QSPR Studies of Carbonyl, Hydroxyl, Polyene Indices, and Viscosity Average Molecular Weight of Polymers under Photostabilization Using ANN and MLR Approaches

H. Maouz,[a]* L. Khaouane,[a] S. Hanini,[a] Y. Ammi,[a,b]
M. Hamadache,[a] and M. Laidi[a]

[a] Laboratory of Biomaterials and Transport Phenomena (LBMPT), University of Médéa, Quartier Aïn d'Heb, 26000, Algeria

[b] University Center, Faculty of Science and Technology, Department of Process Engineering, Relizane, Algeria

Abstract

One of the main disadvantages of the use of synthetic or semi-synthetic polymeric materials is their degradation and aging. The purpose of this study was to use artificial neural networks (ANN) and multiple linear regressions (MLR) to predict the carbonyl, hydroxyl, and polyene indices (I_{CO}, I_{OH}, and I_{OP}), and viscosity average molecular weight ($\overline{M}_v$) of poly(vinyl chloride), polystyrene, and poly(methyl methacrylate). These physicochemical properties are considered fundamental during the study of photostabilization of polymers. From the five repeating units of monomers, the structure of the polymer studied is shown. Quantitative structure-property relationship (QSPR) models obtained by using relevant descriptors showed good predictability. Internal validation {R^2, RMSE, and Q^2_{LOO}}, external validation {R^2, RMSE, Q^2_{pred}, r^2_m, Δr^2_m, k, and k'}, and applicability domain were used to validate these models. The comparison of the results shows that the ANN models are more efficient than those of the MLR models. Accordingly, the QSPR model developed in this study provides excellent predictions, and can be used to predict I_{CO}, I_{OH}, I_{OP}, and $\overline{M}_v$ of polymers, particularly for those that have not been tested.

Keywords

QSPR, photostabilization, polymers, artificial neural network, multiple linear regressions

1 Introduction

Synthetic polymers are among the most widely produced materials and are used in various fields, such as construction, electronics, chemical engineering, packaging and transportation, due to their excellent chemical and physical properties.[1–3] Polyvinyl chloride (PVC), polystyrene (PS) and polymethyl methacrylate (PMMA) are some of the most important industrial-scale polymers. The low cost of production, the wide range of use, and the good performance of PVC, PS, and PMMA products have generated interest for these polymers.[4,5]

However, one of the main disadvantages of the use of synthetic or semi-synthetic polymeric materials is their degradation and aging. Thus, PVC, PS, and PMMA undergo photodegradation when exposed to harsh environments, such as high temperatures, sunlight, fungi, bacteria, yeasts, algae and their enzymes.[6] The consequences of this degradation depend on the nature of the polymer and can cause scission of the polymer chain, rapid yellowing, and loss of gloss, crosslinking accompanied by changes in the physical and chemical properties of the polymer.[7,8] Finally, we end up with useless materials after an unpredictable duration.[9]

Research on the stabilization of polymers against harmful environmental effects is extremely important. They tend to reduce or prevent all kinds of damage to polymers. These polymers are generally protected against such deterioration by the addition of antioxidants, stabilizers against light and heat.[10] Among the stabilizing systems developed are free radical scavengers that have proven effective.[11] It should be noted that photostabilization and aging of polymers are a complex problem to study in practice, as they usually take place slowly, and their service life generally reaches several decades.[12–16] Several studies on photostabilization of polymers are reported in the literature.[17–22] The photostabilization activities of polymers compounds were determined by monitoring the carbonyl, polyene and hydroxyl indices, as well as the variations in the viscosity average molecular weight with the duration exposure. Carbonyl, polyene and hydroxyl groups are used to evaluate/measure the amount of polymer degradation during ultraviolet radiation in the presence of oxygen over time. Growth of carbonyl groups indicates extent of polymer degradation. However, the experimental determination of these indices involves costly experimental studies.

The photostability properties of polymers depend on the molecular structures of the polymers during photodegradation. In addition, the use of in silico predictive methods, based on computer tools, offers a fast and cost-effective alternative to experimentation. These methods include the Quantitative Structure-Property Relationship (QSPR) models. This strategy consists of modelling the properties of the material as functions of the molecular structure using QSPR.[23] The objective of QSPR is to develop mathe-

* Corresponding author: Hadjira Maouz, PhD
  Email: maouz.hadjira@yahoo.com

matical equations capable of establishing the relationships between property and descriptors derived solely from the molecular structure of the polymer. Once a correlation is established and validated, it may be applicable to predicting the property of a new polymer or to discovering new materials with the desired properties.[24]

A few QSPR models have been successfully applied to the correlation of many physicochemical properties of various polymers. For example, *Xu et al.*[25] developed QSPR model, which was built to predict refractive indices of linear polymers by applying four molecular descriptors. Compared with the existing QSPR models, the proposed model requires only four descriptors, which can be obtained by simple calculation making it easier to predict the refractive indices of polymers. QSPR study was elaborated by *Xu et al.*[26] The QSPR model was performed between topological indices representing the molecular structures and lower critical solution temperature (LCST) with a database of 169 data points. A satisfactory mean relative error (MRE) was obtained, and the authors concluded that the model would be very useful in obtaining reliable estimates of LCST in polymer solutions. *Liu et al.*[27] developed a QSPR model for structural units of 35 polymethacrylates. The QSPR of the quantum chemical descriptors and properties, such as molar volumes at room temperature, refractive indices, and glass transition temperature, were obtained by stepwise regression and artificial neural network (ANN) method. The results calculated by ANN method meet the experimental data better than those by the stepwise method. A QSPR to predict the intrinsic viscosity of polymer solutions was developed by *Gharagheizi*.[28] With five relevant descriptors and 65 polymer solutions, a radial based function neural network (RBFNN) with squared correlation coefficient $R^2 = 0.9100$ was constructed. Recently, quantitative structure-property relationship for the thermal decomposition of polymers is suggested by *Toropova et al.*[29] The data on architecture of monomers is used to represent polymers. The average statistical quality of the suggested QSPR for prediction of molar thermal decomposition is the following: RMSE = 4.71 ± 0.1 and $R^2 = 0.97 ± 0.01$. More recently, *Duchowicz et al.*[30] have developed a predictive QSPR for the refractive indices of 234 structurally diverse polymers. The established equations were validated and tested through various well-known techniques, such as the use of an external test set of compounds, the cross-validation method, Y Randomization and applicability domain. They concluded that the developed QSPR could be useful in assisting the development of new polymeric materials.

Unfortunately, in view of the bibliographic research, no application of QSPR studies has been devoted to the prediction of the following parameters: carbonyl index, hydroxyl index, polyene index, and viscosity average molecular weight. Therefore, given the importance of these parameters in photostabilization studies of polymers, the objective assigned to this study was to develop two models based on the exploitation of the relationship between the chemical structure of polymers (PVC, PS, and PMMA) and each of the four parameters using a linear approach (Multiple Linear Regression: MLR) and a non-linear approach (Artificial Neural Network: ANN).

2 Materials and methods

The general methodology adopted for this study consisted of several basic steps to generate valid QSPR models. This methodology is indicated in the diagram of Fig. 1.
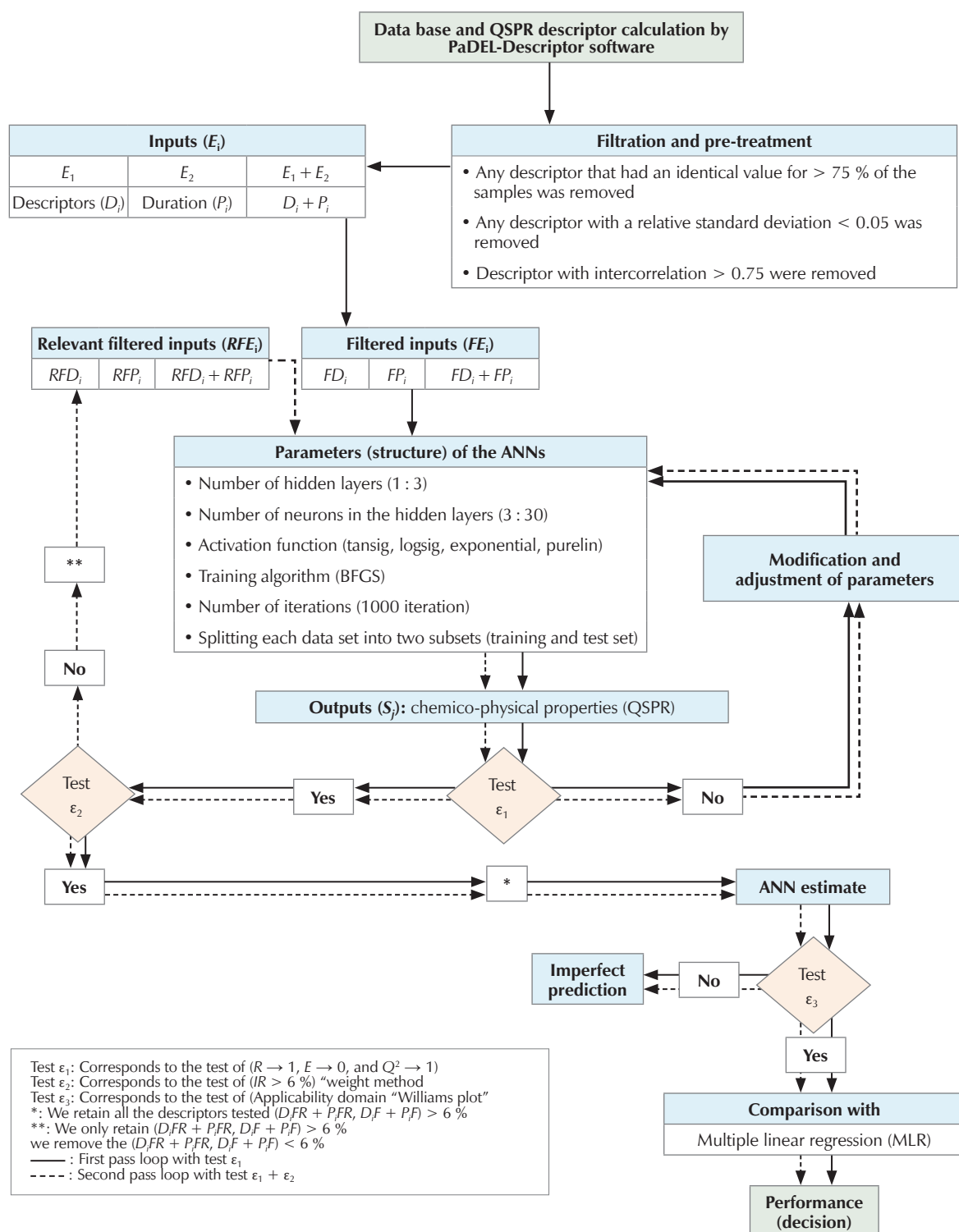
2.1 Datasets

