

Predicting Temperature-Dependent Properties by Correlations Based on Similarity of Molecular Structures □ Application to Liquid Density

Journal:	<i>Industrial & Engineering Chemistry Research</i>
Manuscript ID:	ie-2007-01766m.R1
Manuscript Type:	Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Shacham, Mordechai; Dept. of Chemical Engineering, Ben Gurion University Brauner, Neima; Tel-Aviv University, School of Engineering; Tel-Aviv University, School of Engineering Shore, Haim; Ben-Gurion University, Industrial Engineering and Management Benson-Karhi, Diamanta; Ben-Gurion University, Industrial Engineering and Management



Predicting Temperature-Dependent Properties by Correlations Based on Similarity of Molecular Structures – Application to Liquid Density

Mordechai Shacham^{a*}, Neima Brauner^b, Haim Shore^c, Diamanta Benson-Karhi^c,

^aDepartment of Chemical Engineering, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel

^bSchool of Engineering, Tel-Aviv University, Tel-Aviv 69978, Israel

^cDepartment of Industrial Engineering, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel

Abstract

A novel method for predicting temperature-dependent properties is presented. The method involves the use of measured property values of predictive compounds that are structurally similar to the target compound, and molecular descriptor values. The quantitative structure-structure property relationship (QS2PR) is used to model a linear relationship between property values of the target and the predictive compounds. Whenever necessary, Response Modeling Methodology (RMM) can be employed to obtain a non-linear regression model for representing property data of the predictive compounds. The application of the method is demonstrated under variety of conditions by prediction of temperature-dependent liquid-density variation of 1-butene, toluene, *n*-hexane and *n*-heneicosane. It is shown that straightforward application of the proposed method provides predictions with accuracy within experimental error level. An advantage of the proposed method over other similar prediction models is that it does not require measured property values of the target compound.

Keywords: molecular descriptors; property prediction; QS2PR; RMM; saturated liquid density

* Corresponding author. Tel.: +972-8-6461481, fax: +972-8-6472916 E-mail address: shacham@bgu.ac.il (M. Shacham)

1. Introduction

Pure-compound property data are at present available only for a small fraction of compounds pertaining to diverse areas of chemistry and chemical engineering, environmental engineering and environmental impact assessment, hazard and operability analysis. Therefore, methods for reliable prediction of property data are needed. In particular, prediction of temperature-dependent properties (like vapor pressure, vapor and liquid density, viscosity or specific heat) poses a special challenge because of the limited amount of experimental data available at widely varying temperature ranges.

Current methods used to predict temperature-dependent properties can be classified into "group contribution" methods, methods based on the "corresponding-states principle" (for an extensive review of these methods see Poling *et al.*¹), and "asymptotic behavior" correlations (see, for example, Marano and Holder,²). To predict properties of a target compound, many of these methods require experimental data about that compound (e.g., critical properties), and therefore they are inapplicable for compounds not yet synthesized, or compounds for which some properties cannot be measured.

In recent years, there has been increasing interest in using molecular descriptors integrated into Quantitative Structure Property Relationships (QSPR) for prediction of constant properties or temperature-dependent properties at a given temperature value (Dearden,³). However, very few attempts have been made to develop QSPRs to predict temperature-dependent properties for a wide range of temperature values (Godavarthy,⁴). In this work, a framework for the prediction of temperature-dependent properties is established based on several recently developed techniques (Shacham *et al.*,⁵ Shore,⁶ Shore *et al.*,⁷ Benson-Karhi *et al.*,⁸ Cholakov *et al.*,⁹).

To predict a particular property of a target compound (we will denote it the target property), compounds structurally "similar" to the target are first identified (denoted as predictive compounds). This similarity is represented by structure-structure correlation between molecular descriptors of the target and the predictive compounds. It is assumed that experimental data for the target property of the predictive compounds are available. The QS2PR method of Shacham *et al.*⁵ is

1 used to identify such similar compounds and to derive a linear structure-structure correlation.
2
3
4 Typically, 2-4 predictive compounds are included in the QS2PR model. If the target compound and
5
6 the potential predictive compounds belong to the same homologous series, use of the short-cut
7
8 QS2PR (SC-QS2PR) method of Cholakov *et al.*⁹ can be considered.
9
10

11
12 Two different cases are addressed. In the first case, temperature-dependent models are
13
14 available for the predictive compounds (e.g., Antoine equation for vapor pressure calculations).
15
16 Property values for the target compound can then be calculated, point by point, at any specified
17
18 temperature. Subsequently, a model equation for the target compound can be fitted by linear or non-
19
20 linear regression, using the predicted property values.
21
22

23
24 In the second case, it is assumed that experimental data are available for the predictive
25
26 compounds, however a model to represent these data is not available and empirical (regression)
27
28 models for the predictive compounds must first be derived. One possibility to derive such models is
29
30 via the new Response Modeling Methodology (RMM⁶). Once models for the predictive compounds
31
32 are derived, the prediction procedure for the target compound can proceed as in the previous case.
33
34

35
36 Parts of the calculations associated with the use of QS2PR were carried out with a modified
37
38 version of the stepwise regression program (SROV, Shacham and Brauner,¹⁰). Property data were
39
40 taken from the DIPPR database (Rowley *et al.*,¹¹). The Dragon[®] program (version 5.4), was used to
41
42 calculate molecular descriptors.
43
44

45
46 The structure of the paper is as follows: first, the QS2PR and SC-QS2PR techniques, RMM
47
48 methodologies and the proposed property-prediction algorithm are briefly expounded (Sections 2-4,
49
50 respectively). This is followed by demonstrating, in Sections 5-8, application of the proposed
51
52 approach to (saturated) liquid density of 1-butanol, toluene, n-hexane and heneicosane. Section 9
53
54 compares the new approach to current techniques, and we summarize with some conclusions in
55
56 Section 10.
57
58
59
60

2. The QS2PR and SC-QS2PR Methods

The QS2PR technique and its short-cut version, SC-QS2PR, have been formerly described in detail (Shacham *et al.*,⁵ Cholakov *et al.*,⁹) . Hereunder, we provide a brief outline of the main features.

Let us assume that the vector of properties of the target compound \mathbf{y}_t (the dependent variable) is potentially related to a set of m vectors of properties of m predictive compounds (independent variables), $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$. The following partition into sub-vectors of the \mathbf{y}_t and \mathbf{x} vectors is used:

$$\mathbf{y}_t = \begin{Bmatrix} \mathbf{y}_{ct} \\ \mathbf{y}_{pt} \end{Bmatrix} ; \quad \mathbf{x}_j = \begin{Bmatrix} \mathbf{x}_{ci} \\ \mathbf{x}_{pi} \end{Bmatrix} \quad (1)$$

where \mathbf{y}_{ct} is an N -vector of known properties and \mathbf{y}_{pt} is a K -vector of unknown properties. Both the N vector \mathbf{x}_{ci} and the K vector \mathbf{x}_{pi} contain known properties. Typically, the sub-vectors \mathbf{y}_{ct} and \mathbf{x}_{ci} contain properties directly related to the molecular structure, and can be calculated with high accuracy (molecular descriptors). By contrast, the sub-vectors \mathbf{y}_{pt} and \mathbf{x}_{pi} contain measured properties with various levels of experimental error. We wish to model by a linear regression model the structure-structure relationship between \mathbf{y}_{ct} and m independent variables $\mathbf{x}_{c1}, \mathbf{x}_{c2}, \dots, \mathbf{x}_{cm}$:

$$\mathbf{y}_{ct} = \beta_1 \mathbf{x}_{c1} + \beta_2 \mathbf{x}_{c2} \dots \beta_m \mathbf{x}_{cm} \quad (2)$$

where the weighing factors $\beta_1, \beta_2 \dots \beta_m$ are model parameters to be estimated.

