FULL PAPER

Chemoinformatic approaches to predict the viscosities of ionic liquids and ionic liquid-containing systems

Gonçalo V. S. M. Carrera,*[a] Manuel Nunes da Ponte,[a] Luís P. N. Rebelo[a]

To my parents

Abstract: Modelling, predicting and understanding the factors influencing the viscosities of ionic liquids and related mixtures are sequentially checked in this work. The Molecular maps of atom-level properties (MOLMAP codification system) is adapted for a straightforward inclusion of ionic liquids and mixtures where they are a part of them. Random Forest models have been tested on this context and an optimal model was selected. The interpretability of the selected Random Forest model is highlighted with selected structural features that might contribute to identify low viscosities. The constructed model is able to recognize the influence of different structural variables, temperature, and pressure for a correct classification of the different systems. The codification and interpretation systems are highlighted in this work.

Introduction

The generally high viscosity of ionic liquids (ILs) is a relevant constraint to their general implementation in conventional domains of applicability, ex: solvents in extractions and synthesis, [1-3] materials, [4] cellulose handling, [5] lubricants, [6] bio-active components/carriers, [7] CO_2 capture and utilization, [8] heat transfer fluids [9] among diverse different applications. ILs present heterogeneous structural characteristics and unique dynamic behaviour, which is a challenge for the scientific community to predict their properties and, in the case of this work, the viscosities. [10-17] The enormous number of possible combinations of a cation with an anion [18,19] paves the way for the straightforward application of chemoinformatic approaches. The possibility of combining a pattern recognition methodology with the unique features of ILs is illustrated with examples of compatibility with the two complementary/different concepts on melting points, [20-25] and viscosities [26-29] prediction of these two fundamental properties of ILs.

The situation, respective mixtures, comprises different perspectives. [30,31] The viscosity of ILs-containing systems is modulated by the presence of organic solvents, mostly leading to lower viscosities. The behaviour of combinations of ILs is partially explained by the ideal Arrhenius model, and exceptions are highlighted. A plausible explanation consists on a non-random distribution of components in a system with the irregularities depending on each specific ion involved. This perspective highlights the different behaviour of a single component in the context of a multi ion system and the influence within an IL-based mixture. [32] Temperature changes are fundamental on ILs viscosities. [33] Pressure has a much lower impact, as for any neat liquid; however, systematic incremental values modify progressively the ILs behaviour. [34]

The viscosity prediction/understanding is a challenge for any common IL-based system; [35] however, various attempts have been made with diverse levels of insight. [36,37] Automatic prediction and knowledge-development of IL-based systems require a solid codification system in order to include configurations of different number/nature/distribution of components, and a straightforward curation. [38] The MOLMAP codification system, was constructed for the classification of chemical reactions without any assignment of the reaction centre. [39] The concept was applied for the prediction of chemical reactivity, with the codification of examples that do not react, usually not available in databases, solved with the creation of MOLMAP of virtual components with encoded bonds that don't react or have been created during the reaction. This is based on a contrast/operation between the MOLMAP codes of product and reagent. [40] Other examples illustrate the capacity/flexibility of this technology on the resolution of multiple phenomena. [41-45] The work herein presented consists on the combination of the, MOLMAP codification system with the description-based application on multi-component systems considering different distribution of element/components within an IL mixture/system, accounting for each component's molar fraction, in order to predict viscosities. Moreover, we explain the relationship property characteristic based on Random Forest (RF). Other codification systems are valid on their own fields of application. [46]

Results and Discussion

The work here described comprise the modelling of the viscosity of ionic liquids and their mixtures. It has been tested the Random Forest algorithm to find an intelligible relationship between the characteristics of a certain system, encoded by MOLMAP technology, and the property of interest, the viscosity.

[a] Title(s), Initial(s), Surname(s) of Author(s) including Corresponding Author(s): Doctor G. V. S. M. Carrera, Professor M. Nunes da Ponte, Professor L. P. N. Rebelo
LAQV, Requimte, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa
Address: Faculdade de Ciências e Tecnologia 2829-516 Caparica, Portugal
E-mail: goncalo.carrera@fct.unl.pt

Supporting information for this article is given via a link at the end of the document

The Random Forest algorithm is based on a set of decision trees. Each tree is a sequential partition of objects of the training set, from parent nodes into child nodes. Each partition is obtained with a logical rule from a selected descriptor. The child nodes are purer, regarding the evaluated property, when compared with the parent nodes. The variability among trees is assured considering that each tree is built with a random-selected subset of the train objects, the remaining not used systems are the Out of Bag objects (OOB). The other variability factor concerns to mtry value, the number of random descriptors, from the complete pool of variables that are tested at each node of a given tree. The final class for a given object is the one obtained my majority of votes from the complete set of trees.

The MOLMAP encoding technology mark in a map (Self organizing map Kohonen neural network), in certain positions, atoms of a given system, according atomic properties profile, accounting the molar fractions from the component that a certain atom belong. Identical procedure is carried out for components instead of atoms, however in this case the localization in component's map is based on a component-property's profile The first step on RF modelling is optimization, (Table 1/Experimental section chapter F).

