Factors influencing clathrate hydrate stability in equilibrium with liquid water: Insights from information-based statistical analysis

Matthew Lasich *, Amir H. Mohammadi, Deresh Ramjugernath

Thermodynamics Research Unit, School of Engineering, University of KwaZulu-Natal, Mazisi Kunene Road, Durban 4041, South Africa

ARTICLE INFO

ABSTRACT

Screening of potential hydrate promoter/inhibitor species is essential in the investigation of clathrate hydrates in many areas of industrial interest. However, experimental analyses do not yield sufficient information to determine whether molecular size, dipole moment, molecular mass, or any other parameters influence the thermodynamic stability of clathrate hydrates in general. The present contribution employs a mutual information-based nonparametric statistical analysis to identify trends in a multivariate data set for 21 clathrate hydrate formers. It is showed through the introduction of a simple correlation that prediction of the heat of dissociation of a clathrate hydrate may be sufficient to predict the entire phase equilibrium curve above the ice point. The dipole moment and molecular volume of hydrate forming species influence the dissociation pressure curve, and distinctive behaviours may be observed for cyclic gas species as compared to non-cyclic species.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Clathrate hydrates are an ice-like crystalline inclusion compound formed by intermolecularly hydrogen-bonded water molecules enclosing particles of a guest gas species within a lattice structure consisting of several "cages" or cavities. These materials are significant in offshore natural gas exploitation operations, where methane hydrates frequently form, causing blockages in pipelines [1,2]. Other areas of interest in clathrate hydrate research include their use in industrial separation processes [2–4] and in storage of energy carrier gases such as methane [5,6] and hydrogen [7,8].

With regard to pipeline blockages in natural gas exploitation, it is necessary to determine which chemical species can serve as inhibitors of the hydrate formation process at the pipeline conditions (i.e., T and P). Screening potential inhibitors is therefore an important and necessary task in the natural gas extraction industry. In terms of using clathrate hydrates as an energy storage medium, it is essential to determine which species can serve as hydrate promoters. Since clathrate hydrates form at very high pressures at ambient temperatures—e.g., hydrogen [7]—it is necessary to determine which guest species can serve to lower the pressure at which the clathrate hydrate can form. Therefore, screening of, for instance, hydrogen hydrate promoters would be essential if clathrate hydrates are to serve as a viable energy storage material in any planned future hydrogen economy [8].

Recently, a new tool for analyzing data sets containing large numbers of dependent and independent variables has been developed [9]: The maximal information coefficient (MIC). Firstly, the MIC falls under a broader category of statistics known as maximal information-based non-parametric (MINE) statistics, which are used to identify and classify relationships between variables in data sets. The MIC was originally used [9] to detect novel relationships, as well as to confirm known relationships, within a wide array of available data (global health data, genomic expression information, human gut microbiota, and major-league baseball statistics). This contribution seeks to employ MINE statistics to determine the presence of relationships, if any, between the molecular properties of various hydrate-forming gas species and the thermodynamic stability of the associated clathrate hydrate systems. This is the first time MINE statistics have been used in such a manner for clathrate hydrates.

2. Theory and methods

2.1. Systems considered

Several clathrate hydrate systems were considered in this study. Altogether, 21 guest species were considered. While not all species formed the same clathrate hydrate structure (i.e., sI, sII or sH), the intermolecular forces and molecular properties influencing clathrate hydrate stability are fundamentally similar for all clathrate hydrate systems. In addition, only liquid water + clathrate hydrate phase equilibria were considered (i.e., T > 273.15 K), such that similar phase equilibria are studied in parallel. Table 1 summarizes the data set used in this study.

2.2. Variables considered

Three molecular properties of the clathrate hydrate guest species where considered: Molecular mass, dipole moment, and van der

* Corresponding author.
  E-mail address: lasichm@ukzn.ac.za (M. Lasich).

Table 1
Summary of clathrate hydrate systems considered in this study.

Species	Temperature range/K	Data points
Furan [10]	277.0–277.4	5
Cyclopentane [10]	279.9–280.3	5
Argon [11]	330.8–406.1	49
Methane [12,13]	273.7–298.1	11
Carbon monoxide [14]	274.5–284.9	9
Carbon dioxide [14]	277.5–282.5	3
Ethene [15]	273.3–289.6	11
Ethane [15]	273.8–283.4	3
1,3-Dimethylcyclohexane [6]	276.0–282.5	10
Propane [12,16]	275.0–289.3	11
1,1,1,2-Tetrafluoroethane [17]	273.5–283.1	9
1,1-Dichloro-1-fluoroethane [17]	273.4–281.5	8
1,1-Difluoroethane [17]	273.4–288.2	11
Hydrogen sulfide [18]	277.7–301.3	19
Carbonyl sulfide [18]	274.7–281.5	19
Sulfur dioxide [19]	279.9–285.2	12
Methyl mercaptan [19]	283.5–285.0	5
Isobutane [20,21]	273.4–275.1	6
Nitrogen [13,22–24]	273.4–286.2	12
Hydrogen [25]	294.4–382.8	17
Tetrafluoromethane [26]	276.6–317.1	37

