



# A Machine Learning approach for enhancing permittivity mixing rules of binary liquids with a Gaussian modification and a new interaction factor estimation

Miguel Monteagudo Honrubia\*, Francisco Javier Herraiz-Martínez, Javier Matanza Domingo

Instituto de Investigación Tecnológica (I.I.T.), Universidad Pontificia Comillas, Calle de Santa Cruz de Marcenado, 26, Madrid, 28015, Spain

## ARTICLE INFO

### Keywords:

Binary mixture  
Dielectric characterization  
Machine Learning  
Mixing rules  
Permittivity

## ABSTRACT

The microstructure and solvation mechanics of binary liquids are key for predicting mixture permittivity. However, since traditional mixing rules do not consider this complexity, they must be modified to address the mixture characteristics through an interaction factor ( $k_{int}$ ). This paper evaluates this parameter for several mixing rules, applying Support Vector Regressor models trained with glycerin-water reflective signals acquired with a Dielectric Resonator sensor. The regression error of these models indicates both the optimal interaction factor and the mixing rule that fits the most with experimental permittivity values. Kraszewski and Hashin-Shtrikman mixing rules achieved the best performance with an RMSE of around 1. In addition, this paper suggests that the interaction factor can be estimated through the molar volume and the dielectric contrast between liquids ( $k_{int} = 2.67$ ) without acquiring experimental data. Moreover, after analyzing the physical limitations of a linear modification formula, this paper proposes an alternative based on a Gaussian function that avoids unrealistic volume fractions. Both contributions enhance mixing rule accuracy and improve the flexibility to model mixture dielectric behavior.

## 1. Introduction

Any heterogeneous material is a complex physical-chemical system with inclusions of another material or different phases of the same material. The structural characteristics of these materials such as size, shape, spatial distribution, or the mutual interaction of the heterogeneities, are critical to understanding the macroscopic properties. This paper focuses on dielectric permittivity, which can reveal information about the intramolecular interaction of mixture components, helping to characterize liquid solutions as well as composite materials [1]. It also has particular relevance for designing and validating microwave sensors that measure permittivity changes as a detection method for chemical or biological samples [2].

Nevertheless, estimating the dielectric permittivity in binary mixtures is still challenging, and so far, the best alternative is performing a direct measurement of the mixture of interest. The open-ended probe is the most popular dielectric spectroscopy technique for liquid characterization. Connected to a Vector Network Analyzer (VNA), the coaxial probe is submerged in the liquid sample to measure the

reflective wave's scattering parameters in a long range of frequencies [3]. However, the relationship between the scattering parameters and the medium permittivity relies on algorithms and the probe equivalent circuit. Depending on these models, the measured permittivity could present variations [4]. Moreover, the experimental setup imposes constraints in terms of probe dimensions and positioning, sample size, and frequency range, which affect the method accuracy [5]. The lack of dielectric data for many binary mixtures hinders the validation of the measurements. Therefore, a theoretical permittivity estimation would be useful to contrast any experimental results and detect possible systematic errors. In addition, this methodology requires expensive equipment, which is not always available, and developing an accurate mixing rule would be faster and cost-saving to estimate the permittivity of binary solutions.

First, Section 2 discusses the state-of-the-art of permittivity estimations with mixing rules and their limitations in modeling the dielectric behavior of binary mixtures. Therefore, this section also introduces a novel Machine Learning approach for validating and optimizing mixing rules through an interaction parameter. Section 3 details the data and

\* Corresponding author.

E-mail address: [mmonteagudo@comillas.edu](mailto:mmonteagudo@comillas.edu) (M. Monteagudo Honrubia).

the algorithms employed and discusses two contributions to improve mixing rules: a formula to estimate the interaction factor through pure liquid properties and a new Gaussian modification formula. Finally, Section 4 shows our approach results for the glycerine-water mixture and discusses the effect of both contributions to improve permittivity estimation.

## 2. Theory

### 2.1. The search for a general solution

In an extensive review, Brosseau [6] summarized the historical approaches addressing the permittivity estimation of binary mixtures, from theoretical formulations, using Maxwell's electromagnetic (EM) theory, to relatively recent electromagnetic simulations of these materials. As a result, many formulas have been proposed to estimate the permittivity of heterostructures from the permittivity value of the pure compounds without performing experimental measurements [7]; however, none of them has been proved as a general solution. Indeed, one of the critical issues detected by Brosseau is the excessive number of formulas with uneven estimations that hinder their interpretation and validation.

In general, the formulas aim to calculate the permittivity of an effective homogeneous medium, which conceptually replaces the heterogeneous medium while keeping the same dielectric properties. It should be noted that this approximation is only valid when the incident electromagnetic (EM) waves have a wavelength significantly greater than the dimensions of the inhomogeneities. Besides, many of these formulas are constructed under the assumption of isotropy, which is rarely the case for complex materials. However, regardless of their theoretical grounding, all these formulas have limited success fitting the experimental data, thus giving poor permittivity estimations. In contrast, other formulas follow an empirical approach, and although they allow adjusting experimental data, they are limited to specific systems and cannot be generalized.

As has been observed, most theoretical formulas only work adequately for a minimal fraction of heterogeneities, while for larger volumes of inclusions, the formula's error becomes excessive [6]. It has been suggested that these formulas only consider volume fraction or molar fraction as unique parameters, without indicating the structure of the mixture or the possible interaction between components. In addition to the bibliography examined by Brosseau, later works comparing mixing rules also demonstrate these discrepancies between experimental data and permittivity estimation [7][8]. However, it should be noted that the lack of experimental data limits this mixing rule validation, and the existing data have been extracted with different methods, many of which lack an estimate of the error.

Qin et al. [9] conducted an extensive review of the calculation of the electromagnetic properties of composites based on their structural parameters at the microscopic (atoms and dipoles), mesoscopic (phases, inclusions, agglomerates), and macroscopic (sample volume) scales. Knowing the relationship between scales is essential for understanding the propagation of EM waves in these materials and, therefore, for designing tailored composites for different technological applications. The importance of the mixture structure is strongly related to the very definition of permittivity: a physical parameter that indicates the polarization capability of the material and, therefore, how easily its internal charges can be aligned to the incident electric field [10]. Permittivity depends on the density and the separation of charges; therefore, the heterogeneities in a material cause a non-uniform polarization, which depends not only on the position and distribution of the inclusions, but also on the electric interaction between them and the solvent matrix containing them [9]. In conclusion, there is an effective permittivity, but this can only be accurately calculated by knowing the microstructure of the mixture and the set of electrical interactions between the

phases. However, even knowing the geometry, the permittivity calculation is a high-complexity problem and relies on EM simulation as the only feasible method.

This problem is particularly complex in the case of liquid solutions since the fluid microstructure cannot be observed as easily as a solid material. Tang et al. [11] analyze the severe difficulties in observing and determining cluster structures in aqueous solutions, even when spectral techniques or computational simulations are used. The formation of solute clusters in the primary solvent in aqueous solution defines the mesostructure and microstructure. This solvation process is determined by the network of hydrogen bonds and Coulombic forces, which largely depends on parameters such as chain length and solute concentration, but also on solution thermodynamic properties [11]. The rearrangement process can generate different geometries of clusters: spheres, cubes and chains [12]. The studies reviewed by Tang et al. indicate that conformational changes occur at critical concentrations [11]. Therefore, the microstructure changes (structural transition) should be observed through trend changes in the permittivity-volume fraction function. These insights can be extended to other polar solvents as shown by Pradhan et al. in a review of molecular interactions of non-aqueous solutions [13]. Moreover, if the solute is also a polar substance, the system will undergo a competition between solute-solute, solvent-solute, and solvent-solvent interactions. In that case, the solvation process will be too specific to extract a general rule [14].

In summary, defining the exact dielectric behavior of a solution is a high-complexity problem that has not yet been solved. For this reason, despite the drawbacks discussed in this section, mixing rules seem the only feasible method for estimating the effective permittivity of binary liquids. Therefore, new approaches are required to improve mixing rules, including other parameters to approximate molecular interaction or microstructure.

### 2.2. Improving mixture rules, from the thermodynamic approach to Puranik modification

The works of Reis et al. [15] and Iglesias et al. [1] propose a thermodynamic approach to define the ideal dielectric behavior of a mixture. The electric polarization results in the reorientation of the sample dipoles and is therefore equivalent to an energy change of the system, allowing the application of thermodynamic formulation. According to this approach, the ideal permittivity of the mixture is the volume-fraction weighted average of each component permittivity (1) [15], being  $\phi_{A-B}$  and  $V_{A-B}$  the respective volume fraction and volume of each liquid.

$$\begin{aligned} \epsilon_{Ideal} &= \epsilon_A \phi_A + \epsilon_B \phi_B \\ \phi_A &= \frac{V_A}{V_B + V_A} = 1 - \phi_B \end{aligned} \quad (1)$$

The ideal value assumes a linear, isotropic, and homogeneous medium between the parallel plates of a capacitor [16]. The excess between the ideal and the measured permittivity is strongly related to the deviation from the ideal thermodynamic behavior. Peon Iglesias [1] proposed a method to estimate this excess and concluded that it is mainly attributable to the dielectric contrast ( $\epsilon_A/\epsilon_B$ ) and the effect of the molecular interaction through electrostatic bonds and induced dipoles, which none of the traditional formulas consider. The excess due to variations in the void volume of the mixture is apparently less significant. Every binary mixture is indeed the mixture of each substance and the free space between them. This volume can vary due to the rearrangement of the molecules and interactions upon mixing. If the void volume increases, the dipole density is sparser; if it decreases, the dipoles condense. These findings support the idea of modifying the mixing formulas to model somehow the excess. For example, Puranik et al. [17] proposed a modification formula (2) to replace the solvent volume fraction ( $\phi_A$ ) by an effective volume fraction ( $\phi_{A_{eff}}$ ) through an interaction parameter ( $k_{int}$ ).

$$\phi_{A_{eff}} = [k_{int} * (1 - \phi_A) + \phi_A] * \phi_A \quad (2)$$

$$\phi_{B_{eff}} = 1 - \phi_{A_{eff}}$$

The interaction parameter  $k_{int}$  aims to synthesize how the intramolecular interactions rearrange the molecules in the mixture structure, changing the volume fraction. Moreover, this effect should be practically null in non-polar liquids, giving a  $k_{int}$  value close to 1 [17]. This is consistent with the insights from the bibliography introduced so far, which question the suitability of the volume fraction as a parameter to estimate the mixture permittivity beyond isotropic, ideal, or homogeneous liquids [15]. The modification formula was tested to fit the predictions of Bruggeman's formula to the permittivity measurements of water-alcohol mixtures. Unfortunately, the authors did not provide a deeper theoretical justification for this modification. However, it has been tested by Amooey et al. [7] and Sarami et al. [8] applying the modification to compare an extensive list of mixing rules, with positive results: all the formulas fit their experimental data better. Nevertheless, this methodology has two drawbacks. Firstly, the parameter  $k_{int}$  is calculated with an optimization algorithm in order to reduce the error between the experimental and mixing rules values. Therefore, the reliability of this estimation depends on the quality of the data and the number of data points measured for each concentration. If data are scarce, there is a risk of overfitting, and if the data contain systematic errors, the adjustment will also be wrong. Secondly, each mixture has a unique  $k_{int}$  value, and consequently, the modification formula has no predictive value for new mixtures; therefore, it serves only to validate known permittivity mixtures.