Applying Eq.(2) with property values on its r.h.s (instead of descriptor values) results in a property-property relation. Accordingly, the following equation can then be used to predict the j property of the target compound, y_{ptj} on the basis of data available for that property for the m predictive compounds:

$$y_{ptj} = \beta_1 x_{p1j} + \beta_2 x_{p2j} \dots + \beta_m x_{pmj} \quad (3)$$

The practical application of the QS2PR and SC-QS2PR require preparation of a bank of potential predictive compounds as a database. The same set of molecular descriptors must be defined for all

1
2 compounds included in the database, while the span of the molecular descriptors should reflect the
3
4 difference between the compounds in the database. The corresponding molecular descriptors for the
5
6 target compound, y_{ct} , must be defined as well. Prediction of properties using the QS2PR and SC-
7
8 QS2PR methods may involve the following steps:

- 9
- 10
- 11 1) Identification of the predictive compounds to be included in the structure-structure correlation;
- 12
- 13 2) Identification of a molecular descriptor collinear with the target property;
- 14
- 15 3) Derivation of the structure-structure relationship;
- 16
- 17
- 18 4) Prediction of the desired property by property-property correlation.
- 19
- 20
- 21
- 22

23
24 **2.1 Selection of Predictive Compounds to be Included in the Structure-Structure**
25
26 **Correlation.** The similarity between potential predictive compounds and the target compound is
27
28 measured by the partial correlation coefficient, r_{ii} , between the vector of the molecular descriptors
29
30 of the target compound, y_{ct} , and that of a potential predictive compound x_{ci} . The partial correlation
31
32 coefficient is defined as $r_{ii} = \bar{y}_{ct} \bar{x}_{ci}^T$, where \bar{y}_{ct} and \bar{x}_{ci}^T are row vectors, centered (by subtracting the
33
34 mean) and normalized to a unit length (after dividing by the Euclidean norm of the vector).
35
36 Absolute r_{ii} values close to one ($|r_{ii}| \approx 1$) indicate high correlation between vectors y_{ct} and \bar{x}_{ci}^T , and
37
38 thus point to strong similarity between the molecular structures of the target compound and the
39
40 predictive compound i . The addition of predictive compounds to the structure-structure correlation
41
42 is carried out in a stepwise fashion (using the SROV, stepwise regression program of Shacham and
43
44 Brauner¹⁰). At each step, a new independent variable, say \bar{x}_{ck} is selected to enter the model on the
45
46 basis of its $|r_{ik}|$ value. At the same time the vector \bar{y}_c and the \bar{x}_{ci} vectors, not yet included in the
47
48 model, are updated by removing from them the components which are collinear with \bar{x}_{ck} . Addition
49
50 of predictive compounds to the correlations stops when the signal to noise ratio in $|r_{ii}|$ is reduced
51
52 below the value of 1.
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
Cholakov *et al.*⁹ have shown that this step can be considerably simplified in cases where the database contains predictive compounds belonging to the same homologous series as the target compound. In such cases two or three predictive compounds for which property data are available and which are closest (in terms of the number of carbon atoms) to the target compound are selected to be included in the structure-structure correlation. The number of predictive compounds to be included will be further discussed in Section 5.

17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
2.2 Identification of a Molecular Descriptor Collinear with the Target Property. This step is carried out only when SC-QS2PR is used, following the TQSPR method outlined in Shacham *et al.*,¹². First a "training set" of compounds similar to the target compound are identified. To this aim the $|r_{ii}|$ values are first calculated, as explained in section 2.1. The training set is established by selecting the first n compounds with the highest $|r_{ii}|$ values for which experimental values of the desired property are available at a particular temperature (say, at room temperature). Shacham *et al.*¹² recommended using $n = 10$ for the training set.

34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
To select a descriptor which is collinear with the desired property, the property values of the n compounds (selected to the training set) are introduced into the vector \mathbf{z} and the corresponding values of their molecular descriptors are stored in the vectors ζ_j . The descriptor collinear with the property is selected according to the value of the partial correlation coefficient, $|\rho_{zj}|$. The latter is defined as $\rho_{zj} = \bar{\mathbf{z}}\bar{\zeta}_j^T$, where $\bar{\mathbf{z}}$ and $\bar{\zeta}_j$ are row vectors, centered (by subtracting the mean) and normalized to a unit length. The descriptor with the highest $|\rho_{zj}|$ value is selected as having the highest level of collinearity with the desired property.

55
56
57
58
59
60
2.3 Derivation of the Structure-Structure Correlation. In the case of the QS2PR method, the selection of predictive compounds via the stepwise regression algorithm (as described in Section 2.1) yields the structure-structure relationship of Eq. (2) and the corresponding $\beta_1, \beta_2 \dots \beta_m$ parameter values. However, if the predictive and the target compounds belong to the same

homologous series, the prediction procedure can be simplified significantly by the use of the SC-QS2PR method. In such a case, the minimum information required is $m-1$ non-collinear molecular descriptors for both the predictive and the target compounds (Cholakov *et al.*,⁹). For example, for $m=3$ the coefficients of the structure-structure relation, $y_{ct} = \sum_{i=1}^3 \beta_j x_{cpi}$ are obtained by solving the system of three linear equations:

$$\begin{aligned}\beta_1 + \beta_2 + \beta_3 &= 1 \\ \beta_1 n_{c1} + \beta_2 n_{c2} + \beta_3 n_{c3} &= n_{ct} \\ \beta_1 x_{c1} + \beta_2 x_{c2} + \beta_3 x_{c3} &= y_{ct}\end{aligned}\tag{4}$$

where n_{c1} , n_{c2} and n_{c3} are numbers of carbon atoms of the three predictive compounds, and n_t is the number of carbon atoms of the target compound. The first equation in (4) reflects the influence of the descriptors that have the same value for the target and predictive compounds (e.g. the number of CH₃ groups in the n -alkane series). The molecular descriptor included in the third equation is the one selected in Section 2.2. Alternatively, another property well-correlated with the target property (e.g., normal boiling temperature) can replace the molecular descriptor in the third equation.

If the target property (or the associated descriptor, x_c) is collinear with the vector of the number of carbon atoms, n_c , it is preferable to use two predictive compounds ($m = 2$) rather than three in the structure-structure correlation. The level of the collinearity can be measured, for example, using the "condition number" of the coefficient matrix of Eqs. (4).

2.4 Prediction of the Desired Property by Property-Property Correlation. Applying Eq. (2) with parameter values obtained either via the QS2PR method, or via the SC-QS2PR method with property values on its r.h.s (instead of descriptor values) results in a property-property relationship (Eq. 3).

The determination of the upper error bounds on the predicted property values is discussed and demonstrated in detail in references 5 and 9. The condition number of the coefficient matrix of Eq. (4) provides an estimate of the upper bound on the relative errors in the calculated parameter

values. The maximal error in the predicted property value can be estimated using the reported reliability of the predictive compound data and the error propagation formula.

3. The RMM Method for Modeling Physical Properties

When models for the target property are not available, empirical (regression) models for the predictive compounds must first be derived. One possibility to obtain such models is via the new Response Modeling Methodology (RMM; refer to Shore,⁶ for its introduction, and to Shore *et al.*,⁷ and Benson-Karhi *et al.*,⁸ for demonstrative application to modeling of physical properties). The main advantage of RMM-based models is that they typically deliver representation of the property's temperature-dependent variation, with adequate precision and high level of stability, using only three parameters. Thus, the same model can be used over a wide spectrum of properties, and for data available at various temperature ranges.

With T (temperature) as a single regressor, several types of models can be derived using RMM. The most general RMM model is (Shore⁶):

$$\log(Y) = \log(M) + \left(\frac{\alpha}{\lambda}\right) [(\eta + \varepsilon_1)^\lambda - 1] + \varepsilon_2 \quad (5)$$

where η is an appropriately relocated and rescaled T , namely: $\eta = \beta_0 + \beta_1 T$ (β_0 and β_1 are real numbers), ε_1 and ε_2 are regressor and response errors, respectively, and $\{M, \alpha, \lambda\}$ are the model's parameters. From this model, simplified 2-, 3- and 4-parameter models are derivable (Shore⁶), which have been identified as most appropriate for modeling temperature-dependency of physical properties (Benson-Karhi *et al.*,⁸):

$$\begin{aligned} 2 \text{ Parameters RMM}_2(a, b) : \log(Y) &= \log(M_Y) + \frac{a}{b}(e^{bZ} - 1) + aZ \\ 3 \text{ Parameters RMM}_3(a, b, c) : \log(Y) &= \log(M_Y) + \frac{a}{b}(e^{bZ} - 1) + cZ \\ 4 \text{ Parameters RMM}_4(a, b, c, d) : \log(Y) &= \log(M_Y) + \frac{a}{b/c}[(1 + cZ)^{\frac{b}{c}} - 1] + dZ \end{aligned} \quad (5a)$$

where Y is the modeled property, M_Y is the median of Y , T is assumed to be normally distributed, Z is a standardized T : $Z = (T - M_T)/\sigma_T$, M_T and σ_T are the median and the standard deviation of T , respectively, and $\{a, b, c, d\}$ are four parameters to be estimated.

This model assumes that Y is defined on $[0, \infty]$. If there is reason to assume that Y has a lower non-negative limit which is much larger than zero, an additional location parameter, L , may need to be introduced. For example, RMM_2 then becomes:

$$\log(Y - L) = \log(M_Y - L) + (a/b)[e^{bZ} - 1] + aZ.$$

Earlier investigations (Shore et al.⁷, Benson-Karhi et al.⁸) have shown that RMM_2 is more stable than RMM_3 and RMM_4 (see Eq. 5a). Furthermore, in many cases it can model temperature-dependent variation of a property with sufficient accuracy to allow reliable property prediction. Consequently, RMM_2 is used to model the predictive compound data.

