**K-value program for crude oil components at high pressures based on PVT laboratory data and genetic programming**

K.A. Fattah

*Petroleum and Natural Gas Engineering Department, College of Engineering, King Saud University, P.O. 800, Riyadh 11421, Saudi Arabia*

Received 6 April 2011; accepted 21 June 2011
Available online 8 September 2011

**KEYWORDS**

K-value; Correlation; Genetic program; PVT lab report; Crude oil; High pressures

**Abstract** Equilibrium ratios play a fundamental role in understanding the phase behavior of hydrocarbon mixtures. They are important in predicting compositional changes under varying temperatures and pressures in the reservoirs, surface separators, and production and transportation facilities. In particular, they are critical for reliable and successful compositional reservoir simulation. Several techniques are available in the literature to estimate the K-values. This paper presents a new model for predicting K values with genetic programming (GP). The new model is applied to multicomponent mixtures. In this paper, 732 high-pressure K-values obtained from PVT analysis of 17 crude oil and gas samples from a number of petroleum reservoirs in Arabian Gulf are used. Constant Volume Depletion (CVD) and Differential Liberation (DL) were conducted for these samples. Material balance techniques were used to extract the K-values of crude oil and gas components from the constant volume depletion and differential liberation tests for the oil and gas samples, respectively. These K-values were then used to build the model using the Discipulus software, a commercial Genetic Programming system, and the results of K-values were compared with the values obtained from published correlations. Comparisons of results show that the currently published correlations give poor estimates of K-values for all components, while the proposed new model improved significantly the average absolute deviation error for all components. The
average absolute error between experimental and predicted $K$-values for the new model was 4.355% compared to 20.5% for the Almehaideb correlation, 76.1% for the Whitson and Torp correlation, 84.27% for the Wilson correlation, and 105.8 for the McWilliams correlation.

© 2011 King Saud University. Production and hosting by Elsevier B.V. All rights reserved.

1. Introduction

A phase is defined as that part of a system which is uniform in physical and chemical properties, homogeneous in composition, and separated from other coexisting phases by a definite boundary surface. The most important phases occurring in petroleum production are the hydrocarbon liquid phase and the gas phase. Water is also commonly present as an additional liquid phase. Equilibrium ratios, more commonly known as $K$-values, relate the vapor mole fraction, $y_i$, to the liquid mole fraction, $x_i$, of a component $i$ in a mixture,

$$K_i = \frac{y_i}{x_i} \quad (1)$$

In a fluid mixture consisting of different chemical species at high pressure, $K$-values are dependent on pressure, temperature, and the composition of the mixture.

The objective of this work is to evaluate the published empirical correlations that could possibly be used for computing $K$-values for high-pressure systems, namely Wilson, Whitson and Torp, McWilliams, and Almehaideb correlations, using the experimental $K$-values extracted through material balance techniques from PVT tests performed on Arabian Gulf petroleum samples, and to develop a new model for the Arabian Gulf crudes formulated using the Discipulus software.

2. Literature review

There are a number of methods for predicting $K$ values. Usually these methods compute $K$-values explicitly or iteratively. The explicit methods correlate $K$ values with component parameters (i.e., the critical properties) or the mixture parameters (i.e., the convergence pressure). Iterative methods based on the equation of state (EOS) are usually tuned with binary interaction parameters (Habiballah et al., 1996). In this paper, we will focus on only four widely used correlations that are applicable for high pressure $K$-values: Wilson, Whitson and Torp, McWilliams, and Almehaideb correlations.

Wilson (Wilson, 1968) proposed a simplified expression in the form

$$K_i = \frac{P_{ci}}{P} \exp \left[ \frac{5.37(1 + \omega_i)(1 - T_{ci}/T)}{T_{ci}} \right] \quad (2)$$

where $P_{ci}$ is the critical pressure of component $i$ (psia), $T_{ci}$ is the critical temperature of component $i$ ($^{\circ}$R), $\omega_i$ is the acentric factor of component $i$, $p$ is the system pressure (psia), $T$ is the system temperature ($^{\circ}$R). This correlation is valid typically at low pressures.

