Algorithmic Modeling of Spectroscopic Data to Quantify Binary Mixtures of Vinegars from Different Botanical Origins

José S. Torrecilla*, Regina Aroca-Santos, John C. Cancilla, Gemma Matute

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

Supplementary material

Additional information regarding different optimization procedures followed is detailed in the following sections.

MLR-based models

In Tables S1 1-6 the fit coefficients of five vinegars for each kind of binary blend (six binary possible mixtures) are shown. These coefficients are the result of averaging the ones obtained from the k-fold cross-validation analysis (k=6).

Table SI.1. Values of the fit coefficients of the MLRs designed to estimate the composition of white wine vinegar blends.

Term	Red Wine	Cider	Apple	Molasses	Rice
Constant	-0.03	0.01	-0.01	0.83	-0.02
AUC_1	-39.87	-1.98	4.53	12.65	-0.88
AUC_2	35.64	3.55	-8.45	-16.36	7.38
AUC_3	35.86	-0.17	4.48	-9.58	-4.99
AUC_4	-118.93	-1.49	0.45	14.87	-3.07
AUC_5	166.28	2.06	-1.65	-11.39	2.18
AUC_6	-96.98	-5.02	0.35	21.04	-10.95
AUC_7	29.65	-3.17	-6.33	-21.42	-0.93
AUC_8	75.39	7.12	12.22	42.14	40.69
AUC_9	-193.56	4.56	-8.41	-36.28	-23.16
AUC_{10}	211.35	-7.04	4.23	34.28	-16.91
AUC_{11}	-80.62	4.28	1.09	-35.45	14.12
AUC_{12}	-121.83	-3.28	-6.84	-22.74	-22.05
AUC_{13}	170.87	0.28	2.14	26.53	12.46
AUC_{14}	-64.83	2.47	-4.26	-34.85	0.28
AUC_{15}	-4.73	-1.89	6.93	37.30	7.39

Table SI.2. Values of the fit coefficients of the MLRs designed to estimate the composition of red wine vinegar blends.

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

Term	White Wine	Cider	Apple	Molasses	Rice
Constant	-0.10	-0.02	0.02	1.09	0.02
AUC_1	-40.07	-1.90	1.91	10.03	-3.87
AUC_2	104.90	2.79	-6.11	-24.57	2.18
AUC_3	-24.34	-0.16	9.48	14.51	3.87
AUC_4	-114.71	-0.70	-5.59	11.46	-5.42
AUC_5	160.27	2.92	-12.01	-58.74	-6.40
AUC_6	-159.24	-7.11	15.00	53.45	13.15
AUC_7	35.26	-0.13	11.96	55.38	11.87
AUC_8	47.04	5.55	-13.88	-35.45	0.97
AUC_9	-49.99	4.90	-11.71	-53.58	-12.06
AUC_{10}	96.32	-10.24	20.26	67.59	-9.34
AUC_{11}	-129.39	7.22	-13.10	-60.23	2.37
AUC_{12}	186.08	-3.91	7.93	15.08	-3.24
AUC_{13}	-208.07	1.60	-10.46	-2.87	-0.63
AUC_{14}	118.92	0.79	5.84	-15.00	9.21
AUC_{15}	-22.87	-1.59	0.37	21.72	-2.59

Table SI.3. Values of the fit coefficients of the MLRs designed to estimate the composition of apple cider vinegar blends.

Term	White Wine	Red Wine	Apple	Molasses	Rice
Constant	0.10	0.00	-0.08	0.92	-0.02
AUC_1	-1.25	-0.40	-13.95	-0.09	-3.10
AUC_2	24.67	0.79	-4.00	-9.10	-6.86
AUC_3	-51.72	-1.79	-22.15	24.69	16.27
AUC_4	-8.32	1.30	43.23	8.90	7.13
AUC_5	93.50	1.51	-3.09	-71.16	-24.01
AUC_6	-62.62	-0.66	-1.05	64.24	3.56
AUC_7	25.14	0.91	144.52	-28.47	6.77
AUC_8	-43.83	-5.25	-150.32	15.46	16.06
AUC_9	26.75	0.86	-28.68	-18.79	-0.23
AUC_{10}	-45.23	6.22	67.86	57.56	-23.30
AUC_{11}	102.57	-7.18	-25.40	-90.86	0.48
AUC_{12}	-131.07	5.29	-51.48	99.66	17.18
AUC_{13}	151.02	1.60	85.22	-105.71	-28.45
AUC_{14}	-77.93	-3.34	-39.28	45.78	24.87
AUC_{15}	-4.00	0.78	-9.71	6.10	-4.16

Table SI.4. Values of the fit coefficients of the MLRs designed to estimate the composition of apple vinegar blends.