It is well known that high-quality experimental data are essential for the development of high quality QSPR models.[31] The polymers in the database included two families: 3 polymers (PVC, PS, and PMMA), and 20 polymer mixtures (polymers with stabilizers). The experimental values of carbonyl index (111 polymers), hydroxyl index (141 polymers), polyene index (81 polymers), and viscosity average molecular weight (118 polymers) were collected from the literature.[17,32–35] To develop an ANN model, the polymer database was divided into two sets: a training set, and a test set consisting of 77 % and 23 % of the polymers for I_{CO}, 60 % and 40 % for I_{OH}, 80 %, and 20 % for I_{PO} and $\bar{M}_v$, respectively. The collected photostabilization values used in this work were performed under the influence of accelerated testing technique: Accelerated weatherometer Q.U.V. tester (Q. panel, company, USA).

2.2 Molecular descriptors

The direct calculation of a molecular descriptor for the entire structure is not possible due to the high molecular weight of the polymeric compound. To overcome this drawback, a small number of repeating units (U) was used.[36] In addition, the molecular descriptors calculated from their repeating unit structures end-capped with two hydrogen atoms can be used in the QSPR studies for systems polymer.[25,37] In this study, we chose polymer units to properly represent the interaction between polymers and stabilizers around the photodegradation. The chosen polymer units (UUUUU) are based on existing models in the literature.[17,33] The list of polymer units used in the development of QSPR model is given in Table 1.

One important step in obtaining a QSAR (Quantitative Structure-Activity Relationships) and QSPR model is the numerical representation of the structural features of molecules, which were named molecular descriptors. Currently, there are thousands of molecular descriptors in the literature that can be used to solve different problems in different specialties.[38] All descriptors were obtained through the online program PaDEL-Descriptor (http://www.scbdd.com/padel_desc/index/). PaDEL-Descriptor is one of the most applied softwares in QSPR studies but also for QSAR analyses.[39] In the specific case of this study, for each polymer, 1875 molecular descriptors were calculated, belonging to following classes: Autocorrelation descriptors (346), Basak descriptors (42), BCUT descriptors (6), Burden descriptors (96), Connectivity descriptors (56), Constitutional descriptors (120), E-state descriptors (489), Kappa descriptors (3), Molecular property descriptors (15), Quantum chemical descriptors (6), Topological descriptors (265), CPSA descriptors (29), RDF descriptors (210), Geometrical descrip-

H. MAOUZ et al.: QSPR Studies of Carbonyl, Hydroxyl, Polyene Indices, and Viscosity Average..., Kem. Ind. **69** (1-2) (2020) 1–16

3

Fig. 1 – Basic steps for generating a QSPR model in this study

tors (21), WHIM descriptors (91), and 3D Autocorrelation descriptors (80).

Simplified Molecular Input-Line Entry System (SMILES) notations of polymers were obtained from the ChemBio Ultra Software.

2.3 Selection of relevant descriptors

The pre-processing of the database is to eliminate the irrelevant descriptors in order to avoid the phenomenon of over-fitting. Therefore, we must reduce the variables (descriptors) that do not have or have little influence on

Table 1 – List of structures of the unit polymers (UUUUU) used in the development of QSPR models
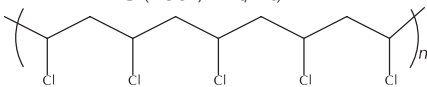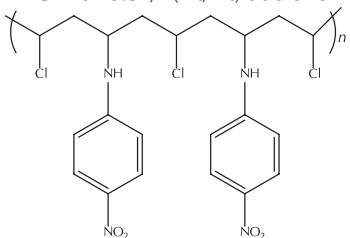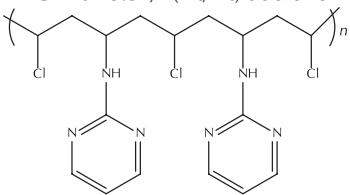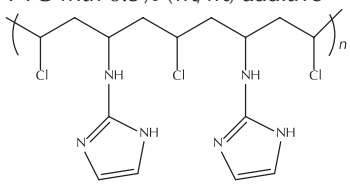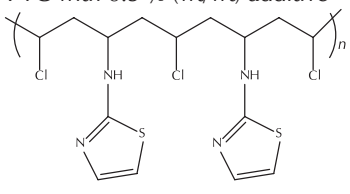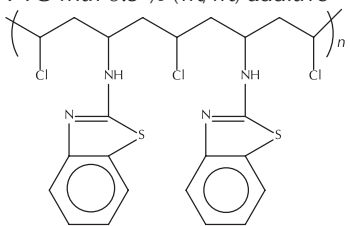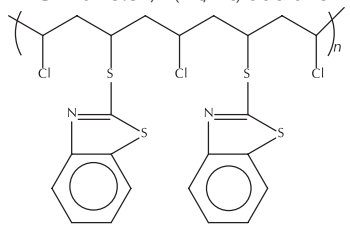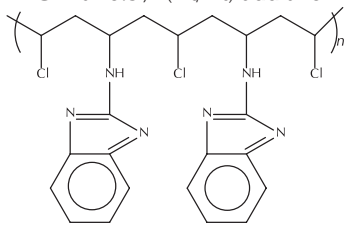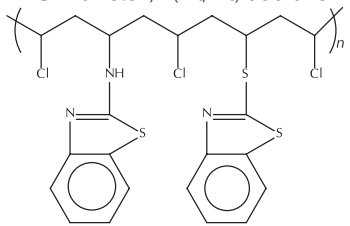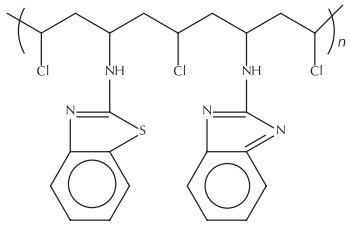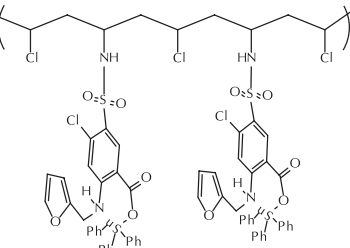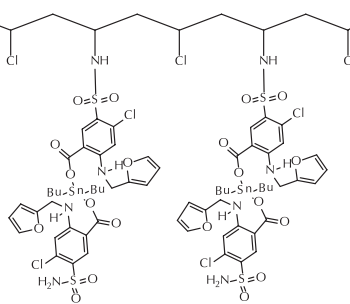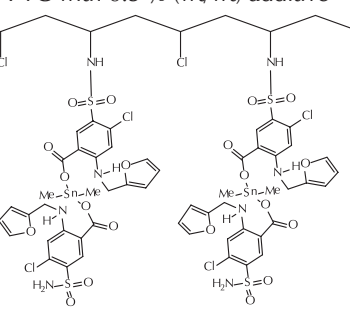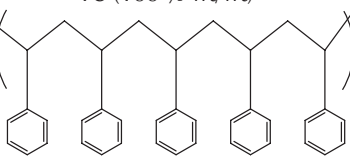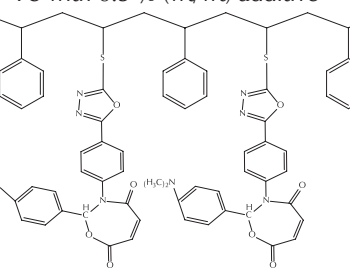
PVC (100 % wt/wt)[14,33,34]	PVC with 0.5 % (wt/wt) additive[33]	PVC with 0.5 % (wt/wt) additive[33]
PVC with 0.5% (wt/wt) additive [33]	PVC with 0.5 % (wt/wt) additive[33]	PVC with 0.5 % (wt/wt) additive[14]
PVC with 0.5 % (wt/wt) additive[14]	PVC with 0.5% (wt/wt) additive [14]	PVC with 0.5 % (wt/wt) additive[14]
PVC with 0.5 % (wt/wt) additive[14]	PVC with 0.5 % (wt/wt) additive[34]	PVC with 0.5% (wt/wt) additive[34]
PVC with 0.5 % (wt/wt) additive[34]	PS (100 % wt/wt)[32]	PS with 0.5 % (wt/wt) additive[32]

H. MAOUZ et al.: QSPR Studies of Carbonyl, Hydroxyl, Polyene Indices, and Viscosity Average..., Kem. Ind. **69** (1-2) (2020) 1–16

