Advances and Applications in Mathematical Sciences
Volume 20, Issue 10, August 2021, Pages 2299-2311

MODELING PHYSICO-CHEMICAL PROPERTIES OF BENZENOID HYDROCARBONS USING TOPOLOGICAL INDICES OF MOLECULAR GRAPHS

RAJU B. JUMMANNAVER, SHASHANK D. SHINDHE and BHARGAVI DESHPANDE

Department of Mathematics
Karnatak College
Dharwad-580003, India
E-mail: rajesh.rbj065@gmail.com

Department of Statistics
Karnatak University
Dharwad-580003, India
E-mail: shashank.shindhe@yahoo.com

Department of Mathematics and Statistics
M. S. Ramaiah University of Applied Sciences
Bangalore-560058, India
E-mail: bhargavi.mdeshpande@gmail.com

Abstract

Predictive potential of Randić index, Sum-connectivity index and Harmonic index is studied. These topological indices are utilized as parameters while predicting the molecular weights (in g/mol), densities (in $^{\circ}C$), boiling points (in $^{\circ}C$) and melting points (in $^{\circ}C$) of Benzenoid hydrocarbons. It is shown that these indices are good predictors of the molecular weight (in g/mol), and boiling point (in $^{\circ}C$). However thy are not sufficient to predict melting point (in $^{\circ}C$) and density (in g/cm^3), when used alone. Usefulness of additional information about the molecular structure of Benzenoid hydrocarbons in improving the performance of the molecular descriptors as predictors is also discussed.

2010 Mathematics Subject Classification: Primary 05C76; Secondary 05C07, 92E10.
Keywords: Randic index, Harmonic index, Sum-connectivity Index, Benzenoid hydrocarbons, Correlation, Regression.
Received March 30, 2021; Accepted April 21, 2021

1. Introduction

It is known that, among a family of chemical compounds, especially in case of organic compounds there seem to be strong correlations between the structures of the compounds and the properties they possess. For instance, in case of alkanes the boiling point increases with increasing number of carbons. These correlations between the structure of a compound and the properties it possesses are modeled as Quantitative Structure-Property Relationships (QSPR) or Quantitative Structure-Activity Relationships (QSAR) models.

QSAR/QSPR models are basically regression models wherein, a response variable (property of a compound, e.g. boiling point of the alkane) is modeled as a function of predictor variable(s) (structural information, e.g. number of carbons in an alkane). Topological indices of molecular graphs are the most frequently used predictors in QSAR/QSPR models since they capture the structural essence of a molecule while being computationally inexpensive as evidenced in [1], [2], [12] and [14].

A molecular graph G of a compound contains a finite nonempty set of n vertices (number of atoms in a compound, e.g. number of carbons in an alkane) called the vertex set, denoted by $V(G)$, along with a pre-defined set of m unordered pairs of different vertices of $V(G)$ called the edge set, denoted by $E(G)$. Such each pair of vertices in $E(G)$ is termed as an edge (a chemical bond between two atoms) of G. Whenever two vertices shares a common edge, those vertices are said to be adjacent and, such an edge is said to be incident on those vertices. If two edges have a common incident vertex then those edges are said to be adjacent. The degree of a vertex u in G is the number of edges incident to it and is denoted by $d_G(u)$. For graph theoretical definitions and notations, we follow [5]. All graphs taken into consideration here are simple, finite and undirected graphs.

The Randić index due to [11], denoted by $R(G)$, also known as product connectivity index is defined as

$$R(G) = \sum_{uv \in E(G)} \frac{1}{\sqrt{d_G(u)\, d_G(v)}}. \tag{1}$$

According to [4], [6], [7], [9] and [13], Randić index is the most successful molecular descriptor and finds applications in multitude of scenarios.