Whitson and Torp (1983) modified Wilson’s equation to accommodate the compositional effects at high pressures by incorporating the convergence pressure, $P_k$ (psia), to obtain

$$K_i = \left( \frac{P_{ci}}{P_k} \right)^{A_{i-1}} \left( \frac{P_{ci}}{P} \right) \exp \left[ \frac{5.37A(1 + \omega_i)(1 - T_{ci}/T)}{T_{ci}} \right] \quad (3)$$

where

$$A = 1 - \left( \frac{P - 14.7}{P_k - 14.7} \right) \quad (4)$$

The convergence pressure, $P_k$, concept is based on the observation that if we hold a hydrocarbon mixture of a certain composition at a constant temperature and increase the pressure, then the equilibrium values for all its components converge toward a common value of unity at certain pressure called the convergence pressure. In computing the high pressure $K$-values using the Whitson–Torp correlation, one has to be careful not to use the correlation for pressures above the convergence pressure.

DePriester (DePriester, 1953) presented $K$-value charts for light hydrocarbons vs. pressure and temperature that are valid up to around 6000 psi pressures or more. McWilliams (McWilliams, 1973) fitted these charts to the following polynomial equation

$$\ln K = \frac{a_{T1}}{T^2} + \frac{a_{T2}}{T} + a_{T3} + a_{p1} \ln p + \frac{a_{p2}}{p^2} + \frac{a_{p3}}{p} \quad (5)$$

where $T$ in $R$ and $p$ in psia. $a_{T1}$, $a_{T2}$, $a_{T3}$, $a_{p1}$, $a_{p2}$, and $a_{p3}$ are constants. McWilliams equation is valid from 365.7 $R$ to 851.7 $R$ and from 14.69 psia to 870.7 psia.

Almehaideb et al. (2003) proposed a correlation for UAE crudes at high pressures. Their correlation was obtained by modifying the McWilliams correlation but with the additional term as a function of $\omega$ used for the $C_7+$ fraction only, and an adjustment for the effect of composition similar to the one suggested by Whitson and Torp. The developed equation is to better fit the UAE data using multi-variable regression technique.

$$K_i = \left( \frac{P_{ci}}{T_k} \right)^{A_{i-1}} \left( \frac{P_{ci}}{P} \right) \exp[A \times k_i] \quad (6)$$

where

$$A = 1 - \left( \frac{P - 14.7}{P_k - 14.7} \right) \quad (7)$$

$$k_i = \frac{a_{T1}}{T^2} + \frac{a_{T2}}{T} + a_{T3} + a_{p1} \ln p + \frac{a_{p2}}{p^2} + \frac{a_{p3}}{p} + \frac{a_\omega}{\omega} \quad (8)$$

3. Extracting $K$-values from CVD and DL tests

Experimental $K$-values can be extracted from three types of PVT tests, namely differential liberation (DL) tests, constant volume depletion (CVD) tests, and separation tests, provided that the gas composition exiting the PVT cell is performed at each pressure stage. The first two tests are normally carried out at high pressures, approximating the reservoir conditions for crude oil and gas condensate systems, respectively, while the separation tests are carried out at pressures approaching surface operating pressures.

To obtain the experimental $K$-values at high pressures, data are extracted from the differential liberation and constant volume depletion using the material balance technique. It is clear that the same methodology will be used in both cases, at high and low pressures, but the equation that will be used for the
extraction of the $K$-values for high pressures will be suitable for the differential liberation and constant volume depletion tests. The equations reported by McCain (William and McCain, 1990) and Whitson and Torp (1983) will be used as the starting point (Almehaideb et al., 2003).

An example of the calculations to extract $K$-values from CVD laboratory data for one of the samples is shown in Table 1. Also, an example of the calculations to extract $K$-values from DL laboratory data for one of the samples is shown in Table 2.

