

Article

Prediction of Molecular Weight of Petroleum Fluids by Empirical Correlations and Artificial Neuron Networks

Dicho Stratiev ^{1,2,*} , Sotir Sotirov ³, Evdokia Sotirova ³, Svetoslav Nenov ⁴, Rosen Dinkov ¹, Ivelina Shishkova ¹ , Iliyan Venkov Kolev ^{1,2}, Dobromir Yordanov ³, Svetlin Vasilev ³, Krassimir Atanasov ^{2,3} , Stanislav Simeonov ³ and Georgi Nikolov Palichev ²

¹ LUKOIL Neftohim Burgas, 8104 Burgas, Bulgaria

² Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences, Georgi Bonchev 105, 1113 Sofia, Bulgaria

³ Intelligent Systems Laboratory, Department Industrial Technologies and Management, University Prof. Dr. Assen Zlatarov, 8010 Burgas, Bulgaria

⁴ Department of Mathematics, University of Chemical Technology and Metallurgy, Kliment Ohridski 8, 1756 Sofia, Bulgaria

* Correspondence: stratiev.dicho@neftochim.bg

Abstract: The exactitude of petroleum fluid molecular weight correlations affects significantly the precision of petroleum engineering calculations and can make process design and trouble-shooting inaccurate. Some of the methods in the literature to predict petroleum fluid molecular weight are used in commercial software process simulators. According to statements made in the literature, the correlations of Lee–Kesler and Twu are the most used in petroleum engineering, and the other methods do not exhibit any significant advantages over the Lee–Kesler and Twu correlations. In order to verify which of the proposed in the literature correlations are the most appropriate for petroleum fluids with molecular weight variation between 70 and 1685 g/mol, 430 data points for boiling point, specific gravity, and molecular weight of petroleum fluids and individual hydrocarbons were extracted from 17 literature sources. Besides the existing correlations in the literature, two different techniques, nonlinear regression and artificial neural network (ANN), were employed to model the molecular weight of the 430 petroleum fluid samples. It was found that the ANN model demonstrated the best accuracy of prediction with a relative standard error (RSE) of 7.2%, followed by the newly developed nonlinear regression correlation with an RSE of 10.9%. The best available molecular weight correlations in the literature were those of API (RSE = 12.4%), Goosens (RSE = 13.9%); and Riazi and Daubert (RSE = 15.2%). The well known molecular weight correlations of Lee–Kesler, and Twu, for the data set of 430 data points, exhibited RSEs of 26.5, and 30.3% respectively.

Keywords: petroleum; molecular weight; modeling; artificial neural network; nonlinear regression; empirical correlation



check for updates

Citation: Stratiev, D.; Sotirov, S.; Sotirova, E.; Nenov, S.; Dinkov, R.; Shishkova, I.; Kolev, I.V.; Yordanov, D.; Vasilev, S.; Atanasov, K.; et al. Prediction of Molecular Weight of Petroleum Fluids by Empirical Correlations and Artificial Neuron Networks. *Processes* **2023**, *11*, 426. <https://doi.org/10.3390/pr11020426>

Academic Editor: Qingbang Meng

Received: 24 November 2022

Revised: 19 January 2023

Accepted: 27 January 2023

Published: 31 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In petroleum engineering simulations to design new equipment or rate the existing one the molecular weight of the petroleum fluids is one of the most important characterization parameters [1–5]. It affects thermodynamic phase equilibrium, reaction kinetics, and vapor density calculations [5]. The measurement of petroleum fluid molecular weight is not a simple job, and, depending on the method used, either vapor pressure osmometry or freezing point depression, some discrepancies may arise [6]. For example, Powers et al. [7] reported repeatability of 15% for molecular weight measurement of asphaltenes, while Yarranton et al. [8] reported repeatability of 12% for saturates/aromatics/resins (SAR), and Lemus et al. [2] reported repeatability of $\pm 15\%$ for the bitumen and its distillation cuts. All these studies [2,7,8] employed vapor pressure osmometry. Thus, petroleum fluid molecular weight has the highest measurement uncertainty of the three main physical

properties: boiling point, specific gravity, and molecular weight [1]. Considering the higher precision of measurement of boiling point and specific gravity, several empirical correlations were developed to predict petroleum fluid molecular weight from boiling point and specific gravity (density) [1,2,6–23]. Schneider [6] summarized the petroleum fluid molecular weight correlations, which appeared in the period 1964–1984. He concluded that the molecular weight correlation selected impacts engineering calculations to a significant degree [6]. During simulation of a crude distillation column for Bachaquero crude feed constant side-draw product rates, constant barrels per day (bpsd) overflash, constant column bottoms true boiling point (TBP) 5% point flash zone temperature variations of 10 to 30 °F, and heater duty differences of 5 to 10%, were obtained when different molecular weight correlations were employed [6]. Such large variations, he concluded, would make the process design and troubleshooting inaccurate. Unfortunately, he did not point out which of the studied molecular weight correlations were most appropriate. Instead he deduced that a molecular weight correlation used for any purpose should not be extended beyond the limits for which it is valid without caution. Goosens [9] announced in his research that a major oil company had reviewed 37 different methods and their marked deviations beyond molecular weights of 300 g/mol. A major reason for the abundance of still less satisfactory correlations, he underlined, is the fact that the principal independent variable for the prediction of the molecular weight, being some average boiling point, is often ill-defined. Lemus in her PhD thesis [1] summarized the petroleum fluid molecular weight correlations, which appeared in the period 1989–2010. She reported average absolute relative deviation for the molecular weight prediction of bitumen and heavy oils ranging between 7.6 and 30.7% for the different literature correlations [1]. The new molecular weight correlation proposed in her PhD thesis shows an average absolute relative deviation of 5.3%. Since 2016, no reports have appeared in the literature to summarize the available information about the molecular weight correlations of petroleum fluids. Moreover, in 2021 Hosseinifar, and Shahverdi [10] proposed a new petroleum fluid molecular weight correlation suggesting that it should be superior to others. Considering the importance of the molecular weight in petroleum engineering calculations, we decided to make a generalization of the correlations published in the literature for the period 1969–2021. Table 1 summarizes 13 published in the literature petroleum fluid molecular weight correlations for the period 1969–2021. Nine of these correlations (Hariu and Sage [21], Kesler and Lee [13], Riazi and Daubert (1980) [15], Twu [20], Rao and Bardon [19], Soreide [12], Goosens [9], Riazi and Daubert for $MW \leq 300$ [18], Riazi and Daubert for $MW \leq 700$ [15]) and their applications have been already discussed in the PhD thesis of Lemus [1]. The additional four correlations included in Table 1 (Liñan et al. [23], Liñan et al. (API) [23], Lemus et al. [2], and Hosseinifar, and Shahverdi [10]) have not been examined with a great number of data points, and can be considered relatively new because the last assessment of molecular weight correlations was reported in 2016 [2]. The correlation of Liñan et al. [23] has been developed to predict molecular weight of petroleum residues and cuts. The correlation of Lemus et al. [2] has been developed to predict molecular weight of heavy oil distillation cuts. The correlation of Hosseinifar, and Shahverdi [10] has been developed for petroleum fluids having a specific gravity from 0.68 to 0.92, boiling points from 340 to 722 K, and molecular weight from 84 to 414 g/mol.

There is no consent in the literature about which correlation can be deemed most appropriate to predict the molecular weight of petroleum fluids. For example, Liu et al. [24] argue that the developed correlations after those of Lee–Kesler correlation [13,14], and the Twu correlation [20], described by Riazi [18] do not have a significant advantage over the Lee–Kesler or Twu correlations. Liu et al. [24] reported that Aspen HYSYS uses the Twu correlation to calculate the molecular weight of petroleum fluids.

Table 1. Summary of correlations available in the literature to estimate molecular weight.