The first objective of this paper is to present a methodology based on Machine Learning models to calculate the  $k_{int}$  value and the effective volume fraction without the risk of overfitting. The second objective is to propose a formula to estimate  $k_{int}$  requiring only pure liquid properties to fulfill the need for a general predictive formula of binary mixtures.

### 2.3. The Machine Learning approach

The starting point for this paper is our previous research [2], in which a Dielectric Resonator (DR) sensor was used to classify glycerine solutions. While techniques such as the open-ended coaxial probe acquire a direct permittivity measure, DR sensors detect substances measuring variations in its resonance frequency, which is closely related to medium permittivity around the sensor. When a liquid sample is dropped in a small cavity on top of the sensor, there is a change of permittivity resulting in a resonance frequency shift that can be detected. Therefore, the signal acquired is an indirect measure of the liquid permittivity within the narrow frequency range of the sensor resonant behavior. This change in technology simplifies the measuring protocol, enabling a faster acquisition and reducing the liquid sample volume. In addition, the DR sensor is designed to be portable, made of economical ceramic materials, and used alongside low-cost electronics. Although the sensor cannot directly measure the permittivity value for each concentration, the signals clearly indicate permittivity variations [2].

The initial objective was to develop a regression model that correlates the reflected DR signal to the corresponding permittivity value of the solution. However, achieving a correct regression prediction requires knowing with precision the real permittivity values as the predictive variable. Which brings back the key issue of this paper: what is the permittivity of a binary mixture? Given the lack of dielectric data for many chemical and biomedical solutions of interest, this can be a critical issue. To the authors knowledge, there is just one reference with glycerine-water permittivity measures in the frequency range of the sensor [18] (Table 1). Consequently, without other experiments to contrast these values, how far is it reasonable to trust them for a regression model? This problem demands either more experimental data, which is not always possible, or theoretical or computational models to contrast with the experimental data.

**Table 1**

Permittivity of glycerin-water solutions extracted from [18] and [25].

Medium	$\epsilon_{\text{experimental}}$	Medium	$\epsilon_{\text{experimental}}$
Air	1.00	Glycerine 40%	58.78
Glycerine 80%	17.00	Glycerine 30%	65.25
Glycerine 70%	27.45	Glycerine 20%	69.23
Glycerine 60%	39.00	Glycerine 10%	74.32
Glycerine 50%	51.55	Water	78.30

Unfortunately, as was briefly discussed in the previous section, the use of suboptimal formulas is mostly forced in the case of liquid mixtures. However, as was stated before, achieving a correct regression requires knowing the real permittivity values, which implies that the regression error will be lower when more realistic permittivity values are considered within the model. In contrast, the regression error will increase when these values deviate more from reality. Using this methodology, this paper aims to evaluate both the experimental data and the formula predictions. In addition, the optimal interaction factor ( $k_{int}$ ) could be estimated by minimizing the regression error. Unlike the works of Amooey et al. [7] and Saramy et al. [8], optimization algorithms such as Levenburg-Marquarts cannot be employed since the DR sensor signals do not directly give the mixture permittivity. Thus, the  $k_{int}$  cannot be fitted to match the mixing rules estimations with the DR data.

It must be noted that the success of this method depends on how this error is measured and the regression technique selected, considering that DR signals are high-dimensionality data. Linear regression is a standard method to analyze the relationship between independent variables or features and a dependent variable, such as the mixture permittivity. As a result, the regression model can predict the outcome of new input data, even interpolating unseen data points [19]. Nevertheless, linear regression is not able to fit non-linear relationships between the dependent and independent variables and requires more advanced methods such as LASSO or Ridge regression, which enable polynomial fitting of m-degree without the risk of overfitting [20]. However, the main disadvantage of traditional regression is the parametric approach: any statistical modeling formalizes the relationships between variables in the form of mathematical equations. In the case of non-linear relationships, finding a useful model could be challenging, especially when the number of features or independent variables is too large, as in any signal analysis problem.

Machine Learning (ML) is a set of computational algorithms that aim to imitate human learning by developing flexible data-driven models for identifying complex patterns, including non-linear relationships [21]. In contrast to linear regression, ML models assume a non-parametric approach where the algorithm is trained to resolve an optimization problem to fit the data with the model architecture. In a parametric model, the number of parameters is limited, and the optimization only finds the set of values that best fits the data. By contrast, in non-parametric approaches, the model complexity and internal parameters change according to the training data [22]. As a result of this flexibility, ML models are able to approach a more complex feature space without human intervention, just learning from the data structure. For this reason, we propose ML regression as the most suitable approach for permittivity regression from DR signals. Our previous work [2] tested the first concept of this approach, but without applying the volume fraction modification [17], and considering only the Maxwell-Garnett (MG) mixing rule [23] (3) to compare and validate the experimental reference with the  $\epsilon_{\text{Maxwell-Garnett}}$  estimation.

$$\epsilon_{\text{Maxwell-Garnett}} = \epsilon_B + 3\epsilon_B\phi_B \frac{\epsilon_A - \epsilon_B}{\epsilon_A + 2\epsilon_B - \phi_B(\epsilon_A - \epsilon_B)} \quad (3)$$

We found that the permittivity value differed drastically between the MG mixing rule and the experimental reference (Table 1), and it was not possible to determine a priori if either was correct. A Support Vector

Regressor (SVR) model was trained for each permittivity set to estimate the mixture permittivity. The model performance differed, with a Root Mean Square Error (RMSE) of 1.119 for the experimental values and 2.091 for the Maxwell-Garnett estimated values, pointing to the mismatch between this formula with the experimental values. In addition, while the experimental error was uniformly distributed, the Maxwell-Garnett permittivity (3) was highly concentrated in the low glycerine concentrations. This different pattern in the error distribution indicates that SVR models can validate the fitness of a mixing rule. In this research, this methodology is extended for other mixing rules to perform a comparative analysis of mixing rules for glycerine solutions with and without the modification formula proposed by Puranik et al. [17].

### 3. Methods

#### 3.1. Dataset description

The main dataset for this work was collected during our previous research [2], testing a microwave Dielectric Resonator (DR) sensor for classifying glycerine solutions. The dataset is available in a GitHub repository and includes 180 reflected signals acquired with a low-cost electronic reader based on Arduino and designed by the authors [24][dataset]. The dataset contains air and nine distinct glycerine concentrations from 0% to 80% in 10% intervals. In order to correlate the sensor response with the permittivity of these concentrations, reference values are extracted from the following studies: Meaney P. et al. [18] for glycerine solutions at 21 °C and Ellison W. et al. for pure water at 20 °C [25] (Table 1). Other data extracted from bibliographic references to validate our approach include: the interaction factors fitted by Puranik for water-alcohol mixtures [17] (Table 3); the permittivity values at several concentrations for anisole-alcohol mixtures [26], propanoic acid mixed with ethyl acetate, ethyl benzoate, and ethyl acrylate [27]; and mixtures of toluene with methanol, ethanol, and propanol [28].

#### 3.2. Machine Learning techniques to predict the mixture permittivity

Due to the high dimensionality of the DR signals, feature reduction techniques are required to improve the training efficiency of ML models. Principal Component Analysis (PCA) is a mathematical technique that detects the directions with greater variability within the data structure and creates a new mathematical space where each signal is projected. This method preserves and condenses the information of each signal in a vector defined by the axes of the new PCA space, which are called principal components [29]. This vector will be the input feature (X) used to train the ML regression model, while the permittivity estimation will be the training target (y) (Fig. 1). For the glycerine-water mixture dataset, the number of principals was established at seven, which condense 95% of the explained variance within the dataset.

Prior screening showed that the most optimal ML model for permittivity regression was Support Vector Regressor (SVR), an extension of the Support Vector Machine (SVM) classification model. While the SVM finds the hyperplane in the feature space, which maximizes the separation between classes [30], the SVR calculates the hyperplane with a flexible error margin that defines the volume containing most of the training data. Thus, the SVR predictions will be inside this volume, which is optimized through model training to make it as narrow as possible [31]. This training is tuned by two hyperparameters: C, which defines the hyperplane exclusion margin, and  $\gamma$ , which regulates the influence distance of a single training point.

The SVR models were implemented with Python using the scikit-learn library [32]. The radial basis function kernel was selected to learn the non-linear relationships within the dataset. Bayesian optimization implemented with scikit-optimize library was used for the hyperparameter tuning [33]. This search algorithm found through training iterations that the hyperparameter combination with the lowest RMSE was  $C=7500$  and  $\gamma=0.003$ . Besides dividing the dataset into train and test

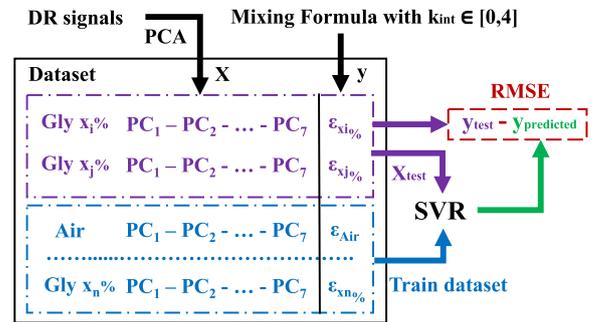


Fig. 1. Summary of the ML approach for a given mixing formula considering a  $k_{int}$  value between 0 and 4. This protocol will be repeated for each formula.

to prevent overfitting (70%-30%), the K-fold Cross-Validation method was applied during the model training with  $k=5$  folds [34].

A set of SVR models were trained to fit the permittivity estimations given by each mixing rule, considering an interaction factor between 0 and 4. After training, each SVR model is fed with the test dataset to generate predictions ( $y_{predicted}$ ) (Fig. 1). The RMSE measures the model accuracy with respect to the test permittivity values ( $y_{test}$ ). Therefore, this score evaluates the performance of each mixing rule for a particular  $k_{int}$  interaction value.

#### 3.3. Mixing rule selection

As the literature indicates, finding the best mixing rule may be not possible since it could depend on the binary mixture. Thus, the aim of this work is proposing new methods to enhance any mixing rule. The mixing rule selection to show the fitting improvement is a subset from the selection made by Sarami et al. [8]. In a preliminary assessment, the whole was tested but all mixing rules show a similar trend but with displaced values, including high-order mixing rules such as Oster [35] or Onsager-Bottcher [36], therefore; for clarity, just the first-order equations were considered. Other mixing rules, such as the Bruggeman asymmetric [37] or Bottcher-Bordewijk [38], achieved a good fitting; however, they were not included to avoid an excessive overlap that made the figures unclear. The final mixing rule selection is a representative set that cover the range of values between the upper a lower bound established by the Ideal (1) and Wiener (9) mixing rules. Besides the Maxwell-Garnett (3) and Ideal (1), the following mixing rules were considered: Hashin-Strikman (4) [39], Looyenga (5) [40], Peon-Iglesias (6) [41], Lichtenecker-Rother (7) [42], Kraszewski (8) [43], Wiener (9) [44].

$$\frac{(\epsilon_{Hashin-Strikman} - \epsilon_A)}{(\epsilon_{Hashin-Strikman} + 2\epsilon_B)} = \frac{\phi_A(\epsilon_A - \epsilon_B)}{3\epsilon_A} \quad (4)$$

$$\epsilon_{Looyenga} = [\epsilon_A^{1/3} + \phi_B(\epsilon_B^{1/3} - \epsilon_A^{1/3})]^3 \quad (5)$$

$$\epsilon_{Peon-Iglesias} = \epsilon_{Ideal} * [1 - \frac{2}{3} \ln \frac{1 + \phi_A(\frac{\epsilon_A}{\epsilon_B}) - 1}{(\frac{\epsilon_A}{\epsilon_B})^{\phi_A}}] \quad (6)$$