4. A generalized algorithm for prediction of temperature-dependent properties

The complexity of this algorithm depends on the amount of information available for the target and predictive compounds. The basic form of the general algorithm includes the following steps:

1. Identifying predictive compounds which are structurally similar to the target compound. It is assumed that for these compounds, experimental data and/or models for the temperature-dependent target property are available (see section 2.1);
2. Identification of a molecular descriptor, which is co-linear with the target property at a particular temperature (Section 2.2);
3. Determination of the number of predictive compounds needed according to the level of co-linearity between the molecular descriptor, identified in (2), and the number of carbon atoms of the predictive compounds (Section 2.3);
4. Determination of the common temperature range (or common reduced temperature range) for the predictive and target compounds, according to the data available for the predictive

1
2 compound, and/or phase transition temperatures (triple point and critical temperature) of the
3
4 predictive and target compounds;

- 5
6
7 5. Representation of the property-temperature variation of the predictive compounds using the
8
9 RMM method (in cases where accurate models are not available);
10
11
12 6. Application of the QS2PR or SC-QS2PR methods for point-by-point calculation of the
13
14 property value for the target compound.
15
16

17
18 This procedure can be simplified if the target and predictive compounds belong to the same
19
20 homologous series and the desired property value is available for the target compound at one
21
22 particular temperature. In this case, the property value can be used instead of a molecular descriptor,
23
24 and step 2 of the procedure can be skipped. Step 5 is unnecessary if regression models more
25
26 accurate than the RMM are available for representing the property-temperature relationship of the
27
28 predictive compounds.
29
30

31
32 Sections 5-8 demonstrate application of the proposed algorithm.
33
34

35 **5. Prediction of liquid density of 1-butanol**

36
37 **5.1 Predictive and target compound data.** For prediction of saturated liquid density (ρ_L)
38
39 of 1-butanol (the target compound), three other compounds were selected as predictive compounds
40
41 from the 1-alcohol homologous series: 1-propanol, 1-pentanol and 1-hexanol. Pertinent data (taken
42
43 from the DIPPR database) are shown in Table 1. The properties included are: the number of carbon
44
45 atoms (n_C), melting point temperature (T_m), normal boiling temperature (T_b), critical temperature
46
47 (T_C), reduced temperature at T_m ($T_{mR} = T_m/T_C$), liquid density at 298.15 K (ρ_L^{25}) and the value of the
48
49 molecular descriptor $VRm2$. The latter belongs to the category of 2D eigenvalue-based indices:
50
51 "average Randic-type eigenvector-based index from mass weighted distance matrix". Additional
52
53 information regarding eigenvalue-based indices can be found in reference¹³. It was identified by the
54
55 TQSPR method as being collinear (highly correlated) with ρ_L^{25} for the 1-alcohol homologous
56
57 series. DIPPR provides also values for the "reliability" (upper limit on experimental error) of the
58
59
60

various properties. For the data shown in Table 1, the reliabilities for T_m , T_b and ρ_L^{25} values is < 1 %, while for $T_C < 0.2$ %.

The following correlation for the variation of ρ_L as function of temperature (T) is recommended by DIPPR:

$$\rho_L = \frac{A}{B^{[1+(1-T/T_C)^C]}} \quad (6)$$

where T is in (K), ρ_L is the liquid density in kmol/m^3 and A , B and C are parameter values for the particular compound (see Table 1). Note, that Eq.(6) is a special case of the general RMM model, given by Eq. (5), with the errors therein ignored. To realize that, re-write Eq.(6) as:

$$\log(\rho_L) = \log(A/B^2) + [-\log(B)][(1 - \frac{T}{T_C})^C - 1]$$

namely, for Eq. (6) the re-scaled and re-located T is: $\eta = 1 - (T/T_C)$.

5.2 Comparing two vs. three predictive compounds QS2PRs. Let us investigate whether linear extrapolation between 1-propanol and 1-pentanol can provide accurate enough prediction for ρ_L^{25} of 1-butanol. Using the ρ_L^{25} values of the predictive compounds 1-propanol and 1-pentanol (shown in Table 1), their n_C values and that of the target, we get:

$$\rho_L^{25} = \frac{4-3}{5-3}(9.19 - 13.318) + 13.318 = 11.254 \text{ (Kmol/m}^3\text{)} \quad (7)$$

Comparing this value with the measured value for 1-butanol (10.87 Kmol/m^3) shows prediction error of 3.53 %, which is appreciably larger than the reliability of the experimental value (< 1 %). The excessive error implies that ρ_L^{25} is non-collinear with the carbon number of the predictive compounds. Indeed, using ρ_L^{25} values in the 3rd row of Eqs. (4), results in the following coefficient matrix \mathbf{A} :

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 \\ 3 & 5 & 6 \\ 13.318 & 9.19 & 7.987 \end{bmatrix} \quad (8)$$

1
2 The condition number of this matrix $\rho(\mathbf{A})= 802.06$ (calculated by MATLAB[®]). Similar
3
4 value of $\rho(\mathbf{A}) = 948$ is obtained replacing ρ_L^{25} by the descriptor $VRm2$.
5
6

7 Let us now carry out similar calculations related to T_b . Introducing the pertinent T_b values
8 into Eq. (7) yields $T_b = (411.1 - 370.3)/2 + 370.3 = 390.7$ K. Comparing with the measured value
9
10 for 1-butanol (390.6 K) shows prediction error < 0.1 %, well below the reliability of the
11
12 experimental value ($< 1\%$). The high level of collinearity in this case is reflected in the condition
13
14 number obtained by introducing the T_b values of the three predictive compounds as the last row in
15
16 matrix \mathbf{A} . The value obtained, $\rho(\mathbf{A}) = 429,000$, is larger by about two and a half orders of
17
18 magnitude than the condition number associated with ρ_L^{25} .
19
20
21
22
23
24

25 **5.3 Determining the temperature range for the prediction.** The correlation for
26 temperature-dependence of ρ_L (Eq. 6) for the predictive compounds, with parameters shown in
27
28 Table 1, is valid between the melting point, T_m , and the critical temperature, T_C . Prediction for the
29
30 target compound will be valid in the temperature range common to all the predictive compounds,
31
32 namely, in the interval $[\max(T_{mi}), \min(T_{Ci})]$, $i = 1, 2, 3$, provided T_C and T_m of the target are not
33
34 included in this common range. If T_C and T_m values are not available for the target compound, their
35
36 prediction may be required. In the particular case of 1-butanol, the common range is [228.55 K,
37
38 536.8 K].
39
40
41
42

43 The plot of ρ_L for the predictive compounds in the common temperature range is shown in
44
45 Figure 1. Irregular behavior of liquid density of 1-propanol near the critical temperature can be
46
47 observed, where ρ_L values fall below those of 1-pentanol and 1-hexanol. This change in trend of the
48
49 variation may cause excessive prediction errors. Thus, the practical temperature range for
50
51 prediction is more limited, with upper temperature limit being ~ 480 K.
52
53
54
55

56 The temperature range can be extended up to the T_C of 1-butanol by using as independent
57
58 variable the reduced temperature $T_R = T/T_C$. The range is then given by $[\text{Max}(T_{mRi}), 1]$, $i = 1, 2, 3$,
59
60 where T_{mR} is the reduced melting point temperature. In this particular case, the T_R range is [0.374,
1]. For the predictive compounds, the ρ_L values are depicted versus T_R in Figure 2. In this

1
2 representation the trend of variation of ρ_L is similar for all the predictive compounds throughout the
3
4 entire region. Note that the decision on using T_R as independent variable instead of T is based
5
6 entirely on predictive compound data.
7

8
9 **5.4 Predicting ρ_L for 1-butanol.** Introducing n_C and the descriptor $VRm2$ into Eq.(4) yields:
10
11 $\beta_1 = 0.28205$, $\beta_2 = 1.1538$ and $\beta_3 = -0.4359$. Inserting these parameter values and the liquid density
12
13 at 298.15 K of the predictive compounds into the property-property correlation (Eq. 3) yields:
14
15 $0.28205*13.318 + 1.1538*9.19 - 0.4359*7.987 = 10.878 \text{ kmol/m}^3$, the predicted ρ_L value for the 1-
16
17 propanol at the same temperature. Comparing it with the experimental value shown in Table 1
18
19 (10.78 kmol/m^3) indicates that the difference is $<0.1\%$. Table 2 shows the predicted ρ_L values of 1-
20
21 butanol at various temperatures. These predicted values were calculated pursuing the procedure
22
23 described for the case of 298.15 K, except that the ρ_L values of the predictive compounds were
24
25 calculated using Eq. (6) with the coefficients provided in Table 1.
26
27
28
29
30

