Modeling Temperature-Dependent Properties of Water via Response Modeling Methodology (RMM) and Comparison with Acceptable Models

Diamanta Benson-Karhi and Haim Shore*

Department of Industrial Engineering, Ben-Gurion University of the Negev, P.O.B. 653, Beer-Sheva 84105, Israel

Mordechai Shacham

Department of Chemical Engineering, Ben-Gurion University of the Negev, P.O.B. 653, Beer-Sheva 84105, Israel

Response Modeling Methodology (RMM) is a new empirical modeling methodology, recently developed. In this paper, a new structured procedure to compare relational models, in terms of goodness-of-fit and stability, is developed and applied to evaluate three types of models: Those obtained by TableCurve2D (a dedicated software for relational modeling), acceptable models recommended by DIPPR (a widely used database for constant and temperature-dependent physical properties), and models derived by RMM. For four pure substances (oxygen, argon, nitrogen, and water), model comparison had been conducted on 14 temperature-dependent physical and thermodynamic properties. Summary tables of ranking the various models are provided. Detailed results are reported in this paper for water. The three variations of the RMM model (two-, three- and four-parameter models) compare favorably with others, both in terms of goodness-of-fit and stability. The unique desirable properties of RMM models are discussed.

1. Introduction

Modeling temperature-dependent physical properties is an ongoing scientific endeavor that spans several centuries. Although some models developed over the years originated in purely theoretical arguments, unsatisfactory accuracy obtained via such models had motivated the development of alternative models that were either empirical or semiempirical in nature. For example, most current acceptable models for vapor pressure are based on an original theory-based model (Clapeyron equation). However, this model had been enhanced empirically, resulting in various semiempirical models, such as Riedel formula (refer, for details, to Daubert[1]).

Response Modeling Methodology (RMM) is a new platform for empirical modeling recently developed (Shore[2] and references therein). RMM has been applied to a myriad of disciplines in science and technology and has been shown to deliver good modeling properties. Examples for such disciplines are chemical engineering,[3,4] quality engineering,[5] hardware reliability,[6,7] software reliability-growth modeling,[8] and forecasting.[9] Furthermore, in some cases, current theoretical, or semiempirical, models have been shown to be special cases of the RMM model.[8,10]

In this paper, and in a forthcoming one by the same authors, results from a research effort, where RMM is used to model properties of pure substances, will be introduced. For four substances (oxygen, argon, nitrogen and water), fourteen temperature-dependent physical and thermodynamic properties are modeled, using available data sets. The resulting RMM models are compared to models derived via two other sources. One source is dedicated software, TableCurve2D (henceforth TC), which allows automatic fitting of mathematical functions to given data sets. TC has a database of nearly 3700 linear and nonlinear functions, which it fits to a given data set, resulting in ranked models according to some goodness-of-fit criteria (note that linearity here refers to “linearity in parameters” and not necessarily to linearity in the regressor variables). A least-squares procedure is used to estimate the parameters. In our analyses of models derived with TC, a normal additive error with constant variance is assumed. This renders the least-squares fitting, offered by TableCurve, into a maximum likelihood procedure. Proper steps are taken (see details later) to ensure that this assumption is valid.

A second source that provides models for comparison is a widely used database for correlating physical and thermodynamic-responses (properties), known as DIPPR (Design Institute for Physical Property Data).[11] Today, it is one of the leading databases for physical properties and their modeling, and it provides both datasets and relational models to represent these data. Based on examining a large number of possible models, DIPPR offers, for every combination of substance and property, the “Acceptable Model”, which is judged by the DIPPR scientific team to capture best the relationship between the response and the affecting factor (temperature). The “Acceptable Model” can be purely theoretical, purely empirical, or a combination thereof. The latter case seems to be typical to most “Acceptable Models” that one can find in the DIPPR database. Note, that while TC provides, in this study, a particular model for each substance−property combination, DIPPR provides a uniform model for each property (although parameter values differ between substances). RMM provides a uniform model (with three variations) to all substances and all properties. Thus, the three approaches compared may perhaps be ranked as moving from the most particular, TC, which offers “tailor-made” model for each modeling endeavor, to DIPPR, which offers the same for each property, to RMM, which is the most general. Consequently, the most interesting comparison in this study is that between models offered by DIPPR and those obtained via RMM.

Models provided by TC, DIPPR, and RMM are compared in this paper, with respect to two dimensions: goodness-of-fit and

* To whom correspondence should be addressed. Tel.: 972-8-6472249. Fax: 972-8-6472958. E-mail address: shor@bgu.ac.il.

10.1021/ie061252x CCC: $37.00
Published on Web 00/00/0000 **PAGE EST: 18**

stability. Several criteria are used to assess each of these dimensions.

The rest of the paper is organized as follows. In the next section (Section 2), we provide a brief overview of the RMM basic models, as used in this paper. In Section 3, we introduce the routine developed to identify the subset of candidate models offered by TC to model a given substance−property pairing, and how the best model is finally selected, using Kullback−Leibler Information Loss criterion (the sample estimate of which is known as Akaike's Information Criterion, AIC). In Section 4, we overview the criteria used in this paper to compare the selected TC model, DIPPR "Acceptable Model", and RMM models, with respect to goodness-of-fit and stability. Section 5 details the water properties modeled in this paper and the source datasets used for each property. Section 6 displays the results (for water). These comprise displaying models obtained by the three approaches (TC, DIPPR, and RMM), and comparisons of the models' quality, in terms of goodness-of-fit and stability. Section 6 comprises 15 subsections, corresponding to the 15 water properties addressed (for solid thermal conductivity, no acceptable measured observations were found and, therefore, this property was excluded from all analyses). Section 7 gives summary comparisons of rankings of the three types of models, in terms of goodness-of-fit and stability for *all* four substances (oxygen, argon, nitrogen, and water). Some practical conclusions are also delineated. Section 8 gives a summary of this paper and some final conclusions.

2. RMM − An Overview

At the heart of RMM is a relational model, which describes a modeled response, Y, in terms of a linear combination of variation-transmitting effects (the linear predictor, η), two possibly correlated normal errors, ϵ_1 and ϵ_2, and a vector of parameters:

$$Y = \exp\left\{\left(\frac{\alpha}{\lambda}\right)[(\eta + \epsilon_1)^{\lambda} - 1] + \mu_2 + \epsilon_2\right\} \qquad (1)$$

where $\{\alpha, \lambda, \mu_2\}$ are three parameters that need to be determined. Assuming that the errors derive from a bivariate normal distribution, and expressing ϵ_2 in terms of ϵ_1 plus an additive *independent* error, the RMM model becomes

$$\begin{aligned} W = \log(Y) &= \left(\frac{\alpha}{\lambda}\right)[(\eta + \sigma_{\epsilon 1} Z_1)^{\lambda} - 1] + \mu_2 + \sigma_{\epsilon 2}\left[\rho\left(\frac{\epsilon_1}{\sigma_{\epsilon 1}}\right) + (1-\rho^2)^{(1/2)} Z_2\right] \\ &= \left(\frac{\alpha}{\lambda}\right)[(\eta + \sigma_{\epsilon 1} Z_1)^{\lambda} - 1] + \mu_2 + \sigma_{\epsilon 2}[\rho Z_1 + (1-\rho^2)^{(1/2)} Z_2] \qquad (2)\end{aligned}$$

where σ_{ϵ_1} and σ_{ϵ_2} are the standard deviations of the errors ϵ_1 and ϵ_2, respectively, Z_1 and Z_2 are two *independent* standard normal variates, and ρ is the correlation between ϵ_1 and ϵ_2. Observing eq 1 or eq 2, one may realize why RMM is flexible in delivering good representation to monotone convex/concave scientific and engineering relationships. Assuming that monotone convexity can generally be represented by the basic "linear, power, exponential" cycle, we realize that (with the random errors ignored and assuming, for convenience, that the scale parameter $\exp(\mu_2) = 1$; also note that as $k \to 0$, $(1/k)(z^k - 1) \to \log(z)$):

- For $\lambda = 0$, $\alpha = 1$: $Y = \eta$ (linear relationship);
- For $\lambda = 0$, $\alpha \neq 1$: $Y = \eta^{\alpha}$ (power relationship, like Newton's kinetic energy formula);
- For $\lambda = 1$: $Y = \exp(\eta)$ (exponential relationship, like the radioactive decay formula);
- For $\lambda \neq 1$: $Y = \exp(\eta^{\lambda})$ (exponential−power relationship, like the Arrhenius formula).

Each successive model represents an increasing intensity of convexity. In fact, as we climb up the "Ladder of Fundamental Monotone Convex Functions" (refer to Shore[2] for details), the basic "linear, power, exponential" cycle can be repeated ad infinitum at the price of introducing additional two parameters for each successive cycle. To demonstrate that, introduce for $\eta + \epsilon_1$ in eq 1 the following two-parameter expression:

$$\exp\left\{\left(\frac{\beta}{\kappa}\right)[(\eta + \epsilon_1)^{\kappa} - 1]\right\} \qquad (3)$$

Introducing $\beta = 1$, $\kappa = 0$ into eq 3 revokes the original $\eta + \epsilon_1$. However, for other values of β and κ, we proceed to "climb the ladder", obtaining an exponential−exponential function, exponential−exponential−power function, and so forth. Modeling monotone convexity thus changes from selecting a model with a specified relationship into an estimation problem, where the values of certain parameters in the RMM model (α and λ in eq 1) determine the form of the final model. Furthermore, the intensity of convexity conveyed by the model becomes a matter of selecting a point on a *continuous spectrum of convexity intensity*, rather than selecting a certain discrete step (model) on the "Ladder of Fundamental Monotone Convex Functions".

To proceed developing the RMM models used in this paper, the model's error quantile function is derived from eq 1, assuming $\eta = 1$. Although a quantile function seems to be suitable to model random variation, however modeling of a systematic variation is required here, a certain adaptation of this function would allow us to use RMM models to model water properties. We have found this adaptation to be convenient in modeling chemical relationships, and both goodness-of-fit and stability indicators testify to its effectiveness (refer also to Shore et al.[3] for an application of this approach).

With $\eta = 1$ (a constant) in eq 2, and assuming, furthermore, that $\rho^2 = 1$ (the two error terms of the RMM model are perfectly correlated), we obtain for eq 2 a quantile function, in terms of a single standard normal variable. Expressed in a re-parametrized form, eq 2 becomes a four-parameter quantile function (henceforth denoted Model RMM_4):

$$W = \log(Y) = \log(M_Y) + \left[\frac{a}{(b/c)}\right][(1 + cZ)^{b/c} - 1] + dZ \qquad (4)$$

where Z is a standard normal variable, M_Y is the median of Y (that is, $Y = M_Y$ for $Z = 0$), and $\{a,b,c,d\}$ are four parameters that need to be determined (estimated). It may be shown that either c or d may be negative but not both (find details in Shore[2]). Note, that Y in eq 4 should be monotone convex (this is a basic assumption of the RMM approach, although adaptation to handle non-monotone curves, such as S-shaped curves, has been developed[9]). Also note that Y is assumed to have a uniform sign (either positive or negative) throughout the modeling range. If this is not the case, a change of location in the data may be required. This may be affected by the addition of a single parameter, δ, to obtain, for eq 4,

$$W = \log(Y + \delta) = \log(M_Y + \delta) + \left[\frac{a}{(b/c)}\right][(1 + cZ)^{b/c} - 1] + dZ \qquad (4a)$$

where e is estimated from the data. This procedure is used here once in modeling water properties. However, the additional parameter is considered only as a linear transformation of the data. Therefore, it is not considered part of the models' parameters, to be so counted in calculating the AIC_C statistic (find details for the latter in Section 4).