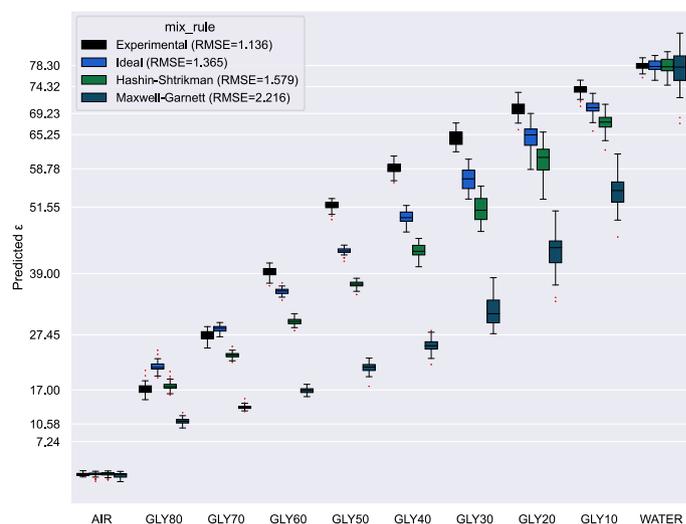
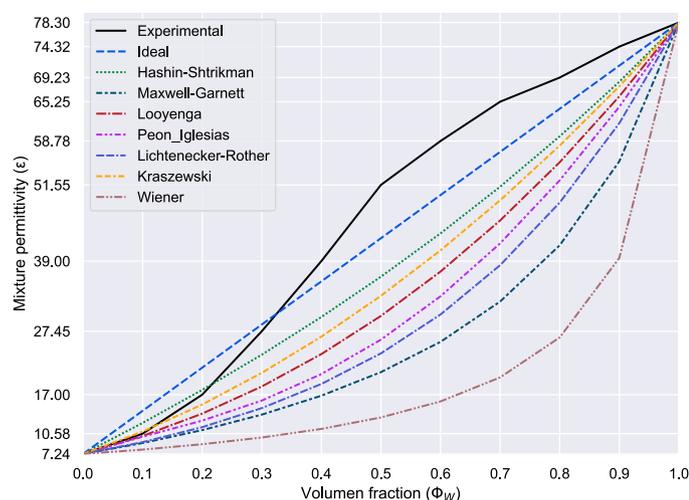
$$\epsilon_{Lichtenecker-Rochter} = \epsilon_A^{\phi_A} * \epsilon_B^{\phi_B} \quad (7)$$

$$\sqrt{(\epsilon_{Kraszewski})} = \phi_A * \sqrt{(\epsilon_A)} + \phi_B * \sqrt{(\epsilon_B)} \quad (8)$$

$$\frac{1}{\epsilon_{Wiener}} = \frac{\phi_A}{\epsilon_A} + \frac{\phi_B}{\epsilon_B} \quad (9)$$

#### 3.4. The Gaussian modification and a novel formula to estimate the interaction factor

However, the proposed methodology still demands a significative amount of data acquisition to train the ML models and find the optimal interaction factor. Therefore, to avoid the need for experimental data,



**Fig. 2.** (a) Comparison of the mixing rules permittivity estimations with the experimental values. (b) Boxplot distribution of permittivity predictions from SVR models trained with different mixing rules.

$k_{int}$  should ideally be estimated with a formula requiring only pure liquid properties, enabling a priori prediction of the mixture permittivity. According to Puranik et al. [17] the change in the effective volume fraction is due to the rearrangement of the molecules within the mixture. This paper proposes that this rearrangement depends on the dielectric strength but also the size of the molecules involved, and therefore  $k_{int}$  could be estimated as a function of the dielectric contrast and molar volume relation between both liquids (10). Consequently, this estimation formula considers that the interaction factor is not a constant since the permittivity depends on the frequency of the incident wave. In addition, it also depends on thermodynamic conditions since the molar volume is proportional to the liquid density, which depends on variables such as temperature or pressure.

$$k_{int} = \frac{\epsilon_A * V m_A}{\epsilon_B * V m_B} \quad (10)$$

To prove this hypothesis, the  $k_{int}$  estimated for glycerin-water mixture is compared with the  $k_{int}$  optimal values obtained with the SVR models for each formula, but also with the interaction factor calculated by Puranik [17] for several polar binary liquids. Moreover, in order to test the whole methodology, additional measurements with DR sensor were taken for water-*PEG*<sub>400</sub> and water-acetone mixtures. Afterwards, the best SVR models trained with the glycerin-water dielectric data and the best mixing rule using our  $k_{int}$  estimation (10) will predict the permittivity of these unseen samples. Finally, the model predictions will be contrasted with the values given by the mixing rule itself. On the other hand, as Puranik et al. state in their article, the volume fraction modification proposed (2) is not based on any theoretical justification [17]. This is a critical issue since all the methods rely on the soundness of the formula. Therefore, the mathematical domain and the fitness with the physical conditions of the problem will be analyzed. After studying its limitations, a new Gaussian modification formula is proposed.

## 4. Results and discussion

### 4.1. Mixing rules comparison

Fig. 2a shows the estimated permittivity of the chosen mixing rules for the glycerin-water mixture. None of them fit either the extracted experimental values from [18], or the curve shape. The Ideal mixing rule (1) has the best approximation, despite being a simple line crossing the curve of the experimental values. From the Hashin-Shtrikman (4) to the Wiener (9) mixing rules, the estimated values distance progressively from the experimental values. The Wiener mixing rule not only

**Table 2**

Comparative performance between standard and modified formulas.

Mixing Rule	Initial RMSE	$k_{opt}$	$RMSE_{min}$ $RMSE_{min+5\%}$	$[k_{opt-range}]$
Ideal (1)	1.365	2.29	0.877 0.921	1.93-2.66
Looyenga (5)	1.820	2.17	1.182 1.241	1.91-2.49
Peon-Iglesias (6)	1.978	2.21	1.304 1.369	1.87-2.45
Lichtenecker-Rother (7)	2.102	1.99	1.461 1.534	1.81-2.37
Kraszewski (8)	1.696	2.23	1.080 1.134	1.93-2.54
Hashin-Shtrikman (4)	1.579	2.23	1.016 1.067	1.93-2.52
Maxwell-Garnett (3)	2.216	1.88	1.69 1.77	1.63-2.12

differs the most from the experimental curve, showing an unrealistic inverted shape, but also, it achieved the worst ML performance, and it was excluded from the discussion for clarity.

The divergences from the experimental curve are correlated with the performance of the trained SVR models (Table 2). The lowest RMSE is reached for the model trained with the experimental values, followed by the Ideal formula (1). As the mixing rule estimation differs from the experimental values in Fig. 2a, the RMSE increases. The Maxwell-Garnett mixing rule (3) used previously in [2] has the second highest RMSE. These results are represented in the boxplot of the predicted permittivity for each concentration (Fig. 2b). As is clearly depicted, lower RMSE implies narrower boxes and a lower standard deviation.

### 4.2. Interaction factor evaluation

Fig. 3 shows the RMSE as a function of the interaction parameter  $k_{int}$ . The experimental set represents a threshold value since it is not modified. All the mixing rules present a similar trend, reaching a minimum RMSE in an optimal region of similar  $k_{int}$  values (Table 2). This insight indicates that the interaction factor is a mixture property with a similar impact on each mixing rule. The mixing rules keep the same performance order described in the previous section. The Ideal formula (1) achieves the lowest RMSE, even surpassing the performance of the experimental value set. The Hashin-Shtrikman (4) and Kraszewski (8)

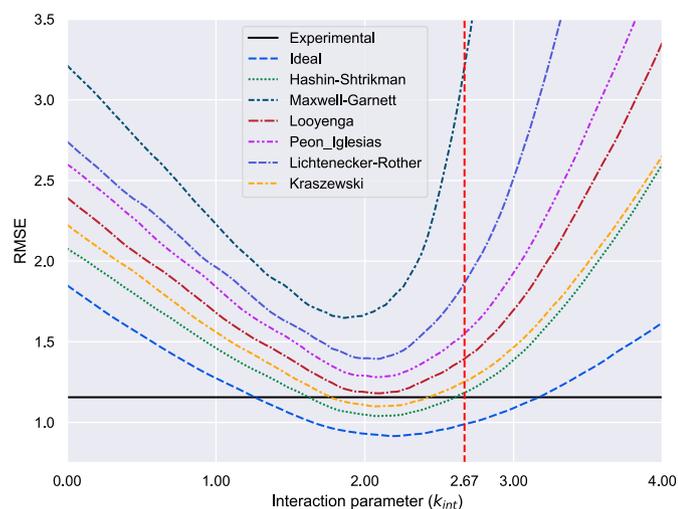


Fig. 3. RMSE variation of SVR models trained with different  $k_{int}$  values for each mixing rule. The vertical red line corresponds to the  $k_{int}$  estimated with equation (10).

mixing rules also achieved a lower RMSE error. This overperformance may indicate that the considered experimental values to train the SVR could be affected by some measured error.

Table 2 shows that after applying the Puranik modification [17] each mixing rule has a significant RMSE reduction of around 0.5-0.6. However, even with this volume fraction modification, the Maxwell-Garnett mixing rule (3) is still achieving suboptimal results and therefore has been excluded from the following discussions. Considering the remaining mixing formulas, the optimal  $k_{int}$  has a mean value of 2.18. According to the parabolic relationship between RMSE and  $k_{int}$  (Fig. 3), the RMSE variation is decreasing around the minimum. Considering a 5% error increment ( $RMSE_{min+5\%}$ ), an optimal range with a low impact on the SVR performance can be defined between the  $k_{int}$  values of 1.9-2.5.

However, considering the mean  $k_{opt} = 2.18$ , which is the center of the optimal range, the new values estimated by the mixing rules do not fit the experimental values set either (Fig. 4a). But in contrast to the unmodified formulas, the new curves tend to surpass the experimental curve. Although the Peon-Iglesias (6) and Lichtenecker-Rother (7) mixing rules achieved a higher RSME, they fit quite well, especially for volume fractions ( $\phi_B$ ) higher than 0.4. Unfortunately, the Ideal formula (1), which obtains the lowest RMSE, does not fit properly with the experimental values.

On the other hand, considering now a  $k_{int} = 1.9$ , at the lower limit of  $k_{int}$  optimal range, there is a significant improvement in the match between mixing rules estimations and experimental data (Fig. 4b). In this scenario, the Looyenga mixing rule (5) obtains the best balance of RMSE-fitting. Nevertheless, it is clear that there is not a pure correlation between the RMSE and the mixing rule fitness. The reason why could be that the regression error reduction is due to the label distribution rather than the label values. Therefore, similar curve shapes with different permittivity values will score a similar RMSE as in the case of Fig. 4a and Fig. 4b. As a consequence, depending on the  $k_{int}$  within the optimal range, this method will find the correct mixture permittivity distribution displaced in a narrow margin.

Following the insights, the boxplot in Fig. 5 compares values 1, 1.9, and 2.18 of  $k_{int}$  for the Ideal formula (1), showing slight differences in standard deviation for each concentration. In contrast, Fig. 6 shows the fitness of the formulas with a similar RMSE to the experimental set. In this case, especially for low glycerine concentrations, the predicted permittivity from SVR models is quite similar, with no significant standard deviation differences by mixing rule. In any case, since the whole analysis is biased by the reliability of the experimental values extracted

Table 3

Summary of interaction factor values for several binary liquids compared with our formula estimation.

