Essential oil profile of *Origanum vulgare* subsp. *vulgare* native population from Rtanj *via* chemometrics tools

Milica Aćimović[1]*, Lato Pezo[2], Stefan Ivanović[3], Katarina Simić[3], Jovana Ljujic[4]

1-Institute of Field and Vegetable Crops Novi Sad – National Institute of the Republic of Serbia, Maksima Gorkog 30, 21000 Novi Sad, Serbia

2-University of Belgrade, Institute of General and Physical Chemistry, Studentski trg 12, 11000 Belgrade, Serbia

3-University of Belgrade, Institute of Chemistry, Technology and Metallurgy – National Institute of the Republic of Serbia, Njegoševa 12, 11000 Belgrade, Serbia

4-University of Belgrade, Faculty of Chemistry, Department of Organic Chemistry, Studentski trg 12-16, 11000 Belgrade, Serbia

Milica Aćimović: milica.acimovic@ifvcns.ns.ac.rs
Lato Pezo: latopezo@yahoo.co.uk
Stefan Ivanović: stefan.ivanovic@ihtm.bg.ac.rs
Katarina Simić: katarina.simic@ihtm.bg.ac.rs
Jovana Ljujić: jovanalj@chem.bg.ac.rs

ABSTRACT

The aim of this study was to predict the retention indices of chemical compounds found in the aerial parts of *Origanum vulgare* subsp. *vulgare* essential oil, obtained by hydrodistillation and analyzed by GC-MS. A total number of 28 compounds were detected in the essential oil. The compounds with the highest relative concentrations were germacrene D (21.5%), 1,8-cineole (14.2%), sabinene (14.0%) and *trans*-caryophyllene (13.4%). The retention time was predicted by using the quantitative structure–retention relationship, using seven molecular descriptors chosen by factor analysis and genetic algorithm. The chosen descriptors were mutually uncorrelated, and they were used to develop an artificial neural network model. A total number of 28 experimentally obtained retention indices (log *RI*) were used to set up a predictive quantitative structure-retention relationship model. The coefficient of determination for the training cycle was 0.998, indicating that this model could be used for predicting retention indices for *O. vulgare* subsp. *vulgare* essential oil compounds.

Keywords: oregano, essential oil, hydrodistillation, GC-MS, QSSR, ANN

Introduction

Origanum is an important genus with multipurpose medicinal and spice plants. It belongs to the family Lamiaceae and is comprised of 42 species divided into 10 sections. Most *Origanum* species are locally distributed within the Mediterranean region where they grow in the mountainous areas on the islands, with high endemism rate (Lukas, 2010). However, among all sections in the genus, only section Origanum is monospecific, consisting of the species *O. vulgare*, but with the largest distribution area. Because of this, *O. vulgare* is an extremely variable species that includes six subspecies (subsp. *vulgare*, subsp. *glandulosum*, subsp. *gracile*, subsp. *hirtum*, subsp. *viridulum* and subsp. *virens*) which are characterized by a high morphological and chemical variability (Chishti et al., 2013; Kosakowska and Czupa, 2018). In general, differences in morphological and chemical features represent environmental adaptation. For example, sessile glands on leaves and the color of bracts and corollas are the main morphological traits (Kokkini et al., 1994). Furthermore, the yield and quality of the essential oil depends on genetics and is strongly affected by the environmental influences (Goliaris et al., 2002;Toncer et al., 2009).

O. vulgare subsp. *vulgare* is the most widespread species in Europe, and has longstanding use in traditional medicine for its carminative, stomachic, emmenagogue, and expectorant effects for treating cramps, flatulence, cough, or menstrual problems (Oniga et al., 2018). The main bioactive components of *O. vulgare* are essential oil and phenolic components, generated from cymyl- pathway such as γ-terpinene, *p*-cymene, carvacrol and thymol (Lukas, 2010;Stanojević et al., 2016). Their ratio represents the quality of the oil and indicates the aroma value (Morsy, 2017).

Quantitative structure retention relationship (QSRR) approach provides a deeper insight into the relation between the chemical compounds, their structure and the physicochemical or biological properties (Wolfender et al., 2015). Gas Chromatography coupled with Mass Spectrometry (GC-MS) extracts a huge amount of data, which could be compared and reproduced, and it also shows the exact retention time indices for large sets of compounds in different biological materials. The chemical compound structure is explained by the mathematical models, described by so-called molecular descriptors, which encode its data by the symbolic representation of a molecule into a numerical value (Héberger, 2007;Micić et al., 2019). Lately, various investigations were assigned to the QSRR coupled with GC-MS data analyses (Kaliszan, 2007;Khezeli et al., 2016;Marrero-Ponce et al., 2018; Wu et al., 2013). The relation between the molecular descriptors and the retention time can be established by using various mathematical tools, such as the artificial neural network (ANN), which was proven to be excellent in solving non-linear problems (Wolfender et al., 2015;Zisi et al., 2017), or by using machine learning algorithms (Tropsha and Golbraikh, 2007).

