Sun for helpful conversations about correlations.

Disclosure statement

The authors declare no financial conflicts of interest associated with this work.

References

[1] S.K. Schnell, P. Englebienne, J.M. Simon, P. Krüger, S.P. Balaji, S. Kjelstrup, D. Bedeaux, A. Bardow and T.J.H. Vlugt, Chem. Phys. Lett. **582**, 154–157 (2013).

[2] S.I. Sandler, *Chemical, Biochemical, and Engineering Thermodynamics*, 4th ed. (Wiley, New York, 2006).

[3] M.J. Moran and H.N. Shapiro, *Fundamentals of Engineering Thermodynamics*, 5th ed. (Wiley, Hoboken, NJ, 2006).

[4] R.A. Alberty, *Thermodynamics of Biochemical Reactions*, 1st ed. (Wiley, Hoboken, NJ, 2003).

[5] T. Engel and P. Reid, *Thermodynamics, Statistical Thermodynamics, and Kinetics*, 3rd ed. (Pearson, Boston, MA, 2012).

[6] B.E. Poling, J.M. Prausnitz and J.P. O'Connell, *The Properties of Gases and Liquids*, 5th ed. (McGraw-Hill Education, New York, 2000).

[7] A. Podgoršek, J. Jacquemin, A.A.H. Pádua and M.F. Costa Gomes, Chem. Rev. **116** (10), 6075–6106 (2016).

[8] D. Almantariotis, O. Fandiño, J.Y. Coxam and M. Gomes, Int. J. Greenhouse Gas Control **10**, 329–340 (2012).

[9] M.B. Shiflett, B.A. Elliott, S.R. Lustig, S. Sabesan, M.S. Kelkar and A. Yokozeki, ChemPhysChem **13** (7), 1806–1817 (2012).

[10] D.S. Firaha, O. Hollczki and B. Kirchner, Angew. Chem., Int. Ed. **54** (27), 7805–7809 (2015).

[11] B. Zhang, A.C.T. van Duin and J.K. Johnson, J. Phys. Chem. B **118** (41), 12008–12016 (2014).

[12] S.P. Šerbanović, B.D. Djordjević and D.K. Grozdanić, Fluid Phase Equilib. **57** (1), 47–65 (1990).

[13] S.P. Balaji, S. Gangarapu, M. Ramdin, A. Torres-Knoop, H. Zuilhof, E.L. Goetheer, D. Dubbeldam and T.J.H. Vlugt, J. Chem. Theory Comput. **11** (6), 2661–2669 (2015).

[14] M. Corea, J.P.E. Grolier and J.M. del Ro, in *Advances in Titration Techniques*, edited by Vu Dang Hoang, Chap. 4 (IntechOpen, Rijeka, 2017).

