PCCP

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/pccp

Neural Networks Applied to Determine Thermophysical Properties of Amino Acid Based Ionic Liquids

John C. Cancilla[1], Ana Perez[1], Kacper Wierzchoś[2], José S. Torrecilla[1,*]

[1] Departamento de Ingeniería Química, Facultad de Ciencias Químicas, Universidad Complutense de Madrid, 28040-Madrid, Spain.

[2] Department of Physics, University of South Florida, Tampa, FL, United States.

Abstract

A series of models based on artificial neural networks (ANNs) have been designed to estimate thermophysical properties of different amino acid-based ionic liquids (AAILs). Three different databases of AAILs were modeled with these algorithms with the goal set to estimate density, viscosity, refractive index, ionic conductivity, and thermal expansion coefficient, by only requiring data regarding temperature and electronic polarizability of the chemicals. Additionally, a global model was designed combining all of the databases to determine the robustness of the method. In general, the results were successful, reaching mean prediction errors below 1% in many cases, as well as a statistically reliable and accurate global model. Attaining these successful models is a relevant fact as AAILs are novel biodegradable and biocompatible compounds which may soon make their way into the health sector forming part of useful biomedical applications. Therefore, understanding the behavior and being able to estimate their thermophysical properties becomes crucial.

Key Words: Neural networks, amino acid-based ionic liquids, thermophysical properties.

*Corresponding author. Tel.: +34 91 394 42 44; Fax: +34 91 394 42 43. E-mail address: jstorre@ucm.es.

1. Introduction

Ionic Liquids (ILs) are commonly formed by organic cations and either organic or inorganic anions, which originates an immense set of potential chemicals, opening the door for their use in numerous fields and almost countless applications. Some of these fields are electrochemistry, catalysis, or specific compound extraction if the cations and anions are properly selected[1].

ILs can be synthesized to possess compelling properties such as having low vapor pressures, not being inflammable, owning high thermal and chemical stabilities, and being recyclable and reusable. All of these potential traits turn these solvents into clear candidates for environmental-friendly or "green" chemistry[2-4].

During the search of new ILs, Fukumoto *et al* (2005)[5] synthesized 20 ILs which contained an amino acid as their anion. These are known as amino acid-based ionic liquids (AAILs). Amino acids are one of the most abundant biomaterials in nature, reason why the cost of the synthesis of AAILs is not relatively high. AAILs are alkaline compounds, which makes them useful to dissolve biomaterials such as cellulose or other carbohydrates. In addition, AAILs are biodegradable and can potentially possess low toxicities[6] and, therefore, their properties may enable them to participate in relevant biomedical or pharmaceutical applications[7-8], as well as being able to transport gases[9] or dissolve biomass[10] safely.

After these first AAILs, other research groups have created new AAILs which differed in their cation[11-15]. The presence of different cations dictates the final thermophysical properties of AAILs such as density, viscosity, refractive index, ionic conductivity, and thermal expansion coefficient, which were all measured by these groups[15]. These physicochemical properties have been selected because, in most cases, they are required to determine over half of the properties of fluids. Additionally, having this information is very useful to define the potential applications of the compounds. The cations that are traditionally employed are imidazolium- or pyridinium-based, although other alternatives exist such as choline, which is an essential nutrient for the correct functioning of cells and, obviously, is biodegradable[16].

Due to the great window of potential applications these novel AAILs cover, it would be very useful to design reliable and robust mathematical models to estimate the thermophysical properties of these chemical compounds [17,18]. In this sense, mathematical tools such as artificial neural networks (ANNs) have proven in the past to be worthy candidates to carry out these kind of tasks[19-24]. ANNs are algorithmic models that search for non-linear relations between independent and dependent variables from databases to originate estimative tools that operate through non-linear interpolation[20-21]. Also, it is important to note that as a deep knowledge of the IL system is not required, the ANN models are more than adequate systems to model systems that are new and not thoroughly described as is the case of the new AAILs that are being produced.