Mixture	$V_{m_A}   V_{m_B}$	$\epsilon_A   \epsilon_B$	$k_{int}$ [17]	$k_{int}   k_{int}^{-1}$ (10)
Methanol (A)	40.65	32.1	1.66	0.90
Water (B)	18.02	80.1		1.11
Ethanol (A)	59.53	24.5	1.53	0.99
Water (B)	18.02	80.1		1.01
Propanol (A)	74.79	17.9	1.33	0.93
Water (B)	18.02	80.1		1.08
Formamide (A)	33.79	111	-0.10	2.597
Water (B)	18.02	80.1		0.385
DMSO (A)	71.03	46.7	2.65	2.29
Water (B)	18.02	80.1		0.437
Acetophenone (A)	116.88	17.4	2.08	1.394
Ethanol (B)	59.53	24.5		0.717

in [18], whose margin of error is unknown, the low dispersion of SVR predictions could be within this error.

#### 4.3. A simple but an effective $k_{int}$ estimation

Despite the considerable improvement after applying the volume fraction modification, the main drawback of this methodology is still its dependency on experimental data to calculate the interaction factor. Therefore, estimating the interaction factor a priori would be a significant advantage. Using our estimation formula (10), the interaction factor for glycerin-water mixtures is close to the optimal range of  $k_{int}$  achieved by the SVR models (11).

$$k_{int} = \frac{\epsilon_{WAT} * V_{m_{WAT}}}{\epsilon_{GLY} * V_{m_{GLY}}} = \frac{78.3 * 18.02}{7.23 * 73.03} = 2.67 \quad (11)$$

$$k_{opt-SVR} = [1.9, 2.5]$$

In addition, the  $k_{int}$  values fitted by Puranik [17] for several binary mixtures were compared with the estimation given by the proposed formula (10). The results seem promising, with a huge level of concordance between both methods (Table 3). However, more comparisons should be made to gather more evidence that our formula could give correct estimations.

#### 4.4. Analyzing the limitations of the volume fraction modification. A new Gaussian equation to improve permittivity estimations

The core of the previous method is the volume fraction modification proposed by Puranik et al. [17], which is not based on any theoretical justification as is stated by the authors. Therefore, its suitability depends on how well the mathematical formulation (2) is able to model the volume fraction change due to the interaction between liquids. Under the assumption of the proposed estimation formula (10), the interaction factor is the relationship between the pure liquid properties that define the molecular arrangement. Therefore, if  $k_{int} = 1$ , the volume fraction does not change since there is an equilibrium between the properties of both liquids. The Puranik modification satisfies this condition, and Fig. 7(a) shows how the effective volume fraction changes when the equilibrium is broken for several initial solvent volume fractions ( $\phi_A$ ). Thus, if  $k_{int} < 1$ , the solute liquid (B) will have more influence over the solvent liquid (A), and  $\phi_{A_{eff}}$  will decrease due to the molecular packing exerted by liquid B. In contrast, if  $k_{int} > 1$ , liquid A will have more influence, and it will expand to compress liquid B, resulting in an increment of its effective volume fraction ( $\phi_{A_{eff}}$ ). The Puranik modification shows that these volume fraction variations follow a linear behavior (Fig. 7(a)). Nevertheless, if  $k_{int} > 2$  with this modification formula, the effective volume fraction will surpass value 1, at a speed depending on the initial volume fraction of liquid A. This behavior cannot be realistic since  $\phi_A + \phi_B = 1$  implies that negative values of  $\phi_B$  do not have

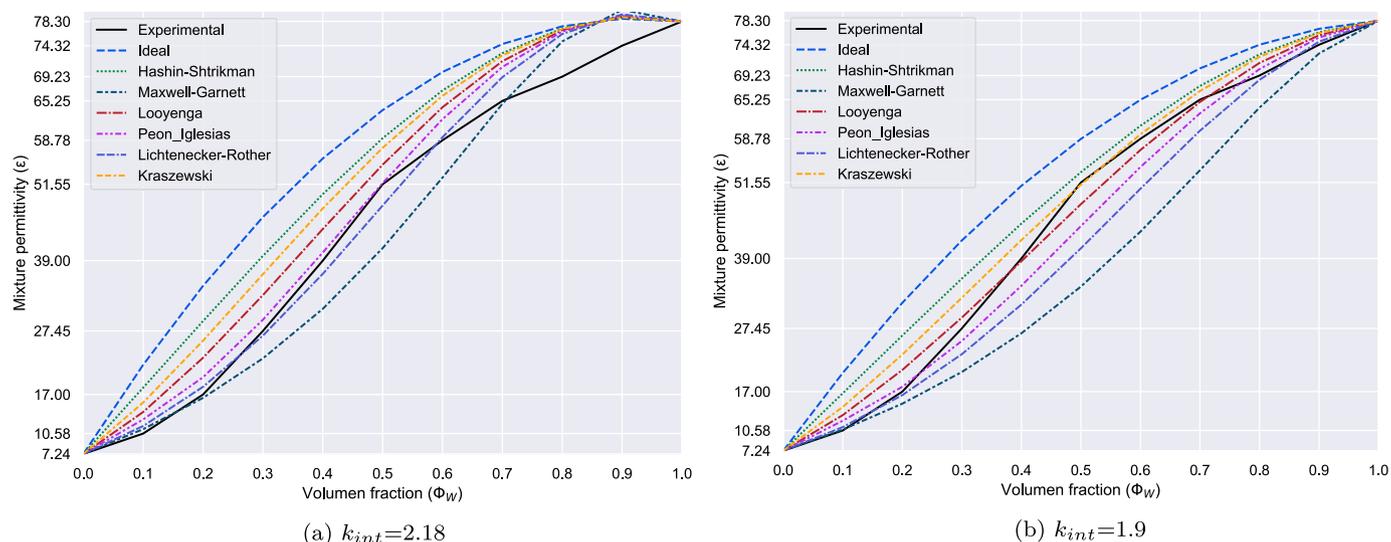


Fig. 4. Comparison of modified mixing rules.

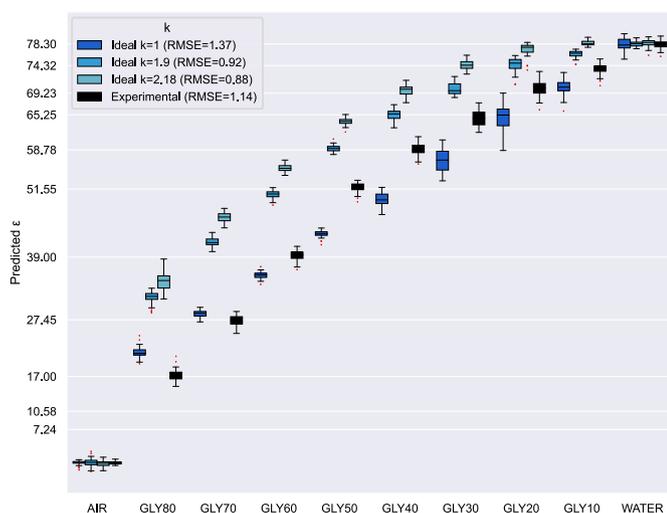
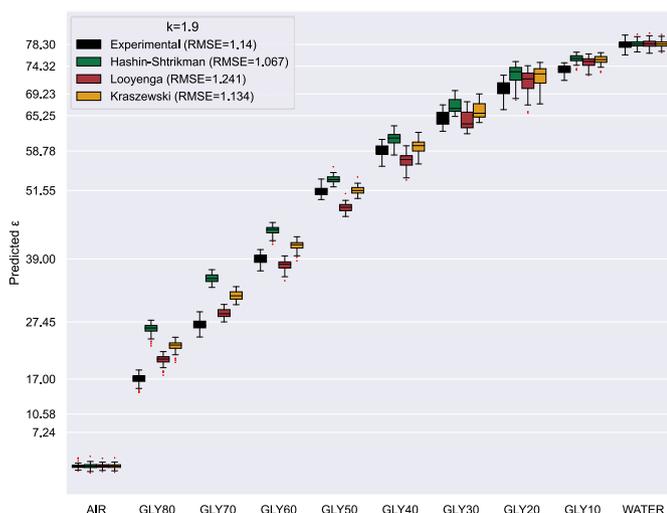


Fig. 5. Boxplot distribution of permittivity predictions of SVR models trained with different modifications of the Ideal mixing rule (1).

Fig. 6. Boxplot distribution of permittivity predictions of SVR models trained with different modified mixing rules with  $k_{int} = 1.9$ .

a physical mean. This insight is a serious concern about applying the Puranik modification for high interaction factors and high volume fractions of solvent.

After studying these limitations, we propose a new volume fraction modification based on a Gaussian formula that satisfies the convergence to a maximum volume fraction for a high interaction factor and the equality between the effective and initial volume fraction when there is no interaction ( $k_{int} = 1$ ). The maximum expansion of  $\phi_A$  due to molecular packing is unknown and should depend on solvent properties. Nevertheless, the value one is the physical limit, which was considered the maximum  $\phi_A$  under the assumption that the  $k_{int}$  estimation avoids unrealistic  $\phi_{A_{eff}}$  values due to the “overpacking” of liquid B.

$$\lim_{k_{int} \rightarrow 1} F(k_{int}, \phi_A) = \phi_A$$

$$\lim_{k_{int} \rightarrow \infty} F(k_{int}, \phi_A) = \phi_{A_{max}} = 1 \quad (12)$$

However, these conditions (12) are not fulfilled by one unique function, in fact there are infinite Gaussian curves defined by the parameter  $\mu$  in the exponent (13). This parameter tunes the steepness of the Gaussian curve alongside the volume fraction and interaction factor. It was determined that the optimal value for the glycerine-water mixture is  $\mu = 2$ .

$$\phi_{A_{eff}} = 1 - \phi_B e^{-\phi_A^\mu (k_{int} - 1)} \quad (13)$$

Fig. 7(b) shows that the new relationship between the interaction factor and the effective volume fraction is a smooth asymptotic curve for high-volume fractions of liquid A. In contrast, the relationship is mostly linear for lower fractions as with the Puranik modification (2) but with a lower slope. In addition, for  $k_{int} < 2$ , both modification formulas perform in a comparable range. It must be noted that the asymptotic steepness is for  $\mu = 2$ , but could be modified to enable a better adaptation to other liquid mixtures.

In the case of glycerine-water, applying the Gaussian modification formula with  $\mu = 2$  and  $k_{int} = 2.67$  improves the fitting for all the mixing rules (Fig. 8a). Compared to the Puranik modification (Fig. 4b), the curve shape of each mixing rule resembles more the sigmoid-like shape of experimental permittivity values. In addition, the Hashin-Shtrikman (4) and Kraszewski (8) are still the mixing rules that achieve the best fitting, as pointed out in previous insights. In summary, the combination of both proposed formulas, the Gaussian modification with the estimation of  $k_{int}$ , improves the dielectric behavior modeling.