[15] D. Henderson, S.D. Hamann and D.Y.C. Chan, Mol. Phys. **96** (7), 1145–1149 (1999).
[16] A.Z. Panagiotopoulos, Mol. Phys. **62** (3), 701–719 (1987).
[17] A.Z. Panagiotopoulos, N. Quirke, M. Stapleton and D.J. Tildesley, Mol. Phys. **63** (4), 527–545 (1988).
[18] B. Smit, P. De Smedt and D. Frenkel, Mol. Phys. **68** (4), 931–950 (1989).
[19] M.G. Martin and J.I. Siepmann, J. Phys. Chem. B **102** (97), 2569–2577 (1998).
[20] W.L. Jorgensen, D.S. Maxwell and J. Tirado-Rives, J. Am. Chem. Soc. **118** (45), 11225–11236 (1996).
[21] A.D. MacKerell, D. Bashford, M. Bellott, R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin and M. Karplus, J. Phys. Chem. B **102** (18), 3586–3616 (1998).
[22] S.K. Schnell, R. Skorpa, D. Bedeaux, S. Kjelstrup, T.J.H. Vlugt and J.M. Simon, J. Chem. Phys. **141** (14), 144501 (2014).
[23] S.M. Walas, *Phase Equilibria in Chemical Engineering*, 1st ed. (Butterworth-Heinemann, Boston, MA, 1984).
[24] P. Sindzingre, G. Ciccotti, C. Massobrio and D. Frenkel, Chem. Phys. Lett. **136** (1), 35–41 (1987).
[25] P. Sindzingre, C. Massobrio, G. Ciccotti and D. Frenkel, Chem. Phys. **129** (2), 213–224 (1989).
[26] J.M. Stubbs and J.I. Siepmann, J. Am. Chem. Soc. **127**, 4722–4729 (2005).
[27] B. Widom, J. Stat. Phys. **19** (6), 563–574 (1978).
[28] K.B. Daly, J.B. Benziger, P.G. Debenedetti and A.Z. Panagiotopoulos, Comput. Phys. Commun. **183** (10), 2054–2062 (2012).
[29] C.H. Bennett, J. Comput. Phys. **22** (2), 245–268 (1976).
[30] D.A. Kofke and P.T. Cummings, Mol. Phys. **92** (6), 973–996 (1997).
[31] I. Nezbeda and J. Kolafa, Mol. Simul. **5** (6), 391–403 (1991).
[32] A. Rahbari, A. Poursaeidesfahani, A. Torres-Knoop, D. Dubbeldam and T.J.H. Vlugt, Mol. Simul. **44** (5), 405–414 (2018).
[33] A. Rahbari, R. Hens, I.K. Nikolaidis, A. Poursaeidesfahani, M. Ramdin, I.G. Economou, O.A. Moultos, D. Dubbeldam and T.J.H. Vlugt, Mol. Phys. **116** (21-22), 3331–3344 (2018).
[34] W. Shi and E.J. Maginn, J. Chem. Theory Comput. **3** (4), 1451–1463 (2007).
[35] W. Shi and E.J. Maginn, J. Comput. Chem. **29** (15), 2520–2530 (2008).
[36] W. Shi and E.J. Maginn, J. Phys. Chem. B **112** (51), 16710–16720 (2008).
[37] W. Shi and E.J. Maginn, AIChE J. **55** (9), 2414–2421 (2009).
[38] A. Poursaeidesfahani, A. Torres-Knoop, D. Dubbeldam and T.J.H. Vlugt, J. Chem. Theory Comput. **12** (4), 1481–1490 (2016).
[39] A. Poursaeidesfahani, R. Hens, A. Rahbari, M. Ramdin, D. Dubbeldam and T.J.H. Vlugt, J. Chem. Theory Comput. **13** (9), 4452–4466 (2017).
[40] J.G. Kirkwood and F.P. Buff, J. Chem. Phys. **19** (6), 774–777 (1951).
[41] P. Krüger, S.K. Schnell, D. Bedeaux, S. Kjelstrup, T.J.H. Vlugt and J.M. Simon, J. Phys. Chem. Lett. **4** (2), 235–238 (2013).
[42] N. Dawass, P. Krüger, J.M. Simon and T.J.H. Vlugt, Mol. Phys. **116** (12), 1573–1580 (2018).
[43] B.L. Eggimann, Y. Sun, R.F. DeJaco, R. Singh, M. Ahsan, T.R. Josephson and J.I. Siepmann, Submitted (2019).
[44] A. Ben-Naim, *Statistical Thermodynamics for Chemists and Biochemists*, 1st ed. (Springer US, New York, 1992).
[45] J.I. Siepmann, et al., *Monte Carlo for Complex Chemical Systems—Minnesota*, versions 16.1; University of Minnesota, Minneapolis, MN, 2011 and 2016.
[46] A.D. Cortés-Morales, I.G. Economou, C.J. Peters and J.I. Siepmann, Mol. Simul. **39** (14-15), 1135–1142 (2013).