Variations of the basic model (eq 4), with a reduced number of parameters, are easily derived, resulting in the following alternative models (find details in Shore[2]):

• The three-parameter RMM model (henceforth, Model RMM_3):

$$W = \log(Y) = \log(M_Y) + (a)[\exp(bZ) - 1] + cZ \quad (5)$$

• The two-parameter RMM model (henceforth, Model RMM_2):

$$W = \log(Y) = \log(M_Y) + \left(\frac{a}{b}\right)[\exp(bZ) - 1] + aZ \quad (6)$$

Note, that while eq 5 is easily derived from eq 4 as an asymptotic case, eq 6 is derived from eq 5, based on an empirical observation that fitting eq 5 to data (a random sample of observations) often results in *c* and (*ab*) being nearly identical in their absolute values (refer, for examples, to Shore,[2] pp 188 and 199). As will be demonstrated by the results in Section 6 and those in two forthcoming papers, RMM_2 is an effective model that exhibits goodness-of-fit and stability properties that are fairly competitive with other models with a much larger-sized parameter set.

Let the regressor variable, temperature (*T*), be expressed by

$$T = M_T + \epsilon = M_T + \sigma_T Z \quad (7)$$

where M_T is the median of *T* and σ_T is the standard deviation. Note that expressing *T* in terms of its median (rather than the mean) ensures that the RMM models preserve the response median (namely, introducing M_T into the models ($Z = 0$) produces the *exact* median of the modeled response; this desirable characteristic has been found to enhance considerably the accuracy of the fitted models).

Introducing for *Z* into eqs 4−6, the RMM models are expressed explicitly in terms of temperature (*T*), and thus are ready for modeling (finding estimates for the parameters) via nonlinear regression.

3. A Procedure To Select the Best Model (Modeling via TC)

Software TC was used to identify candidate models that may best describe a given data set. TC has a bank of nearly 3700 models, which, given a data set, are estimated via least-squares (which is equivalent to maximum likelihood, assuming a normal additive error with constant variance). These models are then ranked according to some selected goodness-of-fit statistic, such as the Pearson correlation. We have found the criteria offered by TC for ranking models wanting, and developed a new procedure. The latter first checks whether the assumptions regarding the error hold. If residuals normality and constancy of variance are in doubt, a logarithmic transformation is applied. Also, the data were screened for outliers (defined as observations exceeding the $\pm 3\sigma_\epsilon$ limits, where σ_ϵ is the standard deviation of the fitted model's error). A third checking was an inspection of the spread of the errors to pinpoint lack of randomness (the existence of a pattern in the scatter plot). This could indicate that the dataset originated in calculated values (rather than empirical observations). We aspired to use only data sets based on real observations, and although in most cases the data source would submit this information, a further check was made for each dataset used.

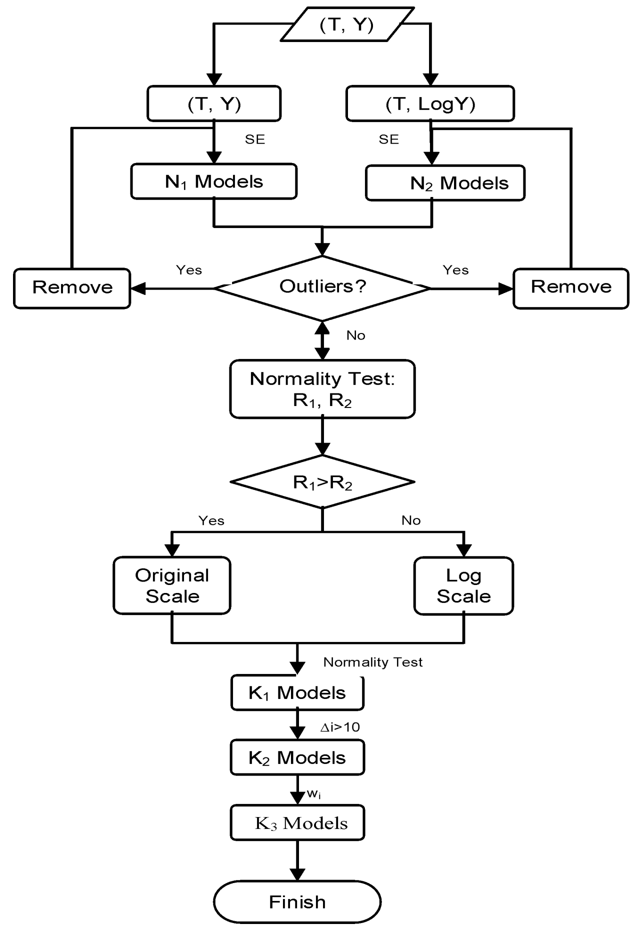

Figure 1. Flowchart of the decision process used to select the best model offered by the TableCurve2D software (TC).

Once preliminary checks have been conducted, the final model to be selected, based on TC-based modeling, essentially pursued the information-theoretic approach of Burnham and Anderson.[12] The latter base their approach on Akaike's Information Criterion (AIC, to be detailed in Section 4), and they rank a subset of candidate models according to the model's "relative likelihood" to be the best model (relative to the data and the complete subset of candidate models):

$$\Delta_i = AIC_{c_i} - \min_K AIC_c$$
$$W_i = \frac{e^{-(1/2)\Delta_i}}{\sum_{i=1}^{K} e^{-(1/2)\Delta_i}} \quad (8)$$

where *K* is the number of candidate models (the subset of best models selected, based on Δ), AIC_c is the *corrected* AIC (find details in Section 4, eqs 10 and 12), and W_i is the "relative likelihood" of model *i* to be the best (in the sense expounded earlier). Only models with $\Delta_i \leq 10$ are included in this set (based on an empirical criterion recommended in Burnham and Anderson[12]). Because the sum-of-likelihood values is 1 (100%), we aspire to have the Pareto principle at work, namely, to have a (preferably) single model with an associated *W* that is much larger than those of all other models in the set. Such a scenario was often encountered in our analyses.

Figure 1 displays the flowchart of the procedure pursued to select the best model based on TC. In this chart, SE is the standard error (of the fitted model), serving as a first criterion to select a subset of N_1 and N_2 models (based on modeling with the response expressed on original and logarithmic scales, respectively), and R_1 and R_2 are the respective numbers of models remaining after conducting the normality test. According to the relative sizes of R_1 and R_2, a decision is made whether to adhere to the original scale or to proceed with a logarithmic scale (for the response). A normality test is then reapplied that screens out models still not compatible with the normality assumption. The remaining K_1 models are then screened based on the Δ_i criterion (eq 8): Any model with $\Delta_i > 10$ is screened out (Burnham and Anderson[12]). For the remaining K_2 models, values of $\{W_i\}$ are calculated anew, and a new subset that has cumulative values of W that exceed a certain threshold (usually 80%) is selected as the final set of K_3 candidate models. From this subset, the best model is selected (the one with the largest W).

Note that, because TC is not restricted by any theoretical consideration (such as the requirement that the same model be used for each property, irrespective of substance), it is expected that TC will *nearly always* provide the best-fitting model (in terms of goodness-of-fit). In fact, this was the case for most analyses. Thus, the selected TC model serves as baseline according to which one may assess the effectiveness of the models offered by DIPPR and by RMM. However, the large number of parameters associated with many of the selected TC models (which often deliver superior goodness-of-fit) is, in many cases, more than compensated for by reduced stability. This will amply be demonstrated in the analyses displayed in this paper and those that follow.

4. Criteria for Goodness-of-Fit and Stability

Measuring the two dimensions of the quality of a model—goodness-of-fit and stability—requires that appropriate statistics be defined and measured for any estimated model. In this section, we expound the criteria used in this paper.

4.1. Criteria for Goodness-of-Fit. All criteria are based on measuring the response in the original scale. Two criteria were used in this study: mean-squared error (MSE) and the corrected Akaike's Information Criterion (AIC_c).

• **Mean-squared error (MSE)** is defined by

$$\mathrm{MSE} = \frac{\mathrm{SSE}}{N - p} \tag{9}$$

where

$$\mathrm{SSE} = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

with $\hat{y}_i$ denoting the predicted value of the ith observation, N being the size of the data set used in fitting the model, and p being the number of parameters estimated. The square root of MSE is denoted RMSE (root-mean-squared error) or, simply, Se.

• **The corrected Akaike's Information Criterion (AIC_C)**, Kullback−Leibler (K-L) Information Criterion, $l(f,g)$, is an information-theoretic measure that expresses the information "loss" incurred when model "g" is used to approximate "reality", given by the "true" model "f". Because the latter is usually unknown, the best we can do is minimize, based on given data, an estimate of the *expected* K−L Information Loss. This is given by AIC, which, assuming for the model's residual an additive normal error with constant variance, is given by

$$\mathrm{AIC} = N \log[\hat{\sigma}_\epsilon^{\,2}(\mathrm{ML})] + 2k \tag{10}$$

where

$$\hat{\sigma}_\epsilon^{\,2}(\mathrm{ML}) = \frac{\mathrm{SSE}}{N}$$

$$k = p + 1$$

with $\hat{\sigma}_\epsilon^{\,2}(\mathrm{ML})$ being a maximum likelihood (ML) estimate of the variance of the residuals. We wish to minimize the *expected* K−L Information Loss, and, therefore, minimization of AIC is required for the selected model. The assumption that the models estimated via TC have an additive normal error (an assumption that is being checked by the procedure outlined in Section 3), allows us to estimate the parameters via the least-squares approach (as offered by TC), and then calculate the associated AIC values by the simple relationship

$$\mathrm{SSE} = N\hat{\sigma}_\epsilon^{\,2}(\mathrm{ML}) = (N - p)\hat{\sigma}_\epsilon^{\,2}(\mathrm{LS}) \tag{11}$$

where $\hat{\sigma}_\epsilon^2(\mathrm{LS})$ is the MSE given by eq 9.

Hurvich and Tsai[13] proposed a correction for the AIC statistic for cases where the number of parameters in the model is large, relative to the size of the data set. This resulted in the corrected AIC:

$$\mathrm{AIC_c} = \mathrm{AIC} + \frac{2k(k+1)}{N - k - 1} \tag{12}$$

The two statistics, MSE and AIC_c, although not statistically independent, were used as the major goodness-of-fit statistics throughout this study. Find further details about these measures in Burnham and Anderson.[12]

4.2. Criteria for Stability. A stable model is one whose parameters are stable, relative to the data used to fit the model. For example, an estimated stable model would submit accurate predicted values, even for temperature ranges different from those used in estimating the model's parameters.

In this study, two major criteria of stability were formulated to judge a model's stability: the relative size of the 95% confidence interval and PRESS (Prediction Error Sum of Squares).

95% Confidence Interval for a Parameter Estimate. This is a go/no-go measure, which judges any fitted model to be nonstable if any of its parameters' estimates has a 95% confidence interval width that exceeds the parameter's estimate. Though this criterion is arbitrary with respect to the associated threshold (namely, that the *relative* confidence interval width should not exceed a unity), it has been shown to provide an effective measure for the model's stability. In the tables showing the results of this study (Section 6), any indication of instability is marked by the confidence intervals being given in *bold* type.

PRESS. This is a traditional statistic used to check a model's stability. It is defined by

$$\mathrm{PRESS} = \sum_{i=1}^{N} e_{(i)}^{\;2} = \sum_{i=1}^{N} (y_i - \hat{y}_{(i)})^2 \tag{13}$$

where $\hat{y}_{(i)}$ is the model's predicted value for observation i, given that the model was fitted with observation i omitted (from the sample of N data points; find details in, for example, Myers et al.[14]). Minimization of PRESS is desired. Large gaps between the root square of PRESS and RMSE (or Se) is indicative of a scenario of overfitting (the model contains too many parameters, relative to the information provided by the data). An overfitted model submits small residuals; however, poor prediction for data not used in the fitting is also encountered. A typical case of overfitting is routinely encountered in fitting a polynomial with too many terms. While PRESS is routinely given (and easily calculated) for linear regression implementation, it is rarely given, in available statistical software packages, when fitting models via nonlinear regression. Therefore, a special program had been written in MATHEMATICA that provided PRESS values for all models compared.

5. Water Properties and Data Sources

Fourteen temperature-dependent physical and thermodynamic water properties have been addressed. These properties, with other relevant information (data sources, ranges and median values for temperature and response, outliers) are displayed in the Appendix. Note that excluded outliers refer to observations, analyzed via TC, with residuals that exceed $\pm 3\sigma_\epsilon$. Also, since modeling with RMM assumes monotone convexity, observations near critical points, when the monotone convexity property was violated, were also excluded (for example, liquid heat capacity, LHC). As related to earlier, modeling non-monotone relationships via RMM has been demonstrated in Shore and Benson-Karhi;[9] however, it is not attempted here.

