



# Correlating multicomponent physical property data with Padé approximants: Part II. Projecting pure component behaviour

Franco Pretorius<sup>3</sup>, Pethile Dzingai, Elizabeth L. du Toit<sup>2</sup>, Walter W. Focke<sup>\*,1</sup>

Department of Chemical Engineering, University of Pretoria, Private Bag X20, Hatfield 0028, Pretoria, South Africa

## ARTICLE INFO

### Keywords:

Density  
Viscosity  
Refractive index  
Surface tension  
Mixture model  
Padé approximant  
Ternary mixture  
Liquid

## ABSTRACT

Physical property data obtained for binary mixtures are conventionally fitted in excess form using the Redlich-Kister equation. The ternary data is then estimated using projections of the binary behaviour on the basis of geometric arguments. It is shown that a more consistent approach entails projection of pure component behaviour onto the simplex region. This generates Padé-type expressions capable of presenting multicomponent data in terms of the pure component temperature dependence using temperature-independent binary parameters. The effectiveness of this concept, for correlating density, viscosity, refractive index and surface tension, was confirmed using ten ternary systems for which data at either three or four different temperatures were available.

## 1. Introduction

The thermophysical properties of mixtures are important for many technological applications. In the chemical industries, knowledge of their variation with temperature, pressure and composition is required for the reliable design and effective operation of chemical processing plants. This information is also relevant for the formulation of chemical products, e.g. pharmaceuticals, cosmetics, transport fuels, etc. The effect of mixture composition on physical property variation can be accounted for by invoking suitable mixture models. While some have been derived on the basis of approximate theories, most are based on empirical expressions. There is ample information on the thermophysical properties of pure components, including their temperature dependence. Unfortunately, valid experimentally-determined mixture property data are rather sparse. While information for binary mixtures is frequently available, there is a dearth of data dealing with ternary and higher mixtures. Therefore, there is a need for procedures that may facilitate the prediction of multicomponent behaviour from binary data.

Sometimes the binary mixture behaviour is highly nonlinear. This may necessitate parameter-rich models in order to accurately represent the complicated composition dependencies. Hence, much research effort was expended, and continues to be spent, on the classic projective

geometric mixture models [1–7]. In these classical methods, attempts are made to express the thermophysical properties of a multicomponent system on the basis of its binary sub-systems [8]. The original impetus for this approach was to model the excess Gibbs free energy of multicomponent systems. However, this approach has also been implemented for correlating physical property data of ternary liquid mixtures [9–12,13–16]. The procedure entails first fitting the excess properties of the constituent binaries. This is usually achieved by applying the Redlich-Kister equation [17]. After obtaining acceptable fits, the excess property for the ternary system is estimated using geometric arguments. In essence, the binary behaviour is projected onto the ternary simplex using special weight functions, either symmetric or asymmetric in nature. A myriad of such projections have been proposed and they are discussed and compared in numerous past reviews [4,18–21] and also in more recent ones [8,22,23].

Unfortunately, there are serious problems associated with these models based on excess properties. Firstly, there is much confusion around the definition of excess properties for other physical properties besides Gibbs free energy. Brocos et al. [24] stated that they should only be used provided the reference state used to calculate this kind of deviation is a thermodynamically ideal mixture property. So, they argued that excess properties make sense for the molar volume  $V$  and the molar

\* Corresponding author.

E-mail address: [walter.focke@up.ac.za](mailto:walter.focke@up.ac.za) (W.W. Focke).

<sup>1</sup> ORCID: <https://orcid.org/0000-0002-8512-8948>.

<sup>2</sup> ORCID: <https://orcid.org/0000-0001-5579-1231>.

<sup>3</sup> ORCID: <https://orcid.org/0000-0002-1370-7158>.

refraction  $R$ , but that it does not make sense for the properties such as the refractive index. In addition, the underlying Redlich-Kister expressions already fail mathematical mixture model consistency tests [25]. For instance, Chou [21] pointed out that both the symmetrical and asymmetrical models have their own inherent consistency problems. The symmetric models do not reduce to binary systems when two out of the three components are identical. Quite disappointingly, the predictions of the asymmetrical models vary depending on how the component indices are assigned. According to Chou [21], this is an unsatisfactory state of affairs. Therefore, better approaches should be sought and developed.

Towards this goal, this communication describes a different way of generating projection-based mixture models. It is proposed that the projection should be based on the properties of the pure components instead of the excess binary behaviour. It is shown that this can lead to parameter-sparse expressions capable of accurately representing real experimental data obtained for ternary systems. The effect of mixture temperature, at least over limited ranges, is effectively incorporated via that of the pure components. Furthermore, the proposed equations do comply with mixture model mathematical consistency requirements [25].

## 2. Model development

Let the physical property value of a pure component labelled  $i$  be denoted by  $p_i$ . In general, it holds that  $p_i = p_i(T, P)$ , i.e. it depends of the system temperature and pressure. For many properties of condensed phases, like solids or liquids, the temperature dependence is important but the pressure dependence is often so weak that it can be neglected. Accepting this assumption, consider a single-phase liquid mixture of  $n$  different compounds at equilibrium at temperature  $T$ . The composition of the mixture is described by the vector of mole fractions:  $\mathbf{x}^T = (x_1, x_2, \dots, x_n)$ . It is postulated that the value which the physical property of this multicomponent mixture assumes, is determined by a weighted average over the pure component property values. The weighting function  $f_i(x_1, x_2, \dots, x_n)$  corresponds to a projection of the property values of pure component  $i$ , onto the simplex region using the following relation:

$$p(\mathbf{x}) = \sum f_i(\mathbf{x}) p_i \quad (1)$$

where  $f_i(\mathbf{x})$  is the projection function while  $p(\mathbf{x})$  and  $p_i$  represent the physical property values exhibited by the mixture and pure component  $i$  respectively. The projection function must satisfy the following consistency requirements:

$$(i) \lim_{x_i \rightarrow 0} f_i(\mathbf{x}) = 0 \quad \forall i \quad (2)$$

$$(ii) \lim_{x_i \rightarrow 1} f_i(\mathbf{x}) = 1 \quad \forall i \quad (3)$$

An assumption of the proposed formulation is that the parameters embedded in the weighting function  $f(\mathbf{x})$  are temperature and pressure independent. In this communication, projection functions are explored that either lead to modified Scheffé K-polynomials [26,27] or Padé-type approximants [28].

This pure component projection concept was inspired by the canonical Padé-type approximant expressions recently proposed for the surface tension of liquid mixtures [29]. They are based on generalisations, to multicomponent mixtures, of a series of empirical and theoretical models originally developed for describing binary mixtures [30–34]. Additionally, the previous analysis [29] incorporated the hypothesis, enunciated by Shardt and Elliott [35], that the temperature dependence of the mixture property arises from that of the pure components. A most general form, which includes all of these models as special cases, is the following  $P(3,3)$  Padé approximant defined by the ratio of two cubic polynomials:

$$\sigma = \frac{\sum_i \beta_i x_i \sigma_i \left( \sum_j \kappa_{ij} x_j \right) \left( \sum_k \lambda_{ik} x_k \right)}{\left[ \sum_i \beta_i x_i \left( \sum_j \tau_{ij} x_j \right) \left( \sum_k \omega_{ik} x_k \right) \right]} \quad (4)$$

where  $\sigma$  and  $\sigma_i$  are the surface tensions of the mixture and of pure component  $i$  respectively;  $\beta_i$  is a characteristic parameter of pure component  $i$ , while the  $\kappa_{ij}$ ,  $\lambda_{ik}$ ,  $\tau_{ij}$  and  $\omega_{ik}$  are adjustable binary parameters. Note that Eq. (4) conforms to the concept of a pure component property value projection, specified by Eq. (1), with the weighting function defined by

$$f_i(\mathbf{x}) = \beta_i x_i \left( \sum_j \kappa_{ij} x_j \right) \left( \sum_k \lambda_{ik} x_k \right) / \left[ \sum_l \beta_l x_l \left( \sum_j \tau_{lj} x_j \right) \left( \sum_k \omega_{lk} x_k \right) \right] \quad (5)$$

An important aspect of Eq. (4) is that it only features single component- and binary coefficients. The implication is that multicomponent behaviour should be predictable from knowledge of binary data. Indeed, Eq. (4) and various simplified versions were successfully applied to fitting binary and ternary surface tension data available over temperature ranges up to 30 °C wide [29]. Accurate representation of the mixture behaviour proved possible even with the adjustable model parameters assumed to be temperature independent. This meant that the temperature dependence was fully incorporated via the pure component properties as per the suggestion originally made by Shardt and Elliott [35]. In essence, this means that this mixture model served as a prescription for projecting the surface tension of the pure components onto the simplex region which defines the range of all possible mixture compositions. The success of this approach for surface tension data correlation is undeniable. The question arose whether this approach could also be applied fruitfully to deal with other physical properties. This communication is an attempt at providing an answer.

Unfortunately, Eq. (4) is rather parameter-rich. There is one parameter for each pure component and eight per binary in the mixture. Also, the pure component characteristics,  $\beta_i$ 's, are ill-defined in the sense that they are pseudo pure-component properties. It would be advantageous if they could be associated with readily measurable quantities representative of the individual pure components. Since parameter-sparse mixture models are preferred, the rest of the present discussion explores strategies aimed at reducing the number of adjustable parameters featuring in Eq. (4).

**Projection functions leading to Scheffé polynomials.** Scheffé-type polynomials are widely used for correlating mixture properties using experimental design [26]. The linear blending rule (LBR) corresponds to the first order Scheffé polynomial. It is generated by postulating  $f_i(\mathbf{x}) = x_i$  for the projection function:

$$p = \sum x_i p_i \quad (6)$$

The linear blending rule, shown in Eq. (6) for mole fractions as the composition descriptor, is a standard form for the molar volume and the molar refraction of ideal solutions [36]. Similarly, the density and Lorentz-Lorenz parameter of ideal solutions follow the LBR with mass fractions as the composition descriptors.

A special form of the quadratic Scheffé model is obtained with a projection function defined by:

$$f_i(\mathbf{x}) = x_i \sum \kappa_{ij} x_j \text{ with } \kappa_{ii} = 1 \quad \forall i \quad (7)$$

For a ternary mixture the corresponding expression for the mixture model is:

$$p = x_1 p_1 (x_1 + \kappa_{12} x_2 + \kappa_{13} x_3) + x_2 p_2 (\kappa_{21} x_1 + x_2 + \kappa_{23} x_3) + x_3 p_3 (\kappa_{31} x_1 + \kappa_{32} x_2 + x_3) \quad (8)$$

In Eq. (7) and Eq. (8), there are two independently adjustable

parameters per binary, i.e.  $\kappa_{ij}$  and  $\kappa_{ji}$ . This contrasts with the conventional second order Scheffé polynomial as it features only one adjustable parameter per binary [27]. A parameter reduction can be achieved if links between the  $\kappa_{ij}$  and the  $\kappa_{ji}$  can be established. Consider the postulate that the property values of a ternary mixture can be described by a quadratic expression defined by the product of two linear polynomials, i.e.

$$p = (a_1x_1 + a_2x_2 + a_3x_3)(b_1x_1 + b_2x_2 + b_3x_3) \quad (9)$$

This expression can be expanded as follows:

$$p = a_1b_1x_1 \left( x_1 + \frac{b_2}{b_1}x_2 + \frac{b_3}{b_1}x_3 \right) + a_2b_2x_2 \left( \frac{b_1}{b_2}x_1 + x_2 + \frac{b_3}{b_2}x_3 \right) + a_3b_3x_3 \left( \frac{b_1}{b_3}x_1 + \frac{b_2}{b_3}x_2 + x_3 \right) \quad (10)$$

Comparing Eq. (8) and Eq. (10) reveals that the pure component property corresponds to the product  $p_i = a_i b_i$ . Additionally, note that the model parameters are now linked via the relation  $\kappa_{ji} = 1/\kappa_{ij}$ . Furthermore, it is clear that the specification  $\kappa_{ii} = 1$  arises naturally. Applying these conditions to Eq. (8) yields a quadratic Scheffé polynomial with just one adjustable parameter per constituent binary.