[47] M.G. Martin and J.I. Siepmann, J. Phys. Chem. B **103** (21), 4508–4517 (1999).
[48] J.I. Siepmann and D. Frenkel, Mol. Phys. **75** (1), 59–70 (1992).
[49] G.C.A.M. Mooij, D. Frenkel and B. Smit, J. Phys.: Condens. Matter **4** (16), L255–L259 (1992).
[50] T.J.H. Vlugt, M.G. Martin, B. Smit, J.I. Siepmann and R. Krishna, Mol. Phys. **94**, 727–733 (1998).
[51] D.Y. Peng and D.B. Robinson, Ind. Eng. Chem. Fundam. **15** (1), 59–64 (1976).
[52] H. Knapp, R. Döring, L. Oellrich, U. Plöcker, J.M. Prausnitz, R. Langhorst and S. Zeck, in *Dechema Chemistry Data Series, Vol. VI.*
[53] D.B. Harwood, C.J. Peters and J.I. Siepmann, Fluid Phase Equilib. **407**, 269–279 (2016).
[54] M.G. Martin and J.I. Siepmann, J. Am. Chem. Soc. **119** (38), 8921–8924 (1997).
[55] Y. Sun, R.F. DeJaco and J.I. Siepmann, Chem. Sci. **10** (16), 4377–4388 (2019).
[56] E.O. Fetisov, I.F.W. Kuo, C. Knight, J. Vande Vondele, T. Van Voorhis and J.I. Siepmann, ACS Central Science **2** (6), 409–415 (2016).
[57] A.D. Becke, Phys. Rev. A **38**, 3098–3100 (1988).
[58] C. Lee, W. Yang and R.G. Parr, Phys. Rev. B **37**, 785–789 (1988).
[59] S. Grimme, J. Antony, S. Ehrlich and H. Krieg, Journal of Chemical Physics **132** (15) (2010).
[60] L.E. Urlic, L.J. Florusse, E.J. Straver, S. Degrange and C.J. Peters, Transp. Porous Media **52** (2), 141–157 (2003).
[61] J.M. Stubbs, D.D. Drake-Wilhelm and J.I. Siepmann, J. Phys. Chem. B **109** (42), 19885–19892 (2005).
[62] S. Bastea and L.E. Fried, J. Chem. Phys. **128** (17), 174502 (2008).
[63] S. Bastea and L.E. Fried, in *Shock Wave Science and Technology Reference Library*, edited by F. Zhang, Vol. 6, Chap. 1 (Springer, Berlin, 2012), pp. 1–31.
[64] G.L. Schott, M.S. Shaw and J.D. Johnson, J. Chem. Phys. **82** (9), 4264–4275 (1985).

