# Prediction of Refractive Index of Polymers Using Artificial Neural Networks

**G. Astray,*,† A. Cid,† O. Moldes,† J. A. Ferreiro-Lage,† J. F. Gálvez,‡ and J. C. Mejuto†**

Department of Physical Chemistry, Faculty of Sciences, University of Vigo, 32004, Ourense, Spain, and Department of Informatics, ESEI, University of Vigo, 32004, Ourense, Spain

This paper was withdrawn on January 11, 2011 (DOI: 10.1021/je200071p).

Density functional theory (DFT) calculations were carried out in the prediction of the refractive index ($n$) of different polymers at the B3LYP/6-31G(d) level. A set of quantum chemical descriptors calculated from monomers of polymers, the energy of the lowest unoccupied molecular orbital ($E_{LUMO}$), molecular average polarizability ($\alpha$), heat capacity at constant volume ($C_V$), and the most positive net atomic charge on hydrogen atoms in a molecule (e) were used to build a general quantitative structure−property relationship (QSPR) model for the refractive index. The proposed model gives the mean error of prediction of 1.048 % for the validation set.

## Introduction

The refractive index is a basic optical property of polymers, which is related to other optical, electrical, and magnetic properties.[1] Therefore, the refractive index is widely used in material science. For example, the specific refractive index increment ($d_n/d_c$) can be used for the determination of molecular weight, size, and shape.[2] A successful quantitative structure−property relationship (QSPR) is valuable due to its quantitative prediction of the refractive index values of as yet unsynthesized polymers. The refractive index can be estimated with the group additive property (GAP) theory.[3] The GAP method provides a rapid and computationally inexpensive approach to the estimation of the refractive index values, but it is purely an empirical approach and limited to systems composed only of functional groups that have been previously investigated. Furthermore, the group contributions method is only approximate since this approach fails to account for the presence of neighboring groups or conformational influences. Some of the shortcomings and limitations of group contribution methods can be avoided by using the theoretical QSPR models. Bicerano[2] developed a 10-descriptor model ($R = 0.977$) for a set of 183 polymers. These descriptors included topological and constitutional descriptors. Other authors[4] correlated the refractive indices of polymers with their graph theoretical indices and also got a 10-descriptor correlation with $R$ of 0.981. There are too many descriptors involved in these two models, since the improvement of results by increasing the number of descriptors in the correlation equation should be considered with care, since over-fitting and chance correlations may in part be due to such an approach. Katritzky et al.[5] used the CODESSA program to develop a correlation coefficient of $R = 0.970$ for a set of the refractive indices of 95 linear polymers with five descriptors. Xu et al.[6] introduced a four-descriptor QSPR model with $R = 0.964$. The 121 experimental refractive index data were taken from the set of 183 polymers. However, only García-Domenech's model has been evaluated with the external validation set.[4] In fact, validation is a crucial aspect of any QSPR/QSAR (quantitative structure−activity relationship) modeling.[7] The quantum-chemical descriptors encode information about the electronic structure of the molecule and thus implicitly account for the cooperative effects between functional groups, charge redistribution, and possible hydrogen bonding in the polymer. Furthermore, the quantum-chemical descriptors have a clearly physical meaning. The aim of this study was to produce an artificial neural network (ANN) that could predict the refractive index of prop-2-enoate polymers.

ANNs are a complete statistical tool for data analysis[8] which try to reproduce artificially the human ability of taking decisions simulating the human brain's basic unit, the neuron, and the interconnections between the neurons that allow them to work together and save information from the experience.[9−11] It is a flexible structure, capable of making a nonlinear mapping between input and output spaces.[12] ANNs were originally an abstract simulation of the biological brain systems, composed by an assembly of units called "neurons" (or "nodes") connected between them. The advantage of ANNs consists of their ability to learn from real cases and relearn when new data are input into the system. They are particularly useful in managing different aspects. In recent years, ANNs have been extended successfully to very different fields, from hydrology to finances, happening through chemistry.[13−22]

## Materials and Methods

***Data Set.*** The refractive index data for 95 polymers (Table 1), measured at room temperature (298 K), were collected from the *Prediction of Polymer Properties*.[2] The 95 polymers are divided among the 73 polymers used to train the ANN and 22 used to verify their proper operation.

***Computational Methods: Quantum Properties Determination.*** Density functional theory (DFT)[23,24] has been used to optimize the geometry of the parent monomers and calculate their electronic and thermal properties. All of the calculations have been performed with the Gaussian 03 suite of programs[25] using Becke's three-parameter exchange functional,[26] the nonlocal correlation functional of Lee, Yang, and Parr,[27] (B3LYP) and Pople's 6-31G(d,p) basis set, which includes a polarization function for every atom.

All minima on the potential energy surface were characterized by harmonic analysis, and the computed frequencies were used

* Corresponding author. E-mail: gastray@uvigo.es.
† Department of Physical Chemistry.
‡ Department of Informatics.