In the current research, multiple ANNs have been developed with the objective set to estimate a series of thermophysical properties (density, viscosity, refractive index, ionic conductivity, and thermal expansion coefficient) of three different sets of AAILs, each one with their own specific cation and resulting database containing the values of the thermophysical properties at various temperatures. Finally, a global ANN-based model was designed combining all three databases, to also estimate several properties. These models employed as independent variables temperature and electronic polarizability[20] values of the AAILs.

2. Materials and Methods

The databases employed to design and optimize the ANN models were gathered from different bibliographical references[11-13]. Their description, as well as a thorough explanation of the different mathematical tools used can be seen in the following subsections.

2.1. Databases

All of the data employed was obtained from three different scientific articles[11-13]. The data from the different papers had in common that it had been attained from the study of different ILs containing five different amino acids as their anions: Glycine (Gly), Alanine (Ala), Serine (Ser), Proline (Pro) and Aspartate (Asp). In each individual database, the AAILs possessed different cations, and the set of thermophysical properties measured was different in each case.

In the work carried out by Muhammad *et al* (2011)[11], they determined density, viscosity, refractive index, and thermal expansion coefficient, at atmospheric pressure, and between a temperature of 293.15 and 333.15 K, for 52 different samples which contained 1-methyl-3-methylimidazolium (Emim) as the cation, and an amino acid as the anion. They studied four AAILs (the purity estimated by the authors indicated in parentheses): EmimGly (97.8%), EmimAla (97.5%), EmimSer (97.3%), and EmimPro (96.9%). This database was used to design the **Emim Model**.

The second research study analyzed was from Ghanem *et al* (2015)[13], where they measured density, viscosity, and thermal expansion coefficient, at atmospheric pressure, and at different temperatures between 293.15 and 373.15 K, for 44 samples containing 1-octyl-3-methylimidazolium (Omim) as the cation, and one of five different amino acids as the anion. The chemical compounds were (the purity estimated by the authors was above 99% for all the AAILs): OmimGly, OmimAla, OmimSer, OmimPro, and OmimAsp. With this data, the **Omim Model** was created.

The last paper from which data was extracted was from Tao *et al* (2013)[12], where information regarding density, viscosity, refractive index, ionic conductivity, and thermal expansion coefficient can be found. These properties were measured at atmospheric pressure and between a temperature range of 298.15 and 343.15 K, for 50 different samples which had choline (Ch) as the cation and an amino acid as the anion. They studied five AAILs (the purity estimated by the authors was above 99% for all the AAILs): ChGly, ChAla, ChβAla, ChSer, and ChPro. These results were employed to develop the **Ch Model**.

Finally, the data from the three studies were brought together to design a fourth model that estimates density, viscosity, and thermal expansion coefficient (common thermophysical properties to all the papers employed to gather data), between the temperatures of 303.15 and 353.15 K. This model was designed using 112 samples of different AAILs with a variety of anions and cations, and will be referred to as the **Global Model**.

As a preliminary calculation, statistical outliers were searched in order to avoid their presence during the modeling phase. No outliers were found in this process.

2.2 Artificial Neural Networks

In order to estimate the thermophysical properties of these three groups of AAILs, a series of multilayer perceptron (MLP) models have been designed and optimized. MLPs are the most commonly employed type of supervised ANN due to their relative simplicity and reliability. The MLPs used are feed-forward ANNs with a back-propagation algorithm, which handles the supervised learning process[25]. As they are supervised mathematical tools, they require the values of the dependent variables (in this case, the thermophysical properties) to be adequately trained.

MLPs possess, as their name suggests, a layered topology or architecture, and each one of these layers has its own name and purpose (**Figure 1**): (i) the input layer is formed by ***i*** nodes, which are solely in charge of presenting the independent variables to the MLP (there will be as many nodes as independent variables). (ii) The hidden layer (or layers) possess ***j*** neurons, which are the calculating units of MLPs and ought to be properly optimized to avoid phenomena such as under-training or over-fitting. (iii) Finally, the output layer, formed by ***k*** neurons, will possess as many units as dependent variables defined for the system. Every unit (node or neuron) is connected to all of the units in the neighboring layers, but not with the ones in the same layer. These connections are controlled by weights, which are optimized during the training process to achieve accurate estimations[20,25].