A special form of the cubic Scheffé model is obtained by defining the projection function as follows:

$$f_i(\mathbf{x}) = x_i \sum \kappa_{ij} x_j \sum \lambda_{ij} x_j \quad (11)$$

Now links between the parameters of this equation can be established by postulating that the property of the mixture is defined by the product of three linear polynomials. Consider again a ternary mixture as the illustrating example:

$$p = (a_1x_1 + a_2x_2 + a_3x_3)(b_1x_1 + b_2x_2 + b_3x_3)(c_1x_1 + c_2x_2 + c_3x_3) \quad (12)$$

This expression can be expanded as follows:

$$p = a_1b_1c_1x_1 \left( x_1 + \frac{b_2}{b_1}x_2 + \frac{b_3}{b_1}x_3 \right) \left( x_1 + \frac{c_2}{c_1}x_2 + \frac{c_3}{c_1}x_3 \right) + a_2b_2c_2x_2 \left( \frac{b_1}{b_2}x_1 + x_2 + \frac{b_3}{b_2}x_3 \right) \left( \frac{c_1}{c_2}x_1 + x_2 + \frac{c_3}{c_2}x_3 \right) + a_3b_3c_3x_3 \left( \frac{b_1}{b_3}x_1 + \frac{b_2}{b_3}x_2 + x_3 \right) \left( \frac{c_1}{c_3}x_1 + \frac{c_2}{c_3}x_2 + x_3 \right) \quad (13)$$

In this case, the pure component properties are defined by the product  $p_i = a_i b_i c_i$  and it also holds that:

$$\kappa_{ji} = 1/\kappa_{ij}; \lambda_{ji} = 1/\lambda_{ij} \text{ and } \kappa_{ii} = \lambda_{ii} = 1 \quad (14)$$

Applying these relationships to the weighting function of Eq. (11) generates a special form of the cubic Scheffé polynomial with just two adjustable binary coefficients. In addition, unlike the conventional cubic Scheffé polynomials, the ternary constants that arises in multicomponent mixtures are fully determined in terms of the binary parameters.

**Projections generating Padé polynomials [28].** The proportional projection defined by Eq. (4) can be modified by postulating the following normalisation:

$$p(\mathbf{x}) = \sum f_i(\mathbf{x}) p_i / \sum f_i(\mathbf{x}) \quad (15)$$

In this case, the projection function must also satisfy the requirements stated in Eq. (2) but the one stated in Eq. (3) can now be relaxed and reformulated as follows:

$$\lim_{x_i \rightarrow 1} f_i(\mathbf{x}) = \beta_i \quad \forall i \quad (16)$$

Where  $\beta_i$ 's are considered to be unique pure component characteristics.

The normalised projection approach also generates the linear

blending rule when it is postulated that  $f_i(\mathbf{x}) = x_i$ , i.e. when it is assumed that  $\beta_i = 1 \quad \forall i$ :

$$p = \sum x_i p_i / \sum x_i = \sum x_i p_i \quad (6b)$$

This results from the simplex constraint which states that:

$$\sum x_i = 1 \quad (17)$$

However, this approach naturally leads to special Padé-type approximations. The lowest order Padé expression, i.e., the  $P(1,1)$  model, is obtained by assuming  $f_i(\mathbf{x}) = \beta_i x_i$ . For a ternary system the resulting expression is:

$$p = \frac{\beta_1 x_1 p_1 + \beta_2 x_2 p_2 + \beta_3 x_3 p_3}{\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3} \quad (18)$$

Eberhart [30] derived this Padé-type approximant for representing the surface tension of binary liquid mixtures. Note that Eq. (18) features pure component properties only. The implication is that multicomponent behaviour should be predictable from knowledge of pure component data provided the  $\beta_i$  values can somehow be ascertained. If this is not possible, binary mixture data are required to fix their values. The  $P(1,1)$  Padé approximant, defined in Eq. (18), has a particularly interesting property in this regard. For an  $n$ -component mixture it suffices to have data for all but one of the constituent binaries. In such a mixture there would be  $n(n-1)$  binaries but data for  $n(n-1)-1$  of the binary subsystems suffices in order to apply this model. The reason is that the model can be scaled using one of the parameters. Due to this scaling effect, the model actually features only  $n-1$  independent parameters. [Supplementary Information](#) includes a proof for ternary mixtures.

A general projection function, which leads to a modified  $P(2,2)$  Padé approximant, is defined by:

$$f_i(\mathbf{x}) = \beta_i x_i \sum_j \kappa_{ij} x_j / \sum_k \beta_k x_k \left( \sum_l \lambda_{kl} x_l \right) \quad (19)$$

with  $\kappa_{ii} = \lambda_{ii} = 1 \quad \forall i$ . The corresponding mixture model is:

$$p = \sum_i \beta_i x_i p_i \left( \sum_j \kappa_{ij} x_j \right) / \left[ \sum_i \beta_i x_i \left( \sum_j \lambda_{ij} x_j \right) \right] \quad (20)$$

This equation features one parameter per pure component and four for each binary in the mixture. There are two ways to reduce the latter to two per binary. The first way is to assume that Eq. (20) resulted from a normalisation approach. Then, it would hold that:

$$\lambda_{ij} = \kappa_{ij} \quad (21)$$

Alternatively, one could implement Eq. (14), i.e. letting  $\kappa_{ji} = 1/\kappa_{ij}$  and  $\lambda_{ji} = 1/\lambda_{ij}$ . The latter approach can also be used to construct a  $P(3,1)$  Padé expression with just two adjustable parameters per binary:

$$p = \sum_i \beta_i x_i \left( \sum_j \kappa_{ij} x_j \right) \left( \sum_k \lambda_{ik} x_k \right) / \sum_l \beta_l x_l \quad (22)$$

with  $\kappa_{ii} = \lambda_{ii} = 1 \quad \forall i$ .

As they stand, the  $\beta_i$ 's in Eq. (4) are "virtual" pure component indicators. It would be preferred if they could be associated with an actual measurable physical property, or at least suitable combinations of some physical properties, of the pure components present in the mixture. Clearly, this would result in a further reduction in the number of adjustable parameters necessary to fit experimental data. It turns out that such assignments of the  $\beta_i$ 's are indeed possible. First note that Eq. (18), 19, 20 and 22 are special cases of Eq. (4). Lower order variants are obtained by selectively setting some of the binary parameters to unity. The simplest form, Eq. (18), is obtained by setting all of them to unity. Together, these expressions form a nested set of equations. The lower order forms can be converted to higher order ones by multiplying, either

the numerator or denominator, with unity in the form of the simplex constraint described by Eq. (17). Eq. (18) is the simplest form devoid of all the binary parameters but it retains the pure component  $\beta_i$  coefficients. It turns out that standard LBR expressions for the properties of “ideal solutions” can be recast in a similar form, i.e. as  $P(1,1)$  Padé-type expressions. Comparing these two equivalent forms reveals the physical property of the pure component that can be associated with the  $\beta_i$  coefficients. This idea will now be explored in more detail.

**Density.** The specific volume ( $v$ ) is the inverse of the density ( $\rho$ ) and it can be expressed as follows:

$$v = 1/\rho = V/M \quad (23)$$

where  $V$  is the molar volume and  $M$  is the molar mass. For ideal solutions, the molar volume follows the linear blending rule:

$$V = \sum_i V_i x_i = \sum_i M_i x_i v_i \quad (24)$$

The linear blending rule applies exactly to the molar mass and therefore a  $P(1,1)$  Padé approximant for the specific volume is:

$$v = \frac{\sum_i M_i x_i v_i}{\sum_i M_i x_i} \quad (25)$$

Comparing Eq. (25) to Eq. (18) indicates that the molar mass  $M_i$  serves the purpose of the  $\beta_i$  characteristic of pure component  $i$  when the specific volume is the physical property of interest. Conveniently, the molar mass is independent of the mixture temperature.

There is an alternative way to model the mixture density of an ideal solution. The density is defined as the ratio of the average molar mass of the mixture divided by the molar volume of the solution, i.e.

$$\rho = \frac{M}{V} = \frac{V_1 x_1 \rho_1 + V_2 x_2 \rho_2}{V_1 x_1 + V_2 x_2} \quad (26)$$

Eq. (26) constitutes Kay's rule [37]. It states that the density of an ideal solution is just a volume fraction average over the densities of the components present. For a ternary mixture this LBR is given by:

$$\rho = \rho_1 \varphi_1 + \rho_2 \varphi_2 + \rho_3 \varphi_3 \quad (27)$$

Eq. (26) indicates that the pure component molar volume  $V_i$  uniquely represents the  $\beta_i$  value when the mixture density is the physical property of interest. Unlike the molar mass, the  $V_i$ 's are temperature dependent but this is not really a problem as the same holds true for density.

**Refractive index.** The Lorentz-Lorenz parameter is defined as [36]:

$$N = \frac{n^2 - 1}{n^2 + 2} \quad (28)$$

where  $n$  is the refractive index. The molar refraction is defined as:

$$R = NM/\rho = NV \quad (29)$$

For ideal solutions, both the molar refraction  $R$  and the molar volume  $V$ , follow the linear blending rule with mole fractions as the composition descriptors [24]. This means that, for ideal solutions the following expression holds:

$$N = \frac{R}{V} = \frac{R_1 x_1 + R_2 x_2}{V_1 x_1 + V_2 x_2} = \frac{V_1 x_1 N_1 + V_2 x_2 N_2}{V_1 x_1 + V_2 x_2} \quad (30)$$

Eq. (30) suggests that, for the physical property  $N$ , the  $\beta_i$ 's can be associated with the molar volumes ( $V_i$ ) of the pure components at the temperature of the solution. This means that, in the context of the present exposition, refractive index data is best correlated using the Lorentz-Lorenz parameter  $N$ . The refractive index is subsequently recovered via the expression:

$$n = \sqrt{\frac{1 + 2N}{1 - N}} \quad (31)$$

Alternatively, the refractive index of an ideal solution can be modelled via a  $P(1,1)$  Padé approximant in which mass fractions serve as the composition descriptors. If this is done, an expression is obtained in which the molar mass  $M_i$  serves as a temperature-independent  $\beta_i$  surrogate. To achieve this, first define a new property:

$$r \equiv N/\rho = R/M \quad (32)$$

The molar mass of the mixture is given by a linear combination of the molar masses of the components. For ideal solutions this also holds for the molar refraction. Therefore:

$$r = \frac{R_1 x_1 + R_2 x_2}{M_1 x_1 + M_2 x_2} = \frac{M_1 x_1 (N_1/\rho_1) + M_2 x_2 (N_2/\rho_2)}{M_1 x_1 + M_2 x_2} = \frac{M_1 x_1 r_1 + M_2 x_2 r_2}{M_1 x_1 + M_2 x_2} \quad (33)$$

In other words, for ideal solutions, it holds that the parameter  $r$  is a mass fraction weighted average over the pure component values:

$$r = r_1 w_1 + r_2 w_2 + r_3 w_3 \quad (34)$$

Eq. (30) and Eq. (33) are equivalent forms but they use different data during the regression analysis. Therefore, there will be differences in the prediction errors. Conveniently, Eq. (33) offers the advantage that all the parameters are temperature independent. However, the actual mixture density must be known in order to determine the value of the refractive index using:

$$n = \sqrt{\frac{1 + 2\rho r}{1 - \rho r}} \quad (35)$$

When using Eq. (30) instead, it is only necessary to know the molar masses and the densities of the pure components as a function of temperature.