### 4. Genetic programming

Genetic algorithms, evolution strategies and genetic programming belong to the class of probabilistic search procedures known as Evolutionary Algorithms that use computational models of natural evolutionary processes to develop computer-based problem solving systems. Solutions are obtained using operations that simulate the evolution of individual structures through mechanism of reproductive variation and fitness based selection. Due to their reported robustness in practical applications, these techniques are gaining popularity and have been used in a wide range of problem domain. The main difference between genetic programming and genetic algorithm is the representation of the solution. Genetic programming creates computer programs as the solution whereas genetic algorithm creates a string of numbers to represent the solution. Genetic programming is based on the Darwinian principle of reproduction and survival of the fittest and analogs of naturally occurring genetic operations such as crossover and mutation (Koza, 1997). Genetic programming uses four steps to solve a problem (Koza, 1992):

1. Generate an initial population of random compositions of the functions and terminals (input) of the problem.

### Table 1 Calculation of $K$-values form CVD data for well G1, pressure step from 3970 to 3559 psia and 245 K.

<table>
<thead>
<tr>
<th>Comp.</th>
<th>$M_w$</th>
<th>$z_i$</th>
<th>$y_i$</th>
<th>$M_w * z_i$</th>
<th>$\sum \Delta n_i Z_i$</th>
<th>$M_w * y_i$</th>
<th>$\omega_i$</th>
<th>$X_{jk}$</th>
<th>$Y_{jk}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>N(_2)</td>
<td>28.01</td>
<td>0.0024</td>
<td>0.25</td>
<td>0.0672</td>
<td>0.0002</td>
<td>0.0700</td>
<td>0.0024</td>
<td>0.0001</td>
<td>20.292</td>
</tr>
<tr>
<td>CO(_2)</td>
<td>44.01</td>
<td>0.0415</td>
<td>4.22</td>
<td>1.8264</td>
<td>0.0033</td>
<td>1.8572</td>
<td>0.0415</td>
<td>0.0324</td>
<td>1.2935</td>
</tr>
<tr>
<td>H(_2)S</td>
<td>34</td>
<td>0.021</td>
<td>2.06</td>
<td>0.7140</td>
<td>0.0016</td>
<td>0.7004</td>
<td>0.0210</td>
<td>0.0256</td>
<td>0.8140</td>
</tr>
<tr>
<td>C(_1)</td>
<td>16.04</td>
<td>0.7095</td>
<td>72.9</td>
<td>11.380</td>
<td>0.0565</td>
<td>11.693</td>
<td>0.7088</td>
<td>0.5365</td>
<td>1.3367</td>
</tr>
<tr>
<td>C(_2)</td>
<td>30.07</td>
<td>0.0595</td>
<td>5.99</td>
<td>1.7892</td>
<td>0.0047</td>
<td>1.8012</td>
<td>0.0595</td>
<td>0.0572</td>
<td>1.0415</td>
</tr>
<tr>
<td>C(_3)</td>
<td>44.1</td>
<td>0.0376</td>
<td>3.67</td>
<td>1.6582</td>
<td>0.0029</td>
<td>1.6185</td>
<td>0.0376</td>
<td>0.0467</td>
<td>0.7965</td>
</tr>
<tr>
<td>IC(_4)</td>
<td>58.12</td>
<td>0.0091</td>
<td>0.87</td>
<td>0.5289</td>
<td>0.0007</td>
<td>0.5056</td>
<td>0.0091</td>
<td>0.0137</td>
<td>0.6518</td>
</tr>
<tr>
<td>NC(_4)</td>
<td>58.12</td>
<td>0.0188</td>
<td>1.79</td>
<td>1.0927</td>
<td>0.0014</td>
<td>1.0403</td>
<td>0.0189</td>
<td>0.0325</td>
<td>0.5607</td>
</tr>
<tr>
<td>IC(_5)</td>
<td>72.15</td>
<td>0.0081</td>
<td>0.75</td>
<td>0.5844</td>
<td>0.0006</td>
<td>0.5411</td>
<td>0.0081</td>
<td>0.0127</td>
<td>0.6243</td>
</tr>
<tr>
<td>NC(_5)</td>
<td>72.15</td>
<td>0.0103</td>
<td>0.95</td>
<td>0.7431</td>
<td>0.0008</td>
<td>0.6854</td>
<td>0.0193</td>
<td>0.0194</td>
<td>0.5101</td>
</tr>
<tr>
<td>C(_6)</td>
<td>84.89</td>
<td>0.0171</td>
<td>1.49</td>
<td>1.4516</td>
<td>0.0013</td>
<td>1.2649</td>
<td>0.0172</td>
<td>0.0308</td>
<td>0.5364</td>
</tr>
<tr>
<td>C(_7)</td>
<td>130.13</td>
<td>0.0651</td>
<td>5.06</td>
<td>8.4715</td>
<td>0.0047</td>
<td>7.5952</td>
<td>0.0656</td>
<td>0.1926</td>
<td>0.3089</td>
</tr>
</tbody>
</table>

