

Correlation Analysis of Experimental Data Applied in the Study of the Extraction Process

Mariana Rusu¹, Aliona Ghendov-Mosanu^{2*}, Rodica Sturza³

¹Department of Applied Informatics, Technical University of Moldova, Chisinau, Moldova

²Department of Food Technology, Technical University of Moldova, Chisinau, Moldova

³Department of Oenology and Chemistry, Technical University of Moldova, Chisinau, Moldova

Email: *aliona.mosanu@tpa.utm.md

How to cite this paper: Rusu, M., Ghendov-Mosanu, A. and Sturza, R. (2021) Correlation Analysis of Experimental Data Applied in the Study of the Extraction Process. *Journal of Applied Mathematics and Physics*, 9, 3019-3031.

<https://doi.org/10.4236/jamp.2021.912195>

Received: November 2, 2021

Accepted: December 7, 2021

Published: December 10, 2021

Copyright © 2021 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The article examines the application of correlation analysis of experimental data in research into the process of extracting bioactive compounds and antioxidant activity in plant extracts from berries and grape pomace. The correlation analysis of the experimental data allowed the establishment of the second order statistical characteristics (autocorrelation function, intercorrelation function and correlation coefficient). Based on the correlation analysis of the experimental data, it was shown that the influencing factor and the measured parameters have zero correlation coefficients for all types of researched extracts. This indicates that they are not independent. Therefore, related mathematical models can be deduced.

Keywords

Correlation Analysis, Autocorrelation Function, Correlation Coefficient, Parameter

1. Introduction

The correlation analysis uses second-order statistical characteristics (second-order probability density) and ensures the assessment of the correlation of the experimental data (whether independent or not and what is the nature of the dependence in the latter case) and the estimation of their nonlinear character (existence of a nonlinear component in experimental series) [1].

Although first-order statistical features that use first-order probability density are frequently used, they do not give a complete picture of the character of a deterministic series. Indeed, two experimental series may have the same mean and dispersion, but their character of variation may be different [2] [3].

The aim of this study is to apply the correlation analysis of experimental data

in the research of the extraction process of bioactive compounds and antioxidant activity in plant extracts from berries and grape marc.

2. Materials and Methods

2.1. Materials

In this study, for the correlation analysis were used experimental data from the extraction of bioactive compounds from chokeberry, rosehip, rowan, hawthorn, and grape marc. Both berries and grape marc are known for their rich content in bioactive compounds [4] [5] [6] [7] [8]. In the obtained hydroalcoholic extracts (ethyl alcohol) the content of biologically active compounds (polyphenol index and total anthocyanin content) and the antioxidant activity were determined by 3 different methods (**Table 1**). The photochemiluminescence test (PCL), the DPPH (2,2-diphenyl-1-picrylhydrazyl) test and the Hydrogen Peroxide Scavenging Activity (HPSA) were applied. Thus, 8 parameters were researched depending on the type of hydroalcoholic extract investigated. 5 concentrations of ethyl alcohol (20%, 40%, 50%, 60%, 80% (v/v)) were applied to obtain extracts of chokeberry, rosehip, rowan, and hawthorn fruit and 6 concentrations (20%, 40%, 50%, 60%, 80%, 96% (v/v)) for grape marc.

The parameters measured according to the type of hydroalcoholic extract are presented in **Table 2**.

Table 1. Notify the parameter description, unit, and parameter cod.

Parameter description	Unit	Parameter code
Antioxidant activity of water-soluble substances	μmol AAE/3g	P1
Antioxidant activity of fat-soluble substances	μmol TE/3g	P2
Total anthocyanin content	mg ME/3g	P3
Total polyphenol index	c. u.	P4
Antiradical activity, DPPH, in the acidic medium	%	P5
Antiradical activity, DPPH, in the basic medium	%	P6
Hydrogen peroxide scavenging activity in the acidic medium	%	P7
Hydrogen peroxide scavenging activity in the basic medium	%	P8

Note: AAE—ascorbic acid equivalents; TE—trolox equivalents; ME—malvidol glycoside equivalents; c.u.—conventional units; DPPH—2,2-diphenyl-1-picrylhydrazyl.

Table 2. The codes of measured parameters and the types of plant hydroalcoholic extracts.

Parameter code	Type of extract
P1, P2, P3, P4, P5, P6, P7, P8.	Chokeberry
P1, P2, P4, P5, P6.	Rose hip
P1, P2, P3, P4, P5, P6.	Hawthorn
P1, P2, P4, P5, P6.	Rowan
P1, P2, P3, P4, P5, P6, P7, P8.	Grape marc

2.2. Methods

The autocorrelation function (R_{xx}) was applied to characterize the internal structure of any series x , which in the discrete domain such as experimental data is determined by knowing that the n values of any series x_i are arranged at equal time intervals h and as a result $x_i = x(ih)$. Considering the time shift $\tau = rh$, the expression of the autocorrelation function results (with m the maximum possible shift):

$$R_{xx}(rh) = R_{xx}(\tau) = \frac{1}{n-r} \sum_{i=1}^{n-r} (x_i - m_x)(x_{i+r} - m_x), r = 0 \dots m \quad (1)$$

The autocorrelation function shows the degree of correlation of the experimental data. Thus, a perfect symmetry on the discrete time axis τ of the graph of the autocorrelation function shows the existence of a perfect linear dependence between data at different time points. In addition, the slower the autocorrelation function tends to zero, the better autocorrelation of the data is obtained.