**Surface tension.** The following expression holds when the vapour density is negligible in comparison to the liquid density [38]:

$$\sigma^{1/4} \approx \frac{\rho P}{M} = \frac{P}{V} \quad (36)$$

where  $\sigma$  is the surface tension,  $M$  is the molar mass,  $\rho$  is the liquid density,  $V$  is the liquid molar volume and  $P$  is the Parachor [39,40]. For nonpolar systems, the expectation is that both  $P$  and  $V$  vary linearly in the mole fractions [41]. For a ternary system this implies that:

$$\sigma^{1/4} \approx \frac{P_1 x_1 + P_2 x_2}{V_1 x_1 + V_2 x_2} = \frac{V_1 x_1 \sigma_1^{1/4} + V_2 x_2 \sigma_2^{1/4}}{V_1 x_1 + V_2 x_2} \quad (37)$$

Similar to the approach explored for the refractive index, a new parameter can be defined that facilitates an equivalent expression in which the molar mass  $M_i$  serves as a temperature-independent  $\beta_i$  surrogate:

$$s \equiv \sigma^{1/4}/\rho \quad (38)$$

This results in the following model for the property  $s$  for an “ideal” solution:

$$s = \frac{s_1 M_1 x_1 + s_2 M_2 x_2}{M_1 x_1 + M_2 x_2} \quad (39)$$

The comments made with respect to the refractive index models apply here too. Actual prediction of the surface tension on the basis of Eq. (39) requires knowledge of the density at the same composition of the liquid mixture under investigation.

**Viscosity.** Lederer [42] proposed the following mixture rule for the viscosity of binary mixtures:

$$\ln \eta = \ln \eta_1 + \frac{\alpha x_2}{x_1 + \alpha x_2} (\ln \eta_2 - \ln \eta_1) \quad (40)$$

This mixture rule for solutions can be recast into an equivalent  $P(1,1)$  Padé-type expression if one defines  $\alpha = \beta_2/\beta_1$ :



$$\ln \eta = \frac{\beta_1 x_1 \ln \eta_1 + \beta_2 x_2 \ln \eta_2}{\beta_1 x_1 + \beta_2 x_2} \quad (41)$$

Therefore, the Lederer [42] equation, for correlating the logarithm of the viscosity data, has a composition dependence similar to that of the Eberhart [30] model for surface tension, i.e. Eq. (18). Interestingly, Azian, et al. [43] proposed that the logarithm of viscosity follows a linear blending rule in terms of volume fractions:

$$\ln \eta = \varphi_1 \ln \eta_1 + \varphi_2 \ln \eta_2 = \frac{V_1 x_1 \ln \eta_1 + V_2 x_2 \ln \eta_2}{V_1 x_1 + V_2 x_2} \quad (42)$$

If this proposal is accepted as the base case, the  $\beta_i$  values are once again associated with the molar volumes  $V_i$ . Clearly, the latter do depend on temperature although only weakly. As done before for the refractive index and for the surface tension, a new viscosity parameter can be defined that will associate the molar mass of pure component  $i$ , i.e.  $M_i$ , with the characteristic  $\beta_i$  values. The required definition, in this case, is:

$$h \equiv \frac{1}{\rho} \ln \eta \quad (43)$$

### 3. Data analysis

The capacity of the canonical  $P(2,2)$  and  $P(3,1)$  Padé approximants, for accurately representing real experimental information, was tested using experimental data gleaned from the literature. Ten ternary systems, shown in Table 1, were used for this purpose. They were selected because they included binary and ternary data generated over a range of temperatures for the mixture density and also for some other properties, e.g. viscosity, refractive index or surface tension. This made it possible to test whether the temperature dependence of a mixture property can be accounted for by the pure component behaviour. This hypothesis was initially put forward by Shardt and Elliott [35] when they considered the surface tension of binary mixtures.

**Table 1**

Data sources used for testing the Padé-based models including the properties reported, the different temperatures applied and the number of binary and ternary data points.

#	Ternary system	Properties	T (°C) <sup>#</sup>	Data points <sup>*</sup>	Reference
I	ethanol, 2-butanone, and 2,2,4-trimethylpentane	$\rho, \eta$	25, <b>35</b> , 45	B: 171 T: 165	[10]
II	diisopropyl ether, ethanol, and methylcyclohexane	$\rho, \eta, n$	25, <b>35</b> , 45	B: 171 T: 55	[11]
III	acetone, ethanol, and 2,2,4-trimethylpentane	$\rho, \eta, n$	<b>15</b> , 25, 35	B: 171 T: 55	[45]
IV	ethanol, 2-methylpropan-2-ol, and 2,2,4-trimethylpentane	$\rho, \eta, n$	25, <b>35</b> , 45	B: 171 T: 55	[12]
V	diisopropyl ether, ethanol, and 2,2,4-trimethylpentane	$\rho, \eta, n$	15, <b>25</b> , 35	B: 171 T: 55	[13]
VI	tetrahydrofuran, 2-propanol, and 2,2,4-trimethylpentane	$\rho, \eta, n, \sigma$	15, <b>25</b> , 35	B: 171 T: 55	[46]
VII	2-propanol, benzyl alcohol, and 2-phenylethanol	$\rho, \eta, n, \sigma$	25, <b>35</b> , 45	B: 171 T: 55	[1447]
VIII	1,3-dioxolane, 2-propanol, and 2,2,4-trimethylpentane	$\rho, \eta, n, \sigma$	15, <b>25</b> , 35	B: 171 T: 55	[48]
IX	2-propanol, tetrahydropyran, and 2,2,4-trimethylpentane	$\rho, \eta, n, \sigma$	20, <b>30</b> , 40, 50	B: 228 T: 55	[15]
X	ethanol, benzyl acetate, and benzyl alcohol	$\rho, \eta, n, \sigma$	20, <b>30</b> , 40, 50	B: 228 T: 55	[16]

<sup>#</sup> The temperatures shown in bold represent those used to fit parameters using just one isothermal data set.

<sup>\*</sup> B: Number of distinct pure component and binary data values; T: Number of ternary data points.

Table 2 shows the different variations of the  $P(2,2)$  Padé projection model for ternary mixtures that were considered in the analysis. At this point it is necessary to address the issue of constraints on the values the parameters are allowed to take on. This issue is dealt with in detail in the [Supplementary Information](#). Only a short outline is presented here. Consider, as illustrating example, a binary mixture described by the  $P(2,2)$  Padé approximant:

$$p = \frac{p_1 \beta_1 x_1 (x_1 + \kappa_{12} x_2) + p_2 \beta_2 x_2 (\kappa_{21} x_1 + x_2)}{\beta_1 x_1 (x_1 + \lambda_{12} x_2) + \beta_2 x_2 (\lambda_{21} x_1 + x_2)} \quad (44)$$

In general, the  $\beta_i$  coefficients can be temperature-dependent. The equivalent Scheffé K-polynomial form, for an isothermal data set, is:

$$p = \frac{a_{11} x_1^2 + 2a_{12} x_1 x_2 + a_{22} x_2^2}{b_{11} x_1^2 + 2b_{12} x_1 x_2 + b_{22} x_2^2} = \frac{a(\mathbf{x})}{b(\mathbf{x})} \quad (45)$$

The links between the coefficients in these two equations are  $a_{ii} = p_i \beta_i$  and  $b_{ii} = \beta_i$  while

$$\begin{aligned} 2a_{12} &= \lambda_{12} \beta_1 p_1 + \lambda_{21} \beta_2 p_2 \\ 2b_{12} &= \lambda_{12} \beta_1 + \lambda_{21} \beta_2 \end{aligned} \quad (46)$$

The second order Scheffé polynomial is also equivalent to a quadratic form. If the mixture composition is defined by a column vector, the expression for multicomponent systems is given by:

$$b(\mathbf{x}) = \mathbf{x}^T \mathbf{B} \mathbf{x} \quad (47)$$

Here  $\mathbf{B}$  is the symmetric matrix of coefficients  $b_{ij}$ . In general, the binary parameters can be allowed to assume negative values. However, the denominator polynomial must always exceed zero in order to avoid prediction of negative or undefined mixture property values. This means that  $b(\mathbf{x})$  must be positive definite. This is guaranteed for all possible compositions if the determinant of the coefficient matrix exceeds zero, i. e.

$$\det \mathbf{B} > 0 \quad (48)$$

For a binary mixture this implies that

$$b_{12} > -\sqrt{b_{11} b_{22}} = -\sqrt{\beta_1 \beta_2} \quad (49)$$

For a ternary mixture, the explicit requirement is that

$$b_{11} b_{22} b_{33} + 2b_{12} b_{13} b_{23} - b_{11} b_{23}^2 - b_{22} b_{13}^2 - b_{33} b_{12}^2 > 0 \quad (50)$$

These constraints on the binary coefficients were imposed during the data regression analysis.