6. Tables and Figures

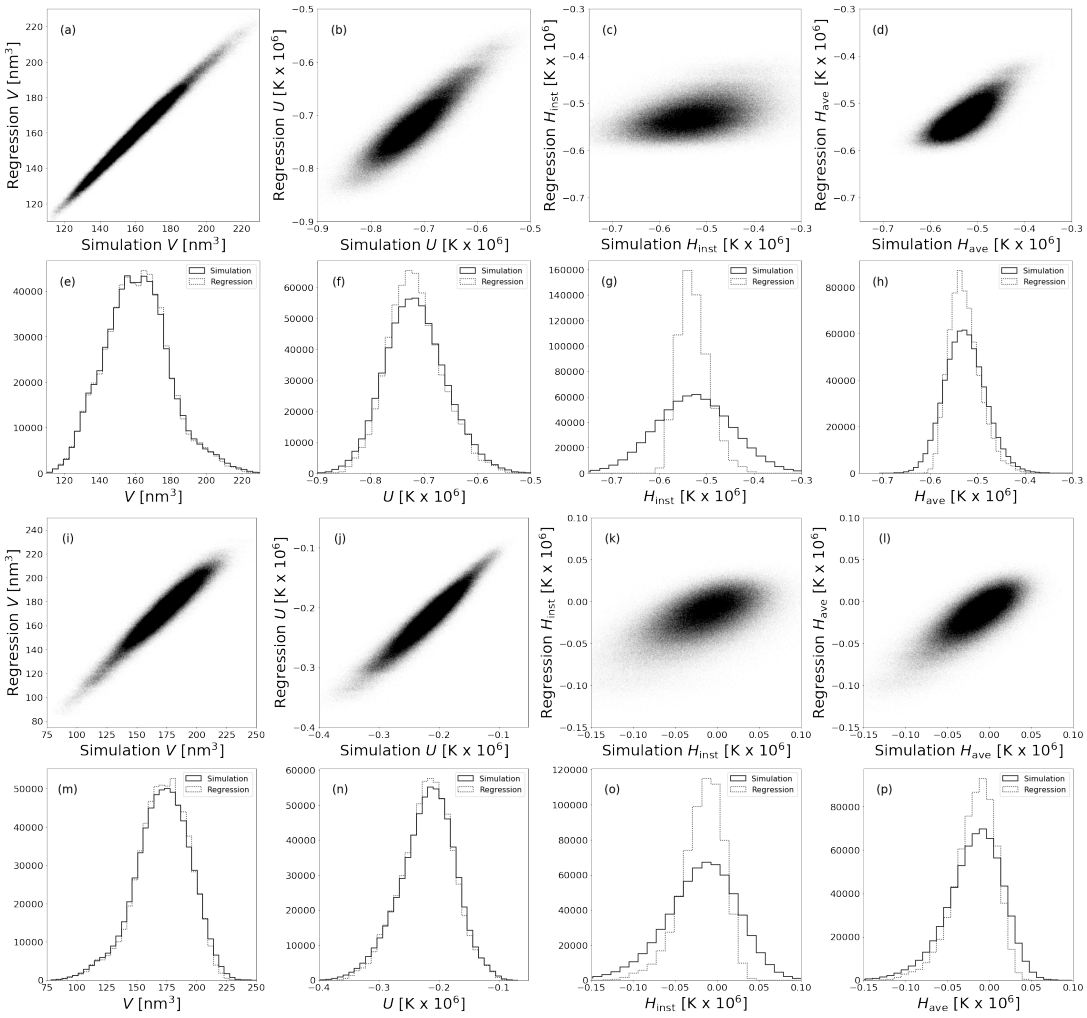


Figure 1. Parity plots and histograms illustrating the quality of the multiple linear regression for the liquid phase (a-h) and the vapor phase (i-p) of the model natural gas condensate in the NpT-Gibbs ensemble. H_{inst} is calculated using the instantaneous U, V, and p for each simulation frame, while H_{ave} is calculated using instantaneous U and V and the ensemble average p. Mean absolute deviations (MAD) of V are 2.1 and 5.4 nm^3 for the liquid and vapor. MAD of U are 2.0 x 10^4 and 1.1 x 10^4 K for the liquid and vapor. MAD of H_{inst} are 5.9 x 10^4 and 2.7 x 10^4 K for the liquid and vapor. MAD of H_{ave} are 2.2 x 10^4 and 1.7 x 10^4 K for the liquid and vapor.

Table 1. Thermodynamic properties of a model natural gas condensate at $T = 333$ K. NVT-Gibbs simulations were performed using an initial box volume of 332.75 nm^3. NpT-Gibbs and NpT simulations were performed at $p = 16220$ MPa. All simulations used 100,000 MC cycles of production. Uncertainties reported are 95% confidence intervals from 64 independent measurements of each quantity. Partial molar properties for GEMC simulations are calculated using MLR; partial molar properties for NpT simulations are calculated using numerical differentiation from 5 state points around the center (6 for n-decane).