The statistical properties of two random experimental series x and y are characterized by the intercorrelation function, denoted R_{xy} , which in the case of a finite discrete series, has the expression (with m the maximum possible shift and $\tau = r\Delta t$):

$$R_{xy}(r \cdot \Delta t) = \frac{1}{n-r} \sum_{i=1}^{n-r} (x_i - m_x)(y_{i+r} - m_y), r = 0 \dots m \quad (2)$$

Frequently, for determining the intercorrelation (interdependence) type between two quantities, that can be linear or nonlinear, direct or indirect, the correlation coefficient (Pearson's coefficient) is used, which for two series x and y is determined from the relation [3] [9]:

$$\rho_{xy} = \frac{R_{xy}(0)}{\sqrt{R_{xx}(0)R_{yy}(0)}}, \quad (3)$$

with values $\rho \in [-1; 1]$.

In expression (3) a maximum possible intercorrelation (a perfect linear dependence) is for $\rho^2 = 1$. If $\rho = 1$. Then there is a perfect direct linear dependence, and if $\rho = -1$ then there is a perfect indirect linear dependence. If $0 < \rho \leq 1$ there is a direct dependence, and if $-1 \leq \rho < 0$ —there is an indirect dependence (when x decreases, y increases and vice versa). Finally, if $\rho = 0$, then the two targeted quantities are independent. Therefore, the further ρ^2 is from the unit value (without reaching the zero value), the more accentuated the nonlinearity.

Correlation analysis was performed in the MATLAB program version R2020b (MathWorks, Inc., Natick, MA, USA) [10].

3. Results and Discussions

Figure 1 shows the autocorrelation functions of parameters P4 and P6 from ro-wan extracts. The discrete time τ , *i.e.* the number of values, was indicated on the abscissa axis. The graphs reveal good temporal autocorrelations of the experimental data because the curves do not suddenly tend to zero. Since the graphs