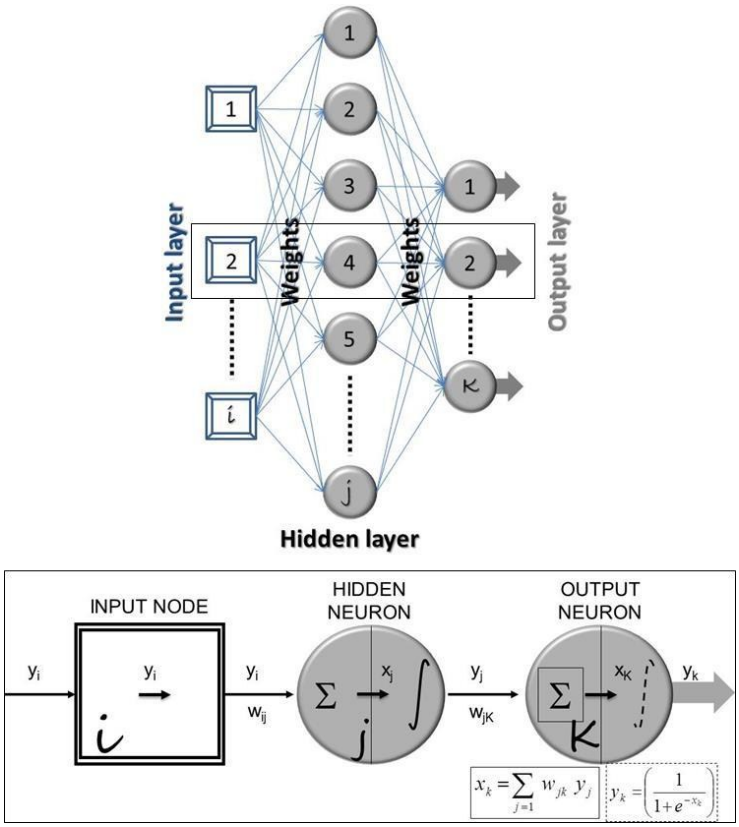

Figure 1. Schematic flow of the MLP used ($\mathbf{w_{ij}}$ and $\mathbf{w_{jk}}$ symbolize the weights which represent the connection between layers **i** and **j**, and **j** and **k**, respectively; ***i*** are the nodes in the input layer; ***j*** and ***k*** are the neurons in the hidden and output layers, respectively; the equations are the activation function (continuous box) and sigmoid transfer function (discontinuous box)).

In general, to design and use an ANN, direct and backpropagation processes are done. The first one is used to estimate the dependent variable values, and the latter is used to optimize every weight (*vide infra*).

Figure 1 shows the direct stage. Here, two consecutive mathematical calculation procedures take place within the hidden and output neurons defined. The first one is carried out by an activation function, which processes the data that enters a neuron. Finally, the second stage, is carried out in a transfer function. One of the most common transfer functions is the sigmoid function that limits the range between 0 and 1 of the resulting values given by a neuron, and was used in this research.

Given the mathematical relation between some of the dependent and independent variables found in the literature [20], in the present scenario, the amount of input nodes for the MLPs that modeled individual databases (**Emim Model**, **Omim Model**, and **Ch Model**) was two, one for the temperature and the other for the polarizability of the anion (amino acids). On the other hand, the **Global Model** possessed three independent variables which were the temperature and the polarizabilities of both anion and cation. The use of two polarizabilities gives the model the capability of estimating not only the properties of ILs which share one cation, but also of completely different ILs (different anion and cation). Additionally, the output neurons were set according to the amount of thermophysical properties estimated in each MLP (four for the **Emim Model**, three for the **Omim Model**, five for the **Ch Model**, and three for the **Global Model**).

