South African Journal of Chemical Engineering 54 (2025) 89–98

Contents lists available at ScienceDirect

South African Journal of Chemical Engineering

journal homepage: www.elsevier.com/locate/sajce

Mixture models inspired by the Kolmogorov-Arnold representation theorem

W.W. Focke

Department of Chemical Engineering, University of Pretoria, Lynnwood Road, Pretoria 0001, South Africa

ARTICLE INFO

Keywords:
Density
Viscosity
Refractive index
Surface tension
Ternary mixture
Liquid

ABSTRACT

Physical property models were developed to predict temperature-dependent multicomponent data using only temperature-independent binary parameters and pure component property temperature dependence. The Kolmogorov-Arnold representation theory was used to extend the linear blending rules and the Padé-like expressions describing the variation of physical properties of ideal solutions with composition. The effectiveness of correlating density, viscosity, refractive index and surface tension using this concept was tested. Ten ternary systems at either three or four different temperatures were regressed and compared to an ideal solution case. It was found that the four-parameter Kolmogorov-Arnold (KA) model performed excellently when the data regression included the full datasets. Unfortunately, the KA model may be too flexible, leading to overfitting binary data when applied to predicting ternary data.

1. Introduction

The physical property (y) of a homogeneous, n-component liquid solution depends on the temperature, the pressure and the composition of the mixture. The present interest is in the variation of selected liquid mixture properties with composition at isobaric conditions. The symbol z_i is used as the placeholder composition descriptor for component i. The set of conventional ones are:

$$z_i \in \{x_i, w_i, \varphi_i\} \tag{1}$$

where x_i is the mole fraction, w_i is the mass fraction and φ_i is the apparent volume fraction of component i in the mixture. All these composition descriptors satisfy the simplex constraint:

$$\sum_{i=1}^{n} z_i = 1 \; with \; 0 \le z_i \le 1 \tag{2}$$

In this communication, a mixture model is a mathematical expression that describes the variation of a physical property with composition, i.e. $y = y(\boldsymbol{z})$, where $\boldsymbol{z}$ is the composition vector (z_1,z_2, …, z_n). The Kolmogorov-Arnold representation theorem (KA) states that such a multivariate continuous function can be represented as a superposition of the two-argument addition of continuous functions of one variable (Kolmogorov, 1957). Formally, this statement is expressed mathematically as follows: For known y:$[0,1]^n \rightarrow \mathbb{R}$ there exists activation functions Φ_i: $\mathbb{R} \rightarrow \mathbb{R}$ and continuous functions ϕ_{ij}: $\mathbb{R} \rightarrow \mathbb{R}$ such that

$$y(z_1, z_2, \ldots, z_n) = \sum_{i=0}^{2n} \Phi_i \left(\sum_{j=1}^{n} \phi_{ij}(z_j) \right) \tag{3}$$

In essence, this theorem states that an arbitrary multivariate function can be represented by a finite composition of continuous functions of a single variable using just the binary operation of addition. Furthermore, because the Kolmogorov–Arnold representation decomposes a multivariate function into an interior and an outer function, it has a structure similar to a neural network with two hidden layers (Liu et al., 2024).

The physical properties of a homogeneous liquid solution are expected to vary continuously with composition. This means that the Kolmogorov-Arnold (KA) theorem applies to this situation and therefore it can, in principle, be used to model the physical property (y) of a liquid mixture. However, the KA representation theorem is a conceptual statement on the existence of such functions. It does not provide explicit information about the nature or structure of the inner- and outer functions. Therefore, imagination was required to identify suitable expressions for the construction of neural networks that implement the KA concept (Liu et al., 2024). This is also true when the objective is the design of an effective mixture model. In the latter case, these functions should be chosen so that the resulting model conforms to a set of consistency rules (Focke, 2006). In this regard, it is noteworthy that many past proposals for mixture models fail the decomposability condition. This requirement demands that the model should be invariant with respect to dividing one component into two or more identical subcomponents (Michelsen and Kistenmacher, 1990). This communication reports on a first attempt towards mixture models, inspired by the KA

E-mail address: walter.focke@up.ac.za.

https://doi.org/10.1016/j.sajce.2025.07.011
Received 18 March 2025; Received in revised form 10 May 2025; Accepted 27 July 2025
Available online 28 July 2025

Table 1
Linear blending rules for liquid mixtures and definitions of derived properties.

Physical property	Mole fractions	Volume fractions	Mass fractions
Molar mass	$M = \sum_i M_i x_i$		
Density	$V = \sum_i V_i x_i$ Molar volume: $V = M/\rho$	$\rho = \sum_i \rho_i \varphi_i$	$v = \sum_i v_i w_i$ Specific volume: $v = 1/\rho$
Refractive index	$R = \sum_i R_i x_i$ Molar refraction: $R \equiv \frac{n^2-1}{n^2+2} V$	$N = \sum_i N_i \varphi_i$ Lorentz-Lorenz: $N \equiv \frac{n^2-1}{n^2+2}$	$r = \sum_i r_i w_i$ $r \equiv N/\rho = R/M$
Surface tension	$P = \sum_i P_i x_i$ Parachor: $P \approx \sigma^{1/4} V = M\sigma^{1/4}/\rho$	$\sigma^{1/4} = \sum_i \varphi_i \sigma_i^{1/4}$	$s = \sum_i s_i w_i$ $s \equiv \sigma^{1/4}/\rho$
Viscosity (i)	$\ln \eta V = \sum_i x_i (\ln \eta_i V_i)$		
Viscosity (ii)	$V \ln \eta = \sum_i x_i (V_i \ln \eta_i)$	$\ln \eta = \sum_i \varphi_i \ln \eta_i$	$\frac{\ln \eta}{\rho} = \sum_i w_i \left(\frac{\ln \eta_i}{\rho_i}\right)$
Thermal conductivity			$k = \sum_i k_i w_i$
Isobaric heat capacity	$C = \sum_i C_i x_i$		

Table 2
Padé expressions valid for ideal solutions with mole fractions as the composition descriptors.

Property & Parameter	"Ideal solution" expression	Eqn.
Density	$\rho = \frac{M}{V} = \frac{\sum M_i x_i}{\sum V_i x_i}$	9
Refractive index Lorentz-Lorenz parameter: $N \equiv \frac{n^2-1}{n^2+2}$	$N = \frac{R}{V} = \frac{\sum R_i x_i}{\sum V_i x_i}$	10
Surface tension Parachor: $P \approx \sigma^{1/4} V = M\sigma^{1/4}/\rho$	$\sigma^{1/4} = \frac{P}{V} = \frac{\sum P_i x_i}{\sum V_i x_i}$	11
Viscosity	$\ln \eta = \frac{\sum (V_i \ln \eta_i) x_i}{\sum V_i x_i}$	12

representation theorem, with this requirement fulfilled.

Mixture models feature parameters linked to the pure component properties and to the magnitude of the interactions between pairs of unlike molecules. In essence, the mixture models can be viewed as a composition weighted average over these parameters (Focke WW et al., 2007). Therefore, it might be possible for the inner functions to be expressed as a transformation of a weighted Kolmogorov-Nagumo-de Finetti mean, also known as the weighted quasi-arithmetic mean (M_n) (Nielsen, 2023). For a set of observations a_i, with associated weights x_i, with $\Sigma x_i = 1$, this mean is defined by:

$$M_n(a_1, a_2, \ldots, a_n; x_1, x_2, \ldots, x_n) = f^{-1}\left(\sum_{i=1}^{n} x_i f(a_i)\right) \tag{4}$$

where the f_i are continuous and strictly monotonic functions. One possible option for these functions are the power means of order p (M_p). They arise when it holds that $f(u) = u^p$ in Eq. 4:

$$M_p(\boldsymbol{a}, \boldsymbol{x}) = \left(\sum_{i=1}^{n} x_i a_i^p\right)^{\frac{1}{p}} \tag{5}$$

Applying these ideas to Eq. 3, and assuming that a common outer function suffices, it leads to the following proposal for a general KA-inspired mixture model:

$$y(\boldsymbol{z}) = T^{-1}\left(\sum_{i=1}^{m} T\left(f^{-1}\left(\sum_{j=1}^{n} z_j f_i(a_{ij})\right)\right)\right) \tag{6}$$

Note that the outer function is defined as the inverse of the transformation used for defining the inner functions. The advantage of this simplification is that it preserves the units of the model parameters. Eq. 6 is a conceptual statement that covers a multitude of possible expressions depending on the choice made for the transformation T and the functions f_i. In this communication previous insights are used to guide the selection of the transfer function T to be used in conjunction with Eq. 5 for the f_i's. This generated a specific version of Eq. 6. Furthermore, its performance was tested using published data for a range of ternary liquid mixtures of organic compounds, i.e. non-electrolytes.