10.1021/je100885f 
Published on Web 10/05/2010

**Table 1. *n* Values for Each Polymer**

| polymer | *n* | polymer | *n* |
|---|---|---|---|
| | Training Data | | |
| poly(heptafluorobutyl prop-2-enoate) | 1.3670 | poly(prop-2-enoate alcohol) | 1.5000 |
| poly(pentafluoropropyl prop-2-enoate) | 1.3850 | poly(1,2-butadiene) | 1.5000 |
| poly(trifluoroethyl prop-2-enoate) | 1.4070 | polyisobutylene | 1.5050 |
| poly(2,2,2-trifluoro-1-methylethylmeth prop-2-enoate) | 1.4185 | poly(cyclohexylmeth prop-2-enoate) | 1.5060 |
| poly(trifluoroethylmeth prop-2-enoate) | 1.4370 | poly(2-hydroxyethylmeth prop-2-enoate) | 1.5119 |
| poly(prop-2-enoate isobutylether) | 1.4507 | poly(1-butene) | 1.5125 |
| poly(prop-2-enoate ethylether) | 1.4540 | poly(prop-2-enoate chloroacetate) | 1.5130 |
| poly(prop-2-enoate *n*-butylether) | 1.4563 | poly(*N*-methylmethacrylamide) | 1.5135 |
| poly(prop-2-enoate *n*-pentylether) | 1.4590 | poly(2-chloroethylmeth prop-2-enoate) | 1.5170 |
| poly(4-fluororo-2-trifluoromethylstyrene) | 1.4600 | polyacrylonitrile | 1.5200 |
| poly(prop-2-enoate *n*-octylether) | 1.4613 | polymethacrylonitrile | 1.5200 |
| poly(prop-2-enoate 2-ethylhexylether) | 1.4626 | poly(prop-2-enoic acid) | 1.5270 |
| poly(prop-2-enoate *n*-decylether) | 1.4628 | poly(1,3-dichloropropylmeth prop-2-enoate) | 1.5270 |
| poly(t-butylmeth prop-2-enoate) | 1.4638 | poly(*N*-prop-2-enoate pyrrolidone) | 1.5300 |
| poly(prop-2-enoate *n*-dodecylether) | 1.4640 | poly(prop-2-enoate chloride) | 1.5390 |
| poly(4-methyl-1-pentene) | 1.4650 | poly(2-bromoethylmeth prop-2-enoate) | 1.5426 |
| poly(n-butyl prop-2-enoate) | 1.4660 | poly(1-phenylethylmeth prop-2-enoate) | 1.5487 |
| poly(prop-2-enoate propionate) | 1.4665 | poly(*p*-isopropylstyrene) | 1.5540 |
| poly(prop-2-enoate methylether) | 1.4670 | poly(benzylmeth prop-2-enoate) | 1.5679 |
| poly(prop-2-enoate acetate) | 1.4670 | poly(phenylmeth prop-2-enoate) | 1.5706 |
| poly(ethyl prop-2-enoate) | 1.4685 | poly(2,3-dibromopropylmeth prop-2-enoate) | 1.5739 |
| poly(1-octadecene) | 1.4710 | poly(prop-2-enoate benzoate) | 1.5775 |
| poly(isopropylmeth prop-2-enoate) | 1.4728 | polystyrene | 1.5849 |
| polypropylene | 1.4735 | poly(*o*-methylstyrene) | 1.5874 |
| poly(prop-2-enoate *sec*-butylether) | 1.4740 | poly(*o*-methoxystyrene) | 1.5932 |
| poly(prop-2-enoate formate) | 1.4757 | poly(*p*-bromophenylmeth prop-2-enoate) | 1.5964 |
| polyethylene | 1.4760 | poly(*N*-benzylmethacrylamide) | 1.5965 |
| poly(2-fluoroethylmeth prop-2-enoate) | 1.4768 | poly(*p*-methoxystyrene) | 1.5967 |
| poly(isobutylmeth prop-2-enoate) | 1.4770 | poly(pentachlorophenylmeth prop-2-enoate) | 1.6080 |
| poly(methyl prop-2-enoate) | 1.4790 | poly(*o*-chlorostyrene) | 1.6098 |
| poly(*n*-hexylmeth prop-2-enoate) | 1.4813 | poly(*N*-prop-2-enoate phthalimide) | 1.6200 |
| poly(*n*-butylmeth prop-2-enoate) | 1.4830 | poly(2,6-dichlorostyrene) | 1.6248 |
| poly(*n*-propylmeth prop-2-enoate) | 1.4840 | poly(2-prop-2-enoate thiophene) | 1.6376 |
| poly(ethylmeth prop-2-enoate) | 1.4850 | poly(α-prop-2-enoate naphthalene) | 1.6818 |
| poly(methylmeth prop-2-enoate) | 1.4893 | poly(*N*-prop-2-enoate carbazole) | 1.6830 |
| poly(prop-2-enoate hexylether) | 1.4951 | poly(pentabromophenylmeth prop-2-enoate) | 1.7100 |
| poly(prop-2-enoate methylketone) | 1.5000 | | |
| | Validation Data | | |
| poly(2-nitro-2-methylpropylmeth prop-2-enoate) | 1.4846 | poly[2-phenylsulfonyl)ethylmeth prop-2-enoate] | 1.5682 |
| poly(3,3,5-trimethylcyclohexylmeth prop-2-enoate) | 1.4850 | poly(*m*-crsylmeth prop-2-enoate) | 1.5683 |
| poly(3-methylcyclohexylmeth prop-2-enoate) | 1.4947 | poly(*o*-crsylmeth prop-2-enoate) | 1.5707 |
| poly(4-methylcyclohexylmeth prop-2-enoate) | 1.4975 | poly(1,2-diphenylethylmeth prop-2-enoate) | 1.5816 |
| poly(2-methylcyclohexylmeth prop-2-enoate) | 1.5028 | poly(*o*-chlorobenzylmeth prop-2-enoate) | 1.5823 |
| poly(1-methylcyclohexylmeth prop-2-enoate) | 1.5111 | poly(*m*-nitrobenzylmeth prop-2-enoate) | 1.5845 |
| poly(*N*-butylmethacrylamide) | 1.5135 | poly[*N*-(2-phenylethyl)methacrylamide] | 1.5857 |
| poly(2-chlorocyclohexylmeth prop-2-enoate) | 1.5179 | poly(diphenylmethylmeth prop-2-enoate) | 1.5933 |
| poly(ethylmercaptylmeth prop-2-enoate) | 1.5470 | poly(β-naphthylmeth prop-2-enoate) | 1.6298 |
| poly(*p*-cyclohexylphenylmeth prop-2-enoate) | 1.5575 | poly(α-naphthylmeth prop-2-enoate) | 1.6300 |
| poly[1-(*o*-chlorophenylethyl)meth prop-2-enoate] | 1.5624 | poly(α-naphthylcarbinyl) | 1.6300 |