The electronic polarizability has been calculated for every anion and cation present in the samples employed to design the models, as it is unique for each chemical compound. The calculation was carried out using the software package Marvin Suite version 5.11.5 and chemicalize.org, both developed by ChemAxom[20,26]. In addition, the software employed to develop the MLPs and carry out the experimental design to optimize its parameters was Matlab version 7.0.1.24704 (R14)[27].

2.3 Multilayer Perceptron Optimization

So as to obtain accurate, applicable, and reliable mathematical tools based on MLPs, they must be properly optimized. To carry out this phase, the databases were divided randomly into two parts. A major one known as the training dataset (80% of the data), which is used by the MLP to intrinsically modify the values of the weights in order to achieve more accurate responses. The second dataset, which is known as the verification dataset, is a smaller yet representative proportion of the database (remaining 20%) and is used by the MLP to ensure that it can generalize well or, in other words, not be over-fit towards the training dataset. While the error for the training dataset can practically reach zero with unlimited training cycles or epochs, the error for the verification dataset will end up increasing as these samples are not involved in the weight modification procedure. When this verification error begins to rise, the model can be seen as optimized, and in this case, six consecutive epochs with verification error increases, defined the end of the training process[27]. It must be noted that this verification dataset must be contained within the limits of the training dataset as ANNs are accurate when interpolating, and lose reliability when forced to estimate results beyond the limits established by the training dataset[28].

The way the model determines if a determined weight is better than the previous one is by calculating the error between the real and the estimated results. In the present case, it operates with the mean prediction error (MPE; **equation 1**), and the modification of the weights attempts to minimize it to reach a better statistical performance.

$$MPE\,(\%) = \frac{1}{N}\sum_{k=1}^{N}\frac{|r_k - y_k|}{r_k}\cdot 100 \qquad (1)$$

In this equation, N stands for the number of samples in the verification dataset, r_k is the real value of a determined sample, and y_k is the estimated value for that same sample.

In addition to the weights (the main optimization process), there are other parameters in a MLP that have to be carefully selected or optimized. They are the training and transfer functions, the hidden neuron number (HNN) (heuristically optimized), the amount of hidden layers (set as one for every model to avoid an excessive weight-to-sample ratio)[29], the Marquardt adjustment parameter (Lc), the decrease factor for Lc (Lcd), and the increase factor for Lc (Lci) [27]. The Lc parameter acts as the learning coefficient in the classic back-propagation algorithms [30]. Its value is respectively increased or decreased by Lci and Lcd parameters until these changes result in a reduced performance value [27]. When this happens, the parameters have acquired their optimal value. These last three MLP parameters were optimized through a thorough experimental design based on "Box-Wilson Central Composite Design 2^3 + star points") [20,25,27,31]. The training functions are in charge of the optimization of the ANN weights in the best way possible. In this paper, the functions chosen for this task were TrainLM and TrainBR because it has been proven that they do not provide overfit models when properly trained [32].

2.4 Validation of the Models

To ensure that the MLP-based models developed can adequately operate inside the range of the database employed to optimize it, two different validation processes have been carried out for every model.

2.4.1 Internal Validation

The basis behind this validation relies on the fact that a series random samples are removed from the databases prior to any type of calculation, and uniquely used to test the final model and determine its statistical performance. The database is initially divided randomly into training, verification, and test datasets, each containing 70%, 20%, and 10% of the data points, respectively. This 10% determines the applicability and performance of the optimized model through its MPE [31].

2.4.2. K-Fold Cross-Validation

In this case, the database is initially divided randomly into K datasets (K=6 in this case) in order to design K models, each one with a different verification dataset. In this case, each model is tested with its verification dataset, which, in the end, leads to the estimation of every single sample in the database. The final statistical performance of a MLP is the average of the MPEs of the K tests [20,31].

3. Results and Discussion

The results of every designed MLP (optimized parameters and statistical performance) will be shown and discussed in this section. In addition to the MPE, the correlation coefficient (R^2) was also calculated, but was not included as it was greater than 0.9 for every test.