2. Model development

The ideal solution is a core concept in thermodynamics. Some properties of ideal solutions vary linearly with composition expressed in terms of mole fractions. More generally, the linear blending rule (LBR) is often applied to estimate the variation of a property with mixture composition. For a ternary mixture it states:

$$y = c_1 z_1 + c_2 z_2 + c_3 z_3 \tag{7}$$

where y and c_i are, respectively, the physical properties of the mixture and component i at the temperature and pressure of the mixture. Table 1 lists examples of such linear blending rules for selected thermodynamic and transport properties.

Consider momentarily, the properties of "ideal solutions". The molar mass M, molar volume V, the molar refraction R, the Parachor P and (perhaps) the composite $\ln \eta V$ might be considered to be the primary properties since they are expected to follow the linear blending rule in mole fractions. The secondary properties are the density ρ, Lorentz-Lorenz parameter N, the fourth root of the surface tension and the logarithm of the dynamic viscosity. All of them can be expressed in terms of ratios of two primary properties. They also follow the LBR in volume factions. Consequently, for ideal solutions, they can also be expressed as a Padé-like expression, i.e. in terms of a ratio of two first order polynomials in the primary properties:

$$y = \sum a_i x_i \Big/ \sum b_i x_i \tag{8}$$

where $a_i/b_i = c_i$ is the pure component property of interest, a_i and b_i are primary properties of component i, usually with $b_i = V_i$, i.e. the molar volume. Table 2 lists such Padé expressions suggested for "ideal" solutions.

The expectation is that, to a first approximation, real solution behaviour follows the trends predicted by the linear- and Padé-like expressions for "ideal solutions". Therefore, mixture models should be

Table 3
The lower order Scheffé K-polynomial expressions for a ternary mixture with mole fractions as composition descriptor.

Order	Scheffé polynomial	Eqn.
Linear	$y = c_1 x_1 + c_2 x_2 + c_3 x_3$	13
Quadratic	$y = c_{11} x_1^2 + c_{22} x_2^2 + c_{33} x_3^2 + 2c_{12} x_1 x_2 + 2c_{13} x_1 x_3 + 2c_{23} x_2 x_3$	14
Cubic	$y = c_{111} x_1^3 + c_{222} x_2^3 + c_{333} x_3^3 + 3c_{112} x_1^2 x_2 + 3c_{113} x_1^2 x_3$ $+ 3c_{122} x_2^2 x_3 + 3c_{223} x_2^2 x_3 + 3c_{133} x_1 x_3^2 + 3c_{233} x_2 x_3^2 + 6c_{123} x_1 x_2 x_3$	15

Table 4
Model expressions for ternary mixtures with the conditions giving rise to the simpler versions.

Model	Expression for y
	Kolmogorov-Arnold inspired models
KA/4p	$\frac{c_1x_1(x_1+\kappa_{12}x_2+\kappa_{13}x_3)}{x_1+\lambda_{12}x_2+\lambda_{13}x_3}+\frac{c_2x_2(\kappa_{21}x_1+x_2+\kappa_{23}x_3)}{\lambda_{21}x_1+x_2+\lambda_{23}x_3}+\frac{c_3x_3(\kappa_{31}x_1+\kappa_{32}x_2+x_3)}{\lambda_{31}x_1+\lambda_{32}x_2+x_3}$
KA/2p	
$\kappa_{ji}=1/\kappa_{ij}$	
$\lambda_{ji}=1/\lambda_{ij}$	
Wassiljewa	$\frac{c_1x_1}{x_1+\lambda_{12}x_2+\lambda_{13}x_3}+\frac{c_2x_2}{\lambda_{21}x_1+x_2+\lambda_{23}x_3}+\frac{c_3x_3}{\lambda_{31}x_1+\lambda_{32}x_2+x_3}$
$\kappa_{ij}=1$	
Scheffé	$c_1x_1(x_1+\kappa_{12}x_2+\kappa_{13}x_3)+c_2x_2(\kappa_{21}x_1+x_2+\kappa_{23}x_3)+c_3x_3(\kappa_{31}x_1+\kappa_{32}x_2+x_3)$
$\lambda_{ij}=1$	
Ideal LBR	$c_1x_1+c_2x_2+c_3x_3$ (i.e. $\kappa_{ij}=1$ and $\lambda_{ij}=1$)
	Projection-based Padé approximant
$P(2,2)/4p$	
$P(2,2)/2p$	$\frac{V_1c_1x_1(x_1+\kappa_{12}x_2+\kappa_{13}x_{23})+V_2c_2x_2(\kappa_{21}x_1+x_2+\kappa_{23}x_3)+V_3c_3x_3(\kappa_{31}x_1+\kappa_{32}x_2+x_3)}{V_1x_1(x_1+\lambda_{12}x_2+\lambda_{13}x_{23})+V_2x_2(\lambda_{21}x_1+x_2+\lambda_{23}x_3)+V_3x_3(\lambda_{31}x_1+\lambda_{32}x_2+x_3)}$
$\kappa_{ji}=1/\kappa_{ij}$	
$\lambda_{ji}=1/\lambda_{ij}$	
$P(1,2)$	$\frac{V_1c_1x_1+V_2c_2x_2+V_3c_3x_3}{V_1x_1(x_1+\lambda_{12}x_2+\lambda_{13}x_{23})+V_2x_2(\lambda_{21}x_1+x_2+\lambda_{23}x_3)+V_3x_3(\lambda_{31}x_1+\lambda_{32}x_2+x_3)}$
$\kappa_{ij}=1$	
$P(2,1)$	$\frac{V_1c_1x_1(x_1+\kappa_{12}x_2+\kappa_{13}x_{23})+V_2c_2x_2(\kappa_{21}x_1+x_2+\kappa_{23}x_3)+V_3c_3x_3(\kappa_{31}x_1+\kappa_{32}x_2+x_3)}{V_1x_1+V_2x_2+V_3x_3}$
$\lambda_{ij}=1$	
$P(1,1)$	$\frac{V_1c_1x_1+V_2c_2x_2+V_3c_3x_3}{V_1x_1+V_2x_2+V_3x_3}$ (i.e. $\kappa_{ij}=1$ and $\lambda_{ij}=1$)

Table 5
Data sources for density (ρ), refractive index (n_D), viscosity (η), and surface tension (σ) used for testing the models shown in Table 4. It lists the different temperatures for which measurements were obtained together with the number of binary and ternary data points.

System A	Acetone (1) - ethanol (2) - 2,2,4-trimethylpentane (3) - 2-butanone (4) - diisopropyl ether (5) - methylcyclohexane (6)	
	Temperatures (°C)	15, **25, 35, 45**
	Data points	Binary: 665 Ternary: 330
	References	(Ku and Tu, 2005); (Ye and Tu, 2005); (Chen HW and Tu, 2005); (Chen H-W and Tu, 2006)
System B	Tetrahydrofuran (1) - 2-propanol (2) - 2,2,4-trimethylpentane (3) - 1,3-dioxolane (4) - tetrahydropyran (5)	
	Temperatures (°C)	15, 20, **25, 30,** 35, 40, 50
	Data points	Binary: 542 Ternary: 165
	References	(Ku et al., 2008); (Kao and Tu, 2011)
System C	Ethanol (1) - benzyl acetate (2) - benzyl alcohol (3) - 2-propanol (4) - 2-phenylethanol (5)	
	Temperatures (°C)	20, 25, **30, 35**
	Data points	Binary: 399 Ternary: 110
	References	(Huang et al., 2008); (Yeh and Tu, 2007); (Ku et al., 2009); (Chen K-D et al., 2012)

The temperatures shown in bold represent those for which ternary data were available.
* B: Number of distinct pure component and binary data values; T: Number of ternary data points.

formulated such that they include these forms as limiting cases. The Scheffé polynomials are the classic mixture models applied in experimental design (Scheffé, 1958; Cornell, 2002). Essentially, they correspond to truncated Taylor series that take the simplex constraint into account. The K-forms are homogeneous versions (Draper and Pukelsheim, 1998). Some of the lower order ones are listed in Table 3 for ternary mixtures with mole fractions as the composition descriptors. The first order versions correspond to the linear blending rule. The quadratic forms reduce to the linear form when the cross parameters can be expressed as the arithmetic averages of the pure component properties, i.e. $c_{ij}=(c_i+c_j)/2$. Similarly, the cubic Scheffé polynomial form reduces to the linear expression when $c_{ijk}=(c_i+c_j+c_k)/3$ for all i, j, k.