For solid thermal conductivity (STC), no acceptable sets of measured data were found. Therefore, this property is excluded from all analyses. Other considerations for exclusion are detailed in the Appendix.

6. Results

Analysis for each property is uniformly structured for all properties according to the following scheme. First, once an appropriate data set was identified, it was plotted to ensure that continuous monotone convexity was not violated. In rare cases where points at either ends of the set indicated non-monotonicity (such as liquid heat capacity (LHC)) or non-continuity (due to proximity to critical points, as with liquid density (LD), vapor pressure (VP), and LHC), those points were not included in the analysis (see details in the Appendix). After the initial screening of the data, the best model suggested by TC was identified (as detailed in Section 3), and then RMM analysis was implemented. Finally, the parameters of DIPPR "Acceptable Model" were re-estimated (although they are given in the DIPPR database) to guarantee that later comparison of models is based on estimates derived from identical data sets.

6.1. Solid Density (SD). A plot of water solid density is given in Figure 2.

Analysis by TC on the original scale showed that 30 models provide cumulative W values of

$$\sum_{i=1}^{K_3} W_i = 0.7 \qquad (K_3 = 30)$$

The large number of models in the set clearly shows that there is no restricted small set of models with superior performance. Among the set of 30 models, the linear model is the leader.

With regard to DIPPR, the "Acceptable Model" for SD is a fourth-degree polynomial. For water, only values for the first two parameters are suggested by DIPPR, which implies that both TC and DIPPR lead to a linear model.

Applying RMM, a first run of model RMM_4 (eq 4) results in values that indicate a linear model ($a \approx 1$, $b \approx 0$, $d \approx 0$). Therefore, the RMM approach also leads to a linear model as the best model.

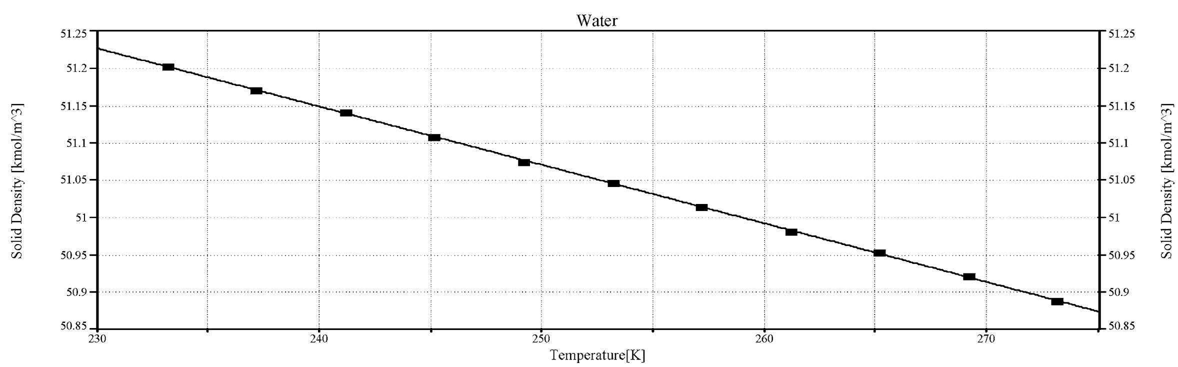

Figure 2. Data for water solid density (SD).

Table 1. Goodness-of-Fit and Stability Statistics for Water Solid Density (SD)

Table 1.1. Goodness-of-Fit Statistics for Water Solid Density

model	number of parameters	normality (?)	MSE	AIC_c
TC#1 = DIPPR	2	+	1.80×10^{-6}	−138.3
RMM2	2	+	3.14×10^{-6}	−132.17
RMM3	3	+	3.50×10^{-6}	−127.01
RMM4	4	+	4.01×10^{-6}	−119.68

Table 1.2. Stability Statistics for Water Solid Density[a]

	TC = DIPPR		RMM2	
parameter	value	width of confidence interval	value	width of confidence interval
a	3.9716	7.1453×10^{-4}	−0.00102	2.484×10^{-5}
b	-1.536×10^{-4}	2.82×10^{-6}	**0.03612**	**0.076601**
PRESS	2.27545×10^{-5}		3.63003×10^{-5}	

[a] Values given in boldface type represent possible instability.

F

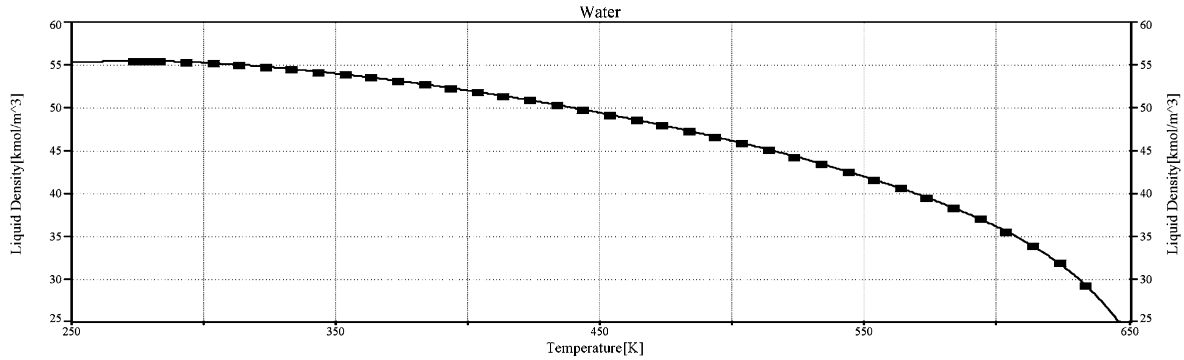

Figure 3. Data for water liquid density (LD).

Table 2. Goodness-of-Fit and Stability Statistics for Water Liquid Density (LD)

Table 2.1. Goodness-of-Fit Statistics for Water Liquid Density

model	number of parameters	normality (?)	MSE	AIC_c
TC#6065	16	+	1.1049×10^{-7}	−993.73
DIPPR	5	+	5.9074×10^{-3}	−320.15
RMM4	4	+	0.022402	−236.21
RMM3	3	+	0.10274	−140.03
RMM2	2	+	0.15634	−114.4

Table 2.2. Stability Statistics for Water Liquid Density[a]

parameter	DIPPR value	DIPPR width of confidence interval	RMM4 value	RMM4 width of confidence interval
a	**1.26357**	**7.53088**	3.41353	3.19840
b	206.961	66.3411	−0.347448	3.92699×10^{-2}
c	−501.607	180.224	−0.394138	1.45888×10^{-3}
d	687.408	209.128	1.28852	1.25809
e	−344.289	89.5246		
PRESS	1.19217		1.59213	

[a] Values given in boldface type represent possible instability.

Tables 1.1 and 1.2 display results obtained for goodness-of-fit and stability. Table 1.1 shows that, among the RMM models, RMM_2 is the best fitting model (with performance nearly identical to the linear model, adopted by TC and DIPPR). However, inspecting Table 1.2 clearly indicates that the RMM_2 parameter *b* is unstable, with the confidence interval covering the “zero point”, which is a clear indication that the true value of *b* is zero. This leads again to the conclusion that the right model is the linear one (as RMM_4 parameters also indicate).

6.2. Liquid Density (LD). A plot of water liquid density is given in Figure 3.

Analysis by TC on the original scale showed that two models are the leaders, with cumulative *W* values of

$$\sum_{i=1}^{K_3} W_i = 0.991 \qquad (K_3 = 2)$$

The leader is a polynomial of order 15 (eq 6065 in TC):

$$Y = a + bT + cT^2 + ... + pT^{15}$$

The DIPPR “Acceptable Model”, unique for water liquid density, is given by

$$Y = a + bt^{0.35} + ct^{2/3} + dt + et^{4/3}$$

where

$$t = 1 - T_r$$

$$T_r = \frac{T}{T_c}$$

$$T_c = 647.096 \text{ K}$$

Applying RMM, all three models (RMM_2, RMM_3, RMM_4) were fitted.

Tables 2.1 and 2.2 display results obtained for goodness-of-fit and stability, respectively. Table 2.2 also displays estimates of the parameter values for DIPPR and RMM, with the associated confidence interval widths. Because of the large number of parameters in the selected TC model, these are omitted from Table 2.2.

6.3. Solid Vapor Pressure (SVP). A plot of water SVP is given in Figure 4.

Analysis by TC on the original scale showed that 17 models are the leaders with cumulative *W* values of 80%. The leader is the following second-degree polynomial (eq 1413 in TC):

$$\log(Y) = a + \frac{b}{T} + \frac{c}{T^2}$$

The DIPPR “Acceptable Model”, uniform for vapor pressure, is given by

$$Y = \exp\left[a + \frac{b}{T} + c\log(T) + dT^e\right]$$

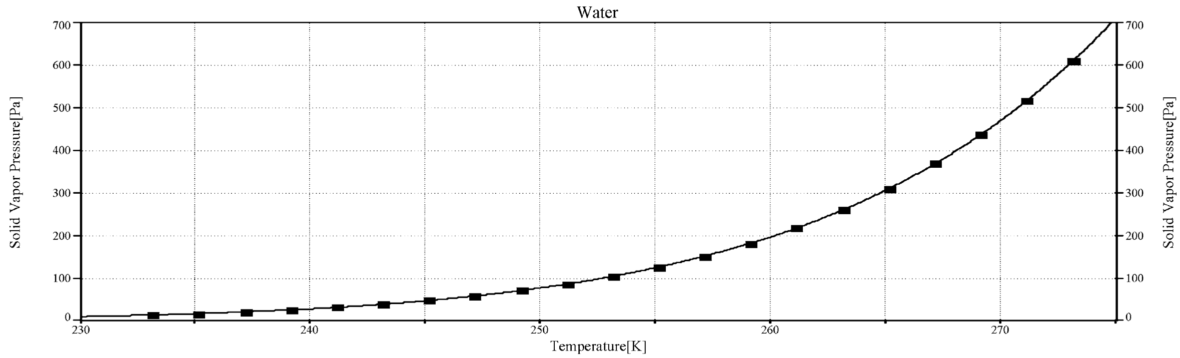

Figure 4. Data for water solid vapor pressure (SVP).

Table 3. Goodness-of-Fit and Stability Statistics for Water Solid Vapor Pressure (SVP)

Table 3.1. Goodness-of-Fit Statistics for Water Solid Vapor Pressure

model	number of parameters	normality (?)	MSE	AIC_c
TC#1413	3	+	3.78112×10^{-4}	−166.239
DIPPR	5	+	6.57917×10^{-4}	−149.254
RMM3	3	+	8.13110×10^{-2}	−48.0808
RMM2	2	+	9.57048×10^{-2}	−46.3862
RMM4	4	+	8.58603×10^{-2}	−44.6755

Table 3.2. Stability Statistics for Water Solid Vapor Pressure[a]

	TC		DIPPR		RMM3	
parameter	value	width of confidence interval	value	width of confidence interval	value	width of confidence interval
a	28.5778	8.9591×10^{-2}	**72.9833**	**2869.32**	**−1.58993**	**1.59762**
b	−5976.45	47.261	**−7422.52**	**56928.1**	−0.28692	0.154401
c	−21139.4	6229.6	**−7.26822**	**524.599**	0.760673	0.213715
d			**4.3647×10^{-5}**	**4.5212×10^{-2}**		
e			**1.84697**	**148.227**		
PRESS	8.9724×10^{-3}		0.0133237		1.71616	

[a] Values given in boldface type represent possible instability.

Applying RMM, all three models (RMM_2, RMM_3, RMM_4) have approximately equal goodness-of-fit, which is inferior to that provided either by TC or DIPPR. This may be realized from the goodness-of fit statistics, displayed in Table 3.1. Table 3.2 shows stability statistics for the TC model, the DIPPR model, and RMM_3 (the RMM with the best goodness-of-fit).

We realize that, although the DIPPR model has small value of PRESS, the confidence interval widths of all its parameters testify to lack of stability (those that are larger than the respective parameter estimate are given in *bold*). This essential instability was also experienced in the difficulties encountered to achieve convergence in the nonlinear regression fitting of the DIPPR model.