3.1. Emim Model

This model has been created using data from four different AAILs (EmimGly, EmimAla, EmimSer, and EmimPro) which had 1-ethyl-3-methylimidazolium as its cation. Information from the temperature and polarizability of the amino acid-based anion was used as inputted data to estimate density, viscosity, refractive index, and thermal expansion coefficient. The final optimized parameters and selected functions can be seen in **Table 1**, while the statistical performance in terms of MPE can be found in **Table 2**.

Table 1. Optimized parameters and functions for the **Emim Model**.

Transfer function	Sigmoid
Learning function	TrainBR
Input node number	2
Hidden neuron number	5
Output neuron number	4
Lc	0.005
Lcd	0.1
Lci	10

The statistical results seem to indicate that the MLP is well suited to estimate these thermophysical properties only requiring temperature and polarizability values as the MPEs are in general terms low. To further attest this statement, two validation tests were carried out (*vide supra*) on this optimized version of the **Emim Model**, and the final results are shown in **Table 2**.

Table 2. Statistical performance of the **Emim Model** using its initial version (verification) and two validation methods (internal validation and k-fold cross validation).

	Verification	**Internal validation**	**K-fold cross validation**
	MPE (%)		**Average MPE (%)**
Density	0.09	0.06	0.08
Viscosity	4.83	5.74	4.92
Refractive index	0.02	0.03	0.03
Thermal expansion coefficient	0.10	0.06	0.10

The performance of both validation procedures are statistically comparable to the initial model, which proves its generalizing capability and applicability within the range covered by the database employed. The MPEs for density, refractive index, and thermal expansion coefficient are very low (in most cases below 0.1%), and only the viscosity is a bit higher, around 5%. This could be due to the fact that the viscosity depends on the water content much

more than the other properties, and its experimental measurement generally has a much larger intrinsic error [33].

3.2. Omim Model

This second MLP was developed employing data from five different AAILs (OmimGly, OmimAla, OmimSer, OmimPro, and OmimAsp) which possessed 1-octyl-3-methylimidazolium as its cation. Density, viscosity, and thermal expansion coefficient were estimated only using temperature and polarizability of the anion as independent variables. The final optimized parameters and selected functions are shown in **Table 3**, while the statistical performance in terms of MPE can be seen in **Table 4**.

Table 3. Optimized parameters and functions for the **Omim Model**.

Transfer function	Sigmoid
Learning function	TrainLM
Input node number	2
Hidden neuron number	8
Output neuron number	3
Lc	0.001
Lcd	0.1
Lci	10

Just like in the previous case, the results appear to be solid. Nevertheless, the **Omim Model** was further validated with the same statistical tests, and the final results can be found in **Table 4**.

Table 4. Statistical performance of the **Omim Model** using its initial version (verification) and two validation methods (internal validation and k-fold cross validation).

	Verification	**Internal validation**	**K-fold cross validation**
	MPE (%)		**Average MPE (%)**
Density	0.05	0.14	0.13
Viscosity	4.96	5.69	5.33
Thermal Expansion coefficient	0.59	0.38	0.52

The results for the **Omim Model** can be interpreted exactly the same way as the first model. The statistical performances seen in the two validation tests are comparable with the initial optimized model, which most likely indicates that the model is trustworthy and can be used for other data contained in the range of the analyzed database. The MPEs for the estimation of density and thermal expansion are low (always below 0.6%), while, for viscosity are a bit higher (around 5%), once again, possibly due to water content and experimental error when measuring [33].

3.3. Ch Model

The third model was created with data gathered from five different AAILs (ChGly, ChAla, ChβAla, ChSer, and ChPro) which had choline as its cation. Five different thermophysical properties (density, viscosity, refractive index, ionic conductivity, and thermal

expansion coefficient) were all estimated solely inputting temperature and polarizability of the anion into the MLP. The resulting optimized parameters and selected functions can be found in **Table 5**, whereas the statistical performance in terms of MPE can be seen in **Table 6**.