The parameter values were fixed using least squares regression analysis executed in Excel spreadsheets. Results obtained using only the binary data for this purpose, and also when the full data sets were considered, are reported. The Akaike information criterion (AIC) [44] was used to rank the relative performance of the models. The governing equation is:

$$AIC = n \ln (SSE) + 2df \quad (51)$$

where  $n$  is the number of data points,  $SSE$  is the sum of the square errors, and  $df$  is the degrees of freedom, i.e. the number of parameters that were fitted. To compare models, it is only the difference between the AIC values that matters. On taking differences, the units cancel out and the result is unitless:

$$\Delta AIC = n \ln (SSE_{ref} / SSE_{model}) + 2\Delta df \quad (52)$$

Equation (52) makes intuitive sense as it balances the change in goodness-of-fit, as assessed by the sum-of-squares, with the change in the degrees of freedom (due to differences in the number of parameters in the models that are being compared). In the present analysis the  $SSE$  values always included the contributions from all data points irrespective as to whether the parameters were determined using just the binary information or the full data sets. Note that the large number of

**Table 2**  
Pad -type projection model variants.

<i>P</i> (2,2) model for ternary mixtures: $p = \frac{\beta_1 x_1 p_1 (x_1 + \kappa_{12} x_2 + \kappa_{13} x_{23}) + \beta_2 x_2 p_2 (\kappa_{21} x_1 + x_2 + \kappa_{23} x_3) + \beta_3 x_3 p_3 (\kappa_{31} x_1 + \kappa_{32} x_2 + x_3)}{\beta_1 x_1 (x_1 + \lambda_{12} x_2 + \lambda_{13} x_{23}) + \beta_2 x_2 (\lambda_{21} x_1 + x_2 + \lambda_{23} x_3) + \beta_3 x_3 (\lambda_{31} x_1 + \lambda_{32} x_2 + x_3)}$		
Binary parameters		
Model variant code	Parameter constraints	# Parameters*
Ideal	$\lambda_{ij} = \kappa_{ij} = 1$	0
A	$\lambda_{ij} = \kappa_{ij}$ & $\kappa_{ji} = 1/\kappa_{ij}$	1
B	$\lambda_{ij} = \kappa_{ij}$	2
C	$\lambda_{ji} = 1/\lambda_{ij}$ & $\kappa_{ji} = 1/\kappa_{ij}$	2
D	None	4
<i>P</i> (3,1) model: $p = \sum_i \beta_i x_i \sum_j \kappa_{ij} x_j \sum_k \lambda_{ik} x_k / \sum_i \beta_i x_i$		
Binary parameters		
Model variant code	Parameter constraints	# Parameters*
E	$\lambda_{ji} = 1/\lambda_{ij}$ & $\kappa_{ji} = 1/\kappa_{ij}$	2
F	None	4
Pure component descriptors		
Physical property	$p_i$ for $\beta_i \equiv V_i$	$p_i$ for $\beta_i \equiv M_i$
Density	$\rho_i$	$v_i = 1/\rho_i$
Viscosity	$\ln \eta_i$	$v_i \ln \eta_i$
Refractive index	$N_i = (\eta_i^2 - 1)/(\eta_i^2 + 2)$	$N_i = v_i (\eta_i^2 - 1)/(\eta_i^2 + 2)$
Surface tension	$\sigma_i^{1/4}$	$v_i \sigma_i^{1/4}$

\* The number of adjustable parameters per binary in the mixture.

data points in the present data sets enhanced the importance of the first term on the right of Eq. (51) at the expense of the second term, which corrects for the increase in the number of model parameters.

The cases employing  $\beta_i = V_i$  and  $\beta_i = M_i$  were considered separately. For each mixture property, the corresponding *P*(1,1) Pad  “ideal solution” expression served as the reference model. The obtained  $\Delta AIC$  values were normalised by taking the ratio with respect to the largest value recorded. Invariably, this was the  $\Delta AIC$  value obtained when the full set of parameters were allowed to vary freely, i.e. model variation D, although less frequently it was model F. These two model variants are the most flexible ones as they feature more adjustable parameters than the others. The normalised  $\Delta AIC$  values varied from zero for the *P*(1,1) Pad  reference equations, expected for ideal solutions, to unity for the best performing model variant. Intermediate values provide relative measures of model performance. When negative values were obtained for a particular model it indicates that it performed worse than the fully predictive “ideal” solution expression. The normalised  $\Delta AIC$  values, obtained when regressing just the binary data, were compared to the values obtained when the full data set were used instead. Often, the value obtained with the former protocol was only slightly less than what is achieved in the latter case. This provided an indication that the ternary response can reliably be predicted from knowledge of binary data. Comprehensive details, including the numerical values assumed by the model parameters, are presented in the [Supplementary Information](#). The Excel spreadsheets are available via the following link:

<https://drive.google.com/drive/folders/1SrCYFHnRLXOENU1sStUl k3fl2fCUySaN?usp=sharing>.

4. Results

Fig. 1 shows a summary of the Akaike information results in a way that facilitates visual comparison of performance across all of the physical properties considered. The plotted values, shown for each physical property, represent the results averaged over all ten ternary data sets. The outcomes obtained for the models employing  $\beta_i = V_i$  and  $\beta_i = M_i$  are presented separately. Better data correlation performance is expected for models with a greater number of adjustable parameters.

This was clearly the case when the regression was performed on the full data sets. Additionally, the variability in the scaled  $\Delta AIC$  values was reduced compared to the situation when only the binary data were considered. However, this was not always the case when the regression was limited to the binary data. Striking examples are provided by the predictions obtained for model F for density in Fig. 1(b) and for surface tension in Fig. 1(h). In these cases, the model fit to the binary data resulted in predictions that were worse than those obtained using the fully predictive ideal expressions. This is indicated by the fact that negative values were obtained for the normalised  $\Delta AIC$  values. Model variant F also performed poorly for refractive index when the regression was limited to the binary data as seen in Fig. 1(f). Interestingly, better performance of model variant F was observed when the pure component molar volumes served as the surrogates for the  $\beta$ -values. A possible explanation for this failure, when the molar mass was used, is as follows. Firstly, overfitting of the binary data must have played a role as model variant E, with fewer adjustable coefficients, performed better for the density and the refractive index. Secondly, with  $M_i$  serving as the  $\beta_i$  parameters, small errors in the measured mixture density could have been amplified into larger errors in the estimates made for viscosity and surface tension. This is because, in the final calculation of viscosity and surface tension, these quantities were obtained by multiplying with the measured density exponentiated and raised to the fourth power respectively. This, combined with the overfitting of the binary data, are the likely causes of the poor performance observed.

Fig. 2 reports the AAD values averaged over the ten data sets for the different model variants and the different choices for the  $\beta$ -parameters. The  $\Delta AIC$  values are based on a statistic affected by the sum of the square errors. In contrast, the AAD values represent absolute deviations between predictions and actual measurements. A proper assessment of the utility of the various model variants should consider both statistics. Fig. 2 shows that, on average, all model variants performed better than the fully predictive “ideal” solution expressions. The exceptions are model E and model F where regression of binary data often resulted in worse AAD values. Additionally, these two model variants were clearly unable to adequately correlate the surface tension data.