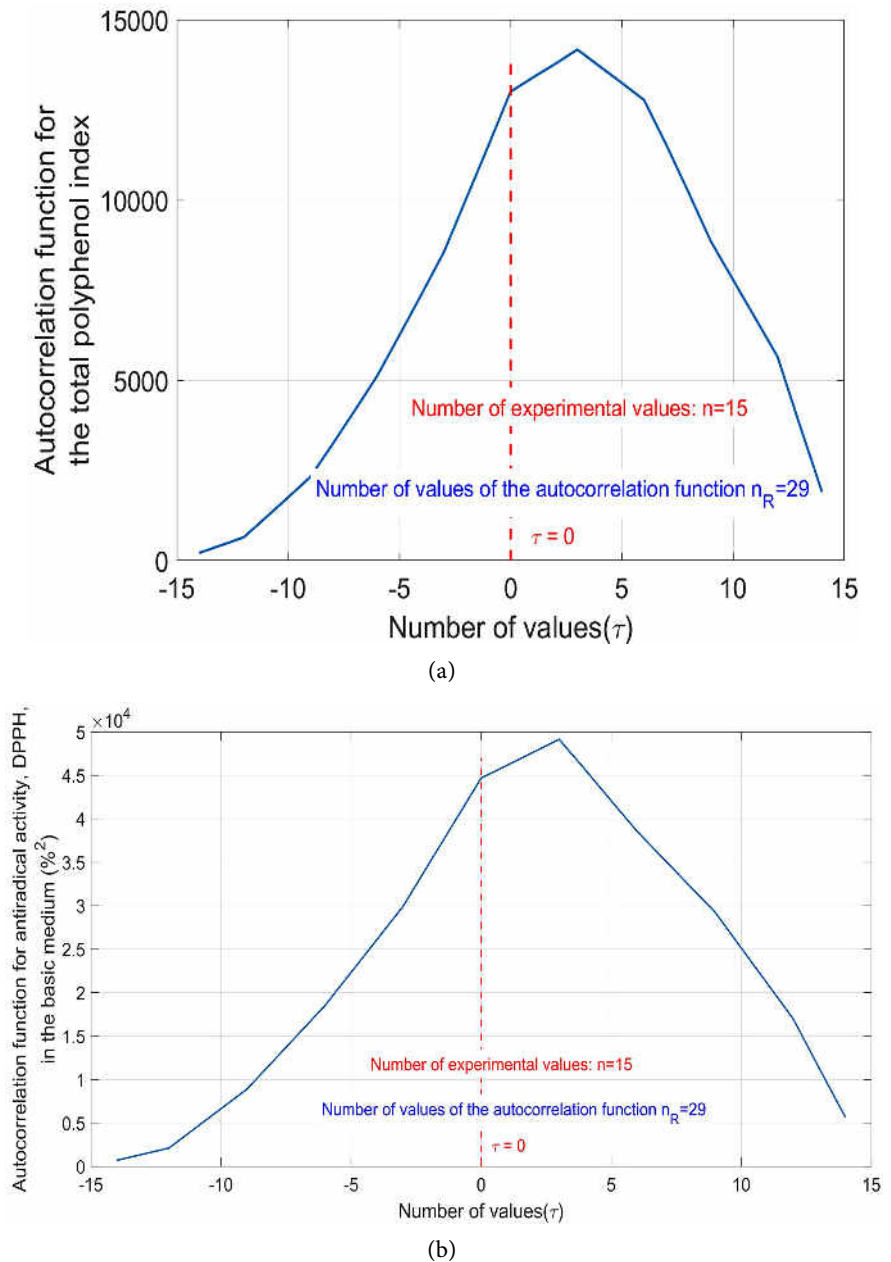


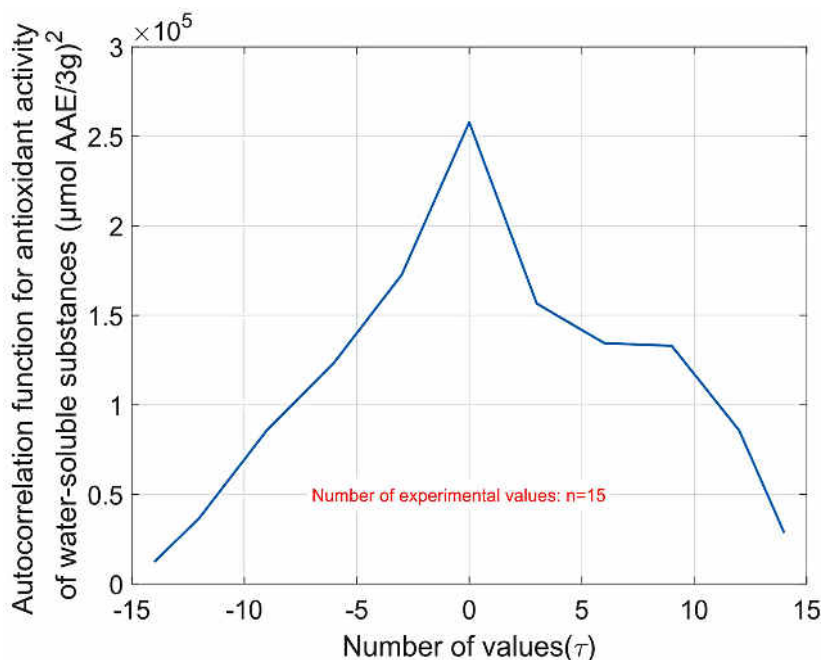
Figure 1. Autocorrelation functions for parameters P4 (a) and P6 (b) of hydroalcoholic extracts from rowan.

are not symmetrical about the vertical axis $\tau = 0$ (dashed lines), it follows that the experimental series is nonlinear. It should be noted that the origin of the discrete time is at $R_{xx}(0)$, so for the value $\tau = 0$, the graphs containing a number as a function of the number n of the values of the experimental series (in this case $n_R = 2n - 1 = 2 \times 15 - 1 = 29$ values).

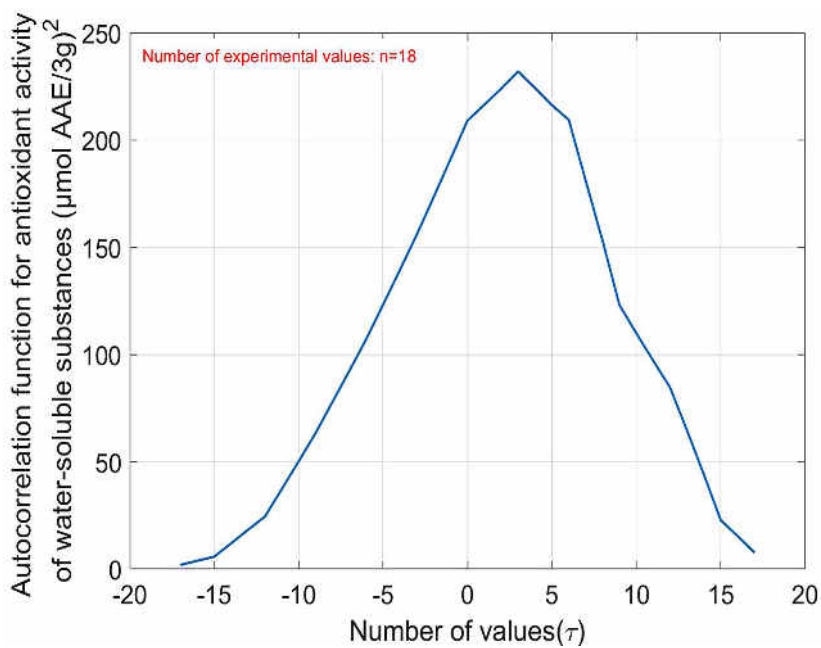
It should also be mentioned that the values and variations of the autocorrelation functions are given by the values and variations of the respective experimental series. This explains the different values and variations of the autocorrelation functions for the 2 parameters referred to in **Figure 1**.

The results presented for the autocorrelation function remain valid for the intercorrelation function as well. If the graph of the intercorrelation function does not suddenly tend to zero, then there is a good intercorrelation of the experimental data. Also, the symmetry of the graph of the intercorrelation function indicates the degree of linear dependence between the two quantities.