Term	White Wine	Red Wine	Cider	Molasses	Rice
Constant	0.67	0.03	0.43	0.46	-0.05

AUC$_1$	-21.27	1.04	-19.45	12.80	3.67
AUC$_2$	37.73	-2.27	43.73	-20.04	-6.18
AUC$_3$	-60.38	0.05	-45.68	41.04	6.07
AUC$_4$	77.85	-0.06	27.49	-55.35	-10.69
AUC$_5$	-0.47	-0.09	7.78	-5.67	-0.44
AUC$_6$	-63.47	7.60	-63.82	50.97	13.81
AUC$_7$	21.60	-4.90	60.97	-13.34	-3.56
AUC$_8$	22.70	-4.34	136.20	-48.34	17.01
AUC$_9$	-42.33	1.34	-214.26	59.68	-11.07
AUC$_{10}$	7.84	5.28	54.42	0.64	-15.56
AUC$_{11}$	80.89	-6.97	44.58	-75.75	2.57
AUC$_{12}$	-140.96	5.43	-102.58	132.05	0.18
AUC$_{13}$	171.55	1.53	138.64	-152.14	-9.82
AUC$_{14}$	-88.42	-3.43	-105.94	70.60	15.17
AUC$_{15}$	-3.00	0.62	40.99	1.13	-0.63

Table SI.5. Values of the fit coefficients of the MLRs designed to estimate the composition of molasses vinegar blends.

Term	**White Wine**	**Red Wine**	**Cider**	**Apple**	**Rice**
Constant	-0.09	-0.02	0.05	0.01	0.08
AUC$_1$	-10.31	0.42	-0.82	3.13	-1.38
AUC$_2$	19.71	-1.30	1.00	-8.31	3.93
AUC$_3$	-15.83	1.88	0.64	8.53	-3.62
AUC$_4$	13.86	-1.26	-0.57	-4.37	1.19
AUC$_5$	6.04	-1.88	-1.37	-2.30	0.43
AUC$_6$	1.89	6.22	-6.35	3.80	-13.72
AUC$_7$	-65.31	-3.55	12.69	7.08	24.10
AUC$_8$	44.69	-8.78	4.23	-9.40	5.65
AUC$_9$	-13.96	7.87	-15.31	3.76	-6.34
AUC$_{10}$	-19.53	1.88	11.40	-3.79	-11.08
AUC$_{11}$	94.04	-1.75	2.17	-5.62	-5.91
AUC$_{12}$	-118.36	-5.33	-24.84	25.58	9.07
AUC$_{13}$	137.71	11.95	26.27	-31.85	-19.66
AUC$_{14}$	-60.48	-8.06	-8.85	11.63	20.17
AUC$_{15}$	-16.38	2.05	0.17	1.76	-0.79

Table SI.6. Values of the fit coefficients of the MLRs designed to estimate the composition of rice vinegar blends.

Term	**White Wine**	**Red Wine**	**Cider**	**Apple**	**Molasses**
Constant	-0.18	-0.04	0.06	-0.05	0.81
AUC$_1$	13.13	1.13	0.57	4.31	-7.09
AUC$_2$	-11.52	-0.59	-2.44	-8.34	18.61