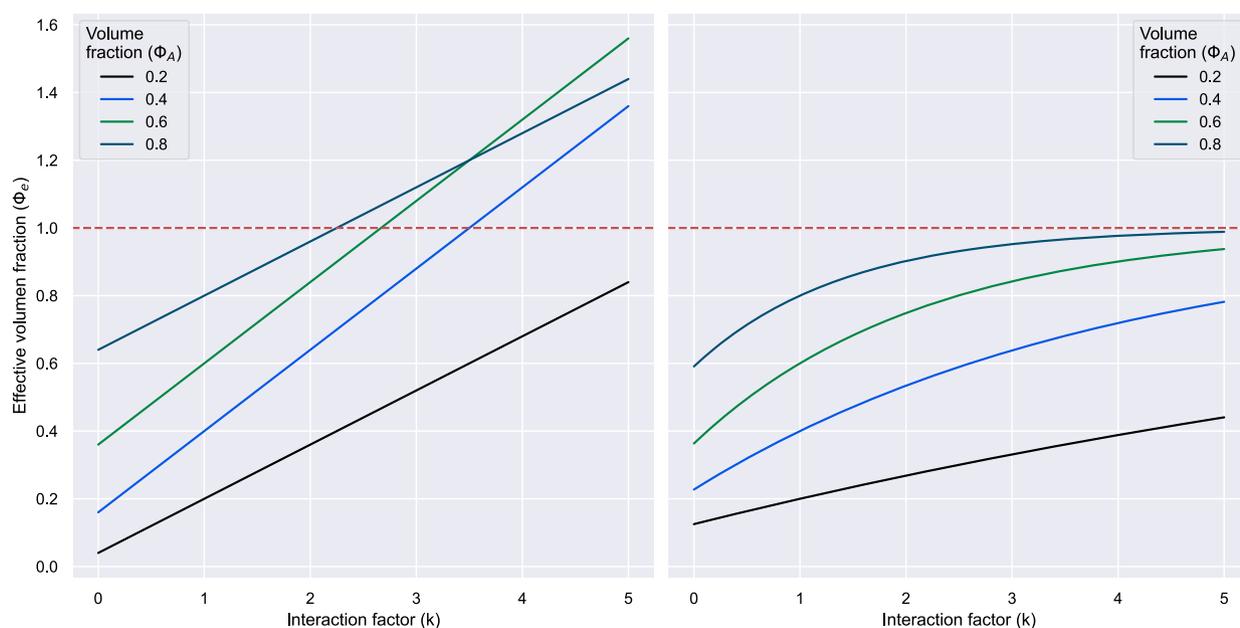


Fig. 7. Volume fraction variation as function of the interaction factor (a) Puranik et al. [17] modification (b) Gaussian modification (10).

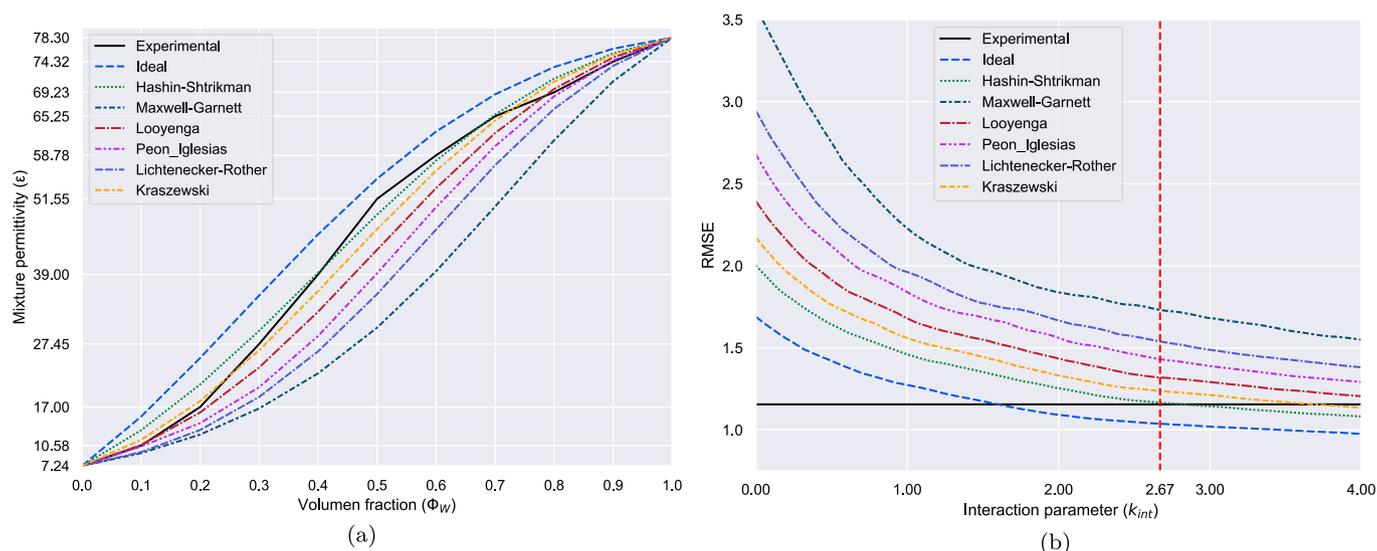


Fig. 8. (a) Comparison of modified mixing rules with  $k_{int} = 2.67$  and applying the Gaussian modification. (b) Asymptotic reduction of RMSE for SVR models trained with mixing rules applying the Gaussian modification. The vertical red line is the  $k_{int}$  threshold estimated by equation (10).

On the other hand, the Gaussian modification also changes the sensitivity of  $k_{int}$  on the regression error of SVR predictions. In contrast to the parabolic behavior from Puranik modification (Fig. 3), the RMSE is reduced asymptotically to a limit value (Fig. 8b). While the Puranik modification defines an optimal region, the Gaussian modification establishes a threshold value. Indeed, for the glycerin-water system, this threshold coincides with  $k_{int} = 2.67$ , estimated with the pure liquid properties (10), for the Hashin-Shtrikman (4) and Kraszewski (8) mixing rules. Below the threshold, the fitting worsens exponentially (Fig. 9a) due to the nature of the Gaussian equation. In comparison, for larger  $k_{int}$  values beyond the threshold, the fitting deviates slightly (Fig. 9b) while the RMSE tends to stabilize. Additionally, it must be noted that the mixing rules keep the same performance order as previous sections, suggesting that the mixing rule suitability could depend mostly on solvation mechanism of each liquid mixture.

#### 4.5. Validation with other liquid mixtures through other experimental data from bibliographic references

In order to evaluate if both proposed formulas could be applied to other liquid mixtures to obtain accurate permittivity estimations, dielectric data for several solutions were extracted from the following references: [26], [27], [28]. In addition, this data was used to estimate the  $k_{int}$  for each solution (Table 4).

The results vary between mixtures of anisole with several alcohols [26]; for the anisole-methanol solution, the experimental data and the modified mixing rules estimations are drastically outfitted (Fig. 10 (a)), while the anisole-hexanol mixture achieves a perfect fitting with the Hashin-Shtrikman (4) and Kraszewski (8) mixing rules (Fig. 10 (d)). Indeed, the fitting seems to worsen as the dielectric contrast increases, as is shown between the anisole-propanol and anisole-butanol mixtures in Fig. 10 (b-c). This insight could indicate that the influence of dielec-

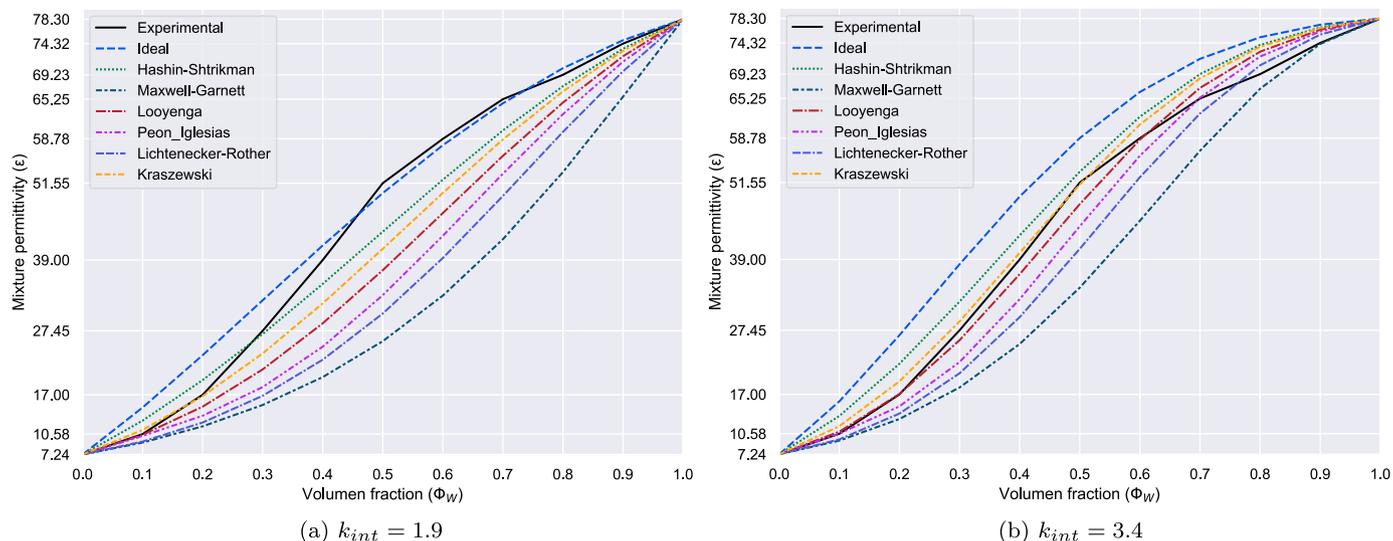


Fig. 9. Comparison of modified mixing rules after applying the Gaussian modification below (a) and above (b) the optimal threshold.

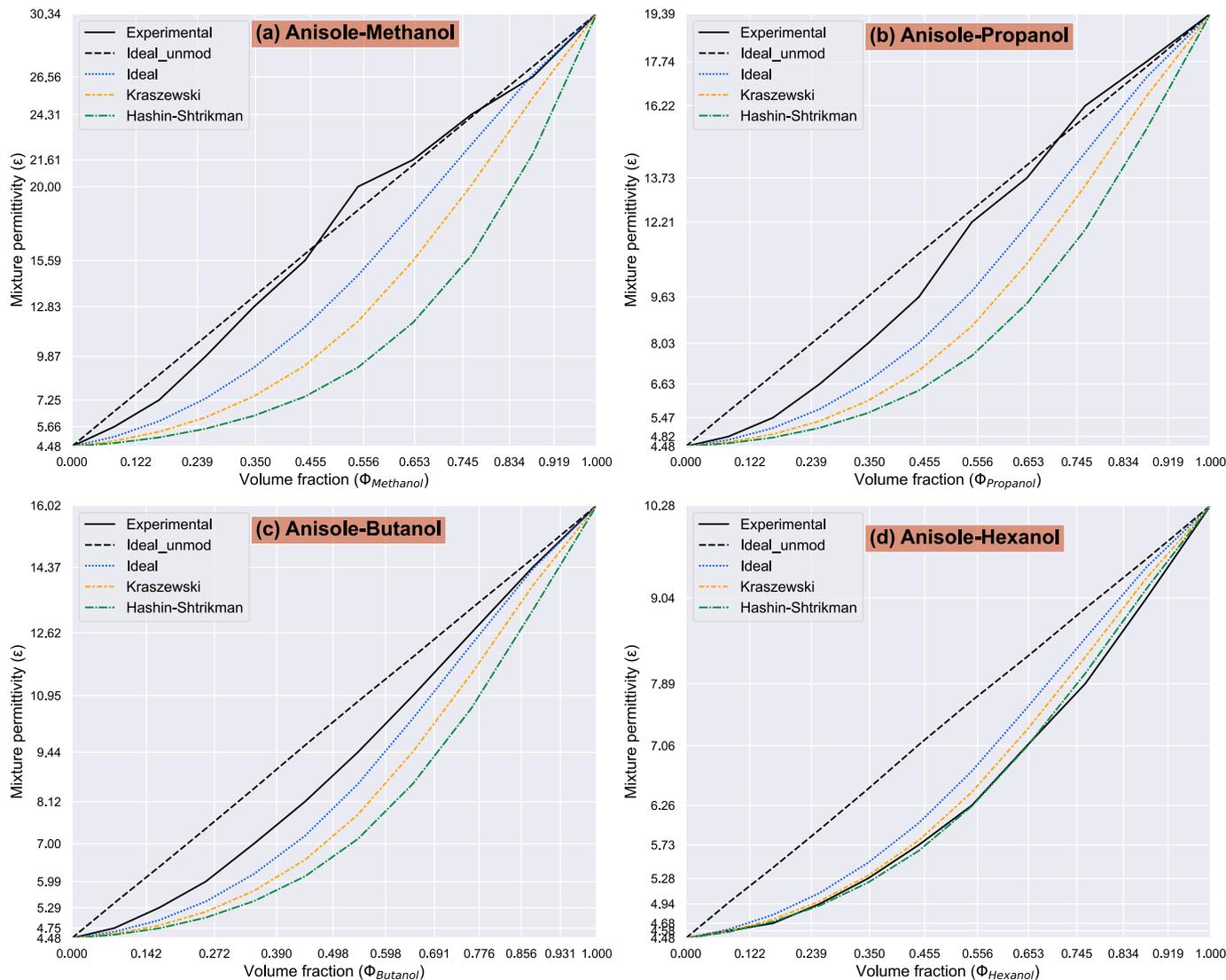


Fig. 10. Fitting of permittivity estimations for anisole mixtures after applying the Gaussian modification with  $\mu = 2$  (13).