Table 5. Optimized parameters and functions for the **Ch Model**.

Transfer function	Sigmoid
Learning function	TrainBR
Input node number	2
Hidden neuron number	6
Output neuron number	5
Lc	0.001
Lcd	0.9
Lci	2

In this case, two of the five estimations are a bit higher than the previous two models, but still are part of a compelling tool that can estimate five properties at a time. Although the **Ch Model** may seem robust, its performance has been validated with the same two approaches, and the statistical results can be found in **Table6**.

Table 6. Statistical performance of the **Ch Model** using its initial version (verification) and two validation methods (internal validation and k-fold cross validation).

	Verification	**Internal validation**	**K-fold cross validation**
	MPE (%)		**Average MPE (%)**
Density	0.08	0.10	0.12
Viscosity	8.37	9.75	14.04
Refractive index	0.04	0.06	0.05
Ionic conductivity	9.36	12.05	13.52
Thermal expansion coefficient	0.16	0.09	0.18

The results for the **Ch Model** must be analyzed with caution. Although there are three estimations that are very accurate (density, refractive index, and thermal expansion coefficient offered estimations with MPEs below 0.2%), the other two are slightly high. Especially when analyzing the results from the K-fold cross validation, the estimations of for viscosity and ionic conductivity raise to around 14%, and, therefore, this MLP should be employed carefully. Perhaps attempting to estimate five properties at once hindered the MLP, and maybe an interesting approach would be to limit the estimations to only the three accurate properties, which would imply an extremely precise and reliable mathematical tool capable of generalizing well.

3.4. Global Model

Finally, a last model that combines all of the databases has been designed. It contains data from 12 distinct AAILs (EmimGly, EmimAla, EmimSer, EmimPro, OmimGly, OmimAla, OmimSer, OmimPro, ChGly, ChAla, ChSer, and ChPro) that include some from each of the prior databases (with the three different cations). In this case, three independent variables were necessary to define the system, and they were temperature, anion polarizability, and cation

polarizability. The final optimized parameters and selected functions are shown in **Table 7**, while the statistical performance in terms of MPE can be seen in **Table 8**.

Table 7. Optimized parameters and functions for the **Global Model**.

Transfer function	Sigmoid
Learning function	TrainBR
Input node number	3
Hidden neuron number	15
Output neuron number	3
Lc	0.005
Lcd	0.1
Lci	10

The results attained indicate that the model may be a useful tool to estimate these three thermophysical properties of a broad set of samples of AAILs from different origins. The **Global Model** was also meticulously validated using the two tests defined previously, and the results are located in **Table 8**.

Table 8. Statistical performance of the **Global Model** using its initial version (verification) and two validation methods (internal validation and k-fold cross validation).

	Verification	**Internal validation**	**K-fold cross validation**
	MPE (%)		**Average MPE (%)**
Density	0.16	0.19	0.10
Viscosity	7.29	8.69	8.11
Thermal expansion coefficient	0.22	0.61	0.28

The results of the validation tests of the **Global Model** are consistent with the previous ones, proving its applicability and generalizing ability. It must be noted, once again, that the estimation of the viscosity is less accurate than the other two properties with an MPE of around 8% [33]. The estimations for density and thermal expansion coefficient are very low, never exceeding 0.7%. The successful results of this MLP model are more than relevant, as it has correctly modeled data from a wide assortment of AAILs that were synthesized by different research groups at different periods of time. This provides robustness to these results, as the MLP model was able to incorporate data from different studies into an accurate and reliable non-linear model based on ANNs.

3.5. Model Comparison

Four different models were designed and optimized during this research. The main difference in the models were that the AAILs possessed different cations.

The first MLP model (**Emim Model**) was developed with data from AAILs that contained 1-ethyl-3-methylimidazolium as its cation, which is one of the most common cations when synthesizing ILs. The second model (**Omim Model**) used AAILs with 1-octyl-3-

methylimidazolium as its cation. These two MLPs were very accurate and reliable according to the validation tests, and the least accurate estimations found were for viscosity at around 5% MPE. The estimations for the other thermophysical properties provided almost negligible error values. Nevertheless, these two cations, and imidazolium in general, present a considerable toxicity [34], reason why, a third model (**Ch Model**) with data from AAILs containing a biodegradable cation such as choline was also designed.