Table 1. RF-Model optimization

20x20A+12x12C (1-0.5)				
mtry-A	27	35	75	150
Accuracy OOB-B	0.8222	0.8226	0.8226	0.8247
20x20A+18x18C (1-0)				
A	27	35	75	150
B	0.7685	0.7765	0.7952	0.8053
20x20A+18x18C (1-0.5)				
A	27	35	75	150
B	0.8226	0.8205	0.8248	0.8243
25x25A+12x12C (1-0.5)				
A	27	35	75	150
B	0.8157	0.8181	0.8205	0.8226
25x25A+15x15C (1-0)				
A	27	35	75	150
B	0.7642	0.7763	0.7902	0.8025
25x25A+15x15C (1-0.5)				
A	27	35	75	150
B	0.8177	0.8200	0.8208	0.8231
30x30A+12x12C (1-0)				
A	27	35	75	150
B	0.7543	0.7634	0.7937	0.7965
30x30A+12x12C (1-0.5)				
A	27	35	75	150
B	0.8149	0.81469	0.8164	0.8147

Diverse parameters have been considered such as the number of randomly chosen descriptors tested at each node of a tree in Random Forest (mtry value), descriptor's dimension and neighbourhood influence. The optimized model corresponds respectively to mtry = 27 and descriptor's dimension 20x20 for atoms + 12x12 for components + 50xTemperature + 50xPressure and 1-0.5 as a winning neuron-neighbourhood influence (Experimental section – D).

The criteria of selection involve high value of Out of Bag OOB accuracy, highest rate of processing and simplest way of interpretability.

The selected model has been externally validated with v1 and v2 sets (conception criteria, described in E - Experimental Section), three-fold cross-validation (Table 2), randomization of the classes in the training set and consequent verification of predictive ability - The OOB accuracy for the best of five randomized models is 0.201, the corresponding v1 and v2 accuracy is 0.169 and 0.153 respectively. Differently, for the optimized model, OOB, v1 and v2 predictions are accurate an indication that an inner order is pre-settled between the structure descriptor-set and viscosity.

Table 2. Modelling and validation results

Form of validation	Accuracy
train	1
out of bag	0.822
V1 - 1st order restriction	0.720
V2 - 2nd order restriction	0.642
cross validation (3-fold)	0.792

The RF algorithm estimates a value of probability for a given prediction based on the fraction of trees that credit a class. Threshold measures have been considered, based on probability, in order to group different systems, with higher probabilities for a concrete class correlating with incremental values of accuracy - Tables 3 & 4.

Table 3. Test Set v1 - 1st order restriction

Number of Objects	Treshold Probability	Accuracy
795	> 0.216	0.7195
716	> 0.4	0.757
662	> 0.45	0.7855
607	> 0.5	0.8056
529	> 0.55	0.845
478	> 0.6	0.864
436	> 0.65	0.8784
374	> 0.7	0.9118
312	> 0.75	0.9295
263	> 0.8	0.9354
209	> 0.85	0.9474
150	> 0.9	0.9733

Table 4. Test Set v2 - 2nd order restriction

Number of Objects	Treshold Probability	Accuracy
366	> 0.216	0.6421
311	> 0.4	0.6881
276	> 0.45	0.7428
247	> 0.5	0.7652
210	> 0.55	0.8238
189	> 0.6	0.8466
172	> 0.65	0.8663
150	> 0.7	0.9
113	> 0.75	0.9292
84	> 0.8	0.9286
71	> 0.85	0.9296
47	> 0.9	0.9362

The sensitivity and predictability among a given class for the different datasets have been determined (Supporting Information) and the results highlight that class A and F, with less neighbor classes lead to higher values of these parameters. Higher representation of a given class in the training set is another factor leading to better performances of that given class along the different datasets.

Beyond bulk predictive criteria the RF selected model is able to distinguish different structural profiles/patterns, figures 2-7 highlight this concept.

The NC4-average-chain and NC10-long-chain MIM ILs illustrate in a straightforward form the influence of pressure on viscosity. The ratio of increment is practically identical. The augment in absolute value is higher for NC10 MIM ILs - Figure 2.

Molecular Dynamics simulations [34] represent a straightforward form to structurally explain this effect on NC4-MIM cation. When submitted to 5000/6000 bars of pressure the alkyl chain bends, the cations get close to each other and the viscosity increases. On this work is speculated that the fraction of bending will be more substantial on longer chains. Considering our own observations on experimental data, part of this observation is adjusted to experimental reality.

The selected model is able to sense the gradation of classes with the pressure increment - Figure 2, even the number of examples at high pressure, submitted to the model is not too high.