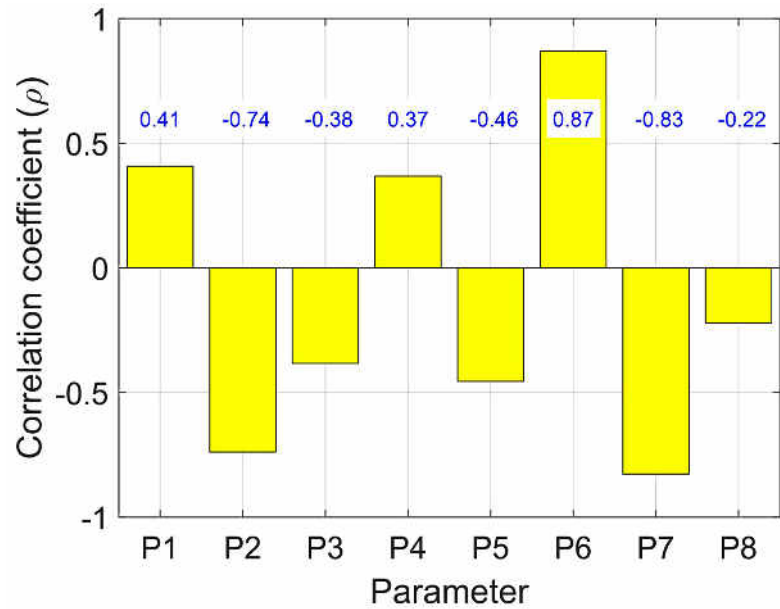
An example of the intercorrelation function is shown in **Figure 2** for chokeberry and grape marc extracts. The graphs in **Figure 2(a)** and **Figure 2(b)** show the autocorrelation functions. Graphs **Figure 2(c)** and **Figure 2(d)** show the



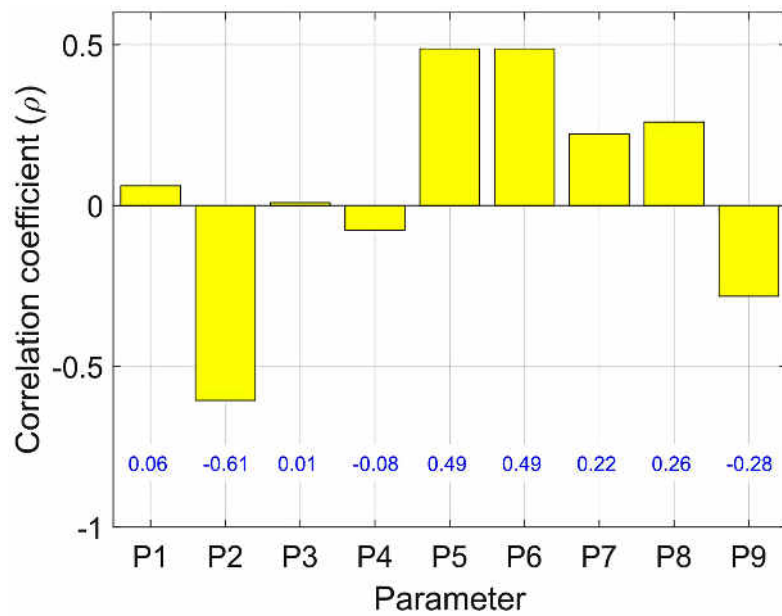
(a)



(b)



(c)



(d)

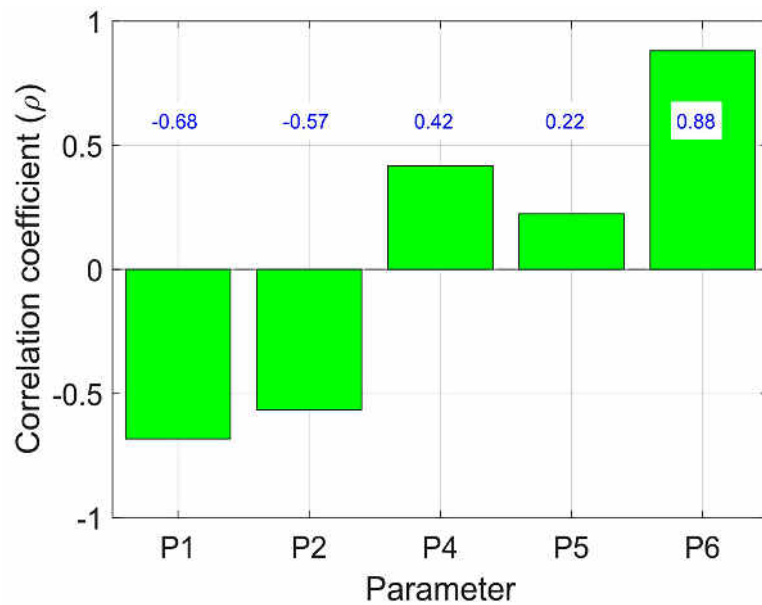
Figure 2. Autocorrelation functions and correlation coefficients between ethyl alcohol concentration and measured parameters: (a) P1-autocorrelation function for chokeberry extracts; (b) P1-autocorrelation function for grape marc extracts; (c) correlation coefficients for chokeberry extracts; (d) correlation coefficients for grape marc extracts.

values of the correlation coefficients between the ethyl alcohol concentration and the measured parameters. As noted, all correlation coefficients have subunit values, so there is a nonlinear dependence between the ethyl alcohol concentration and the measured parameters. Also, all correlation coefficients have non-zero values, so there is a certain dependence between the ethyl alcohol concentration and the measured parameters (they are not independent). Thus, the concentra-

