Dragana Rajković, Ana Marjanović Jeromela, Lato Pezo, Biljana Lončar, Nada Grahovac, & Ankica Kondić Špika (2023) Artificial neural network and random forest regression models for modelling fatty acid and tocopherol content in oil of winter rapeseed. Journal of Food Composition and Analysis, 115, 105020, https://doi.org/10.1016/j.jfca.2022.105020

(https://www.sciencedirect.com/science/article/pii/S088915752200638X)

Available online 1 November 2022, Version of Record 16 November 2022

Artificial neural network and random forest regression models for modelling fatty acid and tocopherol content in oil of winter rapeseed

Dragana Rajković[1*], Ana Marjanović Jeromela[1], Lato Pezo[2], Biljana Lončar[3], Nada Grahovac[1], Ankica Kondić Špika[1]

[1]Institute of Field and Vegetable Crops, 21000 Novi Sad, Serbia;

dragana.rajkovic@ifvcns.ns.ac.rs;

ana.jeromela@ifvcns.ns.ac.rs

nada.grahovac@ifvcns.ns.ac.rs;

ankica.spika@ifvcns.ns.ac.rs

[2] Institute of General and Physical Chemistry, 11000 Belgrade, Serbia; latopezo@yahoo.co.uk

[3] University of Novi Sad, Faculty of Technology Novi Sad, 21000 Novi Sad, Serbia;

biljanacurcicc@gmail.com

[*] Correspondence: dragana.rajkovic@ifvcns.ns.ac.rs; Maksima Gorkog 30, 21 000 Novi Sad, Serbia, +38121 4898 321

Abstract

With the aid of models used in artificial intelligence, a wide range of data can be processed quickly with high accuracy. The quality of rapeseed oil from 40 genotypes cultivated during four consecutive years was analysed. Two machine learning techniques (artificial neural network – ANN, and random forest regression – RFR) were applied for the modelling of fatty acids content (C16:0; C18:0; C18:1; C18:2; C18:3 and C22:1), α-tocopherol, γ-tocopherol and total tocopherols, according to the data of production year and winter rapeseed genotype. The developed models exerted high-quality anticipation features, showing high r^2 during the training cycle. The best fit between the modelled and measured traits for ANN model was observed for erucic acid content. RFR modelling for all fatty acids was more effective than ANN model, with the highest precision for palmitic, stearic, and oleic fatty acids (r^2>0.9). This study emphasized the possibility of using ANN and RFR models to model winter rapeseed quality traits.

Keywords: Mathematical modelling, machine learning, rapeseed, quality traits, fatty acids, tocopherols

1. Introduction

Fatty acids and tocopherols are the main nutritive compounds of rapeseed (*Brassica napus* L.) oil. The composition of fatty acids within plant oil determines its quality and physicochemical properties. Rapeseed oil is the oleic type and often referred to as ideal in terms of omega-6 vs. omega-3 ratio (2:1). Rapeseed oil contains 50–70% oleic, 17–21% linoleic, and 7–10% linolenic acid (Adjonu et al., 2019; Koprna et al., 2006). Although seed quality of rapeseed is determined by genotype, it is also a consequence of complex interactions that occur between plant and environment. Temperature affects the fatty acids content of rapeseed oil. In a study

by Schulte et al. (2013), it was found that the share of oleic acid in oil increased as temperature increased during seed filling, while the share of linoleic and linolenic acids decreased. Increased minimum daily temperature, especially after the end of flowering, during oil accumulation, evoked increased oleic acid content and consequentially decreased linoleic content (Baux et al., 2013). Late sowing and drought reflect on lower oleic acid content in the rapeseed oil (Shirani Rad et al., 2014). Members of *Brassicaceae* family, especially rapeseed and brown mustard seeds (*Brassica juncea* (L.) Czern), are natural sources of erucic acid in which it is present in high amount (Vetter et al., 2020). Drought stress occurring in the late season of rapeseed cultivation leads to an enhanced amount of erucic acid (Gharechaei et al., 2019). Predominantly used technique for determination of fatty acids content in plant oils is gas chromatography, which is officially recommended by the American Oil Chemists' Society (AOCS).