31 Values of ρ_L calculated from Eq.(6), with the coefficients provided by DIPPR for 1-butanol,
32
33 are also shown in Table 2. The differences between the predicted and calculated values are $< 1\%$,
34
35 except for the point closest to the critical temperature of 1-propanol, where the error is 1.29%. The
36
37 reliabilities assigned by DIPPR to the ρ_L values calculated by Eq. (6) are $< 1\%$ for 1-propanol and
38
39 1-butanol, and $< 3\%$ for 1-pentanol and 1-hexanol. Thus prediction accuracy is within the error
40
41 limits of the well established correlation.
42
43
44

45
46 The prediction results for ρ_L when T_R is used as the independent variable are presented in
47
48 Table 3. In this case very accurate predictions are obtained up to the critical point. The prediction
49
50 error is $<1\%$, except at the critical point itself (1.5 %). Even this last value is within the reliability
51
52 limits of the experimental values, as the reliability assigned by DIPPR to the critical volume is $<$
53
54 3% . Using the predicted values and non-linear regression to fit parameter values to Eq. (6) yields
55
56 $A=0.9949387$, $B=0.269691$, $C=0.2520693$, which are practically the same as the original DIPPR
57
58 parameters for 1-butanol.
59
60

Predicting ρ_L as function of T_R (rather than T) has the advantage of extending the prediction range up to the critical point. However, this requires the T_C value of the target compound in order to obtain the T_R values. There are several reliable methods for predicting T_C (see, for example, Yan et al.¹⁴). In the present case, linear interpolation between the two closest neighbors of 1-butanol yields satisfactory prediction (562.45 K instead of the measured value of 563 K).

If experimental data is available for the target compound at a single temperature, say ρ_L^{25} , the use of a molecular descriptor (i.e., *VRM2*) can be avoided. In such a case, the ρ_L^{25} values of the predictive and target compounds can be used in the 3rd row of Eq.(4) to obtain parameter values of the property-property relationship (Eq. 3). For example, introducing ρ_L^{25} values from Table 1 into Eq.(4) yields parameter values: $\beta_1 = 0.277$, $\beta_2 = 1.169$ and $\beta_3 = -0.446$. As these values are very similar to those calculated using the *VRm2* descriptor, the predicted ρ_L values are of similar accuracy. This demonstrates that if a reliable value for the (target) property of the target compound is available at a single temperature, it may save the effort of searching for an appropriate molecular descriptor.

6. Prediction of liquid density of toluene

6.1 Predictive and target compound data. For prediction of liquid density of toluene, three predictive compounds from the alkyl-benzene homologous series were selected: ethylbenzene, pentylbenzene and heptylbenzene. Unlike in the previous example, in this case prediction is done by extrapolation to a lower n_C target compound. The properties n_C , T_m , T_b , T_C , ρ_L^{25} and the values of the descriptor *ARR*, for the three predictive compounds are shown in Table 4. The descriptor *ARR* belongs to the category of "*constitutional descriptors*", and is described by Dragon as "aromatic ratio". The latter is the ratio of the number of aromatic bonds over the total number of non-H bonds. This descriptor was identified by the TQSPR method program as being collinear (highly correlated) with ρ_L^{25} for the alkyl-benzene homologous series. The DIPPR reliabilities for all the

T_m , T_b and ρ_L^{25} values is $< 1\%$, while for T_C the reliability $< 0.2\%$ for toluene and ethylbenzene, $< 1\%$ for pentylbenzene and $< 5\%$ for heptylbenzene.

In this example it is assumed that no correlation to represent ρ_L for the predictive compounds is available and an *RMM* model has to be fitted to experimental ρ_L versus temperature data. The DIPPR database recommends the use of ρ_L data published by the American Petroleum Institute (API¹⁵) and by the Thermodynamic Research Center (TRC¹⁶). These are shown in Table 5 for both the predictive and target compounds. The temperature range covered: 253.15 K – 453.15 K is inside the bounds of T_m and T_C for all the compounds involved.

6.2 Determination of the *RMM*₂ model parameters.

The calculation of *RMM*₂ parameters, a and b , requires nonlinear regression. This was carried out using the *fminsearch* library function of MATLAB, starting with the initial estimates: $a = -0.02$ and $b = 0.2$. The resultant parameter values and the variance of the fit are shown in Table 6. The variance in all cases is smaller by seven or eight orders of magnitude than the median of ρ_L indicating that the data are accurately represented by the *RMM*₂ model.

6.3 Predicting ρ_L for toluene. Calculating the condition number of the **A** matrix, where ρ_L^{25} values of the predictive compounds are included in the 3rd row of this matrix, yields: $\rho(\mathbf{A}) = 906.43$. According to the criterion demonstrated in Section 4.3, three predictive compounds should be used to predict ρ_L for toluene. Introducing n_C and the descriptor *ARR* into Eq.(4) yields: $\beta_1 = 1.8137$, $\beta_2 = -1.5342$ and $\beta_3 = 0.7205$. Using these parameter values and the predictive compounds' ρ_L values into Eq.(3) yields the predicted values shown in Table 7. Comparison of the predicted values with the API (1980) data shows prediction error less than 0.3% in all cases.

7. Prediction of liquid density of *n*-hexane

7.1 Predictive and target compound data. For prediction of liquid density of *n*-hexane, three predictive compounds from the *n*-alkane homologous series were selected: *n*-pentane, *n*-heptane and *n*-octane. It is assumed that only experimental data are available for these compounds (no accurate models are available). Accordingly, an *RMM* model is fitted. The DIPPR database

1
2 recommends the use of ρ_L data published in references^{15,16} (shown in Figure 3). Additional data for
3
4 the predictive and target compounds are given in Table 8. These include n_C , T_m , T_b , T_C , ρ_L^{25} , the
5
6 lowest and the highest temperature for which ρ_L data are available, and the values of descriptor
7
8 $VEA2$. The latter was identified by the TQSPR method as collinear with ρ_L^{25} for the n -alkane
9
10 homologous series in the range $n_C = 4$ through 20. It belongs to the category of 2D eigenvalue-
11
12 based indices (Balaban et al.¹³), and is described by Dragon as an average eigenvector coefficient
13
14 sum from adjacency matrix. The DIPPR reliabilities for all T_m and T_C is $<0.2\%$, and is $< 1\%$ for all
15
16 the ρ_L^{25} and T_C values. The common temperature range covered by the data, 223.15 K – 467.65 K,
17
18 is inside the bounds of T_m and T_C for all the compounds involved.
19
20
21
22
23
24

25 **7.2 Predicting ρ_L for n -hexane for the full temperature range.** Figure 3 indicates a
26
27 different slope of the n -pentane ρ_L data compared to the slopes of ρ_L curves of the other predictive
28
29 compounds, in particular for $T > 450$ K, as the T_C of n -pentane is approached. As shown in Section
30
31 5, such a situation has caused inaccuracy in predicting ρ_L for 1-butanol. Nevertheless, we first
32
33 attempt to predict ρ_L values for n -hexane using T as the independent variable. Parameter values of
34
35 the RMM_2 models and the error variances are shown in Table 9. The variances for two out of the
36
37 three predictive compounds are of the order of 10^{-4} , indicating that the fit is not as good as in the
38
39 case of the toluene.
40
41
42
43

44 The condition number of the \mathbf{A} matrix when ρ_L^{25} values of the predictive compounds are
45
46 introduced in the 3rd row of this matrix is $\rho(\mathbf{A}) = 1590.6$. This value is still low enough to justify the
47
48 use of three predictive compounds to predict ρ_L for n -hexane. Using n_C and the descriptor $VEA2$
49
50 values in Eq.(4) yields: $\beta_1 = 0.27273$, $\beta_2 = 1.1818$ and $\beta_3 = -0.45455$. Introducing these parameter
51
52 values and the $\rho_L(T)$ values of predictive compounds (as obtained by the respective RMM_2 models)
53
54 into Eq.(3) yields the predicted values shown in Table 10. The resulting prediction error is $< 1\%$
55
56 (which is also the reliability of the DIPPR data) throughout the entire temperature region, except for
57
58 the lower temperatures (< 250 K). The relatively large values of the variances in Table 9 imply that
59
60

1
2 this is a result of the discrepancy between the predictive compounds data and its representation by
3
4 the RMM_2 model over the entire temperature range.
5