Waals volume. The dipole moments for each species were estimated in ArgusLab [27–31] at the Hartree-Fock level using the PM3 Hamiltonian [32,33]. The van der Waals volumes were determined by using a modified Bondi method [34]. In order to describe the phase equilibria, a linear trend line was fitted to the natural logarithm of the dissociation pressure (P) data expressed as a function of the reciprocal of the temperature (T):

$$\ln P = a(1/T) + b \tag{1}$$

The magnitude of the slope (a) describes the sensitivity of the dissociation pressure to changes in the system temperature, and the intercept (b) is related to the dissociation pressure itself. The slope a is related to the heat of dissociation ($\Delta H_{diss.}$) and the gas phase compressibility factor (Z) by means of the Clausius-Clapeyron equation [35]:

$$d(\ln P)/d(1/T) = -\Delta H_{diss.}/(ZR) \tag{2}$$

It may be noted that the Clausius-Clapeyron equation can be presented in integral form, assuming an approximately constant heat of dissociation, and so the parameters (a and b) presented in Eq. (1) can be related to the heat of dissociation and a reference temperature (T_R) and a reference pressure (P_R):

$$a = -\Delta H_{diss.}/(ZR) \tag{3}$$

$$b = \ln P_R + \Delta H_{diss.}/(T_R ZR) \tag{4}$$

In order to assess the stability of the clathrate hydrate systems at ambient conditions, T was adjusted so as to be relative to ambient temperature (i.e., T = 298.15 K). The magnitude of the y-intercept (b) of this linear trend line would then serve to describe the. A summary of the molecular properties and phase equilibrium trend line fits are presented in Table 2. It is clear that using a linear trend to describe the natural logarithm of the dissociation pressure as a function of the reciprocal of temperature is reasonable, since the correlation coefficient (R^2) values of such fits are >0.770 in all cases (and above 0.902 in all but one case).

2.3. Data exploration statistics

In general, MINE statistics can be used to identify associations between variables in large data sets, as well to characterize them according to measures such as monotonicity and non-linearity. However, many MINE statistics and functions are not truly general, and with sufficiently large sample sizes, such statistics should capture wide ranges of associations or relationships, or all of the functional relationships. The case of the latter point is especially significant, as many relationships are not well described by a single function (such as a superposition of functions) [36–39]. Moreover, certain MINE statistics fail to show correlation for certain function types; for a parabolic function, the Spearman rank correlation coefficient is −0.01 (suggesting the absence of a relationship), whilst the mutual information coefficient of Kraskov and co-workers [40] would be 3.15 (weakly suggesting the presence of a trend) [9]. Therefore, since the MIC overcomes such limitations, it can be a preferred tool in the search for relationships between variables in multivariate data sets.

The MIC was developed from the concept of drawing up a grid on the scatterplot of two variables, which would encapsulate any relationship existing between said variables. For a given set of data, all possible grids up to a maximum grid resolution (dependent on the sample size) are explored. In the case of a set of bivariate data, for example, for every pair of integers (x,y) the largest possible mutual information achievable by a grid of size x-by-y is computed. These mutual information values are then normalized to compare between different grids and different dimensions, such that the normalized values lie between 0 and 1. The characteristic matrix M is then defined [9]:

$$M = (m_{x,y}) \tag{5}$$

where $m_{x,y}$ is the highest normalized information which can be achieved by any x-by-y grid, and the MIC is then the maximum value in the matrix M. MIC is subject to the constraint [9]:

$$xy < B \tag{6}$$

in which B is related to the sample size, n. The calculation for B, as proposed by Reshef and co-workers [9], is as follows:

$$B = n^{0.6} \tag{7}$$

Since every value in the matrix M is between 0 and 1, so then the MIC will always be between 0 (no relationship present at all) and 1 (a relationship definitely present). A routine developed in the literature [9] to determine the MIC has been employed in the present study. This routine does not optimize over all possible grids for a given data set, but instead makes use of a dynamic programming algorithm to optimize over a subset of all possible grids. This approach has been shown to approximate the true value of the MIC very closely [9].

Additionally, a measure of non-linearity, ν, can also be developed by incorporating the Pearson product-moment correlation coefficient, ρ [9]:

$$\nu = MIC - \rho^2 \tag{8}$$

such that as ν tends to 0 a linear relationship is present, whereas for the case in which it tends to unity then a non-linear relationship is present.

Other quantities of interest in any data exploration exercise which makes use of the MIC are the maximum asymmetry score (MAS), maximum edge value (MEV), and the minimum cell number (MCN). The MAS measures the deviation from monotonicity of the data set in question, and defined by [9]:

$$MAS = \max_{xy < B} | M_{x,y} - M_{y,x} | \tag{9}$$

MEV describes how close the data set appears to originate from a single function, since it is entirely possible that several subsets may