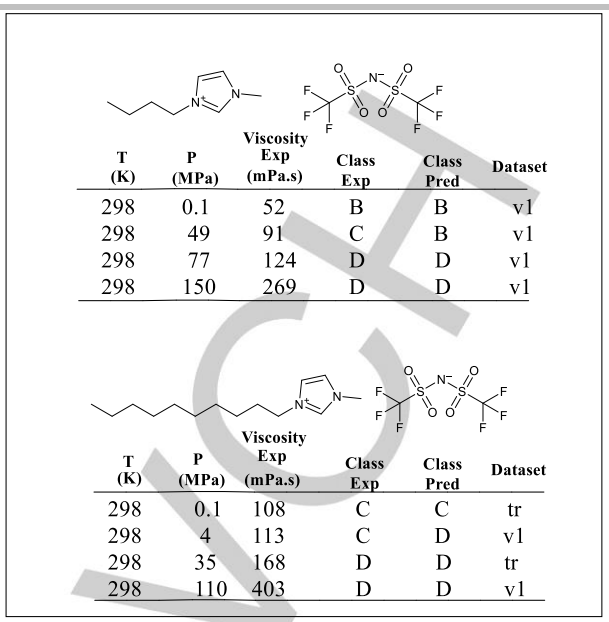


Figure 2. Effect of Pressure on viscosity of ILs of different chain length on cation

When the alkyl chain dimension on MIM ILs increases from C1 to C2 the viscosity is reduced. From C3 on, the viscosity increases. [14] Prediction model captures the increment in a straightforward manner - Figure 3.

At this T, P conditions, regarding NTf_2 anions, there is no imidazolium cations of alkyl chain dimension values near C10, in this context, prediction model learns with specific examples involving different temperatures and pressures considering diverse combinations of cation and anion for a correct/straightforward classification of NC10MIM.NTf2.

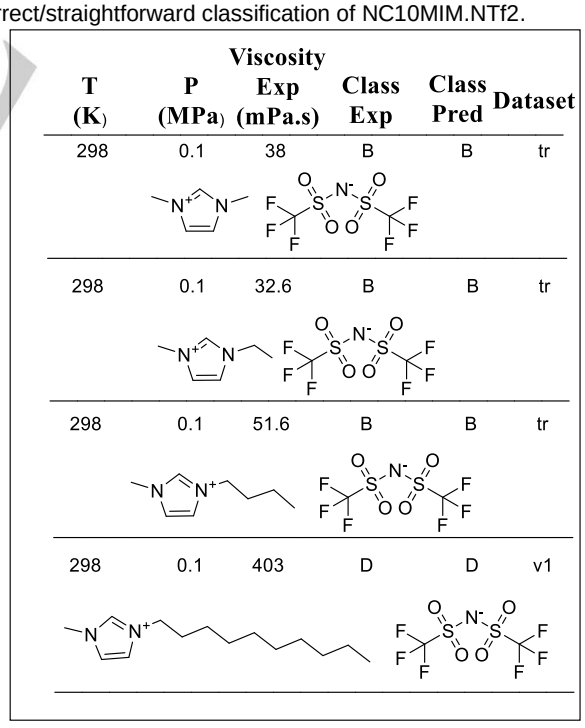


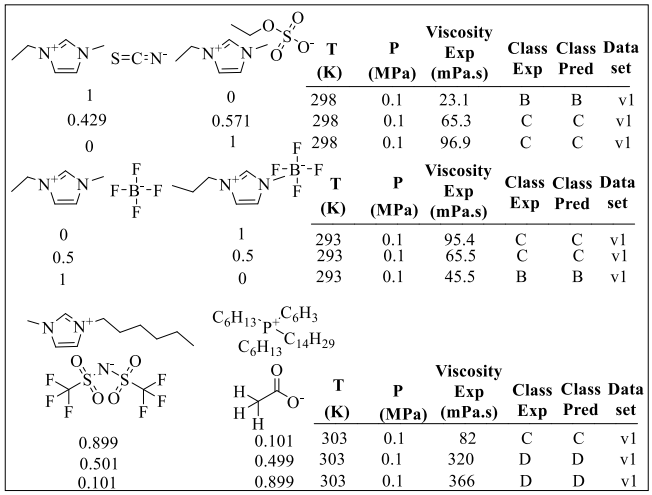
Figure 3. Effect of chain length on MIM-IL-cation's viscosity class

The RF model presents the capacity to learn the effect of different proportions on mixtures of ILs - Figure 4. It's the case of ILs that

share identical cation and different anions. Consistent with reference. [33a]

Similar behaviour is observed when the mixture shares a common anion. [54]

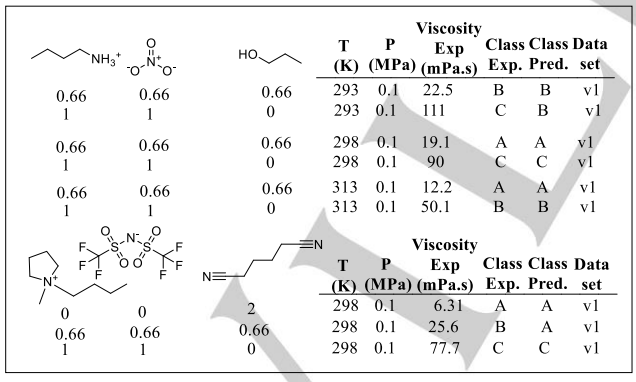
When the mixture contains two different cations and two different anions the model learns/predicts a correct gradation of viscosities in all the situations verified and leads to a correct assignment of classes on most situations. [31]



		T (K)	P (MPa)	Viscosity Exp (mPa.s)	Class Exp	Class Pred	Data set
1	0	298	0.1	23.1	B	B	v1
0.429	0.571	298	0.1	65.3	C	C	v1
0	1	298	0.1	96.9	C	C	v1
0	1	293	0.1	95.4	C	C	v1
0.5	0.5	293	0.1	65.5	C	C	v1
1	0	293	0.1	45.5	B	B	v1
0.899	0.101	303	0.1	82	C	C	v1
0.501	0.499	303	0.1	320	D	D	v1
0.101	0.899	303	0.1	366	D	D	v1

Figure 4. Mixtures of ILs. Effect of 1) same cation, 2) equal anion, 3) different cation and anion

The RF model learns the generalized rule that, the increment of the proportion of a non-charged solvent - in IL-based mixture - Figure 5 - decreases its viscosity. [16] However there are few exceptions to this rule, [54] the model learns with the examples/exceptions of the training set, however it is impossible to predict the correct class of viscosity for the few exceptions/examples on similar situation in the validation v1 set. The model captures the different effect of T on viscosities of IL-solvent.