6.4. Vapor Pressure (VP). A plot of water VP is given in Figure 5.

Analysis by TC on a logarithmic scale showed that a single model is a leader, with a *W* value that reaches nearly 100%. This is an eighth-degree polynomial (eq 6005 in TC):

$$\log(Y) = a + bT + cT^2 + ... + iT^8$$

The DIPPR “Acceptable Model”, uniform for vapor pressure, is given by

$$Y = \exp\left[a + \frac{b}{T} + c\log(T) + dT^e\right]$$

Modeling by RMM shows that RMM_4 delivers a model with the best goodness-of-fit. Table 4.1 shows statistics for goodness-of-fit for all five models, and Table 4.2 shows stability statistics for DIPPR and RMM_4.

We realize that, in terms of both goodness-of-fit and stability, RMM_4 delivers superior performance.

6.5. Heat of Vaporization (HV). A plot of water HV is given in Figure 6.

Analysis by TC on the original scale showed that fewer than 30 models are proposed with acceptable goodness-of-fit (R^2adj > 0.9). A single model is a leader, with a *W* value that reaches nearly 100%. This model is of the “Chebychev Polynomial” type, where the specially defined regressor is restricted to the range $\{-1,1\}$. The model is (eq 6820 in TC)

$$x' = \text{scaled}(x)$$

$$T_n(x') = \cos(n \arccos(x'))$$

$$Y = a + bT_1(x') + cT_2(x') + ... + uT_{20}(x')$$

The DIPPR “Acceptable Model”, uniform for HV across all substances, is given by

$$Y = a(1 - T_r)^{b + cT_r + dT_r^2 + eT_r^3}$$

where

$$T_r = \frac{T}{T_c}$$

Modeling by RMM shows that RMM_4 delivers the best goodness-of-fit, with MSE somewhat lagging behind that of DIPPR. Table 5.1 shows statistics for the goodness-of-fit for all five models. Table 5.2 shows stability statistics for DIPPR and RMM_4.

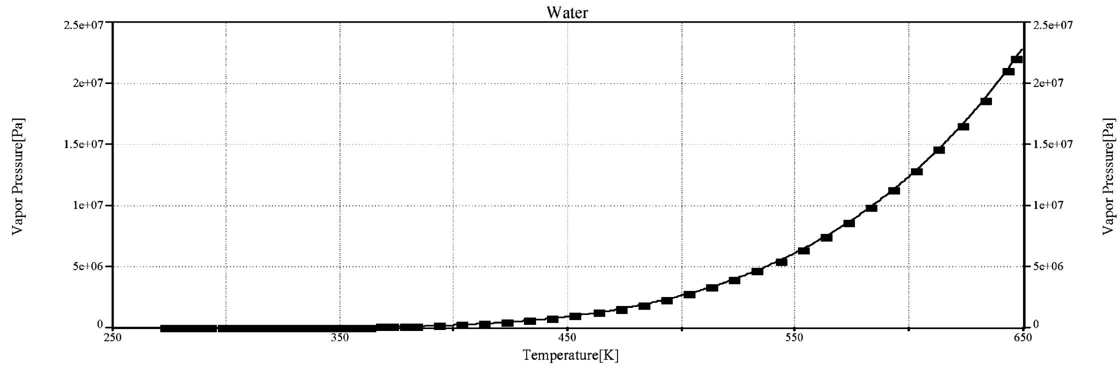

Figure 5. Data for water vapor pressure (VP).

Table 4. Goodness-of-Fit and Stability Statistics for Water Vapor Pressure (VP)

Table 4.1. Goodness-of-Fit Statistics for Water Vapor Pressure

model	number of parameters	normality (?)	MSE	AIC_c
TC#6005	9	+	7.94112×10^{-10}	−1010.89
RMM4	4	+	5.92746×10^{-7}	−695.364
DIPPR	5	+	5.30289×10^{-6}	−586.49
RMM3	3	+	6.66117×10^{-5}	−465.402
RMM2	2	+	5.90870×10^{-3}	−246.944

Table 4.2. Stability Statistics for Water Vapor Pressure[a]

	DIPPR		RMM4	
parameter	value	width of confidence interval	value	width of confidence interval
a	145.587	145.260	15.5283	0.694146
b	−8619.88	2652.92	−0.584088	0.0200233
c	**−26.1799**	**88.3973**	0.187757	0.0104248
d	**4.88384**	**83.3265**	0.586706	0.0331216
e	**0.371493**	**1.40446**		
PRESS	3.15979×10^{-4}		4.19699×10^{-5}	

[a] Values given in boldface type represent possible instability.

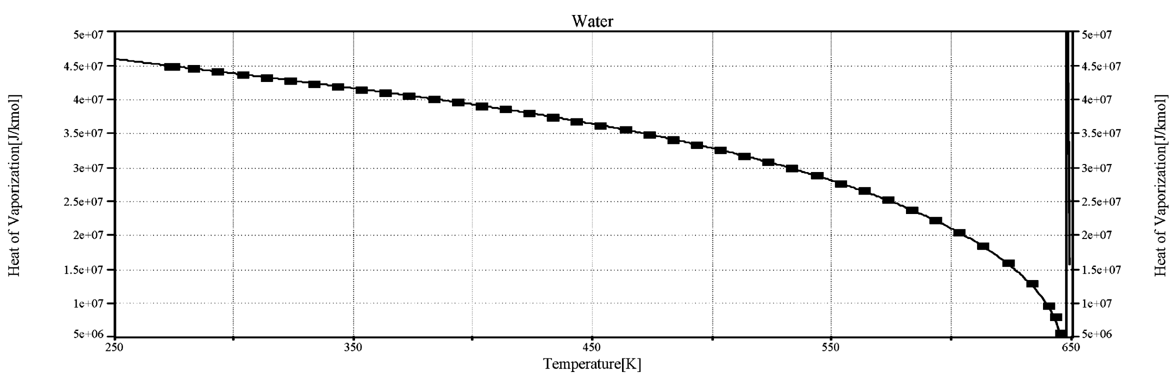

Figure 6. Data for water heat of vaporization (HV).

Table 5. Goodness-of-Fit and Stability Statistics for Water Heat of Vaporization (HV)

Table 5.1. Goodness-of-Fit Statistics for Water Heat of Vaporization

model	number of parameters	normality (?)	MSE	AIC_c
TC#6820	21	+	1.6848×10^{8}	847.4263
DIPPR	5	+	1.10956×10^{9}	863.0547
RMM4	4	+	7.07617×10^{9}	937.3854
RMM3	3	+	5.10653×10^{11}	1111.31
RMM2	2	−	1.11092×10^{12}	1141.78

Table 5.2. Stability Statistics for Water Heat of Vaporization[a]

	DIPPR		RMM4	
parameter	value	width of confidence interval	value	width of confidence interval
a	4.62365×10^{7}	2.9573×10^{6}	0.409244	0.0329019
b	−0.676845	0.410929	**0.0150959**	**0.0429355**
c	2.89586	1.13526	−0.638854	3.77491×10^{-3}
d	−3.44171	1.25276	0.0351164	0.0200207
e	1.56802	0.515306		
PRESS	7.00571×10^{10}		8.34809×10^{11}	

[a] Values given in boldface type represent possible instability.

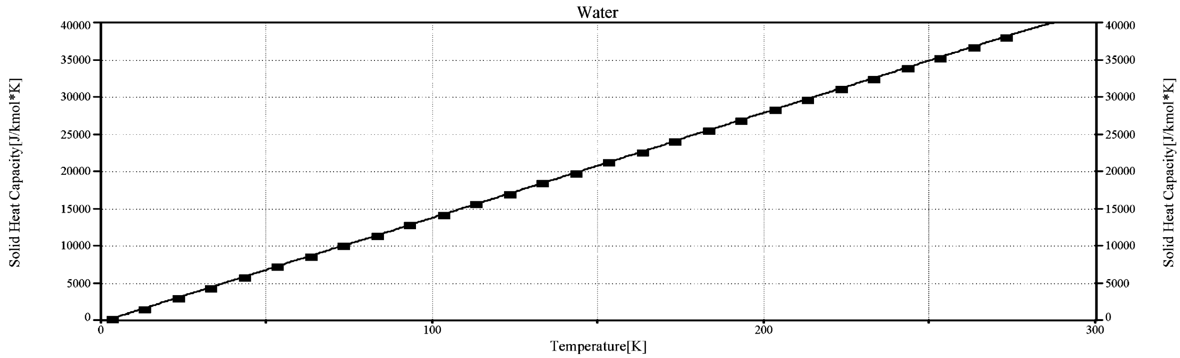

Figure 7. Data for water solid heat capacity (SHC).

Table 6. Goodness-of-Fit and Stability Statistics for Water Solid Heat Capacity (SHC)

Table 6.1. Goodness-of-Fit Statistics for Water Solid Heat Capacity

model	number of parameters	normality (?)	MSE	AIC_c
TC#1=DIPPR	2	+	0.066677	−70.8962
RMM4	2	+	0.066677	−70.8962
RMM3	3	−	113 639	332.508
RMM2	2	−	129 128	334.445

Table 6.2. Stability Statistics for Water Solid Heat Capacity[a,b]

	TC = DIPPR = RMM = Linear		RMM4	
parameter	value	width of confidence interval	value	width of confidence interval
a	−262.489552	0.397447	0.999886	6.89687×10^{-4}
b	140.5184989	0.0024836	**-1.07880×10^{-4}**	**4.56301×10^{-4}**
c			0.603587	7.34060×10^{-5}
d			**6.62731×10^{-5}**	**4.24951×10^{-4}**
PRESS	1.95916			

[a] Values given in boldface type represent possible instability. [b] RMM_4 indicates linearity.

We realize that in terms of both goodness-of-fit and stability, DIPPR delivers superior performance.

6.6. Solid Heat Capacity (SHC). A plot of water SHC is given in Figure 7.

Analysis by TC on the original scale showed that many models (131) display nearly equal performance, with cumulative W values that reaches about 80%. This implies that there is no restricted set of models with superior performance. The leading model is a linear model.

DIPPR also suggests a linear model as the "Acceptable Model", and so is the RMM_4 model (estimating its parameters leads to ($a \approx 1$, $b \approx 0$, $d \approx 0$), as previously encountered with property 1 (SD)). The resulting statistics of goodness-of-fit and stability are displayed in Tables 6.1 and 6.2, respectively. Parameters for RMM_4 are shown in Table 6.2 to demonstrate how close estimates of a, b, and d are to their asymptotic values (1, 0, and 0, respectively), which led us to deduce that, indeed, a linear model is suggested by the RMM_4 model.

6.7. Liquid Heat Capacity (LHC). A plot of water LHC is given in Figure 8.

The first point in the data set (273.15 K) is close to the triple point, which violates the requirement of continuous monotone convexity by RMM. Therefore, this point is deleted from further analyses. Analysis by TC on the original scale showed that there are two leading models with cumulative W values of 87%. The leader from these two models is (eq 7112 in TC)

$$\sqrt{Y} = \frac{a + cT}{1 + bT + dT^2}$$

The DIPPR "Acceptable Model" is a fourth-degree polynomial:

$$Y = a + bT + cT^2 + dT^3 + eT^4$$

Modeling by RMM shows that RMM_4 is the leader, in terms of goodness-of-fit.

The resulting statistics of goodness-of-fit and stability are displayed in Tables 7.1 and 7.2, respectively.

One realizes that RMM_4 is superior, both in terms of goodness-of-fit (unexpectedly, better even than the proposed TC model) and in terms of stability (as judged by PRESS). However, the stability of the parameters' estimates is weak. This suggests that perhaps switching to RMM_3 would provide a more stable and still well-fitting model (with only three parameters versus the five suggested by DIPPR).

6.8. Ideal Gas Heat Capacity (IGHC). A plot of water IGHC is given in Figure 9.

It seems that there is an inflection point that causes the graph of the data to behave similar to an S-shape function. This violates the basic RMM assumption of monotone convexity. Although with certain modifications, S-shaped functions are easily modeled via RMM (refer to Shore and Benson-Karhi[9]), these modifications are not warranted by the moderate deviation from monotone convexity displayed in Figure 9. Therefore, RMM models, as introduced previously, will be employed in this analysis. However, diminished accuracy is expected.