5

Table 1 – (continued)
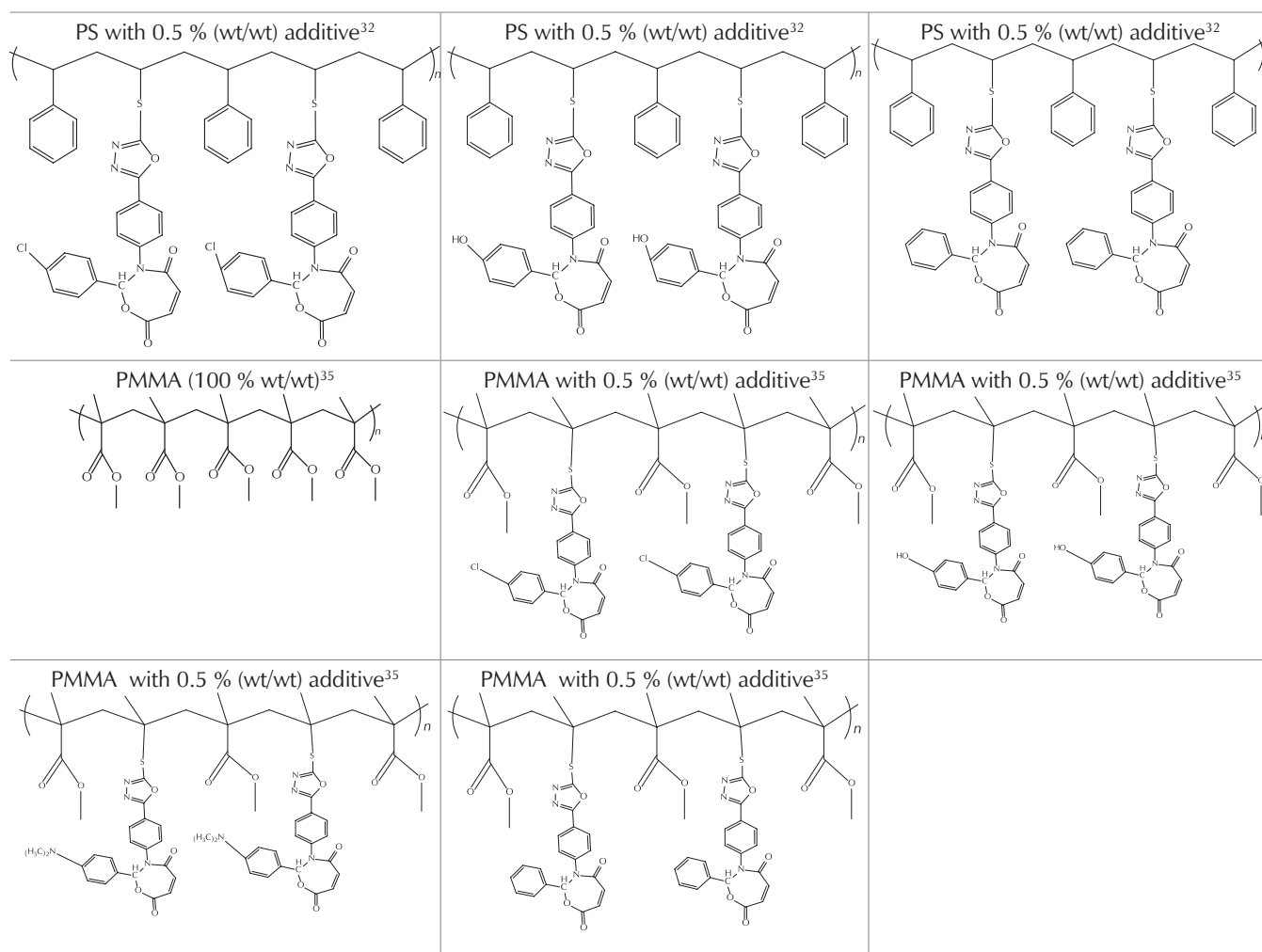
the outputs of network (carbonyl index, hydroxyl index, polyene index, and viscosity average molecular weight). Several methods to simplify the database are available in the literature; for example, principal component analysis (PCA), curvilinear or orthogonalization method of Graam-Schmidt are used. In this present study, the method used to select the most significant descriptors was described by some authors.[40–42] It takes place in several stages: (1) the minimum and maximum are calculated for each descriptor using STATISTICA software, then we remove the descriptors that have the maximum and the minimum equal. (2) The descriptor which has the same value for more than 75 % of the samples is eliminated. (3) Standard deviation (SD) is calculated for each descriptor, and those with SD values less than 0.05 are eliminated. (4) In this stage, we used "Matlab" software; a diagonal matrix is then obtained which represents the correlation between the outputs and the descriptors retained. The descriptors are classified according to the decreasing value of the correlation coefficient. The descriptor with the highest correlation is taken and compared to the other descriptors in the matrix. Those whose correlation coefficient value is greater than 0.75 are eliminated in their turn. The same is repeated with the descriptor ranked just after the first, and so on. The number of descriptors obtained after the selection was 107 for I_{CO}, 102 for I_{OH}, 67 for I_{PO}, and 107 for $\overline{M}_v$. (5) A program based on the stepwise method is used to select the most relevant descriptors from those obtained previously. Finally, the number of descriptors (Table 2) obtained after stepwise selection was: 5 for I_{CO}, 5 for I_{OH}, 3 for I_{PO}, and 4 for $\overline{M}_v$. The relevant descriptors as well as the duration exposure (evaluated in hours) were used to develop the QSPR prediction model.

2.4 Model development

Descriptors obtained after feature pre-screening were used to develop predictive models. Many approaches of model development are widely used. Two different approaches to developing QSPR prediction models were used.

2.4.1 Linear model

The linear model was developed by applying Multiple Linear Regression (MLR). MLR are the most widely used

Table 2 – List of descriptors obtained after stepwise selection and duration of exposure

Parameters predicts	Descriptor	Description	Category	VIF	*t*-Test
Carbonyl index (I_{CO})	MATS2m	Moran autocorrelation – lag 2 / weighted by mass	Autocorrelation descriptors	1.37	−11.96
	minHBd	Minimum E-States for (strong) Hydrogen Bond donors	E-state descriptors	1.20	−5.62
	MATS2s	Moran autocorrelation – lag 2 / weighted by I-state	Autocorrelation descriptors	1.44	−4.72
	VE3_Dzm	Logarithmic coefficient sum of the last eigenvector from Barysz matrix / weighted by mass	Topological descriptors	1.34	5.06
	MATS7i	Moran autocorrelation – lag 7 / weighted by first ionization potential	Autocorrelation descriptors	1.44	3.65
	t(h)	duration of exposure	duration of exposure	1.01	19.97
Hydroxyl index (I_{OH})	TDB5i	3D topological distance based autocorrelation – lag 5 / weighted by first ionization potential	3D Autocorrelation descriptors	1.38	−12.32
	geomShape	Petitjean geometricshape index	Geometrical descriptors	1.41	−9.69
	minHBd	Minimum E-States for (strong) Hydrogen Bond donors	E-state descriptors	1.04	−5.82
	ATSC7m	Centered Broto-Moreau autocorrelation – lag 7 / weighted by mass	Autocorrelation descriptors	1.53	5.18
	P2i	2nd component shape directional WHIM index / weighted by relative first ionization potential	WHIM descriptors	1.28	−3.85
	t(h)	duration of exposure	duration of exposure	1.01	17.9
Polyene index (I_{PO})	SpMin2_Bhv	Smallest absolute eigenvalue of Burden modified matrix – n 2 / weighted by relative van der Waals volumes	Burden descriptors	2.13	−9.33
	VE3_Dzi	Logarithmic coefficient sum of the last eigenvector from Barysz matrix / weighted by first ionization potential	Topological descriptors	1.11	6.70
	VR3_Dzv	Logarithmic Randic-like eigenvector-based index from Barysz matrix / weighted by van der Waals volumes	Topological descriptors	2.08	−7.01
	t(h)	duration of exposure	duration of exposure	1.02	17.03
Viscosity average molecular weight ($\overline{M}_v$)	RDF25p	Radial distribution function – 025 / weighted by relative polarizabilities	RDF descriptors 3D	1.12	21.56
	TDB4i	3D topological distance based autocorrelation – lag 4 / weighted by first ionization potential	Autocorrelation descriptors	1.03	−4.99
	minHBint8	Minimum E-State descriptors of strength for potential Hydrogen Bonds of path length 8	E-state descriptors	1.12	4.71
	ATSC3c	Centred Broto-Moreau autocorrelation – lag 3 / weighted by charges	Autocorrelation descriptors	1.01	2.67
	t(h)	duration of exposure	duration of exposure	1.00	−16.19

and known modelling methods, and used as the basis for a number of multivariate methods.[42] MLR is a commonly used method in QSPR due to its simplicity, transparency, reproducibility, and easy interpretability. MLR consists of a quantitative relationship between a group of variables X_i (descriptors) and a response *Y,* as shown in Eq. (1):

$$Y = a_0 + \sum_{i=1}^{n} a_i X \qquad (1)$$

where *Y* is the response or dependent variable (outputs), X_i represents the molecular descriptors (inputs), and a_0 is a constant (intercept). MLR calculations were performed using STATISTICA v. 8.0 (StatSoft, Inc.) and XLSTAT software.

2.4.2 Nonlinear model

Nonlinear model was then developed by submitting the relevant descriptors to a statistical learning method: the Artificial Neural Network (ANN). ANN is particularly well suited for QSPR/QSAR models because of their capability to take out nonlinear information from the data set.[43] MLP-ANN is considered the easiest and most commonly used ANN type in literature.[42] The architecture of an MLP-ANN consists of an input layer encompassing the inputs, one or more hidden layers (intermediate), and an output layer including the outputs. The layers are connected to each other linearly by the weights corresponding to the neurons in the neighbouring layers upstream and downstream. In

this work, the tangent sigmoid (tansig), the sigmoid log (logsig), and the exponential transfer function were used as a transfer function of the hidden layer, while the exponential function and the linear function (Purelin) were used as a transfer function for the output layer. The number of hidden neurons was optimized (from 3 to 30) by trial and error procedure in the training process. One output neuron was used to represent the experimental values of I_{CO}, I_{OH}, I_{PO}, and $\bar{M}_v$. The network was trained using the quasi-Newton BFGS (Broyden–Fletcher–Goldfarb–Shanno) algorithm.