to obtain zero-point energies and thermodynamic parameters, applying the free particle, harmonic oscillator, and rigid rotor approximations at the high temperature limit in a canonical ensemble ($T = 298.15$ K, $P = 0.101325$ MPa). Frequency values were uncorrected, based on the Scott and Radom estimation of correction coefficients very close to unity for calculations using B3LYP/6-31G(d).[28]

In Table 2, we show the values of the average polarizability (α), the most positive net atomic charge on hydrogen atoms in a molecule (*e*), the energy of the lowest unoccupied molecular orbital ($E_{LUMO}$), and the heat capacity at constant volume ($C_V$).

***Computational Methods: Artificial Neural Network.*** For the implementation of ANNs a commercial software package was used, provided by Neural Planner Software Ltd. All of the component parts are implemented as C++ reusable classes to simplify future development. We will use a perceptron neural network (Figure 1) which could be described as follows: each neuron from the primary layer collects the data of the input variables and presents them according to an input vector (eq 1), which spreads toward the intermediate layer by means of the following propagation rule (eq 2)

$$x^p = (x_1^p, x_2^p, ..., x_N^p)^T \quad (1)$$

$$s_i^p = \sum_{j=1}^{N} w_{ji} x_j^p + b_i \quad (2)$$

where $N$ is the number of the network input neurons, $w_{ji}$ is the weight value of the connection between the neuron $j$ from the input layer and the neuron $i$ from the intermediate layer, and $b_i$ is the value of the "bias" associated to the neuron $i$. If it is assumed that the activation state of the neuron $i$ is the function of the network input vector, then the network output derives from eq 3:

**Table 2. Quantum Chemical Descriptors for Each Polymer Obtained by DFT Calculations**

| polymer | α / $C^2 \cdot m^2 \cdot J^{-1} \cdot 10^{-39}$ | $e$ / $C \cdot 10^{-20}$ | $E_{LUMO}$ / $kJ \cdot mol^{-1}$ | $C_V$ / $kJ \cdot mol^{-1}$ |
|---|---|---|---|---|
| | Training Data | | | |
| poly(1-octadecene) | 3.2367 | 2.2811 | −1.3480 | 356.2925 |
| poly(prop-2-enoate *n*-butylether) | 1.0493 | 2.2815 | 1.1043 | 114.0602 |
| polyethylene | 0.3281 | 2.2876 | 0.8201 | 33.9097 |
| poly(1,2-butadiene) | 0.7275 | 2.3189 | −0.9809 | 65.8668 |
| poly(4-methyl-1-pentene) | 1.0540 | 2.3339 | 1.0494 | 116.6245 |
| poly(prop-2-enoate *n*-dodecylether) | 2.6134 | 2.3855 | 1.8115 | 282.3809 |
| poly(prop-2-enoate *n*-pentylether) | 1.3280 | 2.3861 | 1.8028 | 141.0522 |
| poly(prop-2-enoate hexylether) | 1.5107 | 2.3874 | 1.8054 | 161.2522 |
| poly(prop-2-enoate *n*-octylether) | 1.8775 | 2.3883 | 1.7988 | 209.9567 |
| poly(prop-2-enoate *n*-decylether) | 2.2451 | 2.3884 | 1.8001 | 250.3274 |
| poly(*p*-isopropylstyrene) | 1.9562 | 2.4073 | −1.2146 | 177.0401 |
| poly(1-butene) | 0.6997 | 2.4321 | 1.1279 | 73.9451 |
| polyisobutylene | 0.7003 | 2.4433 | −1.3585 | 76.5220 |
| poly(α-prop-2-enoate naphthalene) | 2.0491 | 2.4481 | −1.9048 | 157.4057 |
| polypropylene | 0.5180 | 2.4944 | 1.2334 | 54.1180 |
| poly(prop-2-enoate 2-ethylhexylether) | 1.8561 | 2.5018 | 1.7539 | 212.6844 |
| poly(*N*-prop-2-enoate carbazole)) | 2.5485 | 2.5130 | −1.2286 | 192.5389 |
| poly(*o*-methoxystyrene) | 1.5865 | 2.5936 | −1.2730 | 147.6766 |
| poly(2,6-dichlorostyrene) | 1.6730 | 2.6117 | −1.7997 | 141.2784 |
| poly(prop-2-enoate ethylether) | 0.7826 | 2.6302 | 1.7631 | 80.3349 |
| poly(prop-2-enoate formate) | 0.5963 | 2.6369 | −0.7420 | 70.5722 |
| poly(prop-2-enoate isobutylether) | 1.1186 | 2.6509 | 1.9022 | 130.7280 |
| poly(*o*-methylstyrene) | 1.5296 | 2.6982 | −1.1776 | 132.4166 |
| poly(prop-2-enoate methylether) | 0.5959 | 2.7028 | 1.6772 | 68.2342 |
| poly(prop-2-enoate *sec*-butylether) | 1.1316 | 2.7081 | 1.8241 | 131.7713 |
| poly(*o*-chlorostyrene) | 1.5163 | 2.7268 | −1.7082 | 124.7070 |
| poly(*p*-methoxystyrene) | 1.6655 | 2.7317 | −0.9347 | 147.4587 |
| poly(4-fluororo-2-trifluoromethylstyrene) | 1.5317 | 2.7528 | −1.9989 | 174.0819 |
| poly(*n*-hexylmeth prop-2-enoate) | 1.8643 | 2.7635 | −1.5586 | 217.0336 |
| poly(*n*-butylmeth prop-2-enoate) | 1.4984 | 2.7644 | −1.5713 | 176.6923 |
| poly(*n*-propylmeth prop-2-enoate) | 1.3166 | 2.7658 | −1.5826 | 156.5677 |
| poly(cyclohexylmeth prop-2-enoate) | 1.7664 | 2.7658 | −1.5817 | 193.1548 |
| poly(ethylmeth prop-2-enoate) | 1.1355 | 2.7662 | −1.5983 | 136.2420 |
| poly(methylmeth prop-2-enoate) | 0.9529 | 2.7710 | −1.6650 | 115.6440 |
| poly(2-prop-2-enoate thiophene) | 1.2736 | 2.7780 | −1.6571 | 100.6606 |
| poly(isobutylmeth prop-2-enoate) | 1.4874 | 2.7788 | −1.6048 | 179.2231 |
| poly(1-phenylethylmeth prop-2-enoate) | 2.0734 | 2.7876 | −1.6489 | 215.1188 |
| poly(ethyl prop-2-enoate) | 0.9546 | 2.7947 | −1.9039 | 113.0672 |
| poly(*n*-butyl prop-2-enoate) | 1.3152 | 2.7950 | −1.8764 | 153.5090 |
| poly(isopropylmeth prop-2-enoate) | 1.3097 | 2.7951 | −1.5734 | 159.4295 |
| poly(2-fluoroethylmeth prop-2-enoate) | 1.1384 | 2.8002 | −1.9837 | 143.6123 |
| poly(methyl prop-2-enoate) | 0.7736 | 2.8181 | −1.9719 | 92.5068 |
| poly(*p*-bromophenylmeth prop-2-enoate) | 2.0967 | 2.8209 | −2.3726 | 190.1548 |
| polymethacrylonitrile | 0.7282 | 2.9082 | −1.9667 | 77.2175 |
| poly(phenylmeth prop-2-enoate) | 1.7166 | 2.9224 | −2.1520 | 173.2523 |
| poly(prop-2-enoate methylketone) | 0.7136 | 2.9225 | −2.4811 | 81.1016 |
| poly(pentabromophenylmeth prop-2-enoate) | 3.2982 | 2.9253 | −3.0762 | 258.5691 |
| poly(pentachlorophenylmeth prop-2-enoate) | 2.7234 | 2.9295 | −2.6564 | 251.9405 |
| poly(prop-2-enoate propionate) | 0.9691 | 2.9583 | −0.4412 | 114.5253 |
| poly(prop-2-enoate benzoate) | 1.6301 | 2.9599 | −2.3656 | 148.9042 |
| poly(*t*-butylmeth prop-2-enoate) | 1.4784 | 2.9975 | −1.5181 | 184.3977 |
| poly(*N*-prop-2-enoate phthalimide) | 1.8802 | 3.0020 | −3.7816 | 164.4198 |
| poly(*N*-prop-2-enoate pyrrolidone) | 1.1484 | 3.0284 | −0.0432 | 116.5951 |
| poly(benzylmeth prop-2-enoate) | 1.7696 | 3.0315 | −2.1363 | 164.2229 |
| polyacrylonitrile | 0.5480 | 3.0362 | −2.4563 | 54.4700 |
| poly(trifluoroethylmeth prop-2-enoate) | 1.1401 | 3.0398 | −2.2566 | 163.8039 |
| poly(pentafluoropropyl prop-2-enoate) | 1.1583 | 3.0404 | −2.5587 | 183.8111 |
| poly(prop-2-enoate chloride) | 0.4864 | 3.0511 | −0.0567 | 44.7827 |
| poly(heptafluorobutyl prop-2-enoate) | 1.3222 | 3.0767 | −2.5696 | 226.8173 |
| poly(prop-2-enoate acetate) | 0.8021 | 3.2022 | −0.5833 | 94.8197 |
| poly(2,2,2-trifluoro-1-methylethylmeth prop-2-enoate) | 1.3120 | 3.2095 | −2.1402 | 187.2302 |
| poly(2-bromoethylmeth prop-2-enoate) | 1.4147 | 3.2169 | −2.0260 | 149.2813 |
| poly(2-chloroethylmeth prop-2-enoate) | 1.2965 | 3.2890 | −2.0556 | 147.3246 |
| poly(trifluoroethyl prop-2-enoate) | 0.9715 | 3.3325 | −2.5321 | 139.9963 |
| poly(2,3-dibromopropylmeth prop-2-enoate) | 1.8453 | 3.3831 | −2.0356 | 185.2399 |
| poly(1,3-dichloropropylmeth prop-2-enoate) | 1.6433 | 3.5221 | −2.6525 | 181.9172 |
| poly(prop-2-enoate chloroacetate) | 0.9072 | 3.7379 | −1.9737 | 104.9930 |
| poly(*N*-methylmethacrylamide) | 1.1080 | 5.4263 | −1.9715 | 133.0325 |
| poly(*N*-benzylmethacrylamide) | 1.9909 | 5.4549 | −1.8198 | 208.7542 |
| poly(2-hydroxyethylmeth prop-2-enoate) | 1.1928 | 6.4044 | −1.7090 | 151.7325 |
| poly(prop-2-enoate alcohol) | 0.3966 | 6.5500 | 1.9074 | 49.7018 |
| poly(prop-2-enoic acid) | 0.5921 | 6.5811 | −2.2056 | 71.0373 |
| polystyrene | 1.3373 | 2.3131 | −1.3306 | 109.0531 |