Source (Year of Publication)	Correlation	Range of Applicability	Eq.
Hariu, and Sage [21] (1969)	$MW = 0.6670202 + 0.1552531Kw - 0.005378496Kw^2 + 0.004583705T_bKw + 0.00002500584T_bKw^2 + 0.000002698693T_b^2 + 0.0000387595T_b^2Kw - 0.0000001566228T_b^2 \times Kw^2$	$30 < T_b < 800$	(1)
Kesler and Lee [13] (1976)	$MW = [D] + \frac{[A]}{T_b} + \frac{[B]}{T_b^3}$ $D = -12272.6 + 9486.4 \times SG + (4.6523 - 3.3287 \times SG) \times T_b$ $A = (1 - 0.77084 \times SG - 0.02058 \times SG^2) \times \left(1.3437 - \frac{720.79}{T_b}\right) \times 10^7$ $B = (1 - 0.80882 \times SG + 0.02226 \times SG^2) \times \left(1.8828 - \frac{181.92}{T_b}\right) \times 10^{12}$	$T_b < 750$	(2)
Riazi and Daubert [15] (1980)	$MW = 0.000045673 \times T_b^{2.1962} \times SG^{-1.0164}$	$300 < T_b < 610$	(3)
Twu [20] (1984)	$\ln(M) = \ln M^\circ \left[\frac{(1 + 2f_M)}{(1 - 2f_M)} \right]^2$ $f_M = \Delta SG_M \left[\chi + \left(-0.0175691 + \frac{0.143979}{T_b^{1/2}} \right) \Delta SG_M \right]$ $\chi = \left[0.012342 - 0.2445541 / T_b^{1/2} \right]$ $\Delta SG_M = \exp \left[5 \left(SG^\circ - SG \right) \right] - 1$ $SG^\circ = 0.843593 - 0.128624\alpha - 3.36159\alpha^3 - 13749.5\alpha^{12}$ $\alpha = 1 - T_b / T_b^\circ$	$MW < 600$	(4)
Rao and Bardon [19] (1985)	$\ln MW = (1.27 + 0.071Kw) \ln \left(\frac{1.8T_b}{22.31 + 1.68Kw} \right)$	$361 < T_b < 830$	(5)
Soreide [12] (1989)	$T_b = [1928.3 - 169500 \times MW^{-0.03522} \times SG^{3.266} D] / 1.8$ $D = \text{EXP}(0.004922 \times MW - 4.7685 \times SG + 0.003462 * MW * SG)$	$361 < T_b < 830$	(6)
Goosens [9], (1996)	$MW = 0.010770T_b^{[1.52869 + 0.06486 \ln(\frac{T_b}{1078 - T_b})]} / d$	$306 < T_b < 1012$	(7)
Riazi and Daubert for MW ≤ 300 g/mol [18] (2005)	$M = 1.6607 \times 10^{-4} T_b^{2.1962} SG^{-1.0164}$	$300 < T_b < 610$	(8)
Riazi and Daubert for MW ≤ 700 g/mol [18], (2005)	$M = 42.965 \left[\exp(2.097 \times 10^{-4} T_b - 7.78712SG + 2.08476 \times 10^{-3} T_b SG) \right] T_b^{1.26007} SG^{4.98308}$	$300 < T_b < 900$	(9)
Liñan et al. [23] (2011)	$M_{WadC} = 284.75 \left[\exp(0.00322(t_{VABP} + 273.15)) \right] \left[\exp(-2.52SG) \right] \times (t_{VABP} + 273.15)^{0.083} SG^{2.44}$	$673 < T_b < 1235$	(10)
Liñan et al. [23] (2011) (API)	$M_{WAPI} = 219.05 \left[\exp(0.0039(t_b + 273.15)) \right] \left[\exp(-3.07SG) \right] \times (t_b + 273.15)^{0.118} SG^{1.88}$		(11)
Lemus et al. [2] (2016)	$T_b = [1805 - 21131 \times MW^{-0.049} \times SG^{1.5258} D] / 1.8$ $D = \text{EXP}(-0.005 \times MW - 2.675 \times SG + 0.003 \times MW \times SG)$	$300 < T_b < 900$	(12)
Hosseiniifar, and Shahverdi [10] (2021)	$M_w(T_b, SG) = \left d_1 \times T_b^{d_2} \left(\frac{3+2SG}{3-SG} \right)^{\frac{d_3}{2}} + d_4 \times T_b^{d_5} \left(\frac{3+2SG}{3-SG} \right)^{\frac{d_6}{2}} \right ^{d_7}$	$340 < T_b < 722$	(13)

Goosens [9] stated that the correlation developed in his work was superior to that of API procedure 2B2.1 imposed as a standard that was limited to molecular weights of 700. The correlation of Goosens covers the full practical range of molecular weights of 75–1700 [9]. Hosseiniifar and Shahverdi [10] have developed recently a new petroleum fluid molecular weight correlation, which is suggested to have advantage over the published (until 2021) other molecular weight correlations. In addition, Hosseiniifar and Shahverdi [25] developed a method to generate TBP distillation data from molecular weight, or $T_{50\%}$ and density, that allowed us, in our recent study, to employ artificial neural network (ANN) to predict petroleum viscosity [26]. The same approach could be applied to predict petroleum fluid molecular weight from boiling point and density using the method of Hosseiniifar and Shahverdi [25] to generate more than two (density and boiling point) inlet parameters to allow the employment of the ANN method. In our recent research, the ANN method using the approach discussed above demonstrated the best prediction of petroleum viscosity [26]. In order to evaluate the molecular weight prediction ability of

the available empirical correlations in the literature, data from 430 petroleum fluids and individual hydrocarbons with molecular weight variation between 70 and 1685 g/mol were selected from 21 literature sources [1,2,9,22,23,27–42] and shown in Table S1. The data generated by the use of the Hosseinifar and Shahverdi [25] method, described in our recent research [26], along with specific gravity, boiling point, and the corresponding molecular weight of the 430 petroleum fluids to be used for molecular weight modeling by the ANN method, are presented Supplementary Table S2. The wide range of variation of molecular weight, boiling points, and density of the 430 petroleum fluids and individual hydrocarbons allows us to explore the capabilities of two commonly used techniques to model petroleum properties: the nonlinear regression and the ANN. This can give a notion as to which method is better to model petroleum fluid molecular weight. Comparisons between capabilities of the nonlinear regression and the metaheuristic methods to model oil properties have been reported in several studies [43–48]. However, no such study has been reported yet, to the best of our knowledge, for modelling of petroleum fluid molecular weight by the use of ANN. That was the reason for us to conduct this research.

The aim of this investigation is to develop two new methods to predict petroleum fluid molecular weight using two different techniques, nonlinear regression and ANN, and contrast them with the empirical correlations available in the literature.

2. Materials and Methods

The petroleum fluids, whose measured density, boiling point, and molecular weight were extracted from the literature, vary between light naphtha and vacuum residue. The carbon number of the individual hydrocarbons, extracted from the literature, and used in this study, varies between C₆ and C₅₂. The boiling point of the 430 petroleum fluids and individual hydrocarbons varies between 30 (303K) and 739 °C (1012K), and specific gravity at 15.6 °C varies between 0.631 and 1.527. The boiling point, specific gravity, and measured molecular weight of the 430 oils are summarized in Table S1.

The method of Hosseinifar and Shahverdi [25] and its use to generate TBP boiling point at 5, 10, 30, 50, 70, and 90 vol.% and its application for making more inlet parameters required by the ANN method is detailed in our recent research [26]. The equations used to generate TBP boiling point at 5, 10, 30, 50, 70, and 90 vol.% are shown below:

$$T_{50} = 1.037003 * T_b - 16.0825 \quad (14)$$

$$T_{70} = -1.06487 * T_b + 18.49267 \quad (15)$$

$$T_{30} = \left[(1.679547 * T_b - 31.7009)^3 - T_{50}^3 - 2T_{70}^3 \right]^{1/3} \quad (16)$$

$$T_{10} = \frac{[Abs((-0.94473 * Exp(SG) + 3.178959) * Kw^{-1.52054})]^{1.011152}}{T_{50}^{-1.58808}} \quad (17)$$

$$T_{90} = \frac{[-1.01857 * 0.5 * (T_b + T_{70}) + 11.24528]^2}{0.5 * (T_{50} + T_{10})} \quad (18)$$

$$T_5 = Abs\left((1.42946 * T_b * T_{50})^{0.5} - 7.25905 \right)^3 - T_{70}^3 - T_{90}^3)^{0.343176} \quad (19)$$

where,

T₅₀–TBP boiling point at 50% evaporate, °C;

T₇₀–TBP boiling point at 70% evaporate, °C;

T₃₀–TBP boiling point at 30% evaporate, °C;

T₁₀–TBP boiling point at 10% evaporate, °C;

T₉₀–TBP boiling point at 90% evaporate, °C;

T₅–TBP boiling point at 5% evaporate, °C;

It is worth mentioning here that the use of the method of Hosseinifar and Shahverdi [25] has not been made to construct the real TBP curve, but instead to make more than two

inlet parameters (specific gravity, and boiling point) to apply for ANN modeling purposes. Kw-characterization factor of the studied oils was estimated as shown in our recent article [26]. Table S2 summarizes the data employed for ANN modeling that includes nine inlet characterizing parameters: specific gravity, boiling point (K), $T_{5\%}$, $T_{10\%}$, $T_{30\%}$, $T_{50\%}$, $T_{70\%}$, $T_{90\%}$, and Kw. The Kw-characterization factor for the studied oils varies between 7.4 and 14.5.

The computer algebra system (CAS) Maple and NLPsolve with Modified Newton Iterative Method as described in [43] was employed to develop a new empirical correlation predicting molecular weight of petroleum fluids using the data extracted from the literature and discussed above.

The artificial neural network (ANN) modeling approach used in this investigation is described in our earlier research [26].