6
7 It is a common practice to divide a wide temperature range to smaller intervals if a
8
9 correlation with a small number of parameters is used to model the property (like in the case where
10
11 the Antoine equation is used to model vapor pressure data). This practice is followed here to
12
13 improve the prediction, whereby the temperature range is divided into two intervals: low-
14
15 temperature range (223.15 K – 353.15 K) and a high-temperature range (353.15 K – 467. 65 K).
16
17 Note that the decision to divide the temperature range into smaller intervals is based entirely on
18
19 predictive compound data.
20
21

22
23 Modeling separately the low-temperature range yields ρ_L predictions for n-hexane all accurate
24
25 within the experimental error level.
26
27

28 29 30 31 **8. Prediction of liquid density of *n*-heneicosane**

32
33 **8.1 Predictive and target compound data.** This example demonstrates a case where the
34
35 prediction involves extrapolation in the high carbon number range, where the target compound is *n*-
36
37 heneicosane ($n_C = 21$) and the predictive compounds are *n*-pentadecane ($n_C = 15$), *n*-heptadecane (n_C
38
39 = 17) and *n*-nonadecane ($n_C = 19$). Pertinent data for the target and predictive compounds (taken
40
41 from the DIPPR database) are shown in Table 11. The properties included are: n_C , T_m , T_b , T_C , ρ_L^{25}
42
43 and the values of the molecular descriptor $HNar$. This descriptor belongs to the 2D topological
44
45 indices category, "Narumi harmonic topological index¹⁷". It has been identified by the TQSPR
46
47 method as collinear with ρ_L^{25} for the *n*-alkane homologous series in the range $n_C = 14$ through 25
48
49 range. For the data shown in Table 11, the reliabilities for all the T_b and ρ_L^{25} values are < 1 %, for
50
51 T_C the reliability is <0.2% (except for *n*-heneicosane for which the reliability < 3%), and for T_m the
52
53 reliability is < 0.2% (except for *n*-heneicosane for which the reliability is < 3%).
54
55
56
57
58
59

60
In this example Eq. (6) is used to calculate ρ_L for the predictive (and the target) compounds,
with the parameter values A , B and C taken from the DIPPR database (see Table 13). The reliability

1
2 assigned by DIPPR to the ρ_L values calculated by Eq. (6) is <3 % for all the predictive compounds
3
4 and <5 % for *n*-heneicosane.
5
6

7 **8.2 Predicting ρ_L for *n*-heneicosane.** When the *HNar* descriptor values of the predictive
8
9 compounds are introduced in the 3rd row of the A matrix, the condition number of this matrix is
10
11 $\rho(\mathbf{A}) = 27112$. The large value indicates high collinearity between n_C and the descriptor *HNar*. The
12
13 parameter values obtained when n_C and the descriptor *HNar* are used in Eq.(4) are: $\beta_1 = 1.66667$, β_2
14
15 = -4.33333 and $\beta_3 = 3.66667$. Using these parameter values and the $\rho_L(T)$ values of predictive
16
17 compounds (as obtained by Eq.(6)) in Eq.(3) yields the predicted values shown in Figure 6.
18
19 Comparison of the predicted ρ_L values for the target (*n-heneicosane*) with those obtained by Eq. (6)
20
21 (the A, B C coefficients are included in Table 13) shows that the prediction error is below
22
23 experimental error level (< 3%) for the temperature region up to 650 K. However, for *n*-
24
25 pentadecane approaching T_C (708 K), the prediction error increases considerably.
26
27
28
29
30

31 Considering the high collinearity between the descriptor *HNar* and n_C (as implied by the
32
33 high condition number), it is expected that accurate predictions may be obtained using only two
34
35 predictive compounds. Removing *n*-pentadecane from the set of predictive compounds and carrying
36
37 out linear extrapolation using n_C yields: $\beta_1 = 0$, $\beta_2 = -1$ and $\beta_3 = 2$. The prediction of ρ_L of *n*-
38
39 heneicosane using these parameter values yields the results shown in Figure 4. In this case, the
40
41 prediction error is below experimental error level (<2.5 %) throughout the entire temperature range.
42
43
44

45 9. Comparison with Existing Methods

46
47 Extensive review and comparison of the methods used for prediction of ρ_L over 79 families
48
49 of compounds was recently carried out by Mulero *et al.*,^{18,19}. Their comparison (like most other
50
51 published comparisons) is based on the average prediction error for the entire temperature range for
52
53 one compound (the notation AAD, absolute average percentage deviation, is used by Mulero *et al.*
54
55 ^{18,19}). AAD mean value for an entire family of compounds (MAPD) is also shown in these
56
57 publications. While those measures are convenient for presenting comparison results, they do not
58
59 take into consideration the experimental error that may exist in the data. Obviously, measures of
60

1 prediction accuracy must consider the experimental error (reliability) of the measured property
2 values, as predicted values cannot be more accurate than the data used to generate them. In fact, all
3 predicted values that are within the experimental error bounds should be considered as of equivalent
4 reliability. The comparison must be based on particular compounds for which the prediction error
5 exceeds the experimental error.
6
7

8 The case studies shown in the previous sections demonstrate the use of the proposed methods
9 assuming various deficiencies in the availability of data (for example the RMM_2 model is used to
10 model the data, even through the DIPPR recommended correlation exists). The objective was to
11 demonstrate the achievable of prediction accuracy under such conditions, rather than optimizing the
12 prediction using all the available data. In spite of that, comparison of the predicted values with
13 existing techniques yields favorable results.
14
15

16 For the n -alcohol series only the MAPD values are reported by Mulero *et al.*¹⁸. Among the
17 nine methods tested, the lowest MAPD value is 2.4 %. Comparing this value with the prediction
18 errors shown in Table 3 for 1-butanol indicates that using the proposed method yields appreciably
19 lower errors, and all predicted values are within the experimental error bounds (reliability < 1% for
20 most of the region, and <3% at the critical point).
21
22

23 For toluene the results of the comparison are very similar. The lowest MAPD value reported
24 by Mulero *et al.*¹⁷ for the n -alkylbenzene series is 2.2 %. The prediction error obtained by the
25 proposed method (see Table 7) is well below experimental error level (reliability <1%).
26
27

28 For the n -hexane, AAD values are reported by Mulero *et al.*¹⁸ for four methods that yield
29 the lowest AAD values. These are within the range of 1.5% to 3.0 %, which is higher than the
30 reliability of the data (<1%) and the prediction errors obtained by the proposed method (see Table
31 12).
32
33

34 In the case of n -heneicosane, the AAD values, reported by Mulero *et al.*¹⁸ for the four
35 leading prediction methods, are below the data reliability (<5%). Similar results are obtained by the
36 proposed prediction method using two predictive compounds (Fig. 6).
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

10. Conclusions

A new method for predicting temperature-dependent properties of chemical compounds has been presented. The method is based on the identification of predictive compounds, which are structurally similar to the target compound and for which experimental data and/or models for the property-temperature dependence are available. The principal objective of this paper was to test the proposed method in a variety of conditions.

In the current paper, the suggested approach has been applied for prediction of liquid-density variation with temperature. The compounds studied are 1-alcohols, alkylbenzenes, low carbon number n-alkanes and high carbon number n-alkanes. Four case studies were presented to cover a variety of conditions: in two cases the prediction requires extrapolation (in terms of the number of the carbon atoms) toward the target compound, and two cases involve interpolation; in three of the cases 3-compound QS2PRs are required to obtain adequate prediction, while in one of the cases a linear extrapolation is accurate enough. The examples given include also identification of cases (based entirely on predictive compound data) where division of temperature range, or the use of reduced temperature, are required in order to obtain reliable property prediction over the entire temperature range. Thus, the method has been tested thoroughly for the cases where the QS2PR and the SC-QS2PR methods are applicable. Depending on the data available for the predictive compounds, the test cases demonstrate how the method can be adapted for optimizing the prediction for the target compound.

An advantage of proposed method over other prediction models for temperature-dependent properties is that it does not require measured property values of the target compound. If such values are available, the flexibility of the method enables their beneficial use for simplifying the calculations and enhancing confidence in the predicted values. The proposed algorithm can be successfully applied to various temperature-dependent properties. Its performance has been recently tested, for example, on prediction of vapor pressure and liquid viscosity (Shore *et al.*²⁰).