**Table 4**  
Summary of interaction factor values for several binary liquids from the bibliography [23], [24], [25].

Mixture	$Vm_A   Vm_B$	$\epsilon_A   \epsilon_B$	$k_{EST}$	Mixture	$Vm_A   Vm_B$	$\epsilon_A   \epsilon_B$	$k_{EST}$
Methanol (A)	40.46	30.34	2.521	Propanoic acid (A)	74.98	3.36	0.296
Anisole (B)	108.68	4.48		Ethyl benzoate (B)	143.03	5.94	
Propanol (A)	74.8	19.39	2.98	Propanoic acid (A)	74.98	3.36	0.381
Anisole (B)	108.68	4.48		Ethyl acrylate (B)	106.45	6.21	
Butanol(A)	91.51	16.02	3.011	Toluene (A)	106.29	2.46	0.193
Anisole (B)	108.68	4.48		Methanol (B)	40.46	33.19	
Hexanol (A)	124.61	10.28	2.63	Toluene (A)	106.29	2.46	0.181
Anisole (B)	108.68	4.48		Ethanol (B)	59.53	24.26	
Propanoic acid (A)	74.98	3.36	0.428	Toluene (A)	106.29	2.46	0.175
Ethyl acetate (B)	97.68	6.01		Propanol (B)	74.793	19.96	

tric contrast should be adjusted to estimate  $k_{int}$ . However, the dielectric contrast between glycerine-water is greater than anisole-methanol, yet the adjustment worked perfectly. Likewise, the interaction factor could also depend on the solvation mechanisms and therefore, a formula to estimate  $k_{int}$  should be designed to address how different types of compounds interact. Although our proposed formula (11) estimates  $k_{int}$  correctly for the glycerine-water mixture, improving the formula to include other parameters could lead to better results for other liquids.

Similar discrepancies can be found in the rest of the liquid mixtures analyzed. While the fitting for propanoic acid with ethyl acetate, ethyl benzoate, and ethyl acrylate [27] was promising, the mismatch for the mixtures of toluene with methanol, ethanol, and propanol [28] was very significant. However, changing the exponent in the Gaussian modification formula can improve the fitting of the mixing rules to the experimental data. For example, after applying the equation with  $\mu = 3$  (14), the anisole-propanol and anisole-butanol estimations using the Ideal mixing rule (1) lead to a better fitting (Fig. 11 (b-c)). In contrast, this exponent change hinders the fitting of the anisole-hexanol mixture (Fig. 11 (d)). In fact, the increment of  $\mu$  moves the curve, approaching the unmodified ideal mixing rule. On the other hand, the estimations for the anisole-methanol mixture are slightly improved for high volume fractions, especially for the Ideal mixing rule (1), but it is still a bad fitting (Fig. 11 (a)).

$$\phi_{A_{eff}} = 1 - \phi_B e^{-\phi_A^3 (k_{int} - 1)} \quad (14)$$

$$\phi_{A_{eff}} = 1 - \phi_B e^{-\phi_A^{1.5} (k_{int} - 1)} \quad (15)$$

In the same way, the mixtures of propanoic acid with ethyl acetate, ethyl benzoate, and ethyl acrylate slightly improve their fitting with the Gaussian modifications with  $\mu = 1.5$  (15) (Fig. 12 (a-c)). In addition, for the mixtures of toluene-ethanol and toluene-methanol, the Gaussian modifications with  $\mu = 3$  (14) achieve a reasonable fitting with Looyen-ga's mixing rule (5) (Fig. 12 (d-f)).

These results reinforce the idea that the asymptotic steepness from the Gaussian modification could depend on the mixing behavior. As the literature indicates, the main drawback to achieving a general mixing rule is probably the different solvation processes, which are the key to defining the mixture microstructure and, therefore, the volume fraction and relative permittivity. Probably, the interaction factor cannot explain by itself the complexity of the mixture permittivity. Therefore, a multiparametric approach could be the best option to enhance mixing rules. Additional parameters, such as the factor  $\mu$ , must be considered to correlate the mixture permittivity with the solvation mechanics and liquid properties. This approach could be applied to other mathematical expressions beyond the Gaussian formulation. For example, another parameter could be included in the linear formulation proposed by Purnik et al. [17].

#### 4.6. Validation with other liquid mixtures through DR measures and SVR predictions

Most articles report dielectric measurements in static and near-optic frequencies, to address minimum and maximum relaxation, respectively. As a result, finding dielectric data at specific frequency ranges for many liquid mixtures is not always possible. Thus, the DR sensor, in combination with the Machine Learning approach, can be a low-cost method to validate mixing rules estimations and/or the modification formulas proposed in this article. To prove this concept, new measurements were performed with our DR sensor for  $PEG_{400}$  and acetone in water at two concentrations: 30% and 60%. The estimated interaction factor for the  $PEG_{400}$ -water mixture (16) differs from the glycerine-water mixture (11); although the  $PEG_{400}$  has a lower permittivity at 2.45 GHz [45] than glycerol, the molecular volume is significantly larger, resulting in a  $k_{int} < 1$ . On the other hand, the permittivity of acetone at 2.45 GHz [46] is four times lower than water, while the molecular volume ratio is just the inverse of the dielectric contrast. Thus, the interaction factor for the acetone-water mixture is close to  $k_{int} = 1$  (17), indicating that the volume fraction does not change upon mixing.

$$k_{int} = \frac{\epsilon_{WAT} * Vm_{WAT}}{\epsilon_{PEG} * Vm_{PEG}} = \frac{78.3 * 18.02}{6.7 * 354.61} = 0.594 \quad (16)$$

$$k_{int} = \frac{\epsilon_{WAT} * Vm_{WAT}}{\epsilon_{ACE} * Vm_{ACE}} = \frac{78.3 * 18.02}{19.4 * 74.034} = 0.982 \quad (17)$$

After applying PCA, the new mixture signals were analyzed using two SVR models trained with the glycerin-water mixture. The first was trained with the permittivity values from the experimental reference [18]. The second was trained with the estimations from the modified Hashin-Shtrikman (4) mixing rule. This formula was selected because it achieved the best performance for the glycerine-water mixture and, therefore, has the best correlation between the sensor response and the relative permittivity of the liquid sample. The predictions of these models for both liquid mixtures were compared with the estimations from the whole set of mixing rules considered in this paper. In agreement with previous results, there are significant differences between mixing rules depending on the liquid mixture and its solvation mechanisms. In the case of the  $PEG_{400}$  solution, the Ideal formula (1) achieved the best results Fig. 13a. In contrast, for the Acetone solution, the Kraszewski mixing rule (8) obtained the best fitting (Fig. 13b), with a trend quite similar to the results. Both SVR models considered for each binary mixture gave similar predictions; once again showing how well the modified Hashin-Shtrikman mixing rule (4) models glycerine-water permittivity.

Finally, it must be noted that these new measurements for acetone and  $PEG_{400}$  were taken a year after the glycerine-water dataset, which was used to train the SVR models. However, since the DR sensor was used for other projects during this period, the air reference signal without a sample has been displaced. This sensor drift was detected after

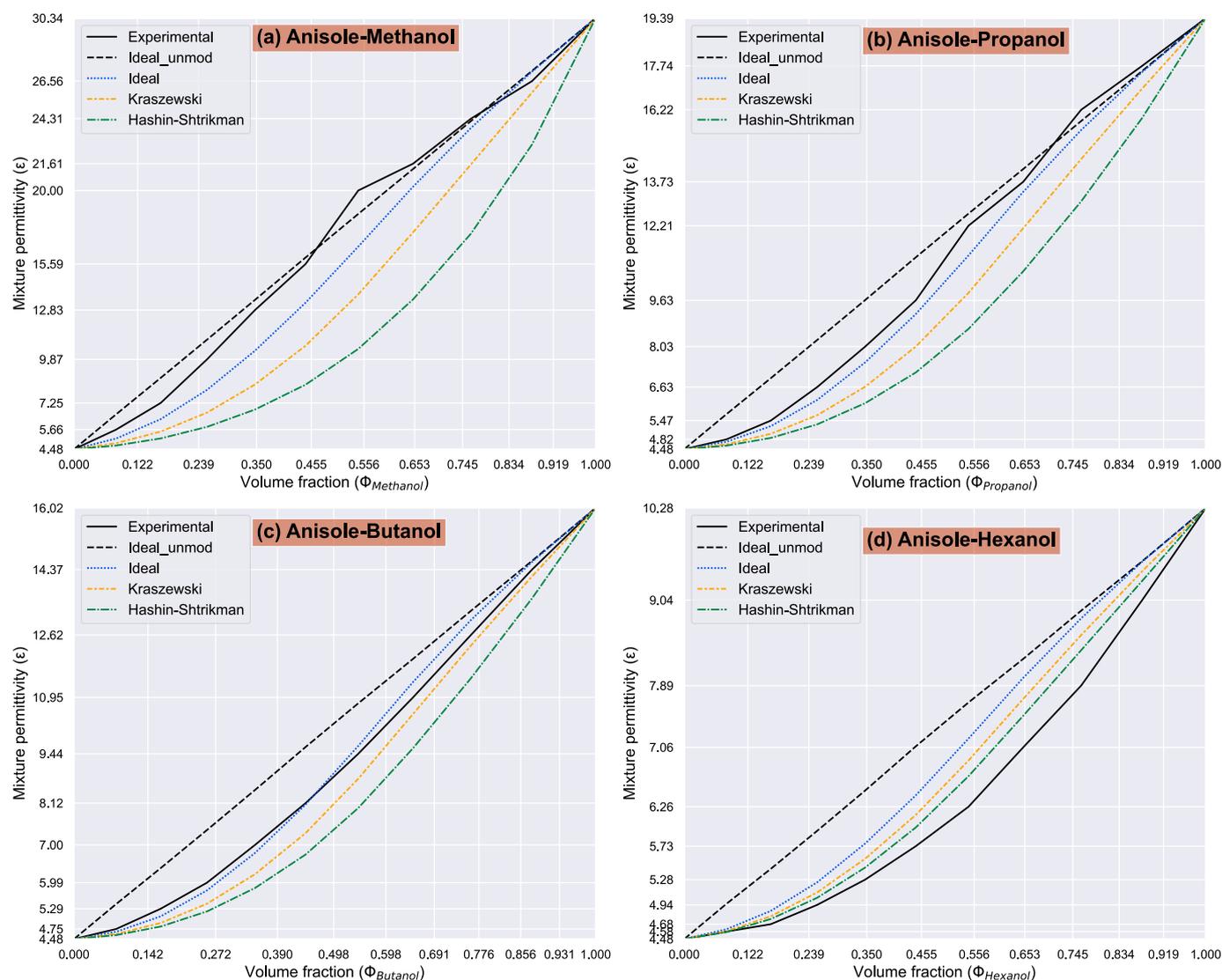


Fig. 11. New fitting of permittivity estimations for anisole mixtures after applying the Gaussian modification with  $\mu = 3$  (14).

analyzing the results. Consequently, the predictions for the air reference present a small shift from the real value (Fig. 13a, Fig. 13b); this deviation must affect the rest of the measurements. Therefore, SVR predictions for water- $PEG_{400}$  and water-acetone could be adjusted by reducing the permittivity predictions around two points, fitting the reference air prediction. This modification would slightly improve the match between the mixing rule estimated  $\epsilon$  and the SVR predicted  $\epsilon$  for acetone 60% and both  $PEG_{400}$  60% and 30%, while acetone 30% deviates slightly. However, despite the sensor drift, the original results were not modified due to its qualitative information about the potential of the ML approach to validate mixing rules estimations.