The aim of this study was to establish a new QSRR model for the prediction ofthe retention times of chemical compounds found in *O. vulgare* subsp. *vulgare* essential oil, obtained by hydrodistillation and analyzed by GC-MS using the coupled genetic algorithm (GA) and factor analysis (FA) variable selection method and the artificial neural network (ANN) model.

Experimental

Plant material

Origanum vulgare subsp. *vulgare* was collected on the 7th July 2018, on Mt. Rtanj. The plant species were at full flowering by this date. The plant aboveground parts were harvested manually at around 2-3 cm above the soil surface, and the biomass was placed in an air-dryer until constant weight at 35 °C to avoid essential oil losses. Voucher specimens were confirmed and deposited at the Herbarium BUNS, the University of Novi Sad, Faculty of Sciences, Department of Biology and Ecology, under the acquisition number 2-1450.

Essential oil isolation

Air-dried aerial parts of *O. vulgare* subsp. *vulgare* were submitted to hydrodistillation according to Ph. Eur. 5.0(Ph. Eur. 5.0) by using the Clevenger apparatus. The 30 g of the plant material was placed in round-bottomed flask of 1 L and 400 mL of distilled water was added. Then it was heated to the boiling point. The steam in combination with the essential oils was distilled into a graduated tube for 2h. After separation of essential oil from aqueous phase it was dried over anhydrous Na_2SO_4 and stored in a dark glass vial at 4 °C for further analysis. The essential oils yields were calculated on dry-weight basis, and average content of essential oil was 0.12%.

GC-MS analysis

GC-MS analysis was carried out using an HP 5890 gas chromatograph coupled to an HP 5973 MSD and fitted with a capillary column HP-5MS. The carrier gas was helium, and its inlet pressure was 25 kPa and linear velocity of 1 mL/min at 210 °C. The injector temperature was 250 °C, and analysis was conducted under splitless injection mode. Mass detection was carried out under source temperature conditions of 200 °C and interface temperature of 250 °C. The EI mode was set at electron energy, 70 eV with mass scan range of 40–350 amu. Temperature was programmed from 60 °C to 285 °C at a rate of 4.3 °C/min. The components were identified based on their linear retention index relative to C_8-C_{32} *n*-alkanes, by the comparison with data reported in the literature (Wiley and NIST databases). Quantification was done by external standard method using calibration curves generated by running GC analysis of representative authentic compounds.

Artificial neural network (ANN)

A multi-layer perceptron model (MLP) consisted of the three layers (input, hidden and output) was used in this paper, having in mind that it is well known and proven as being capable of approximating nonlinear functions (Aalizadeh et al., 2016). Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm was used for ANN modelling. The experimental database was randomly divided into: train, testing and validation parts (60, 20 and 20%, respectively) for ANN modelling. A series of different neural network topologies was tested. The number of hidden neurons varied from 1 to 20 and 1,000,000 networks were tested, using random initial values of weights and biases. The weight coefficient was calculated during the training period, with the initial assumptions of parameters, which were adjusted using ANN structure and fitting (Kojic and Omorjan, 2018; Xuet al., 2015). The optimization process was performed on the basis of validation error minimization. Statistical investigation of the data has been performed mainly by the Statistica 10 software (Statistica, 2010).

Molecular descriptors

Coupled factor analysis and genetic algorithm were used to select the most relevant molecular descriptors for the representation of the retention indices (Goldberg, 1989; Tropsha, 2010), and a calculation was performed using Heuristic Lab (HeuristicLab, 3.3). The correlation between the obtained descriptors was examined and collinear descriptors were detected using factor analysis. GA was used to select the most appropriate molecular descriptors to develop a reliable model for the prediction of retention times of the compounds found in *O. vulgare* subsp. *vulgare* essential oil.