**Density.** Fig. 1(a) and Fig. 1(b) show the normalised  $\Delta AIC$  results

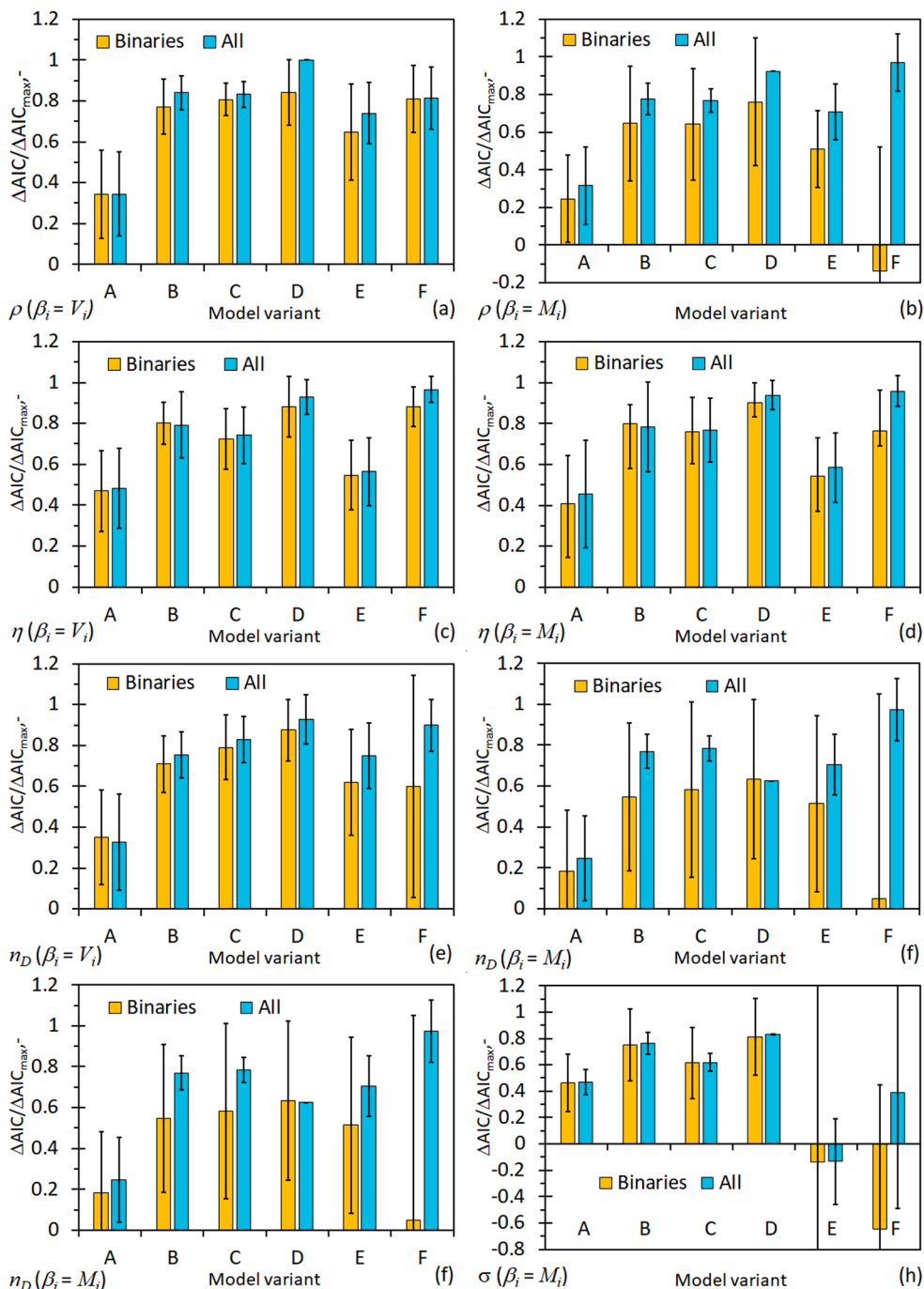


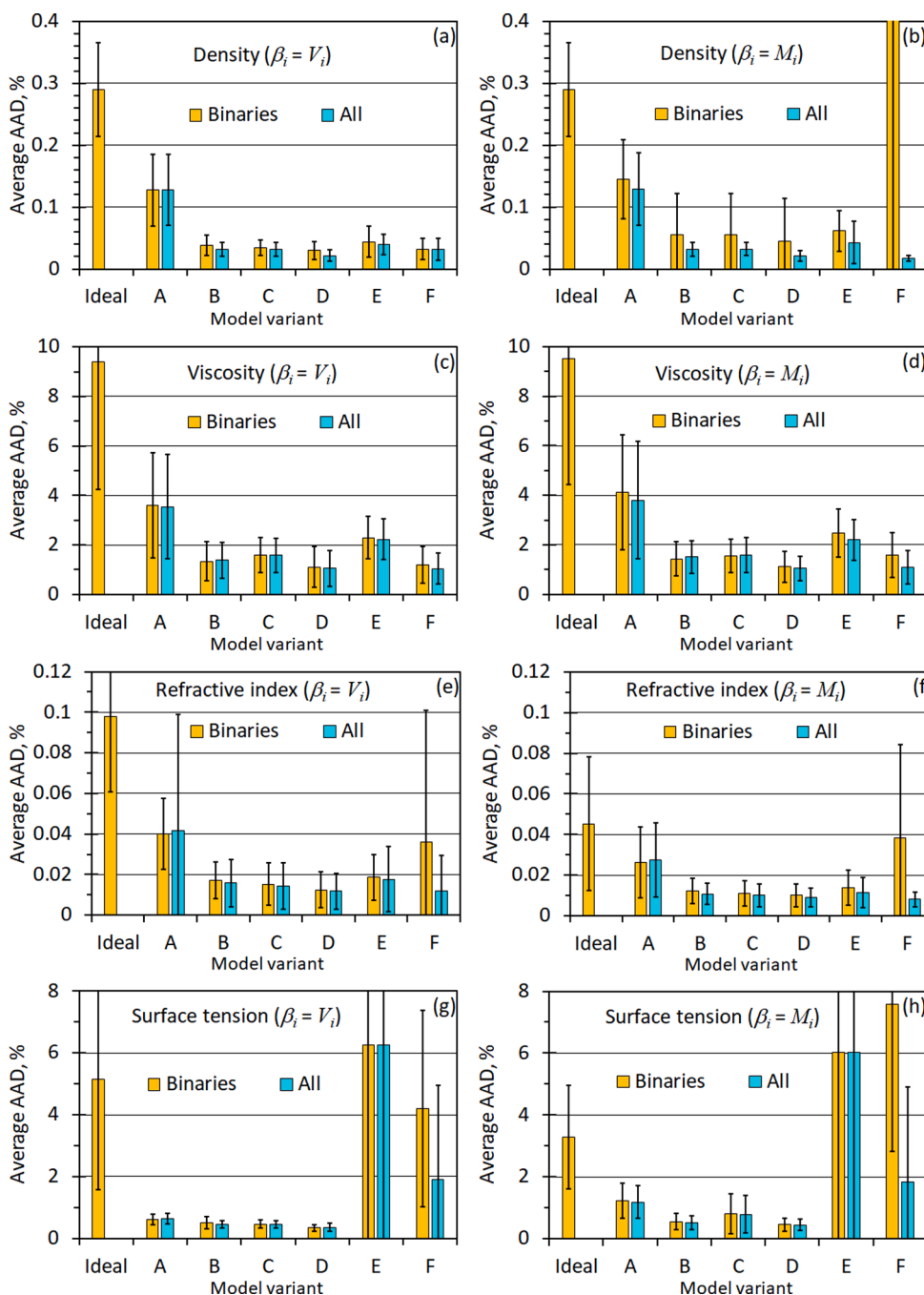
Fig. 1. Scaled  $\Delta AIC$  values for the six model variants averaged over the ten ternary systems (listed in Table 1) for the four different physical properties.

obtained for density with the choices  $\beta_i = V_i$  and  $\beta_i = M_i$  respectively. The corresponding AAD values are shown in Fig. 2(a) and Fig. 2(b). In general, the  $\Delta AIC$  values increased with the number of adjustable parameters in the model variants. The performance of model variants B and C were roughly on par. For model F, using the molar volume of the pure components for the  $\beta$ -values appears to be a better option than molar mass. This was the case when attention was focussed on the regression of the binary data only. Otherwise, there was little difference in performance for density as the physical property. Considering regression analysis performed on the full data sets, for  $\beta_i = V_i$  model D performed best while for  $\beta_i = M_i$  it was actually model F. The corresponding AAD values for the latter were  $0.021 \pm 0.009\%$  and  $0.017 \pm 0.004\%$ .

**Viscosity.** Fig. 1(c) and Fig. 1(d) show the normalised  $\Delta AIC$  results

obtained for viscosity with the choices  $\beta_i = V_i$  and  $\beta_i = M_i$  respectively. The corresponding AAD values are shown in Fig. 2(c) and Fig. 2(d). The problems experienced with model F for correlating the density data were not found when dealing with the viscosity data. Models B and C performed similarly. The viability of predicting ternary data on the basis of binary results was confirmed. The prediction errors were only marginally higher compared to those obtained when the full data sets were regressed. Considering regression analysis performed on the full data sets, for  $\beta_i = V_i$  model F performed best while for  $\beta_i = M_i$  it was model variant F. The corresponding AAD values for the latter were  $1.04 \pm 0.63\%$  and  $1.05 \pm 0.49\%$ .

**Refractive index.** Fig. 1(e) and Fig. 1(f) show the normalised  $\Delta AIC$  results obtained for the refractive index with the choices  $\beta_i = V_i$  and  $\beta_i = M_i$  respectively. The corresponding AAD values are shown in Fig. 2(e)



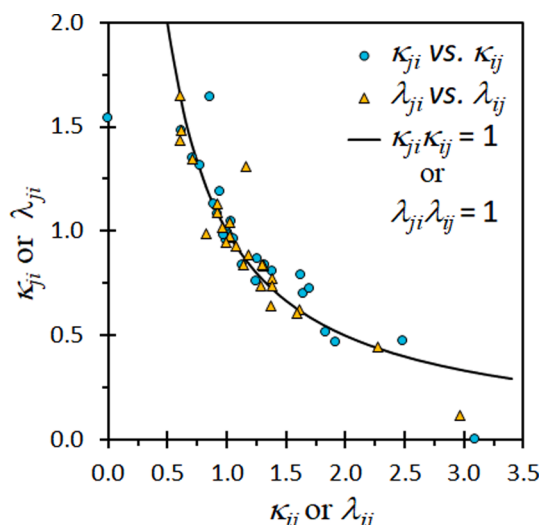
**Fig. 2.** Comparing the average absolute deviations (AAD) values achieved with the six model variants for the four physical properties in the data sets for the ten ternary systems listed in Table 1.

and Fig. 2(f). Once again, Model F performed poorly when the data regression was limited to the binaries. However, the predictions were, on average, slightly better than those obtained using the ideal model forms. Model variant D performed best when using  $\beta_i = V_i$  with AAD equal to  $0.0124 \pm 0.0088$  % when the full data set was regressed. However, applying model variant F and setting  $\beta_i = M_i$  resulted in a lower AAD value of  $0.0080 \pm 0.0036$  %. For refractive index, model variant C clearly outperformed model variant B irrespective of the choice made for the  $\beta$ -value. The AAD value achieved using this variant was  $0.0143 \pm 0.0093$  %  $\beta_i = V_i$  and  $0.0100 \pm 0.0056$  % for  $\beta_i = M_i$ . In other words, its performance compared very favourably with the best performing models which had twice the number of adjustable coefficients. It was therefore of interest whether the constraints imposed

on the parameters in model variant C are inherent to the data sets under study. Therefore, the model coefficients obtained using the full data sets were plotted against each other. The results, for the case where no restrictions were imposed on the binary coefficients except that they assume positive values, are shown in Fig. 3. Interestingly, with few exceptions, the data points fall on the line describing the parameter constraint imposed in model variant C.

**Surface tension.** Fig. 1(g) and Fig. 1(h) show the normalised  $\Delta AIC$  results obtained for the surface tension with the choices  $\beta_i = V_i$  and  $\beta_i = M_i$  respectively. The corresponding AAD values are shown in Fig. 2(g) and Fig. 2(h). For this physical property, the choice of  $\beta_i = V_i$  is clearly better than the alternative. All model variants, except variants E and F, performed well in reducing the prediction error. This was the case



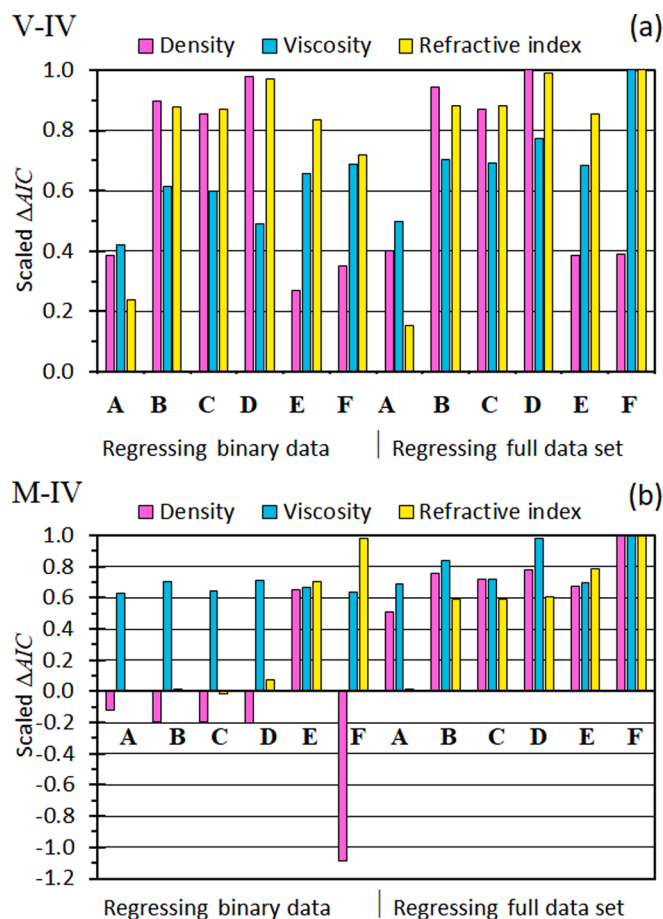


**Fig. 3.** Plots of  $\kappa_{ji}$  vs.  $\kappa_{ij}$  and  $\lambda_{ji}$  vs.  $\lambda_{ij}$ . The parameters were obtained on regressing the full data sets using model variant D with  $\beta_i = M_i$ , and forcing all the binary coefficients to be positive.

irrespective as to whether just the binary, or the full data set was regressed. This means that the assumption that ternary data can be predicted from binary results holds. Model variant D performed best with the AAD value of  $0.36 \pm 0.13$  % for  $\beta_i = V_i$  and  $0.44 \pm 0.18$  % for  $\beta_i = M_i$ .

More detailed plots, showing how the scaled  $\Delta AIC$  values differ with model variant and physical property under consideration per ternary system, are presented in the [Supplementary Information](#). Interestingly, negative  $\Delta AIC$  values were occasionally recorded for some properties when the regression was limited to the binary data. The most negative values were recorded for density as the property in System IV: Ethanol (1), 2-methylpropan-2-ol (2), 2,2,4-trimethylpentane (3). The scaled  $\Delta AIC$  values for this system are shown in [Fig. 4](#). Negative  $\Delta AIC$  values were obtained for all model variants (except model variant E) when it held that  $\beta_i = M_i$  and the regression analysis was solely based on the binary data. This was not the case when the full data set were regressed. For the molar mass serving as  $\beta$ -value surrogate, the  $\Delta AIC$  values for the refractive index predicted on the basis of the binary data, were also very low and even negative for model variant C. Once again, this was not the case when the molar volumes of the pure components were used instead. This indicates that the models considered presently might be unable to properly account for one or more of the binary data trends. Regressing the full data set with model variant D, led to excellent data fits for density for the ethanol (1) + 2,2,4-trimethylpentane (3) and the 2-methylpropan-2-ol (2) + 2,2,4-trimethylpentane (3) binaries. However, the predictions for the ethanol (1) + 2-methylpropan-2-ol (3) binary were poor as shown in [Fig. 5](#). The composition dependence of density for this binary, although fairly weak, shows an S-type variation with composition. The  $P(2,2)$  Padé models appear unable to represent such behaviour. The conclusion is that this was the root cause of the poor predictions obtained for density and refractive index when the regression was limited to the binary data. Note however, that model F was effective at representing the density data as seen in [Fig. 5\(b\)](#) when the ternary data were included in the regression exercise.

