Supplementary Information

Accurate Estimation of Physicochemical Properties of Ternary Mixtures Containing Ionic Liquids via Artificial Neural Networks

John C. Cancilla, Pablo Díaz-Rodríguez, Gemma Matute, José S. Torrecilla*

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

Database Utilized

The database that has been employed to design and test the multiple linear regression (MLR) models as well as to train, optimize, and test the artificial neural network (ANN) based multilayer perceptron (MLP) model was obtained from the work of Navarro *et al*, 2012, and can be seen here (**Table S1**). They obtained the data from various experimental measurements of ternary mixtures of the IL 1-butyl-3-methylimidazolium tetrafluoroborate, 2-propanol, and water at 298.15 K.

Table S1. Database used to design the MLR models and to train, optimize, and test the ANN (1). Ternary mixture parameters: $\mathbf{w_1}$, mass fraction of 2-propanol; $\mathbf{w_2}$, mass fraction of water; $\mathbf{n_D}$, refractive index; **ρ**, density (g/cm^3). Data points labelled with a "T" when used in training dataset, with a "V" when used in verification dataset, and with an "S" when used in simulation dataset.

Ternary Mixtures and Dataset	w_1	w_2	n_D	ρ (g/cm^3)
1-T	0.0210	0.3882	1.3816	1.102
2-T	0.0225	0.4948	1.3726	1.081
3-V	0.0228	0.6827	1.3577	1.048
4-T	0.0254	0.8671	1.3437	1.010
5-T	0.0315	0.1981	1.3986	1.134
6-T	0.0324	0.2944	1.3898	1.115
7-T	0.0334	0.2433	1.3941	1.124
8-T	0.0341	0.0161	1.4183	1.176
9-V	0.0341	0.0565	1.4134	1.165
10-T	0.0341	0.0949	1.4093	1.155
11-T	0.0348	0.0382	1.4154	1.168
12-T	0.0375	0.0094	1.4187	1.176
13-T	0.0390	0.0197	1.4177	1.162
14-S	0.0418	0.1476	1.4033	1.140
15-T	0.0788	0.1282	1.4037	1.125
16-T	0.0957	0.0647	1.4095	1.131

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

17-T	0.1023	0.3560	1.3837	1.073
18-V	0.1047	0.0244	1.4135	1.136
19-T	0.1049	0.4530	1.3760	1.056
20-V	0.1088	0.0433	1.4112	1.129
21-T	0.1089	0.0258	1.4160	1.139
22-T	0.1089	0.6225	1.3630	1.027
23-S	0.1124	0.7897	1.3507	0.999
24-T	0.1906	0.1495	1.3996	1.068
25-V	0.1926	0.3202	1.3852	1.041
26-V	0.1946	0.4077	1.3788	1.027
27-T	0.1971	0.0795	1.4040	1.077
28-T	0.1990	0.5595	1.3679	1.003
29-V	0.1999	0.2354	1.3906	1.048
30-T	0.2027	0.7093	1.3577	0.977
31-T	0.2039	0.0210	1.4088	1.084
32-T	0.2074	0.0459	1.4132	1.062
33-T	0.2082	0.0398	1.4067	1.075
34-V	0.2117	0.2289	1.3913	1.045
35-T	0.2144	0.3227	1.3798	1.021
36-S	0.2176	0.2535	1.3899	1.041
37-T	0.2487	0.1734	1.3940	1.030
38-T	0.2551	0.1609	1.3952	1.035
39-V	0.2598	0.1397	1.3987	1.032
40-V	0.2649	0.0996	1.4024	1.037
41-S	0.2711	0.2956	1.3849	1.010
42-T	0.2887	0.0254	1.4044	1.045
43-T	0.3152	0.2716	1.3858	0.992
44-T	0.3236	0.3424	1.3809	0.985
45-T	0.3265	0.2020	1.3900	1.000
46-V	0.3303	0.4678	1.3726	0.967
47-T	0.3376	0.5893	1.3643	0.948
48-T	0.3542	0.2071	1.3888	0.987
49-S	0.3759	0.2674	1.3845	0.971
50-T	0.3876	0.0592	1.3974	0.992
51-T	0.3969	0.1302	1.3924	0.980
52-T	0.3982	0.0350	1.3983	0.990
53-T	0.4105	0.2126	1.3869	0.964
54-V	0.4318	0.2727	1.3831	0.949
55-T	0.4423	0.1492	1.3898	0.959
56-T	0.4588	0.2598	1.3826	0.941
57-T	0.5058	0.0288	1.3941	0.948
58-T	0.5144	0.0249	1.3888	0.945
59-T	0.5577	0.1042	1.3883	0.920
60-T	0.5826	0.3418	1.3744	0.892
61-S	0.5974	0.0689	1.3884	0.908
62-T	0.6124	0.0472	1.3890	0.905