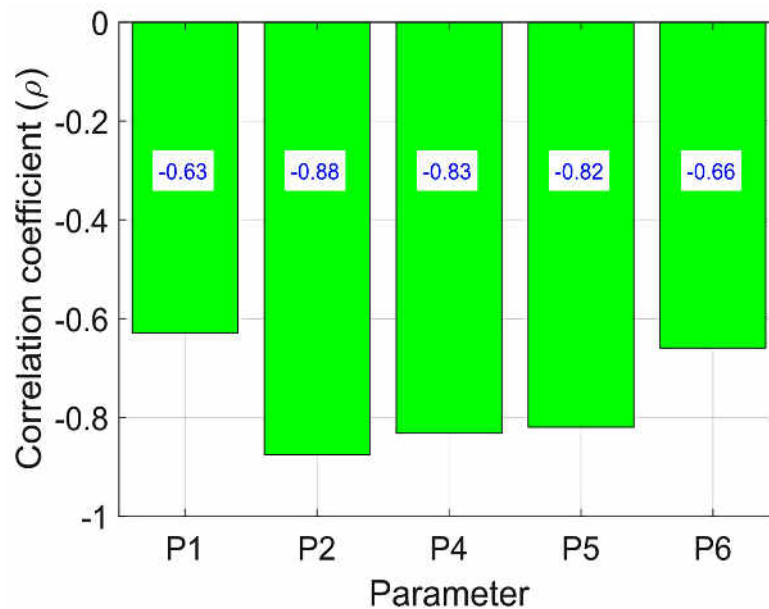
tion of ethyl alcohol is indeed a factor influencing all the measured parameters.

As it can be seen from **Figure 2(c)**, at parameter P6 the correlation coefficient is positive and therefore overall, when the concentration of ethyl alcohol increases, the values of the parameter increase and vice versa. On the other hand, at parameter P7 the correlation coefficient is negative and therefore overall, when the concentration of ethyl alcohol increases, the values of the parameter decrease and vice versa.

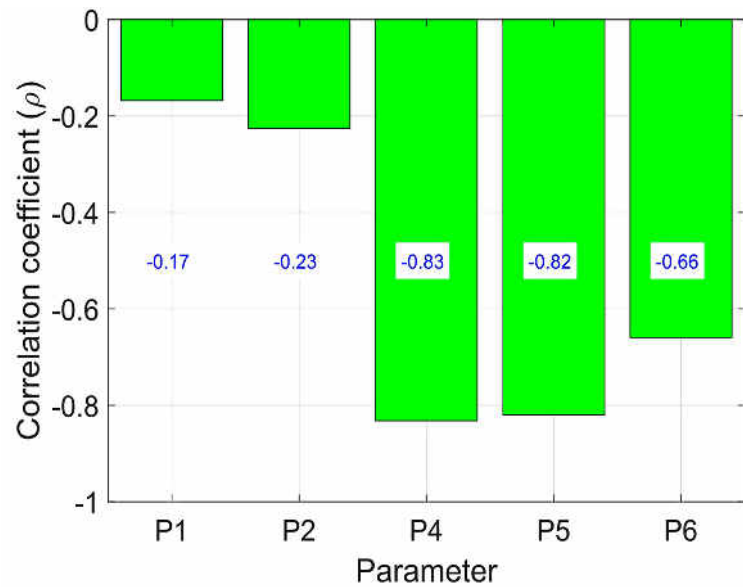
Similarly, **Figure 3** shows the values of the correlation coefficients for the other 4 types of plant extracts (sea buckthorn, rosehip, rowan, and hawthorn). And in these cases, all the correlation coefficients have subunitary and non-zero



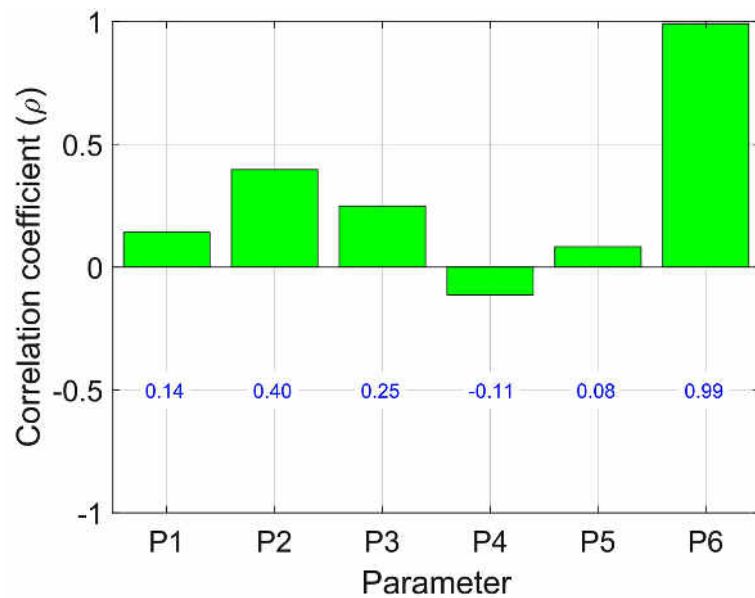
(a)



(b)



(c)



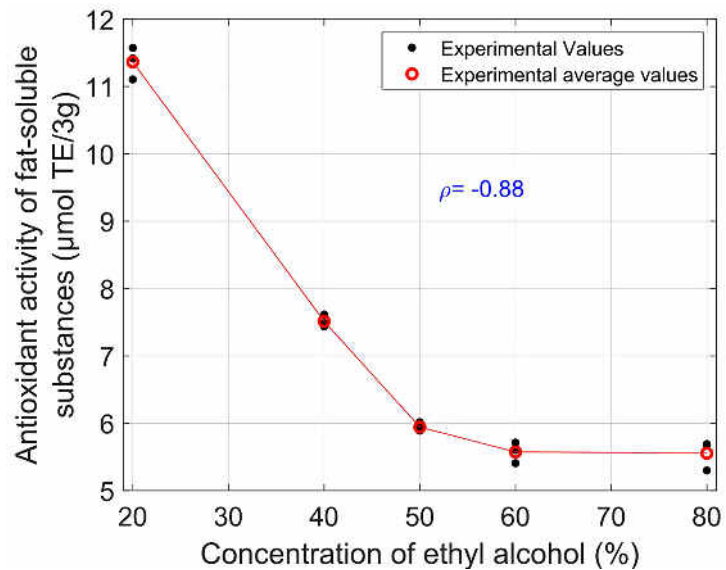
(d)

Figure 3. Correlation coefficients between the concentration of ethyl alcohol and the parameters measured for different types of extracts: (a) sea buckthorn; (b) rosehip; (c) rowan; (d) hawthorn.