2.5 Model validation

The relevance of the QSPR/QSAR models is judged on the basis of the results of statistical validation. The statistical validation of models consists of internal and external validations. Recent studies[44–48] have indicated that internal validation is essential for the validation of a QSPR/QSAR model. In our study, the most important traditional validation metrics were applied: root mean square error (RMSE), determination coefficient (R^2), cross validated correlation coefficient (Q^2_{LOO}), in addition to the use of the parameters (r^2_m; Δr^2_m) introduced by *Roy et al.*[49] External validation is essential to judge the predictive power of a model.[50] For this study, the predictive power of the QSPR model was tested on the test set not used for model development, using the Q^2_{pred} parameter and the validation criteria by *Golbraikh and Tropsha*.[51] The statistical parameters are collected in Eqs. (2–7), and the terms utilized in these equations are defined in the following:

$$\text{RMSE} = \sqrt{\frac{\sum \left(Y_{obs} - Y_{pred}\right)^2}{n}} \tag{2}$$

$$R^2 = 1 - \frac{\sum \left(Y_{obs} - Y_{pred}\right)^2}{\sum (Y_{obs} - \bar{Y}_{obs})^2} \tag{3}$$

$$Q^2_{LOO} = 1 - \frac{\sum \left(Y_{obs\ (training)} - Y_{pred\ (training)}\right)^2}{\sum (Y_{obs\ (training)} - \bar{Y}_{obs(training)})^2} \tag{4}$$

$$Q^2_{pred} = R^2_p = 1 - \frac{\sum \left(Y_{obs\ (test)} - Y_{pred\ (test)}\right)^2}{\sum (Y_{obs\ (test)} - \bar{Y}_{obs(training)})^2} \tag{5}$$

$$\bar{r}^2_m = \frac{\left(r^2_m + r'^2_m\right)}{2} \tag{6}$$

$$\Delta r^2_m = \left| r^2_m - r'^2_m \right| \tag{7}$$

where Y_{obs} is observed (experimental) value of Y, Y_{pred} is predicted (calculated) Y-value of training set, test set or validation set, n is the number of compounds in the data set (training, test, validation), $\bar{Y}_{obs}$ is the average of Y_{obs}, and r^2 is the squared correlation coefficient between the observed and predicted value of polymers with intercept.

To see the contribution of each parameter in the explanation of the dependent variable Y in the MLR models, we used the test of significance of each parameter t-Student "T-test" statistic. From this statistic, it is possible to test one by one the nullity of the different parameters of the multiple linear regression models, and build confidence intervals on these parameters, very useful during the interpretation phase of the model.[52–53]

$$\left| t_i \right| = \left| \frac{a_i}{s\left(a_i\right)} \right| > t^{n-p-2}_{1-\frac{\alpha}{2}} \tag{8}$$

where: t_i is the T-test for descriptor "i"; a_i is the coefficient associated with descriptor "i" in the model; s is the standard error associated with descriptor "i"; α is the confidence interval, n is the number of observations (size of database); p is the number of independent variables (descriptors).

2.6 Sensitivity analysis

To see the contribution of each input variable (descriptors with duration exposure) on the outputs (I_{CO}, I_{OH}, I_{PO}, and $\bar{M}_v$), a sensitivity analysis was carried out using the "weight" method. This method, proposed by *Garson*,[54] provides a quantification of the relative importance (RI) of the different inputs (variables) on the outputs of each neural network. The process of calculating relative importance by the "weight" method unrolls as follows: We calculate the product of input–hidden layer and hidden–output layer connection weights between each input neuron and output neuron and sum the products across all hidden neurons.[55]

2.7 Applicability domain

Any model of prediction must have a range of accuracy satisfactory for its application. This range is defined by the applicability domain of the model. Outside this domain, the application of the model can lead to erroneous predictions.[50] It should be noted that there are several approaches to determining the application domains. In this present work, we used the leverage approach (Williams plot). The influence of a sample on the model is measured at the leverage (h_i). The leverage of a compound in the original variable space is defined as follows:[50]

$$H = X\left(X^T X\right)^{-1} X^T \tag{9}$$

where X is the model matrix derived from the training set descriptor values, and the leverage values of training set are diagonal elements of the hat or influence matrix ($h_i = \text{diag}\ (H)$). The leverage values are always between 0 and 1. The warning leverage h^* is, generally, fixed at $3(p + 1)/n$, where n is the total number of samples in the training set, and p is the number of descriptors involved in the correlation.[50]

3 Results and discussion

3.1 MLR predictive model (linear model)

The linear models obtained for the prediction of I_{CO}, I_{OH}, I_{PO}, and $\bar{M}_v$ of different polymers are represented by the following Eqs. (10–13) with the reported statistical parameters:

$I_{CO} = 0.084 + 5.075\text{E} - 04\ t\ \text{(h)} - 0.473\ \text{MATS2m} - $
$\quad - 0.118\ \text{minHBd} - 9.849\text{E} - 02\ \text{MATS2s} + $
$\quad + 2.904\text{E}\ 04\ \text{VE3\_Dzm} + 0.181\ \text{MATS7i}$ (10)
$\quad (N = 111;\ R^2 = 0.862;\ \text{RMSE} = 0.024;$
$\quad F = 106.905;\ p < 0.0001)$

$I_{OH} = 0.585 + 1.933\text{E} - 04\ t\ \text{(h)} - 5.824\text{E} - 04\ \text{TDB5i} - $
$\quad - 9.632\text{E} - 02\ \text{geomShape} - 3.987\text{E}\ 02\ \text{minHBd} + $
$\quad + 1.850\text{E} - 06\ \text{ATSC7m} - 8.135\text{E} - 02\ \text{P2i}$ (11)
$\quad (N = 141;\ R^2 = 0.802;\ \text{RMSE} = 0.011;$
$\quad F = 89.907;\ p < 0.0001)$

$I_{PO} = 1.319 + 6.382\text{E} - 04\ t\ \text{(h)} - 0.546\ \text{SpMin2_Bhv} + $
$\quad + 5.742\text{E} - 04\ \text{VE3\_Dzi} - 9.128\text{E} - 04\ \text{VR3\_Dzv}$ (12)
$\quad (N = 81;\ R^2 = 0.898;\ \text{RMSE} = 0.030;$
$\quad F = 165.866;\ p < 0.0001)$

$\bar{M}_v = 547869.690 - 324.233\ t\ \text{(h)} + 1594.286\ \text{RDF25p} - $
$\quad - 804.937\ \text{TDB4i} + 5277.705\ \text{minHBint8} + $
$\quad + 1293.069\ \text{ATSC3c}$ (13)
$\quad (N = 118;\ R^2 = 0.891;\ \text{RMSE} = 0.0650;$
$\quad F = 181.371;\ p < 0.0001)$

The large F ratio (106.905; 89.907; 165.866; 181.371) indicates that Eqns. (10), (11), (12), and (13) were powerful in predicting I_{CO}, I_{OH}, I_{PO}, and $\bar{M}_v$.

For each pair of descriptors, the value of the correlation coefficient (see Table S1 in the supplementary materials) of the four parameters studied was < 0.717, which leads to the independence of the selected descriptors. In addition, the multi-collinearity of the descriptors used in the MLR model can be evaluated by calculating their Inflation Variation Factors (VIF). If the calculation of VIF gives a value between 1 and 5, the associated model is acceptable.[54] As Table 2 shows, all variables have VIF values < 2.13, which shows that the resulting model has statistical significance, and therefore, the descriptors represent some orthogonality.

As shown in Table 2, the t-Test value of the exposure time (t) of the carbonyl number (I_{CO}) is 19.97 h, the latter value is greater than the other descriptors, and therefore, it corresponds to a significant duration of exposure. Negative regression coefficients for the MATS2m, minHBd, and MATS2s descriptors have a negative impact on I_{CO}. On the contrary, a positive influence for the descriptors VE3_Dzm and MATS7i, due to the positive signs of the regression coefficients, will favour the improvement of the carbonyl index. We maintain the same reasoning for the other parameters (I_{OH}, I_{PO}, and $\bar{M}_v$), where some descriptors have positive signs (good impact), while others have negative signs (impact obsolete).

3.2 ANN predictive model (nonlinear model)

The optimization of the artificial neural network architecture is essential to obtain an optimal network. In addition, the database distribution, the activation functions (for hidden neurons and output neurons), the number of neurons in the hidden layer, and the learning algorithms were optimized after several trials. The optimal performance of the model is evaluated in terms of RMSE.[38] The results of optimization of the nonlinear model are shown in Table S2 (supplementary materials). Thus, the networks architectures obtained are the following: {6-18-1} for I_{CO}, {6-23-1} for I_{OH}, {4-13-1} for I_{PO}, and {5-24-1} for $\bar{M}_v$. The predicted values of I_{CO}, I_{OH}, I_{PO}, and $\bar{M}_v$ in the training and test sets have been plotted *versus* their observed values in Fig. 2. As may be seen, a close correlation between the predicted and the observed values was found. The main performance parameters of ANN model are shown in Table 3. As may be seen from Table 3, all values of the statistical parameters (R^2, Q^2_{LOO}, and RMSE) of the training set are acceptable. The non-linear ANN model also gives good results for the test set. Therefore, these results reveal that the ANN model not only performed well in model development, but also had excellent prediction. Moreover, all these results confirm the existence of a non-linear relationship between the relevant descriptors of the model and the predicted physicochemical properties.

3.2.1 Sensitivity analysis

The weight method was used to calculate the relative importance IR (%) of variables (descriptors + duration exposure) for ANN model. Fig. 3 gives a graphica representation of the relative importance IR (%) of each variable on the properties of photostabilization (I_{CO}, I_{OH}, I_{OP}, and $\bar{M}_v$). According to this figure, all input relevant variables have a significant contribution (IR > 6 %) to the photostabilization properties. This sensitivity analysis by the weight method successfully identified the true importance of all the variables used for the modelling of physicochemical properties during the photostabilization of PVC, PS, and PMMA, and therefore, proves the correctness of the choice of variables that were used in this study.