63-T	0.6275	0.2504	1.3773	0.880
64-T	0.6863	0.2655	1.3751	0.863
65-T	0.6909	0.1747	1.3796	0.869
66-T	0.6970	0.1675	1.3793	0.866
67-V	0.7588	0.0936	1.3810	0.851
68-S	0.7641	0.1822	1.3768	0.844
69-V	0.7970	0.0479	1.3816	0.842
70-T	0.8316	0.1100	1.3778	0.827
71-T	0.8831	0.0549	1.3781	0.814
72-T	0.9079	0.0283	1.3780	0.807

* Standard uncertainties: u (n_D) = 0.0008; u(ρ) = 0.001 g/cm^3.

Hidden Neuron Number Selection

Different hidden neuron numbers (HNNs) were tested in order to define the optimal number in terms of ANN performance, which is measured through the mean prediction error (MPE) and R^2 correlation coefficient of the verification process. Correctly defining the HNN is important because a high HNN may lead to an over-fit network, while a low one might create a MLP with a diminished learning capability. Due to the size of the database, HNNs from two to ten were tried, and the results can be seen in **Table S2**.

Table S2. HNN tests using the verification dataset to simulate the network. MPEs and R^2 for the estimation of both physicochemical properties studied are shown. The selected HNN is marked in bold.

HNN	n_D MPE (%)	n_D R^2	ρ MPE (%)	ρ R^2
2	0.166	0.980	0.915	0.985
3	0.124	0.987	0.537	0.992
4	0.115	0.987	0.768	0.990
5	0.067	0.995	0.459	0.997
6	0.086	0.995	0.406	0.996
7	0.086	0.995	0.327	0.997
8	**0.050**	**0.997**	**0.140**	**0.999**
9	0.055	0.997	0.152	0.999
10	0.078	0.994	0.312	0.998

The HNN selected was eight because the MPE values for the estimation of density and refractive index were lower than the other tests, and, therefore, it is the most accurate option.

Artificial Neural Network Parameter Optimization

The optimization of the Marquardt adjustment parameter (Lc), the decrease factor for Lc (Lcd), and the increase factor for Lc (Lci) using a thorough experimental design based on the "Box-Wilson Central Composite Design 2^3 + star points" was carried out (2). The range of values tested was from 0.001 to 1 for Lc and Lcd, and from 2 to 100 for Lci. The values of these parameters to optimize the statistical performance of the network are shown in **Table S3**. Four different combinations of Lc, Lcd, and Lci were obtained to give preference and optimize a specific statistical result ($\mathbf{n_D}$ MPE, $\mathbf{n_D}$ R^2, **ρ** MPE, or **ρ** R^2).

Table S3. Optimal ANN parameter values obtained through an experimental design based on the "Box-Wilson Central Composite Design 2^3 + star points".

Combination	Optimized	Lc	Lcd	Lci
1	$\mathbf{n_D}$ MPE	0.46	0.85	53.32
2	$\mathbf{n_D}$ R^2	0.56	1	52.66
3	**ρ** MPE	0.40	1	100
4	**ρ** R^2	0.51	1	100

In order to select the best combination of parameters (Lc, Lcd, and Lci), all four options have been tested. The new MLPs were created using these values (**Table S3**) and their performance was evaluated analyzing the MPE and R^2 values obtained after employing the simulation dataset to test the ANNs designed (**Table S4**).

Table S4. Statistical results of the different MLPs designed with the four combinations of ANN parameters (**Table S3**) attained after the experimental design. Selected option is shown in bold.

Combination	$\mathbf{n_D}$ **MPE** (%)	$\mathbf{n_D}$ $\mathbf{R^2}$	**ρ MPE** (%)	**ρ** $\mathbf{R^2}$
1	0.069	0.997	0.360	0.998
2	0.052	0.998	0.447	0.996
3	**0.050**	**0.998**	**0.227**	**0.999**
4	0.135	0.982	0.387	0.998

The ANN parameters selected were the ones from combination 3 (Lc=0.40; Lcd=1; Lci=100) which gives preference to the optimization of the MPE of the density estimation. This is the option selected because it offers the best statistical results in terms of both, MPE and R^2 values.

References

(1) P. Navarro, M. Larriba, S. García, J. García, F. Rodríguez, *Journal of Chemical Engineering Data*, 2012, **57**, 1165.

(2) H. Demuth, M. Beale, M. Hagan, *Neural Network Toolbox for Use with MATLAB® User's Guide*. Version 4.0.6. Ninth printing Revised for Version 4.0.6 (Release 14SP3), Natick, MA **2005**.