**Table 2. Continued**

| polymer | α / $C^2 \cdot m^2 \cdot J^{-1} \cdot 10^{-39}$ | $e$ / $C \cdot 10^{-20}$ | $E_{LUMO}$ / $kJ \cdot mol^{-1}$ | $C_V$ / $kJ \cdot mol^{-1}$ |
|---|---|---|---|---|
| Validation Data | | | | |
| poly(4-methylcyclohexylmeth prop-2-enoate) | 1.9361 | 2.5263 | −1.5778 | 216.2040 |
| poly(3-methylcyclohexylmeth prop-2-enoate) | 1.9438 | 2.7669 | −1.5852 | 217.2347 |
| poly(*β*-naphthylmeth prop-2-enoate) | 2.5591 | 2.8013 | −2.2060 | 221.7641 |
| poly(*m*-crsylmeth prop-2-enoate) | 1.9567 | 2.8017 | −1.9959 | 197.9859 |
| poly(*o*-crsylmeth prop-2-enoate) | 1.9296 | 2.8031 | −2.0112 | 197.6172 |
| poly(3,3,5-trimethylcyclohexylmeth prop-2-enoate) | 2.2950 | 2.8159 | −1.6101 | 264.6907 |
| poly(α-naphthylcarbinyl) | 2.6555 | 2.8418 | −1.9340 | 239.8901 |
| poly(2-methylcyclohexylmeth prop-2-enoate) | 1.9573 | 2.8482 | −1.5281 | 216.6942 |
| poly(1,2-diphenylethylmeth prop-2-enoate) | 2.9978 | 2.8557 | −1.7160 | 289.9480 |
| poly(diphenylmethylmeth prop-2-enoate) | 2.8788 | 2.8592 | −1.6846 | 270.4855 |
| poly(*p*-cyclohexylphenylmeth prop-2-enoate) | 2.8882 | 2.8791 | −1.9375 | 274.7970 |
| poly(2-chlorocyclohexylmeth prop-2-enoate) | 1.9425 | 2.8942 | −2.2056 | 207.6941 |
| poly(*m*-nitrobenzylmeth prop-2-enoate) | 2.2144 | 2.9755 | −2.2056 | 227.0268 |
| poly(1-methylcyclohexylmeth prop-2-enoate) | 1.9496 | 2.9918 | −2.2056 | 218.8521 |
| poly(α-naphthylmeth prop-2-enoate) | 2.4447 | 3.0012 | −2.2056 | 222.0784 |
| poly[1-(*o*-chlorophenylethyl)meth prop-2-enoate] | 2.2384 | 3.0346 | −2.2056 | 231.0869 |
| poly(ethylmercaptylmeth prop-2-enoate) | 1.3936 | 3.1153 | −2.2056 | 145.6947 |
| poly(*o*-chlorobenzylmeth prop-2-enoate) | 2.0794 | 3.1564 | −2.2056 | 208.0084 |
| poly(2-nitro-2-methylpropylmeth prop-2-enoate) | 1.6951 | 3.2214 | −2.2056 | 215.1397 |
| poly[2-phenylsulfonyl)ethylmeth prop-2-enoate] | 2.4448 | 3.5313 | −2.2056 | 261.4602 |
| poly(*N*-butylmethacrylamide) | 1.5601 | 5.1610 | −2.2056 | 182.2105 |
| poly[*N*-(2-phenylethyl)methacrylamide] | 2.1535 | 5.2102 | −2.2056 | 217.8549 |

$$y_i^p = F_i(s_i^p) \tag{3}$$

Similarly, for any neuron $k$ from the output layer, the equations that determine its activation state are:

$$s_k^p = \sum_{i=1}^{L} w_{ik} y_i^p + b_k \tag{4}$$

$$y_k^p = F_k(s_k^p) \tag{5}$$

where $L$ is the neuron number of the intermediate layer, $w_{ik}$ is the weight value of the connection between the neuron $i$ from the intermediate layer and the neuron $k$ from the output layer, and $b_k$ is the value of the "bias" associated with the neuron $k$. The term of error for the output neuron is calculated by means of the following equation