3.2.2 Applicability domain

The applicability domain of the ANN model was analysed using a Williams plot (Fig. 4). The vertical line is the critical leverage value (h^*). As seen in Fig. 4, none of the polymers of the training set and the validation set had a leverage effect higher than the warning value h^* (0,2035 for I_{CO}, 0,2059 for I_{OH}, 0,1923 for I_{PO}, and 0,1579 for $\bar{M}_v$). For polyene index and in the training set, one polymer was underestimated. In the Williams plot of the carbonyl index and the viscosity average molecular weight, two polymers belonging to the test set and one polymer for the training set can be considered as response outliers. One of which is overestimated, while another is underestimated. In the domain applicability of the hydroxyl index, there polymers
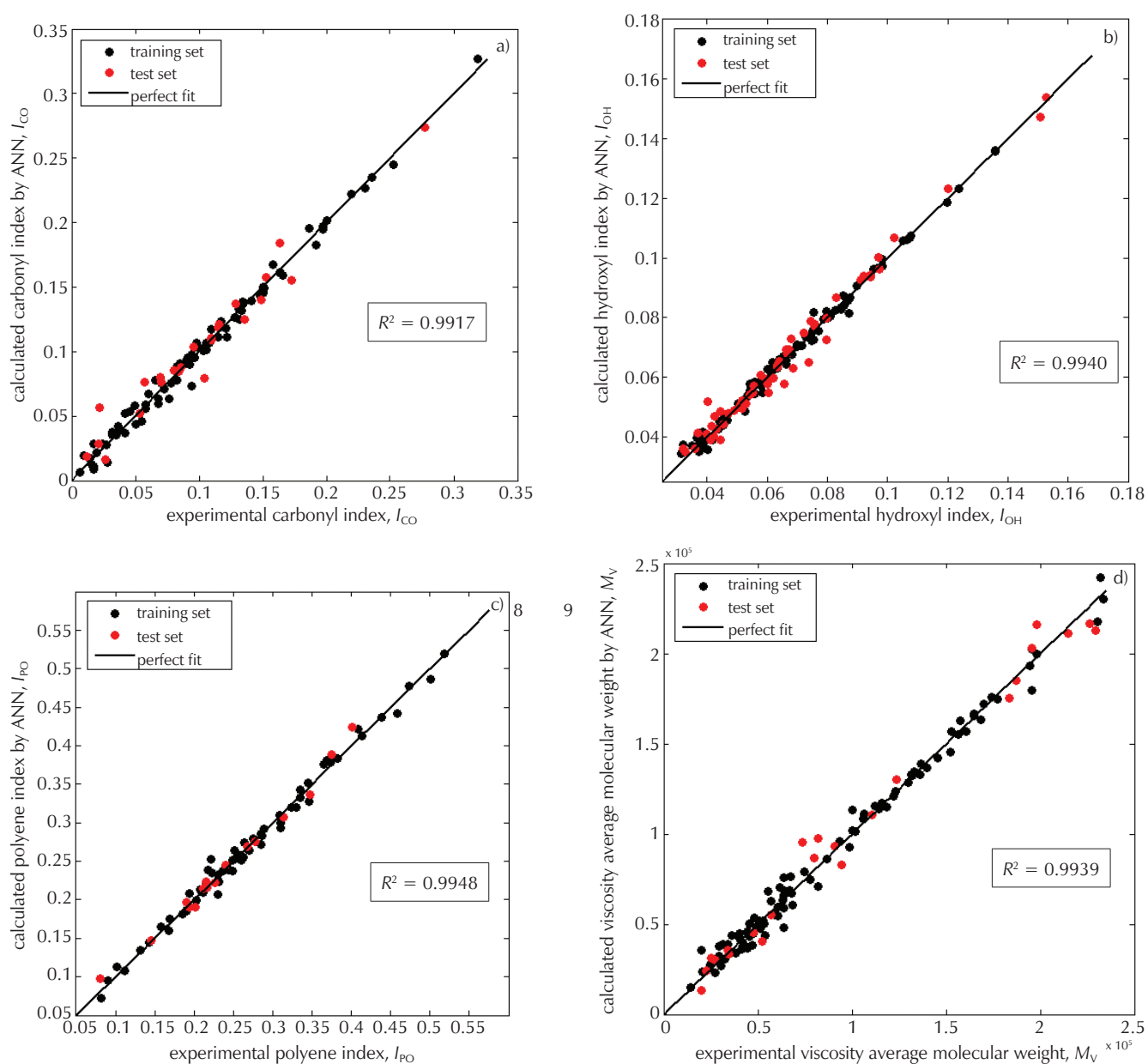
Fig. 2 – Plot of observed *vs.* predicted: (a) I_{CO}, (b) I_{OH}, (c) I_{PO}, and (d) $\bar{M}_v$ values from the ANN model

Table 3 – Performance of the ANN model

Models		Internal validation				External validation							
		n	R^2	RMSE	Q^2_{LOO}	n	R^2	RMSE	Q^2_{pred}	$\overline{r^2_m}$	Δr^2_m	k	k'
ANN	I_{CO}	86	0.99	0.006	0.99	25	0.96	0.012	0.95	0.89	0.04	0.96	1.00
	I_{OH}	85	0.99	0.002	0.99	56	0.98	0.004	0.98	0.97	0.01	0.99	1.00
	I_{PO}	65	0.99	0.009	0.99	16	0.99	0.010	0.99	0.98	0.01	0.99	1.01
	$\bar{M}_v$	95	0.98	0.057	0.99	23	0.98	0.093	0.98	0.97	0.01	0.99	1.00
Threshold value[50,56]			> 0.6		> 0.5		> 0.6		> 0.5	> 0.5	< 0.2	0.85 < k < 1.15	0.85 < k' < 1.15

Fig. 3 – Plot of the relative importance of the descriptors for ANN models: carbonyl index, hydroxyl index, polyene index, and viscosity average molecular weight

Fig. 4 – Projection of the training set and the test set in the Williams plot for ANN: (a) I_{CO}, (b) I_{OH}, (c) I_{PO}, and (d) $\bar{M}_v$