Establishing the values of the coefficients of the ternary and higher terms in cubic and higher order Scheffé polynomials is of practical concern. Fixing them, on the basis of experimental results, requires data for ternary or higher order mixtures. Such data are scarce or often not even available. It would be convenient if pure component and binary mixture data would suffice to fix all model coefficients. Consequently, alternative versions of cubic Scheffé polynomials have been proposed (Focke W et al., 2007; Pretorius et al., 2024). One approach considered the cubic Scheffé polynomial defined by the product of three linear expressions (Pretorius et al., 2024):

$$y(\mathbf{z})=\left(\sum_{i=1}^{n}\alpha_i z_i\right)\left(\sum_{i=1}^{n}\beta_i z_i\right)\left(\sum_{i=1}^{n}\gamma_i z_i\right) \tag{16}$$

Where the α_i, β_i, and γ_i are adjustable parameters. They are not independent quantities since the pure component property values are defined by the products $c_i=\alpha_i\beta_i\gamma_i$. Conveniently, Eq. 16 can be recast in a more useful form:

$$y(\mathbf{z})=\sum_{i=1}^{n}c_i z_i\left(\sum_{i=1}^{n}\omega_{ij}z_i\right)\left(\sum_{i=1}^{n}\tau_{ij}z_i\right) \tag{17}$$

with $\omega_{ij}=\beta_j/\beta_i$, and $\tau_{ij}=\gamma_j/\gamma_i$. Note that this also implies that $\omega_{ii}=\tau_{ii}=1$. Furthermore, in this projective form, the mixture property is expressed fully in terms of the pure component property values c_i and binary interactions coefficients (Pretorius et al., 2024). Consequently, the ternary constant, which arises when ternary and higher mixtures are considered, is defined in terms of the latter parameters.

It is noteworthy to realise that expression 17 conforms to the formalism of Eq. 6 since it can be written in the following equivalent form:

$$y(\mathbf{z})=\sum_{i=1}^{n}\exp\left[\ln(c_i z_i)+\ln\left(\sum_{j=1}^{n}\omega_{ij}z_j\right)+\ln\left(\sum_{k=1}^{n}\tau_{ik}z_k\right)\right] \tag{18}$$

Eq. 18 employs natural logarithms for the transformation T. Hence the outer function is defined by the exponential function. Note that each of the terms in brackets can be interpreted as weighted arithmetic means taken over the coefficient vectors $\boldsymbol{\omega}_i$ and $\boldsymbol{\tau}_i$. Note that Eq. 18 only applies n such summations whereas the KA representation theorem prescribes a total of $2n+1$. Therefore, additional logarithmic terms can be added and this will also facilitate improved data representation. The disadvantage is that the number of adjustable model parameters becomes very large. However, there is another troubling matter associated with Eq. 17. The last two terms have exactly the same form, which means that the κ_{ij} and the λ_{ij} coefficients are strongly correlated. This causes difficulties when regressing actual experimental data. This problem is alleviated by

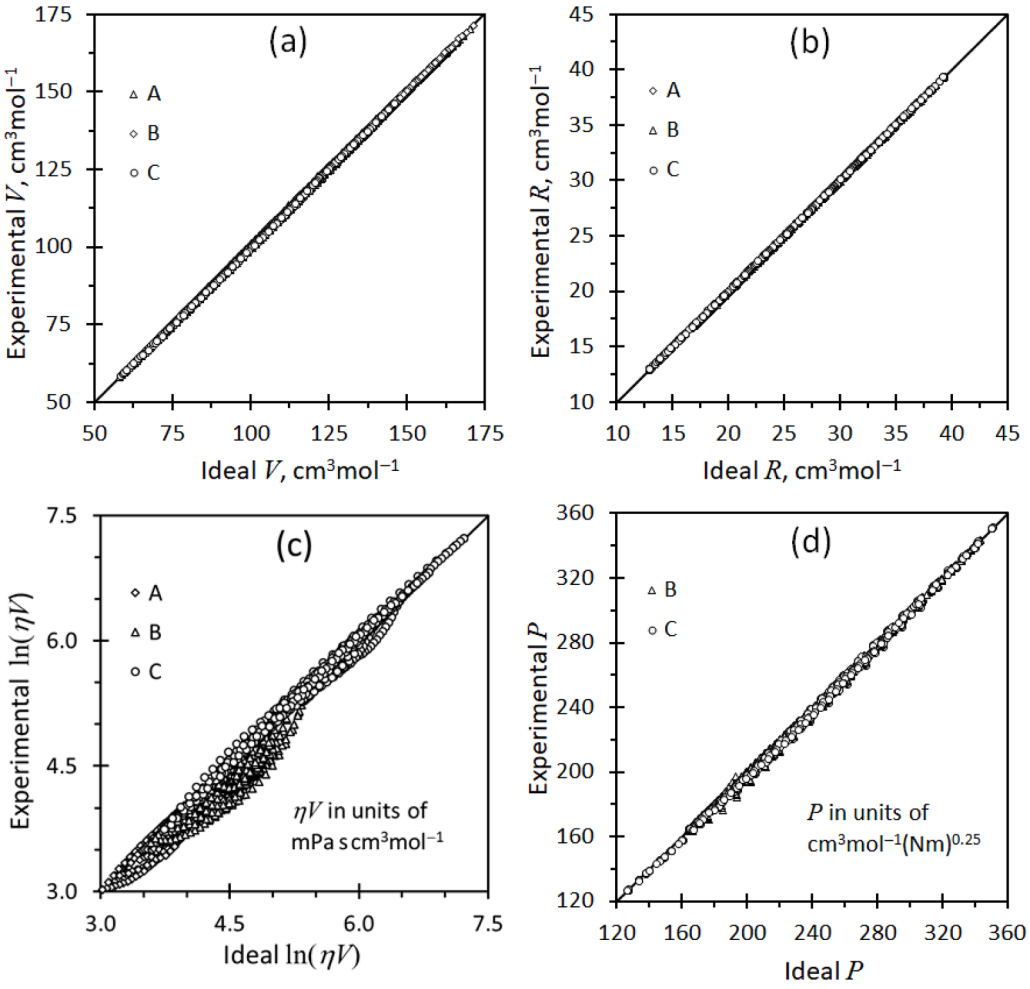

Fig. 1. Predictions for (a) molar volume, V, (b) molar refraction, R, (c) logarithm of the product of the viscosity and the molar volume, ηV and (d), Parachor, P, using the mole fraction-based LBR expressions listed in Table 1.

replacing those terms, which are in the form of arithmetic means, by sequences of different power means as suggested by Eq. 5. Reformulating extensions of Eq. 18, using this idea, results in:

$$y(\mathbf{z}) = \sum_{i=1}^{n} \exp\left[\ln(c_i z_i) + p_1 \ln\left(\sum_{j=1}^{n} \kappa_{ij} z_j\right) + p_2 \ln\left(\sum_{k=1}^{n} \lambda_{ik} z_k\right) + \ldots\right] \quad (19)$$

Where $\kappa_{ij} = \omega_{ij}^{1/p_1}$ and $\lambda_{ik} = \tau_{ik}^{1/p_2}$ are the revised adjustable constants with $\kappa_{ii} = \lambda_{ii} = 1$; and the p_i are power mean exponents.