One of the goals in quality breeding of rapeseed is focused on improving oil quality through changes in the content of certain fatty acids. These are namely oleic and erucic acids. Breeding for the purpose of human consumption led to the creation of canola quality cultivars and hybrids whose oil has maximally 2% of erucic acid and less than 30 μmol of glucosinolates per gram of defatted meal (Opinion of the Scientific Panel on Contaminants in the Food Chain on a request from the European Commission on glucosinolates as undesirable substances in animal feed, 2008). Such a low concentration of erucic acid in modern rapeseed cultivars is relatively low for *Brassicaceae* species.

Apart from fatty acids, the content and composition of tocopherols in rapeseed oil are important for its stability. Tocopherols and tocotrienols are monophenols, and represent forms of vitamin E. Alfa isoform of vitamin E is most potent and powerful regarding biological activity. Vitamin E is an essential micronutrient for humans. Tocopherols, as important antioxidants, protect polyunsaturated fatty acids from lipid peroxidation (Lebold & Traber, 2014). The presence of these natural antioxidants in vegetable oils and processed products (margarine, salad dressings,

and mayonnaise) is important for the benefit of human health. There are several methods for determination of individual tocopherols in vegetable oils. Normal, or reverse phase high-performance liquid chromatography (HPLC) is mostly used with fluorescence or ultraviolet-visible detection (Bakre et al., 2015; Gruszka & Kruk, 2007).

Many chemical components of the oil are strongly correlated. Therefore, modelling the chemical properties of rapeseed oil is essential to guide breeders towards on-time information about genotypes with the highest amount of certain desirable compounds to make timely decisions and quickly evaluate large-scale samples. Especially when it comes to erucic acid content in food-based products, the need for quality monitoring of rapeseed oil is evident.

Within the last decade, machine learning has been successfully used in agriculture. Most literature points out the benefits of these technologies for crop yield modelling, detection of different crop conditions such as diseases and mineral insufficiency (Iniyan et al., 2020; Niedbała et al., 2019; Yu et al., 2020). Machine learning, as a nonlinear and nonparametric method, has higher efficiency over classical statistical methods in analysing data related to complex relationships in living organisms. Machine learning methods are prosperous for the analysis of crop quality and contribute to higher revenue, because they reduce the number of required analyses. The capability of modelling seed quality on specific farms/locations based on the information about genotypes, management practice and environmental conditions is challenging, but achievable task. Published literature on machine learning related to *Brassica* species mostly covers papers that refer to image analysis with the aid of different learning algorithms. Crop quality was the subject of only 3% of studies that are related to machine learning in agriculture (Benos et al., 2021).

The classical machine learning models such as: artificial neural network (ANN), random forest regression (RFR), support vector machine (SVM), extreme learning machine (ELM), K-nearest neighbors (KNN) and decision tree (DT) are extensively used in modelling in various branches

of science. The SVM is widely used discriminant technique based on the statistical learning theory, well recognized for its strong generalization ability. The optimal network is obtained by exploring the balance among the complexity of the model and the training error (Ma et al., 2022). The ELM designs a single-layer feedforward network by randomly generating the input weights and biases of the hidden layers (Wang et al., 2022).

The vast variety of state-of-the-art machine learning techniques are suitable for sequence data like ensemble learning models, such as: XGBoost (Su et al., 2022) and LightGBM (Jawad et al., 2022) and CatBoost. XGBoost model exerts its advantages especially for high prediction accuracy and interpretability. LightGBM model enables large amounts of data and GPU training. The LightGBM models are proven to be more accurate and faster than XGBoost. Data fusion enables stronger forecasting accuracy, according to the integration of gradient boosting based categorical attributes supported by CatBoost algorithm (Dutta & Roy, 2022).

In Imahara et al. (2006), modelling procedure was established to determine the optimal fatty acid composition of vegetable oil for biodiesel production. Campbell et al. (2021) used trait-specific genomic relationship matrices to model fatty acids and lipid content in oat seed and reported advantages of this approach over conventional genomic prediction. In the experiment of Niedbała et al. (2020), ANN was developed with the aim to estimate ferulic acid concentration in wheat. Similar to our study, they also created a model on the basis of cultivar and weather data. Data on the fatty acids content in rapeseed oil can be used to estimate the oxidative stability of oil using ANN (Dehghani et al., 2012). Chemical composition, sensory properties, as well as verification of the authenticity and geographical origin of olive oil can be predicted with the means of ANN (Gonzalez-Fernandez et al., 2018).