The accuracy of petroleum fluid molecular weight prediction has been evaluated by the statistical parameters shown as Equations (20)–(25).

$$\text{Error (E)} : E = \left(\frac{MW_{exp} - MW_{calc}}{MW_{exp}} \right) \times 100 \quad (20)$$

$$\text{Standarderror (SE)} : SE_i = \left(\sum \left(\frac{(MW_{exp} - MW_{calc})^2}{N - 2} \right) \right)^{\frac{1}{2}} \quad (21)$$

$$\text{Relativestandarderror (RSE)} : RSE_i = \frac{SE}{\text{mean of the sample}} \times 100 \quad (22)$$

$$\text{Sumofsquareerrors (SSE)} : SSE = \sum \frac{1}{MW_{exp}^2} (MW_{exp} - MW_{calc})^2 \quad (23)$$

$$\text{Averageabsolutedeviation (%AAD)} : \%AAD = \frac{1}{N} \frac{\sum |MW_{exp} - MW_{calc}|}{MW_{exp}} \times 100 \quad (24)$$

$$\text{Sumofrelativeerrors (SRE)} : SRE = \sum \left(\frac{MW_{exp} - MW_{calc}}{MW_{exp}} \right) \times 100 \quad (25)$$

3. Results

Goosens [9] mentioned in his work that a great number of petroleum fluid molecular weight correlations exhibited marked deviations beyond molecular weights of 300 g/mol. That was the reason for us to select a data base with a very wide range of variation of molecular weight, boiling points, and specific gravity. Figure 1 shows a graph of dependence of molecular weight on boiling point for the selected data base of 430 petroleum fluids and individual hydrocarbons. It is evident from this data that molecular weight exponentially increases with boiling point augmentation. It can be also seen that some individual polynuclear aromatic hydrocarbons deviate from the general exponential dependence, displaying lower molecular weight at the same boiling point.

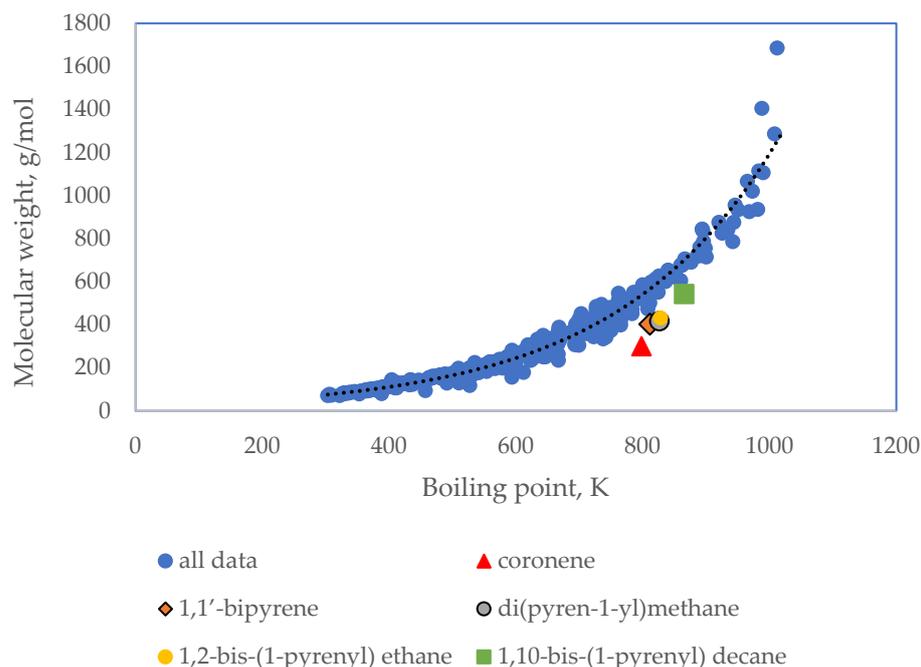


Figure 1. Dependence of molecular weight on boiling point for petroleum fluids and individual hydrocarbons.

The specific gravity, which well reflects the aromatic structure content [45], accounts for this deviation of the polynuclear aromatic compounds from the general exponential dependence of molecular weight on boiling point. By the use of CAS Maple and NLPsolve with Modified Newton Iterative Method and 306 points (the data with numbers from 1 to 306 from Table S1) out of 430, the following new empirical correlation was developed:

$$MW = -552.982 + 453.095 * EXP(0.19239 * EXP\left(0.000421163 * \frac{BP^{1.22097}}{SG^{0.297075}}\right)) \quad (26)$$

where,

BP = boiling point of the petroleum fluid or individual hydrocarbon, K;

SG = specific gravity at 15.6 °C.

The remaining 124 data points were used to test the validity of Equation (26). For the development of the artificial neural network, a 3-layer deep learning neural network with a structure of 9 inputs (SG , BP , $T_{5\%}$, $T_{10\%}$, $T_{30\%}$, $T_{50\%}$, $T_{70\%}$, $T_{90\%}$, and K_w), 12 neurons in the first layer, 1 neuron in the output layer, and 1 output was used. Figure 2 presents the structure of the neural network employed in this study.

The entire training process took 18 iterations, and the performance was 0.0031481. The training process is shown in the Figure 3.

The regression coefficients of the artificial neural network are: Training—0.99705, Testing—0.99118, Validation—0.99638, All—0.99615 (Figure 4).

It is worth mentioning here that, during ANN modeling, the logarithm of molecular weight was modeled and the output of the ANN model was converted in molecular weight by the use of exponential function.

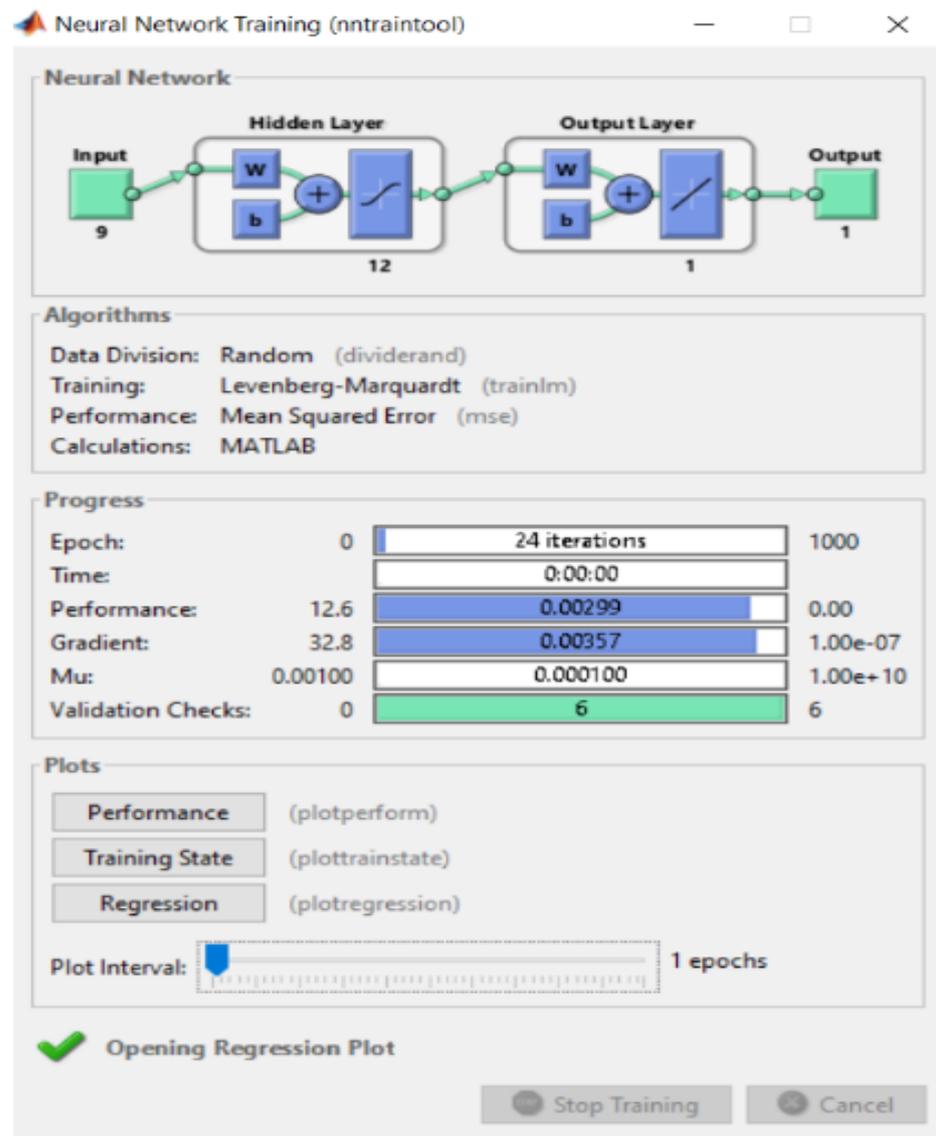


Figure 2. Structure of the neural network employed in this study.