### Nomenclature

- $a_T$, $a_p$: correlation parameters
- $B_o$: oil formation volume factor (bbl/STB)
- $K_i$: equilibrium ratio for component $i$
- $M$: molecular weight (lbm/lb mol)
- $n$: mole fraction
- $n_{ji}$: mole fraction of gas separated in separator stage based on 1 lb mol of feed
- $n_{ji}$: mole fraction of liquid separated in pressure stage $j$ based on 1 lb mol of feed
- $p$: pressure (psia)
- SD: percent standard deviation
- AAD: percent average absolute deviation
- $R$: universal gas constant
- $R_s$: gas to oil ratio in pressure stage $j$ where $j$ is 1, 2, etc.
- SCF/STB: standard cubic foot/bbl
- $T$: temperature ($^\circ$F)
- $V$: volume (ft$^3$)
- $x_i$: mole fraction of component $i$ in liquid phase
- $y_i$: mole fraction of component $i$ in gas (vapor) phase
- $z_i$: mole fraction of component $i$ in feed

### Greeks

- $\rho_{STO}$: density of stock-tank oil (lbm/ft$^3$)
- $\rho_o$: density of reservoir oil (lbm/ft$^3$)
- $\omega_i$: acentric factor for component $i$
Execute each program in the population and assign a fitness value.

Create a new offspring population of computer programs by copying the best programs and creating new ones by mutation and crossover.

Designation of the best computer program in the generation.

5. Developed $K$-value correlation using genetic program

The Discipulus software, a commercial Genetic Programming system, was used to correlate the new $K$-value correlation. Discipulus wrote computer programs from data given to it. These given data were classified into “training data,” “validation data,” and “test data.”
data” and “testing data” that provided to Discipulus program. These data files contained matched inputs and outputs data. From them, Discipulus created models that allow us to predict outputs from similar inputs. The models were created as computer programs in Java, C, or assembler program. The input data for our new correlation are:

- Pressure, psi
- Reservoir temperature, R
- Component molecular weight, lb/mole
- \( C_{7+} \) molecular weight, lb/mole
- Component boiling temperature, R
- Component critical temperature, R
- Component critical pressure, psi
- Component acentric factor
- Constant \( P_k \)

\[
P_k = -2381.8542 + 46.341487(MW \times \gamma)_{C_{7+}} \]
\[
+ \sum_{i=1}^{3} a_i[(MW \times \gamma)_{C_{7+}}/(T - 460)]^i
\]  \hspace{1cm} (9)

where \( (MW)_{C_{7+}} \) = molecular weight of \( C_{7+} \), \( (\gamma)_{C_{7+}} \) = specific gravity of \( C_{7+} \), \( a_1 - a_2 \) = correlation coefficient with the

---

**Figure 5**  The best program fitness improvement with time.

**Figure 6**  The extracted vs. calculated \( K \)-value for training data from genetic program.

**Figure 7**  The \( K \)-value for training data from new correlation.
following values, \( a_1 = 6124.3049, \ a_2 = -2753.2538, \ a_3 = 415.42049, \ T = \) reservoir temperature in \( R^\circ \).

The extracted \( k \)-value data were classified into three semi-equal groups, “training data,” “validation data” and “testing data”. After uploading the data files into the program with the ‘run’ command issued, the program gives different types of data that show how the run in progress improved its performance.