QSRR analysis

The molecular structure was introduced in the quantitative structure retention relationship (QSRR) calculation in the form of *.smi* files, which represented the structure of a molecule in a simplified molecular input line (Matyushin, et al. 2019). The calculation of the specified molecular descriptors for each chemical compound obtained in the GC-MS analysis was performed using PaDel-descriptor software (Dong et al., 2015; Yap, 2011). The PaDel-descriptor software was used

to calculate the 1875 molecular descriptors (1444 1D and 2D descriptors and 431 3D descriptors), which included: constitutional descriptors, topological descriptors, connectivity indices, information indices, 2D and 3D autocorrelations descriptors, Burden eigenvalues descriptors, eigenvalue-based indices, geometrical descriptors, WHIM descriptors, functional group counts, atom-centered fragments and molecular properties.

Global sensitivity analysis

Global sensitivity analysis was used to explore the relative influence of molecular descriptors on retention time (Yoon et al., 2017). This method was applied on the basis of the weight coefficients of the developed ANN.

Results and Discussion

Essential oil composition

A total number of 28 compounds were detected in the *O. vulgare* subsp. *vulgare* essential oil in this study, representing 99.5% of the total oil composition (Table 1). The compounds with the highest relative concentration in *O. vulgare* subsp. *vulgare* essential oil were germacrene D (21.5%), 1,8-cineole (14.2%), sabinene (14.0%) and *trans*-caryophyllene (13.4%). Out of these, 15 compounds had average relative concentrations over 1.0%. Monoterpene hydrocarbons (47.9%) and sesquiterpene hydrocarbons (42.9%) were the dominant classes. According to the obtained results, *O. vulgare* subsp. *vulgare* collected at Mt. Rtanj can be classified as germacrene D chemotype. This chemotype is already described (Mockute et al., 2001). Differences among the oregano accessions with respect to morphological traits and chemical constituents of essential oils, indicate the existence of intraspecific variations and chemical polymorphism (Aćimović et al., 2020; Radusiene et al., 2005). Subspecies which accumulate carvacrol and/or thymol and their precursors (*γ*-terpinene and *p*-cymene) contain low amounts of other monoterpenes (Kosakowska and Czupa, 2018).

Table 1. Chemical composition of *O. vulgare* subsp. *vulgare* essential oil from dry aerial parts.

No	Compound	RI[a]	RI[b]	%
1	*α*-Thujene	915	924	0.1
2	*α*-Pinene	926	932	1.5
3	Camphene	944	946	0.9
4	Sabinene	967	969	14.0
5	*β*-Pinene	971	974	3.9
6	*β*-Phellandrene	1026	1025	1.6
7	1,8-Cineole	1028	1026	14.2
8	*cis*-*β*-Ocimene	1033	1032	6.8
9	*trans*-*β*-Ocimene	1043	1044	4.5
10	*γ*-Terpinene	1052	1054	0.4
11	Borneol	1160	1165	1.2
12	Terpinen-4-ol	1172	1174	0.3

13	*α*-Terpineol	1187	1190	0.7
14	Bornyl acetate	1286	1287	0.1
15	*β*-Bourbonene	1384	1387	1.9
16	*trans*-Caryophyllene	1420	1408	13.4
17	*α*-Humulene	1454	1452	2.1
18	9-*epi-trans*-Caryophyllene	1461	1464	0.5
19	Germacrene D	1486	1484	21.5
20	Bicyclogermacrene	1497	1500	1.7
21	(*trans,trans*)-*α*-Farnesene	1509	1505	0.7
22	*δ*-Cadinene	1524	1513	1.1
23	Germacrene D-4-ol	1575	1574	0.8
24	Spathulenol	1577	1577	0.6
25	Caryophyllene oxide	1583	1582	3.9
26	Humulene epoxide II + *β*-Oplopenone	1606	1608	0.3
27	*epi-α*-Murrolol (=*tau*-muurolol)	1640	1640	0.4
28	*α*-Cadinol	1654	1652	0.9
	Monoterpene hydrocarbons			47.9
	Oxygenated monoterpenes			2.3
	Sesquiterpene hydrocarbons			42.9
	Oxygenated sesquiterpenes			6.9
	Total identified			99.5

RI^a – Retention Index calculated; RI^b – Retention Index from the NIST webbook database.