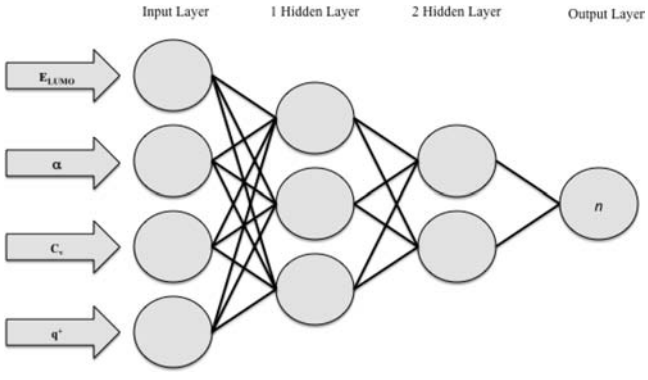


**Figure 1.** Diagram of a perceptron network constituted by four neurons in input layer, two middle layers (with three and two neurons respectively), and one output neuron.

$$E^p = \frac{1}{2} \sum_{k=1}^{M} (d_k^p - y_k^p)^2 \tag{6}$$

and if it adjusts to the previously established value, the training of the neural network finishes here. On the contrary, if it does not adjust to the previously established margins of error, the process would be repeated until reaching the desired error value. The activation equation used in this article is the sigmoid or logistic.

$$F_k(s_k^p) = \frac{1}{1 + e^{-s_k^p}} \tag{7}$$

A back-propagation rule (BP), which is a typical gradient-based learning algorithm, was used as a learning rule in the present work.

$$\Delta_p w_{ik} = -\eta \frac{\partial E^p}{\partial w_{ik}} = \eta (d_k^p - y_k^p) F_k'(s_k^p) y_i^p = \eta \delta_k^p y_i^p \tag{8}$$

This learning rule presents an important limitation, which is the large number of input cases for the training process, but for our study, a large amount of cases was available. This application could be interpreted as an ANN constituted by a primary neural layer (where the data of the input variables would be collected), an output neural layer (where the collected value would be obtained), and one or various intermediate layers (where the convergence work of the neural network would be facilitated) (see Figure 1).

As usual, the architecture of ANN is denoted as the following code: $N_{in} - [N_{h1} - N_{h2}]_e - N_{out}$, where $N_{in}$ and $N_{out}$ are the number of neurons in the input layer and output layer, respectively, $N_{h1}$ and $N_{h2}$ are the numbers of neurons in the first and second intermediate layer, respectively, and $e$ is the number of hidden layers.

**Table 3. Root Mean Square Errors (RMSE) of Training and Validation Using Different ANN Architectures**

| no. | topology | training (T) | validation (V) | $R^2_T$ | $R_T$ | $RMSE_T$ | $R^2_V$ | $R_V$ | $RMSE_V$ | $RMSE_{sum}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 4-[3-2]$_2$-1 | 73 | 22 | 0.9190 | 0.9586 | 0.0230 | 0.8700 | 0.9327 | 0.0172 | 0.0401 |
| 2 | 4-[7-1]$_2$-1 | 73 | 22 | 0.9019 | 0.9497 | 0.0235 | 0.9229 | 0.9607 | 0.0180 | 0.0415 |
| 3 | 4-[3-3]$_2$-1 | 73 | 22 | 0.9110 | 0.9545 | 0.0262 | 0.9100 | 0.9539 | 0.0158 | 0.0419 |
| 4 | 4-[3-1]$_2$-1 | 73 | 22 | 0.9118 | 0.9549 | 0.0262 | 0.8868 | 0.9417 | 0.0158 | 0.0419 |
| 5 | 4-[7-3]$_2$-1 | 73 | 22 | 0.8380 | 0.9154 | 0.0253 | 0.7501 | 0.8661 | 0.0175 | 0.0427 |
| 6 | 4-[8]$_1$-1 | 73 | 22 | 0.9145 | 0.9563 | 0.0242 | 0.8772 | 0.9366 | 0.0190 | 0.0432 |
| 7 | 4-[8-6]$_2$-1 | 73 | 22 | 0.9149 | 0.9565 | 0.0243 | 0.8035 | 0.8964 | 0.0213 | 0.0456 |
| 8 | 4-[7-2]$_2$-1 | 73 | 22 | 0.9206 | 0.9595 | 0.0251 | 0.8385 | 0.9157 | 0.0208 | 0.0459 |
| 9 | 4-[8-2]$_2$-1 | 73 | 22 | 0.9132 | 0.9556 | 0.0276 | 0.8694 | 0.9324 | 0.0184 | 0.0460 |
| 10 | 4-[8-7]$_2$-1 | 73 | 22 | 0.9080 | 0.9529 | 0.0287 | 0.8389 | 0.9159 | 0.0214 | 0.0502 |
| 11 | 4-[4]$_1$-1 | 73 | 22 | 0.9006 | 0.949 | 0.0292 | 0.8435 | 0.9184 | 0.0212 | 0.0504 |
| 12 | 4-[1]$_1$-1 | 73 | 22 | 0.8885 | 0.9426 | 0.0227 | 0.8521 | 0.9231 | 0.0322 | 0.0549 |
| 13 | 4-[7-5]$_2$-1 | 73 | 22 | 0.8744 | 0.9351 | 0.0333 | 0.82283 | 0.9071 | 0.0220 | 0.0553 |
| 14 | 4-[8-3]$_2$-1 | 73 | 22 | 0.8998 | 0.9486 | 0.0306 | 0.6142 | 0.7837 | 0.0363 | 0.0668 |