Choline is an essential micronutrient required by cells to operate correctly, and, therefore, it is biodegradable and biocompatible. Nevertheless, the MLP for this database provides higher estimation errors for two properties (viscosity and ionic conductivity) while still providing very accurate results for the other three (density, refractive index, and thermal expansion coefficient). This was the only model that attempted to estimate up to five thermophysical properties at the same time, so perhaps this fact hurt the learning capability of the model, or it simply required more information to provide more accurate results.

Finally, a **Global Model** was developed to estimate density, viscosity, and thermal expansion coefficient of all the AAILs studied in the prior models. These results were significantly promising, as they were accurate, reliable, and robust according to the validation tests. Given the different nature of the data employed, it turns this model into an attractive tool that is not limited to data from single research groups, which could lead to over-fit results. It must be noted, that the error in the estimation of viscosity was higher than the other two estimations.

Mathematically speaking, in three of the four MLPs designed, the training function TrainBR was employed, while TrainLM was used in the other one (**Omim Model**). Regarding their network parameters, the values of Lc, Lcd and Lci optimized are ranged between 0.001 and 0.005, 0.1 and 0.9, and 2 and 10, respectively. The parameter values of the **Global** and **Emim Models** are the same and they are very similar to those of the **Omim Model**. Nevertheless, as the topology in all the ANN models tested is different, it is very complicated to establish a relationship between the optimal parameters and the architecture in each case.

The authors have not found in the literature any research articles that show ANNs estimating thermophysical properties of AAILs, and, therefore, the designed models open an appealing path to continue analyzing these promising compounds. The specific models that have been designed and optimized during the present research may not be appropriate to estimate properties of every AAIL at any temperature, as MLPs are tools that rely on interpolation to perform their non-linear calculations. Nevertheless, these algorithms can be re-optimized and updated with data from new samples to reach tools with a wider operational window than the original ones.

4. Conclusions

Four different models based on MLPs have been developed to estimate a series of thermophysical properties of different AAILs. Three distinct databases were employed, each of which contained information regarding AAILs with different cations (yet common to all samples from a single database).

Three of the properties (density, refractive index, and thermal expansion coefficient) were estimated with MPEs below 1% in every case. Nevertheless, the estimations were less accurate when analyzing the viscosity and ionic conductivity results.

Every model was thoroughly validated, which implies the reliability of the statistical performance offered in this research. Given the non-linear mathematical relations found during this research, it can be concluded that ANNs can be employed to create robust and reliable models for the estimation of the properties of a wide range of AAILs.