Artificial neural network (ANN)

Graphical representation of experimentally obtained retention time indices of *O. vulgare* subsp. *vulgare* essential oil composition (RI^a), the retention time indices found in NIST database (RI^b) and the retention time indices predicted by the ANN model ($RI_{pred.}$) were presented in Figure 1.
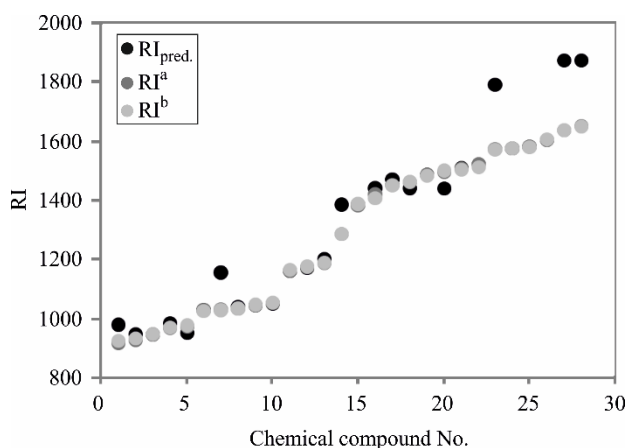

Figure 1. Retention time indices of the *O. vulgare* subsp. *vulgare* essential oil composition, from: experimentally obtained GC-MS data (RI^a); NIST database (RI^b) and predicted by the ANN ($RI^{pred.}$).

105

The nonlinear relationship between *RI*s and the selected descriptors, applying the ANN technique was used in this paper. The statistical results of the MLP 7-12-1 network are shown in Table 2.

Table 2. ANN model summary (performance and errors), for training, testing and validation cycles

Net. name	Performance			Error			Train. algor.	Error funct.	Hidden activat.	Output activat.
	Train.	Test.	Valid.	Train.	Test.	Valid.				
MLP 7-12-1	0.998	0.958	0.999	53.764	2311.444	16294.67	BFGS 38	SOS	Tanh	Exponential

*Performance term represent the coefficients of determination, while error terms indicate a lack of data for the ANN model. ANN cycles: Train. – training, Test. – testing, Valid. – validation, algor. –algorithm, funct. – function, activat. – activation.

The better prediction of *RI*s was obtained in the training cycle, which was expected, because more chemical compounds retention time indices were used in the calculation compared with testing cycle. This is also obvious from Table 2, where the training set performance reached r^2 of 0.998, while the r^2 for testing set was lower. Also, better results for r^2 were obtained in training cycle, due to the fact that these data were used for the modelling of ANN, while the data in testing and verification cycles were used for testing purposes and to explore the quality of the ANN model created in training cycle. Obtained results reveal the reliability of the ANN models for predicting the *RI*s of compounds in *O. vulgare* subsp. *vulgare* essential oil determined by GC-MS.

Molecular descriptors

Seven molecular descriptors were chosen by FA and GA analyses for predictions of *RI* in the obtained ANN model.

- *Autocorrelation descriptors*
  1. ATSC3v - Centered Broto-Moreau autocorrelation - lag 3 / weighted by van der Waals volumes;
  2. AATSC5c - Average centered Broto-Moreau autocorrelation - lag 5 / weighted by charges;
  3. AATSC1v - Average centered Broto-Moreau autocorrelation - lag 1 / weighted by van der Waals volumes;
  4. AATSC1e -Average centered Broto-Moreau autocorrelation - lag 1 / weighted by Sanderson electronegativities;
  5. GATS5p - Geary autocorrelation - lag 5 / weighted by polarizabilities,
- *Information content descriptors:*
  6. BIC2 - Bond information content index (neighbourhood symmetry of 2-order);
  7. MIC0 - Modified information content index (neighbourhood symmetry of 0-order).

The above mentioned molecular descriptors encode different aspects of the molecular structure and they were used to develop a QSRR model for prediction of retention indices of

compounds found in *O. vulgare* subsp. *vulgare* essential oil. The values of the selected descriptors were displayed in Table 3.