Another variant of Randić index, called the Harmonic index which appeared for the first time in [3] is defined as

$$H(G) = \sum_{uv \in E(G)} \frac{2}{\sqrt{d_G(u) + d_G(v)}}. \tag{2}$$

One other molecular descriptor that is similar to Randić index in its predictive prowess, as shown in [8] is the sum-connectivity index originally defined in 2009 and given by

$$SCI(G) = \sum_{uv \in E(G)} \frac{1}{\sqrt{d_G(u) + d_G(v)}}. \tag{3}$$

Persuaded by work in [10], in this paper we explore the predictive abilities of Randić index, Sun-connectivity index and Harmonic index in predicting the physico-chemical properties of Benzenoid hydrocarbons. For computations, we consider the indices of 29 benzenoid hydrocarbons given in [10] and record their molecular weight (MW in gm/mol), density (ρ_B in gm/cm^3), boiling point (T_B in $^\circ C$) and melting point (T_M in $^\circ C$) from selected sources [15]. Tools of R programming language ([16], [17], [18]) are employed to carry out the analysis. We explore the relationship between the indices and the physico-chemical properties of the Benzenoid hydrocarbons in section 2 and build the predictive models based on the correlation analysis in section 3 using R-programming language. Section 4 deals with findings and conclusions.

2. Correlation Analysis

Correlation analysis is an important step in modeling a response variable as a function of predictor(s). It provides the necessary information regarding the viability of the relationship between the response(s) and predictor(s). The correlation between two variables is contemplated either graphically through scatter plot or quantified as a coefficient (R) and its value ranges from -1 (negatively correlated) to +1 (positively correlated). A positive R implies variability of the two variables in the same direction, a negative R is an

indication of variability in opposite directions and $R = 0$ signifies no relationship between the variables.

Figure 1 presents the relationship between the $R(G)$ and the properties of Benzenoid hydrocarbons under study. According to scatter plots, the Randić index positively correlates with all the four properties considered. However, the relationship with the *MW* seems to be stronger as compared to other properties. The Pearson correlation coefficients (R) of $R(G)$ with T_B, T_M, ρ_B and *MW* are 0.9047 $(p - \text{value} = 1.64 \times 10^{-11})$, 0.6120 $(p - \text{value} = 0.0004)$, 0.7531 $(p - \text{value} = 2.43 \times 10^{-06})$ and 0.9998 $(p - \text{value} = 2.2 \times 10^{-16})$.

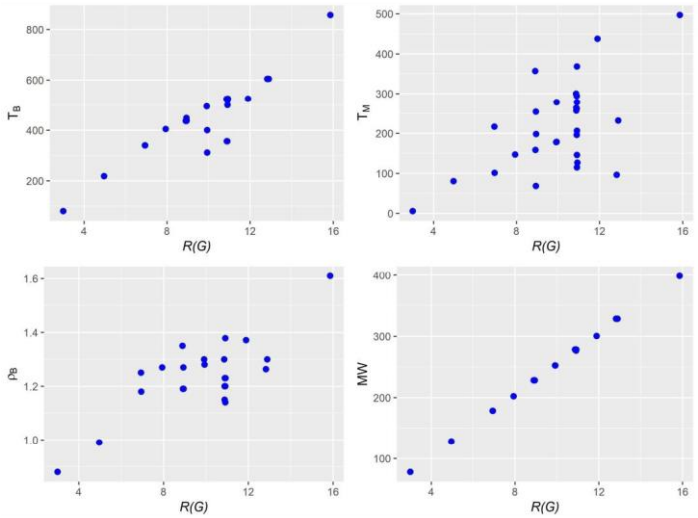

Figure 1. Scatter plot of $R(G)$ against T_B, T_M, ρ_B and *MW*.

A similar relationship can be observed between the harmonic index and the properties of hydrocarbons. Based on the scatter plots in figure 2, it is deduced that, the relationship between $H(G)$ and T_B, T_M, ρ_B and *MW* are positive in nature with respective correlation coefficients computed to be equal to 0.9047 $(p - \text{value} = 1.63 \times 10^{-11})$, 0.6135 $(p - \text{value} = 0.0004)$, 0.7543 $(p - \text{value} = 2.28 \times 10^{-06})$ and 0.9995 $(p - \text{value} = 2.2 \times 10^{-16})$.

According to figure 3, the relationship between the sum-connectivity

index and the properties of hydrocarbons is positive in nature with correlation coefficients being 0.9067 $(p = 1.25 \times 10^{-11})$, 0.6259 $(p = 0.0002)$, 0.7686 $(p = 1.12 \times 10^{-06})$ and 0.9982 (2.202×10^{-16}) respectively for T_B, T_M, ρ_B and MW.