Table 4 – External validation of the ANN and MLR models

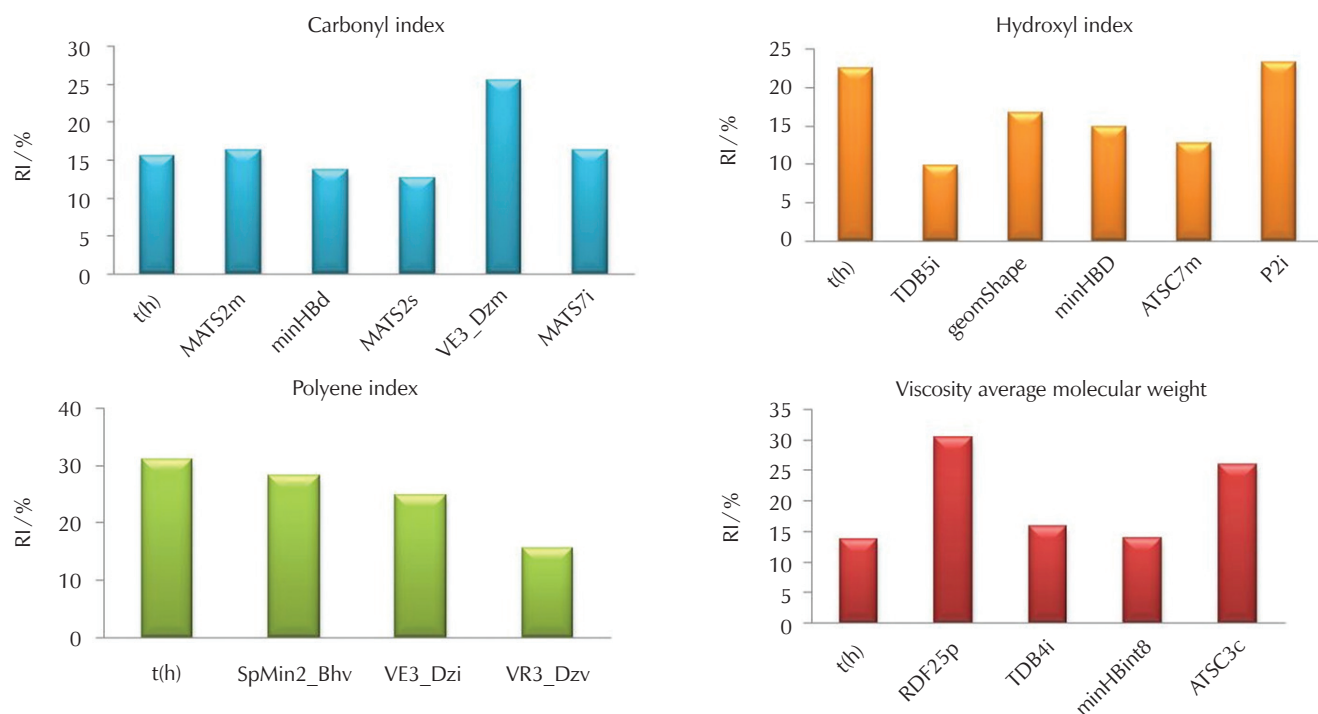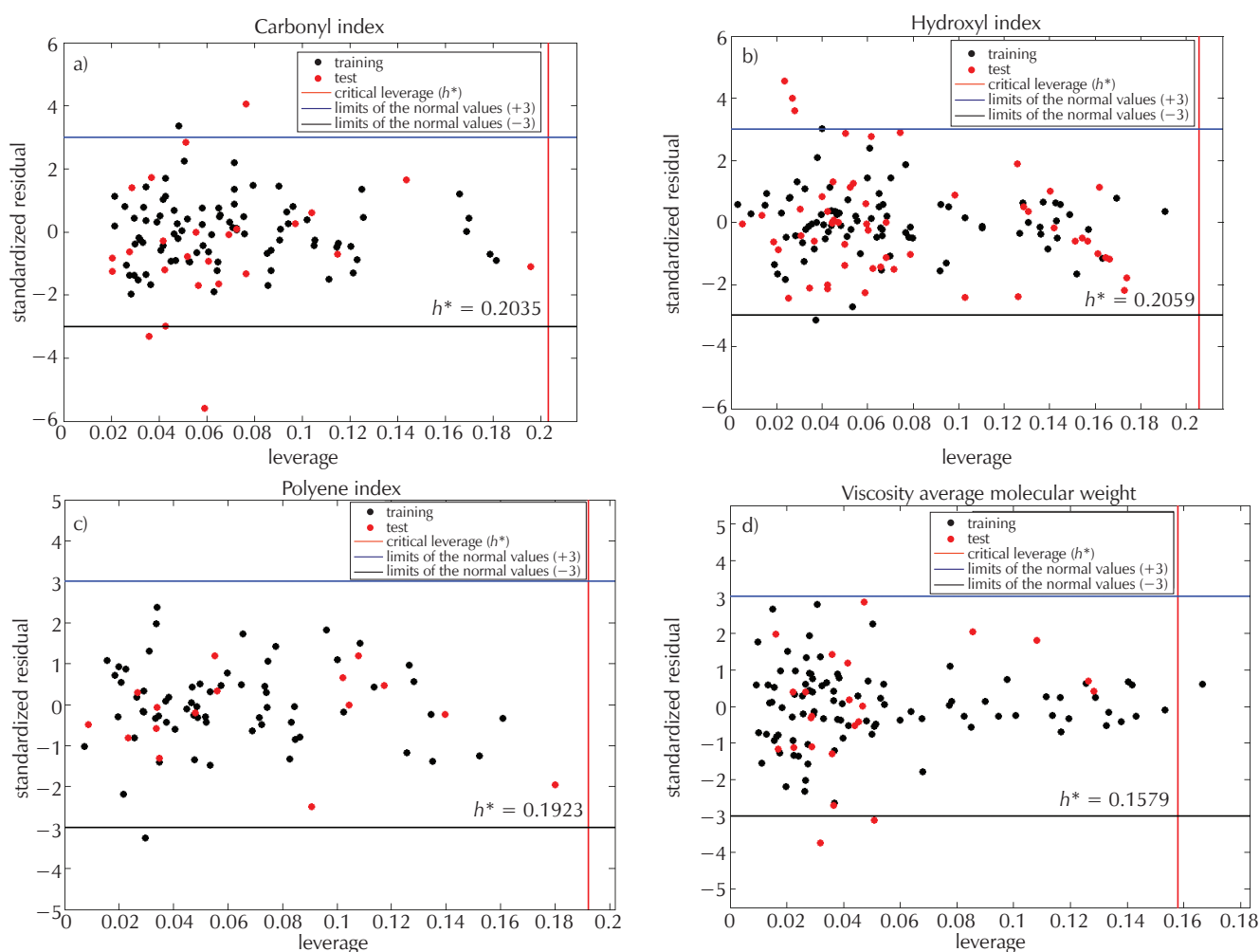|	Models	n	R^2	RMSE	Q^2_{pred}	$\overline{r^2_m}$	Δr^2_m	k	k'
I_{CO}	ANN	25	0.96	0.012	0.95	0.89	0.04	0.96	1.00
	MLR		0.80	0.025	0.80	0.69	0.16	1.00	0.95
I_{OH}	ANN	56	0.98	0.004	0.98	0.97	0.01	0.99	1.00
	MLR		0.76	0.013	0.76	0.65	0.17	1.00	0.96
I_{PO}	ANN	16	0.99	0.010	0.99	0.98	0.01	0.99	1.01
	MLR		0.88	0.031	0.87	0.81	0.09	0.97	1.02
$\overline{M_v}$	ANN	23	0.98	0.093	0.98	0.97	0.01	0.99	1.00
	MLR		0.84	0.283	0.85	0.77	0.10	1.05	0.91

of the test set is underestimated. These eleven response outliers (3 for I_{CO}, 4 for I_{OH}, 1 for I_{PO}, and 3 for $\overline{M_v}$) could be associated with errors in the experimental values. The results of the ANN models correspond to the third principle of the Organization for Economic Cooperation and Development (OECD).

of the statistical parameters for the ANN model can be noted. In conclusion, the ANN model developed in this study meets all OECD principles for QSPR validation, and can be used to predict I_{CO}, I_{OH}, I_{OP}, and $\overline{M_v}$ of polymers, particularly those that have not been tested, and thus help reduce experimental determination of these indices, which involves costly experimental studies.

3.3 Comparison of the result of MLP-ANN with MLR models

To compare the performance and the predictive quality of the ANN and MLR models, the statistical parameters are summarized in Table 4. According to Table 4, all the values of the statistical parameters seem satisfactory for the two models, which show a good robustness. However, a substantial improvement of the statistical parameters for the ANN model can be noted. Thus, it can be concluded that the ANN model has better predictive power than the MLR model. This means that the model obtained with an ANN suggested the existence of a non-linear correlation between the outputs of the network (I_{CO}, I_{OH}, I_{PO}, and $\overline{M_v}$) and the selected variables (descriptors + duration exposure).

4 Conclusion

The aim of the present work was to develop a QSPR study and to predict the carbonyl, hydroxyl, and polyene indices (I_{CO}, I_{OH}, I_{OP}), and viscosity average ($\overline{M_v}$) of poly (vinyl chloride), polystyrene and poly (methyl methacrylate). These physicochemical properties were considered a fundamental property during the study of the photodegradation of the polymers. This QSPR study, which involved 111, 141, 81, and 118 structurally diverse polymers, and a series of descriptors calculated by PaDEL-Descriptor software selected by a stepwise method, was based on the artificial neural network (ANN) and multiple linear regression (MLR). The built ANN and MLR models were assessed comprehensively (by internal and external validation). They showed good values of R^2 and Q^2_{LOO} for the training set, and high predictive R^2 and Q^2_{pred} values for the validation set. All the validations indicated that the built QSPR models were robust and satisfactory. However, a substantial improvement

List of abbreviations
Popis kratica

ANN – artificial neural network

$\overline{M_v}$ – viscosity average molecular weight

D_i – descriptors

E_i – inputs

FD_i – filtrated descriptors

FE_i – filtrated inputs

P_i – physical properties

Q^2 – cross-validated correlation coefficient

BFGS – Broyden-Fletcher-Goldfarb-Shanno

I_{CO} – carbonyl index

I_{OH} – hydroxyl index

I_{OP} – polyene index

LCST – lower critical solution temperature

MLR – multiple linear regressions

MRE – mean relative error

OECD – organization for economic cooperation and development

PMMA – polymethyl methacrylate

PS – polystyrene

PVC   – polyvinyl chloride

QSAR   – quantitative structure–activity relationships

QSPR   – quantitative structure–property relationship

RBFNN – radial based function neural network

RI   – relative importance

RMSE   – root mean squared error
             – korijen srednje kvadratne pogreške

SD   – standard deviation

SMILES – simplified molecular input-line entry system

VIF   – variation inflation factor

E – error

FP_i – filtrated physical properties

IR   – relative importance

R   – correlation coefficient

RFD_i – relevant filtrated descriptors

RFE_i – relevant filtrated input

RFP_i – relevant filtrated physical properties

References
Literatura

1. *E. Yousif, N. Salih, J. Salimon*, Improvement of the Photostabilization of PVC Films in the Presence of 2N-Salicylidene-5-(Substituted)-1,3, 4-Thiadiazole, J. Appl. Polym. Sci. **120** (2011) 2207–2214, doi: https://doi.org/10.1002/app.33463.

2. *M. Ito, K. Nagai*, Degradation issues of polymer materials used in railway field, Polym. Degrad. Stabil. **93** (2008) 1723–1735, doi: https://doi.org/10.1016/j.polymdegradstab.2008.07.011.

3. *C. C. Wang, G. Pilania, S. A. Boggs, S. Kumar, C. Breneman, R. Ramprasad*, Computational strategies for polymer dielectrics design, Polymer **55** (2014) 979–988, doi: https://doi.org/10.1016/j.polymer.2013.12.069.

4. *E. Yousif, A. A. Al-Amiery, A. Kadihum, A. A. H. Kadhum, A. B. Mohamad*, Photostabilizing Efficiency of PVC in the Presence of Schiff Bases as Photostabilizers, Molecules **20** (2015) 19886–19899, doi: https://doi.org/10.3390/molecules201119665.

5. *E. A. Yousif, S. A. Aliwi, A. A. Ameer, J. R. Ukal*, Induced photodegradation effect on the functionalized FE (III) complex additive-poly(vinyl chloride) thin film, Turk J Chem. **33** (2009) 399–410, doi:10. https://doi.org/10.3906/kim-0711-4.

6. *G. Q. Ali, G. A. El-Hiti, I. H. R. Tomi, R. Haddad, A. J. Al-Qaisi, E. Yousif*, Photostability and Performance of Polystyrene Films Containing 1,2,4-Triazole-3-thiol Ring System Schiff Bases, Molecules **21** (2016) 1699, doi: https://doi.org/10.3390/molecules21121699.

7. *R. R. Mohamed*, "Photostabilization of Polymers." Polymers and Polymeric Composites: A Reference Series, 2014, pp. 1–10, doi: https://doi.org/10.1007/978-3-642-37179-0_74-1.

8. *E. Yousif*, Triorganotin (IV) complexes photo-stabilizers for rigid PVC against photodegradation, J. Taibah. Uni. Sci. **7** (2013) 79–87, doi: https://doi.org/10.1016/j.jtusci.2013.04.007.