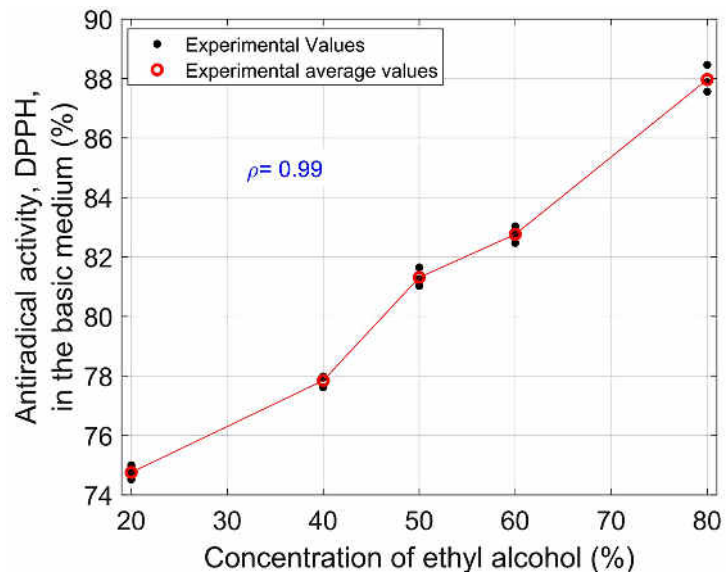
values, so the previous conclusions are valid here as well (there are nonlinear dependencies, and they are not independent).

To illustrate the significance of the correlation coefficient from **Figure 3**, parameter P2 from rosehip extracts (**Figure 3(b)**) and parameter P6 from hawthorn extracts (**Figure 3(d)**) were chosen.

Figure 4(a) demonstrates that the strictly decreasing curve of parameter P2 is confirmed by the high negative value of the correlation coefficient $\rho = -0.88$ (when the concentration of ethyl alcohol increases, the values of parameter P2



(a)



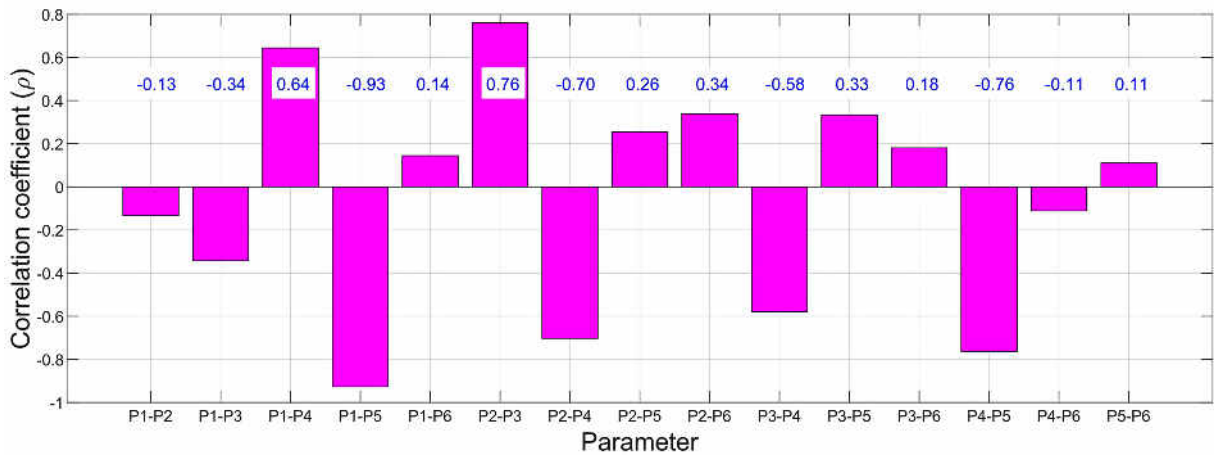
(b)

Figure 4. Correlation coefficients between the ethyl alcohol concentration and the measured parameters: (a) P2 for rosehip extracts; (b) P6 for hawthorn extracts.