Table 3. Molecular descriptors chosen by a genetic algorithm

Descriptors	Autocorrelation descriptors					Information content	
No	ATSC3v	AATSC5c	AATSC1v	AATSC1e	GATS5p	BIC2	MIC0
1	-881.860	1.2E-04	3.158	3.3E-04	0.563	0.768	6.797
2	-1182.918	-1.1E-04	3.158	3.3E-04	0.762	0.768	6.797
3	-1209.560	4.2E-04	3.158	3.3E-04	0.431	0.698	6.797
4	-868.539	3.1E-04	3.158	3.3E-04	0.594	0.742	6.797
5	-1169.597	-2.1E-04	3.158	3.3E-04	0.766	0.742	6.797
6	-694.032	2.5E-04	0.000	8.4E-17	0.798	0.790	6.797
7	-984.187	3.2E-05	1.156	2.1E-03	0.973	0.545	9.466
8	-823.247	2.7E-04	-3.410	-3.6E-04	0.717	0.686	6.797
9	-753.977	2.4E-04	0.000	8.4E-17	0.745	0.688	6.797
10	-1240.975	-2.0E-04	4.092	-2.6E-03	0.069	0.647	9.466
11	-901.854	9.7E-04	1.111	-2.8E-03	0.634	0.763	9.466
12	-1197.667	2.0E-04	0.865	1.0E-03	0.313	0.679	10.663
13	-1164.268	1.5E-04	4.159	4.4E-04	0.932	0.704	6.797
14	-1354.761	5.4E-05	2.131	2.2E-04	0.963	0.753	6.797
15	-1530.600	1.4E-04	0.000	8.4E-17	0.930	0.683	6.797
16	-1354.761	5.4E-05	2.131	2.2E-04	0.963	0.753	6.797
17	-739.324	1.4E-04	0.000	8.4E-17	0.941	0.772	6.797
18	-1398.721	4.3E-05	2.131	2.2E-04	0.960	0.711	6.797
19	-1124.305	2.2E-04	-2.244	-2.4E-04	0.810	0.673	6.797
20	-731.331	-4.4E-06	2.131	2.2E-04	0.970	0.711	6.797
21	-985.861	2.1E-04	0.779	-1.7E-03	0.972	0.735	8.851
22	-1352.199	9.3E-04	4.804	-1.6E-03	1.039	0.719	8.941
23	-1468.406	2.1E-04	2.708	1.7E-03	0.988	0.711	8.941
24	-667.006	-3.3E-05	2.042	1.2E-04	1.068	0.735	8.941
25	-986.239	-1.6E-04	2.828	-1.6E-03	0.984	0.737	8.851
26	-986.239	-1.6E-04	2.828	-1.6E-03	0.984	0.737	8.851

The most comprehensive explanation about the molecular descriptors could be found in the Handbook of Molecular Descriptors (Todeschini and Consonni, 2000). Table 4 represents the correlation matrix among these descriptors. There were no statistically significant correlation between selected molecular descriptors; therefore, they could be used for QSRR model building.

Table 4. The correlation coefficient matrix for the selected descriptors by GA

	AATSC5c	AATSC1v	AATSC1e	GATS5p	BIC2	MIC0
ATSC3v	0.000	-0.318	-0.108	0.003	0.188	-0.051
	p=0.999	p=0.113	p=0.599	p=0.988	p=0.359	p=0.804
AATSC5c		-0.097	-0.251	-0.040	0.100	0.135
		p=0.638	p=0.216	p=0.846	p=0.628	p=0.511
AATSC1v			-0.083	-0.062	0.157	0.177
			p=0.688	p=0.765	p=0.444	p=0.386
AATSC1e				0.179	-0.259	-0.277
				p=0.382	p=0.202	p=0.171
GATS5p					0.161	-0.166
					p=0.433	p=0.419
BIC2						-0.334
						p=0.096

QSRR model validation

The factor analysis was performed on the molecular descriptor data obtained from PaDel-descriptor software, in order to eliminate the descriptors with equal or almost equal factor values. Only one of the correlated descriptors remained in the GA calculation. GA was used to select the most appropriate set of molecular descriptors which were left in the calculation, while the selection of the most relevant set of descriptors was used in the evolution simulation(Mohammadhosseini, 2013; Nekoei et al., 2015). The number of elements was equal to the number of the molecular descriptors obtained in the PaDel-descriptor, and the population of the first generation in the GA calculation was selected randomly. The probability of generating zero for the element was set at least 60%. The operators used in the simulation were: crossover (90% probability) and mutation (0.5%). A population size of 100 elements was chosen for GA, and evolution was allowed for over 50 generations. The evolution of the generations was stopped when 90% of the generations took the same fitness.

The calibration and predictive capability of a QSRR model should be tested through the model validation. The most widely used squared correlation coefficient (r^2) can provide a reliable indication of the fitness of the model; thus, it was employed to validate the calibration capability of a QSRR model. The quality of the model fit was tested in Table 5, in which the lower reduced chi-square (χ^2), mean bias error (*MBE*), root mean square error (*RMSE*), mean percentage error (*MPE*) are presented (Arsenović et al., 2015).

Table 5. The "goodness of fit" tests for the developed ANN model

χ^2	*RMSE*	*MBE*	*MPE*
7519.609	85.032	-36.684	3.250

χ^2 - reduced chi-square, *MBE* - mean bias error, *RMSE* - root mean square error, *MPE* - mean percentage error.