In recent study of Rajković et al. (2022), ANN and RFR models were used to estimate the seed yield, oil and protein yield, oil and protein content, and 1000 seed weight, based on the year of production and genotype. The exploration of the ANN and RFR models in this new article was

with the goal to model the fatty acids and tocopherols content, according to data of production year and rapeseed genotype. Therefore, this article should be considered as a second part of the same study. To the best of the author`s knowledge, no previous studies have addressed the modelling of fatty acids and tocopherols content in rapeseed using a machine learning approach. The main objective of this investigation was to explore the potential of forecasting the fatty acids (C16:0; C18:0; C18:1; C18:2; C18:3 and C22:1) and tocopherols content (α-tocopherol, γ-tocopherol and total tocopherols), according to data of production year and rapeseed genotype (introduced as categorical variables), developing two machine learning methods, such as artificial neural network model (ANN) and random forest regression (RFR) models.

2. Materials and methods

2.1.Rapeseed samples

Cold-pressed rapeseed oil was obtained from 40 winter rapeseed genotypes. Three genotypes were experimental hybrids (NS-H-R-1, NS-H-R-2, NS-H-R-3), 37 were lines from which 14 are registered cultivars (Banaćanka, Slavica, Zlatna, Branka, Express, Nevena, Valesca, Ilia, Kata, Nena, Svetlana, Jasna, Zorica and Jelena) and 21 are experimental lines (NS-L-74, NS-L-7, NS-L-31, NS-L-126, NS-L-33, NS-L-128, NS-L-101, NS-L-102, NS-L-134, NS-L-32, NS-L-136, NS-L-137, NS-L-138, NS-L-251, NS-L-210, NS-L-44, NS-L-45, NS-L-46, NS-L-47, Forward and Maidan). The remaining two genotypes were improved lines derived from the existing variety Valesca (Valeska tamna and Valeska svetla). The genotypes Valesca, Valeska tamna and Valeska svetla originate from Sweden, Express is from Germany, and the remaining genotypes are from Serbia. Upon pressing on the hydraulic oil press machine, oil samples were immediately stored in the dark at -40 °C for one to four weeks until the moment of analysis.

2.2.Trial design

A field trial was set up during four vegetation seasons in the period 2014-2018 as a randomized complete block design on location Rimski šančevi (45°19'53.7" N 19°50'12.6" E), Vojvodina province, Serbia. Field trial consisted of 40 rapeseed genotypes and was set up in three replicates with a plant protection (rapeseed cultivar Slavica was sown) around the whole trial to reduce border effect. Dimension of experimental plot was 6 m^2 (4 m × 1.5 m). Sowing was performed with manual single-row planter at 2-3 cm depth. Sowing and standard cultivation practices throughout the years were applied at the optimum time. Each year, NPK fertilizer (nitrogen, phosphorus, potassium) was applied prior to sowing (Table 1). Plant protection was performed in accordance with the pathogen and pest infestation. Climate variables such as average daily temperatures, precipitations and sunhine hours were obtained from the Republic hydrometeorological service of Serbia (Fig. 1a and 1b) and Ogimet weather service (Fig. 1c) for meteo station ``Rimski šančevi``.

Fig. 1.

Table 1.

2.3.Weather conditions

During the winter season of 2014/15, precipitation levels were higher than the multi-year average (1964-2013), resulting in moisture reserves in the soil's deeper layers. The second half of April, when rapeseed was in flowering phase, was quite dry, which was unfavorable for the plant development. The spring was warm, with three times more precipitations in May as compared to the long-term average. The winter of 2015/16 was warmer than the long-term

mean. Days with the least amount of sunshine were observed during the winter months, as expected. In December 2015, even there was decrease in sunshine hours, amount of precipitations was much lower than average. Average daily temperatures at the end of January and in February were above 5℃, which influenced the earlier start of vegetation. Average temperatures for this period of the year with a higher amount of precipitation were recorded during spring. Around two times more precipitations than the multi-year average rainfall occurred in June and extended the seed filling stage. In the autumn of the production year 2016/17, there was a lot of precipitations, while the winter was mostly dry, with the average precipitation below the multi-year mean level. A warmer production year than usual was observed during 2017/18. Decrease in sunhine hours in March may be reflected in a decrease of photosynthesis efficiency. April and May were warmer than average, with 8 and 9.5 average month sun hours respectively, resulting in faster rapeseed growth.