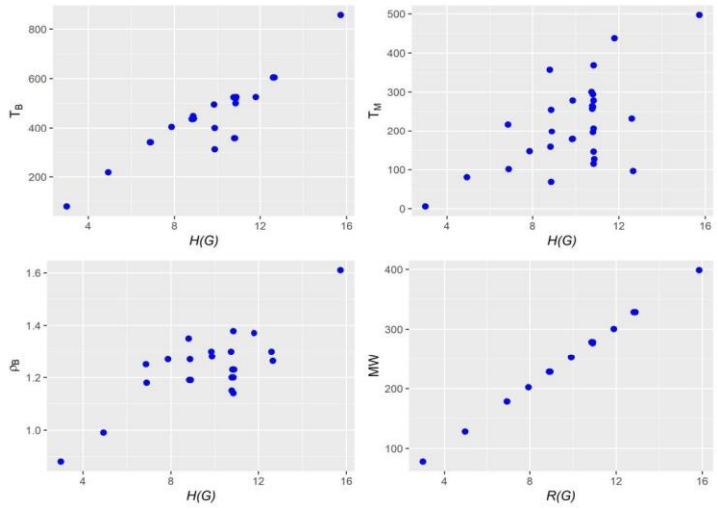

Figure 2. Scatter plot of $H(G)$ against T_B, T_M, ρ_B and MW.

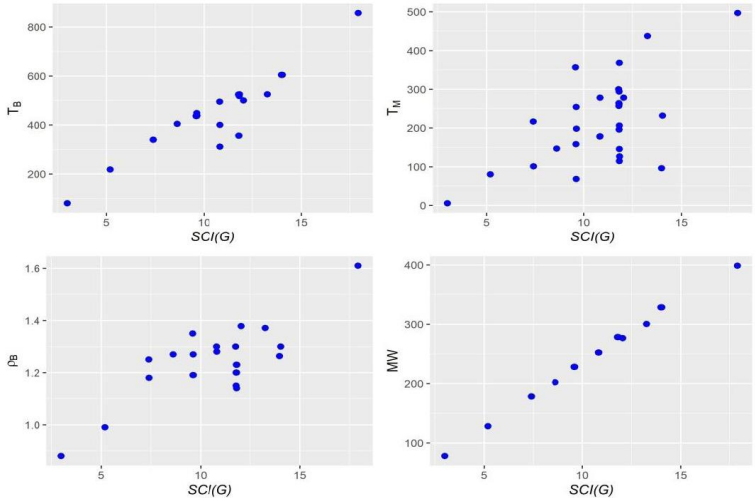

Figure 3. Scatter plot of SCI (G) against T_B, T_M, ρ_B and MW.

3. Predicting Properties of Benzenoid Hydrocarbons using Molecular Descriptors

After establishing the relationship between the descriptors and the properties of the compounds under consideration, a linear relation of the form

$$Y = b_0 + b_1 X \tag{4}$$

can be obtained so that Y, an unknown response can be estimated from a given X, a descriptor such that b_0 and b_1 known as the regression coefficients account for the relation between X and Y. Once the regression coefficients are computed by the method of least squares, the models are compiled by substituting them in equation (4).

The models to estimate the T_B of Benzenoid hydrocarbons from $R(G)$, $H(G)$ and $SCI\ (G)$ are respectively

$$T_B = -47 + 51\ R(G) \tag{5}$$

$$T_B = -49 + 52\ H(G) \tag{6}$$

and

$$T_B = -24 + 45\ SCI\ (G). \tag{7}$$

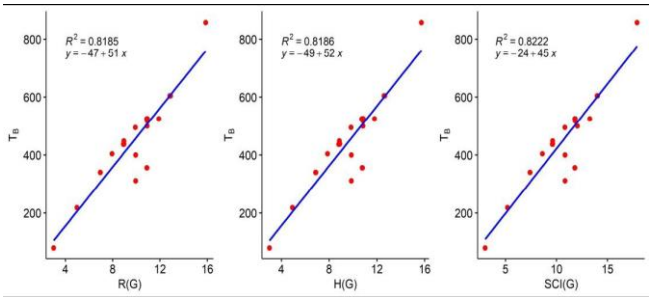

Figure 4. Regression lines for predicting T_B from $R(G)$, $H(G)$ and $SCI\ (G)$.