1
2 For target compounds for which these methods are not applicable, because of lack of
3
4 predictive compounds with available property data belonging to the same homologous series, the
5
6 Targeted QSPR method¹² may be applicable. Work is currently underway to extend the Targeted
7
8 QSPR method for prediction of temperature-dependent properties.
9
10

11 12 13 14 15 **Acknowledgments.**

16
17
18 © DRAGON is copyrighted by TALETE srl (<http://www.talete.mi.it>)

19
20
21 ® MATLAB is a registered trademark of The MathWorks Inc (<http://www.mathworks.com>)
22
23

24 25 26 27 **Literature Cited**

- 28
29 (1) Poling, B.E.; Prausnitz, J. M.; O'Connell, J. P. *Properties of Gases and Liquids*; 5th Ed.,
30
31 McGraw-Hill: New York, 2001.
32
33 (2) Marano, J.J.; Holder, G.D. General Equations for Correlating the Thermo-physical
34
35 Properties of n-Paraffins, n-Olefins and other Homologous Series. 2. Asymptotic Behavior
36
37 Correlations for PVT Properties, *Ind. Eng. Chem. Res.* **1997**, 36, 1887.
38
39 (3) Dearden, J. C. Quantitative Structure–Property Relationships for Prediction of Boiling
40
41 Point, Vapor Pressure, and Melting Point. *Environmental Toxicology and Chemistry.* **2003**, 22,
42
43 1696.
44
45 (4) Godavarthy, S.S.; Robinson, R.L.; Gasem, K.A.M. SVRC-QSPR model for predicting
46
47 saturated vapor pressures of pure fluids, *Fluid Phase Equilib.* **2006**, 246 (1-2): 39.
48
49 (5) Shacham, M.; Brauner, N.; Cholakov, G.St.; Stateva R.P. Property Prediction by
50
51 Correlations Based on Similarity of Molecular Structures. *AIChE J.* **2004**, 50, 2481.
52
53 (6) Shore, H. *Response Modeling Methodology (RMM): Empirical Modeling for Engineering*
54
55
56
57
58
59
60 *and Science*; World Scientific Publishing Co. Ltd; Singapore, 2005.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- (7) Shore, H.; Brauner, N.; Shacham, M. Modeling physical and thermodynamic properties via inverse normalizing transformations; *Ind. Eng. Chem. Res.* **2002**, 41, 651.
- (8) Benson-Karhi, D.; Shore, H.; Shacham, M. Modeling Temperature-Dependent Properties of Water via Response Modeling Methodology (RMM) and Comparison with Acceptable Models. *Ind. Eng. Chem. Res.* **2007** 46, 3446.
- (9) Cholakov, G.St.; Stateva, R.P.; Shacham M.; Brauner, N. Identifying Equations that Represent Properties in Homologous Series using Structure-Structure Relations, *AIChE Journal* , **2007**. 53, 150.
- (10) Shacham, M.; Brauner, N. The SROV Program for Data Analysis and Regression Model Identification. *Computers Chem. Engng.* **2003**, 27, 701.
- (11) Rowley, R.L.; Wilding, W.V.; Oscarson, J.L.; Yang, Y.; Zundel, N.A. DIPPR Data Compilation of Pure Chemical Properties Design Institute for Physical Properties. <http://dippr.byu.edu>, Brigham Young University Provo Utah, 2006.
- (12) Shacham, M.; Kahrs, O.; St Cholakov, G.; Stateva, R.; Marquardt, W.; Brauner, N. The Role of the Dominant Descriptor in Targeted Quantitative Structure Property Relationships", *Chem. Eng. Sci.* **2007**, 62, (22), 6222-6233.
- (13) Balaban, A. T.; Ciubotariu, D.; Medeleanu, M. Topological indices and real number vertex invariants based on graph eigenvalues or eigenvectors. *J. Chem. Inf. Comput. Sci.* **1991**, 31, 517-523.
- (14) Yan, X.; Dong, Q.; Hong, X. Reliability analysis of group-contribution methods in predicting critical temperatures of organic compounds, *J. Chem. Eng. Data*, **2003**, 48, 374-380.
- (15) American Petroleum Institute Research Project 44; *Selected Values of Properties of Hydrocarbons and Related Compounds* ; Thermodynamic Research Center, Texas A&M University, College Station, Texas, 1980
- (16) Thermodynamics Research Center; *TRC - Thermodynamic Tables -Hydrocarbons*; The Texas A&M University System, College Station, TX, 1994.

- 1
2 (17) Narumi, H. New topological indices for finite and infinite systems. MATCH (Comm. Math.
3
4 Comp. Chem.) **1987**, 22, 195-207.
5
6 (18) Mulero, A.; Cachadin, I.; Parra, M. I. Liquid Saturation Density from Predictive
7
8 Correlations Based on the Corresponding States Principle. Part 1: Results for 30 Families of Fluids,
9
10 *Ind. Eng. Chem. Res.* **2006**, 45, 1840
11
12 (19) Mulero, A.; Cachadin, I.; Parra, M. I. Liquid Saturation Density from Predictive
13
14 Correlations Based on the Corresponding States Principle. Part 2: Results for 49 Families of Fluids,
15
16 *Ind. Eng. Chem. Res.* **2006**, 45, 6864-6873
17
18 (20) Shore, H.; Benson-Karhi, D.; Brauner, N.; Shacham, M. Prediction of Temperature-
19
20 Dependent Properties by Correlations Based on Similarity of Molecular Structures, AIChE's 2007
21
22 Annual Meeting in Salt Lake City Utah, Nov. 4 -9, 2007.
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 1. Prediction of Liquid Density of 1-Butanol – Data for the Predictive and the Target Compounds (from the DIPPR Database, Rowley et al.¹¹)

	Predictive Compounds			Target
	1-propanol	1-pentanol	1-hexanol	1-butanol
No. of C Atoms	3	5	6	4
Melting Point Temp. (T_m , K)	146.95	195.56	228.55	184.51
Normal Boiling Temp. (K)	370.3	411.1	430.55	390.6
Critical Temp. (K)	536.8	588.1	610.3	563
Reduced Temperature at T_m	0.274	0.333	0.374	0.328
Liquid Density at 298.15 K (kmol/m ³)	13.318	9.190	7.987	10.870
Descriptor $VRm2$	0.681	0.778	0.807	0.738
Parameter A	1.246	0.81754	0.68425	0.97473
Parameter B	0.27254	0.26732	0.26477	0.26799
Parameter C	0.24108	0.25348	0.24395	0.24486

Table 2. Prediction of Liquid Density of 1-Butanol, T is the Independent Variable

Temp.(K)	Liquid Density (Kmol/m ³)		
	DIPPR Equation	Predicted value	Error (%)
228.55	11.59	11.65	-0.45
243.96	11.44	11.49	-0.42
259.38	11.28	11.33	-0.40
274.79	11.12	11.16	-0.37
290.2	10.96	10.99	-0.34
305.61	10.79	10.82	-0.32
321.02	10.61	10.64	-0.29
336.44	10.43	10.46	-0.26
351.85	10.25	10.27	-0.23
367.26	10.05	10.07	-0.19
382.67	9.85	9.87	-0.16
398.09	9.64	9.65	-0.12
413.5	9.42	9.43	-0.08
428.91	9.19	9.19	-0.03
444.32	8.94	8.94	0.02
459.74	8.68	8.67	0.09
475.15	8.39	8.37	0.19
490.56	8.07	8.04	0.33
505.97	7.71	7.67	0.60
521.39	7.29	7.20	1.29

Table 3. Prediction of Liquid Density of 1-Butanol, T_R is the Independent Variable

Reduced Temperature	Liquid Density (Kmol/m ³)			
	Temp.(K)	DIPPR Equation	Predicted value	Error (%)
0.4	225.2	11.63	11.67	-0.42
0.44	247.72	11.40	11.45	-0.39
0.48	270.24	11.17	11.21	-0.34
0.52	292.76	10.93	10.96	-0.30
0.56	315.28	10.68	10.71	-0.26
0.6	337.8	10.42	10.44	-0.22
0.64	360.32	10.14	10.16	-0.17
0.68	382.84	9.85	9.86	-0.12
0.72	405.36	9.54	9.55	-0.08
0.76	427.88	9.20	9.21	-0.03
0.8	450.4	8.84	8.84	0.03
0.84	472.92	8.43	8.42	0.08
0.88	495.44	7.96	7.95	0.13
0.92	517.96	7.39	7.38	0.18
0.96	540.48	6.62	6.60	0.21
1	563	3.64	3.69	-1.50

Table 4. Prediction of Liquid Density of Toluene – Data for the Predictive and the Target Compounds (from the DIPPR Database, Rowley et al.¹¹)

	Predictive Compounds			Target
	ethylbenzene	pentylbenzene	heptylbenzene	toluene
No. of C Atoms	8	11	13	7
Melting Point (K)	178.2	198.15	225.15	178.18
Normal Boiling Temp. (K)	409.35	478.61	519.25	383.78
Critical Temp. (K)	617.15	679.9	714	591.75
Liquid Density at 298.15 K (kmol/m ³)	8.134	5.765	4.844	9.376
Descriptor ARR	0.75	0.545	0.462	0.857