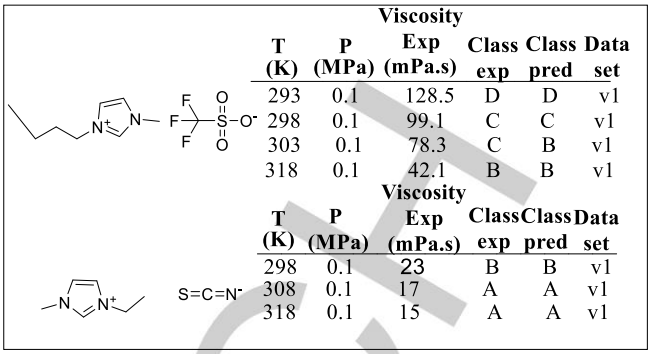


			T (K)	P (MPa)	Viscosity Exp (mPa.s)	Class Exp.	Class Pred.	Data set
0.66	0.66	0.66	293	0.1	22.5	B	B	v1
1	1	0	293	0.1	111	C	B	v1
0.66	0.66	0.66	298	0.1	19.1	A	A	v1
1	1	0	298	0.1	90	C	C	v1
0.66	0.66	0.66	313	0.1	12.2	A	A	v1
1	1	0	313	0.1	50.1	B	B	v1

			T (K)	P (MPa)	Viscosity Exp (mPa.s)	Class Exp.	Class Pred.	Data set
0	0	2	298	0.1	6.31	A	A	v1
0.66	0.66	0.66	298	0.1	25.6	B	A	v1
1	1	0	298	0.1	77.7	C	C	v1

Figure 5. Effect of molecular solvents on IL systems

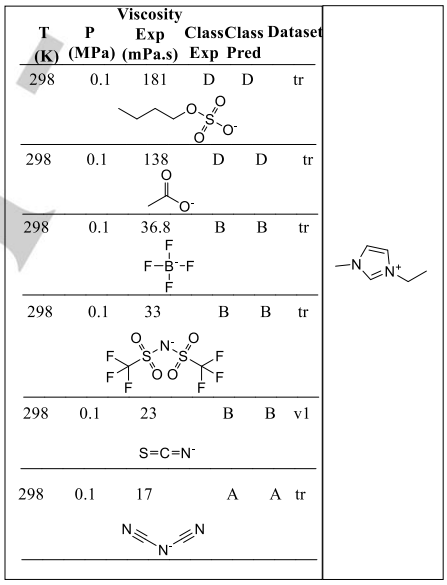
The increment of temperature reduces the viscosity on IL-based systems. The degree of influence is different considering the structure of the system - Figure 6. The RF model presents predictive ability evaluating correctly the structural effect on viscosity.



	T (K)	P (MPa)	Viscosity Exp (mPa.s)	Class exp	Class pred	Data set
	293	0.1	128.5	D	D	v1
	298	0.1	99.1	C	C	v1
	303	0.1	78.3	C	B	v1
	318	0.1	42.1	B	B	v1
S=C=N⁻	298	0.1	23	B	B	v1
	308	0.1	17	A	A	v1
	318	0.1	15	A	A	v1

Figure 6. Effect of temperature on MIM-ILs viscosity class

Concerning the anion influence, Figure 7, the hydrogen-bond-interaction-based anions lead to higher viscosities, fluorine-based anions to intermediate values and N-based delocalized anions to lower viscosities [33a] with the minimum values obtained with the dicyanamide anion. [53] The RF model captures the correct gradation of viscosities and predict, in a straightforward form, the class of the isothiocyanate ion.



T (K)	P (MPa)	Viscosity Exp (mPa.s)	Class Exp	Class Pred	Dataset
298	0.1	181	D	D	tr
298	0.1	138	D	D	tr
298	0.1	36.8	B	B	tr
298	0.1	33	B	B	tr
298	0.1	23	B	B	v1
298	0.1	17	A	A	tr

Figure 7. Influence of different anions on ILs viscosity

The ten most influent descriptors selected by the RF model, (Table 5) comprise common atoms/components.