	Liquid			Vapor		
	NVT-Gibbs	NpT-Gibbs	NpT	NVT-Gibbs	NpT-Gibbs	NpT
N_{C1}	561_{15}	580_{17}	561	715_{15}	696_{17}	715
N_{C4}	303_5	308_5	303	121_5	116_5	122
N_{C10}	116.5_6	117.0_5	117	8.5_6	8.0_5	8
x_{C1}	0.570_4	0.574_4	0.5719	0.848_3	0.851_3	0.8462
x_{C4}	0.3099_{17}	0.308_2	0.3089	0.142_2	0.140_2	0.1444
x_{C10}	0.1200_{19}	0.118_2	0.1193	0.0098_4	0.0095_4	0.0095
p [kPa]	16230_{50}	16240_{20}	16235_{19}	16220_{40}	16230_{10}	16224_6
V [nm^3]	157_3	161_3	157.23_4	175_3	171_4	175.90_6
U [K x 10^3]	-706_8	-716_9	-708.1_3	-228_8	-220_9	-226.86_{10}
$E_{pV,\mathrm{inst}}$ [K x 10^3]	184_3	189_4		206_4	201_5	
$E_{pV,\mathrm{ave}}$ [K x 10^3]	185_3	190_4	184.9_2	206_4	201_5	206.70_9
H_{inst} [K x 10^3]	-521_5	-527_5		-22_4	-18_3	
H_{ave} [K x 10^3]	-521_5	-526_5	-523.2_4	-22_4	-18_3	-20.16_{15}
$\bar{V}_{\mathrm{C1}}$ [L/mol]	0.0954_{12}	0.0892_9	0.0889_9	0.1413_{17}	0.1536_5	0.1545_{14}
$\bar{V}_{\mathrm{C4}}$ [L/mol]	0.079_2	0.0927_{19}	0.0940_{18}	0.034_9	-0.018_3	-0.0184_{15}
$\bar{V}_{\mathrm{C10}}$ [L/mol]	0.146_4	0.140_4	0.1403_{17}	0.021_{17}	-0.249_{12}	-0.257_5
$\bar{U}_{\mathrm{C1}}$ [kJ/mol]	0.07_7	-0.64_4	-0.60_7	-1.35_4	-1.09_2	-1.06_4
$\bar{U}_{\mathrm{C4}}$ [kJ/mol]	-12.79_{19}	-11.36_{15}	-11.35_{15}	-7.2_2	-8.27_{11}	-8.47_4
$\bar{U}_{\mathrm{C10}}$ [kJ/mol]	-17.5_4	-17.8_4	-18.06_{14}	-6.2_5	-12.9_4	-12.66_{11}
$\bar{H}_{\mathrm{C1,inst}}$ [kJ/mol]	0.86_7	0.84_6		1.16_4	1.40_2	
$\bar{H}_{\mathrm{C4,inst}}$ [kJ/mol]	-10.10_{18}	-9.90_{19}		-7.6_2	-8.53_{13}	
$\bar{H}_{\mathrm{C10,inst}}$ [kJ/mol]	-15.1_4	-15.6_4		-10.4_5	-16.9_5	
$\bar{H}_{\mathrm{C1,ave}}$ [kJ/mol]	1.62_9	0.81_5	0.85_{11}	0.95_6	1.41_2	1.47_6
$\bar{H}_{\mathrm{C4,ave}}$ [kJ/mol]	-11.5_2	-9.85_{18}	-9.77_{21}	-6.7_4	-8.55_{14}	-8.76_6
$\bar{H}_{\mathrm{C10,ave}}$ [kJ/mol]	-15.1_4	-15.5_4	-15.89_{19}	-5.9_7	-16.9_5	-16.77_{18}

Table 2. Comparison of observable thermodynamic properties from MC simulation to those calculated using the PR EOS, both with and without the binary interaction parameters k_{ij}. Molecular simulations with the TraPPE force field have both the instantaneous and the ensemble average observables reported for $\Delta\bar{H}_{tr}$.

		PR with k_{ij}	PR without k_{ij}	TraPPE	
Pressure [kPa]		16900	15800	16220_{40}	
Vapor Composition	C1	0.8543	0.8642	0.851_3	
	C4	0.137	0.1285	0.140_2	
	C10	0.0087	0.0073	0.0095_4	
ρ_{liq} [kg/m^3]		433.2	433.2	453.9_{39}	
ρ_{vap} [kg/m^3]		192.2	174.5	184.1_{25}	
Liquid $\bar{V}_i$ [L/mol]	C1	0.0881	0.0881	0.0892_9	
	C4	0.1091	0.1091	0.0927_{19}	
	C10	0.1502	0.15	0.140_4	
Vapor $\bar{V}_i$ [L/mol]	C1	0.144	0.155	0.1536_5	
	C4	-0.01270	-0.02568	-0.018_3	
	C10	-0.211	-0.2926	-0.249_{12}	
$\Delta\bar{U}_{tr}$ [kJ/mol]	C1	0.650	0.655	0.45_5	
	C4	-3.754	-4.731	-3.1_2	
	C10	-7.670	-9.283	-4.9_5	
				inst	ave
$\Delta\bar{H}_{tr}$ [kJ/mol]	C1	-0.291	-0.395	-0.56_7	-0.60_7
	C4	-1.694	-2.601	-1.4_2	-1.3_2
	C10	-1.563	-2.292	1.4_6	1.4_6

Table 3. Partial molar volumes $\bar{V}_i$ [L/mol], partial molar internal energies $\bar{U}_i$ and enthalapy of transfer $\Delta\bar{H}_{\text{tr,ave}}$ [kJ/mol] for components in the LLE systems. $\bar{H}_i \approx \bar{U}_i$ because the pV contribution to enthalpy is negligible.