Implementing TC, we have restricted the candidate models to those having "a constant sign in the first derivative". This is an option provided by TC, which implies that the first derivative of Y has the same sign, but it is not necessarily monotone throughout the whole range. Applying TC to the data on the

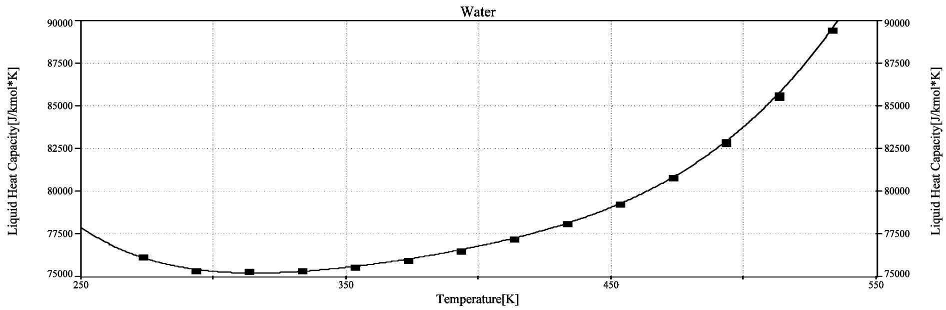

Figure 8. Data for water liquid heat capacity (LHC).

Table 7. Goodness-of-Fit and Stability Statistics for Water Liquid Heat Capacity (LHC)

Table 7.1. Goodness-of-Fit Statistics for Water Liquid Heat Capacity

model	number of parameters	normality (?)	MSE	AIC_c
RMM4	4	+	3829.43	143.405
TC#7112	4	+	6179.68	151.061
DIPPR	5	+	6062.27	154.696
RMM3	3	+	15 879.80	163.079
RMM2	2	+	196 413	200.871

Table 7.2. Stability Statistics for Water Liquid Heat Capacity[a]

	TC		DIPPR		RMM4	
parameter	value	width of confidence interval	value	width of confidence interval	value	width of confidence interval
a	284.477	29.9666	258 538	121 841	**−0.84319**	**4.74913**
b	-1.180×10^{-3}	1.1502×10^{-3}	−1915.72	1220.97	**−0.64807**	**0.90130**
c	−0.391195	0.151624	7.49368	4.52824	−0.73879	0.34968
d	$\mathbf{-4.123 \times 10^{-7}}$	$\mathbf{8.9413 \times 10^{-7}}$	−0.0131131	7.3706×10^{-3}	**−0.56282**	**3.79389**
e			8.78003×10^{-6}	4.4457×10^{-6}		
PRESS	227 149		458 102		110 015	

[a] Values given in boldface type represent possible instability.

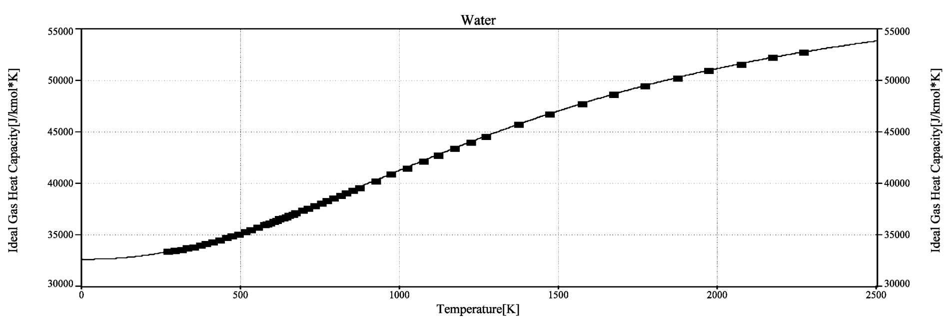

Figure 9. Data for water ideal gas heat capacity (IGHC).

Table 8. Goodness-of-Fit and Stability Statistics for Water Ideal Gas Heat Capacity (IGHC)

Table 8.1. Goodness-of-Fit Statistics for Water Ideal Gas Heat Capacity

model	number of parameters	normality (?)	MSE	AIC_c
TC#6061	12	+	30.4913	206.07
DIPPR	5	+	154.498	280.71
RMM3	3	+	50943.70	591.01
RMM4	4	+	161 001	654.51
RMM2	2	+	232 073	671.16

Table 8.2. Stability Statistics for Water Ideal Gas Heat Capacity (DIPPR Model)

parameter	value	width of confidence interval
a	33397.1	34.4961
b	26822.2	113.952
c	2631.87	29.4743
d	9047.77	252.125
e	1185.46	19.5051

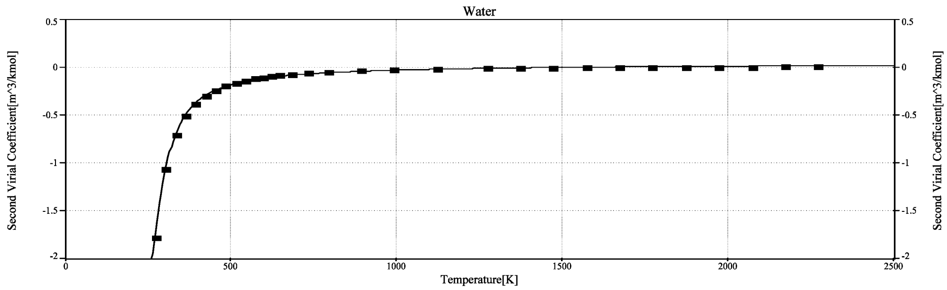

Figure 10. Data for water second virial coefficient (SVC).

Table 9. Goodness-of-Fit and Stability Statistics for Water Second Virial Coefficient (SVC)

Table 9.1. Goodness-of-Fit Statistics for Water Second Virial Coefficient

model	number of parameters	normality (?)	MSE	AIC_c
TC#6505	8	+	1.85×10^{-9}	−606.08
RMM4	4	+	1.65×10^{-7}	−475.94
DIPPR	5	+	1.85×10^{-7}	−470.5
RMM2	2	+	7.46×10^{-5}	−289.8
RMM3	3	+	7.50×10^{-5}	−288.07

Table 9.2. Stability Statistics for Water Second Virial Coefficient

	DIPPR		RMM4	
parameter	value	width of confidence interval	value	width of confidence interval
a	0.0222237	1.37939×10^{-3}	−1.61272	3.02414×10^{-2}
b	−26.3803	1.93427	−0.459742	6.47907×10^{-2}
c	-1.6750×10^{7}	3.83×10^{5}	1.03261	1.73570×10^{-2}
d	-3.8939×10^{19}	3.462×10^{18}	0.0658035	5.05439×10^{-2}
e	3.13317×10^{21}	8.4254×10^{20}		
PRESS	2.37529×10^{-4}		2.88376×10^{-4}	

original scale, two leading models have cumulative W values of 94%. The leading model is an 11th-degree polynomial (eq 6061 in TC):

$$Y = a + bT + cT^2 + ... + lT^{11}$$

The "Acceptable Model" recommended by DIPPR is uniform for IGHC, and is given by

$$Y = a + b\left[\frac{c/T}{\sinh(c/T)}\right]^2 + d\left[\frac{e/T}{\cosh(e/T)}\right]^2$$

RMM modeling, as expected, delivers inferior performance for all models. Tables 8.1 and 8.2 display, respectively, the goodness-of-fit and stability statistics.

The first table compares all five models, while the second table shows statistics associated with the DIPPR model only. Although the parameters' estimates for the selected TC model are not shown (there are 12 of these), stability checking indicates strong multi-collinearity, a sure sign of overfitting (using too many parameters in the model).

6.9. Second Virial Coefficient (SVC). A plot of water SVC is given in Figure 10.

TC modeling on the original scale identifies a single model with a W value of 86%. This is a seventh-degree polynomial (eq 6505 in TC):

$$Y = a + bX + cX^2 + ... + hX^7$$

where

$$X = \frac{1}{T}$$

The DIPPR "Acceptable Model" is uniform for SVC, and is given by

$$Y = a + \frac{b}{T} + \frac{c}{T^3} + \frac{d}{T^8} + \frac{e}{T^9}$$

Modeling with RMM requires that all data have the same sign (the sign of the predicted values is solely determined by the response median, as an inspection of the RMM models in eqs 4−6, with the response expressed on the original scale, can attest). Because some of the data in Figure 10 are negative, a location parameter is added to the RMM models to allow for such values. As related to previously (Section 2, following eq 4), the addition of a location parameter is a linear transformation of the data, and therefore in the pursuing calculation of goodness-of-fit statistics for the various RMM models, it is not considered an additional estimated parameter. Running RMM_2 with the additional parameter δ, we obtain $\delta = -0.031638$. Because the maximum response value in the data is 0.009548, this value of δ implies that, after a location change in the data, all response observations are negative, namely, having same sign (as required). Statistics for goodness-of-fit and stability are displayed by Tables 9.1 and 9.2, respectively.

One realizes that RMM_4 (four parameters) and DIPPR (five parameters) exhibit about equal goodness-of-fit and stability, although RMM_4 is slightly better on both counts. Note, once again, the uniformly centered values of the RMM parameters' estimates, which are confined to the $\{-2,2\}$ range.

6.10. Liquid Viscosity (LV). A plot of water LV is given in Figure 11.

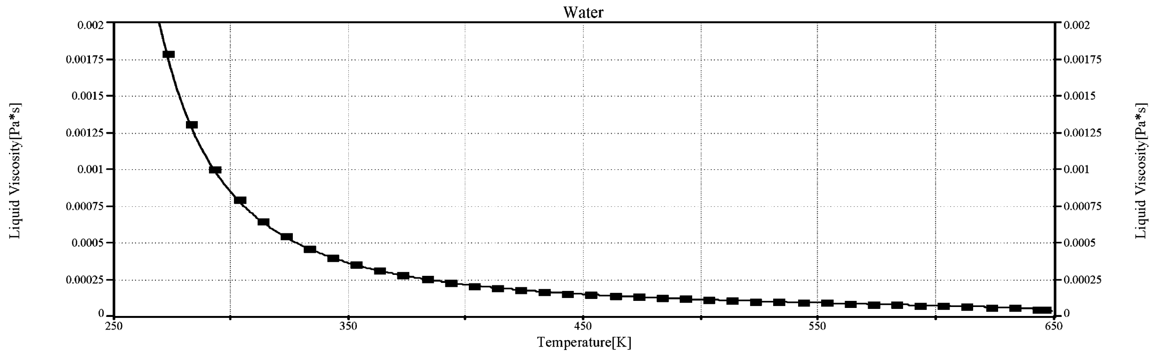

Figure 11. Data for water liquid viscosity (LV).

Table 10. Goodness-of-Fit and Stability Statistics for Water Liquid Viscosity (LV)

Table 10.1. Goodness-of-Fit Statistics for Water Liquid Viscosity

model	number of parameters	normality (?)	MSE	AIC_c
TC#6112	5	+	1.07125×10^{-11}	−1026.50
RMM3	3	+	1.96781×10^{-11}	−1004.70
RMM4	4	+	2.10406×10^{-11}	−1000.50
RMM2	2	+	1.34560×10^{-10}	−927.29
DIPPR	3	+	3.40034×10^{-10}	−887.89

Table 10.2. Stability Statistics for Water Liquid Viscosity

	TC		DIPPR		RMM3	
parameter	value	width of confidence interval	value	width of confidence interval	value	width of confidence interval
a	27115.6	9374.07	−46.6025	10.9421	0.065194	0.0222563
b	−307.064	104.321	3473.60	558.397	−1.92501	0.156856
c	1.19303	0.426125	4.90836	1.59114	−0.55063	0.0671661
d	-1.854×10^{-3}	7.5493×10^{-4}				
e	1.0953×10^{-6}	4.8779×10^{-7}				
PRESS	2.25847×10^{-9}		3.39968×10^{-8}		2.27547×10^{-9}	

TC modeling on the original scale identifies a single leading model with a *W* value of nearly 100%. This is a reciprocal fourth-degree polynomial (eq 6112 in TC):

$$\frac{1}{Y} = a + bT + cT^2 + dT^3 + eT^4$$

The DIPPR "Acceptable Model" is uniform for LV, and is given by

$$Y = \exp\left[a + \frac{b}{T} + c\log(T) + dT^e\right]$$

Because *d* is actually zero, a three-parameter model is suggested.