AUC_3	-46.51	-2.75	0.85	12.63	11.48
AUC_4	63.28	2.55	2.95	-16.56	-42.72
AUC_5	39.65	2.13	-3.06	5.02	-4.77
AUC_6	-83.16	-2.20	-3.26	7.67	39.38
AUC_7	32.57	0.20	8.73	-12.15	-13.30
AUC_8	-137.62	-9.60	1.89	24.22	27.29
AUC_9	199.98	10.92	-1.60	-36.23	-57.19
AUC_{10}	-93.28	0.65	-7.40	26.90	54.25
AUC_{11}	19.17	-6.90	5.80	-6.50	-29.85
AUC_{12}	78.97	9.30	-3.57	-8.65	-28.48
AUC_{13}	-97.17	-4.49	2.72	6.87	36.20
AUC_{14}	62.00	0.00	-1.06	-3.10	-30.96
AUC_{15}	-36.01	0.56	-0.44	2.98	24.10

Optimization process: training function and MLP parameters

The optimization of an MLP comprises more than the modification of the value of the weights. There are several elements that must be also considered, such as the training function selected, the hidden neuron number (HNN), and the learning coefficients.

The training function controls the training process. It is an algorithm responsible of the modification of the value of the weights during the training phase [Demuth *et al*. 2007]. Despite of the existence of several of these functions, in the case considered the different MLPs constructed employed trainLM, as it is considered the fastest training function due to its memory reduction feature which is used when the computational requirements are huge. In addition, it presents a compromise between conjugate gradient and quasi-Newtonian methods in the learning process [Parmar *et al*. 2011, Demuth *et al*. 2007].

On the other hand, the HNN was optimized following a heuristic method based on several trial-and-error test, and choosing the one that offers the best results, in terms of low error and high R^2 value. Due to the traits of the systems studied (15 input nodes, five output neurons, and 105 samples per binary blend), the range considered was from two to five neurons, so the number of weights was always below the number of data instances that were available. It must be mentioned, that during this optimization, the learning coefficients were maintained constant (Lc = 0.001; Lcd = 0.1; Lci = 10). The results are shown in **Table SI.7** to **Table SI.12**.

Table SI.7. Statistical results attained in the HNN optimization process for the white wine vinegar blends, in terms of R^2 and MAE. The best HNN appears in bold.

HNN	R^2					MAE (% v/v)				
	Red Wine	**Cider**	**Apple**	**Molasses**	**Rice**	**Red Wine**	**Cider**	**Apple**	**Molasses**	**Rice**

2	0.99	0.96	0.94	0.96	0.98	1.2	2.6	1.0	1.9	2.2
3	0.98	0.99	0.99	>0.99	0.99	0.7	1.0	0.4	0.2	1.2
4	0.98	0.92	0.98	0.99	>0.99	0.8	3.1	1.0	1.3	1.1
5	**0.98**	**>0.99**	**>0.99**	**0.99**	**>0.99**	**1.1**	**0.5**	**0.1**	**0.5**	**0.8**

Table SI.8. Statistical results attained in the HNN optimization process for the red wine vinegar blends, in terms of R^2 and MAE. The best HNN appears in bold.

HNN	R^2					MAE (% v/v)				
	White Wine	**Cider**	**Apple**	**Molasses**	**Rice**	**White Wine**	**Cider**	**Apple**	**Molasses**	**Rice**
2	0.97	0.99	0.98	0.97	>0.99	3.0	2.2	1.9	3.5	1.6
3	0.99	0.98	>0.99	0.99	>0.99	1.8	2.3	0.5	2.4	1.8
4	**0.98**	**>0.99**	**>0.99**	**0.99**	**>0.99**	**2.7**	**0.7**	**0.2**	**1.8**	**0.5**
5	0.99	>0.99	>0.99	0.99	>0.99	2.2	1.3	0.5	1.4	1.3

Table SI.9. Statistical results attained in the HNN optimization process for the apple cider vinegar blends, in terms of R^2 and MAE. The best HNN appears in bold.

HNN	R^2					MAE (% v/v)				
	White Wine	**Red wine**	**Apple**	**Molasses**	**Rice**	**White Wine**	**Red wine**	**Apple**	**Molasses**	**Rice**
2	0.68	0.99	0.96	0.85	0.96	7.9	1.1	2.8	2.0	2.5
3	0.49	0.98	0.68	0.95	0.98	10.2	1.8	5.8	1.6	1.6
4	0.97	0.99	0.88	0.84	>0.99	2.5	1.1	4.1	2.7	1.1
5	**0.98**	**0.98**	**0.99**	**0.93**	**0.99**	**2.6**	**1.3**	**1.1**	**0.9**	**1.2**

Table SI.10. Statistical results attained in the HNN optimization process for the apple vinegar blends, in terms of R^2 and MAE. The best HNN appears in bold.