The predicted *RI*s which were presented in Figure 2 confirmed the adequate prediction of the retention indices, for constructed ANN, by showing the relationship between the predicted and experimental retention values.
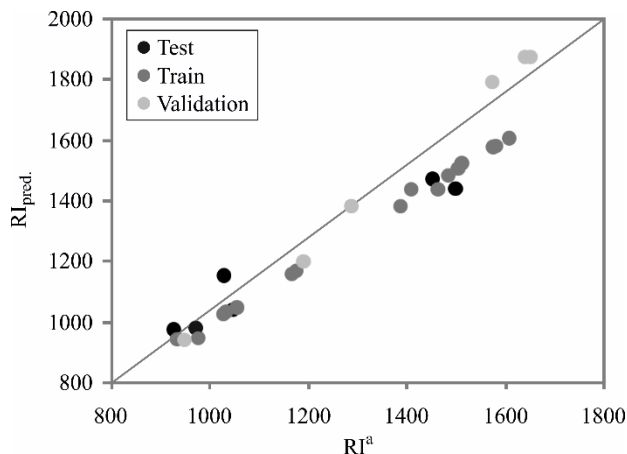

Figure 2. Comparison of experimentally obtained RIs with ANN predicted values

Global sensitivity analysis- Yoon's interpretation method

In this section the influence of seven the most important input variables, identified using genetic algorithm on *RI* was studied. According to the Figure 3, ATSC3v was the most influential parameter with approximately relative importance of 18.8%, while the influence of AATSC1v, AATSC1e, GATS5p and AATSC5c were 14.9, 14.6%, 13.6% and 13.3%, respectively. MIC0 and BIC2 were influential at levels 13.2% and 11.6%, respectively.
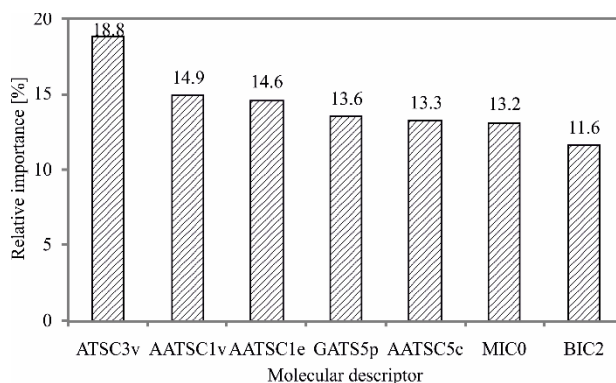
Figure 3 The relative importance of the molecular descriptors on RI, determined using Yoon interpretation method

Conclusion

The QSRR model for the estimation of retention times of *O. vulgare* subsp. *vulgare*essential oil compounds was developed for 28 compounds using the ANN modelling approach. The results demonstrated that the ANN model was adequate in predicting retention times

of found chemicalcompounds. A suitable model with high statistical quality and low prediction errors was derived.

The following five molecular descriptors were suggested by genetic algorithm: five 2D autocorrelation molecular descriptors (ATSC3v, AATSC1v, AATSC1e, AATSC5c and GATS5p) and two Information content descriptors (MIC0 and BIC2), that predicted retention times of the obtained compounds. Selected molecular descriptors were not mutually correlated and the obtained descriptors were suitable for QSRR model building. The results demonstrated that the ANN model was adequate to predict the *RIs* of the compounds in *O. vulgare* subsp. *vulgare*essential oil obtained by hydrodistillation and analysed by GC-MS. The coefficient of determination for training cycle was 0.998, which is a good indication that this model could be used for the prediction of retention time.

Acknowledgment

Conflict-of-Interest Statement

Authors declare no conflict of interest.

References

Aalizadeh, R., Thomaidis, N.S., Bletsou, A.A.,& Gago-Ferrero, P. (2016). Quantitative structure–retention relationship models to support nontarget high-resolution mass spectrometric screening of emerging contaminants in environmental samples. Journal of Chemical Information and Modeling, 56, 1384-1398; https://doi.org/10.1021/acs.jcim.5b00752.

Aćimović, M., Zorić, M., Zheljazkov, V., Pezo, L., Čabarkapa, I., Stanković Jeremić, J.,& Cvetković, M. (2020). Chemical characterization and antibacterial activity of essential oil of medicinal plants from Eastern Serbia. Molecules, 25, 5482; DOI:10.3390/molecules25225482.

Arsenović, M., Pezo, L., Stanković, S.,& Radojević, Z. (2015). Factor space differentiation of brick clays according to mineral content: Prediction of final brick product quality. Applied Clay Science, 115, 108-114; https://doi.org/10.1016/j.clay.2015.07.030.