Note that the first term in the square brackets must always be retained as it captures the contribution of the relevant pure component property. If only this first term is retained, expression 19 reduces to the linear blending rule. When only one additional linear term is included, a quadratic Scheffé polynomial is obtained. A particularly fruitful approach allows negative values for the p_i's, the power mean exponents. If, only a single term with p = -1 is retained, Eq. 19 reduces to the Wassiljewa model (Wassiljewa, 1904) which features the same composition dependence as the well-known NRTL equation (Renon and Prausnitz, 1968):

$$y(\mathbf{z}) = \frac{c_1 z_1}{z_1 + \lambda_{12} z_2 + \lambda_{13} z_3} + \frac{c_2 z_2}{\lambda_{21} z_1 + z_2 + \lambda_{23} z_3} + \frac{c_3 z_3}{\lambda_{31} z_1 + \lambda_{32} z_2 + z_3} \quad (20)$$

A simple extension is obtained from this KA approach by just keeping three terms with $p_1 = 1$ and $p_2 = -1$. This yields the following LBR-based and KA-inspired expression for a multicomponent mixture:

$$y(\mathbf{z}) = \sum_{i=1}^{n} c_i z_i \sum_{k=1}^{n} \kappa_{ik} z_k \Big/ \sum_{j=1}^{n} \lambda_{ij} z_j \quad (21)$$

Eq. 21 defines a new mixture model, not previously considered, that arises from a cropped KA approach.

Recently it was found that the effect of temperature on some mixture properties is primarily determined by the temperature dependencies exhibited by the pure constituents of the mixture (Shardt and Elliott, 2017; Pretorius et al., 2024). If this is actually the case, the adjustable binary coefficients in Eq. 19, Eq. 20 and Eq. 21, should be temperature independent. This implies that experimental data for physical properties, e.g. density, can be regressed directly using these equations or, where necessary, their extensions.

There is another possibility for regressing experimental data, especially when the density of the mixture has also been measured. This is because the molar volume of the mixture plays a central role in the Padé expressions listed in Table 2. Perusal of Table 1 shows that the LBR expressions for molar mass M, the molar volume V, the molar refraction R, the Parachor P, and the composite property $V \ln \eta$, are all expressed in terms of linear blending rules based on mole fractions. The LBR forms are just the lowest order expressions for these quantities. Via the KA expressions, more accurate data representations should be achievable. When these are established, the density, refractive index, surface tension and viscosity of the mixture are readily calculated from the relevant primary property ratios listed in Table 2. The usefulness of this

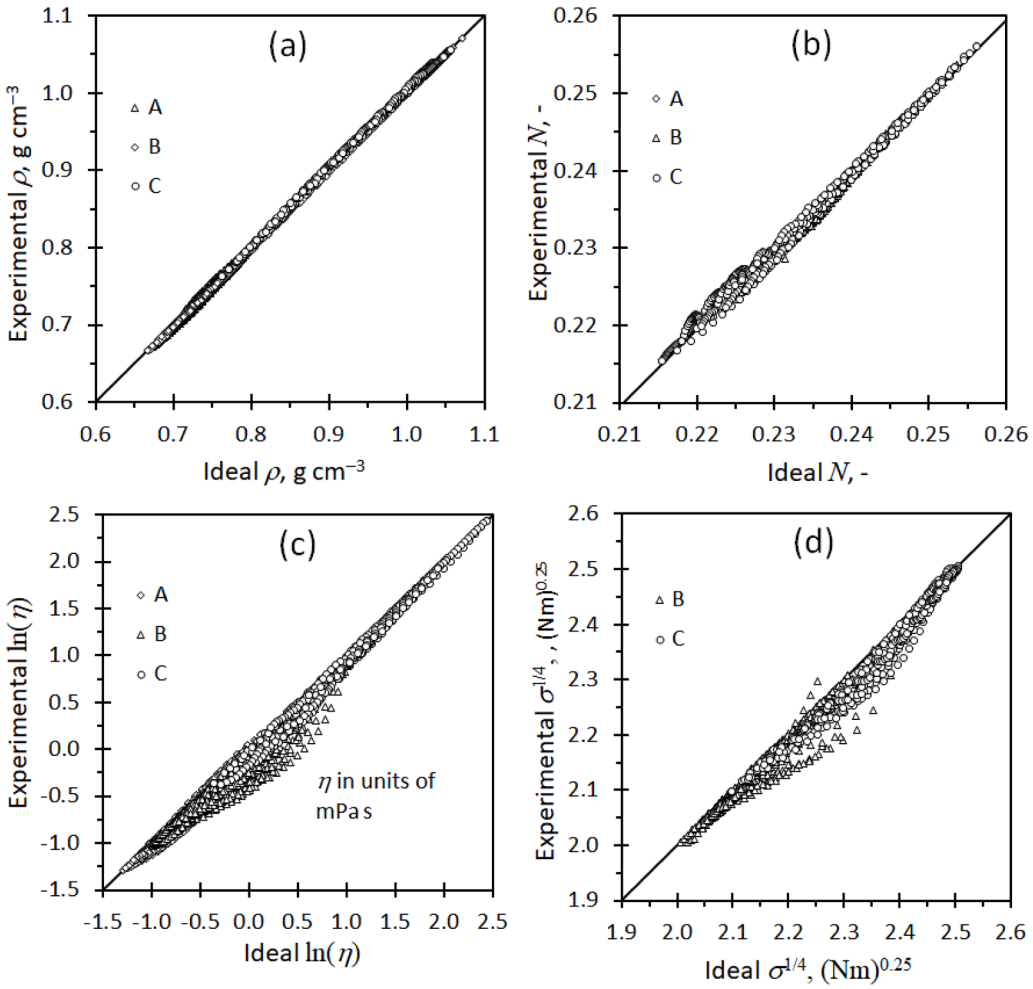

Fig. 2. Predictions for (a) density, ρ, (b) Lorentz-Lorenz parameter, N, (c) logarithm of the viscosity, $\ln(\eta)$ and (d), $\sigma^{¼}$, using the volume fraction-based LBR expressions listed in Table 1.

Table 6
Performance of the fully predictive linear blending rules listed in Table 1.

Property	Molar volume	Molar refraction	"Arrhenius"	Parachor
Equation	$V = \Sigma x_i V_i$	$R = \Sigma x_i R_i$	$\ln\eta V = \Sigma x_i \ln(\eta_i V_i)$	$P = \Sigma x_i P_i$
AD (%)	0.05 ± 0.37	0.03 ± 0.13	-2.17 ± 2.55	-0.73 ± 0.91
AAD (%)	0.28 ± 0.23	0.09 ± 0.11	3.02 ± 2.71	0.78 ± 0.86
Correlation	0.99992	0.999987	0.9861	0.9565
Property	Density	"Lorentz-Lorenz"	Viscosity	Surface tension
Equation	$\rho = \Sigma\varphi_i\rho_i$	$N = \Sigma\varphi_i N_i$	$\ln\eta = \Sigma\varphi_i \ln\eta_i$	$\sigma^{1/4} = \sum \varphi_i \sigma_i^{1/4}$
AD (%)	-0.05 ± 0.36	0.03 ± 0.40	n.a.	-0.72 ± 0.81
AAD (%)	0.28 ± 0.23	0.30 ± 0.27	n.a.	0.74 ± 0.79
Correlation	0.99971	0.99939	0.9927	0.9911

procedure, for experimental data reduction, is explored in this communication.