The meaning and a visual interpretation of each descriptor is carried out in the Supporting Information,

Most relevant descriptors-order	Descriptor	MeanDecreaseAccuracy
1	A312	25.79
2	A162	25.34
3	C18	24.97
4	A161	24.61
5	A143	24.30
6	A59	23.98
7	A180	23.41
8	A264	23.29
9	A96	22.75
10	A361	22.57

Table 5. The most influent descriptors from RF model

This form of codification accounts for the physico-chemical characteristics of atoms/components and, for a specific descriptor the standard rule is that diverse components/atoms that contain similar physicochemical profile share the winning neuron. The neighbour neurons account for the value of a particular descriptor position-Experimental section - D. This physico-chemical profile indexing procedure is straightforward and may resolve hidden relations gathering different items. This application resolves mixture and related system-inclusion in the context of information curation. [38]

The interpretation of relevant and meaningful descriptors (Experimental Section - G) reveals indicative and intelligible relationships - Figures S1-S7 (Supporting Information).

The viscosity class depend effectively on the corresponding value of the average descriptor. The winning position and neighbour cells influence the final value.

The straightforward interpretation of the RF model (Figures S1-S7 Supporting information and Results and Discussion) leads to the conclusion that structural features such as: neutral components, the presence of specific anions such as dicyanamide, as the best illustrative example, delocalization of charge/small-size chain on both cation and anion, and proper CO_2 levels contribute to a clear viscosity reduction.

Conclusions

A novel codification system has been designed in order to include Ionic Liquids and their mixtures in a global model. A Random Forest model has been optimized for the modelling and prediction of viscosities. The interpretative system from Random Forest information permits to unveil the influence of a given descriptor on the viscosity of a given system. This method adds value to the Random Forest algorithm solving previous interpretability limitations

The Random Forest model is able to identify correct characteristics, leading to viscosity reduction such as the presence of molecular components, dicyanamide and bistriflimide as anions, non-centred delocalized charge, considerable CO_2 levels and small chains on both anion and cation.

This work contributes to computer-assisted research for generic characteristics influencing a given property value.

Experimental Section

A - Database structuration

NIST ILThermo database [47] gathers information on diverse features of a generic IL and related systems. A work-selection involves 13798 samples including single IL systems, mixtures IL/IL and IL/Organic compound. A system includes different components, a component is an ion or a molecule. The work environment Chemfinder [48] has been the platform to conclude this aspect of the project. This work-selection includes structural information, viscosity, temperature, pressure and molar fractions. The structural information is encoded in smiles form.

B - Clean-up and extraction of information

The structures have been standardized following the sequence 1) Mesomerize 2) Add explicit Hydrogens and 3) Clean three-dimensional structures. Ion/Molecule and atomic properties have been extracted from the file: a) weight, b) surface area, c) volume and d) polarizabilities within a component and i) hydrogen bond donor, ii) hydrogen bond acceptor, iii) total charge, iv) polarizabilities, v) sigma charge, vi) pi charge, vii) orbital electronegativity sigma, viii) hindrance and ix) atomic number from elements. The programs standardizer and cxcalc from Chemaxon [49] have been tested on this sequence for standardization of chemical structures and estimation of atomic and component properties.

All the properties have been normalized considering the formula:

$$Prop.normalized = \frac{Prop - min}{max - min}$$

Equation 1: Prop – original value of the property, min – minimum value of the property considering all the objects components/elements, max – maximum value of the property.

C - Kohonen Neural Network training

Two different types of Kohonen Neural Networks have been trained - atoms and components, which have been organized considering their properties profile, characterized on the previous section. 5000 components have been randomly chosen - 2413 anions, 2307 cations and 280 molecules. 10000 atoms have been selected in an identical form - 4500 atoms from cation, 4000 from anion and 1500 atoms from molecules. The component's Kohonen NNs comprise 12x12 (C1-C144 component descriptors), 15x15 (C1-C225 component descriptors) and 18x18 (C1-C324 component descriptors) Self-Organizing maps (SOMs). 20x20 (A1-A400 atom descriptors), 25x25 (A1-A625 atom

descriptors) and 30x30 (A1-A900 atom descriptors) are the dimensions of element's SOMs.

D - System's codification method

First step: MOLMAP (MOLecular Maps of Atom level Properties) for atoms/components

MOLMAP descriptors encode the presence/absence of specific atom or component types in the system. The atom/component types are defined by *Kohonen* neural networks on the basis of empirical physicochemical properties of atoms or even components.

MOLMAPs applied to this work involve atoms and components. Each component/atom (item) of a system was submitted to the respective trained *Kohonen* Neural Network – Self Organizing Map (SOM) obtained in the previous C-Section. Each item activates a neuron (a wining position in the SOM). The program CUTMAPZ3N converts the *Kohonen* Neural Network profile (atom or component's activation profile) into a MOLMAP pattern (numerical representation of activation in a grid), each winning neuron/activation represents a specific/concrete value, the respective neighbour neurons activate a different/lower value.

Second step: System codification

The next step includes the two-level conversion of compound molar fraction to component correspondence and that converted molar fraction weighting for the respective MOLMAPs of atoms of a component (by multiplication). Identical procedure for MOLMAP of component. The weighted MOLMAPs of a certain type (atoms or component) are summed up for all components of a system, resulting in unified MOLMAPs, one for atoms and another for the components of a system.