5. References

1. J. S. Torrecilla, J. Palomar, J. García, F. Rodríguez, *J. Chem. Eng. Data*. 2009, **54**, 1297.
2. M. J. Earle, K. R. Seddon, *Pure Appl. Chem*. 2000, **72**, 1391.
3. H. Olivier-Bourbigou, L. Magna, D. Morvan, *Appl. Catal. A: Gen*. 2010, *373*, 1.
4. T. Welton, *Chem. Rev.* 2011, **111**, 3508.
5. K. Fukumoto, M. Yoshizawa, H. Ohno, *J. Am. Chem. Soc.* 2005, **127**, 2398.
6. K. M. Docherty, J. C. F. Kulpa, *Green Chem.* 2005, **7**,185.
7. H. Ohno, K. Fukumoto, *Acc. Chem. Res*. 2007, **40**, 1122.
8. Y. X. An, M. H. Zong, H. Wu, N. Li, *Bioresource Technol*. 2015, **192**, 165.
9. S. Kasahara, E. Kamio, A. Yoshizumi, H. Matsuyama, *Chem. Commun*. 2014, **50**, 2996.
10. K. Ohira, K. Yoshida, S. Hayase, T. Itoh, *Chem Lett.* 2012, 41, 987.
11. N. Muhammad, Z. B. Man, M. A. Bustam, M. I. A. Mutalib, C. D. Wilfred, S. Rafiq, *J. Chem. Eng. Data.* 2011, **56**, 3157.
12. D. J. Tao, Z. Cheng, F. F. Chen, Z. M. Li, N. Hu, X. S. Chen, *J. Chem. Eng. Data*. 2013, **58**, 1542.
13. O. Ben Ghanem, M. A. Mutalib, J. M. Lévêque, G. Gonfa, C. F. Kait, M. El-Harbawi, *J. Chem. Eng. Data*. 2015, **60**, 1756.
14. X. Chen, X. Li, A. Hu, F. Wang, *Tetrahedron Asymmetry*. 2008, **19**, 1.
15. R. L. Gardas, R. Ge, P. Goodrich, C. Hardacre, A. Hussain, D. W. Rooney, *J. Chem. Eng. Data* 2010, **55**, 1505.
16. X-D. Hou, Q-P Liu, T. J. Smith, N. Li, M-H. Zong, PLoS ONE. 2013, **8**, e59145.
17. M. C. Palancar, J. M. Aragon, J. A. Miguens, J. S. Torrecilla, *Ind. Eng. Chem. Res.* 1996, 35, 4100.
18. Vladan Mlinar. Ann. Phys. (Berlin) 527, No. 3–4, 187–204 (2015)
19. J. O. Valderrama, A. Reátegui, R. E. Rojas, *Ind. Eng. Chem. Res.* 2009, **48**, 3254.
20. P. Díaz-Rodríguez, J. C. Cancilla, N. V. Plechkova, G. Matute, K. R. Seddon, J. S. Torrecilla, *Phys. Chem. Chem. Phys.* 2014, **16**, 128.
21. J. C. Cancilla, P. Díaz-Rodríguez, G. Matute, J. S. Torrecilla, *Phys. Chem. Chem. Phys.* 2015, 17, 4533.
22. K. B. Lipkowitz, D. B. Boyd. Reviews in computational chemistry (Artificial Neural Networks and Their use in Chemistry, Keith L. Peterson) Wiley-VCH, Inc 2000. MA
23. M. Alonso, B. Herradón. *Phys. Chem. Chem. Phys.* 2010, 12, 1305.
24. M. Gastegger, P. Marquetand, J. Chem. Theory Comput. 2015, 11, 2187.
25. K. Knoerzer, P. Juliano, P. Roupas C. Versteeg, Innovative Food Processing Technologies: Advances in Multiphysics Simulation, Wiley-Blackwell, Oxford, UK. 2011.
26. M. Swain, *J. Chem. Inf. Model.* 2012, **52**, 613–615.

27. H. Demuth, M. Beale M. Hagan, M. Neural Network Toolbox for Use with MATLABs User's Guide. Version 5. Ninth printing Revised for Version 5.1 (Release 2007b), 2007 (online only).
28. J. S. Torrecilla, J. García, E. Rojo, F. Rodríguez, *J. Hazard. Mater*. 2009, **164**, 182.
29. Cybenko, G. Math. Control Signals Systems 1989, 2, 303.
30. M.C. Palancar, J.M. Aragon, J.S. Torrecilla, *Ind. Eng. Chem. Res.*, 1998, 37, 2729.
31. J. C. Cancilla, S. C. Wang, P. Díaz-Rodríguez, G. Matute, J. D. Cancilla, D. Flynn, J. S. Torrecilla, *J. Agric. Food Chem.* 2014, **62**, 10661.
32. J. S. Torrecilla, J. M. Aragon, M. C. Palancar, *Ind. Eng. Chem. Res.* 2008, **47**, 7072.
33. J. S. Torrecilla, T. Rafione, J. García, F. Rodríguez, *J. Chem. Eng. Data* 2008, 53, 923.
34. S. Studzińska, B. Buszewski, *Anal Bioanal Chem.* 2009, **393**, 983.