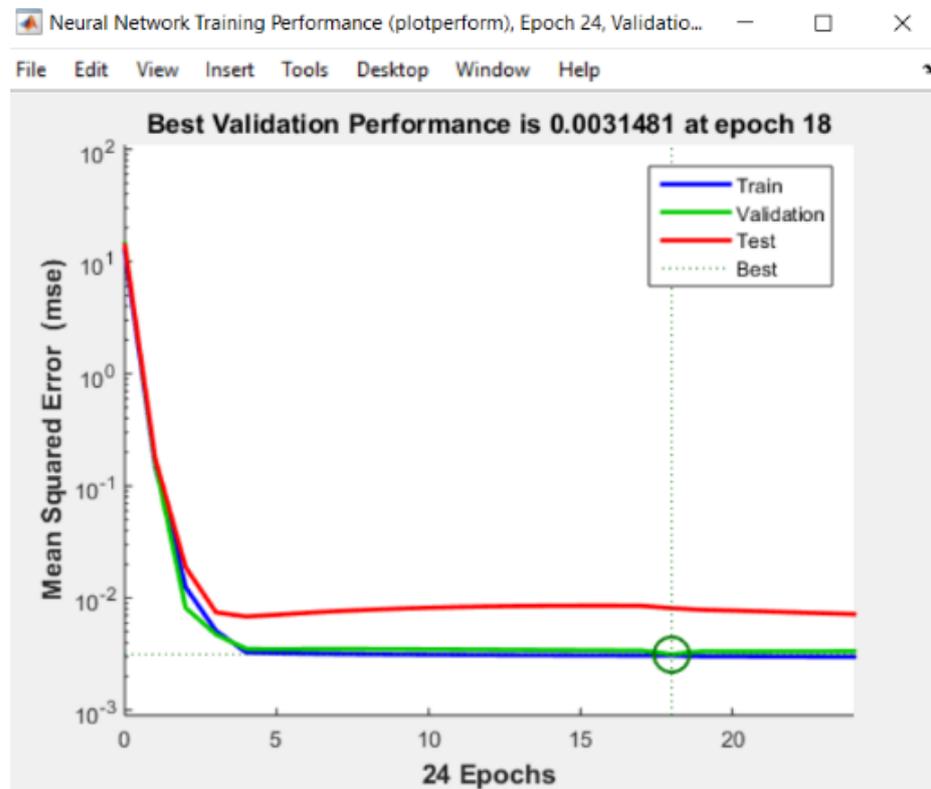


Figure 3. Training process of the neural network.

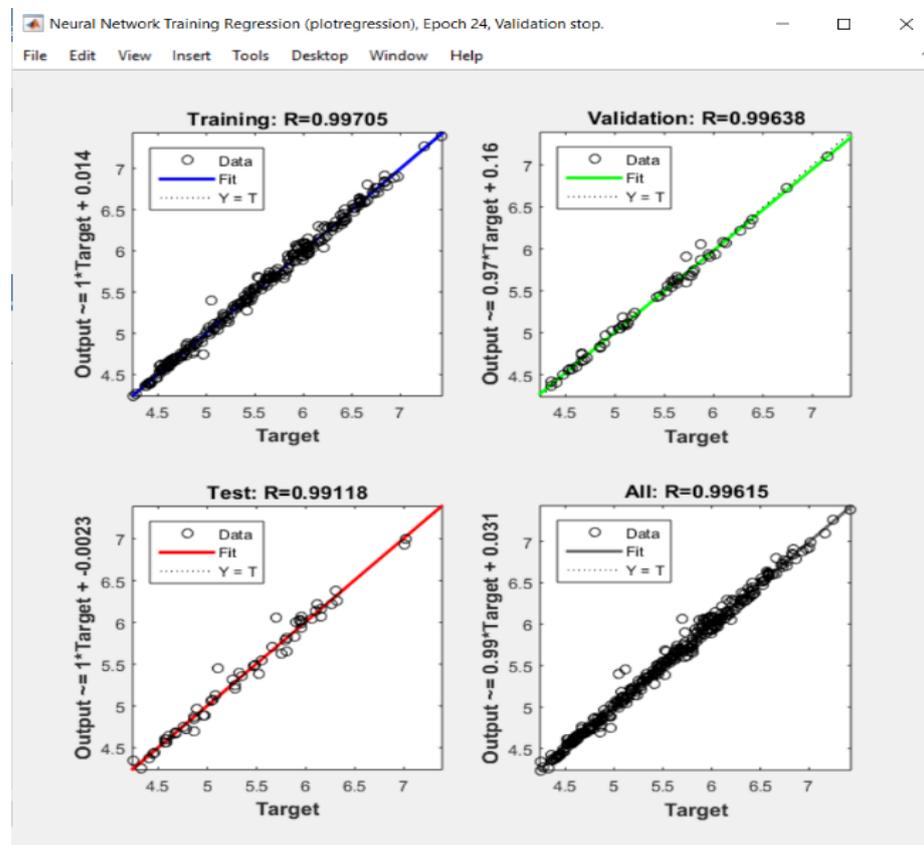


Figure 4. Regression coefficients of the learning process of the ANN.

Figure 5 presents parity graphs for both new methods to predict molecular weight of petroleum fluids using nonlinear regression (Figure 5a) and ANN (Figure 5b).

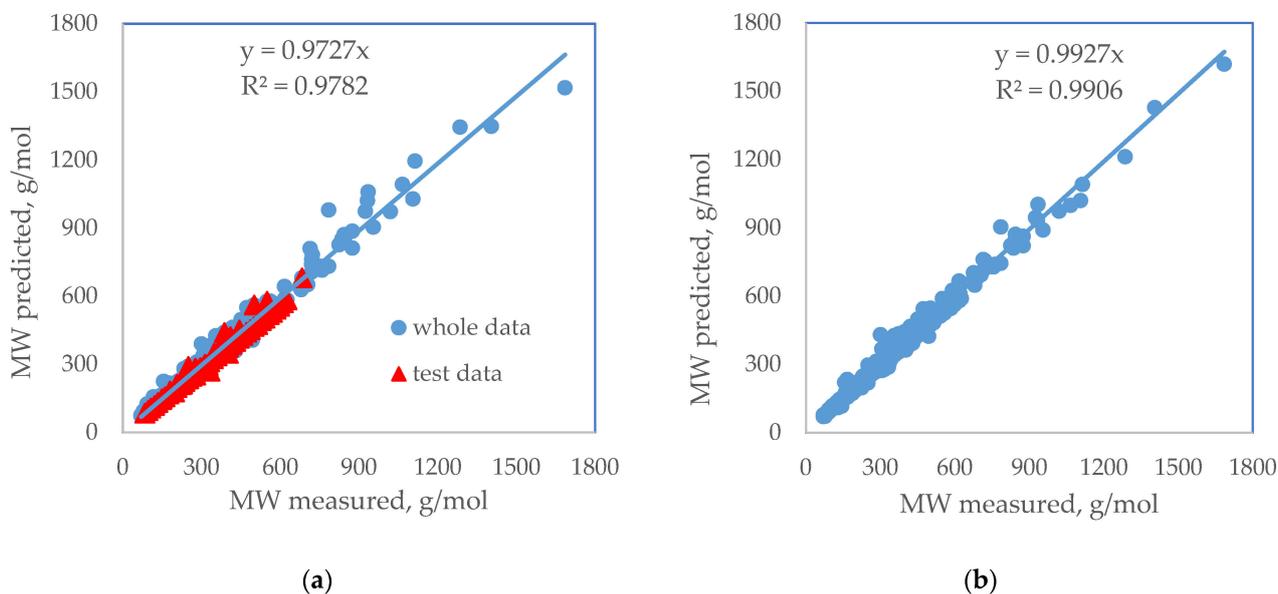


Figure 5. Parity graph for predicted versus measured molecular weight by Equation (20) (a) and by the ANN model (b).

Figures 6–9 presents parity graphs of predicted versus measured molecular weight by the empirical correlations of Goosens (Figure 6a), the API method reported in [23], Riazi and Daubert correlations for $MW \leq 300$ (Figure 7b) and 700 g/mol (Figure 7a), Kesler and Lee [14] (Figure 8a), Twu (Figure 8b), Liñan et al. [23] (Figure 9 a), and Hosseinifar and Shahverdi [10] (Figure 9b).