## Results and Discussion

The four variables were used like parameters of the information entrance. The ANN was trained with data corresponding to 73 prop-2-enoate polymers. The number of neurons in the intermediate layer was tested between $n/2 + 1$ and $2n + 1$, where $n$ corresponds with the input variables. Once trained, the ANN correct functioning has been tested with the validation data corresponding to 22 polymers.

In Table 3 are shown the training root-mean-square error (RMSE) value ($RMSE_T$) and the validation RMSE value ($RMSE_V$) for each ANN architecture. To evaluate the accuracy of the ANN model we use the sum of root-mean-square errors ($E_{sum}$) of the training set ($RMSE_T$) and the validation set ($RMSE_V$). The value of $E_{sum}$ (see Table 3) can be expressed as $E_{sum} = RMSE_T + RMSE_V$. As we can see in Table 3, the best ANN architecture consists in four input neurons [(i) the average polarizability (α), (ii) the most positive net atomic charge on hydrogen atoms in a molecule (e), (iii) the energy of the lowest unoccupied molecular orbital ($E_{LUMO}$), and (iv) the heat capacity at constant volume ($C_V$)], two middle layers (with three and two neurons, respectively), and one output neuron (see Figure 1). For the training of this, the ANN has established a target error of 0.01 %; the maximum number of training cycles was established as 2500, and the learning rate was set at 0.70 and the momentum value at 0.8. RMSEs are 0.023 ($R = 0.9586$) for the training set and 0.017 ($R = 0.9327$) for the prediction set (Figure 2).

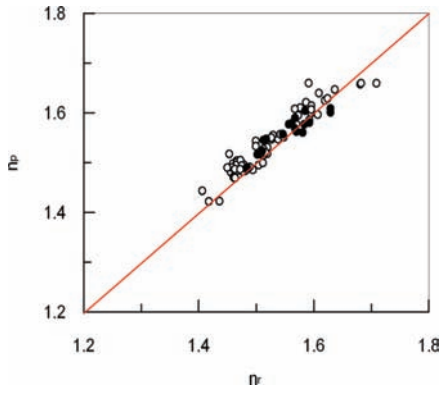


**Figure 2.** Plot of experimental data ($n_r$) versus predicted data ($n_p$) for training values (○) and validation values (●).

**Table 4. Weight Matrix of the ANN**

| | | | |
|---|---|---|---|
| −2.68524 | 7.85501 | −13.81874 | 91.998355 |
| −1.55745 | 23.82747 | −33.21980 | 5.099988 |
| −3.72450 | 32.75384 | −30.86519 | 10.729451 |
| −15.82583 | −23.86407 | −2.71565 | |
| −2.88338 | −8.84742 | −35.56258 | |
| −2.60864 | | | |
| −15.87540 | | | |

**Table 5. Importance of Variables Considered for ANN**

| value | importance | value | importance |
|---|---|---|---|
| $e$ | 107.823 | α | 64.436 |
| $C_V$ | 77.900 | $E_{LUMO}$ | 7.967 |

**Table 6. Adjustments for the Validation Values after Training and Validation of the ANN, *R*, and RMSE**

| value | $R$ | RMSE |
|---|---|---|
| training | 0.9586 | 0.0230 |
| validation | 0.9327 | 0.0172 |

Table 4 shows the ANN weight matrix. The values of the connections between the neurons are statistical weights of the connections, and corresponding to the synaptic strength of neuron connections they exceed a pre-set threshold value; the neuron fires. In any other case the neuron does not fire. Table 5 shows the importance of the variables selected for the ANN; this value is the sum of weights of the input neurons to all intermediate neurons.

Once the functioning of the ANN was verified, it was also checked by using the previously reserved validation data corresponding to 22 prop-2-enoate polymers. As shown in Figure 2, the data predicted by the ANN were confronted with the previously reserved validation data. The correlation coefficient is 0.93.

## Conclusions

ANN has been implemented to predict the values of $n$ of 95 polymers. Four variables were used from quantum chemistry, α, $e$, $E_{LUMO}$, and $C_V$, obtained from the monomer structures in the DFT.

Overall the ANN developed for this study consists of an input layer with four neurons, two intermediate layers formed by three and two neurons, respectively, and one neuron in the output layer which showed the best results obtained (Table 6). $n$ values predicted on training cases had a RMSE of 0.0230 ($R = 0.96$) and on validation cases of 0.0172 ($R = 0.93$).