## 5. Conclusion

Determining the dielectric behavior of a liquid solution is still an unsolved problem due to the high-complexity relationships between the liquid microstructure and the solvation mechanics. Traditional mixing rules must be modified to include this complexity in order to enhance their accuracy. The literature has proved that considering volume fraction change caused by the molecular interaction can improve the fitting of mixing rules estimations to the experimental values. This paper presents a novel approach to calculating this interaction factor ( $k_{int}$ ) by

applying ML regression models trained with glycerin-water reflective signals acquired with a Dielectric Resonator sensor.

Modifying the mixing rules by applying the volume fraction modification improved both the SVR performance and the fitting between estimations and experimental values. Therefore, the model RMSE can indicate both the optimal range of  $k_{int}$  and the mixing rule that obtains better estimations for the studied binary mixture. However, this methodology is limited to validation and is not suitable for a priori estimations, since it depends on acquired dielectric data. Therefore, we propose an estimation formula for  $k_{int}$ , considering only the molar volume of each liquid and their dielectric contrast ( $\epsilon_A/\epsilon_B$ ), which approximately fits both the SVR results and the  $k_{int}$  estimated by Puranik et al. for water-alcohol mixtures [17].

After studying the mathematical domain of the Puranik et al. modification [17], we proposed an alternative method to calculate the effective volume fraction through a Gaussian function. This new modification formula fulfills the conditions to obtain a physically realistic volume fraction for any  $k_{int}$ . For the glycerine-water mixture, combining the formulas proposed in this article improved both the experimental fitting of the Puranik modification [17] and the match between the minimum RMSE and  $k_{int}$ . Afterward, the suitability of the Gaussian modification was tested for several binary mixtures, achieving an optimal fitting for most of them. However, the Gaussian modification

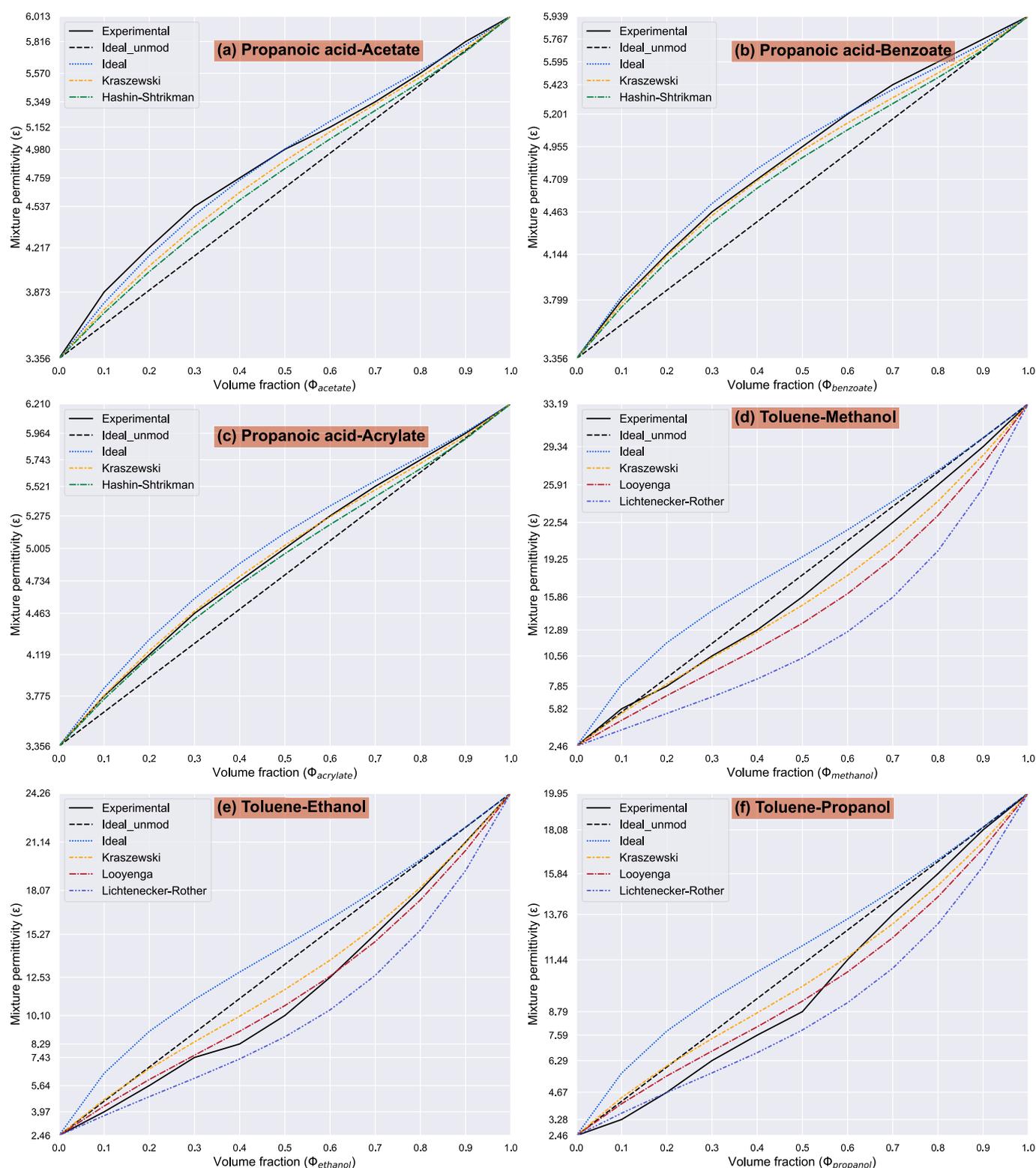


Fig. 12. Fitting of permittivity estimations for Propanoic acid mixtures (a-c) after applying the Gaussian modification with  $\mu = 1.5$  (15) and Toluene mixtures (d-f) after applying  $\mu = 3$  (14).

required the additional parameter  $\mu$  to model the dielectric behavior. The multiparametric approach can help to improve the understanding of the mixture dielectric behavior. However, the relationship between the proposed formulas and the nature of the molecular interaction and the resulting liquid microstructure is still unknown and demands further studies.

Finally, the ML approach was also tested for acetone-water and  $PEG_{400}$ -water mixtures, proving that ML models trained with a specific mixture, such as glycerine-water, can obtain reasonable permittivity predictions to validate the mixing rules estimations for other mixtures. In conclusion, the ML approach is a promising method to validate any permittivity characterization. Using the SVR models, the interaction fac-

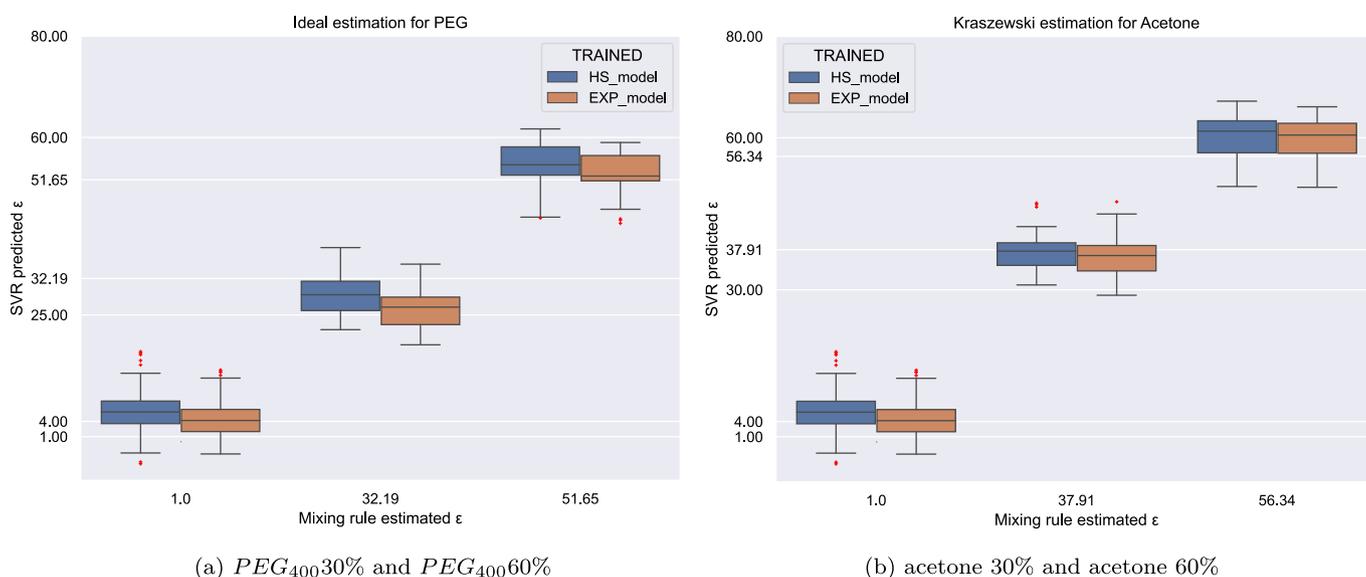


Fig. 13. Distribution of SVR permittivity predictions compared to the (a) Ideal (1) and (b) Kraszewski (8) estimations.

tor can be estimated for non-direct permittivity measures from resonant techniques without the risk of overfitting. Therefore, this approach is a sound alternative to other optimization methods for enhancing mixing rules.

#### CRediT authorship contribution statement

**Miguel Monteagudo Honrubia:** Conceptualization, Formal analysis, Investigation, Software, Validation, Writing – original draft, Writing – review & editing. **Francisco Javier Herraiz-Martínez:** Project administration, Supervision, Validation, Writing – review & editing. **Javier Matanza Domingo:** Project administration, Supervision, Validation, Writing – review & editing.

#### 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.

#### Data availability

The data is available in a public repository and referenced in the article.

#### Acknowledgement

The authors would like to express their gratitude to Christopher Michael Turner for proofreading and suggesting valuable insights about the writing.