9. *E. Yousif, R. Haddad*, Photodegradation and photostabilization of polymers, especially polystyrene: review, Springerplus **2** (2013) 398, doi: https://doi.org/10.1186/2193-1801-2-398.

10. *E. Yousif, J. Salimon, N. Salih*, Mechanism of photostabilization of poly (methy methacrylate) films by 2-thioacetic acid benzothiazol complexes, Arab. J. Chem. **7** (2014) 306–311, doi: https://doi.org/10.1016/j.arabjc.2010.11.003.

11. *E. Yousif, J. Salimon, N. Salih*, New Stabilizers for Polystyrene Based on 2-Thioacetic Acid Benzothiazol Complexes, J. Appl. Polym. Sci. **125** (2012) 1922–1927, doi: https://doi.org/10.1002/app.36307.

12. *X. Colin, G. Teyssedre, M. Fois*, Ageing and degradation of multiphase polymer systems, Handbook of Multiphase Polymer Systems, 2011, pp. 797–841, doi: https://doi.org/10.1002/9781119972020.ch21.

13. *N. S. Allen, M. Edge, T. Corrales, F. Catalina*, Stabiliser interactions in the thermal and photooxidation of titanium dioxide pigmented polypropylene films, Polym. Degrad. Stabil. **61** (1998) 139–149, doi: https://doi.org/10.1016/S0141-3910(97)00143-2.

14. *J. M. Peña, N. S. Allen, M. Edge, C. M. Liauw, I. Roberts, B. Valange*, Triplet quenching and antioxidant effect of several carbon black grades in the photodegradation of LDPE doped with benzophenone as a photosensitiser, Polym. Degrad. Stabil. **70** (2000) 437–454, doi: https://doi.org/10.1016/S0141-3910(00)00140-3.

15. *J. M. Peña, N. S. Allen, M. Edge, C. M. Liauw, B. Valange*, Studies of synergism between carbon black and stabilisers in LDPE photodegradation, Polym. Degrad. Stabil. **72** (2001) 259–270, doi: https://doi.org/10.1016/S0141-3910(01)00033-7.

16. *N. S. Allen, M. Edge, A. Ortega, G. Sandoval, C. M. Liauw, J. Verran, J. Stratton, R. B. McIntyre*, Degradation and stabilisation of polymers and coatings: nano versus pigmentary titania particles, Polym. Degrad. Stabil. **85** (2004) 927–946, doi: https://doi.org/10.1016/j.polymdegradstab.2003.09.024.

17. *E. Yousif, A. Hameed, R. Rasheed, H. Mansoor, Y. Farina, A. Graisa, N. Salih, J. Salimon*, Synthesis and Photostability Study of Some Modified Poly(vinyl chloride) Containing Pendant Benzothiazole and Benzimidozole Ring, Int. J. Chem. **2** (2010) 65–80.

18. *I. H. R. Tomi, G. Q. Ali, A. H. Jawad, E. Yousif*, Photostabilizing efficiency of PVC based on epoxidized oleic acid, J. Polym. Res. **24** (2017) 119, doi: https://doi.org/10.1007/s10965-017-1283-7.

19. *E. Yousif, J. Salimon, N. Salih*, New photostabilizers for PVC based on some diorganotin (IV) complexes, J. Saudi. Chem. Soc. **19** (2015) 133–141, doi: https://doi.org/10.1016/j.jscs.2012.01.003.

20. *E. Yousif, J. Salimon, N. Salih*, New stabilizers for polystyrene based on 2-N-salicylidene-5-(substituted)-1,3,4-thiadiazole compounds, J. Saudi. Chem. Soc. **16** (2012) 299–306, doi: https://doi.org/10.1016/j.jscs.2011.01.011.

21. *E. Yousif, E. Bakir, J. Salimon, N. J. Salih*, Evaluation of Schiff bases of 2,5-dimercapto-1,3,4- thiadiazole as photostabilizer for poly(methyl methacrylate), J. Saudi. Chem. Soc. **16** (2012) 279–285, doi: https://doi.org/10.1016/j.jscs.2011.01.009.

22. *E. Yousif, J. Salimon, N. Salih, A. Ahmed*, Improvement of the photostabilization of PMMA films in the presence 2N-salicylidene-5-(substituted)-1,3,4-thiadiazole, J. King. Saud. University (Science). **24** (2012) 131–137, doi: https://doi.org/10.1016/j.jksus.2010.09.001.

23. *A. Afantitis, G. Melagraki, H. Sarimveis, P. A. Koutentis, J. Markopoulos, O. Igglessi-Markopoulou*, Prediction of intrinsic viscosity in polymer–solvent combinations using a QSPR

model, Polymer **47** (2006) 3240–3248, doi: https://doi.org/10.1016/j.polymer.2006.02.060.

24. *M. Hamadache, A. Amrane, S. Hanini, O. Benkortbi*, Multilayer Perceptron Model for Predicting Acute Toxicity of Fungicides on Rats, International Journal of Quantitative Structure-Property Relationships (IJQSPR) **3** (2018) 100–118, doi: https://doi.org/10.4018/IJQSPR.2018010106.

25. *J. Xu, B. Chen, Q. Zhang, B. Guo*, Prediction of refractive indices of linear polymers by a four-descriptor QSPR model, Polymer **45** (2004) 8651–8659, doi: https://doi.org/10.1016/j.polymer.2004.10.057.

26. *J. Xu, L. Liu, W. Xu, S. Zhao, D. Zuo*, A general QSPR model for the prediction of θ (lower critical solution temperature) in polymer solutions with topological indices, J. Mol. Graph. Model. **26** (2007) 352–359, doi: https://doi.org/10.1016/j.jmgm.2007.01.004.

27. *W. Liu, P. Yi, Z. Tang*, QSPR Models for Various Properties of Polymethacrylates Based on Quantum Chemical Descriptors, QSAR. Comb. Sci. **25** (2006) 936–943, doi: https://doi.org/10.1002/qsar.200510177.

28. *F. Gharagheizi*, QSPR analysis for intrinsic viscosity of polymer solutions by means of GA-MLR and RBFNN, Comput. Mater. Sci. **40** (2007) 159–167, doi: https://doi.org/10.1016/j.commatsci.2006.11.010.

29. *A. P. Toropova, A. A. Toropov, K. O. Valentin, L. Leszczynska, L. Jerzy*, Optimal descriptors as a tool to predict the thermal decomposition of polymer, J. Math. Chem. **52** (2014) 1171–1181, doi: https://doi.org/10.1007/s10910-014-0323-3.

30. *P. R. Duchowicz, S. E. Fioressi, D. E. Bacelo, L. M. Saavedra, A. P. Toropova, A. A. Toropov*, QSPR studies on refractive indices of structurally heterogeneous polymers, Chemometr. Intell. Lab. Syst. **140** (2015) 86–91, doi: https://doi.org/10.1016/j.chemolab.2014.11.008.

31. *M. T. D. Cronin, T. W. Schultz*, Pitfalls in QSAR, J. Mol. Struct. **622** (2003) 39–51, doi: https://doi.org/10.1016/S0166-1280(02)00616-4.

32. *E. Yousif, A. Hameed, N. Salih, J. Salimon, B. M. Abdullah*, New photostabilizers for polystyrene based on 2, 3-dihydro-(5-mercapto-1, 3, 4-oxadiazol-2-yl)-phenyl-2-(substituted)-1, 3, 4-oxazepine-4,7-dione compounds, SpringerPlus, **2** (2013) 104, doi: https://doi.org/10.1186/2193-1801-2-104.

33. *E. Yousif, A. Ahmed, R. Abood, N. Jaber, R. Noaman, R. Yusop*, Poly(vinyl chloride) derivatives as stabilizers against photodegradation, J. Taibah. Uni. Sci. **9** (2015) 203–212, doi: https://doi.org/10.1016/j.jtusci.2014.10.003.

34. *M. M. Ali, G. A. El-Hiti, E. Yousif*, Photostabilizing Efficiency of Poly(vinyl chloride) in the Presence of Organotin (IV) Complexes as Photostabilizers, Molecules **21** (2016) 1151, doi: https://doi.org/10.3390/molecules21091151.

35. *A. Hameed*, Microwave Synthesis of Some New 1, 3-Oxazepine Compounds as Photostabilizing Additives for Pmma Films, JAUS **15** (2012) 47–59.

36. *P. R. Duchowicz, S. E. Fioressi, D. E. Bacelo, L. M. Saavedra, A. P. Toropova, A. A. Toropov*, QSPR studies on refractive indices of structurally heterogeneous polymers, Chemometr. Intell. Lab. Syst. **140** (2015) 86–91, doi: https://doi.org/10.1016/j.chemolab.2014.11.008.

37. *A. R. Katritzky, S. Sild, V. Lobanov, M. Karelson*, Quantitative Structure-Property Relationship (QSPR) Correlation of Glass Transition Temperatures of High Molecular Weight Polymers, J. Chem. Inf. Comput. Sci. **38** (1998) 300–304, doi: https://doi.org/10.1021/ci9700687.

38. *M. Hamadache, O. Benkortbi, S. Hanini, A. Amrane, L. Khaouane, C. Si-Moussa*, A Quantitative Structure Activi-

ty Relationship for acute oral toxicity of pesticides on rats: Validation, domain of application and prediction, J. Hazard. Mater. **303** (2016) 28–40, doi: https://doi.org/10.1016/j.jhazmat.2015.09.021.

39. *J. C. Dearden*, The Use of Topological Indices in QSAR and QSPR Modeling. In: Roy K. (Eds) Advances in QSAR Modeling. Challenges and Advances in Computational Chemistry and Physics, vol 24, Springer, Cham (2017) pp. 57–88, doi: https://doi.org/10.1007/978-3-319-56850-8_2.

40. *D. A. Konovalov, N. Sim, E. Deconinck, Y. Vander Heyden, D. Coomans*, Statistical confidence for variable selection in QSAR models, J. Chem. Inf. Model. **48** (2008) 370–383, doi: https://doi.org/10.1021/ci700283s.