2.4.Fatty acid composition

Fatty acids were methylated and chemically converted into their volatile esters in transesterification reaction. Fatty acid methyl esters (FAME) were prepared according to the method by Kravić et al. (2010) with some modifications. Oil of rapeseed (170 µL) and n-hexane (2.4 mL) were added to a test tube with a stopper. Then, 0.6 mL of 2 moL/dm^3 KOH in methanol was added and shaken for 20 seconds. Following this, the closed test tube was placed in a heated water bath at 70˚C for one minute, after which it was removed from the water bath and shaken for 20 seconds. Afterwards, 1.2 mL of 1 moL/dm3 HCl in methanol was added to the tube and left until separated into two phases. After phase separation, 1 µL of fatty methyl esters in n-hexane (upper phase) was injected into the gas chromatograph with flame ionizing detector (GC-FID). The composition and relative content of individual fatty acids were determined with a gas chromatograph (Konik HRGC 4000) equipped with a fused silica capillary column

(Supelco Omegawax® 250, 30 m × 0.25 mm ID, film thickness 0.25 µm), and poly(ethylene glycol) stationary phase. An oven temperature of 150 °C was used, then the temperature was raised to 250 °C at a rate of 12 °C/min for 8 minutes. The injector and detector temperatures were 250 ℃. Flow rate of helium (carrier gas) was 1 mL/min with a split ratio of 1:70. Fatty acids were identified by comparing their relative retention time in analysed samples with retention time of pure commercial standard fatty acid methyl esters solution (multistandard from Supelco, Cat. No 07256-AMP, 07756-1AMP) under the same conditions. Data processing was performed with Konikrom plus software (DataApex, ver. 2.3.0.195).

2.5.Tocopherol composition

Tocopherol composition was determined by HPLC chromatograph with a fluorescent detector according to Lazzez et al. (2008) method with slight modification. Rapeseed oils (300 µL) and n-hexane were added to a 2 mL volumetric flask with stirring. Aliquot (1 mL) of this solution was filtered into the reaction vial for HPLC analysis through a regenerated cellulose filter (0.22 µm). Sykam HPLC system normal-phase liquid chromatography was used to separate tocopherols. Tocopherols from oil were separated on a Nucleosil 100-5 NH 2 amino column (Machery Nagel, 250 × 4.6 mm, 5 µm particle size, 100 Å pore size). As a mobile phase, a mixture of n-hexane/ethyl acetate (70:30, v/v) with a flow rate of 1 mL/min was used. The temperature of the detector was 30 ℃. The eluent was monitored using the fluorescence detector set at excitation wavelength 280 nm and emission wavelength 340 nm. Tocopherols were identified and quantified by comparing retention times of samples with retention times of commercial standards in hexane (dl α-tocopherol [Cat-No.4-7783], rac β-tocopherol [Cat.No. 46401-U], γ- [Cat.No. 47784] and δ-tocopherol [Cat.No. 4-7785], manufactured by Sigma-Aldrich). Total tocopherols are represented as a sum of alpha- and gamma-tocopherols. Clarity

Chromatography Station (DataApex, ver. 7.4.1.88) software was used to process the obtained data.

2.6.ANN modelling

In this investigation, the ANN modelling technique was chosen for modelling purposes, due to its proven efficiency in approximating nonlinear functions (Agatov, 2019; Anitescu et al., 2019; Basir et al., 2021; Kleijnen, 2018; Kujawa & Niedbała, 2021; Samaniego et al., 2020). The ANN model building structure was based on the multi-layer perceptron model (MLP) scheme, comprising of three layers (input, hidden, and output) to forecast the fatty acids content (C16:0; C18:0; C18:1; C18:2; C18:3 and C22:1) and tocopherols content (α-tocopherol, γ-tocopherol and total tocopherols), relied on the year of production and rapeseed genotype.