#### References

- [1] T.P. Iglesias, J.C.R. Reis, Separating the contributions of the volume change upon mixing, permittivity contrast and molecular interactions in the excess relative permittivity of liquid mixtures, *Phys. Chem. Chem. Phys.* 17 (2015) 13315–13322, <https://doi.org/10.1039/c4cp05987e>.
- [2] M.M. Honrubia, J.M. Domingo, F.J. Herraiz-Martínez, R. Giannetti, Low-cost electronics for automatic classification and permittivity estimation of glycerin solutions using a dielectric resonator sensor and machine learning techniques, *Sensors* 23 (2023) 3940, <https://doi.org/10.3390/s23083940>, <https://www.mdpi.com/1424-8220/23/8/3940>.
- [3] M.A. Stuchly, S.S. Stuchly, Coaxial line reflection methods for measuring dielectric properties of biological substances at radio and microwave frequencies—a review, *IEEE Trans. Instrum. Meas.* 29 (1980) 176–183, <https://doi.org/10.1109/TIM.1980.4314902>.
- [4] G. Ruvio, M. Vaselli, V. Lopresto, R. Pinto, L. Farina, M. Cavagnaro, Comparison of different methods for dielectric properties measurements in liquid sample media, *Int. J. RF Microw. Comput.-Aided Eng.* 28 (2018) e21215, <https://doi.org/10.1002/mmce.21215>.
- [5] M. Cavagnaro, G. Ruvio, Numerical sensitivity analysis for dielectric characterization of biological samples by open-ended probe technique, *Sensors* 20 (2020), <https://doi.org/10.3390/s20133756>.
- [6] C. Brosseau, Modelling and simulation of dielectric heterostructures: a physical survey from an historical perspective, *J. Phys. D, Appl. Phys.* 39 (2006) 1277–1294, <https://doi.org/10.1088/0022-3727/39/7/S02>.
- [7] A.A. Amooy, Improved mixing rules for description of the permittivity of mixtures, *J. Mol. Liq.* 180 (2013) 31–33, <https://doi.org/10.1016/j.molliq.2013.01.003>.
- [8] M.A. Sarami, M. Moghadam, A.G. Gilani, Modified dielectric permittivity models for binary liquid mixture, *J. Mol. Liq.* 277 (2019) 546–555, <https://doi.org/10.1016/j.molliq.2018.12.149>.
- [9] F. Qin, M. Peng, D. Estevez, C. Brosseau, Electromagnetic composites: from effective medium theories to metamaterials, *J. Appl. Phys.* 132 (2022), <https://doi.org/10.1063/5.0099072>.
- [10] D.J. Griffiths, *Introduction to Electrodynamics*, Cambridge University Press, 2017, <https://www.cambridge.org/highereducation/product/9781108333511/book>.
- [11] H. Tang, J. Cai, C.Y. Zhu, G.J. Chen, X.H. Wang, C.Y. Sun, Review on the clustering behavior in aqueous solutions, *J. Mol. Liq.* 367 (2022), <https://doi.org/10.1016/j.molliq.2022.120382>.
- [12] R. Li, C. D'Agostino, J. McGregor, M.D. Mantle, J.A. Zeitler, L.F. Gladden, Mesoscopic structuring and dynamics of alcohol/water solutions probed by terahertz time-domain spectroscopy and pulsed field gradient nuclear magnetic resonance, *J. Phys. Chem. B* 118 (2014) 10156–10166, <https://doi.org/10.1021/jp502799x>.
- [13] S. Pradhan, S. Mishra, An eye on molecular interaction studies of non-aqueous binary liquid mixtures with reference to dielectric, refractive properties and spectral characteristics, *J. Mol. Liq.* 279 (2019) 317–326, <https://doi.org/10.1016/j.molliq.2019.01.138>.
- [14] M. Homocianu, A. Airinei, Intra-/inter-molecular interactions – identification and evaluation by optical spectral data in solution, *J. Mol. Liq.* 225 (2017) 869–876, <https://doi.org/10.1016/j.molliq.2016.11.013>.
- [15] J.C.R. Reis, T.P. Iglesias, G. Douhéret, M.I. Davis, The permittivity of thermodynamically ideal liquid mixtures and the excess relative permittivity of binary dielectrics, *Phys. Chem. Chem. Phys.* 11 (2009) 3977–3986, <https://doi.org/10.1039/b820613a>.
- [16] S.E. Wood, R. Battino, *Thermodynamics of Chemical Systems*, Cambridge University Press, 1990.
- [17] S.M. Puranik, A.C. Kumbharkhane, S.C. Mehrotra, The static permittivity of binary mixtures using an improved Bruggeman model, *J. Mol. Liq.* 59 (1994) 173–177.
- [18] P.M. Meaney, C.J. Fox, S.D. Geimer, K.D. Paulsen, Electrical characterization of glycerin: water mixtures: implications for use as a coupling medium in microwave tomography, *IEEE Trans. Microw. Theory Tech.* 65 (2017) 1471–1478, <https://doi.org/10.1109/TMTT.2016.2638423>.
- [19] N. Altman, M. Krzywinski, Points of significance: simple linear regression, *Nat. Methods* 12 (2015) 999–1000, <https://doi.org/10.1038/nmeth.3627>.
- [20] J. Friedman, T. Hastie, R. Tibshirani, Regularization paths for generalized linear models via coordinate descent, *J. Stat. Softw.* 33 (2010), <http://www.jstatsoft.org/>.

- [21] A. Singh, A. Sharma, A. Ahmed, A.K. Sundramoorthy, H. Furukawa, S. Arya, A. Khosla, Recent advances in electrochemical biosensors: applications, challenges, and future scope, *Biosensors* 11 (2021), <https://doi.org/10.3390/bios11090336>.
- [22] Z. Ghahramani, *Probabilistic Machine Learning and Artificial Intelligence*, vol. 521, Nature Publishing Group, 2015.
- [23] A. Sihvola, Mixing rules with complex dielectric coefficients, vol. 1, <https://doi.org/10.1023/A:1026511515005>, 2000.
- [24] M.M. Honrubia, Vna\_er\_glycerinsolutions, [https://github.com/MigMH/VNA\\_ER\\_GlycerinSolutions](https://github.com/MigMH/VNA_ER_GlycerinSolutions), 2023.
- [25] W.J. Ellison, K. Lamkaouchi, J.-M. Moreau, Liquids water: a dielectric reference water: a dielectric reference chapter I, *J. Mol. Liq.* 68 (1996) 171–279, [https://doi.org/10.1016/0167-7322\(96\)00926-9](https://doi.org/10.1016/0167-7322(96)00926-9).
- [26] H.A. Chaube, V.A. Rana, D.H. Gadani, Dielectric absorption in mixtures of anisole with some primary alcohols at microwave frequency, *Philos. Mag.* 91 (2011) 4465–4473, <https://doi.org/10.1080/14786435.2011.610374>.
- [27] T. Thenappan, A.P. Devaraj, Dielectric studies on binary polar mixtures of propanoic acid with esters, *J. Mol. Liq.* 123 (2006) 72–79, <https://doi.org/10.1016/j.molliq.2005.04.010>.
- [28] C.V. Ramana, A.B. Kumar, A.S. Kumar, M.A. Kumar, M.K. Moodley, Dielectric and excess dielectric constants in non polar + polar binary liquid mixtures of toluene with alcohols at 303, 313 and 323 K, *Thermochim. Acta* 566 (2013) 130–136, <https://doi.org/10.1016/j.tca.2013.05.022>.
- [29] J. Lever, M. Krzywinski, N. Altman, Points of significance: principal component analysis, *Nat. Methods* 14 (2017) 641–642, <https://doi.org/10.1038/nmeth.4346>.
- [30] D. Bzdok, M. Krzywinski, N. Altman, Points of significance: machine learning: supervised methods, *Nat. Methods* 15 (2018) 5–6, <https://doi.org/10.1038/nmeth.4551>.
- [31] F. Zhang, L.J. O'Donnell, *Support Vector Regression*, Elsevier, 2019.
- [32] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, *Scikit-learn: machine learning in Python*, *J. Mach. Learn. Res.* 12 (2011) 2825–2830.
- [33] T. Head, M. Kumar, H. Nahrstaedt, G. Louppe, I. Shcherbatyi, scikit-optimize, <https://doi.org/10.5281/zenodo.5565057>, 2021.
- [34] G.C. Cawley, N.L.C. Talbot, On over-fitting in model selection and subsequent selection bias in performance evaluation, *J. Mach. Learn. Res.* 11 (2010) 2079–2107.
- [35] G. Oster, The dielectric properties of liquid mixtures, *J. Am. Chem. Soc.* 68 (1946) 2036–2041, <https://doi.org/10.1021/ja01214a050>.
- [36] L. Onsager, Electric moments of molecules in liquids, *J. Am. Chem. Soc.* 58 (1936) 1486–1493.
- [37] D.A.G. Bruggeman, Berechnung verschiedener physikalischer konstanten von heterogenen substanzen. i. dielektrizitätskonstanten und leitfähigkeiten der mischkörper aus isotropen substanzen, *Ann. Phys.* 416 (1935) 636–664, <https://doi.org/10.1002/andp.19354160705>.
- [38] C.J.F. Böttcher, O.C. van Belle, P. Bordewijk, A. Rip, D.D. Yue, Theory of electric polarization, *J. Electrochem. Soc.* 121 (1974) 211Ca.
- [39] Z. Hashin, S. Shtrikman, A variational approach to the theory of the effective magnetic permeability of multiphase materials, *J. Appl. Phys.* 33 (2004) 3125–3131, <https://doi.org/10.1063/1.1728579>.
- [40] H. Looyenga, Dielectric constants of heterogeneous mixtures, *Physica* 31 (1965) 401–406, [https://doi.org/10.1016/0031-8914\(65\)90045-5](https://doi.org/10.1016/0031-8914(65)90045-5).
- [41] J. Peón-Fernández, T. Iglesias, On the permittivity of mixtures: a low-fluctuation approach, *J. Electrostat.* 32 (1994) 113–122, [https://doi.org/10.1016/0304-3886\(94\)90002-7](https://doi.org/10.1016/0304-3886(94)90002-7).
- [42] K. Lichtenecker, K. Rother, Die herleitung des logarithmischen mischungsgesetzes aus allgemeinen prinzipien der stationären stromung, *Phys. Z* 32 (1931) 255–260, <https://cir.nii.ac.jp/crid/1572543025156436352>.
- [43] A. Kraszewski, Prediction of the dielectric properties of two-phase mixtures, *J. Microw. Power* 12 (1977) 216–222, <https://doi.org/10.1080/16070658.1977.11689049>.
- [44] O. Wiener, Theorie des mischkörpers für das feld der stationären stromung, *Abhandl. Sachs. Ges. Wiss. Math. Phys. K 1* (1912) 509–598, <https://cir.nii.ac.jp/crid/1570854174137180800>.
- [45] A.V. Sarode, A.C. Kumbharkhane, Dielectric relaxation study of poly(ethylene glycols) using TDR technique, *J. Mol. Liq.* 164 (2011) 226–232, <https://doi.org/10.1016/j.molliq.2011.09.020>.
- [46] A.Y. Zarubina, S.G. Kibets, A.A. Politiko, V.N. Semenenko, K.M. Baskov, V.A. Chistyayev, Complex permittivity of organic solvents at microwave frequencies, *IOP Conf. Ser., Mater. Sci. Eng.* 862 (2020), <https://doi.org/10.1088/1757-899X/862/6/062085>.