41. *A. S. Reddy, S. Kumar, R. Garg*, Hybridgenetic algorithm based descriptor optimization and QSAR models for predicting the biological activity of tipranavir analogs for HIV protease inhibition, J. Mol. Graph. Model. **28** (2010) 852–862, doi: https://doi.org/10.1016/j.jmgm.2010.03.005.

42. *A. Habibi-Yangjeh, M. Danandeh-Jenagharad*, Application of a genetic algorithm and an artificial neural network for global prediction of the toxicity of phenols to Tetrahymena pyriformis, Monatsh. Chem. **140** (2009) 1279–1288, doi: https://doi.org/10.1007/s00706-009-0185-8.

43. *M. Hamadache, L. Khaouane, O. Benkortbi, C. Si Moussa, S. Hanini, A. Amrane*, Prediction of Acute Herbicide Toxicity in Rats from Quantitative Structure–Activity Relationship Modeling, Environ. Eng. Sci. **31** (2014) 243–252, doi: https://doi.org/10.1089/ees.2013.0466.

44. *M. Hamadache, S. Hanini, O. Benkortbi, A. Amrane, L. Khaouane, C. Si-Moussa*, Artificial neural network-based equation to predict the toxicity of herbicides on rats, Chemometr. Intell. Lab. Syst. **154** (2016) 7–15, doi: https://doi.org/10.1016/j.chemolab.2016.03.007.

45. *R. Wang, J. Jiang, Y. Pan, H. Cao, Y. Cui*, Prediction of impact sensitivity of nitro energetic compounds by neural network based on electro topological-state indices, J. Hazard. Mater. **166** (2009)155–186, doi: https://doi.org/10.1016/j.jhazmat.2008.11.005.

46. *I. Mitra, A. Saha, K. Roy*, Chemometric QSAR Modeling and In Silico Design of Antioxidant NO Donor Phenols, Sci. Pharm. **79** (2011) 31–57, doi: https://doi.org/10.3797/scipharm.1011-02.

47. *V. Consonni, D. Ballabio, R. Todeschini*, Evaluation of model predictive ability by external validation techniques, J. Chemometr. **24** (2010) 194–201, doi: https://doi.org/10.1002/cem.1290.

48. *S. Bitam, M. Hamadache, S. Hanini*, Prediction of therapeutic potency of tacrine derivatives as BuChE inhibitors from quantitative structure–activity relationship modeling, SAR. QSAR. Environ. Res. **29** (2018) 213–230, doi: https://doi.org/10.1080/1062936X.2018.1423640.

49. *K. Roy, I. Mitra, S. Kar, P. K. Ojha, R. N. Das, H. Kabir*, Comparative studies on some metrics for external validation of QSPR models, J. Chem. Inf. Model. **52** (2012) 396–408, doi: https://doi.org/10.1021/ci200520g.

50. *M. Hamadache, O. Benkortbi, S. Hanini, A. Amrane*, QSAR modeling in ecotoxicological risk assessment: application to the prediction of acute contact toxicity of pesticides on bees (*Apis mellifera* L.), Environ. Sci. Pollut. Res. Int. **25** (2018) 896–907, doi: https://doi.org/10.1007/s11356-017-0498-9.

51. *A. Golbraikh, A. Tropsha*, Beware of q2!, *J. Mol. Graph. Model*. **20** (2002) 269–276, doi: https://doi.org/10.1016/S1093-3263(01)00123-1.

52. *K. Bellifa*, Etude des relations quantitatives structure–toxicité des composés chimiques à l'aide des descripteurs molécu-

laires. Modélisation QSAR (Doctoral dissertation) (2015).

53. *S. Chtita,* Modélisation de molécules organiques hétérocycliques biologiquement actives par des méthodes QSAR/QSPR. Recherche de nouveaux médicaments (Doctoral dissertation) (2017).

54. *G. D. Garson,* Interpreting neural network connection weights, Artif. Intell. Expert. **6** (1991) 46–51.

55. *M. Gevrey, I. Dimopoulos, S. Lek,* Review and comparison of methods to study the contribution of variables in artificial neural network models, Ecol. Model. **160** (2003) 249–264, doi: https://doi.org/10.1016/S0304-3800(02)00257-0.

56. *S. Bitam, M. Hamadache, S. Hanini,* QSAR model for prediction of the therapeutic potency of *N*-benzylpiperidine derivatives as AChE inhibitors, SAR. QSAR. Environ. Res. **28** (2017) 471–489, doi: https://doi.org/10.1080/1062936X.2017.1331467.

SAŽETAK

QSPR studije karbonilnih, hidroksilnih, polienskih indeksa i prosječne molekulske težine polimera pod fotostabilizacijom pristupom ANN i MLR

Hadjira Maouz,[a*] *Latifa Khaouane,*[a] *Salah Hanini,*[a] *Yamina Ammi,*[a,b] *Mabrouk Hamadache,*[a] and *Maamar Laidi*[a]

Jedan od glavnih nedostataka upotrebe sintetičkih ili polusintetičkih polimernih materijala je njihova razgradnja i starenje. Svrha ove studije je primjena umjetnih neuronskih mreža (ANN) i višestrukih linearnih regresija (MLR) za predviđanje karbonilnih, hidroksilnih i polienskih indeksa (I_{CO}, I_{OH} i I_{OP}) i prosječne molekulske mase viskoznosti ($\overline{M}_v$) poli(vinil-klorida), polistirena i poli(metil metakrilata). Ta fizikalno-kemijska svojstva smatraju se važnim tijekom proučavanja fotostabilizacije polimera. Iz pet ponavljajućih jedinica monomera prikazana je struktura ispitivanog polimera. Kvantitativni modeli odnosa strukture-svojstava (QSPR) dobiveni primjenom relevantnih deskriptora pokazali su dobru predvidljivost. Za potvrdu tih modela provedene su: interna provjera {$\overline{R}^2$, RMSE i Q^2_{LOO}}, vanjska provjera {R^2, RMSE, Q^2_{pred}, $\overline{r}^2_m$, Δr^2_m, k i k'} i domena primjenjivosti. Usporedba rezultata pokazuje da su modeli ANN učinkovitiji od modela MLR. Prema tome, model QSPR razvijen u ovoj studiji pruža izvrsna predviđanja i može se primjenjivati za predviđanje I_{CO}, I_{OH}, I_{OP} i $\overline{M}_v$ polimera, posebno za one koji nisu testirani.

Ključne riječi
QSPR, fotostabilizacija, polimeri, umjetna neuronska mreža, višestruke linearne regresije

[a] *Laboratory of Biomaterials and Transport Phenomena (LBMPT), University of Médéa, Quartier Aïn d'Heb, 26000, Algeria*

[b] *University Center, Faculty of Science and Technology, Department of Process Engineering, Relizane, Algeria*

Supplementary material

Table S1 (a) – Correlation matrix of the descriptors used for the carbonyl index

	t(h)	MATS2m	minHBd	MATS2s	VE3_Dzm	MATS7i
t(h)	1	−0.056	−0.015	0.053	0.041	−0.038
MATS2m	−0.056	1	0.014	−0.447	0.176	−0.200
minHBd	−0.015	0.014	1	0.254	−0.029	0.297
MATS2s	0.053	−0.447	0.254	1	0.072	0.187
VE3_Dzm	0.041	0.176	−0.029	0.072	1	−0.441
MATS7i	−0.038	−0.200	0.297	0.187	−0.441	1

Table S1 (b) – Correlation matrix of the descriptors used for the hydroxyl index

	t(h)	TDB5i	geomShape	minHBd	ATSC7m	P2i
t(h)	1	0.062	0.034	−0.016	0.080	0.015
TDB5i	0.062	1	−0.198	0.093	0.396	0.063
geomShape	0.034	−0.198	1	−0.091	0.205	−0.333
minHBd	−0.016	0.093	−0.091	1	−0.108	0.039
ATSC7m	0.080	0.396	0.205	−0.108	1	0.213
P2i	0.015	0.063	−0.333	0.039	0.213	1

Table S1 (c) – Correlation matrix of the descriptors used for the polyene index

	t(h)	SpMin2_Bhv	VE3_Dzi	VR3_Dzv
t(h)	1	0.097	0.061	0.109
SpMin2_Bhv	0.097	1	0.306	0.717
VE3_Dzi	0.061	0.306	1	0.263
VR3_Dzv	0.109	0.717	0.263	1

Table S1 (d) – Correlation matrix of the descriptors used for the viscosity average molecular weight

	t(h)	RDF25p	TDB4i	minHBint8	ATSC3c
t(h)	1	−0.007	0.009	−0.005	0.000
RDF25p	−0.007	1	0.069	0.313	−0.016
TDB4i	0.009	0.069	1	−0.063	−0.119
minHBint8	−0.005	0.313	−0.063	1	0.004
ATSC3c	0.000	−0.016	−0.119	0.004	1

Table S2 – Selected parameters of the optimal ANN model

ANN models	I_{CO}	I_{OH}	I_{PO}	$\bar{M}_v$
Number of input layers	1	1	1	1
Number of hidden layers	1	1	1	1
Number of output layers	1	1	1	1
Number of input neurons	6	6	4	5
Number of hidden neurons	18	23	13	24
Number of output neurons	1	1	1	1
Transfer function of the hidden neurons	Exponential	Tanh	Tanh	Tanh
Transfer function of the output neurons	Identity	Exponential	Identity	Identity
Training algorithm	BFGS	BFGS	BFGS	BFGS
Training set	77 % ($n = 86$)	60 % ($n = 85$)	80 % ($n = 65$)	80 % ($n = 95$)
Test set	23 % ($n = 25$)	40 % ($n = 56$)	20 % ($n = 16$)	20 % ($n = 23$)
RMSE	0.0079	0.0027	0.0095	0.0650