The regression lines fitted to the data are presented in figure (4). The coefficients of determination (R^2) for these models are respectively 0.8185, 0.8186 and 0.8222. The R^2 is the amount of variance in response variable that is predicted from the predictors. In other words, the coefficient of

determination measures the efficiency with which the descriptors estimate the responses. Another way of representing the efficiency of a model is through the fraction of variance unexplained (FVU), the FVUs in T_B that ware not explained by $R(G)$, $H(G)$ and $SCI\ (G)$ are 0.1815, 0.1814 and 0.1778 respectively. This implies that, all the three molecular descriptors considered are similar to one another as far as their potential in predicting the T_B is concerned.

Next, the predictive prowess of the descriptors in predicting the T_M of the hydrocarbons is considered. The models for estimating the T_M from the Randić, harmonic and sum-connectivity indices are respectively

$$T_M = -60 + 28\ R(G) \tag{7}$$

$$T_M = -62 + 28\ H(G) \tag{8}$$

and

$$T_M = -53 + 25\ SCI\ (G). \tag{9}$$

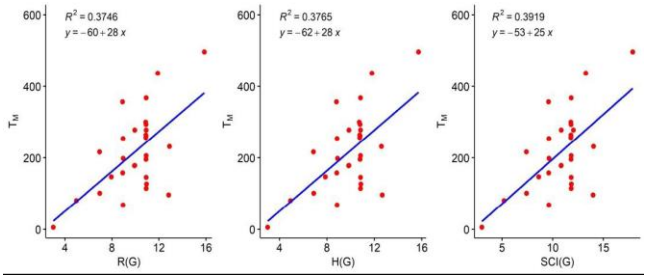

Figure 5. Regression lines for predicting T_M from $R(G)$, $H(G)$ and $SCI\ (G)$.

Figure 5 contains the regression lines. The R^2 for models (7), (8) and (9) are respectively 0.3746, 0.3765 and 0.3919. Since, the model (9) has the highest R^2 among the three models we may say that it is the best out of the three. Also, it has a FVU of 0.6081 which is the lowest among others.

Similarly, we obtain the models for estimating ρ_B based on the molecular descriptors as given in (10), (11) and (12) with coefficients of determinations equaled at 0.5672, 0.5691 and 0.5908 respectively. These models are presented in figure 6. The FUVs for these models are 0.4328, 0.4309 and

0.4092 respectively. Hence, the model (15) which corresponds to the sum connectivity index is the best descriptor of density among Benzenoid hydrocarbons with respect to R^2 and FUV.

$$\rho_B = 0.86 + 0.038\ R(G) \tag{10}$$

$$\rho_B = 0.86 + 0.039\ H(G) \tag{11}$$

$$\rho_B = 0.87 + 0.034\ SCI\ (G) \tag{12}$$

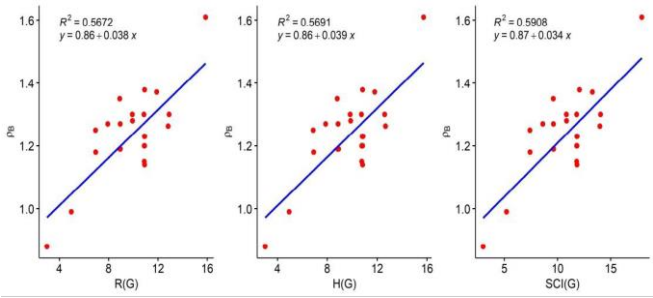

Figure 6. Regression lines for predicting ρ_B from $R(G)$, $H(G)$ and $SCI\ (G)$.

The regression equations for predicting the molecular weights of Benzenoid hydrocarbons are given in equations (13), (14) and (15). The regression lines fitted to the data are displayed in figure. The R^2 and FUV for these models are 0.9996, 0.9991, 0.9966 and 0.0004, 0.0009, 0.0034 respectively. All these models are a good fit for estimating the molecular weights of hydrocarbons under consideration since there is a small proportion of variance which is unexplained.

$$MW = 4.1 + 25\ R(G) \tag{13}$$

$$MW = 2.8 + 25\ H(G) \tag{14}$$

$$MW = 16 + 22\ SCI\ (G) \tag{15}$$

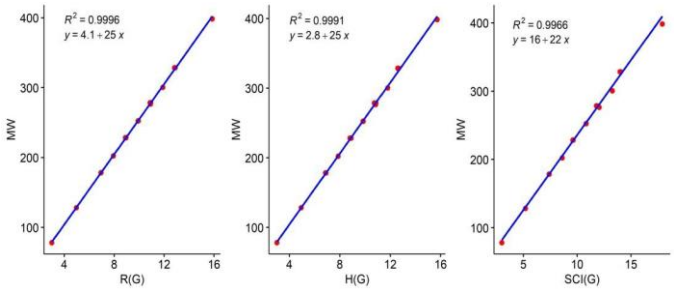

Figure 7. Regression lines for predicting MW from $R(G)$, $H(G)$ and SCI (G).