	$x_{\text{H2O}}^{\text{total}}$	Phase	EtOH	n-Hexane	n-Dodecane	Water
$\bar{V}_i$	0.1	EtOH-rich	0.0612_6	0.138_3	0.240_2	0.019_3
		C12-rich	0.0636_9	0.134_3	0.2335_4	0.023_5
	0.25	EtOH-rich	0.0608_{19}	0.143_5	0.245_5	0.019_4
		C12-rich	0.0648_8	0.139_5	0.23329_{12}	0.023_4
	0.36	EtOH-rich	0.0611_{13}	0.143_4	0.242_7	0.0180_{15}
		C12-rich	0.0646_{11}	0.135_3	0.23334_{10}	0.025_5
$\bar{U}_i$	0.1	EtOH-rich	-33.8_3	-10.1_{13}	-11.1_{11}	-37.1_{18}
		C12-rich	-31.5_6	-12.9_{17}	-16.5_3	-32.3_{41}
	0.25	EtOH-rich	-33.5_6	-4.1_{43}	-6.1_{25}	-38.5_{12}
		C12-rich	-30.3_8	-11_2	-16.53_{11}	-33.1_{36}
	0.36	EtOH-rich	-33.0_5	-4.2_{25}	-5.7_{31}	-39.1_6
		C12-rich	-30.4_4	-11.9_8	-16.51_5	-33_2
$\Delta\bar{H}_{\text{tr,ave}}$	0.1		-2.3_7	3_2	5_1	-5_4
	0.25		-3.1_{11}	7_6	10_2	-6_4
	0.36		-2.5_{10}	7_3	11_3	-6_2