Other approaches for formulating effective mixture models have been explored. This includes the projection-based Padé expression (Dzingai et al., 2024); (Pretorius et al., 2024):

$$y(\mathbf{z}) = \sum_i b_i c_i x_i \left(\sum_j \kappa_{ij} x_j\right) \Big/ \left[\sum_i b_i x_i \left(\sum_j \lambda_{ij} x_j\right)\right] \tag{22}$$

The projection-based expression, defined by Eq. 22, was up to this point in time the only model that was suitable for correlating a range of different liquid properties with temperature independent binary coefficients. It, as well as its simplified forms, proved to be highly effective for correlating ternary mixture data for density, refractive index, viscosity and surface tension with the temperature dependence determined by those of the pure component properties. Therefore, it is used as a yardstick to gauge the relative performance of the KA-derived model, Eq. 21, for representing the physical properties of ternary mixtures. Table 4 lists the expressions, valid for ternary mixtures, for the models tested presently. Note that full versions of Eq. 21 and Eq. 22 feature four adjustable binary parameters. Setting $\kappa_{ji} = 1/\kappa_{ij}$ and $\lambda_{ji} = 1/\lambda_{ij}$ reduces them to two per binary and this option is also explored. In addition to the models listed in Table 4, the KA/8p model was also tested using the full data sets. It is defined by the following expression:

$$y(\mathbf{z}) = \sum_{i=1}^{3} b_i c_i x_i \frac{\left(\sum_{j=1}^{3} \kappa_{ij} x_j\right)\left(\sum_{j=1}^{3} k_{ij} x_j\right)^2}{\left(\sum_{j=1}^{3} \lambda_{ij} x_j\right)\left(\sum_{j=1}^{3} l_{ij} x_j\right)^2} \tag{23}$$

Table 7
Summary of absolute average deviations (%). Number in bold and italic indicate best and worst performance respectively for the models listed in Table 4.

Model	System	Ideal	*P*(1,1)	Scheffé	Wassiljewa	KA/2p	KA/4p	*P*(1,2)	*P*(2,1)	*P*(2,2)/2p	*P*(2,2)/4p	KA/8p
Density												
Binaries	A	0.301	***0.302***	0.053	0.098	0.309	0.286	**0.026**	0.031	0.046	0.042	
	B	0.223	0.223	0.049	0.038	0.489	***0.437***	0.040	0.042	0.041	**0.037**	
	C	0.341	0.341	0.018	0.053	0.046	***0.817***	0.046	0.036	0.027	**0.021**	
Ternaries	A	0.301	***0.302***	0.052	0.036	0.059	**0.020**	0.025	0.031	0.030	0.022	0.015
	B	***0.223***	***0.223***	0.048	0.036	0.028	**0.017**	0.041	0.041	0.037	0.033	0.018
	C	***0.341***	***0.341***	0.018	0.023	0.021	**0.011**	0.045	0.036	0.019	0.013	0.010
Refractive index												
Binaries	A	0.085	***0.087***	0.016	0.015	0.012	0.027	**0.008**	0.009	**0.008**	0.011	
	B	0.107	0.070	0.176	0.134	***0.206***	0.085	0.023	0.024	0.023	**0.018**	
	C	***0.159***	0.159	0.008	0.012	0.023	0.013	0.025	0.018	0.009	**0.006**	
Ternaries	A	0.085	***0.087***	0.016	0.011	0.009	0.008	0.008	0.009	0.008	**0.007**	0.007
	B	***0.107***	0.070	0.059	0.066	0.072	0.027	0.024	0.023	0.022	**0.017**	0.025
	C	***0.159***	0.159	0.008	0.008	0.009	**0.005**	0.024	0.018	0.009	0.006	0.005
Viscosity												
Binaries	A	5.77	***5.78***	1.80	1.88	3.07	1.63	2.33	1.67	1.42	**0.94**	
	B	***15.36***	15.32	2.27	2.89	1.41	0.91	13.27	1.88	1.39	**0.84**	
	C	***7.38***	***7.38***	1.14	1.12	1.50	1.02	2.65	0.99	1.35	**0.91**	
Ternaries	A	5.77	5.78	1.78	1.66	1.28	1.09	2.28	1.64	1.38	**0.84**	0.50
	B	***15.36***	15.32	2.26	2.89	1.33	0.66	13.27	1.86	1.36	**0.72**	0.05
	C	***7.38***	***7.38***	1.11	1.09	1.41	0.79	2.61	0.97	1.35	**0.88**	0.69
Surface tension												
Binaries	B	2.86	***15.32***	0.73	0.79	0.65	**0.40**	13.27	1.88	1.39	0.84	
	C	3.39	***3.40***	0.60	1.08	1.68	0.72	0.48	0.46	**0.46**	0.51	
Ternaries	B	***2.86***	2.85	0.71	0.60	0.54	**0.40**	0.93	0.92	0.49	**0.40**	0.37
	C	3.39	***3.40***	0.60	0.42	0.51	**0.25**	0.46	0.46	0.45	0.29	0.25

3. Data analysis

The capacity of the models listed in Table 4 to accurately represent real experimental information, was tested using experimental data gleaned from the literature. Table 5 shows the three data sets, comprising of several ternary systems, that were used for this purpose. These data sets were selected because they included experimental results for binary and ternary mixtures 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 (2017) when they considered the surface tension of binary mixtures.

The parameter values were fixed using least squares regression analysis executed in Excel spreadsheets. Results obtained when using only the binary data for this purpose, and also when the full data sets were considered, are reported. The Akaike information criterion (AIC) (Akaike, 1983) was used as one of the measures for ranking the relative performance of the models. The governing equation is:

$$AIC = n\ln(SSE) + 2df \tag{24}$$

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 dimensionless:

$$\Delta AIC = n\ln\left(SSE_{\text{ref}} / SSE_{\text{model}}\right) + 2\Delta df \tag{25}$$

Eq. 25 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 data points in the present data sets enhanced the importance of the first term on the right of Eq. 24 at the expense of the second term, which corrects for the increase in the number of model parameters.

For each mixture property, the corresponding "ideal solution" LBR expression served as the reference model. The obtained ΔAIC values were normalised by taking the ratio with respect to the largest value recorded. Invariably, this was the ΔAIC value obtained when the full set of parameters were allowed to vary freely, i.e. model variation KA/8p, KA/4p or *P*(2,2)/4p. These model variants are the most flexible ones as they feature more adjustable parameters than the others. This approach provided a relative scale with best performance indicated by ΔAIC values equal to unity. In some instances, negative values for ΔAIC were obtained when the regression was limited to binary data. This indicates that the model predictions were worse than those obtained using the corresponding fully predictive "ideal solution" LBR expression. The normalised ΔAIC values, obtained when regressing just the binary data, were compared to the values obtained when the full data set were used instead. For some models, 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 sometimes 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/1OMlX_SP86gxwV7WCtwkk_VU13EBCwvgt?usp=drive_link

4. Results

In order to establish a baseline for comparison, the ability of ideal-solution-based expressions (Table 1) to represent the data of the three ternary systems was first considered. Fig. 1 shows the mole-fraction-based LBR predictions for the molar volume, the molar refraction, the Parachor, and the product of the viscosity and the molar volume, while Fig. 2 shows the predictions for the properties listed in Table 1 that vary linearly with volume fraction. The average relative errors (AD) and the average absolute relative errors (AAD), as well as the correlation coefficients obtained for the full available data sets are summarised in Table 6. The definitions of AD and AAD for a given property P are:

$$\text{AD} = \left(P_{\text{exp}} - P_{pred}\right)/P_{\text{exp}} \tag{26}$$

and

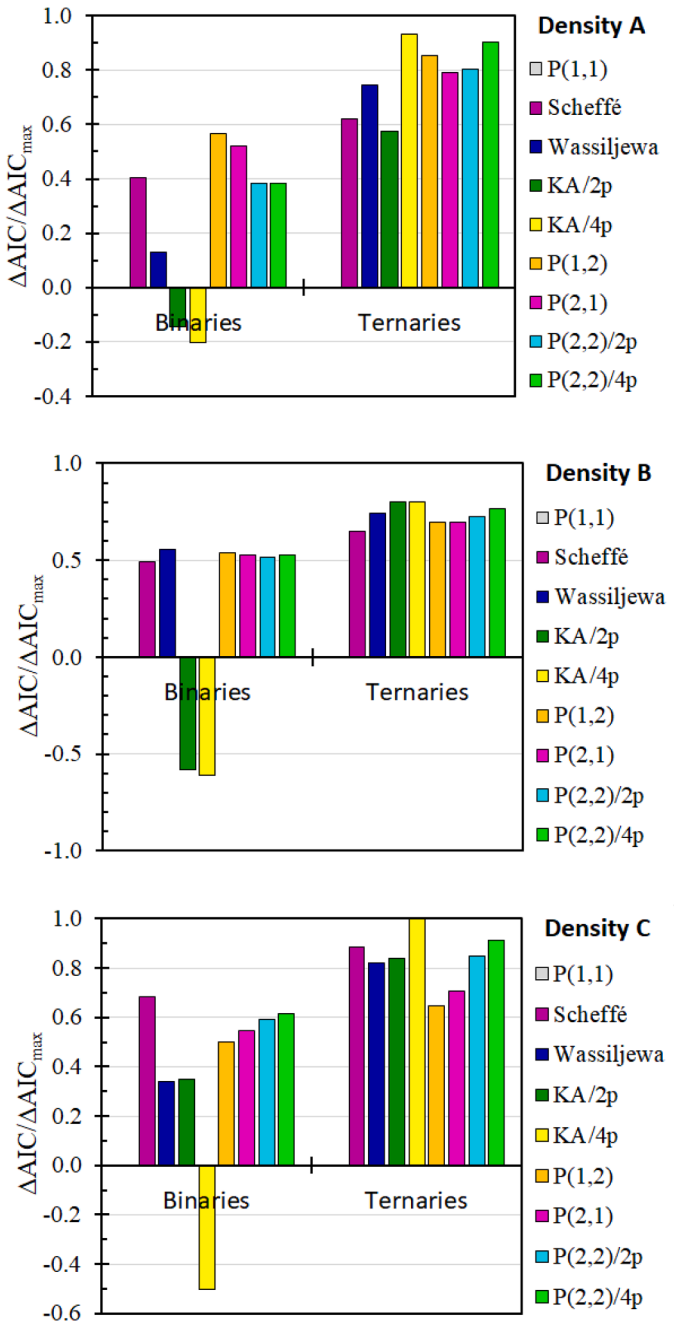

Fig. 3. Scaled ΔAIC values for system density prediction obtained by fitting the various models to the three ternary data systems listed in Table 4 when either the binary data sets (Binaries) or full data sets (including Ternaries) were used to fix the model parameters.

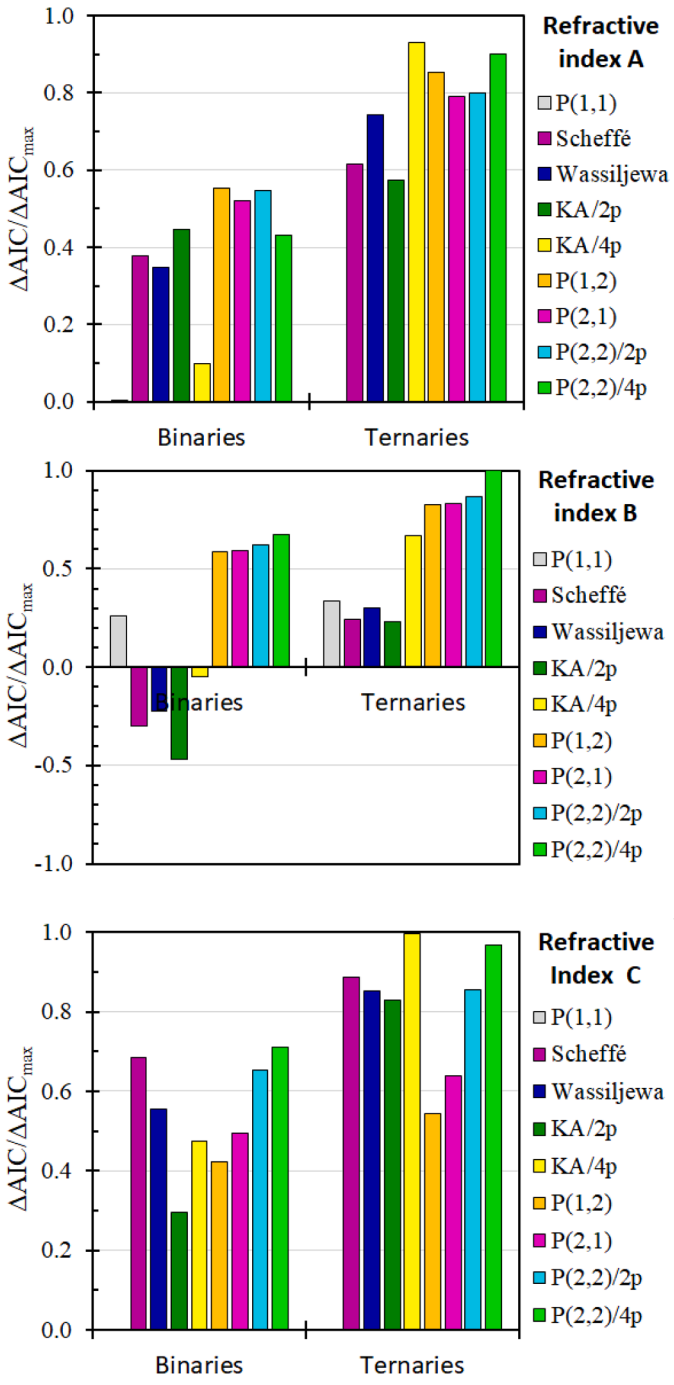

Fig. 4. Scaled ΔAIC values for the refractive index predictions obtained by fitting the various models to the three ternary data systems listed in Table 4 when either the binary data sets (Binaries) or full data sets (including Ternaries) were used to fix the model parameters.

$$\mathrm{AAD} = \left|P_{\exp} - P_{pred}\right| / P_{\exp} \tag{27}$$

Perusal of the results in Table 6 shows that the LBR predictions for the molar volume (V), the molar refraction (R), the Parachor (P), the density (ρ), the Lorentz-Lorenz parameter (N) and the fourth root of the surface tension are surprisingly accurate. In all these cases, the average deviation is less than 1 %. The logarithm of the Arrhenius viscosity parameter (ηV) and the logarithm of the viscosity also show a strong linear dependence on mole fractions and volume fractions respectively. However, in the latter case the average deviation exceeds 3 % while in the former case the relative deviations could not be determined as the range includes both positive and negative values. One may conclude that the linear blending rules are surprisingly accurate and probably provide acceptable estimates for many engineering applications.

Table 7 shows the AAD values for the secondary properties, achieved by regressing the data using the different models. By design, the mixture models listed in Table 5 are extensions of the "ideal solution" expressions. This means that there is limited scope for improving accuracy of the data correlation. Nevertheless, using the four parameter models the AAD values were significantly lowered, by about an order of magnitude, when compared to the "ideal" expressions. Provided the full data sets, including the ternary results, are regressed, it appears possible to predict density, refractive index, viscosity and surface tensions to better than 0.042 %, 0.017 %, 0.94 % and 0.84 % respectively. The corresponding