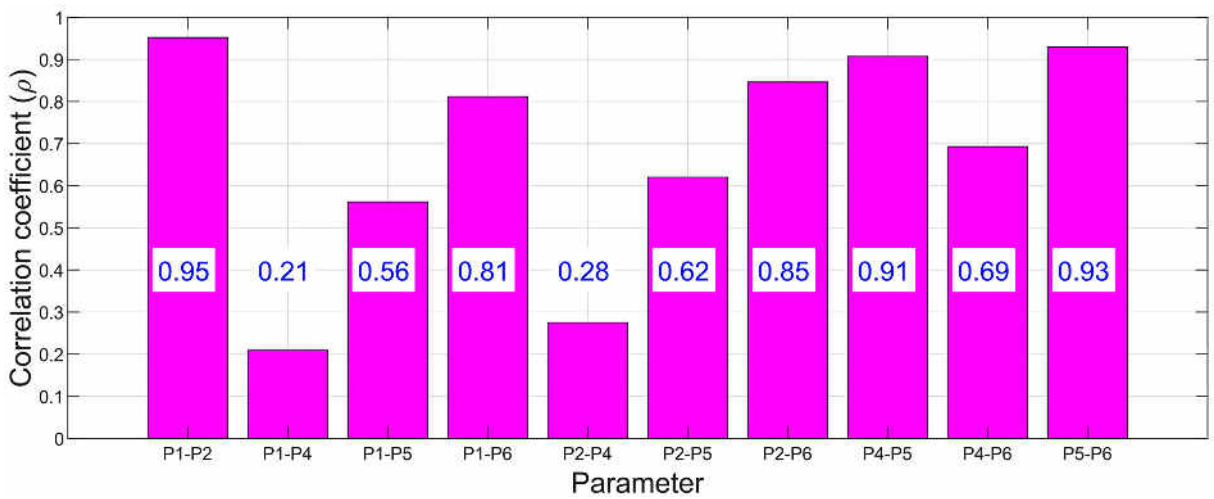
decrease and vice versa). Also, from **Figure 4(b)**, the strictly increasing curve of the parameter P6 is confirmed by the high positive value of the correlation coefficient $\rho = 0.99$ (when the concentration of ethyl alcohol increases, the values of the parameter P6 increase and vice versa).

The graphs in **Figure 5** confirm the aspect mentioned above, that between the measured parameters there are dependencies not only between the alcohol concentration and the determined parameters. Indeed, the non-zero values of the correlation coefficients (here in hawthorn and chokeberry extracts) show that the parameters are not independent.

Analogous to the graph in **Figure 4**, **Figure 6** provides two examples with the

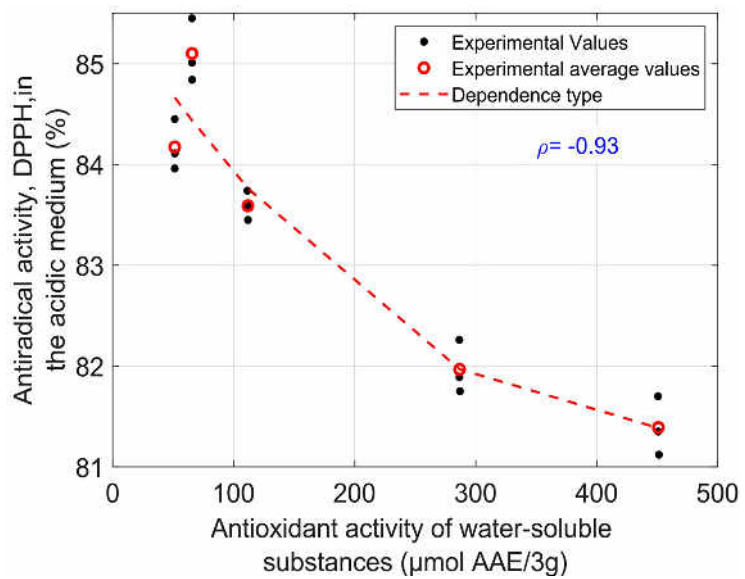


(a)



(b)

Figure 5. Correlation coefficients between the measured parameters: (a) for hawthorn extracts; (b) for chokeberry extracts.



(a)

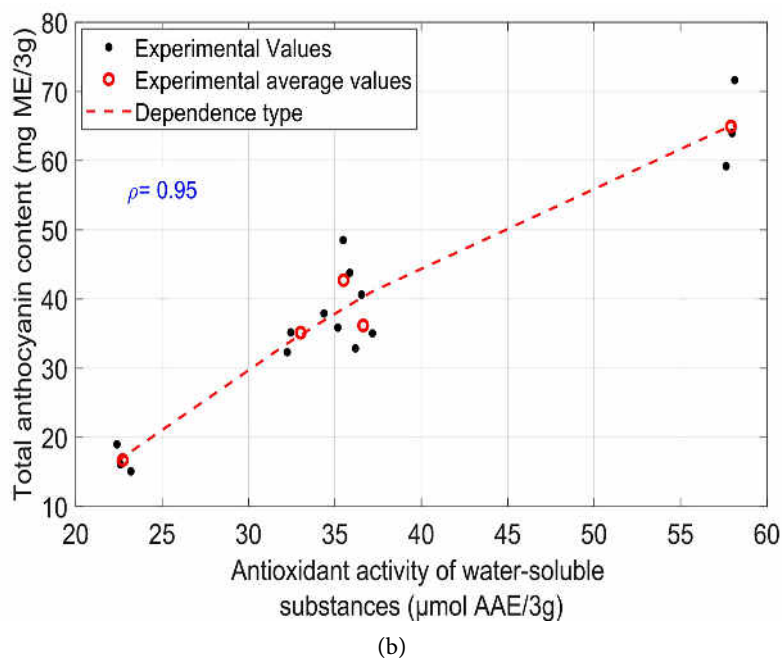


Figure 6. Correlation coefficients between the parameters measured in hawthorn (a) and chokeberry extracts (b).

significance of the correlation coefficient set in **Figure 5**. Thus, **Figure 6(a)** confirms the existence of an indirect quasilinear dependence between parameters P1 and P5 in hawthorn extracts, the correlation coefficient having a negative value close to the unit value. Similarly, **Figure 6(b)** confirms the existence of a direct quasilinear dependence between parameters P1 and P3 in the chokeberry extracts, the correlation coefficient having a positive value close to the unit.