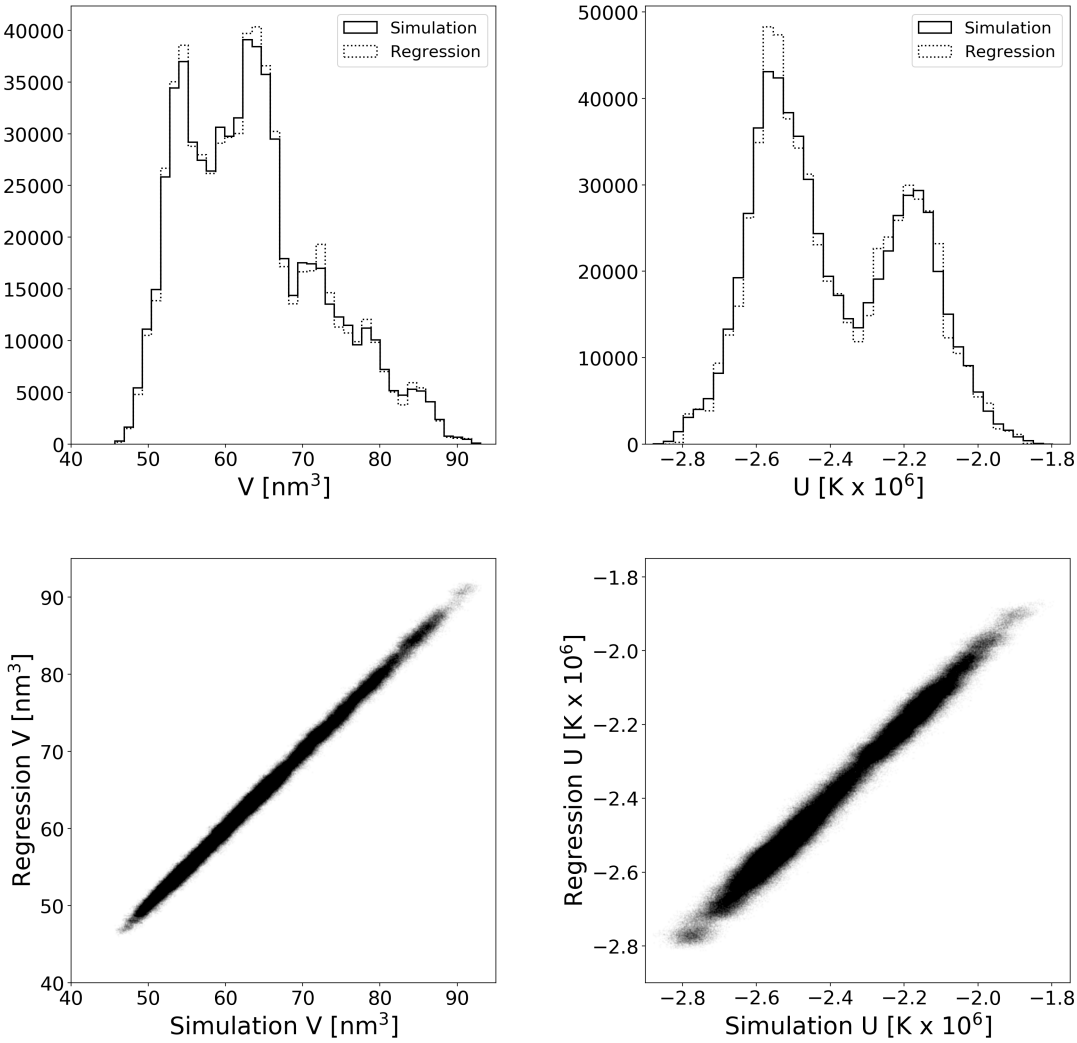


Figure 2. Parity plots and histograms illustrating the quality of the multilinear regression for the EtOH-rich liquid in the LLE system, with $x_{\mathrm{H2O}}^{\mathrm{total}} = 0.1$. Mean absolute deviations of V and U are 0.52 nm^3 and 2.2×10^4 K, respectively.

Table 4. Thermodynamic properties of 1,5-pentanediol and ethanol adsorbing into MFI at 323 K and 1 bar, from solutions with low (7.4%) and high (42%) concentrations of diol.

	$x_{\text{Diol}} = 7.4\%$		$x_{\text{Diol}} = 42\%$	
	Zeolite	Liquid	Zeolite	Liquid
N_{EtOH}	127.3_{10}	295.8_{14}	32_3	150_3
N_{Diol}	21.7_7	23.5_7	79.9_{16}	110.4_{17}
x_{EtOH}	0.854_5	0.926_2	0.28_2	0.576_9
x_{Diol}	0.146_5	0.074_2	0.72_2	0.424_9
V [nm^3]		33.86_{13}		34.5_3
U [K x 10^3]	-936_4	-1414_4	-1059_6	-1455_6
$E_{pV,\text{ave}}$ [K x 10^3]		0.245_1		0.250_2
$\bar{V}_{\text{EtOH}}$ [L/mol]		0.0605_4		0.059_3
$\bar{V}_{\text{Diol}}$ [L/mol]		0.103_4		0.107_4
$\bar{U}_{\text{EtOH}}$ [kJ/mol]	-46.2_4	-34.77_{18}	-49_3	-37_2
$\bar{U}_{\text{Diol}}$ [kJ/mol]	-88_2	-63_2	-90.7_{15}	-60_3
	EtOH	Diol	EtOH	Diol
ΔG_{tr} [kJ/mol]	5.9_3	2.53_{11}	3.97_4	1.92_{19}
ΔU_{tr} [kJ/mol]	-12_3	-31_3	-11.4_3	-25_2
$\Delta H_{\text{tr,ave}}$ [kJ/mol]	-12_3	-31_3	-11.4_3	-25_2
$\Delta S_{\text{tr,ave}}$ [J/mol K]	-56_{11}	-104_{10}	-47.5_{10}	-84_8