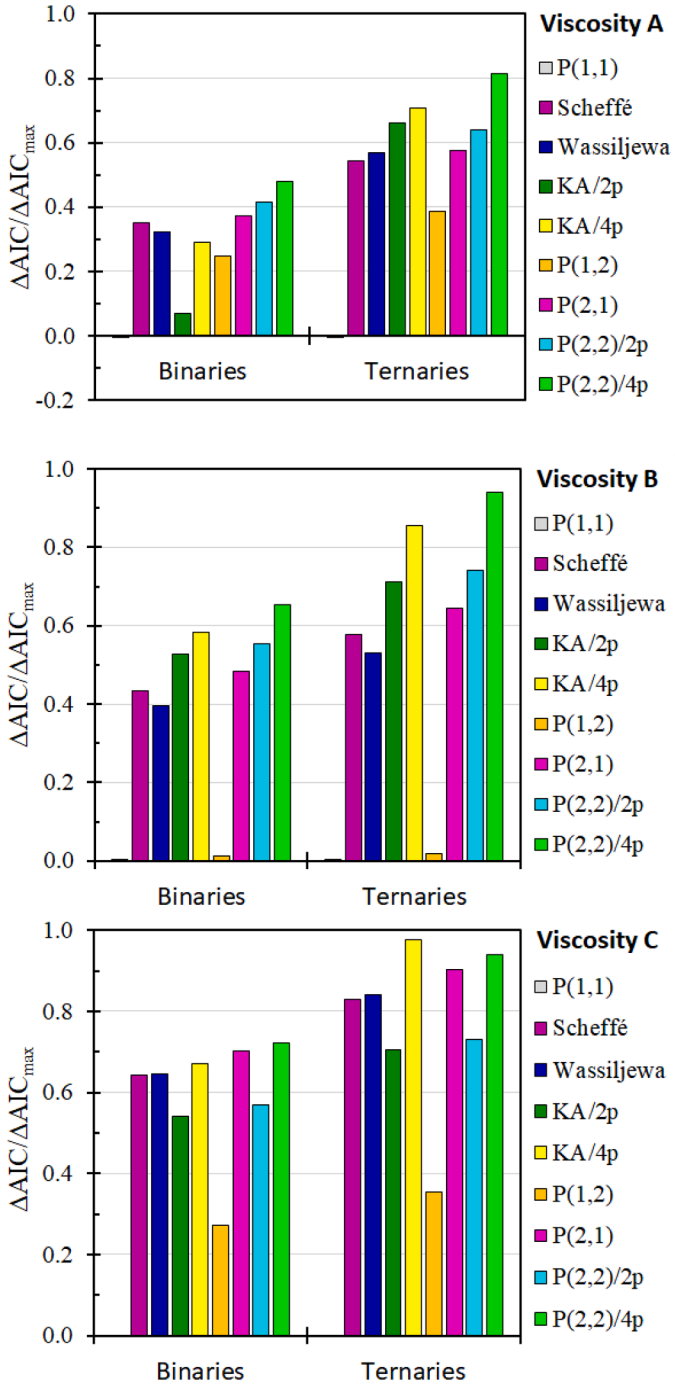

Fig. 5. Scaled ΔAIC values for the viscosity predictions obtained by fitting the various models to the three ternary data systems listed in Table 4 when either the binary data sets (Binaries) or full data sets (including Ternaries) were used to fix the model parameters.

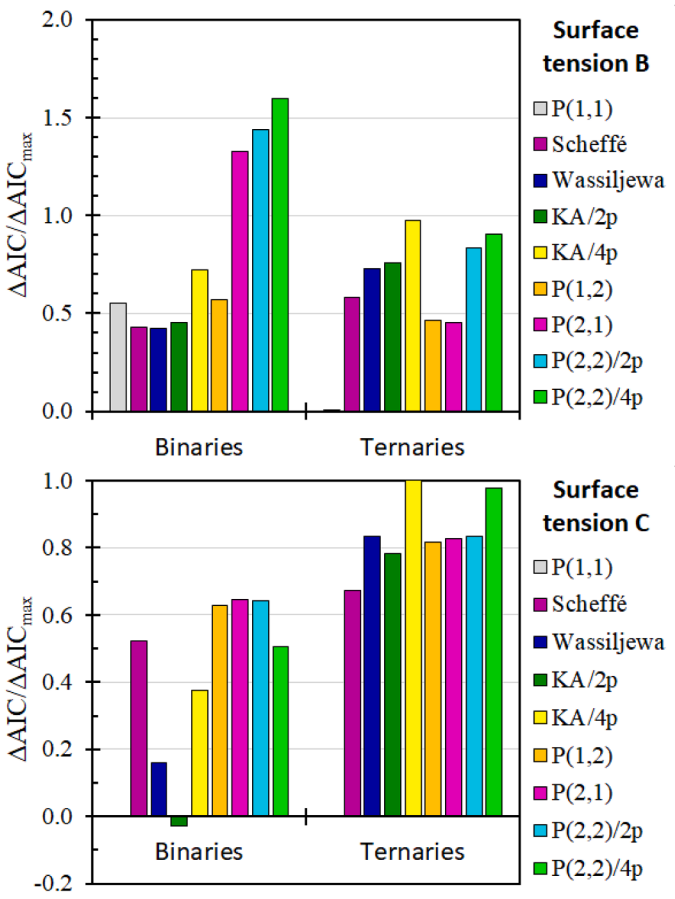

Fig. 6. Scaled ΔAIC values for the surface tension predictions obtained by fitting the various models to the three ternary data systems listed in Table 4 when either the binary data sets (Binaries) or full data sets (including Ternaries) were used to fix the model parameters.

values for the "ideal" expressions were 0.341 %, 0.159 %, 15.36 % and 3.39 %.

The results presented in Table 7 are best considered in conjunction with Fig. 3 to Fig. 6. These show a summary of the Akaike information results in a way that facilitates visual comparison of performance across all of the models considered for the secondary physical properties. In these Figures the ΔAIC values, normalised with respect to the best performing model, are plotted. Negative values show model performance that was worse than what was achieve using the fully predictive "ideal" expressions. In Table 7 the values in red indicate the worst AAD values while those presented in bold show the lowest AAD values achieved.

Table 8
Absolute average deviations for density, refractive index, viscosity and surface tension. These properties were calculated from the "ideal" Padé expressions and fits done using the KA/4p model for the primary properties molar mass (*M*), molar volume (*V*), Lorentz-Lorenz parameter (*N*), Parachor (*P*), and viscosity function $V \ln \eta$.

Property	Model	System A	System B	System C
Density	Ideal	0.301 ± 0.237	0.223 ± 0.167	0.341 ± 0.281
	KA/4p	0.016 ± 0.015	0.014 ± 0.020	0.011 ± 0.014
Refractive index	Ideal	0.085 ± 0.069	0.070 ± 0.060	0.159 ± 0.129
	KA/4p	0.010 ± 0.009	0.011 ± 0.017	0.007 ± 0.007
Viscosity	Ideal	6.66 ± 6.48	15.3 ± 15.6	7.38 ± 6.84
	KA/4p	0.57 ± 0.66	0.66 ± 1.14	1.08 ± 1.12
Surface tension	Ideal	-	2.84 ± 3.37	1.43 ± 1.38
	KA/4p	-	0.41 ± 0.61	0.33 ± 0.37

Density. Fig. 3 and Table 7 show that the KA/4p model not only provided the best predictions when the full data sets were regressed, but this model also gave some of the worst predictions when only the binary data was used to fix the model parameters. This disappointing outcome is tentatively attributed to overfitting of the binary data resulting in poor predictions for the ternary data. The implication is that the KA/4p model is too flexible. This notion is supported by the fact that, when only binary data was regressed, both the Scheffé and the Wassiljewa models outperformed the KA/2p and KA/4p versions. Overall, it would appear that any of the two-parameter projection models are preferred as they all

gave good predictions irrespective to whether just the binary data or the full data sets were used to find the model parameter values.

Refractive index. Fig. 4 and Table 7 show that, overall, the *P*(2,2)/4p model gave the best predictions for the refractive index. The exceptions were data set C where model KA/4p was best when the full data were regressed and data set A where model *P*(1,2) performed slightly better when just the binary data were considered. Except for system B, for which model KA/2p performed worse that the "ideal" expressions, the KA model types performed quite well, even when the model parameters were determined on the basis of the binary data alone. However, the projection-based models generally performed better.

Viscosity. The P(2,2)/4p model was a clear winner with the lowest AAD values (reported in Table 7) irrespective of the way the parameters were determined. This is confirmed by Fig. 5. The P(2,2)/2p and the P(2,1) models also did well, but for System B, the P(1,2) model faired rather poorly. The KA/4p model was second best except for System A when the regression considered binary data only.

Surface tension. The results for surface tension in Table 7 and Fig. 6 show that model KA/4p performed best when the full data set was used to fix the model parameters, whereas the Projection models were the best when just the binary data were considered. The KA/2p performed poorly for System C.

Predicting the secondary- from primary properties. Table 8 shows that it is indeed possible to calculate the properties ρ, N, $\ln\eta$ and $\sigma^{¼}$ from the molar mass (M), the molar refraction (R), the viscosity group $V\ln\eta$, and the Parachor (P) ratio'ed with respect to the volume fraction (V). Good results were found even when the linear blending rules listed in Table 6 were used for this purpose. Significantly more accurate predictions were obtained using the KA/4p model. The absolute average deviations are presented in Table 8. The results show that this approach can provide highly accurate data representation for the secondary physical properties. The Supplementary Information includes additional graphical visualisations that support this observation.