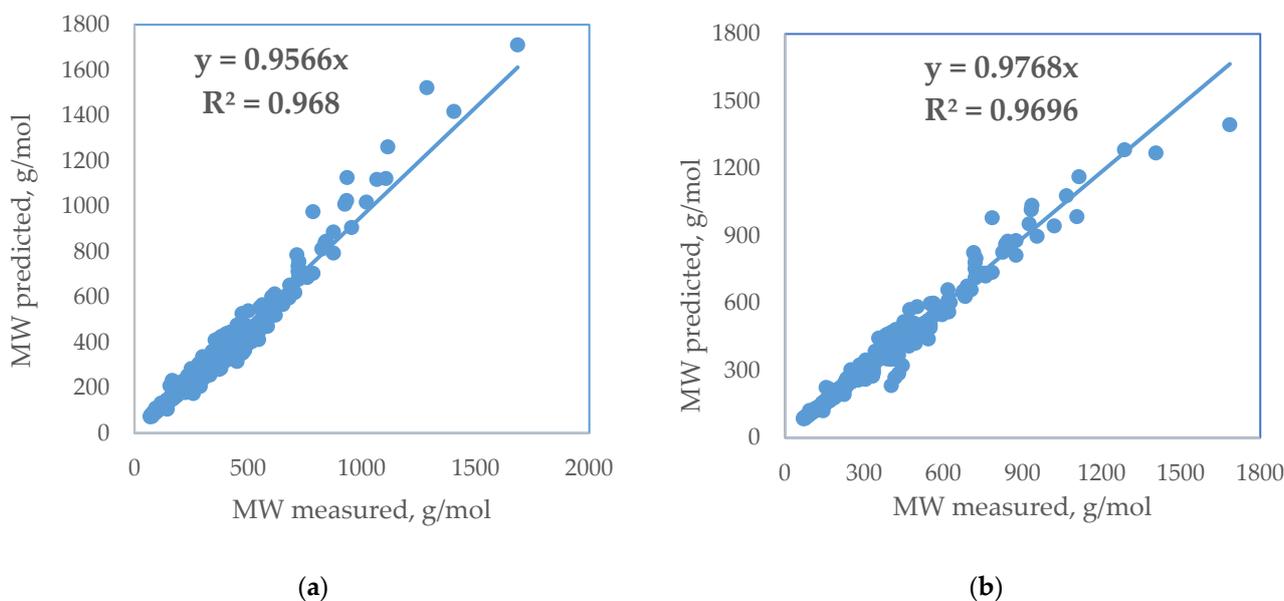


Figure 6. Parity graph for predicted versus measured molecular weight by Equation (7) (Goosens correlation) (a) and by Equation (11) (API method, 2011) (b).

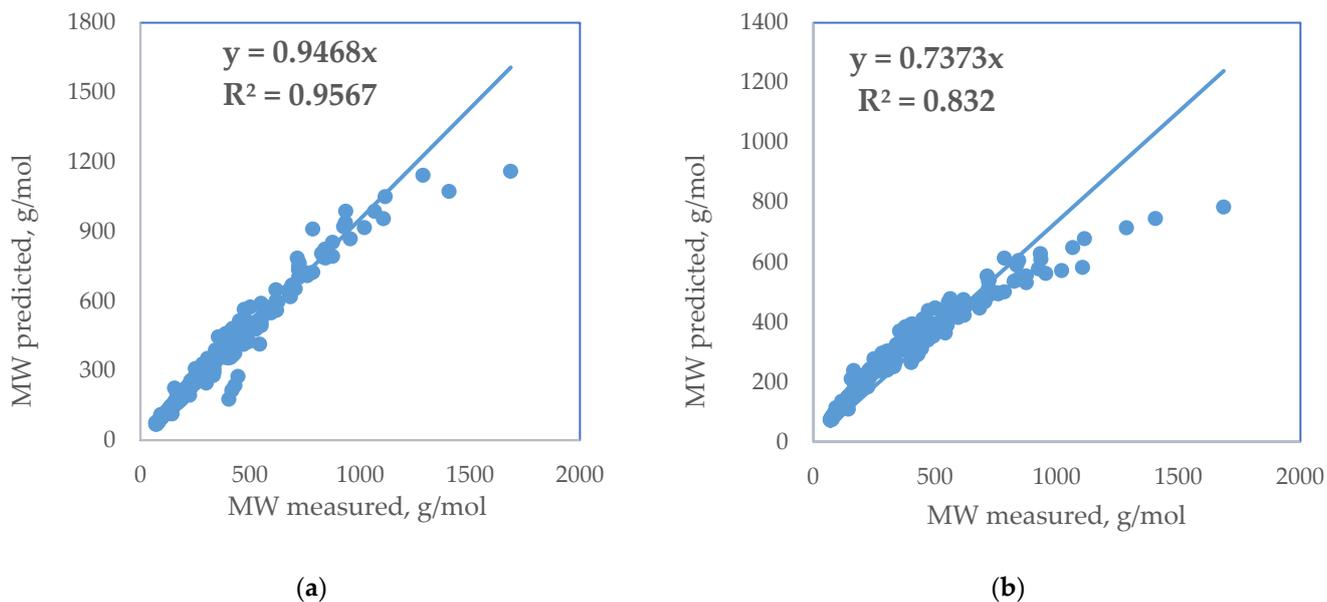


Figure 7. Parity graph for predicted versus measured molecular weight by Equation (9) (Riazi and Daubert correlation up to $MW \leq 700$ g/mol) (a) and by Equation (8) (Riazi and Daubert correlation up to $MW \leq 300$ g/mol) (b).

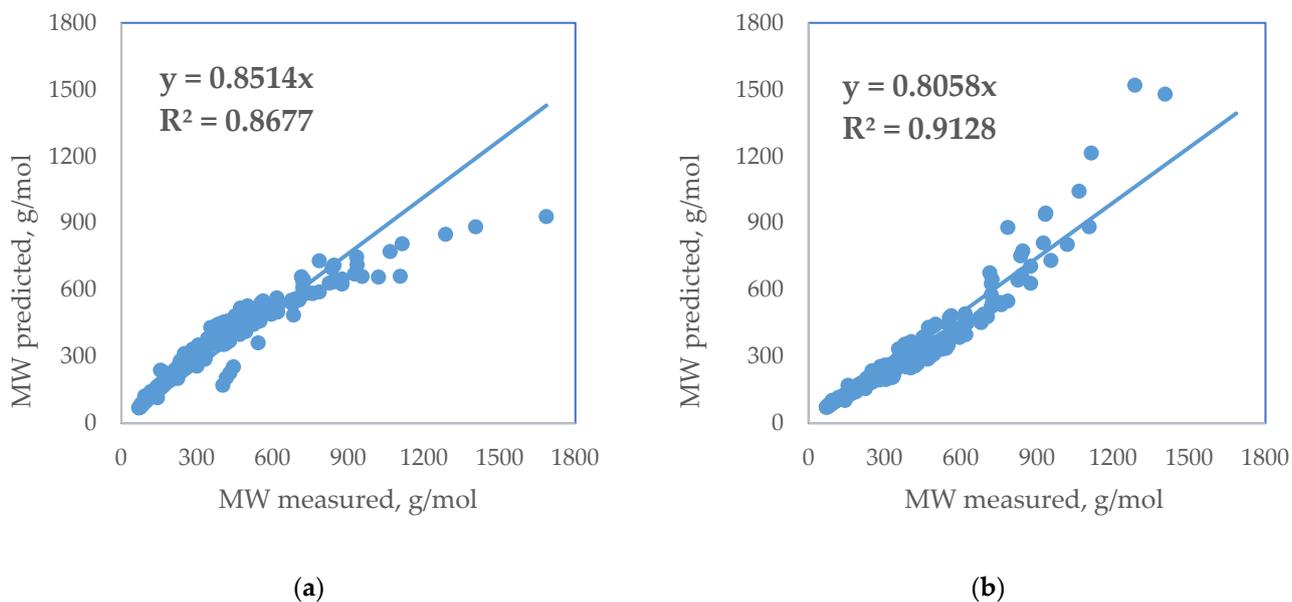


Figure 8. Parity graph for predicted versus measured molecular weight by Equation (2) (Kesler and Lee) (a) and by Equation (4) (Twu) (b).

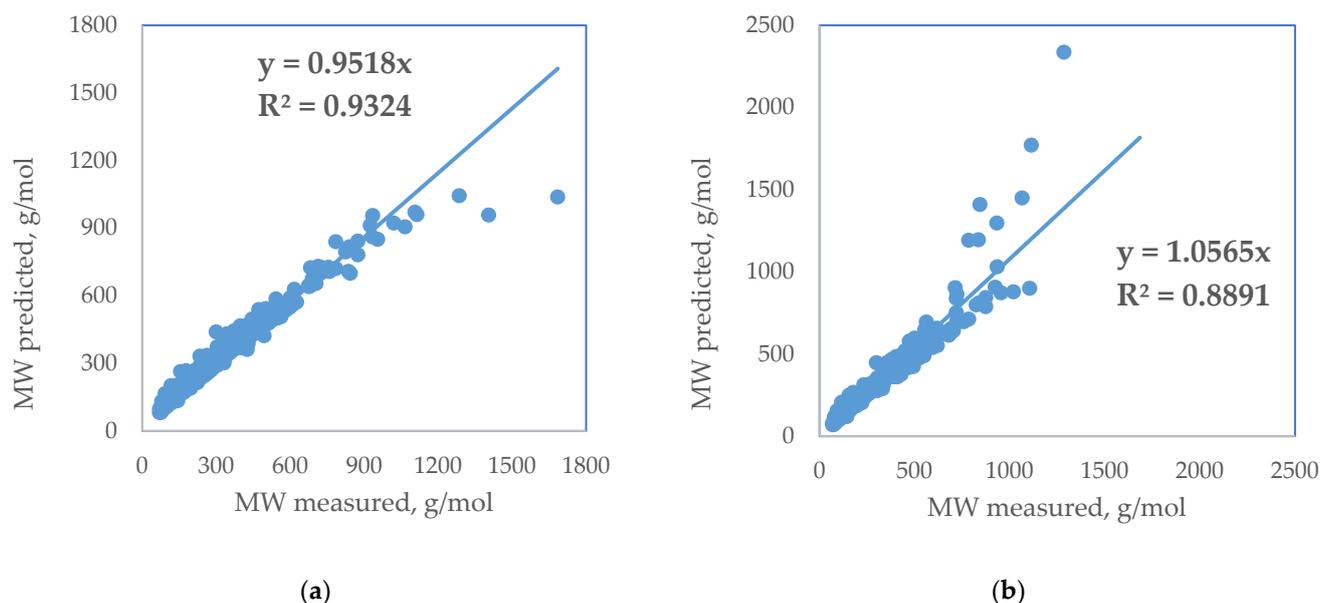


Figure 9. Parity graph for predicted versus measured molecular weight by Equation (11) (Liñan et al.) (a) and by Equation (13) (Hosseiniifar, and Shahverdi) (b).