Table 5. Prediction of Liquid Density of Toluene – ρ_L Data for the Predictive and the Target Compounds (from the API Project¹²)

Temp.(K)	Liquid Density (Kmol/m ³)			
	ethylbenzene	pentylbenzene	heptylbenzene	toluene
253.15	8.497	6.0015	5.0289	9.8068
263.15	8.4141	5.9488	4.9863	9.7081
273.15	8.3312	5.8962	4.9444	9.6082
283.15	8.2483	5.8436	4.9018	9.5084
293.15	8.1654	5.791	4.8593	9.4079
298.15	-	5.7647	-	9.3575
303.15	8.0835	5.7384	4.8167	9.3068
313.15	8.0006	5.6858	4.7748	9.2056
323.15	7.9168	5.6332	4.7322	9.1025
333.15	7.832	5.5805	4.6908	8.9993
343.15	7.7463	5.5279	4.6477	8.8941
353.15	7.6606	5.4753	4.6046	8.7888
363.15	7.573	5.4227	4.5609	8.6814
373.15	7.4844	5.3701	4.5178	8.5728
383.15	7.395	5.3154	4.4753	8.4621
393.15	7.3036	5.2682	4.4299	8.3503
403.15	7.2113	5.2143	4.3902	8.2353
413.15	7.1171	5.1603	4.3448	8.1181
423.15	7.022	5.1063	4.2994	7.9976
433.15	6.924	5.0524	4.2597	7.8739
443.15	6.8242	5.0051	4.2144	7.7469
453.15	6.7224	4.9512	4.1747	7.6188

Table 6. Prediction of Liquid Density of Toluene – RMM_2 Model Parameters

RMM_2 parameter	Predictive Compounds			Target
	ethylbenzene	pentylbenzene	heptylbenzene	toluene
M_T	353.15	348.15	353.15	348.15
σ_T	62.0484	61.6779	62.0484	61.6779
M_Y	7.6606	5.5016	4.6046	8.8415
a	-0.03548	-0.029462	-0.028855	-0.037325
b	0.28978	0.11405	0.14139	0.31864
variance	1.1662E-07	6.1855E-08	9.5982E-08	2.2317E-07

Table 7. Prediction of Liquid Density of Toluene

Temp.(K)	Liquid Density (Kmol/m ³)		
	API 44 Data	Predicted value	Error (%)
253.15	9.80	9.81	-0.15
263.15	9.70	9.72	-0.18
273.15	9.61	9.63	-0.21
283.15	9.51	9.53	-0.23
293.15	9.41	9.43	-0.25
303.15	9.31	9.33	-0.26
313.15	9.21	9.23	-0.27
323.15	9.10	9.13	-0.28
333.15	9.00	9.03	-0.28
343.15	8.89	8.92	-0.28
353.15	8.79	8.81	-0.27
363.15	8.68	8.70	-0.25
373.15	8.57	8.59	-0.23
383.15	8.46	8.47	-0.20
393.15	8.34	8.36	-0.17
403.15	8.23	8.24	-0.13
413.15	8.11	8.12	-0.08
423.15	7.99	8.00	-0.02
433.15	7.87	7.87	0.05
443.15	7.75	7.74	0.12
453.15	7.63	7.61	0.21

Table 8. Prediction of Liquid Density of *n*-hexane – Data for the Predictive and the Target Compounds (from the DIPPR Database, Rowley et al.¹¹)

	Predictive Compounds			Target
	n-pentane	n-heptane	n-octane	n-hexane
No. of C Atoms	5	7	8	6
Melting Point (K)	143.42	182.57	213.36	177.83
Normal Boiling Temp. (K)	309.22	371.58	398.83	341.88
Critical Temp. (K)	469.7	540.2	568.7	507.6
Lowest Temp. in Data (K)	143.46	182.58	223.15	193.15
Highest Temp. in Data (K)	467.65	535.15	553.15	502.35
Liquid Density at 298.15 K (kmol/m ³)	8.617	6.802	6.152	7.613
Descriptor VEA2	0.431	0.359	0.334	0.39

Table 9. Prediction of Liquid Density of *n*-Hexane – *RMM*₂ Model Parameters. Full Temperature Range.

<i>RMM</i> ₂ parameters	Predictive Compounds			Target
	n-pentane	n-heptane	n-octane	n-hexane
<i>M</i> _T	348.15	343.15	383.15	358.15
<i>σ</i> _T	77.1829	75.6148	99.2795	84.4572
<i>M</i> _Y	7.8587	6.385	5.4817	6.9311
a	-0.079169	-0.054882	-0.073931	-0.072733
b	1.3866	0.54119	1.0069	1.18E+00
variance	0.000449	5.514E-06	0.000134	0.000321

Table 10. Prediction of Liquid Density of *n*-Hexane in the Full Temperature Range

Temp.(K)	Liquid Density (Kmol/m ³)		
	API 44 Data	Predicted value	Error (%)
223.15	8.20	8.31	-1.34
235.38	8.10	8.20	-1.19
247.60	8.00	8.08	-1.02
259.82	7.90	7.97	-0.85
272.05	7.79	7.85	-0.68
284.27	7.69	7.72	-0.50
296.50	7.57	7.60	-0.32
308.73	7.46	7.47	-0.14
320.95	7.34	7.34	0.03
333.17	7.21	7.20	0.20
345.40	7.08	7.05	0.36
357.63	6.94	6.90	0.50
369.85	6.79	6.74	0.62
382.07	6.63	6.58	0.72
394.30	6.45	6.40	0.79
406.52	6.26	6.21	0.81
418.75	6.06	6.01	0.77
430.97	5.84	5.80	0.67
443.20	5.60	5.57	0.47
455.42	5.34	5.33	0.12
467.65	5.05	5.07	-0.42

Table 11. Prediction of Liquid Density of *n*-Heneicosane – Data for the Predictive and the Target Compounds (from the DIPPR Database, Rowley et al.¹¹)

	Predictive Compounds			Target
	<i>n</i> -pentadecane	<i>n</i> -heptadecane	<i>n</i> -nonadecane	heneicosane
No. of C Atoms	15	17	19	21
Melting Point (K)	283.072	295.134	305.04	313.35
Normal Boiling Temp. (K)	543.835	575.3	603.05	629.65
Critical Temp. (K)	708	736	758	778
Liquid Density at 298.15 K (kmol/m ³)	3.600	3.215	2.893	2.623
Descriptor <i>HNar</i>	1.765	1.789	1.81	1.826
Parameter A	0.25142	0.21897	0.19199	0.1716
Parameter B	0.23837	0.23642	0.23337	0.23153
Parameter C	0.28571	0.28571	0.28571	0.28571

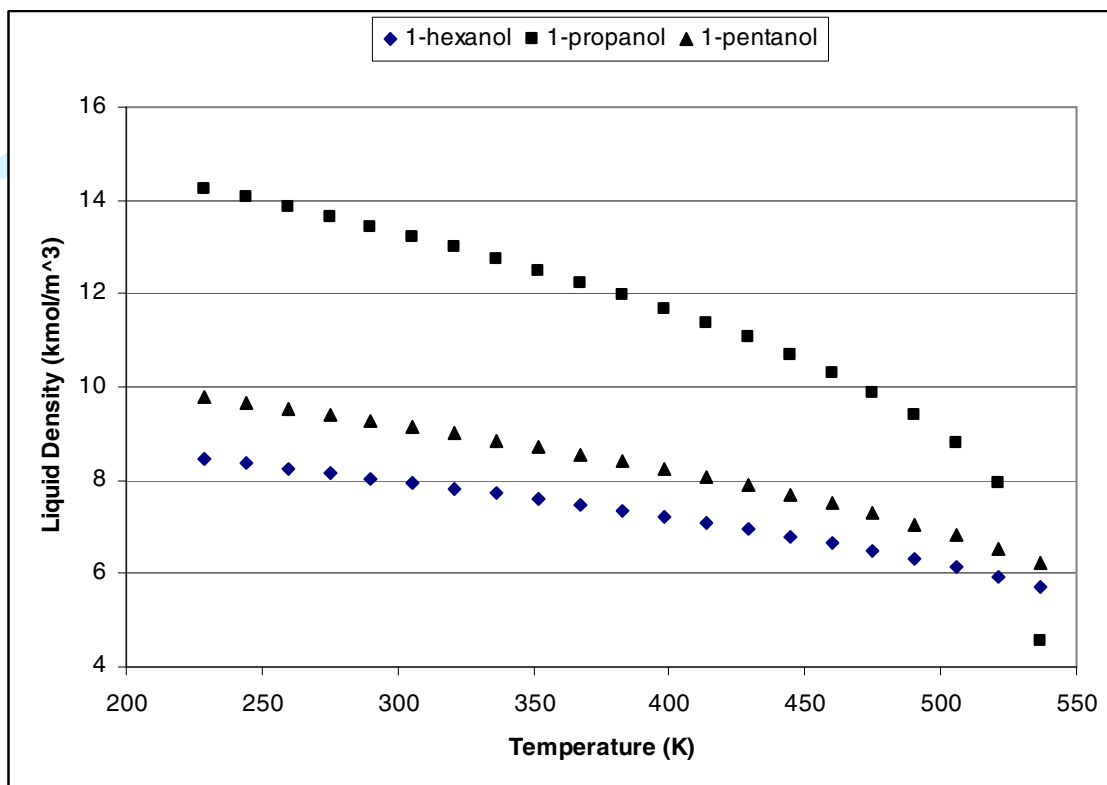


Figure 1. Liquid density of the predictive compounds in the 228.55 K – 536.8 K temperature range