These results indicate that the proposed model, with its various variants, is able to represent data trends as a function of composition and, to some extent, also over narrow temperature ranges. Compared to previous approaches, the current model proposals are, relatively speaking, parameter-sparse. System X, comprising binary and ternary mixtures of ethanol, benzyl acetate, and benzyl alcohol at four different temperatures, provides an illustrating example. The authors [16] performed the data analysis using the Cibulka [7] method. First, they



**Fig. 4.** Scaled  $\Delta AIC$  values for System IV: Ethanol, 2-methylpropan-2-ol, and 2,2,4-trimethylpentane. (a) Molar volume and (b) molar mass used for the  $\beta$ -parameters.

defined the excess properties of the property  $p$  for each binary subsystem:

$$\Delta p_{ij} = p - x_i p_i - x_j p_j \quad (53)$$

The corresponding excess property of the ternary mixture is defined as:

$$\Delta p_{123} = p - x_1 p_1 - x_2 p_2 - x_3 p_3 \quad (54)$$

Next, they correlated the excess data with a Redlich-Kister equation for each of the three binaries at each of the four temperatures:

$$\Delta p_{ij} = x_i x_j \sum_{k=0}^{k=d} a_k (x_i - x_j)^k \quad (55)$$

Finally, the data for the ternary mixtures was separately correlated using the Cibulka [7] equation:

$$\Delta p_{123} = \Delta p_{12} + \Delta p_{13} + \Delta p_{23} + x_1 x_2 x_3 (C_0 + C_1 x_1 + C_2 x_3) \quad (56)$$

This was only done at the one temperature for which ternary data were generated. This means that a total of seventeen adjustable constants were deemed necessary to represent the binary and ternary data obtained at just one temperature. Furthermore, with this approach there is no way to predict the ternary behaviour at other temperatures. This is even true for temperatures for which binary data were collected. Also, it is not known how this approach can be extended to multicomponent systems with more than three components.

By comparison, the present approach using the highest order models, e.g. either model variant D or variant F, only features twelve adjustable model coefficients. Furthermore, in some cases lower order model

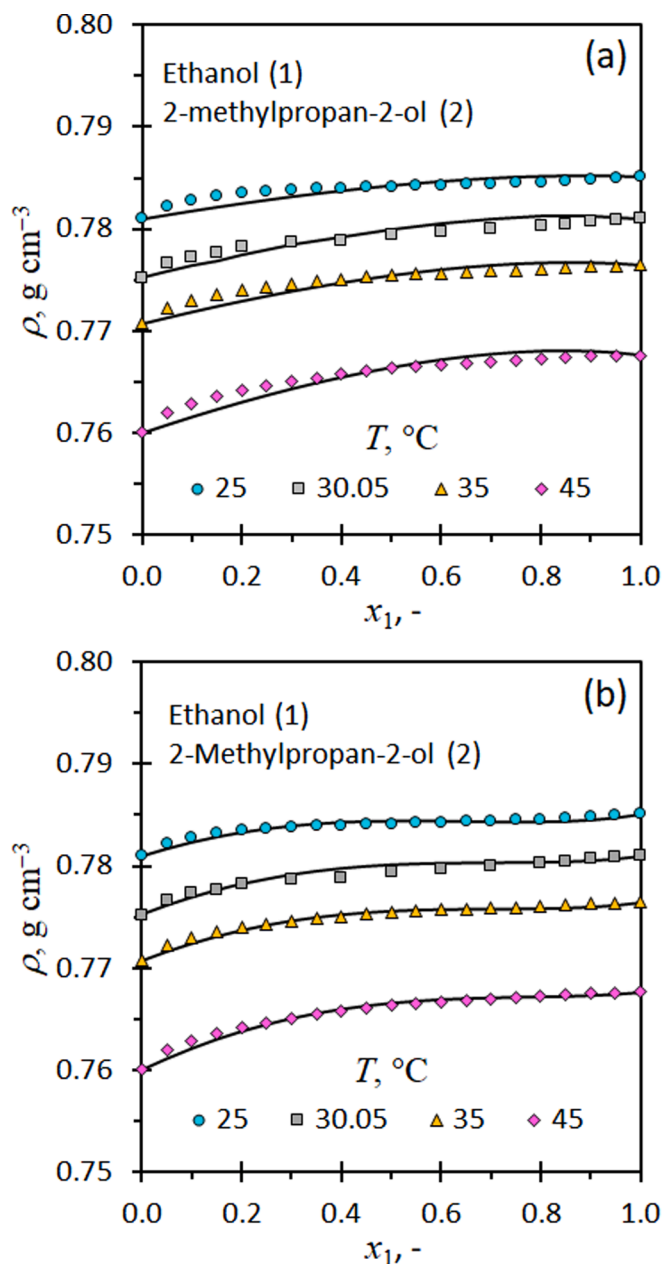


Fig. 5. Density data for System IV: ethanol (1) + 2-methylpropan-2-ol (2) binary. The solid lines show global fits obtained using (a) model variant D, and (b) model variant F with  $\beta_i = M_i$  in both cases.

variants, with less than half the number of free parameters, provided acceptable data representation over the full composition range and at all four temperatures. It even proved possible to predict ternary behaviour on the basis of just binary information. Clearly, this is not possible with the Cibulka [7] approach. Additionally, the current formulation allows facile extension of the equations to higher orders in order to cater for more complex composition dependencies in multicomponent mixtures.

For the data sets considered presently, the  $P(2,2)$  Padé forms, even those with just two-parameters per binary - e.g. model variants B and C - usually gave good data trend representations. The present formulation was specifically developed for correlating data gathered at more than one temperature. However, for special situations these latter two models can be fitted to an isothermal data set and the results can be extrapolated in order to predict performance at other, near-by, temperatures. This possibility is now illustrated for a binary mixture. The  $P(2,2)$  Padé model for this case can be cast in two equivalent forms as shown in equations

(44) and Eq. (45) with the link between the parameters given by Eq. (46). For the Model B variant, the parameter constraint is  $\lambda_{ij} = \kappa_{ij}$ . Solving for the cross-parameters yields:

$$\kappa_{12} = 2 \frac{a_{12} - p_2 b_{12}}{\beta_1 (p_1 - p_2)} \kappa_{21} = 2 \frac{a_{12} - p_1 b_{12}}{\beta_2 (p_2 - p_1)} \quad (57)$$

These two equations show that it is possible to extract the values for the binary coefficients  $\kappa_{12}$  and  $\kappa_{21}$  from an isothermal data set. Note though, that problems can be anticipated when the difference in the pure component property values, i.e.  $p_2 - p_1$ , is small. If this is not the case, the implication is that the isothermal data can be used to predict behaviour at other temperatures that are close to the one at which the data was collected.

For the Model C variant, the parameter constraint is  $\kappa_{ji} = 1/\kappa_{ij}$  and  $\lambda_{ji} = 1/\lambda_{ij}$ . Solving for the cross-parameters yields two possible solutions, both of which are valid for representing the composition behaviour:

$$\lambda_{12} = \frac{b_{12}}{\beta_1} \left( 1 \pm \sqrt{1 - \beta_1 \beta_2 / b_{12}^2} \right) \lambda_{21} = \frac{1}{\lambda_{12}} = \frac{b_{12}}{\beta_2} \left( 1 \mp \sqrt{1 - \beta_1 \beta_2 / b_{12}^2} \right) \quad (58)$$

$$\kappa_{12} = \frac{a_{12}}{\beta_1 p_1} \left( 1 \pm \sqrt{1 - \beta_1 \beta_2 p_1 p_2 / a_{12}^2} \right) \quad (59a)$$

And

$$\kappa_{21} = \frac{1}{\kappa_{12}} = \frac{a_{12}}{\beta_2 p_2} \left( 1 \mp \sqrt{1 - \beta_1 \beta_2 p_1 p_2 / a_{12}^2} \right) \quad (59b)$$

If the two solutions each for the  $\kappa_{ij}$  and the  $\lambda_{ij}$  are labelled using the hash and star symbols, the following relationships between them hold:

$$\kappa_{12}^{\#} \kappa_{12}^* = \frac{\beta_2 p_2}{\beta_1 p_1} \quad \text{and} \quad \kappa_{21}^{\#} \kappa_{21}^* = \frac{\beta_1 p_1}{\beta_2 p_2} \quad (60)$$

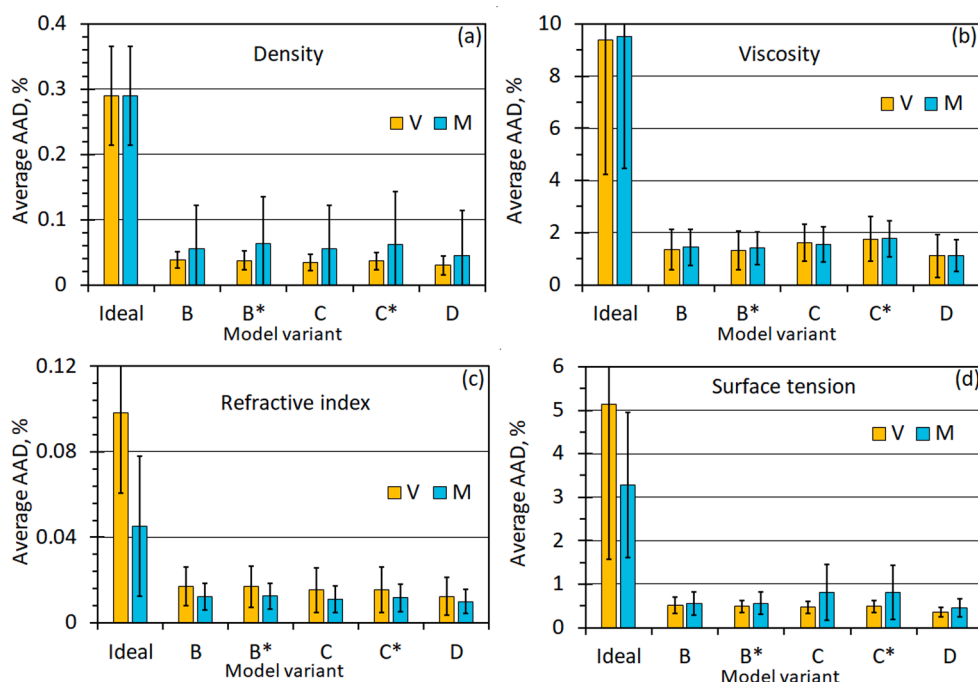
$$\lambda_{12}^{\#} \lambda_{12}^* = \frac{\beta_2}{\beta_1} \quad \text{and} \quad \lambda_{21}^{\#} \lambda_{21}^* = \frac{\beta_1}{\beta_2} \quad (61)$$

(NB. Note that the hash (#) labelled parameters correspond to the addition of the square root term and the star (\*) labelled parameters correspond to the subtraction of the square root term.)