The MLP-formed ANN model could be presented using matrix notation, with weight and bias coefficients associated with the hidden and output layer written in matrices W_1, B_1, W_2 and B_2, with Y as the output variables matrix, f_1 and f_2 as activation functions in the hidden and output layers, and with X as the matrix of input variables (Kollo & von Rosen, 2005):

$$Y = f_1(W_2 \cdot f_2(W_1 \cdot X + B_1) + B_2) \quad (1)$$

Prior to the calculation, the experimentally obtained database which consisted of measured input and output parameters was transformed using a min-max normalization scheme. This database was randomly divided into training and testing groups (70% and 30%, respectively). Throughout the learning procedure, ANN inputs were supplied with a training set of parameters, in order to establish the optimal number of neurons in the hidden layer, to estimate the weights and bias coefficients and non-linear activation functions for every neuron in the ANN model. The Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm was employed during the iterative process of weights and biases coefficients calculation (Berrueta et al., 2007; Doumpos &

223 Zopounidis, 2011). A sequence of distinct MLP-formed ANN layouts were investigated,
224 altering the number of hidden neurons (between 5 to 20) introducing random initial values of
225 weights and biases coefficients. The learning procedure of the network was repeated 100,000
226 times (Pezo et al., 2013).
227
228 ***2.6.1. Global sensitivity analysis***
229 The Yoon's method for global sensitivity analysis was employed to estimate the relative
230 influence of the inputs on the output variables, enumerating the weighting coefficients within
231 the ANN model (Yoon et al., 1993):

232 $$RI_{ij}(\%) = \frac{\sum_{k=0}^{n}(w_{ik} \cdot w_{kj})}{\sum_{i=0}^{m}\left|\sum_{k=0}^{n}(w_{ik} \cdot w_{kj})\right|} \cdot 100\% \quad (2)$$

233 where the parameters abbreviations were: w – weight coefficient in the ANN model, i – input
234 variable, j - output variable, k - hidden neuron, n - number of hidden neurons, m - number of
235 inputs.
236
237 ***2.7.RFR modelling***
238 Random forest regression (RFR) modelling is a widely accepted machine learning
239 mathematical tool developed according to the decision trees principle, with an intention of
240 modelling the output variables corresponding to the inputs (Breiman, 2001). The RFR
241 modelling is employed to foresee the structure of each individual tree, according to developed
242 decision trees computed utilizing the training dataset (Rasaei & Bogaert, 2019). In the course
243 of RFR modelling, a huge number of decision trees were grown and tested, and a single tree
244 was modelled based on the unique bootstrap sample within a training dataset (Khanal et al.,
245 2018). The fatty acids content (C16:0; C18:0; C18:1; C18:2; C18:3 and C22:1) and tocopherols

content (α-tocopherol, γ-tocopherol and total tocopherols) were modelled, according to the year of production and genotype. The same sets of training and testing data were used for ANN and RFR modelling. In this investigation, the bootstrap function was utilized to randomly divide the dataset into two uniform subsets (training and test) which outlined 70% and 30% of the entire data (Zhang et al., 2021). Input sample dataset was used for designing new sub-samples, and multiple trees were associated with the RFR structure to fit thus obtained sub-samples. Throughout the training sequence, the RFR model averaged the outcomes of the grown trees, with the intention of diminishing the error of anticipation (Yang et al., 2021). In the course of RFR calculation, the count of trees was adjusted to 100, 200, 300, 400, 500, and 10000, whereas the random training data proportion count was set to 70% and the test sample proportion was 30%.

2.8.The accuracy of the model

The numerical confirmation of the obtained ANN and RFR models was performed using statistical tests, such as coefficient of determination (r^2), reduced chi-square (χ^2), mean bias error (MBE), root mean square error (RMSE), mean percentage error (MPE), sum of squared errors (SSE) and average absolute relative deviation (AARD). These commonly used parameters were calculated as in Puntarić et al., (2022):

$$\chi^2 = \frac{\sum_{i=1}^{N}(x_{\exp,i} - x_{pre,i})^2}{N-n}, \quad (3)$$

$$RMSE = \left[\frac{1}{N}\cdot\sum_{i=1}^{N}(x_{pre,i} - x_{\exp,i})^2\right]^{1/2}, \quad (4)$$