E - Training and validation setup

The original 13798 samples have been reduced and a compact 8501-sample remained. The criteria of deletion involve redundancies and incoherencies. The reduced set was distributed into a 7706-system training set (tr) and a 795-system validation set (v1). This validation set includes systems with non-identical MOLMAP and/or temperature and/or pressure profile when compared to the training set. This is the first level of restriction evolving tr an v1 sets. The second level of restriction comprises a v2 validation set of 366 systems, from v1, where, at least one component is not present in a generic system of the training set. [46c]. The v2 dataset is included in v1 collection of systems. The integral pool of systems includes different: 611 ILs, 288 cations, 99 anions, 45 molecules. Temperature interval from 253-438 K. Pressure range [0.07-300MPa]. Check Supporting Information for a detailed datasets composition's report.

F - Random forest classification model - RF - Model building

Classification models have been constructed comprising six different levels of viscosity - Class A: 0.28-20.59 mPa.s, Class B: 20.6-51.6 mPa.s, Class C: 51.7-122.9 mPa.s, Class D: 123-414 mPa.s, Class E: 415-1035.5 mPa.s and Class F: 1036-140000 mPa.s. The concept of these six classes considers different classes with substantial representation in order to the Random Forest – RF algorithm learn general rules. This algorithm has been tested to find a straightforward link between the codification system/descriptors and the viscosities of IL-based systems. The R environment [50] has been used as a platform to build up the models. The optimization comprises descriptor's dimensions and neighbourhood influence: /20x20/-/30x30/ as /dimension/ interval for elements /12x12/-/18x18/ for components, and {1-0} or {1-0.5} concerning {winning neuron-neighbour neuron weights for activation} and to conclude the *mtry* verification (27-150).

The out of bag OOB [51] accuracy of the training set was verified as criteria to select the optimum model:

$$Accuracy = \frac{number\ of\ correct\ assignments\ of\ the\ data\ set}{total\ number\ of\ systems}$$

Equation 2

The chosen model comprehends a measure of descriptor-importance, predictive ability for validation set and probability of prediction. [52]

G - Interpret RF model

The ten most relevant descriptors selected by RF model and the form of codification developed in chapters C and D have been combined:

The trained Kohonen neural networks in chapter C create an output file with information of item position on the network. This position corresponds to a contribution on a determined descriptor - chapter D. Each concrete system has their own patterns of activation component/atom with the weight of molar fractions embedded. Each concrete item (component/atom) on this context has its own value and situation for description. Each specific item/descriptor contributor has the influence of the neighbourhood cells in a selective proportion.

The items (atoms or components) of a generic system, of v1 validation set, mapped on a winning or neighbour neuron are gathered in excel file. This procedure is carried out for a given descriptor of the ten most important for the viscosity model. The value of activation, the predicted and experimental classes of that item are included. The systems/items are divided in different sheets according to the experimental class and tested by Equation 3, and it permits to recognize all items of a given system and when a different system starts being evaluated. The application of that equation permits the determination, on the next step, of the average value of that descriptor for the correct assignments considering a given class (A-F). The final step consists on finding a straightforward/intelligible relationship between descriptor/property, a significant pattern between the different classes of viscosity and the value of the descriptor.

$$SystemDescriptor = IF(OR\left(AND(I2 = I1; NOT(I2 = I3)); AND(I2 = I1; I2 = I3)\right); 0; IF(AND\left(NOT(I2 = I1); NOT(I2 = I3)\right); M2; IF(AND\left(NOT(I2 = I1); I2 = I3; NOT(I3 = I4)\right); SUM(M2{:}M3); IF(AND\left(NOT(I2 = I1); I2 = I3; I3 = I4; NOT(I4 = I5)\right); SUM(M2{:}M4); \ldots$$

Equation 3: I System where an M item belongs with a certain value

H - Measures of Validation/Probability

-Viscosity class-randomization in the training set, model-build-up and prediction for validation sets

-3-fold cross validation

-Out of bag OOB prediction

- v1 and v2 predictions

- Probability of correct assignment

- Sensitivity within a given class (Supporting Information): Percentage of a given experimental class objects, effectively classified as that class by the model.

- Predictability for a given class (Supporting information):

Percentage of objects classified as a given class that effectively belong to that class experimentally.

I - Experimental Procedure/Figure

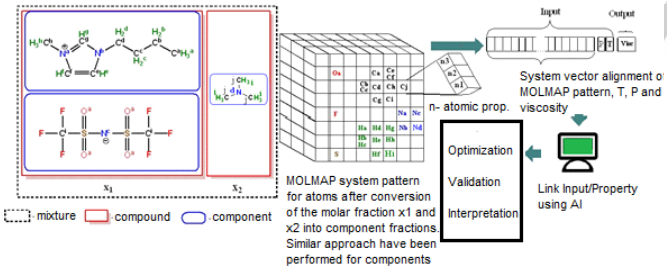


Figure 1. Schematic representation of the experimental procedure

Acknowledgements

This work was financially supported by FCT/MCTES-Portugal-Project-PTDC/EQU-EQU/30060/2017-Grant-SFRH/BPD/72095/2010.

This work was supported by the Associate Laboratory for Green Chemistry-LAQV which is financed by national funds from FCT/MCTES (UID/QUI/50006/2019).

Professor João Aires-de-Sousa is acknowledged for fruitfull discussions.