It can be seen from these results that, the models for estimating T_M and ρ_B are underperforming with respect to R^2 and FUV. Hence, in these cases the molecular descriptors alone are not sufficient to predict the properties. However, these models can be improved by considering additional information pertaining to the molecular structure of Benzenoid hydrocarbons.

4. Corrected Models for Predicting T_B and ρ_B

The molecular graphs of Benzenoid hydrocarbons consists of hexagons, therefore, there are only two types of vertices present in their structure. The graph of these hydrocarbons has vertices of degree 2 and vertices of degree 3. This implies that, there are three types edges present in their molecular structures, namely e_{22}, e_{23} and e_{33}. The e_{22} kind of edges has vertices of degree 2 at both ends. The e_{23} kind of edges has a vertex of degree 2 at one end and a vertex of degree 3at the other end. The e_{33} kind of edges has vertices of degree 3 at both ends.

Suppose that, a Benzenoid graph has n vertices, h hexagons and r inlets, then the number of these edges can be obtained as given in [8]. They are

$$e_{22} = n - 2h - r + 2$$

$$e_{23} = 2r$$

and

$$e_{33} = 3h - r - 3.$$

Models for predicting T_M and ρ_B based on this information about the structure of hydrocarbons can be obtained such that $c_0, c_1, \ldots$ are regression coefficients associated with each of the covariates and they are of the following forms.

$$T_M \text{ or } \rho_B \approx c_0 R(G) + c_1 e_{22} + c_2 e_{23} + c_3 e_{33} \tag{16}$$

$$T_M \text{ or } \rho_B \approx c_0 H(G) + c_1 e_{22} + c_2 e_{23} + c_3 e_{33} \tag{17}$$

$$T_M \text{ or } \rho_B \approx c_0 SCI(G) + c_1 e_{22} + c_2 e_{23} + c_3 e_{33}. \tag{18}$$

The regression coefficients in these models are estimated using least squares method and are back substituted in (16), (17) and (18) to obtain following models for T_M

$$T_M = 173.5 + 2904.5 R(G) - 1470.7 e_{22} - 1178.8 e_{23} - 949.9 e_{33};\ R^2 = 0.5315 \tag{19}$$

$$T_M = 165.95 + 60 H(G) - 48.13 e_{22} - 16.86 e_{23} - 2.31 e_{33};\ R^2 = 0.5254 \tag{20}$$

$$T_M = 173.7 + 3595.6 SCI(G) - 1816.2 e_{22} - 1601.1 e_{23} - 1449.6 e_{33};\ R^2 = 0.5317 \tag{21}$$

and for ρ_B as

$$\rho_B = 1.1183 - 3.0427 R(G) + 1.5044 e_{22} + 1.2516 e_{23} + 1.0382 e_{33};\ R^2 = 0.7426 \tag{22}$$

$$\rho_B = 1.1354 - 0.1673 H(G) + 0.0650 e_{22} + 0.0758 e_{23} + 0.0801 e_{33};\ R^2 = 0.7390 \tag{23}$$

$$\rho_B = 1.1183 - 3.5697 SCI(G) + 1.7679 e_{22} + 1.6058 e_{23} + 1.4813 e_{33};\ R^2 = 0.7422. \tag{24}$$

The comparison between the original and corrected models is given in table 1. It can be seen that adding more structural information into the model improves their predictive potential. In this particular study, using the number of different kinds of edges present in the molecular structure of the Benzenoid hydrocarbons results in at least 40% increase in the proportion of variance in the response that is explained by the predictors.

Table 1. Comparison of original to corrected models.