RMM modeling shows that all three versions of the RMM approach deliver better goodness-of-fit statistics, relative to the DIPPR "Acceptable Model". This is shown in Table 10.1, where, as expected, the TC model is the leader.

Table 10.2 shows stability statistics for models derived by TC, DIPPR, and RMM_3 (the latter being the leader among the RMM models).

One realizes that, in terms of PRESS, TC and RMM_3 models show the highest stability, followed by the DIPPR model. All the parameters' estimates are stable.

6.11. Vapor Viscosity (VV). A plot of water VV is given in Figure 12.

TC modeling on the original scale identifies three leading models with a cumulative *W* value of 87%. The leader among these is (eq 2099 in TC)

$$Y = a + bT + cT^3 + de^{-T}$$

DIPPR recommends an "Acceptable Model" that is uniform for VV:

$$Y = \frac{aT^b}{1 + (c/T) + (d/T^2)}$$

RMM modeling delivers three models that all show good performance. This is shown in Table 11.1, where both RMM_3 and RMM_4 are better than the TC and DIPPR models. Stability comparison shown in Table 11.2 indicates that the TC model is extremely nonstable, as evidenced by the large PRESS value. The DIPPR model similarly shows that the parameters' estimates are unstable (these are shown in *bold*). By contrast, RMM_3 is well-behaved, with respect to both measures of stability.

6.12. Vapor Thermal Conductivity (VTC). A plot of water VTC is given in Figure 13.

Modeling with TC on the original scale identifies six leading models with a cumulative *W* value of 80%. The leader among these is (eq 6401 in TC)

$$Y = a + bX + \frac{c}{X} + dX^2 + \frac{e}{T^2}$$

where

$$X = \log(T)$$

DIPPR recommends an "Acceptable Model" that is identical to that for VV (refer to Section 6.11). Goodness-of-fit statistics for the TC model, for the DIPPR "Acceptable Model", and for the three versions of RMM are shown in Table 12.1. RMM_3 is the leading model; however, all three RMM models and the

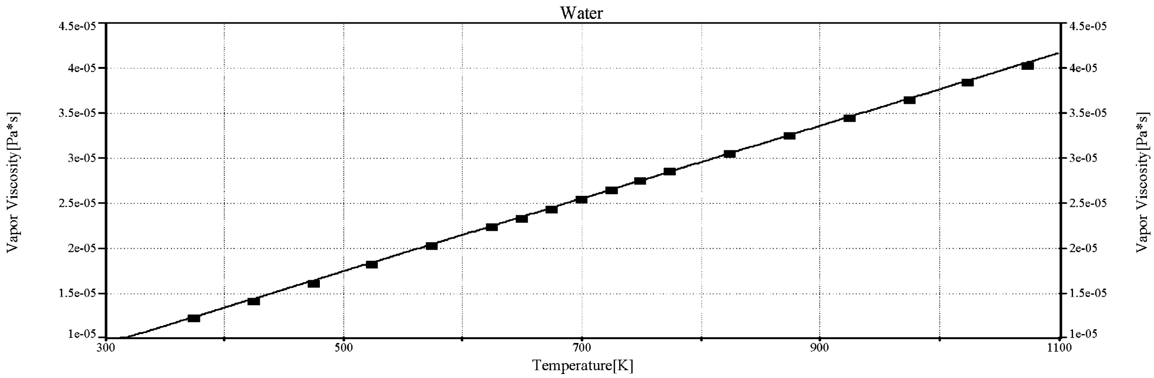

Figure 12. Data for water vapor viscosity (VV).

Table 11. Goodness-of-Fit and Stability Statistics for Water Vapor Viscosity (VV)

Table 11.1. Goodness-of-fit Statistics for Water Vapor Viscosity

model	number of parameters	normality (?)	MSE	AIC_c
RMM3	3	+	2.009×10^{-17}	−684.24
RMM4	4	+	1.175×10^{-16}	−649.76
TC#2099	4	+	1.934×10^{-15}	−599.35
DIPPR	4	+	2.444×10^{-15}	−595.13
RMM2	2	+	5.574×10^{-15}	−585.18

Table 11.2. Stability Statistics for Water Vapor Viscosity[a]

	TC		DIPPR		RMM3	
parameter	value	width of confidence interval	value	width of confidence interval	value	width of confidence interval
a	-3.872×10^{-6}	5.102×10^{-7}	**2.5228×10^{-6}**	**1.9711×10^{-4}**	−0.42644	0.019606
b	4.2645×10^{-8}	1.0446×10^{-9}	**0.505063**	**8.72236**	−0.51959	0.011679
c	-1.189×10^{-15}	5.746×10^{-16}	**1215.97**	**38404.9**	0.10278	4.809×10^{-3}
d	3.314×10^{155}	2.830×10^{155}	**−18293.8**	**886 960**		
PRESS	8.30514×10^{29}		8.68965×10^{-14}		7.4821×10^{-16}	

[a] Values given in boldface type represent possible instability.

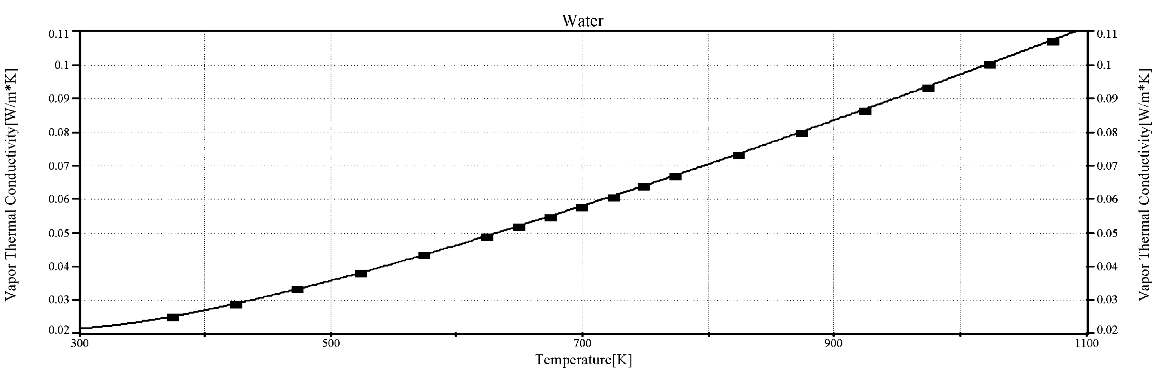

Figure 13. Data for water vapor thermal conductivity (VTC).

Table 12. Goodness-of-Fit and Stability Statistics for Water Vapor Thermal Conductivity (VTC)

Table 12.1. Goodness-of-Fit Statistics for Water Vapor Thermal Conductivity

model	number of parameters	normality (?)	MSE	AIC_c
TC#6401	5	+	2.258×10^{-11}	−427.47
RMM3	3	+	1.397×10^{-8}	−317.76
RMM4	4	+	1.785×10^{-8}	−310.67
DIPPR	4	+	2.976×10^{-8}	−298.17
RMM2	2	+	6.891×10^{-8}	−291.24

Table 12.2. Stability Statistics for Water Vapor Thermal Conductivity[a]

	TC		DIPPR		RMM3	
parameter	value	width of confidence interval	value	width of confidence interval	value	width of confidence interval
a	128.620	7.25260	213.729	117.269	−1.25758	1.08259
b	−14.5865	0.75117	−1.05397	0.188396	−0.29772	0.129142
c	−509.829	31.0920	−2702.94	271.399	**0.042788**	**0.156095**
d	0.628682	0.0291481	3.21769×10^{6}	1.2582×10^{6}		
e	765.545	49.9368				
PRESS	8.53741×10^{-10}		1.29641×10^{-6}		5.94577×10^{-7}	

[a] Values given in boldface type represent possible instability.

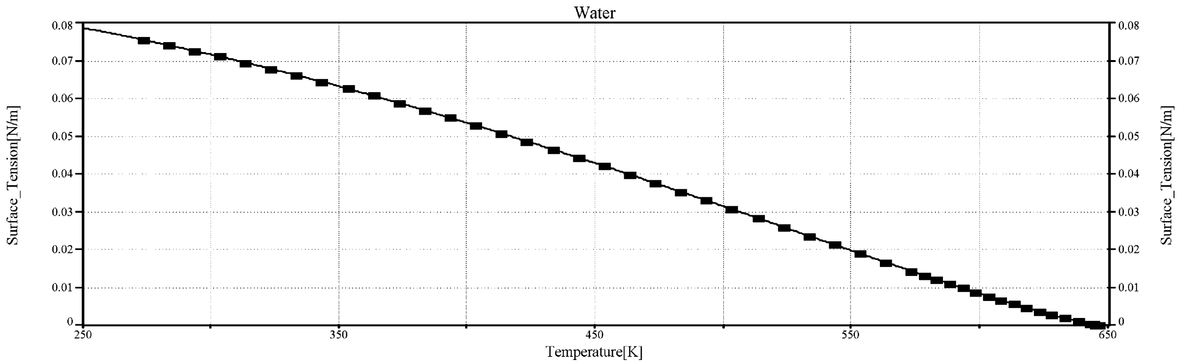

Figure 14. Data for water surface tension (ST).

Table 13. Goodness-of-Fit and Stability Statistics for Water Surface Tension (ST)

Table 13.1. Goodness-of-Fit Statistics for Water Surface Tension

model	number of parameters	normality (?)	MSE	AIC_c
RMM4	4	+	1.4288×10^{-9}	−970.34
DIPPR	4	+	7.3771×10^{-9}	−891.54
TC#8090	5	+	1.1321×10^{-8}	−869.47
RMM3	3	+	5.6686×10^{-7}	−684.56
RMM2	2	+	2.9670×10^{-6}	−606.44

Table 13.2. Stability Statistics for Water Surface Tension

	TC		DIPPR		RMM4	
parameter	value	width of confidence interval	value	width of confidence interval	value	width of confidence interval
a	0.130363	0.0128576	0.110659	7.3576×10^{-3}	1.39578	0.0184236
b	−0.146534	9.183×10^{-3}	0.775001	0.276907	−0.07216	0.0158593
c	382.557	44.4609	−0.560210	0.470395	−0.86369	0.0040016
d	68.5938	16.3014	0.848176	0.219517	0.256943	0.0170185
e	0.186669	0.0808549				
PRESS	0.00440225		4.11705×10^{-7}		7.33516×10^{-8}	

DIPPR model display goodness-of-fit statistics that are nearly equal. Table 12.2 shows that, in terms of PRESS, RMM$_3$ is more stable than the DIPPR model. However, one of its parameter estimates is unstable (given in *bold*). Parameter *d*, in the DIPPR model, is also bordering instability (in terms of the required relative confidence interval width, as defined in Section 4.2).

6.13. Surface Tension (ST). A plot of water ST is given in Figure 14.

One can realize that the graph displays a mirror of an S-shaped plot, being moderately concave for low-temperature values and convex for high values. Consequently, RMM may be anticipated to perform poorly (results are not compatible with this prediction).

Modeling with TC on the original scale identifies a single leading model, with a *W* value of 100% (namely, all other potential models do not fulfill the requirement for $\Delta < 10$). This is an asymmetric sigmoid (eq 8090 in TC):

$$Y = a + \frac{b}{\left\{1 + \exp\left[-\frac{T - d\log(2^{1/e} - 1) - c}{d}\right]\right\}^{e}}$$

The DIPPR-recommended “Acceptable Model” is uniform for ST, and is identical to that recommended for HV (Section 6.5). A four-parameter version of this model is proposed in DIPPR, and the critical temperature needed to fit this model is: T_c = 647.13 K.

Comparison of the five models of TC, DIPPR and RMM is given in Table 13.1, for goodness-of-fit, and in Table 13.2, for stability. With regard to both dimensions, RMM$_4$ outperforms all other models.