Table 2 summarizes the statistical analyses for the new empirical correlation, the ANN model, and the ten studied empirical correlations available from the literature.

Table 2. Statistical analysis of studied methods to predict petroleum fluid molecular weight for the whole range of studied molecular weights.

	Standard Error	Rel. St. Error	Sum of Squared Errors	%AAD	SRE
ANN	23.0	7.2	1.6	4.3	−126.8
New empirical correlation (this work)	34.8	10.9	3.4	6.4	695.5
Hosseiniifar (2021) [10]	84.4	26.6	8.6	9.4	−1851.7
Goosens (1996) [9]	44.1	13.9	4.5	7.6	1970.2
Riazi and Daubert MW ≤ 700 g/mol (2005) [18]	48.3	15.2	3.9	6.3	358.2
API (2011) [23]	39.3	12.4	5.1	8.6	−999.1
Liñan (2011) [23]	54.4	17.1	10.2	11.3	−2973.1
Twu (1984) [20]	96.4	30.3	20.6	18.7	7840.0
Lee–Kesler (1976) [14]	84.2	26.5	7.1	9.7	345.9
Riazi and Daubert ≤ 300 g/mol (2005) [18]	117.3	36.9	12.4	12.5	4155.7
Lemus et al. (2016) [2]	103.0	32.4	14.0	10.8	864.9
Soreide (1989) [12]	84.1	26.5	12.2	9.4	−598

The data in Table 2 indubitably indicates that the ANN method provides the best accuracy of molecular weight prediction of petroleum fluids, following by the new empirical correlation developed in this work. If one compares the data for accuracy of molecular weight prediction of the methods tested in this work with the 430 petroleum fluid sample points with those reported by the authors who have developed the correlations, some discrepancies will be seen. For example, Goosens [9] reported standard deviation for his method of about 2%, while the relative standard error, that can be assumed equivalent to the standard deviation shown in Table 2, is 13.9%. The %AAD of Riazi and Daubert MW ≤ 700 g/mol was reported as 4.7 [18], while that shown in Table 2 for this method is 6.3%. Liñan et al. [23] reported %AAD of 2.2, while that shown in Table 2 for this method is 11.3. Lemus et al. [2] reported %AAD of 5.3, while that shown in Table 2 for this method is 10.8. The reason for these discrepancies may lie in the range for which the discussed correlations were developed, confirming again the statement of Schneider [6] that any

molecular weight correlation should not be extended beyond the limits for which it is valid. Considering that some of the tested correlations were developed for a narrower range of boiling point and molecular weight variations, a statistical analysis of the studied methods was performed for the petroleum fluids whose molecular weight is ≤ 300 g/mol.

Table 3 presents data of standard error, relative standard error, and %AAD for the predictive methods employing only these petroleum fluid data whose molecular weight is ≤ 300 g/mol.

Table 3. Statistical analysis of studied methods to predict molecular weight for petroleum fluids and individual components whose molecular weight ≤ 300 g/mol.

	Standard Error	Rel. St. Error	%AAD
ANN	13.4	7.9	3.6
New empirical correlation (this work)	16.5	9.7	5.1
Hosseiniifar (2021) [10]	18.1	10.7	8.1
Goosens (1996) [9]	17.8	10.5	5.4
Riazi and Daubert MW ≤ 700 g/mol (2005) [18]	16.0	9.4	4.0
API (2011) [23]	17.0	10.0	5.4
Twu (1984) [20]	38.8	22.9	13.0
Lee–Kesler (1976) [14]	18.3	10.8	7.3
Riazi and Daubert MW ≤ 300 g/mol (2005) [18]	11.5	6.8	4.6
Lemus et al. (2016) [2]	29.4	17.3	8.0
Soreide (1989) [12]	32.0	18.8	7.6

The data in Table 3 indicates that the correlation of Riazi and Daubert MW ≤ 300 g/mol and the ANN model demonstrate the highest accuracy in molecular weight prediction, followed by the new correlation developed in this work. It is worth mentioning here that a great part of the data points in Table S1 which have molecular weight ≤ 300 g/mol 83 points out of 240 (35%), is taken from Riazi's work [27]. Therefore, the highest accuracy of the correlation of Riazi and Daubert MW ≤ 300 g/mol may be attributed to this fact. It is also worth noting here that the accuracy of prediction of the lower molecular weight petroleum fluids (MW ≤ 300 g/mol) is understandably higher than that of the higher molecular weight petroleum fluids (MW ≥ 300 g/mol), because the heavy oil molecular weight is measured with a lower exactitude. Nevertheless, the correlations developed on the basis of extended molecular weight range also exhibit a satisfactory prediction for the lower molecular weight petroleum fluids (see, for example, the data for the ANN model, the new correlation, and Riazi and Daubert MW ≤ 700 g/mol from Table 3).

4. Discussion

Artificial neural networks are a mathematical model inspired by biological neural networks [49]. They are a class of artificial intelligence algorithms addressing different aspects or elements of learning, such as how to learn, how to induce, and how to deduce [50]. In our study, a classical three-layered neural network, as depicted in Figure 10, was availed.

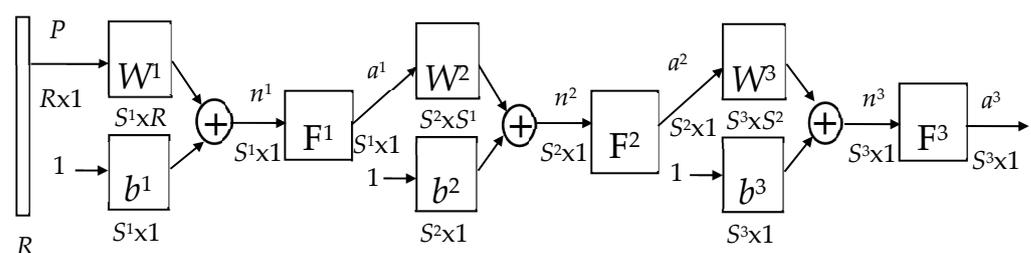


Figure 10. Flow chart of the multilayer neural network.

where

- P is an entry network's vector;
- a_m is the exit of the m-th layer of the neural network, where;
- w_m is a matrix of the coefficients of all inputs;
- b_m is neuron's input bias;
- F_m is the transfer function of the m-th layer exit.

In multilayer networks, the outputs of the previous (or first) layer become inputs to the next. In tutored learning (supervised learning), the neural network must achieve a result that is labeled as a goal and is predefined. Weighting factors are accordingly calculated to target values. After training, the neural network is tested—only input signals are given, without the one to be received. The main algorithm for training neural networks with a teacher is back-propagation. It is designed for multilayer networks with direct transfer.

The ANN molecular weight model developed in this research, as shown in the data in Figure 5b and Tables 2 and 3, distinguishes with the highest accuracy of prediction among all investigated empirical correlations. These findings support the opinion of Hadavimoghaddam et al. [51] that the artificial intelligence models provide the lowest average absolute relative error of petroleum property prediction. The use of Hossenifar and Shahverdi [25] method to generate seven additional inputs for the ANN showed again that it is a successful approach to model not only viscosity, as reported in our recent research [26], but also molecular weight of petroleum fluids.

The data in Table 2 and in Figures 5–9 and Figure 11 indicates that the new empirical correlation (Equation (26)) demonstrates the best accuracy among the other empirical correlations available in the literature when the whole data base of 430 points is concerned. Then the correlations of Goosens (Equation (7)), Riazi and Daubert (Equation (9)), and that of API (Equation (11)) distinguish with better accuracy in molecular weight prediction than the remaining seven correlations. However, the new correlation (Equation (26)) and those of Goosens and Riazi and Daubert have been developed based on a great amount of data in Table S1. That is why their comparison with data not included in their development may give a better indication of their molecular weight prediction ability. This was performed with 124 data points, and the statistical analysis of this comparison is summarized in Table 4.