Response	Predictor	Original models		Corrected models		Percent change in R^2
		R^2	FUV	R^2	FVU	
T_M	$R(G)$	0.3746	0.6254	0.5315	0.4685	41.89%
	$H(G)$	0.3765	0.6235	0.5254	0.4746	39.54%
	$SCI\ (G)$	0.3919	0.6081	0.5317	0.4683	35.67%
	$R(G)$	0.5672	0.4328	0.7426	0.2574	30.92%
	$H(G)$	0.5691	0.4309	0.7390	0.2610	29.85%
T_M	$SCI\ (G)$	0.5908	0.4092	0.7422	0.2578	25.62%

5. Conclusions

The current study focused on exploring the predictive potential of $R(G), H(G)$ and $SCI\ (G)$ in estimating the T_B, T_M, ρ_B and MW found that the molecular descriptors alone are capable of predicting the physico-chemical properties considered except while predicting T_M. Based on the R^2 and FUV values, it is found that

- The models fitted using the three molecular descriptors perform equitably in predicting boiling point.
- For estimating melting point, the model fitted using sum-connectivity index emerges as the best.
- The model constructed using sum-connectivity index is the best descriptor of density.
- All the three descriptors have equal strengths to predict the molecular weight.
- Individually the descriptors considered are not sufficient to predict and single handedly.
- However, using additional information about the molecular structure of

the hydrocarbons results in significant improvement of the models while predicting T_M and ρ_B.

References

[1] J. C. Dearden, The use of topological indices in QSAR and QSPR modeling, In: K. Roy, (eds.) Advances in QSAR modeling-Applications in pharmaceutical, chemical, food, agricultural and environmental sciences, 57-88. Springer International Publishing, 2017.

[2] J. Devillers and A. T. Balaban, Topological indices and related descriptors in QSAR and QSPAR, In: J. Devillers, A. T. Balaban, (eds.) Topological indices and related descriptors in QSAR and QSPAR, CRC Press, London, 2000.

[3] S. Fajtlowicz, On conjectures of Graffiti-II, Congr. Number. 60 (1987), 187-197.

[4] R. Garcia-Domenech, J. G. Alvez, J. V. de Julian-Ortiz and L. Pogliani, Some new trends in chemical graph theory, Chemical Reviews 105 (2008), 1127-1169.

[5] F. Harary, Graph Theory, Addison-Wesley Publishing Company, Reading, 1969.

[6] L. B. Kier and L. H. Hall, Molecular connectivity in Chemistry and Drug research, Academic press, New York, 1976.

[7] L. B. Kier and L. H. Hall, Molecular connectivity in structure activity analysis, Wiley, New York, 1986.

[8] B. Lucic, N. Trinajstic and B. Zhou, Comparison between the sum-connectivity index and product-connectivity index for benzenoid hydrocarbons, Chemical Physical Letters, 475(1) (2009), 146-148.

[9] L. Pogliani, From molecular connectivity indices to semi empirical connectivity terms: Recent trends in graph theoretical descriptors, Chemical Reviews 100(10) (2000), 3827-3858.

[10] H. S. Ramane, V. B. Joshi, R. B. Jummannaver and S. D. Shindhe, Relationship between Randi index, Sum-connectivity index, Harmonic index and π-electron energy for Benzenoid hydrocarbons 42(6) (2019), 519-524.

[11] M. Randić, On characterization of molecular branching, Journal of the American Chemical society 97 (1975), 6609-6615.

[12] A. Talevi, C. L. Bellera, M. Di Ianni, P. R. Duchowicz, L. E. Bruno-Blanch and E. A. Castro, An integrated drug development approach applying topological descriptors 8(3) (2012), 172-181.

[13] R. Todeschini and V. Consonni, Handbook of molecular descriptors, Wiley-VCH Verlag GmbH, Germany, 2000.

[14] R. Zanni, M. Galvez-Llompart, R. Garcia-Domenech and J. Galvez, Latest advances in molecular topology applications for drug discovery, Expert Opinion on Drug Discovery 10 (2015), 945-957.

[15] Chem Spider, Search and share Chemistry: Available online at http://www.chemspider.com/. Last accessed: 22/10/2020.

[16] R. Core and R. Team, A language and environment for statistical computing, R. Foundation for statistical computing, Vienna, Austria. URL: https://www.R-project.org/, (2020).

[17] H. Wickham, ggplot2: Elegant Graphics for Data Analysis, Springer-Verlag, New York, 2016.

[18] A. Kassambara, ggpubr: 'ggplot 2' Based Publication Ready Plots, *R* package version 0.4.0. https://CRAN.R-project.org/package=ggpubr, (2020).