6.14. Liquid Thermal Conductivity (LTC). A plot of water LTC is given in Figure 15.

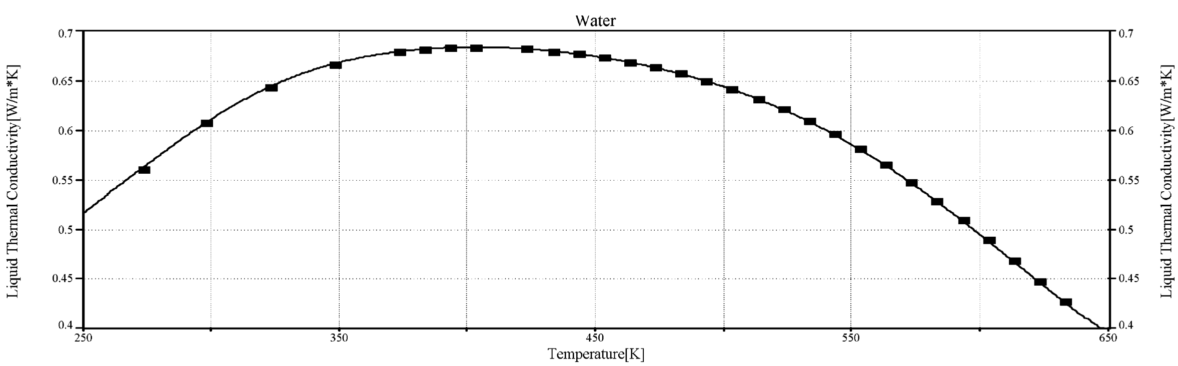

Figure 15. Data for water liquid thermal conductivity (LTC).

Table 14. Sample Sizes Used for All Four Substances

		Sample Size, *N*			
No.	property	oxygen	argon	nitrogen	water
1	solid density				11
2	liquid density	49		31	64
3	solid vapor pressure				22
4	vapor pressure	177	57	68	49
5	heat of vaporization	22	31	20	41
6	solid heat capacity				28
7	liquid heat capacity	24		11	16
8	ideal gas heat capacity	19		19	54
9	second virial coefficient	23	24	33	31
10	liquid viscosity	9	13	23	41
11	vapor viscosity	51	28	93	18
12	vapor thermal conductivity	40	28	40	18
13	surface tension				41
14	liquid thermal conductivity	49	8	23	
15	solid thermal conductivity		19	32	

Because *standard* application of RMM requires monotone convexity, no attempt was made to model this property. However, because, with some adaptation, RMM has been shown to model well S-shaped diffusion processes (refer to Shore and Benson-Karhi[9]), it is not ruled out that nonstandard application of RMM may ultimately result in models with properties comparable to those that are currently widely used.

6.15. Solid Thermal Conductivity (STC). No measured data were found.

7. Summary Comparisons (all Four Substances) and Some Practical Conclusions

7.1. Summary Comparisons (Models' Rankings). Table 14 displays sample sizes used for each property for all four substances. Table 15 displays the models' rankings, in terms of goodness-of-fit and stability, as obtained by TC, DIPPR, and RMM (with its three versions). This summary table relates to all properties and to all four substances (namely, oxygen, argon, nitrogen, and water). Entries use “1” to denote the best model. Values for PRESS are occasionally missing for TC because the large size of the model's parameters sometimes made PRESS numerically prohibitive to calculate. Property 8 is not monotone convex throughout the whole range. Although this property occasionally delivered good fit for RMM models, the lack of monotone convexity violated one of the basic assumptions of RMM. Property 8 was therefore excluded from the summary tables showing rankings across all properties (refer to the next two tables). Table 16 is a frequency table showing rankings, in terms of MSE and AIC_c (combined) for all substances across properties (property 8 was excluded for reasons just explained). Table 17 similarly displays frequencies for the stability measure, PRESS.

We realize that, on comparison of DIPPR's acceptable models to models derived via RMM (which is the comparison of relevance, as explained in the Introduction), RMM fares quite favorably. For example, on comparing the three approaches in terms of being best or second best, in terms of goodness-of-fit, TC appears 69 times, RMM 58 times, and DIPPR 25 times (Table 16). Similarly, in terms of stability, RMM appears 32 times, TC 24 times, and DIPPR 20 times (Table 17).

7.2. Some Practical Conclusions and Recommendations. This section describes some of the conclusions and recommendations determined during this experiment.

• For each of the properties modeled, exact values of the goodness-of-fit and stability statistics are given for each of the three approaches compared (occasionally, statistics for only the relevant DIPPR and RMM approaches are given). These should serve as guidelines for the selection of the best model, dependent on the most important model perspective that the user looks for (goodness-of-fit or stability).

• Observing models offered by TC, and the associated stability statistics, it seems that, in many cases, models offered by TC are overfitting the data, resulting in models with a number of parameters that is larger than required or justified, in view of the available data. Although the AIC_c statistic is supposed to guard against such overfitting, it is the research team's feeling that this measure is not sufficient for the intended purpose, and stability checking is always very much desired. Here, the statistic PRESS, rarely used in chemical engineering research (as inspection of related literature can attest), is very effective in identifying ill-conditioning.

To realize that, let us refer to modeling water vapor viscosity (property 11). Here, neither MSE nor AIC_c indicate ill-conditioning for the TC model (Table 11.1). However, the associated huge PRESS value (Table 11.2) obviously indicates ill-conditioning. In this case, only RMM_3 seems to indicate a stable model.

As another example for TC tendency to offer overfitted models with too many parameters, let us observe water liquid density (property 2). Here, TC offers a model with 16 parameters. Although stability statistics for this model are not shown, one would doubt that a model with such a high number of parameters is useful and can ever be considered a stable model.

• Stability statistics employed in this paper indicate, in the strict statistical sense, interpolation capability only. A routine practice to check for extrapolation capabilities of a model is to divide the data set into two unequal parts (e.g., $^2/_3$ and $^1/_3$), fitting with one set (for example, $^2/_3$ of the data) and then check for errors in the other set (data not participating in the fitting). Our adherence to high-quality data in this study caused some datasets to be too small for implementation of this practice. It is our experience that quite often, the stability indicators used in this study also serve as good indicators of the extrapolation capabilities of the model. Exercising caution in using a fitted model for extrapolation is, notwithstanding, always good advice.

• There is no justification to use RMM models where a linear relationship is suggested by the data. Because RMM_4 includes the linear model as a special case, it seems good practical advice that, if the RMM platform is selected for modeling, RMM_4 should first be used. If the parameters' estimates, together with the associated confidence intervals, point toward a linear relationship (for example, $b = d = 0$ is indicated), then obviously a linear relationship is suggested by the model, and simple linear regression should be used to derive estimates for the linear model. Similar practice should be followed if any of the other special cases of the general RMM model are indicated by the parameters' estimates and the respective confidence intervals.

• Because TC fits a different model to each property−substance combination, while DIPPR attempts to deliver a uniform model for each property, it is natural to expect that, in most cases, TC would deliver better goodness-of-fit. However, comparing DIPPR and RMM, the latter approach goes a step further (relative to the former) in that it offers uniform models, not only for each property (as DIPPR does) but also to all properties. As realized from Tables 15−17, the RMM general platform for modeling monotone convex relationships (and, perhaps, other types, such as S-shaped relationships) is effec-

P

Table 15. Ranking the Models According to the Goodness-of-Fit (MSE, AIC_c) and Stability (PRESS) Statistics[a]

Property	מדד	Oxygen			Argon			Nitrogen			Water		
		TC	DIPPR	RMM	TC	DIPPR	RMM	TC	DIPPR	RMM	TC	DIPPR	RMM
1	p	-	-	-	-	-	-	-	-	-	2	2	2
	MSE	-	-	-	-	-	-	-	-	-	1	1	1
	AIC_c	-	-	-	-	-	-	-	-	-	1	1	1
	PRESS	-	-	-	-	-	-	-	-	-	1	1	1
2	p	15	4	4	-	-	-	11	4	4	16	5	4
	MSE	1	3	2	-	-	-	1	3	2	1	2	3
	AIC_c	1	3	2	-	-	-	1	3	2	1	2	3
	PRESS	-	2	1	-	-	-	-	1	2	-	1	2
3	p	-	-	-	-	-	-	-	-	-	3	5	3
	MSE	-	-	-	-	-	-	-	-	-	1	2	3
	AIC_c	-	-	-	-	-	-	-	-	-	1	2	3
	PRESS	-	-	-	-	-	-	-	-	-	1	2	3
4	p	7	5	2	5	5	2	7	5	2	9	5	4
	MSE	1	3	2	1	3	2	1	3	2	1	3	2
	AIC_c	1	3	2	1	3	2	1	3	2	1	3	2
	PRESS	1	3	2	1	2	3	3	2	1	-	2	1
5	p	4	4	4	2	2	2	9	4	4	21	5	4
	MSE	1	2	3	1	3	2	2	1	3	1	2	3
	AIC_c	1	2	3	1	3	2	2	1	3	1	2	3
	PRESS	1	2	3	1	3	2	-	1	2	-	1	2
6	p	-	-	-	-	-	-	-	-	-	2	2	2
	MSE	-	-	-	-	-	-	-	-	-	1	1	1
	AIC_c	-	-	-	-	-	-	-	-	-	1	1	1
	PRESS	-	-	-	-	-	-	-	-	-	1	1	1
7	p	3	5	4	-	-	-	4	5	2	4	5	4
	MSE	2	3	1	-	-	-	1	3	2	3	2	1
	AIC_c	2	3	1	-	-	-	2	3	1	2	3	1
	PRESS	2	3	1	-	-	-	1	3	2	2	3	1
8	p	11	5	2	-	-	-	9	5	-	12	5	3
	MSE	1	2	3	-	-	-	2	1	-	1	2	3
	AIC_c	1	2	3	-	-	-	2	1	-	1	2	3
	PRESS	-	-	-	-	-	-	-	-	-	-	-	-
9	p	4	5	4	4	5	4	7	5	4	8	5	4
	MSE	2	3	1	2	3	1	1	3	2	1	3	2
	AIC_c	2	3	1	2	3	1	1	3	2	1	3	2
	PRESS	2	3	1	2	3	1	1	3	2	-	1	2
10	p	3	3	2	4	5	2	4	3	2	5	3	3
	MSE	1	2	3	1	3	2	1	3	2	1	3	2
	AIC_c	1	2	3	1	3	2	1	3	2	1	3	2
	PRESS	1	2	3	1	3	2	1	3	2	1	3	2
11	p	4	3	4	4	4	4	6	3	4	4	4	3
	MSE	1	2	3	1	2	3	1	3	2	2	3	1
	AIC_c	1	2	3	1	2	3	1	3	2	2	3	1
	PRESS	1	3	2	1	2	3	1	3	2	3	2	1
12	p	10	3	4	4	3	4	9	4	4	5	4	3
	MSE	1	3	2	1	3	2	1	3	2	1	3	2
	AIC_c	1	3	2	1	3	2	1	3	2	1	3	2
	PRESS	-	2	1	1	3	2	-	2	1	1	3	2
13	p	-	-	-	-	-	-	-	-	-	5	4	4
	MSE	-	-	-	-	-	-	-	-	-	3	2	1
	AIC_c	-	-	-	-	-	-	-	-	-	3	2	1
	PRESS	-	-	-	-	-	-	-	-	-	3	2	1
14	p	4	2	2	-	3	2	3	2	2	-	-	-
	MSE	1	3	2	-	2	1	1	3	2	-	-	-
	AIC_c	1	3	2	-	2	1	1	3	2	-	-	-
	PRESS	1	3	2	-	1	2	1	3	2	-	-	-

[a] RMM refers to any of its models.

Table 16. Frequency Table of the Model Rankings Across All Properties, Using the Goodness-of-Fit Statistics[a]

rank	Goodness of Fit[b] (MSE, AIC_c) "1"	"2"	"3"	missing	sum
Oxygen					
TC	14	4	0	8	26
DIPPR	0	6	12	8	26
RMM	4	8	6	8	26
Argon					
TC	10	2	0	14	26
DIPPR	0	4	10	12	26
RMM	4	8	2	12	26
Nitrogen					
TC	15	3	0	8	26
DIPPR	2	0	16	8	26
RMM	1	15	2	8	26
Water					
TC	18	3	3	2	26
DIPPR	4	9	11	2	26
RMM	10	8	6	2	26
Summary					
TC	57	12	3	32	104
DIPPR	6	19	49	30	104
RMM	19	39	16	30	104

[a] Property 8 (IGHC) has been excluded. [b] Goodness of fit is defined as MSE + AIC.