5. Conclusions

The ideal solution is an important concept considered in thermodynamics. The primary properties describing ideal solution behaviour are the molar volume, the molar refraction, the Lorentz-Lorenz parameter and the Parachor. These properties are all expected to follow the linear blending rule (LBR) with mole fractions as the composition descriptors. It was found that the behaviour of the real mixtures, considered in this communication, did indeed closely agree with these expectations. Therefore, it makes sense to design more accurate mixture models such that they include these expressions as limiting cases. This was recently done using pure component projections. The Kolmogorov-Arnold representation theory inspired another approach to models for predicting the density, viscosity, refractive index and surface tension for multicomponent mixtures, on the basis of knowledge of binary data. A very general form is presented by Eq. 6 in the main text. In this communication, a specific example described by Eq. 21 was considered. The ability of this equation to represent ternary data for density, refractive index, viscosity and surface tension, over a limited temperature range was tested. It was found that it performed excellently when the data regression included the full data sets. Unfortunately, poor data representation resulted when only the binary data were considered for fixing the model parameters. By comparison, the binary-predictive capacity of the projective mixture models previously proposed, fared better. The implication is that the Kolmogorov-Arnold inspired model may be too flexible, leading to overfitting of the binary data, which results in poor prediction of ternary behaviour.

Declarations

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 contributions

Walter W Focke: Conceptualization, Methodology, Software, Validation, Formal analysis, Data Curation, Writing - Original Draft.

Funding

This research was not funded

Availability of data and materials

Supplementary Information and the Excel spreadsheets containing the full data sets with calculations are available from the corresponding author.

CRediT authorship contribution statement

W.W. Focke: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The author expresses his gratitude to Elizbe du Toit for editorial assistance.

References

Akaike, H., 1983. Information measures and model selection. Bull. Int. Stat. Inst. 50, 277–290.

Chen, H.-W., Tu, C.-H., 2006. Densities, viscosities, and refractive indices for binary and ternary mixtures of diisopropyl ether, ethanol, and 2,2,4-trimethylpentane. J. Chem. Eng. Data 51 (1), 261–267. https://doi.org/10.1021/je050367p.

Chen, H.W., Tu, C.H., 2005. Densities, viscosities, and refractive indices for binary and ternary mixtures of acetone, ethanol, and 2,2,4-trimethylpentane. J. Chem. Eng. Data 50 (4), 1262–1269. https://doi.org/10.1021/je050010l.

Chen, K.-D., Lin, Y.-F., Tu, C.-H., 2012. Densities, viscosities, refractive indexes, and surface tensions for mixtures of ethanol, benzyl acetate, and benzyl alcohol. J. Chem. Eng. Data 57 (4), 1118–1127. https://doi.org/10.1021/je201009c.

Cornell, J.A., 2002. Experiments with Mixtures: Designs, Models, and the Analysis of Mixture Data. John Wiley & Sons, New York.

Draper, N.R, Pukelsheim, F., 1998. Mixture models based on homogeneous polynomials. J. Statist. Res. 71 (1), 303–311. https://doi.org/10.1016/S0378-3758(98)00012-3.

Dzingai, P., Pretorius, F., Focke, W.W., 2024. Correlating multicomponent surface tension data with Padé approximants. Part I. surface tension. J. Mol. Liq. 396. https://doi.org/10.1016/j.molliq.2024.124003.

Focke, W., Sandrock, C., Kok, S., 2007. Weighted-power-mean mixture model: application to multicomponent liquid viscosity. Ind. Eng. Chem. Res. 46 (13), 4660–4666. https://doi.org/10.1021/ie061465m.

Focke, W.W., 2006. Mixture models based on neural network averaging. Neural Comput. 18 (1), 1–9. https://doi.org/10.1162/089976606774841576.

Huang, T.T., Yeh, C.T., Tu, C.H., 2008. Densities, viscosities, refractive indices, and surface tensions for the ternary mixtures of 2-propanol + benzyl alcohol + 2-phenylethanol at T = 308.15 K. J. Chem. Eng. Data 53 (5), 1203–1207. https://doi.org/10.1021/je7007445.

Kao, Y.-C., Tu, C.-H., 2011. Densities, viscosities, refractive indices, and surface tensions for binary and ternary mixtures of 2-propanol, tetrahydropyran, and 2,2,4-trimethylpentane. J. Chem. Thermodyn. 43 (2), 216–226. https://doi.org/10.1016/j.jct.2010.08.019.

Kolmogorov, A.N., 1957. On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition. Dokl. Akad. Nauk.

Ku, H.-C., Tu, C.-H., 2005. Densities and viscosities of binary and ternary mixtures of ethanol, 2-butanone, and 2,2,4-trimethylpentane at T = (298.15, 308.15, and 318.15) K. J. Chem. Eng. Data 50 (2), 608–615. https://doi.org/10.1021/je049655w.

Ku, H.-C., Wang, C.-C., Tu, C.-H., 2008. Densities, viscosities, refractive indexes, and surface tensions for binary and ternary mixtures of tetrahydofuran, 2-propanol, and 2,2,4-trimethylpentane. J. Chem. Eng. Data 53 (2), 566–573. https://doi.org/10.1021/je700626v.

Ku, H.-C., Wang, C.-C., Tu, C.-H., 2009. Densities, viscosities, refractive indices, and surface tensions for the mixtures of 1,3-dioxolane + 2-propanol or + 2,2,4-trimethylpentane at (288.15, 298.15, and 308.15) K and 1,3-dioxolane + 2-propanol + 2,2,4-trimethylpentane at 298.15 K. J. Chem. Eng. Data 54 (1), 131–136. https://doi.org/10.1021/je800664z.

Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Soljačić, M., Hou, T.Y., Tegmark, M., 2024. Kan: Kolmogorov-Arnold networks. ArXiv Prepr. arXiv:240419756.

Michelsen, M.L., Kistenmacher, H., 1990. On composition-dependent interaction coefficients. Fluid Phase Equilib. 58 (1–2), 229–230.

Nielsen, F., 2023. Beyond scalar quasi-arithmetic means: quasi-arithmetic averages and quasi-arithmetic mixtures in information geometry. ArXiv Prepr. arXiv:230110980.

Pretorius, F., Dzingai, P., du Toit, E.L., Focke, W.W., 2024. Correlating multicomponent physical property data with Padé approximants: part II. Projecting pure component behaviour J. Mol. Liq. 397. https://doi.org/10.1016/j.molliq.2024.124205.

Renon, H., Prausnitz, J.M., 1968. Local compositions in thermodynamic excess functions for liquid mixtures. AIChE J. 14 (1), 135–144. https://doi.org/10.1002/aic.690140124.

Scheffé, H., 1958. Experiments with mixtures. J. R. Stat. Soc. B 20 (2), 344–360.

Shardt, N., Elliott, J.A.W., 2017. Model for the surface tension of dilute and concentrated binary aqueous mixtures as a function of composition and temperature. Langmuir 33 (41), 11077–11085. https://doi.org/10.1021/acs.langmuir.7b02587.

Wassiljewa, A., 1904. Waermeleitung in Gasgemischen. Phys. Z. 5, 737.

Ye, J.-D., Tu, C.-H., 2005. Densities, viscosities, and refractive indices for binary and ternary mixtures of diisopropyl ether, ethanol, and methylcyclohexane. J. Chem. Eng. Data 50 (3), 1060–1067. https://doi.org/10.1021/je050031f.

Yeh, C.-T., Tu, C.-H., 2007. Densities, viscosities, refractive indexes, and surface tensions for binary mixtures of 2-propanol + benzyl alcohol, + 2-phenylethanol and benzyl alcohol + 2-phenylethanol at T = (298.15, 308.15, and 318.15) K. J. Chem. Eng. Data 52 (5), 1760–1767. https://doi.org/10.1021/je700140j.