Chishti, S., Kaloo, Z.A.,& Sultan P. (2013). Medicinal importance of genus *Origanum*: A review. Journal of Pharmacognosy and Phytotherapy, 5, 170-177; DOI: 10.5897/JPP2013.0285

Council of Europe (2004). European Pharmacopoeia. 5 ed. Strasbourg, 1206-1208.

Dong, J., Cao, D.S., Miao, H.Y., Liu, S., Deng, B.C., Yun, Y.H., Wang, N.N., Lu, A.P., Zeng, W.B.,& Chen, A.F. (2015). ChemDes: an integrated web-based platform for molecular descriptor and fingerprint computation. Journal of Cheminformatics,7, 60. doi: 10.1186/s13321-015-0109-z.

Goldberg, D.E. (1989). Genetic algorithms in search, optimisation and machine learning. Addison-Wesley, Massachusetts, Boston, USA.

Goliaris A.H., Chatzopoulou P.S.,& Katsiotis S.T. (2002). Production of new Greek oregano clones and analysis of their essential oils. Journal of Herbs, Spices and Medicinal Plants, 10, 29-35; https://doi.org/10.1300/J044v10n01_04.

Héberger, K. (2007). Quantitative structure–(chromatographic) retention relationships. Journal of Chromatography A,1158, 273-305;https://doi.org/10.1016/j.chroma.2007.03.108.

HeuristicLab, 3.3 (Heuristic and Evolutionary Algorithms Laboratory, ver. 3.3.16) (Accessed 20 December 2020). https://dev.heuristiclab.com/trac.fcgi/

Hollas, B., Gutman, I.,& Trinajstić, N. (2005). On reducing correlations between topological indices. Croatica Chemica Acta, 78, 489-492.

Kaliszan, R. (2007). QSRR: Quantitative Structure - (Chromatographic) Retention Relationships. Chemical Reviews, 107, 3212-3246; https://doi.org/10.1021/cr068412z.

Khezeli, T., Daneshfar, A.,& Sahraei, R. (2016). A green ultrasonic-assisted liquid–liquid microextraction based on deep eutectic solvent for the HPLC-UV determination of ferulic, caffeic and cinnamic acid from olive, almond, sesame and cinnamon oil. Talanta, 150, 577-585; doi: 10.1016/j.talanta.2015.12.077.

Kojic, P.,& Omorjan, R. (2018). Predicting hydrodynamic parameters and volumetric gas–liquid mass transfer coefficient in an external-loop airlift reactor by support vector regression. ChemicalEngineering and Research and Design,125, 398-407; https://doi.org/10.1016/j.cherd.2017.07.029.

Kokkini S., Karousou R.,&Vokou D. (1994). Pattern of geographic variations of*Origanum vulgare* trichomes and essential oil content in Greece. Biochemical Systematics and Ecology, 22, 517-528; https://doi.org/10.1016/0305-1978(94)90046-9.

Kosakowska O.,& Czupa W. (2018). Morphological and chemical variability of common oregano (*Origanum vulgare* L. subsp. *vulgare*) occurring in eastern Poland. Herba Polonica, 64, 11-21; DOI: 10.2478/hepo-2018-0001

Lukas B. (2010). Molecular and phytochemical analyses of the genus *Origanum* L. (Lamiaceae). Doctoral Dissertation, Department of Botany and Biodiversity Research, Faculty of Life Sciences, Vienna University.

Marrero-Ponce, Y., Barigye, S.J., Jorge-Rodriguez, M.E.,& Tran-Thi-Thu, T. (2017). QSRR prediction of gas chromatography retention indices of essential oil components. Chemical Papers, 72, 57-69; https://doi.org/10.1007/s11696-017-0257-x.

Matyushin, D., Sholokhova, A.Y.,& Buryak, A.K. (2019). A deep convolutional neural network for the estimation of gaschromatographic retention indices. Journal of Chromatography A, 1607, 460395; https://doi.org/10.1016/j.chroma.2019.460395.

Micić, D., Ostojić, S., Pezo, L., Blagojević, S., Pavlić, B., Zeković, Z.,& Đurović, S. (2019). Essential oils of coriander and sage: Investigation of chemical profile, thermal properties and QSRR analysis. Industrial Crops and Products, 138, 111438; https://doi.org/10.1016/j.indcrop.2019.06.001.

Mockute, D., Bernotiene, G.,& Judzentiene A. (2001). The essential oil of *Origanum vulgare* L.ssp. *vulgare* growing wild in Vilnius district (Lithuania). Phytochemistry, 57, 65-69; https://doi.org/10.1016/S0031-9422(00)00474-X.