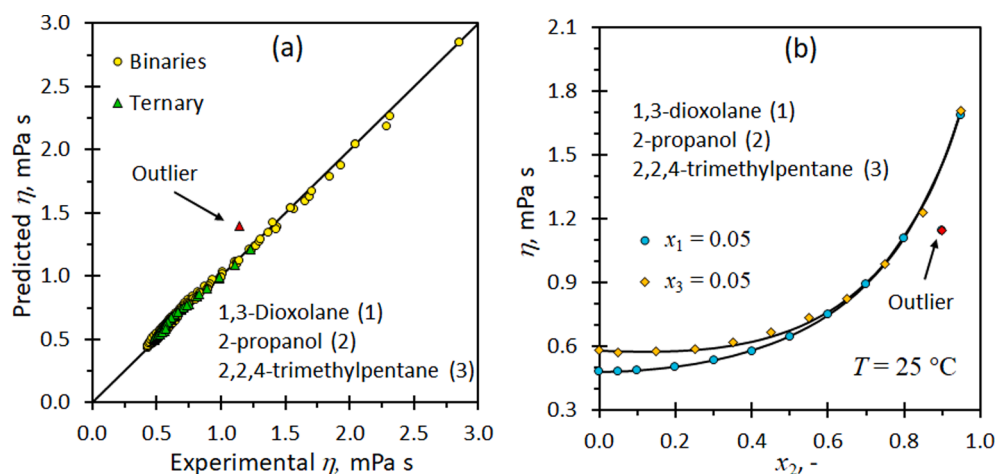
The ability of these two-parameter  $P(2,2)$  Padé model variants for predicting behaviour at temperatures beyond the isothermal temperature at which parameters were fixed, was tested. Binary data obtained at one temperature in the measured range were used to fix the parameter values. Fig. 6 then shows that, more often than not, it is indeed possible to extrapolate the results to predict ternary behaviour over a range of temperatures. However, it should be noted that the temperature range, that was used to test this assertion, was limited to either ten degrees lower or higher than the temperature at which the data were collected.

The advantage of a global property interpolation formula extends beyond simple data representation. It can also aid identification of erroneous measurements present in a data set. There are strong indications that two of the ten ternary data sets, considered presently, include outlier data points. There are well-established statistical methods for identifying such outliers. However, the present higher order model variants can provide visual indications of their potential presence.

Fig. 7 shows that System VIII likely includes an outlier corresponding to a ternary composition of  $\mathbf{x} = (0.05, 0.90, 0.05)$ . The outlier appears on the plot of predicted vs. experimental viscosity values shown in Fig. 7(a). This point is located at the intersection of two composition contours along which data are available. In the ternary simplex, they correspond to contours along which  $x_1 = 0.05$  and  $x_2 = 0.05$  respectively. The expectation is that the viscosity should vary smoothly with the mole fraction of component 2 along both of these two composition contours. Fig. 7(b) shows that the supposed outlier does not lie on either of the two predicted lines implying that it indeed represents an incorrect value.



**Fig. 6.** Comparing AAD values for the full data sets averaged over the ten ternary systems. The parameters of the  $P(2,2)$  Padé approximants were fixed considering only the binary data. In the case of model variants B\* and C\* they were determined using just one isothermal data set. The corresponding temperatures are shown in bold in Table 1.



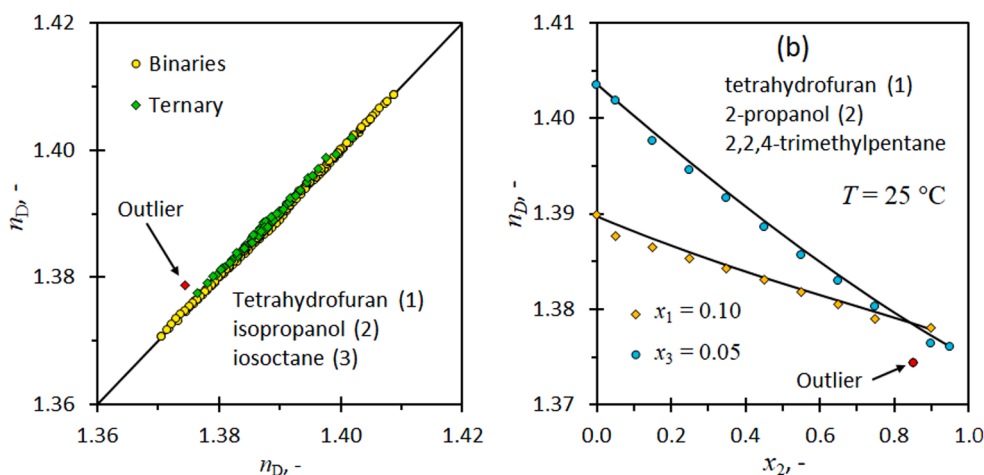
**Fig. 7.** (a) Plot of predicted viscosity against measured values for system VIII 1,3-dioxolane (1) + 2-propanol (2) + 2,2,4-trimethylpentane. The full data set was regressed using model variant F with  $\beta_i = V_i$ . The red triangle indicates a possible outlier corresponding to the composition (0.050, 0.900, 0.050). The variation of the measured and predicted viscosity values along the contour lines corresponding to  $x_1 = 0.050$  and  $x_3 = 0.050$  are shown in (b). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

A similar situation is illustrated in Fig. 8 in which an outlier is identified in the refractive index values. The red dot indicates a possible outlier corresponding to the composition  $\mathbf{x} = (0.10, 0.85, 0.05)$ . The variation of the measured and predicted viscosity values along the contour lines corresponding to  $x_1 = 0.10$  and  $x_2 = 0.05$  are shown in Fig. 8(b). The expected values for the refractive index at the indicated composition corresponds to the intersection of the two solid lines. The location of the actual reported value deviates significantly from this expectation.

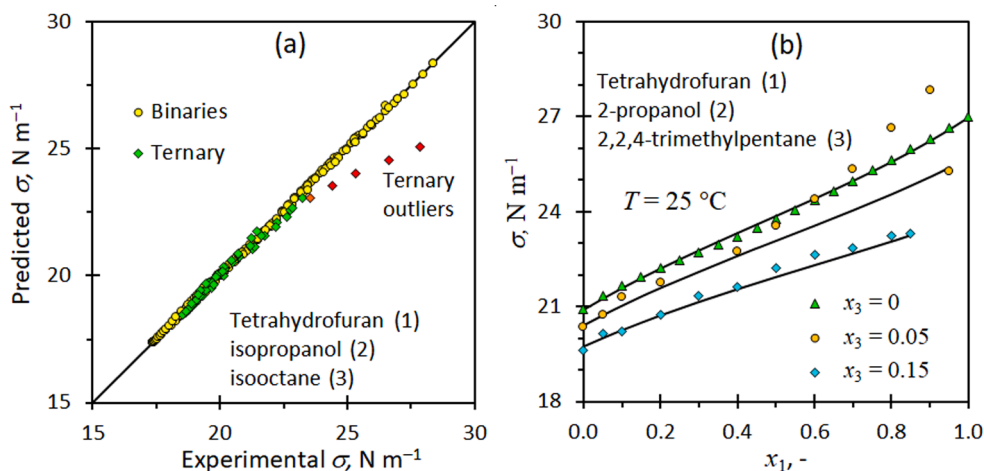
Fig. 9 shows that there are several outliers in System VI. The variation of the measured and predicted surface tension values along three contour lines, corresponding to  $x_3 = 0, 0.05$  and  $0.15$ , are shown in (b). The experimental values, plotted as orange dots in this Figure, strongly deviate from the expected data trends. The expectation is that the values

along the composition contour corresponding to  $x_3 = 0.05$  should be intermediate to the values observed and predicted along the other two contour lines. Since this is not the case, there are strong indications that the points in orange represent outliers, i.e. errors in the data set.

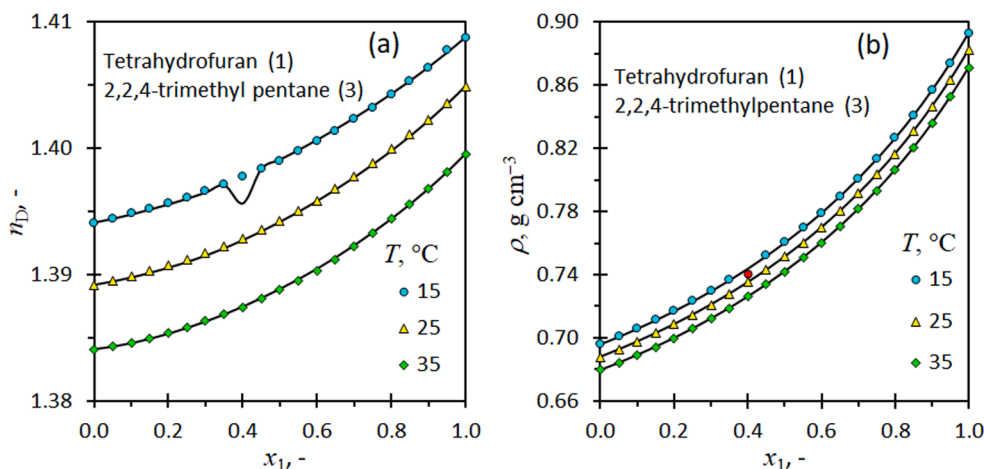
The aforementioned examples dealt with data points that deviated significantly from the expected trends. Fig. 10(a) shows a strange dip in the predicted refractive index curve for the binary mixture of tetrahydrofuran + 2,2,4-trimethylpentane at a temperature of 15 °C. This dip occurs at a composition of  $\mathbf{x} = (0.40, 0, 0.60)$  despite the fact that the refractive index measurement at this point was correct. Note that the predicted response was plotted as a smoothed curve in Excel. The full data set was regressed using model variant F with  $\beta_i = M_i$ . In this approach, the final estimates for the refractive index are obtained from Eq. (35). This requires knowledge of the density of the mixture at the



**Fig. 8.** (a) Plot of predicted refractive index against experimental values for System VI tetrahydrofuran (1) + 2-propanol (2) + 2,2,4-trimethylpentane. The outlier corresponds to the composition (0.05, 0.85, 0.05). The full data set was regressed using model variant F with  $\beta_i = V_i$ .



**Fig. 9.** (a) Predicted surface tension values plotted against experimental observations for System VI tetrahydrofuran (1) + 2-propanol (2) + 2,2,4-trimethylpentane. The full data set, except for the outliers, was regressed using model variant F with  $\beta_i = V_i$ . The compositions of the outlier points are provided in the Excel spreadsheet for this system.



**Fig. 10.** Results for the binary mixtures of tetrahydrofuran (1) + 2,2,4-trimethylpentane after System VI was regressed using the full ternary information using model variant F with  $\beta_i = M_i$ . The red dot highlights a small error in a density measurement (at a temperature of 25 °C) for the sample with composition (0.40, 0, 0.6). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

composition of interest. Clearly, if an experimental value for the density is used that is in error, the outcome will be an incorrect prediction for the refractive index. As seen in Fig. 10(b), the dip arose from a small error in the density measurement. A similar dip occurred in the surface tension plot at 15 °C. The same two effects were also observed when the other model variants were used together with  $\beta_i = M_i$ . The high sensitivity of the surface tension and the refractive index to small errors in density measurements makes this approach less attractive for representing experimental data. However, it makes these models eminently suitable for highlighting small precision errors in density measurements.