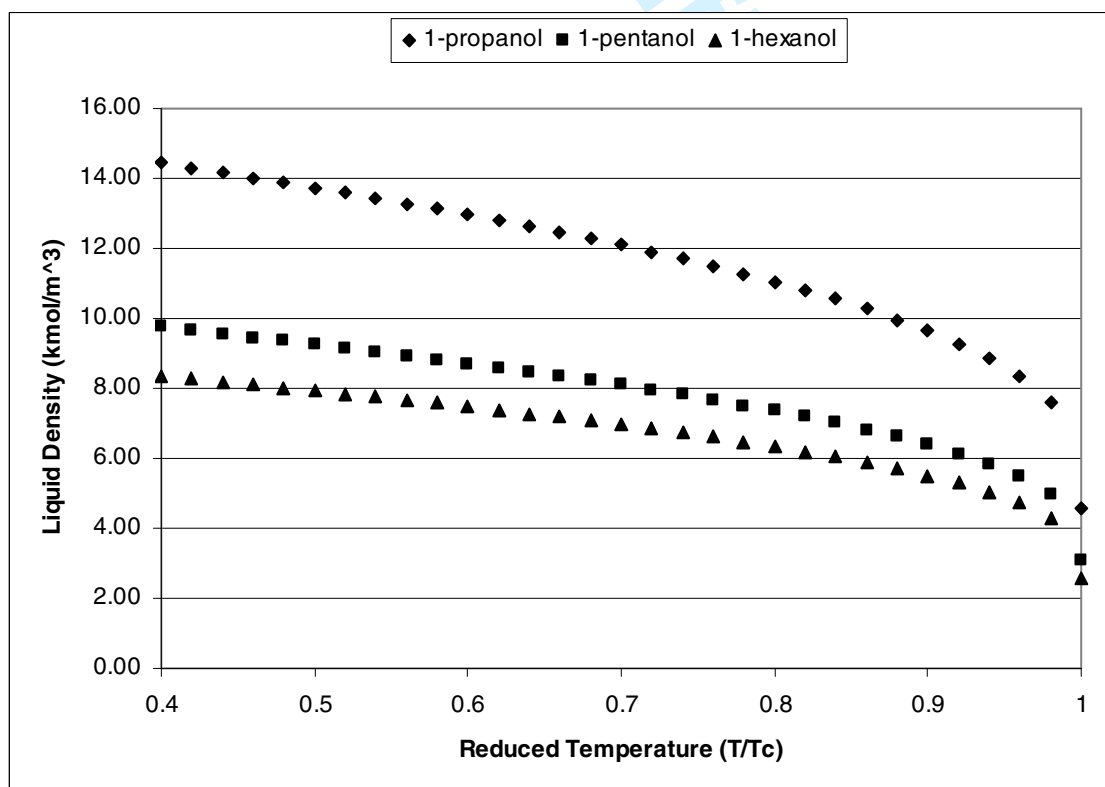


Figure 2. Liquid density of the predictive compounds in the 0.4 – 1.0 reduced-temperature range

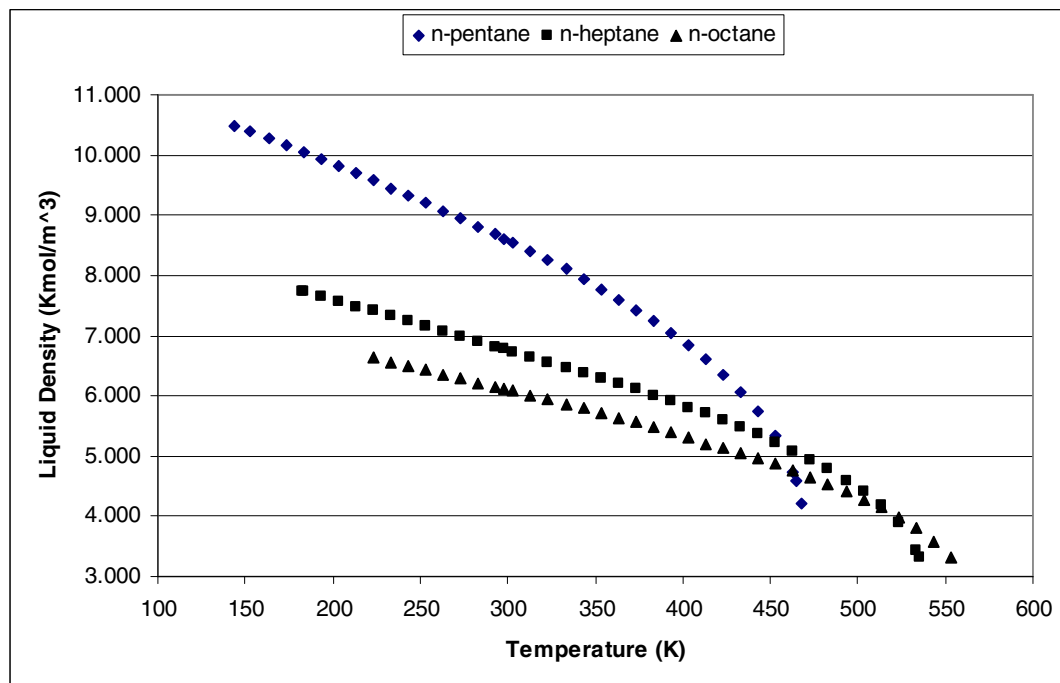


Figure 3. Liquid density of the predictive compounds of the n-alkane group (Data from API¹⁴ and TRC¹⁵).

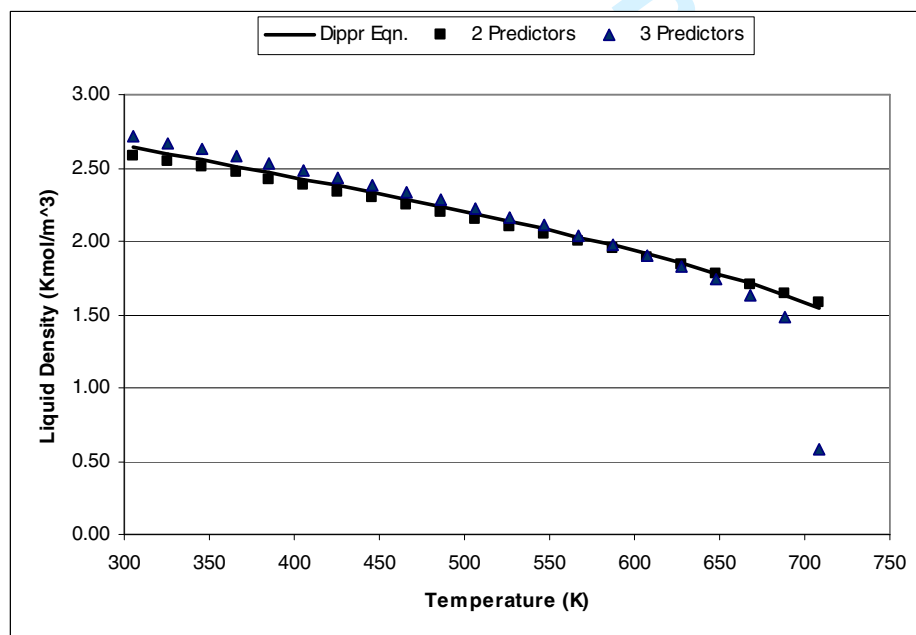


Figure 4. Liquid density data for n-heneicosane predicted by 2-predictor and 3-predictor SC-QS2PR methods.