Mohammadhosseini, M. (2013). Novel PSO-MLR algorithm to predict the chromatographic retention behaviors of natural compounds. Analytical Chemistry Letters, 3, 226-248; https://doi.org/10.1080/22297928.2013.861164.

Moreau, G.,& Broto, P. (1980). The autocorrelation of a topological structure: A new molecular descriptor. Nouveau Journal De Chimie, 4, 359-360.

Morsy N.F.S. (2017). Chemical structure, quality indices and bioactivity of essential oil constituents. In: Active Ingredients from Aromatic and Medicinal Plants, Edited by: El-Shemy H. InTechOpen, London, UK.

Nekoei, M., Mohammadhosseini, M.,& Pourbasheer, E. (2015).QSAR study of VEGFR-2 inhibitors by using genetic algorithm-multiple linear regressions (GA-MLR) and genetic

algorithm-support vector machine (GA-SVM): a comparative approach. Medicinal Chemistry Research,24, 3037-3046; https://doi.org/10.1007/s00044-015-1354-4.

Oniga, I., Puscas, C., Silaghi-Dumitrescu, R., Olah, N.K., Sevastre, B., Marica, R., Marcus, I., Sevastre-Berghian, A.C., Benedec, D., Pop, C.E.,& Hanganu, D. (2018). *Origanum vulgare* ssp. *vulgare*: chemical composition and biological studies. Molecules, 23, 2077; doi:10.3390/molecules23082077

Radusiene, J., Ivanauskas, L., Janulis, V.,& Jakštas V. (2008) Composition and variability of phenolic compounds in*Origanum vulgare* from Lithuania. Biologija, 54, 45-49

Stanojević, Lj.P., Stanojević, J.S., Cvetković, D.J.,& Ilić, D.P. (2016). Antioxidant activity of oregano essential oil (*Origanum vulgare* L.)*.* Biologica Nyssana, 7, 131-139.

Statistica 10 software (StatSoft, Inc. STATISTICA, ver. 10, data analysis software system) (Accessed 15 December 2018).

Todeschini, R.,& Consonni, V. (2000). Handbook of Molecular Descriptors, Methods and Principles in Medicinal Chemistry. Wiley-VCH Verlag GmbH, Weinheim, Germany.

Toncer, O., Karaman, S., Kizil, S.,& Diraz E. (2009). Changes in essential oil composition of oregano (*Origanum onites* L.) due to diurnal variations at different development stages. Notulae Botanicae Horti Agrobotanici Cluj-Napoca, 37, 177-181; https://doi.org/10.15835/nbha3723188.

Tropsha, A. (2010). Best practices for QSAR model development, validation, and exploitation.Molecular Informatics, 29, 476-488; https://doi.org/10.1002/minf.201000061.

Tropsha, A.,&Golbraikh, A. (2007). Predictive QSAR modeling workflow, model applicability domains, and virtual screening. Current Pharmaceutical Design, 13, 3494-3504; DOI:10.2174/138161207782794257

Wolfender, J.L., Martia, G., Thomas, A.,& Bertranda, S. (2015). Current approaches and challenges for the metabolite profiling of complex natural extracts. Journal of Chromatography A, 1382, 136-164; https://doi.org/10.1016/j.chroma.2014.10.091

Wu, L., Gong, P., Wu, Y., Liao, K., Shen, H., Qi, Q., Liu, H., Wang, G.,& Hao, H. (2013). An integral strategy toward the rapid identification of analogous nontarget compounds from complex mixtures. Journal of Chromatography A, 1303, 39-47; https://doi.org/10.1016/j.chroma.2013.06.041.

Xu, Q., Wei, C., Liu, R., Gu, S., &Xu, J. (2015). Quantitative structure–property relationship study of β-cyclodextrin complexation free energies of organic compounds. Chemometrics and Intelligent Laboratory Systems, 146, 313-321; https://doi.org/10.1016/j.chemolab.2015.06.001

Yap, C.W. (2011).PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. Journal of Computational Chemistry, 32, 1446-1474; doi: 10.1002/jcc.21707.

Yoon, Y., Swales, G., Margavio, T.M. (2017). A comparison of discriminant analysis versus artificial neural networks. Journal of the Operational Research Society, 44, 51-60; DOI: 10.2307/2584434

Zisi, C., Sampsonidis, I., Fasoula, S., Papachristos, K., Witting, M., Gika, H.G., Nikitas, P.,& Pappa-Louisi A. (2017). QSRR modeling for metabolite standards analyzed by two different chromatographic columns using multiple linear regression. Metabolites, 7, 7;https://doi.org/10.3390/metabo7010007.