## 5. Conclusions

Models for predicting the density, viscosity, refractive index and surface tension for multicomponent mixtures, on the basis of knowledge of binary data, were addressed. One well-established and widely used conventional method estimates the physical properties of a multicomponent mixture by extrapolating the excess properties of the sub-binary systems using a variety of geometric arguments. It was shown that an alternative approach, based on the projection of the pure component properties onto the simplex region, yields more consistent mixture models. The starting points are expressions proposed for the case of ideal solution behaviour taking the form of linear blending rules. These can be recast in low-order  $P(1,1)$  Padé forms defined by the ratio of two first order Scheffé K-polynomials. They only feature pure-component property values which means that they are fully predictive for mixtures. Interestingly, the pure component property can either be associated with the molar mass or the molar volumes of constituents present in the mixture. Higher order model forms are then obtained using a pure component projection approach. This results in nested sets of special Scheffé polynomials or Padé approximants which only feature temperature-independent binary parameters. The temperature dependence, indicated by the ideal solution forms as being contributed by the pure component properties, is subsequently retained. The utility of the Padé approximant model variants was explored by correlating density, viscosity, refractive index and surface tension information contained in ten ternary data sets. Model performance was characterised using the Akaike information criterion and the mean absolute deviation between predicted and measure property values. Best representation of the composition and temperature dependence of the physical property trends was achieved when the binary model coefficients were fixed by regressing the full data sets. However, with the pure component parameter associated with the molar volume, good predictions of ternary trends were also obtained when limiting the regressions to the binary data. The present approach therefore provides accurate global interpolation of ternary data with parameter sparse expressions. Additionally, it was shown that the higher order models aid identification of data values that might be in error, i.e. that are outliers.

## Ethical approval

No ethical clearance was required for this study.

## Consent to participate

Consent to participate was not necessary as human subjects were not involved in any experiments.

## Author statement

During the preparation of this work, the authors did not make use of any generative AI or AI-assisted technologies in the writing process nor in the analysis of the data.

## Funding

This research was funded, in part, by a scholarship grant from the German Academic Exchange Service (DAAD) to PD at the University of Pretoria.

## CRediT authorship contribution statement

**Franco Pretorius:** Software, Validation, Visualization, Writing – review & editing. **Pethile Dzingai:** Data curation, Investigation, Writing – review & editing. **Elizabeth L. du Toit:** Supervision, Writing – review & editing. **Walter W. Focke:** Conceptualization, Data curation, Investigation, Methodology, Resources, Software, Validation, Writing – original draft.

## Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Pethile Dzingai reports financial support was provided by German Academic Exchange Service. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data will be made available on request.

## Acknowledgment

The authors express their gratitude to the Deutsche Forschungsgemeinschaft (DFG) and the German Academic Exchange Service (DAAD) for providing financial support.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.molliq.2024.124205>.

## References

- [1] F. Kohler, Monatshefte für Chemie und verwandte Teile anderer Wissenschaften 91 (1960) 738; 10.1007/BF00899814.
- [2] Y.-M. Muggianu, M. Gambino, J.-P. Bros, J. Chim. Phys. 72 (1975) 83.
- [3] G. Toop, Trans. TMS-AIME 223 (1965) 850.
- [4] M. Hillert, Calphad 4 (1980) 1, [https://doi.org/10.1016/0364-5916\(80\)90016-4](https://doi.org/10.1016/0364-5916(80)90016-4).
- [5] K.T. Jacob, K. Fitzner, Thermochim. Acta 18 (1977) 197, doi: 10.1016/0040-6031(77)80019-1.
- [6] A.D. Pelton, Calphad 25 (2001) 319, doi: 10.1016/S0364-5916(01)00052-9.
- [7] I. Cibulka, Collect. Czech. Chem. Commun. 47 (1982) 1414, doi: 10.1135/cccc19821414.
- [8] Z. Yu, H. Leng, Q. Luo, J. Zhang, X. Wu, K.-C. Chou, Mater. Des. 192 (2020) 108778, <https://doi.org/10.1016/j.matdes.2020.108778>.
- [9] B. Orge, M. Iglesias, J. Tojo, J.L. Legido, J. Chem. Thermodyn., 26 (1994) 121, doi: 10.1006/jcht.1994.1028.
- [10] H.-C. Ku, C.-H. Tu, Journal of Chemical & Engineering Data 50 (2005) 608, doi: 10.1021/jc049655w.
- [11] J.-D. Ye, C.-H. Tu, Journal of Chemical & Engineering Data 50 (2005) 1060, doi: 10.1021/jc050031f.
- [12] C.C. Wang, H.W. Chen, C.H. Tu, J. Chem. Eng. Data, 50 (2005) 1687, doi: 10.1021/jc0501639.
- [13] H.-W. Chen, C.-H. Tu, J. Chem. Eng. Data, 51 (2006) 261, doi: 10.1021/jc050367p.
- [14] T.T. Huang, C.T. Yeh, C.H. Tu, J. Chem. Eng. Data, 53 (2008) 1203, doi: 10.1021/jc07007445.
- [15] Y.-C. Kao, C.-H. Tu, J. Chem. Thermodyn., 43 (2011) 216, doi: 10.1016/j.jct.2010.08.019.
- [16] K.-D. Chen, Y.-F. Lin, C.-H. Tu, J. Chem. Eng. Data, 57 (2012) 1118, doi: 10.1021/jc201009c.
- [17] O. Redlich, A.T. Kister, Ind. Eng. Chem. 40 (1948) 345, doi: 10.1021/ie50458a036.
- [18] I. Ansara, C. Bernard, L. Kaufman, P. Spencer, Calphad 2 (1978) 1, [https://doi.org/10.1016/0364-5916\(78\)90002-0](https://doi.org/10.1016/0364-5916(78)90002-0).



- [19] G.H. Zhang, K.C. Chou, *J. Sol. Chem.* 39 (2010) 1200, doi: 10.1007/s10953-010-9570-5.
- [20] D.V. Malakhov, *Calphad* 35 (2011) 142, doi: 10.1016/j.calphad.2010.10.001.
- [21] K.-C. Chou, *Calphad* 19 (1995) 315, doi: 10.1016/0364-5916(95)00029-E.
- [22] T. Ju, X. Ding, X. Yan, C. Liu, X. Zhang, L. Zhang, *J. Mol. Liq.* 320 (2020), <https://doi.org/10.1016/j.molliq.2020.114469>.
- [23] K.C. Chou, *Int. J. Miner. Metall. Mater.* 29 (2022) 577 : 10.1007/s12613-022-2411-x.
- [24] P. Brocos, Á. Piñeiro, R. Bravo, A. Amigo, *Phys. Chem. Chem. Phys.* 5 (2003) 550, <https://doi.org/10.1039/B208765K>.
- [25] W.W. Focke, S. Endres, E.L. du Toit, M.T. Loots, R.L.J. Coetzer, *Ind. Eng. Chem. Res.* 60 (2021) 5639, doi: 10.1021/acs.iecr.0c06330.
- [26] H. Scheffé, *J. Roy. Stat. Soc.: Ser. B (Methodol.)* 20 (1958) 344.
- [27] N.R. Draper, F. Pukelsheim, *J. Stat. Plann. Inference* 71 (1998) 303, doi: 10.1016/S0378-3758(98)00012-3.
- [28] W.W. Focke, B. Du Plessis, *Ind. Eng. Chem. Res.* 43 (2004) 8369, doi: 10.1021/ie049415+.
- [29] P. Dzingai, F. Pretorius, W.W. Focke, *J. Mol. Liq.* 396 (2024), <https://doi.org/10.1016/j.molliq.2024.124003>. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85182391489&doi=10.1016%2fj.molliq.2024.124003&partnerID=40&md5=f1947b05cd6e32eee61f7d3c5907eb0e>.
- [30] J.G. Eberhart, *J. Phys. Chem.* 70 (1966) 1183, doi: 10.1021/j100876a035.
- [31] K.A. Connors, J.L. Wright, *Anal. Chem.*, 61 (1989) 194, doi: 10.1021/ac00178a001.
- [32] R. Belda, *Fluid Phase Equilib.* 282 (2009) 88, <https://doi.org/10.1016/j.fluid.2009.04.023>.
- [33] Á. Piñeiro, P. Brocos, A. Amigo, J. Gracia-Fadrique, M.G. Lemus, *Langmuir* 17 (2001) 4261, doi: 10.1021/la001210s.
- [34] C. Das, K. Bhattacharyya, *Chem. Phys. Lett.*, 374 (2003) 676, doi: 10.1016/S0009-2614(03)00813-3.
- [35] N. Shardt, J.A.W. Elliott, *Langmuir* 33 (2017) 11077, doi: 10.1021/acs.langmuir.7b02587.
- [36] F. Pretorius, W.W. Focke, R. Androsch, E. du Toit, *J. Mol. Liq.*, 332 (2021) 115893, doi: 10.1016/j.molliq.2021.115893.
- [37] W. Kay, *Ind. Eng. Chem.*, 28 (1936) 1014, doi: 10.1021/ie50321a008.
- [38] R. Patiño-Camino, A. Cova-Bonillo, M. Lapuerta, J. Rodríguez-Fernández, L. Segade, *Fluid Phase Equilibria* 555 (2022) 113363, doi: 10.1016/j.fluid.2021.113363.
- [39] D.B. Macleod, *Trans. Faraday Soc.* 19 (1923) 38, <https://doi.org/10.1039/TF9231900038>.
- [40] S. Sugden, *J. Chem. Soc. Trans.* 125 (1924) 1177, <https://doi.org/10.1039/CT9242501177>.
- [41] S.T. Bowden, E.T. Butler, *J. Chem. Soc. (resumed)* 79 (1939), <https://doi.org/10.1039/JR9390000079>.
- [42] E. Lederer, *Kolloid-Beihfte* 34 (1931) 270.
- [43] M.N. Azian, A.A. Mustafa Kamal, F. Panau, W.K. Ten, *J. Am. Chem. Soc.*, 78 (2001) 1001, doi: 10.1007/s11746-001-0378-9.
- [44] H. Akaike, *Bull. Int. Stat. Inst.* 50 (1983) 277.
- [45] H.W. Chen, C.H. Tu, *J. Chem. Eng. Data* 50 (2005) 1262, <https://doi.org/10.1021/je050010l>.
- [46] H.-C. Ku, C.-C. Wang, C.-H. Tu, *J. Chem. Eng. Data*, 53 (2008) 566, doi: 10.1021/je700626v.
- [47] C.-T. Yeh, C.-H. Tu, *J. Chem. Eng. Data* 52 (2007) 1760, <https://doi.org/10.1021/je700140j>.
- [48] H.-C. Ku, C.-C. Wang, C.-H. Tu, *J. Chem. Eng. Data*, 54 (2009) 131, doi: 10.1021/je800664z.