## Literature Cited

(1) Knoll, W. Interfaces and thin films as seen by bound electromagnetic waves. *Annu. Rev. Phys. Chem.* **1998**, *49*, 569–638.
(2) Bicerano, J. J. *Prediction of Polymer Properties*; Marcel Dekker: New York, 1996.
(3) Van Krevelen, D. W. *Properties of Polymers: Correlation with Chemical Structure*, 2nd ed.; Elsevier: Amsterdam, 1976.
(4) García-Domenech, R.; de Julián-Ortiz, J. V. Prediction of indices of refraction and glass transition temperatures of linear polymers by using graph theoretical indices. *J. Phys. Chem. B* **2002**, *106*, 1501–1507.

(5) Katritzky, A. R.; Sild, S.; Karelson, M. Correlation and Prediction of the Refractive Indices of Polymers by QSPR. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 1171–1176.
(6) Xu, J.; Chen, B.; Zhang, Q. J.; Guo, B. Prediction of refractive indices of linear polymers by a four-descriptor QSPR model. *Polymer* **2004**, *45*, 8651–8659.
(7) Golbraikh, A.; Tropsha, A. Beware of q2! *J. Mol. Graphics Modell.* **2002**, *20*, 269–276.
(8) Bishop, M. C. *Neural Networks for Pattern Recognition*; Oxford University Press: New York, 1995.
(9) Rosenblatt, F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psych. Rev.* **1958**, *65*, 386–408.
(10) Haykin, S. *Neural Networks*; Macmillan College Publishing Company: New York, 1994.
(11) Xu, K.; Xie, M.; Tang, L.; Ho, S. L. Application of neural networks in forecasting engine system reliability. *Appl. Soft Comput.* **2003**, *2*, 255–268.
(12) Rumelhart, D. E.; McClelland, J. L. *Parallel distributed processing: Exploration in the microstructure of cognition*; MIT Press: Cambridge, MA, 1986.
(13) Castillo, E.; Gutiérrez, J. M.; Hadi, A. S.; Lacruz, B. Some applications of functional networks in statistics and engineering. *Technometrics* **2001**, *43*, 10–24.
(14) Castellano-Méndez, M.; Aira, M. J.; Iglesias, I.; Jato, V.; González-Manteiga, W. Artificial neural networks as a useful tool to predict the risk level of *Betula* pollen in the air. *Int. J. Biometeorol.* **2005**, *49*, 310–316.
(15) Rodríguez-Rajo, F. J.; Astray, G.; Ferreiro-Lage, J. A.; Aira, M. J.; Jato-Rodriguez, V.; Mejuto, J. C. Evaluation of atmospheric *Poaceae* pollen concentration using a neural network applied to a coastal Atlantic climate region. *Neural Networks* **2010**, *23*, 419–425.
(16) Astray, G.; Castillo, X.; Ferreiro-Lage, J. A.; Gálvez, J. F.; Mejuto, J. C. Artificial neural networks: a promising tool to evaluate the authenticity of wine. *J. Food Sci.* **2010**, *8*, 79–86.
(17) Astray, G.; Caderno, P. V.; Ferreiro-Lage, J. A.; Gálvez, J. F.; Mejuto, J. C. Prediction of Ethene + Oct-1-ene Copolymerization Ideal Conditions Using Artificial Neuron Networks. *J. Chem. Eng. Data* **2010**, *55*, 3542–3547.
(18) Chen, X.; Sztandera, L.; Cartwright, H. M. A neural network approach to prediction of glass transition temperature of polymers. *Int. J. Intell. Syst.* **2008**, *23*, 22–32.
(19) Wanqiang, L.; Chenzhong, C. Artificial neural network prediction of glass transition temperature of polymers. *Colloid Polym. Sci.* **2009**, *287*, 811–818.
(20) Astray, G.; Cid, A.; Ferreiro-Lage, J. A.; Galvez, J. F.; Mejuto, J. C.; Nieto-Faza, O. Prediction of Prop-2-enoate Polymer and Styrene Polymer Glass Transition Using Artificial Neural Networks. *J. Chem. Eng. Data* **2010**; DOI: 10.1021/je100573n.
(21) Astray, G.; Rodríguez-Rajo, F.; Ferreiro-Lage, J. A.; Fernández-González, M.; Jato, V.; Mejuto, J. C. The use of artificial neural networks to forecast biological atmospheric allergens or pathogens only as *Alternaria* spores. *J. Environ. Monit.* DOI: 10.1039/C0EM00248H.
(22) Astray, G.; Bendaña, R. Computer system for rural valuation based in Neural Networks Technology. *Elect. J. Environ. Agric. Food Chem.* **2007**, *6*, 1927–1933.
(23) Parr, R. G.; Yang, W. *Density functional theory of atoms and molecules*; Oxford University Press: New York, 1989.
(24) Johnson, B. G.; Gill, P. M. W.; Pople, J. A. The performance of a family of density functional methods. *J. Chem. Phys.* **1993**, *98*, 5612–5626.
(25) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision B.01; Gaussian, Inc.: Pittsburgh PA, 2003.
(26) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
(27) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785–789.
(28) Scott, A. P.; Radom, L. J. Harmonic Vibrational Frequencies: An Evaluation of Hartree-Fock, Møller-Plesset, Quadratic Configuration Interaction, Density Functional Theory, and Semiempirical Scale Factors. *J. Phys. Chem.* **1996**, *100*, 16502–16513.

Received for review August 26, 2010. Accepted September 22, 2010. G.A. thanks the "Ministerio de Educación" of Spain for a FPU grant P.P. 0000 421S 14006.

JE100885F