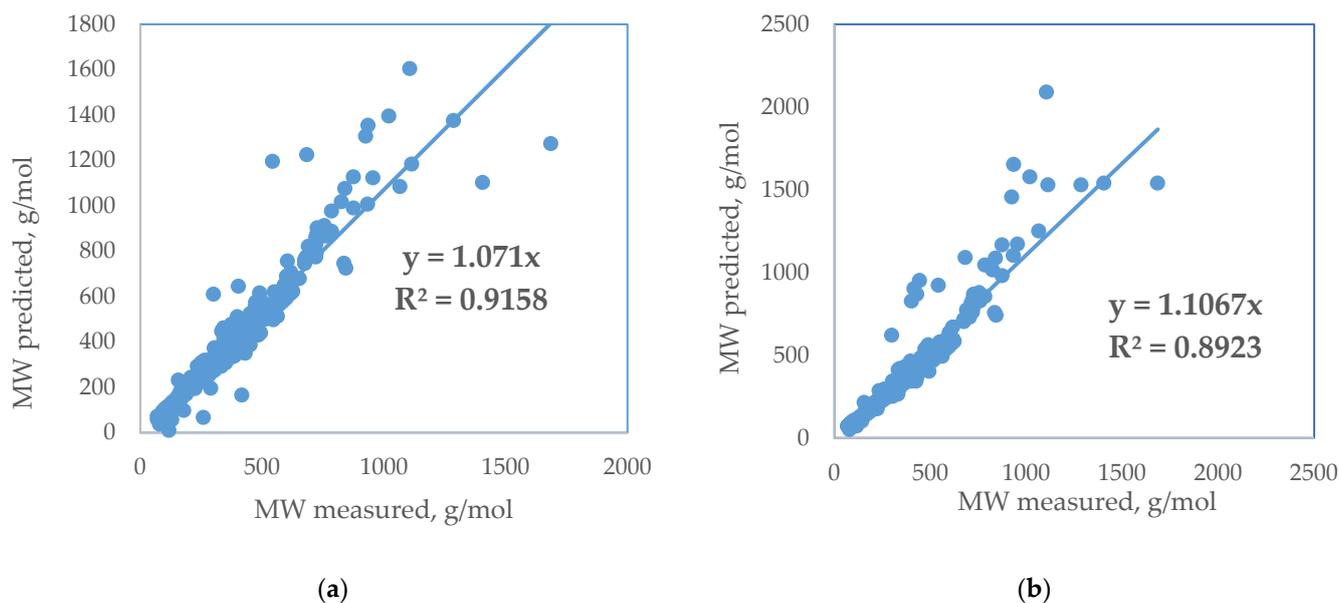


Figure 11. Parity graph for predicted versus measured molecular weight by Equation (6) (Soreide) (a) and by Equation (12) (Lemus et al.) (b).

Table 4. Statistical analysis of the new empirical correlation developed in this work, and of the correlations of Goosens, Riazi and Daubert, and API with 124 data points not used in their development.

	Standard Error	Rel. St. Error	%AAD
New empirical correlation (this work, test data)	42.1	11.6	8.8
Goosens (1996) [9]	56.7	15.6	12.1
Riazi and Daubert MW \leq 700 g/mol (2005) [18]	52.7	14.5	8.7
API (2011) [23]	47.1	13.0	9.0

The data in Table 4 again confirms the best performance of Equation (20) among the empirical correlations. It deserves mentioning here that 52 data points (42%) (the data points with numbers from 379 to 430 from Table S1) of all tested 124 data points (the data points with numbers from 307 to 430 from Table S1) were reported to be measured with repeatability of 15% [1]. This could explain the higher inaccuracy of molecular weight prediction for this test data set compared with that observed for the whole 430 data points and shown in Table 2.

5. Conclusions

The measurement of molecular weight of heavy oils is a difficult task associated with an error of about 15%. Thus, the correlations which satisfactorily predict molecular weight of various petroleum fluids can be an appropriate substitute for measurement. Among the correlations existing in the literature examined in this work, three distinguished with a better precision of prediction: Goosens, Riazi and Daubert, and API. The new correlation developed in this work using a nonlinear regression technique demonstrated a higher accuracy of molecular weight prediction than that of the three correlations mentioned above. However, the molecular weight prediction by the ANN three-layered neural network with a back-propagation teacher model exhibited much better exactitude than the new nonlinear regression correlation. This observation confirms again that the ANN method can predict petroleum properties with a higher accuracy than the regression correlations. In opposite of some statements in the literature that the other correlations developed after those of Lee–Kesler and Twu are not superior, the results of this research showed that the correlations of Goosens, Riazi and Daubert, API, and the new correlation predicted the molecular weight with a higher accuracy.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/pr11020426/s1>, Table S1: 430 data points for boiling point, specific gravity, and molecular weight of petroleum fluids, and individual hydrocarbons to be used for nonlinear regression modeling; Table S2: 430 data points for boiling point, specific gravity, Kw-factor, and TBP distillation characteristics and molecular weight of petroleum fluids, and individual hydrocarbons to be used for ANN modeling.

Author Contributions: Conceptualization, D.S.; methodology, E.S.; software, S.N. and S.S. (Sotir Sotirov); validation, S.S. (Stanislav Simeonov), R.D. and D.Y.; formal analysis, G.N.P.; investigation, I.V.K. and S.V.; data curation, I.S.; writing, D.S.; supervision, K.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Asen Zlatarov University—Burgas, Project: Center of Excellence UNITE BG05M2OP001-1.001-0004 /28.02.2018 (2018–2023).

Conflicts of Interest: The authors declare no conflict of interest.

Nomenclature

ANN	Artificial neural network
%AAD	% Average absolute deviation
E	Error
Kw	Kw-characterization factor of petroleum fluids
MW	Molecular weight
RSE	Relative standard error
SE	Standard error
SRE	Sum of relative errors
SSE	Sum of squared errors
RSE	Dichloromethane
Tb	Boiling point

References

- Lemus, M.C.S. Extended Distillation and Property Correlations for Heavy Oil. Ph.D. Thesis, University of Calgary, Calgary, AB, Canada, December 2015.
- Lemus, M.C.S.; Schoeggl, F.; Taylor, S.D.; Yarranton, H.W. Physical properties of heavy oil distillation cuts. *Fuel* **2016**, *180*, 457–472. [CrossRef]
- Nji, G.N. Characterization of heavy oils and bitumens. Ph.D. Thesis, University of Calgary, Calgary, AB, Canada, January 2010.
- Al-Mhanna, N.M. Simulation of High Pressure Separator Used in Crude Oil Processing. *Processes* **2018**, *6*, 219. [CrossRef]
- Aladwani, H.A.; Riazi, M.R. Some guidelines for choosing a characterization method for petroleum fractions in process simulators. *Trans IChemE Part A Chem. Eng. Res. Des.* **2005**, *83*, 160–166. [CrossRef]
- Schneider, D.F. Select the Right Hydrocarbon Molecular Weight Correlation. Available online: <https://www.stratusengr.com/Articles/MoleWt.pdf> (accessed on 17 November 2022).
- Powers, D.P.; Sadeghi, H.; Yarranton, H.W.; van den Berg, F.G.A. Regular solution based approach to modeling asphaltene precipitation from native and reacted oils: Part 1, molecular weight, density, and solubility parameter distributions of asphaltenes. *Fuel* **2016**, *178*, 218–233. [CrossRef]
- Yarranton, H.W.; Powers, D.P.; Okafor, J.C.; van den Berg, F.G.A. Regular solution based approach to modeling asphaltene precipitation from native and reacted oils: Part 2, molecular weight, density, and solubility parameter of saturates, aromatics, and resins. *Fuel* **2018**, *215*, 766–777. [CrossRef]
- Goosens, A.G. Prediction of molecular weight of petroleum fractions. *Ind. Eng. Chem. Res.* **1996**, *35*, 985–988. [CrossRef]
- Hosseinfar, P.; Shahverdi, H. A predictive method for constructing the distillation curve of petroleum fluids using their physical bulk properties. *J. Petrol. Sci. Eng.* **2021**, *200*, 108403. [CrossRef]
- Altgelt, K.H.; Boduszynski, M.M. *Composition and Analysis of Heavy Petroleum Fractions*; Marcel Dekker: New York, NY, USA, 1994; pp. 1–495.
- Soreide, I. Improved Phase Behavior Predictions of Petroleum Reservoir Fluids from a Cubic Equation of State. Ph.D. Thesis, Norwegian Institute of Technology, Department of Petroleum Technology and Applied Geophysics, Trondheim, Norway, 1989.
- Lee, B.I.; Kesler, M.G. A Generalized thermodynamic correlation based on the three-parameter corresponding states. *AIChE J.* **1975**, *21*, 510–527. [CrossRef]
- Kesler, M.G.; Lee, B.I. Improve prediction of enthalpy of fractions. *Hydrocarb. Process* **1976**, *55*, 153–158.
- Riazi, M.R.; Daubert, T.E. Simplify property predictions. *Hydrocarb. Process* **1980**, *59*, 115–116.
- Riazi, M.R.; Daubert, T.E. Analytical correlations interconvert distillation-curve types. *Oil Gas J.* **1986**, *84*, 50–57.
- Riazi, M.R.; Daubert, T.E. Characterization parameters for petroleum fractions. *Ind. Eng. Chem. Res.* **1987**, *26*, 755–759. [CrossRef]
- Riazi, M.R. *Characterization and Properties of Petroleum Fractions, 1st ed*; ASTM International: West Conshohocken, PA, USA, 2005; pp. 1–407.
- Rao, V.K.; Bardon, M.F. Estimating the molecular weight of petroleum fractions. *Ind. Eng. Chem. Proc. Des. Dev.* **1985**, *24*, 498. [CrossRef]
- Twu, C.H. An internally consistent correlation for predicting the critical properties and molecular weight of petroleum and coal-tar liquids. *Fluid Phase Equilibria* **1984**, *16*, 137–150. [CrossRef]
- Hariu, O.H.; Sage, R.C. Crude split figured by computer. *Hydrocarb. Process* **1969**, *4*, 143–148.
- Katz, D.L.; Firoozabadi, B. Predicting phase behavior of condensate/crude oil systems using methane interaction coefficients. *J. Pet. Tech.* **1978**, *228*, 1649–1655. [CrossRef]
- Liñan, L.Z.; Lima, N.M.N.; Maciel, M.R.W.; Filho, R.M.; Medina, L.C.; Embiruçu, M. Correlation for predicting the molecular weight of Brazilian petroleum residues and cuts: An application for the simulation of a molecular distillation process. *J. Pet. Sci. Eng.* **2011**, *78*, 78–85. [CrossRef]
- Liu, Y.A.; Chang, A.-F.; Kiran, P. Chapter 1: Characterization, physical and thermodynamic properties of oil fractions. In *Petroleum Refinery Process Modeling: Integrated Optimization Tools and Applications*; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2018; pp. 1–58.