The fact that the measured parameters are not independent of each other allows the establishment of mathematical models not only between the concentration of ethyl alcohol (as an influencing factor) and these parameters but even between the measured parameters [11] [12] [13] [14].

4. Conclusions

The correlation analysis of the experimental data allowed to establish the second order statistical characteristics (autocorrelation function, intercorrelation function and correlation coefficient). The correlation of the experimental data was assessed whether they are independent or dependent and what is the nature of the dependence. The nonlinear nature of the experimental series was estimated.

Based on the correlation analysis of the experimental data it was shown that the influencing factor (ethyl alcohol concentration) and the measured parameters show zero correlation coefficients for all types of researched extracts. This fact indicated that they are not independent and therefore afferent mathematical models can be deduced.

Between the measured parameters, there are dependencies on all 6 types of extracts, the afferent correlation coefficients having non-zero values. The allure

of the autocorrelation and intercorrelation functions, as well as the non-unitary values of the correlation coefficients, indicate the existence of nonlinear dependencies, with implications in establishing the mathematical models.

For rosehip and rowan extracts, the correlation coefficients between the measured parameters and the concentration of ethyl alcohol are negative, which means that on the whole of the measurements for these products there is an indirect dependence between the mentioned quantities.

The non-zero values of the correlation coefficients between the various parameters show that in all 6 types of extracts there are interdependencies between the measured parameters, so they are not independent. The non-unit values of the correlation coefficients between the various measured parameters indicate that there are nonlinear interdependencies between them.

Acknowledgements

This work was funded through AUF-MECC Project “Intelligent models to improve the training process” running at the Technical University of Moldova.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] van Wieringen, W.N. (2021) Lecture Notes on Ridge Regression, 129. <https://arxiv.org/pdf/1509.09169;Lecture>
- [2] Yuan, M. and Lin, Y. (2006) Model Selection and Estimation in Regression with Grouped Variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **68**, 49-67. <https://doi.org/10.1111/j.1467-9868.2005.00532.x>
- [3] Saleh, A.K.M.E., Arashi, M., and Kibria, B.M.G. (2019) Theory of Ridge Regression Estimation with Applications. John Wiley & Sons, New Jersey, 365 p. <https://doi.org/10.1002/9781118644478>
- [4] Ghendov-Mosanu, A., Cristea, E., Patras, A., Sturza, R. and Niculaua, M. (2020) Rose Hips, a Valuable Source of Antioxidants to Improve Gingerbread Characteristics. *Molecules*, **25**, Article No. 5659. <https://doi.org/10.3390/molecules25235659>
- [5] Ghendov-Mosanu, A., Cristea, E., Patras, A., Sturza, R., Padureanu, S., Deseatnicova, O., Turculet, N., Boestean, O. and Niculaua, M. (2020) Potential Application of Hippophae Rhamnoides in Wheat Bread Production. *Molecules*, **25**, Article No. 1272. <https://doi.org/10.3390/molecules25061272>
- [6] Opreș, O.I., Lung, I., Soran, M.L., Sturza, R. and Ghendov-Moșanu A. (2020) Fondant Candies Enriched with Antioxidants from Aronia Berries and Grape Marc. *Revista de Chimie*, **71**, 74-79. <https://doi.org/10.37358/RC.20.2.7895>
- [7] Cristea, E., Ghendov-Mosanu, A., Patras, A., Socaciu, C., Pinteau, A., Tudor, C. and Sturza, R. (2021) The Influence of Temperature, Storage Conditions, pH, and Ionic Strength on the Antioxidant Activity and Color Parameters of Rowan Berry Extracts. *Molecules*, **26**, Article No. 3786. <https://doi.org/10.3390/molecules26133786>
- [8] Ghendov-Moșanu, A., Cojocari, D., Balan, G. and Sturza, R. (2018) Antimicrobial

-
- Activity of Rose Hip and Hawthorn Powders on Pathogenic Bacteria. *Journal of Engineering Science*, **4**, 100-107.
- [9] Anatolyev, S. (2020) A Ridge to Homogeneity for Linear Models. *Journal of Statistical Computation and Simulation*, **90**, 2455-2472. <https://doi.org/10.1080/00949655.2020.1779722>
- [10] Barnes, B. and Fulford, G.R. (2011) *Mathematical Modelling with Case Studies: A Differential Equations Approach using Maple and MATLAB*. 2nd Edition, CRC Press, 368 p.
- [11] Andrei, T., Stancu, S. and Pele, D.T. (2002) *Statistics. Theory and Applications*. Economic, Bucharest, 576 p. (In Romanian)
- [12] Shao, J. and Deng, X. (2012) Estimation in High-Dimensional Linear Models with Deterministic Design Matrices. *The Annals of Statistics*, **40**, 812-831. <https://doi.org/10.1214/12-AOS982>
- [13] Van De Wiel, M.A., Lien, T.G., Verlaat, W., van Wieringen, W.N., and Wilting, S.M. (2016) Better Prediction by Use of Co-Data: Adaptive Group-Regularized Ridge Regression. *Statistics in Medicine*, **35**, 368-381. <https://doi.org/10.1002/sim.6732>
- [14] Kang, L., Yang, S., Peng, Y., Dai, J. and Ying, X. (2015) Research on Extraction Process of Gallic Acid from *Penthorum chinense* Pursh by Aqueous Ethanol. *Green and Sustainable Chemistry*, **5**, 63-69. <https://doi.org/10.4236/gsc.2015.52009>