6. Results and discussion

Extracted \( K \)-values were statistically compared in this study with the results predicted using the Wilson correlation, Whitson and Torp correlation, McWilliams correlation, Almehaideb correlation and the new correlation.

The following is a statistical comparison of the results obtained by the four published correlations and the new correlation with the extracted values of equilibrium ratios.

Figs. 1–4 provide a general comparison between experimental and calculated \( K \)-values for all components using the
Wilson correlation, Whitson and Torp correlation, McWilliams correlation and Almehaideb correlation.

Fig. 5 shows how the fitness of the best program for the new genetic correlation created by the genetic program was improved with time. Fig. 6–11 show the match between the extracted $K$-value and the predicted $K$-values from the new genetic correlation.

Table 3 is a comparison of the statistical measures for the four published correlations with the new correlation. It shows a significant overall improvement in the predictions of $K$-values using the new Almehaideb correlation over the Almehaideb correlation, the Wilson correlation, Whitson and Torp correlation, and McWilliams correlation for the extracted data.

The average absolute error (AAE) between extracted and predicted $K$-values form the new correlation created by genetic program was 4.355%, and the standard deviation (SD) was 0.2%. The C++ code of the genetic program to calculate the new correlation was given in the appendix.

7. Conclusions

As a result of this study, the following conclusions can be cited:

1. $K$-values extracted from the CVD and DL experiments provided a direct comparison between experimental and correlated $K$-values for Arabian Gulf crude oil at high pressure.
2. $K$-values obtained from four correlations, namely the Wilson correlation, Whitson and Torp correlation, McWilliams, and the Almehaideb correlation, are poorly compared with the extracted $K$-values.
3. A new correlation Program, for the Arabian Gulf crude oil, is proposed. The Discipulus software, a commercial Genetic Programming system, was used to develop the new $K$-value correlation program. It is based on the concept of genetic algorithm.
4. The statistical comparison shows that the new correlation compares favorably well with the results from the other four correlations included in this study.

Acknowledgment

The author expresses his thanks and appreciation to the Research Center at King Saud University, College of Engineering for supporting this research.

Appendix A

This appendix gives the C++ code of the genetic program to calculate the new correlation.
float DiscipulusCFunction(float v[]) {
    long double f[8];
    long double tmp = 0;
    int cflag = 0;