Table 17. Frequency Table of the Model Rankings Across All Properties, Using the Stability Statistics[a]

rank	Stability[b] (PRESS) "1"	"2"	"3"	missing	sum
Oxygen					
TC	5	2	0	6	13
DIPPR	0	4	5	4	13
RMM	4	3	2	4	13
Argon					
TC	5	1	0	7	13
DIPPR	1	2	4	6	13
RMM	1	4	2	6	13
Nitrogen					
TC	5	0	1	7	13
DIPPR	2	2	5	4	13
RMM	2	7	0	4	13
Water					
TC	5	1	2	5	13
DIPPR	5	4	3	1	13
RMM	6	5	1	1	13
Summary					
TC	20	4	3	25	52
DIPPR	8	12	17	15	52
RMM	13	19	5	15	52

[a] Property 8 (IGHC) has been excluded. [b] Stability is defined as PRESS.

tive in both goodness-of-fit and stability comparisons. The "uniformity of practice", offered by RMM, and the variety of special cases it offers (only a certain version, assuming a single error term, was used here) may qualify it to serve as a general-purpose platform for empirical modeling of chemical relationships.

• The uniformity of the RMM approach extends to its parameters' values. As realized from the RMM estimates that appear in the various tables in section 6, most values are confined to the {−2, 2} range. This stands in stark contrast to the parameters' estimates offered by the other approaches (TC and DIPPR). The practical advantage of this observation can hardly be overestimated. When a model is fit to the data, restricting the search routine to a uniform range for all parameters would almost surely guarantee an optimum global solution (that is, globally best estimates). Such an assertion cannot be extended to platforms, where a different model is used either for each property or for each substance−property combination.

8. Summary and Conclusions

Some results are introduced from a comprehensive research study, aimed at using RMM to model chemical relationships and comparing these to models obtained by other approaches. Two forthcoming papers by the same authors would detail the other results obtained in this study. One paper would show results from modeling the temperature dependence of three vapors (oxygen, argon, and nitrogen), similar to the results displayed in section 6 for water. The other paper would display results from using RMM to model temperature-dependent properties of pure compounds, using molecular descriptors. This modeling is conducted in the framework of structure−structure correlation (S−S−C).

In this paper, we have introduced a newly developed routine to select, in terms of goodness-of-fit, the best empirical model out of a set of given models. The routine was applied to select the best model from a set of models offered by TC. The selected model then served a baseline for further comparisons between models offered by DIPPR and by RMM. Two dependent criteria

Table A1.1. Water Properties: Response Data

property, Y	source[a]	Response Data: sample size, N[b]	T range (K)	Y range
1. solid density, SD (kmol/m^3)	B-1	11	233.15−273.15	51.203−50.888
2. liquid density, LD (kmol/m^3)	B-2	65	273.16−647.13	55.496−29.314
3. solid vapor pressure, SVP (Pa)	B-3	22	33.15−273.16	12.9−611.3
4. vapor pressure, VP (Pa)	B-4	51	273.16−647.13	6.1173×10^2−2.2055×10^7
5. heat of vaporization, HV (J/kmol)	B-5	41	273.16−645.65	4.5047×10^7−5.7450×10^6
6. solid heat capacity, SHC (J kmol^{-1} K^{-1})	B-6	28	3.1500−273.15	180.15−38120
7. liquid heat capacity, LHC (J kmol^{-1} K^{-1})	B-7	17	273.15−533.15	76168−89481
8. ideal gas heat capacity, IGHC (J kmol^{-1} K^{-1})	B-8	54	273.15−2273.15	33490−52803
9. second virial coefficient, SVC (m^3/kmol)	B-9	31	273.15−2273.15	(-1.7828×10^0)−9.5480×10^{-3}
10. liquid viscosity, LV (Pa s)	B-10	41	273.16−646.15	0.001792×10^0−4.801×10^{-5}
11. vapor viscosity, VV (Pa s)	B-11	18	373.15−1073.15	1.227×10^{-5}−4.037×10^{-5}
12. vapor thermal conductivity, VTC (W m^{-1} K^{-1})	B-12	18	373.15−1073.15	0.02508−0.1073
13. surface tension, ST	B-13	48	273.16−646.15	0.07565×10^0−7.000×10^{-6}
14. liquid thermal conductivity, LTC (W m^{-1} K^{-1})	B-14	34	373.15−633.15	0.5610−0.4272
15. solid thermal conductivity, STC	B-15		no measured data found	

[a] See Appendix A2. [b] N refers to the data set before the exclusion of observations. T denotes temperature, Y represents the response (property measured).

Table A1.2. Water Properties: RMM and TC Analysis Results

property, Y	source[a]	RMM Analysis: T median, M_T	RMM Analysis: Y median, M_Y	RMM Analysis: T STD, STD_T	TC Analysis: outliers
1. solid density, SD (kmol/m^3)	B-1	253.15	$M_{SD} = 51.047$	13.266499	none
2. liquid density, LD (kmol/m^3)	B-2	338.15	$M_{LD} = 54.424$	116.27069	$T = 647.13$ K (critical temperature)
3. solid vapor pressure, SVP (Pa)	B-3	254.15	$M_{SVP} = 114.35$	12.840042	none
4. vapor pressure, VP (Pa)	B-4	393.15	$M_{VP} = 198\,480$	110.62597	$T = 643.15$ K, 647.13 K (at or near the critical temperature)
5. heat of vaporization, HV (J/kmol)	B-5	463.15	$M_{HV} = 35\,638\,000$	118.15466	none
6. solid heat capacity, SHC (J kmol^{-1} K^{-1})	B-6	138.5	$M_{SHC} = 19\,150.5$	82.259751	none
7. liquid heat capacity, LHC (J kmol^{-1} K^{-1})	B-7	443.15	$M_{LHC} = 78\,708.5$	79.288503	$T = 273.15$ (near triple point)
8. ideal gas heat capacity, IGHC (J kmol^{-1} K^{-1})	B-8	703.15	$M_{IGHC} = 37\,535$	522.11622	none
9. second virial coefficient, SVC (m^3/kmol)	B-9	733.15	$M_{SVC} = -0.056928$	635.11217	none
10. liquid viscosity, LV (Pa s)	B-10	473.15	$M_{LV} = 1.344 \times 10^{-4}$	117.70502	none
11. vapor viscosity, VV (Pa s)	B-11	710.65	$M_{VV} = 2.600 \times 10^{-6}$	203.86883	none
12. vapor thermal conductivity, VTC (W m^{-1} K^{-1})	B-12	710.65	$M_{VTC} = 0.059255$	203.86883	none
13. surface tension, ST (N/m)	B-13	508.15	$M_{ST} = 0.029575$	119.22493	none
14. liquid thermal conductivity, LTC (W m^{-1} K^{-1})	B-14		no RMM modeling attempted		last four observations (Classified as "Acceptable" only)
15. solid thermal conductivity, STC	B-15				

[a] See Appendix A2. [b] T denotes temperature, Y represents the response (property measured).

Table A2. Data Sources for the Analyses of the 14 Properties of Water[a]

source code	reference/comment
B-1	*Steam Tables, S.I.*; New York, 1978
B-2	*NBS/NRC Steam Tables*, Hemisphere Publishing Corp.: Washington, DC, 1984
B-3	*Steam Tables S.I.*; Wiley: New York, 1978
B-4	*NBS/NRC Steam Tables*, Hemisphere Publishing Corp.: Washington, DC, 1984
B-5	*NBS/NRC Steam Tables*, Hemisphere Publishing Corp.: Washington, DC, 1984
B-6	*Engineering Sciences Data. Physical Data, Chemical Engineering*: Engineering Science Data: London, 1966
B-7	*NBS/NRC Steam Tables*, Hemisphere Publishing Corp.: Washington, DC, 1984
B-8	*NBS/NRC Steam Tables*, Hemisphere Publishing Corp.: Washington, DC, 1984
B-9	*NBS/NRC Steam Tables*, Hemisphere Publishing Corp.: Washington, DC, 1984
B-10	*J. Phys. Chem. Ref. Data*, 1986
B-11	*J. Phys. Chem. Ref. Data*, 1986
B-12	*J. Phys. Chem. Ref. Data*, 1986
B-13	*NBS/NRC Steam Tables*, Hemisphere Publishing Corp.: Washington, DC, 1984
B-14	*J. Phys. Chem. Ref. Data*, 1986
B-15	no measured data available

[a] For all data sources, the data quality classification is "Primary Accepted" (refer to DIPPR for further details).

were used for goodness-of-fit (MSE and AIC$_c$), and two criteria were used for stability (relative width of 95% confidence intervals of estimated parameters and PRESS). PRESS was found to be an effective measure for lack of stability, particularly for models with a large number of parameters, for which the goodness-of-fit statistic, AIC$_c$, was unable to detect overfitting.

The RMM models, their small number of parameters and their uniform models notwithstanding, seem to provide an effective tool for empirical modeling of temperature-dependent physical and thermodynamic properties.

Appendix

The water properties are detailed in Tables A1.1 and A1.2, whereas the data sources for the analyses of 14 properties of water are listed in Table A2. (No measured data for solid thermal conductivity (STC) was found, so no data are reported for this property.) The source codes given in Table A2 are used in Tables A1.1 and A1.2 to identify the source of the data.

Literature Cited

(1) Daubert, T. E. Evaluated Equation Forms for Correlating Thermodynamic and Transport Properties with Temperature. *Ind. Eng. Chem. Res.* **1998**, *37*, 3260−3267.

(2) Shore, H. *Response Modeling Methodology (RMM): Empirical Modeling for Engineering and Science*; World Scientific Publishing Co., Ltd.: Singapore, 2005.

(3) Shore, H.; Brauner, N.; Shacham, M. Modeling physical and thermodynamic properties via inverse normalizing transformations. *Ind. Eng. Chem. Res.* **2002**, *41*, 651−656.

(4) Shore, H. Response Modeling Methodology (RMM)—A new approach to model a chemo-response for a monotone convex/concave relationship. *Comput. Chem. Eng.* **2003**, *27* (5), 715−726.

(5) Shore, H. General control charts for variables. *Int. J. Prod. Res.* **2000**, *38* (8), 1875−1897.

(6) Shore, H. The Random Fatigue Life Model as a special case of the RMM model—A comment on Pascual (2003). *Commun. Stat.—Simul. Comput.* **2004**, *33* (2), 537−539.

(7) Ladany, S.; Shore, H. Profit Maximizing Warranty Period with Sales Expressed by a Demand Function. *Qual. Reliab. Eng. Int.* **2007**, *23* (3), 291−301.

(8) Shore, H. Response Modeling Methodology (RMM)—Validating Evidence from Engineering and the Sciences. *Qual. Reliab. Eng. Int.* **2004**, *20*, 61−79.

(9) Shore, H.; Benson-Karhi, D. Forecasting S-shaped diffusion processes via response modeling methodology. *Oper. Res.* **2006**, in press (available via Internet site).

(10) Shore, H. Response Modeling Methodology (RMM): Current distributions, transformations, and approximations as special cases of the RMM error distribution. *Commun. Stat.—Theory Methods* **2004**, *33* (7), 1491−1510.

(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, Brigham Young University, Provo, UT, 2006. (Available via the Internet at http//dippr.byu.edu.)

(12) Burnham, K. P.; Anderson, D. R. *Model Selection and Multimodel Inference*; Springer−Verlag: New York, 2004.

(13) Hurvich, C. M.; Tsai, C.-L. Regression and time series model selection in small samples. *Biometrika* **1989**, *76*, 297−307.

(14) Myers, R. H.; Montgomery, D. C.; Vining, G. G. *Generalized Linear Models: With Applications in Engineering and the Sciences*; Wiley: New York, 2002.

Received for review September 27, 2006
Revised manuscript received January 14, 2007
Accepted February 16, 2007

IE061252X