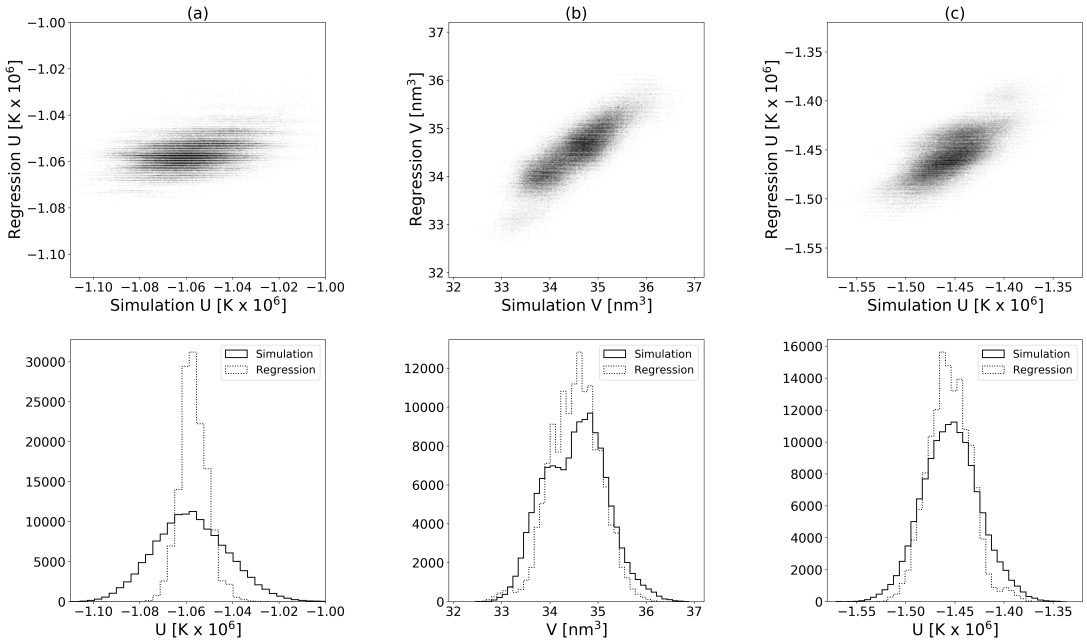
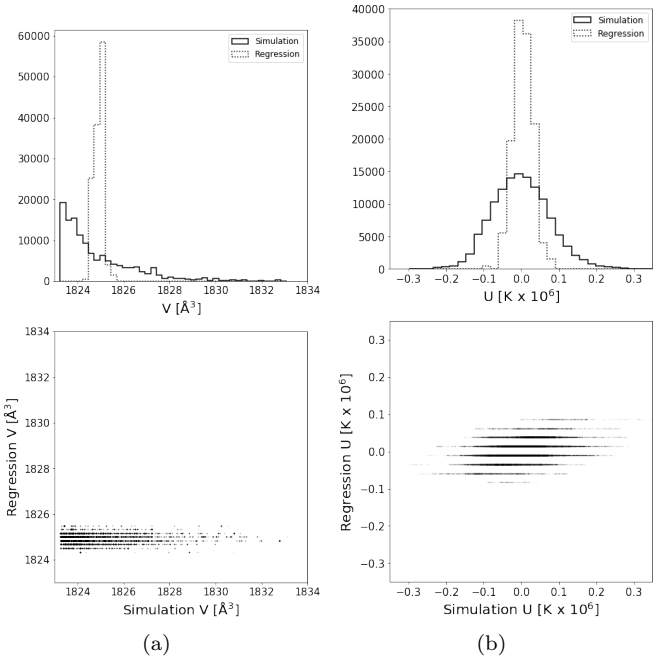


Figure 3. Parity plots and histograms illustrating the quality of the MLR for the adsorbed phase (a) and the liquid phase (b) and (c) for 1,5-pentanediol + ethanol adsorption into MFI from a solution with high (42%) concentration of diol. Mean absolute deviations of V and U for the adsorbed phase are 0.28 nm^3 and 1.6×10^4 K respectively and for the liquid phase are 0.28 nm^3 and 1.15×10^4 K respectively.

Table 5. Partial molar properties of the reactive N/O system in the reactive NpT ensemble, and reaction properties for $N_2 + O_2 \rightarrow 2NO$.

	NO	O_2	N_2
N_i	13.3_5	25.4_3	9.4_3
$\bar{V}_i$ [Å^3/molec]	38.08_7	37.99_3	38.003_{17}
$\bar{U}_i$ [kJ/mol]	29_{11}	-11_5	-13_5
$\Delta\bar{V}_{\text{rxn}}$ [Å^3/molec]	0.1_2		
$\Delta\bar{U}_{\text{rxn}}$ [kJ/mol]	81_{32}		



(a) (b)

Figure 4. Parity plots and histograms illustrating the quality of the ridge regression fit for the reactive N/O system.