    L0: f[0] -= v[8];
    L1: f[0] += v[7];
    L2: f[0] = sin(f[0]);
    L3: f[0] = -0.9636838436126709f;
    L4: f[0] = -1.907608032226563f;
    L5: f[0] = -0.1387641429901123f;
    L6: f[0] = sin(f[0]);
    L7: f[0] = fabs(f[0]);
    L8: f[0] = -f[0];
    L9: f[0] += f[0];
    L10: f[0] += v[5];
    L11: if (cflag) f[0] = f[1];
    L12: f[0] += -1.364008665084839f;
    L13: f[0] += f[0];
    L14: f[0] *= f[1];
    L15: tmp = f[1]; f[1] = f[0]; f[0] = tmp;
    L16: f[0] = -f[0];
    L17: f[0] = 1.530829906463623f;
    L18: tmp = f[0]; f[0] = f[0]; f[0] = tmp;
    L19: tmp = f[0]; f[0] = f[0]; f[0] = tmp;
    L20: if (cflag) f[0] = f[0];
    L21: f[0] = fabs(f[0]);
    L22: f[0] = sin(f[0]);
    L23: f[0] = fabs(f[0]);
    L24: cflag = (f[0] < f[0]);
    L25: f[0] = f[0];
    L26: f[0] += -1.427085638046265f;
    L27: f[0] = f[0];
    L28: tmp = f[0]; f[0] = f[0]; f[0] = tmp;
    L29: f[0] = f[0];
    L30: f[0] = 1.450522422790527f;
    L31: cflag = (f[0] < f[0]);
    L32: f[0] = -f[0];
    L33: f[0] = fabs(f[0]);
    L34: f[0] = v[7];
    L35: f[0] *= f[0];
    L36: f[0] = f[0];
    L37: f[0] = -f[0];
    L38: if (cflag) f[0] = f[0];
    L39: f[0] *= 0.6342074871063232f;
    L40: cflag = (f[0] < f[0]);
    L41: f[0] = fabs(f[0]);
    L42: f[0] = 1.086833715438843f;
    L43: f[0] = v[7];
    L44: f[0] = v[7];
    L45: f[0] *= -0.494312047958374f;
    L46: tmp = f[0]; f[0] = f[0]; f[0] = tmp;
    L47: f[0] = v[7];
    L48: f[0] = v[7];
    L49: f[0] = v[7];
    L50: f[0] = v[7];
    L51: f[0] = v[7];
    L52: f[0] = v[7];
    L53: f[0] = v[7];
    L54: f[0] = v[7];
    L55: f[0] = v[7];
    L56: f[0] = v[7];
    L57: f[0] = v[7];
    L58: f[0] = v[7];
    L59: f[0] = v[7];
    L60: f[0] = v[7];
    L61: f[0] = v[7];
    L62: f[0] = v[7];
    L63: f[0] = v[7];
    L64: f[0] = v[7];
    L65: f[0] = v[7];
    L66: f[0] = v[7];
    L67: f[0] = v[7];
    L68: f[0] = v[7];
    L69: f[0] = v[7];
    L70: cflag = (f[0] < f[0]);
    L71: f[0] = fabs(f[0]);
    L72: f[0] *= 0.00262165095800781f;
    L73: f[0] = 0.138764129901123f;
    L74: f[0] = v[2];
    L75: f[0] = f[0];
    L76: tmp = f[0]; f[0] = f[0]; f[0] = tmp;
    L77: f[0] = v[8];
    L78: f[0] = 0.00262165095800781f;
    L79: f[0] = f[0];
    L80: f[0] = v[2];
    L81: f[0] = sqrt(f[0]);
    L82: f[0] = v[2];
    L83: f[0] = f[0];
    L84: f[0] = sqrt(f[0]);
    L85: f[0] = sqrt(f[0]);
    L86: f[0] = sqrt(f[0]);
    L87: f[0] = 1.258495330810547f;
    L88: f[0] = sin(f[0]);
    L89: f[0] = v[8];
    L90: f[0] = v[8];
    L91: if (cflag) f[0] = f[0];
    L92: tmp = f[0]; f[0] = f[0]; f[0] = tmp;
    L93: f[0] = f[0];
    L94: f[0] += -0.949312047958374f;
    L95: f[0] = v[2];
    L96: f[0] = sqrt(f[0]);
    L97: f[0] = f[1]; f[1] = f[0]; f[0] = tmp;
    L98: f[0] = v[0];
    L99: f[0] = fabs(f[0]);
    L100: f[0] = f[0];
    L101: f[0] = f[0];
    L102: f[0] = f[0];
    L103: f[0] = f[0];
    L104: f[0] = f[0];
    L105: f[0] = f[0];
    L106: if (cflag) f[0] = f[0];
    L107: f[0] = fabs(f[0]);
    L108: f[0] = f[0];
    L109: f[0] = f[0];
    L110: f[0] = f[0];
    L111: f[0] = f[0];
    L112: f[0] = f[0];
    L113: f[0] = f[0];
    L114: f[0] = f[0];
    L115: f[0] = f[0];
    L116: f[0] = f[0];
    L117: f[0] = f[0];
    L118: cflag = (f[0] < f[1]);
    L119: f[1] = f[0];
    L120: f[0] = v[5];
    L121: f[0] = 0.287938747406060f;
    L122: f[0] = 1.086833715438843f;
    L123: f[0] = v[0];
    L124: f[1] = f[0];
    L125: f[0] = 0.1401152610778809f;
    L126: if (cflag) f[0] = f[1];
    L127: f[0] += v[0];
    L128: cflag = (f[0] < f[0]);
    L129: f[0] = f[0];
    L130: f[1] += f[0];
    L131: f[0] = f[0];
    L132: f[0] = f[0];
    L133: f[0] = f[0];
    L134: f[0] = v[e];
References