25. Hosseinifar, P.; Shahverdi, H. Prediction of the ASTM and TBP distillation curves and specific gravity distribution curve for fuels and petroleum fluids. *Can. J. Chem. Eng.* **2022**, *100*, 3288–3310. [[CrossRef](#)]
26. Stratiev, D.; Shishkova, I.; Dinkov, R.; Nenov, S.; Sotirov, S.; Sotirova, E.; Kolev, I.; Ivanov, V.; Ribagin, S.; Atanassov, K.; et al. Prediction of petroleum viscosity from molecular weight and density. *Fuel* **2023**, *331*, 125679. [[CrossRef](#)]
27. Riazi, M.R.; Daubert, T.E. Prediction of molecular-type analysis of petroleum fractions and coal liquids. *Ind. Eng. Chem. Res.* **1986**, *25*, 1009–1015. [[CrossRef](#)]
28. Riazi, M.R.; Daubert, T.E. Improved characterization of wide boiling range undefined petroleum fractions. *Ind. Eng. Chem. Res.* **1987**, *26*, 629–632. [[CrossRef](#)]
29. White, C.M.; Perry, M.B.; Schmidt, C.E.; Douglas, L.J. Relationship between refractive indices and other properties of coal hydrogenation distillates. *Energy Fuels* **1987**, *1*, 99–105. [[CrossRef](#)]
30. Bollas, G.M.; Vasalos, I.A.; Lappas, A.A.; Iatridis, D.K.; Tsioni, G.K. Bulk molecular characterization approach for the simulation of FCC feedstocks. *Ind. Eng. Chem. Res.* **2004**, *43*, 3270–3281. [[CrossRef](#)]
31. Vargas, F.M.; Chapman, W.G. Application of the one-third rule in hydrocarbon and crude oil systems. *Fluid Phase Equilib.* **2010**, *290*, 103–108. [[CrossRef](#)]
32. Yarranton, H.W.; Okafor, J.C.; Ortiz, D.P.; van den Berg, F.G.A. Density and refractive index of petroleum, cuts, and mixtures. *Energy Fuels* **2015**, *29*, 5723–5736. [[CrossRef](#)]
33. Carbognani, L.; Díaz-Gómez, L.; Oldenburg, T.B.P.; Pereira-Almao, P. Determination of molecular masses for petroleum distillates by simulated distillation. *Cienc. Tecnol. Futuro* **2012**, *4*, 43–55.
34. Wang, S.; Dong, X.; Sun, R. Predicting saturates of sour vacuum gas oil using artificial neural networks and genetic algorithms. *Expert Syst. Appl.* **2010**, *37*, 4768–4771.
35. Wang, G.; Liu, Y.; Wang, X.; Xu, C.; Gao, J. Studies on the catalytic cracking performance of coker gas oil. *Energy Fuels* **2009**, *23*, 1942–1949. [[CrossRef](#)]
36. Pitault, I.; Nevicato, D.; Forissier, M.; Bernard, J.-R. Kinetic model based on a molecular description for catalytic cracking of vacuum gas oil. *Chem. Eng. Sci.* **1994**, *49*, 4249–4262. [[CrossRef](#)]
37. Sheng, Q.; Wang, G.; Duan, M.; Ren, A.; Yao, L.; Hu, M.; Gao, J. Determination of the hydrogen-donating ability of industrial distillate narrow fractions. *Energy Fuels* **2016**, *30*, 10314–10321. [[CrossRef](#)]
38. Altgelt, K.H.; Boduszynski, M.M. Composition of heavy petroleum. 3. An improved boiling point-molecular weight relation. *Energy Fuels* **1992**, *6*, 68–72. [[CrossRef](#)]
39. Dominguez, M. FCC feed fractionation. *Pet. Coal* **2003**, *45*, 113–118.
40. Van Camp, C.E.; Van Damme, P.S.; Froment, G.F. Thermal cracking of kerosene. *Ind. Eng. Chem. Process Des. Dev.* **1984**, *23*, 155–162. [[CrossRef](#)]
41. Nace, D.M.; Voltz, S.E.; Weekman, V.W., Jr. Application of a kinetic model for catalytic cracking. Effects of charge stocks. *Ind. Eng. Chem. Process Des. Develop.* **1971**, *10*, 530–538. [[CrossRef](#)]
42. Zhang, Y.; Schuler, B.; Fatayer, S.; Gross, L.; Harper, M.R.; Kushnerick, J.D. Understanding the effects of sample preparation on the chemical structures of petroleum imaged with non-contact atomic force microscopy. *Ind. Eng. Chem. Res.* **2018**, *57*, 15935–15941. [[CrossRef](#)]
43. Stratiev, D.; Nenov, S.; Nedanovski, D.; Shishkova, I.; Dinkov, R.; Stratiev, D.D.; Stratiev, D.D.; Sotirov, S.; Sotirova, E.; Atanassova, V.; et al. Different Nonlinear Regression Techniques and Sensitivity Analysis as Tools to Optimize Oil Viscosity Modeling. *Resources* **2021**, *10*, 99. [[CrossRef](#)]
44. Stratiev, D.; Nenov, S.; Sotirov, S.; Shishkova, I.; Palichev, G.; Sotirova, E.; Ivanov, V.; Atanassov, K.; Ribagin, S.; Angelova, N. Petroleum viscosity modeling using least squares and ANN methods. *J. Pet. Sci. Eng.* **2022**, *212*, 110306. [[CrossRef](#)]
45. Sinha, U.; Dindoruk, B.; Soliman, M. Machine learning augmented dead oil viscosity model for all oil types. *J. Pet. Sci. Eng.* **2020**, *195*, 107603. [[CrossRef](#)]
46. Sinha, U.; Dindoruk, B.; Soliman, M.Y. Physics augmented correlations and machine learning methods to accurately calculate dead oil viscosity based on the available inputs. *SPE J.* **2022**, *27*, 3240–3253. [[CrossRef](#)]
47. Stratiev, D.; Marinov, I.; Dinkov, R.; Shishkova, I.; Velkov, I.; Sharafutdinov, I.; Nenov, S.; Tsvetkov, T.; Sotirov, S.; Mitkova, M.; et al. Opportunity to improve diesel fuel cetane number prediction from easy available physical properties and application of the least squares method and the artificial neural networks. *Energy Fuels* **2015**, *29*, 1520–1533. [[CrossRef](#)]
48. Shishkova, I.; Stratiev, D.; Kolev, I.V.; Nenov, S.; Nedanovski, D.; Atanassov, K.; Ivanov, V.; Ribagin, S. Challenges in Petroleum Characterization—A Review. *Energies* **2022**, *15*, 7765. [[CrossRef](#)]
49. D’Addona, D.M. Neural Network. In *CIRP Encyclopedia of Production Engineering*; Laperrière, L., Reinhart, G., Eds.; Springer: Berlin/Heidelberg, Germany, 2014. [[CrossRef](#)]
50. Yang, Z.R.; Yang, Z. 6.01—Artificial Neural Networks. *Compr. Biomed. Phys.* **2014**, *6*, 1–17. [[CrossRef](#)]
51. Hadavimoghaddam, F.; Ostadhassan, M.; Heidaryan, E.; Sadri, M.A.; Chapanova, I.; Popov, E.; Cheremisn, A.; Rafieepour, S. Prediction of dead oil viscosity: Machine learning vs. classical correlations. *Energies* **2021**, *14*, 930. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.