HNN	R^2					MAE (% v/v)				
	White Wine	**Red wine**	**Cider**	**Molasses**	**Rice**	**White Wine**	**Red wine**	**Cider**	**Molasses**	**Rice**
2	0.99	0.99	0.97	>0.99	0.90	2.5	0.8	2.0	1.1	4.4
3	0.99	0.99	0.86	0.96	0.97	1.2	0.5	2.4	4.0	1.8
4	**>0.99**	**>0.99**	**0.98**	**0.99**	**>0.99**	**1.5**	**0.4**	**1.2**	**1.8**	**0.6**
5	>0.99	>0.99	0.97	0.99	0.99	1.2	0.2	1.6	1.6	1.2

Table SI.11. Statistical results attained in the HNN optimization process for the molasses vinegar blends, in terms of R^2 and MAE. The best HNN appears in bold.

HNN	R^2					MAE (% v/v)				
	White Wine	**Red wine**	**Cider**	**Apple**	**Rice**	**White Wine**	**Red wine**	**Cider**	**Apple**	**Rice**
2	>0.99	0.99	>0.99	0.97	0.99	1.3	1.0	1.1	1.7	1.7
3	>0.99	0.99	>0.99	0.97	0.99	1.9	1.7	1.5	2.4	2.6

4	**>0.99**	**0.99**	**>0.99**	**0.99**	**>0.99**	**0.4**	**0.9**	**0.1**	**0.8**	**0.4**
5	>0.99	0.99	>0.99	0.99	0.99	1.1	1.1	0.2	0.9	1.5

Table SI.12. Statistical results attained in the HNN optimization process for the rice vinegar blends, in terms of R^2 and MAE. The best HNN appears in bold.

HNN	R^2					MAE (% v/v)				
	White Wine	**Red wine**	**Cider**	**Apple**	**Molasses**	**White Wine**	**Red wine**	**Cider**	**Apple**	**Molasses**
2	0.98	0.95	0.99	0.98	0.68	1.6	1.1	2.1	1.4	2.9
3	0.99	>0.99	>0.99	>0.99	0.99	1.4	0.2	0.7	0.9	1.2
4	0.99	0.98	>0.99	0.99	0.99	1.2	1.1	1.3	1.9	1.1
5	**0.99**	**>0.99**	**>0.99**	**>0.99**	**0.99**	**1.2**	**0.3**	**0.6**	**0.7**	**1.0**

The optimal HNN was always four or five. In every case, the number of weights was below the number of data instances (76% and 95%, when four and five neurons are used, respectively). When the performance is similar, the combination with the less HNNs must be selected, as the computational requirements of such model will be smaller.

The three parameters regarding the learning coefficient (Lc, Lcd and Lci) were optimized through a Box-Wilson Central Composite design 2^3 + star points experimental design. The ranges tested were from 0.001 to 1 in the case of Lc and Lcd, and from 2 to 100, for Lci. The studied responses were the R^2 and the estimation error, in terms of MAE. During this optimization, the HNN was set to the optimal value determined during the previous analysis for each binary mixture. The results can be seen in the Results and discussion section of the main article.

References

Ashrafi, S., Jahanbakhsh, O., Alizadeh, D. (2014). Application of artificial neural network in non-destructive Compton scattering densiometry. *Nuclear Intruments and Methods in Physics Research A, 760*, 1-4.

Demuth, H., Beale, M., Hagan, M. (2007). Neural network toolbox for use with MATLAB® user’s guide. Version 5. Ninth printing Revised for Version 5.1 (online only)].

Parmar, H., Jindoliya, D.A. (2011).Artificial neural network based modelling of desiccant wheel. *Energy and Buildings*, *43*, 3505-3513.

Torrecilla, J.S., Mena, M.L., Yáñez-Sedeño, P., García, J. (2007). Application of artificial neural network to the determination of phenolic compounds in olive oil mill wastewater. *Journal of Food Engineering*, *81*, 544-552.