Keywords: Ionic Liquids • Viscosity • Chemoinformatics • MOLMAP• Random Forest

[1] S. Werner, M. Haumann, P. Wasserscheid, *Annu. Rev. Chem. Biomol. Eng.* **2010**, *1*, 203-230

[2] M. G. Freire, C. L. S. Louros, L. P. N. Rebelo, J. A. P. Coutinho, *Green Chem.* **2011**, *13*, 1536-1545

[3] a) Z. Hu, C. J. Margulis, *Acc. Chem. Res.* **2007**, *40*, 1097-1105 b) S. Tiwari, N. Khupse, A. Kumar, *J. Org. Chem.* **2008**, *73*, 9075-9083

[4] E. I. Izgorodina, Z. L. Seeger, D. L. A. Scarborough, S. Y. S. Tan, *Chem. Rev.* **2017**, *117*, 6696-6754

[5] a) H. Cruz, M. Fanselow, J. D. Holbrey, K. R. Seddon, *Chem. Commun.* **2012**, *48*, 5620-5622 b) Y. Li, J. Wang, X. Liu, S. Zhang, *Chem. Sci.* **2018**, *9*, 4027-4043

[6] F. Zhou, Y. Liang, W. Liu, *Chem. Soc. Rev.* **2009**, *38*, 2590-2599

[7] K. S. Egorova, E. G. Gordeev, V. P. Ananikov, *Chem. Rev.* **2017**, *117*, 7132-7189

[8] a) G. V. S. M. Carrera, N. Jordão, M. M. Santos, M. N. da Ponte, L. C. Branco, *RSC Adv.* **2015**, *5*, 35564-35571 b) L. C. Tomé, C. Florindo, C. S. R. Freire, L. P. N. Rebelo, I. M. Marrucho, *Phys. Chem. Chem. Phys.* **2014**, *16*, 17172-17182 c) G. V. S. M. Carrera, N. Jordão, L. C. Branco, M. Nunes da Ponte, *Faraday Discuss.* **2015**, *183*, 429-444 d) M. Alvarez-Guerra, J. Albo, E. Alvarez-Guerra, A. Irabien, *Energy Environ. Sci.* **2015**, *8*, 2574-2599 [9] K. Dong, X. Liu, H. Dong, X. Zhang, S. Zhang, *Chem. Rev.* **2017**, *117*, 6636-6695

[10] E. J. Maginn, *Acc. Chem. Res.* **2007**, *40*, 1200-1207

[11] F. F. Canova, H. Matsubara, M. Mizukami, K. Kurihara, A. L. Shluger, *Phys. Chem. Chem. Phys.* **2014**, *16*, 8247-8256

[12] Z. Hu, C. J. Margulis, *Acc. Chem. Res.* **2007**, *40*, 1097-1105

[13] Z-P. Zheng, W-H. Fan, S. Roy, K. Mazur, A. Nazet, R. Buchner, M. Bonn, J. Hunger, *Angew. Chem. Int. Ed.* **2015**, *54*, 687-690

[14] J. Jacquemin, P. Husson, A. A. H. Padua, V. Majer, *Green Chem.* **2006**, *8*, 172-180

[15] C. A. N. de Castro, *J. Mol. Liq.* **2010**, *156*, 10-17

[16] K. R. Seddon, A. Stark, M. J. Torres, *Pure Appl. Chem*. **2000**, *72*, 2275-2287

[17] J. S. Wilkes, *J. Mol. Catal. A*. **2004**, *214*, 11-17

[18] A. R. Katritzky, R. Jain, A. Lomaka, R. Petrukhin, M. Karelson, A. E. Visser, R. D. Rogers, *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 225-231

[19] D. M. Eike, J. F. Brennecke, E. J. Maginn, *Green. Chem.* **2003**, *5*, 323-328

[20] A. Varnek, N. Kireeva, *J. Chem. Inf. Model.* **2007**, *47*, 1111-1122

[21] I. Lopez-Martin, E. Burello, P. N. Davey, K. R. Seddon, G. Rothenberg, *ChemPhysChem.* **2007**, *8*, 690-695

[22] R. Bini, C. Chiappe, C. Duce, A. Micheli, R. Solaro, A. Starita, M. R. Tiné, *Green Chem.* **2008**, *10*, 306-309

[23] S. Trohalaki, R. Pachter, *QSAR Comb. Sci.* **2005**, *24*, 485-490

[24] S. Trohalaki, R. Pachter, G. W. Drake, T. Hawkins, *Energy Fuels.* **2005**, *19*, 279-284

[25] G. Carrera, J. Aires-de-Sousa, *Green. Chem.* **2005**, *7*, 20-27

[26] K. Paduszynski, U. Domanska, *J. Chem. Inf. Model.* **2014**, *54*, 1311−1324

[27] Y. Zhao, Y. Huang, X. Zhang, S. Zhang, *Phys. Chem. Chem. Phys.* **2015**, *17*, 3761-3767

[28] S. Martin, H. D. Pratt, III, T. M. Anderson, *Mol. Inf.* **2017**, *36*, 1600125

[29] R. Alcalde, G. García, M. Atilhan, S. Aparicio, *Ind. Eng. Chem. Res.* **2015**, *54*, 10918-10924

[30] M. Tariq, T. Altamash, D. Salavera, A. Coronas, L. P. N. Rebelo, J. N. C. Lopes, *ChemPhysChem*. **2013**, *14*, 1956-1968

[31] M. T. Clough, C. R. Crick, J. Grasvik, P. A. Hunt, H. Niedermeyer, T. Welton, O. P. Witaker, *Chem. Sci.* **2015**, *6*, 1101-1114

[32] J. J. Fillion, J. F. Brennecke, *J. Chem. Eng. Data*. **2017**, *62*, 1884-1901

[33] a) H. F. D. Almeida, J. N. C. Lopes, L. P. N. Rebelo, J. A. P. Coutinho, M. G. Freire, I. M. Marrucho, *J. Chem. Eng. Data.* **2016**, *61*, 2828-2843 b) X. Wang, F. W. Heinemann, M. Yang, B. U. Melcher, M. Fekete, A-V. Mudring, P. Wasserscheid, K. Meyer, *Chem. Commun.* **2009**, 7405-7407

[34] Y. Zhao, X. Liu, X. Lu, S. Zhang, J. Wang, H. Wang, G. Gurau, R. D. Rogers, L. Su, H. Li, *J. Phys. Chem. B.* **2012**, *116*, 10876-10884

[35] S. Tang, G. A. Baker, H. Zhao, *Chem. Soc. Rev.* **2012**, *41*, 4030-4066

[36] C. Han, G. Yu, L. Wen, D. Zhao, C. Asumana, X. Chen, *Fluid Phase Equil.* **2011**, *300*, 95-104

[38] D. Fourches, E. Muratov, A. Tropsha, *J. Chem. Inf. Model.* **2010**, *50*, 1189-1204

[39] Q-Y. Zhang, J. Aires-de-Sousa, *J. Chem. Inf. Model.* **2005**, *45*, 1775-1783

[40] G. V. S. M. Carrera, S. Gupta, J. Aires-de-Sousa, *J. Comput. Aided Mol. Des.* **2009**, *23*, 419-429

[41] D. A. R. S. Latino, J. Aires-de-Sousa, *Angewandte, Chem. Int. Ed.* **2006**, *45*, 2066-2069

[42] D. A. R. S. Latino, Q-Y. Zhang, J. Aires-de-Sousa, *Bioinformatics* **2008**, *24*, 2236-2244

[43] S. Gupta, S. Matthew, P. M. Abreu, J. Aires-de-Sousa, *Bioorg. Med. Chem*. **2006**, *14*, 1199-1206

[44] B. Hemmateenejad, A. R. Mehdipour, P. L. A. Popelier, *Chem. Biol. Drug Des*. **2008**, *72*, 551-563

[45] Q-Y. Zhang, J. Aires-de-Sousa, *J. Chem. Inf. Model.* **2007**, *47*, 1-8

[46] a) J. Palomar, J. S. Torrecilla, J. Lemus, V. R. Ferro, F. Rodriguez, *Phys. Chem. Chem. Phys.* **2010**, *12*, 1991-2000 b) E. Mokshyna, V. I. Nedostup, P. G. Polishchuk, V. E. Kuzmin, *Mol Inform.* **2014**, *33*, 647-654 c) I. Oprisiu, S. Novotarskyi, I. V. Tetko, *J. Chemoinformatics*. **2013**, *5*:4 d) P. Polishchuk, T. Madzhidov, T. Glimadiev, A. Bodrov, R. Nugmanov, A. Varnek, *J. Comput. Aided Mol. Des.* **2017**, *31*, 829-839

[47] https://ilthermo.boulder.nist.gov/

[48] http://www.cambridgesoft.com

[49] https://chemaxon.com/

[50] a) https://cran.r-project.org/; b) L. Breiman, *Machine Learning*, **2001**, *45*, 5-32

[51] https://www.r-project.org/conferences/useR-2009/slides/Cutler.pdf

[52] https://cran.r-project.org/web/packages/randomForest/randomForest.pdf

[53] D. R. MacFarlane, J. Golding, S. Forsyth, M. Forsyth, G. B. Deacon, *Chem. Commun.* **2001**, 1430-1431

[54] Y. Zhao, X. Zhang, S. Zeng, Q. Zhou, H. Dong, X. Tian, S. Zhang, *J. Chem. Eng. Data.* **2010**, *55*, 3513-3519

[55] X. Wang, Y. Chi, T. Mu, *J. Mol. Liq.* **2014**, *193*, 262-266

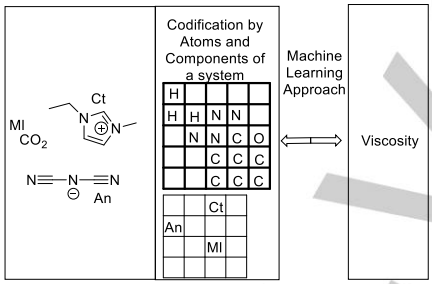
Entry for the Table of Contents (Please choose one layout)

Layout 1:

FULL PAPER

Text for Table of Contents

A unified form of codification of pure compounds and mixtures is presented in this work. A Machine-learning technology has been tested to model, predict and interpret the viscosity of ionic liquids and related systems



Gonçalo V. S. M. Carrera, Manuel Nunes da Ponte, Luís P. N. Rebelo*

Page No. – Page No.

Chemoinformatic approaches to predict the viscosities of ionic liquids and ionic liquid-containing systems