$$MBE = \frac{1}{N}\cdot\sum_{i=1}^{N}(x_{pre,i} - x_{\exp,i}), \quad (5)$$

$$MPE = \frac{100}{N}\cdot\sum_{i=1}^{N}\left(\frac{\left|x_{pre,i} - x_{\exp,i}\right|}{x_{\exp,i}}\right), \quad (6)$$

$$SSE = \sum_{i=1}^{N} (x_{pre,i} - x_{\exp,i})^2 , \quad (7)$$

$$AARD = \frac{1}{N} \cdot \sum_{i=1}^{N} \left| \frac{x_{\exp,i} - x_{pre,i}}{x_{\exp,i}} \right| , \quad (8)$$

where $x_{exp,i}$ were experimental values and $x_{pre,i}$ were the model predicted values, N and n are the number of observations and constants, accordingly.

2.9.Statistical analysis

For comparison of the mean values of individual fatty acids and tocopherols concentrations in different oil samples, analysis of variance (ANOVA) with the Duncan's post-hoc test was used. The correlation analysis of fatty acids and tocopherols was performed using R software v.4.0.3 (64-bit version). Data processing for ANN and RFR modelling was performed with the StatSoft Statistica, ver. 10.0, Palo Alto, CA, USA.

3. Results

3.1.Fatty acids and tocopherols content

Following fatty acids were identified by gas chromatography in rapeseed oil samples: myristic (C14:0), palmitic (C16:0), stearic (C18:0), oleic (C18:1), linoleic (C18:2), linolenic (C18:3), arachidic (C20:0), arachidonic (C20:4), behenic (C22:0), erucic and lignoceric acid (C24:0). Due to the small amount of myristic (C14:0), arachidic, arachidonic, behenic and lignoceric acids, which individually accounted for less than 1% of total fatty acids, their content is not presented. In all analysed samples, oleic acid was the dominant fatty acid with an average of 58.64% (Fig. 2, Suplementary Table 1). Then follows linoleic (average 20.12%) and linolenic

acid (average 11.48%), which was around two times less represented in oil than linoleic. Among saturated fatty acids, palmitic acid was present in the highest amount.

Fig. 2

Genotypic value for C18:1 of analysed genotypes in the period 2015–2018 varied from 37.59±0.26% (NS-L-102) to 61.90±0.22% (NS-L-7). Kata, Jasna, NS-L-31, NS-L-74, NS-L-210, NS-L-45 and Zlatna also had a high four-year average of C18:1 of over 61%. The mean C18:1 value of 58.64% indicates that rapeseed oil is of high quality. Each analysed year, including the four-year mean, line NS-L-102 had the lowest C18:1 content of 37.59±0.26%. A low four-year mean of C18:1, below 50%, was also determined in Valeska tamna. Cultivar Slavica, the standard for DUS (distinctness, uniformity, stability) and VCU (value for cultivation and use) tests in Serbia, had a higher oleic acid content (59.07±0.24%) than the grand mean (58.64%). Between the analysed years, average values of C18:1 varied the most in the line NS-L-251 (38.68±0.28 – 62.10±0.27%). The biggest deviation in the variation of this line was in 2018, when the share of C18:1 in the total mixture of fatty acids was about 20% lower compared to its content in other analysed years. The smallest variation in C18:1 content between years was observed in lines NS-L-136, NS-L-32, NS-L-138 and NS-L-210. In relation to linoleic acid content, the average value during the four-year period was 20.12%. NS-L-102 had the lowest mean C18:2 of 17.08±0.22%, while Jasna had the lowest content of C18:3 10.49±0.27%. Valeska tamna, Valeska svetla, Nevena, Nena, NS-L-102 and NS-L-251 had mean erucic acid content above the maximum threshold (2%) value allowed for canola quality of the oil. Content of C22:1 in each year, including the four-year average, was the highest in line NS-L-102, 14.13±0.28%.

Amount of total tocopherols in rapeseed genotypes ranged from 397.06±16.25 mg/kg (Kata) to 514.18±11.84 mg/kg (Valesca). Gamma-tocopherols content varied from 249.95 to 325.42 mg/kg (Valesca tamna and Valesca, respectively), while alpha-tocopherols represented 135.93±8.37 – 218.51±8.51 mg/kg (Kata and Valeska tamna, respectively) [Fig. 3, Supplementary Table 1].

Fig. 3

3.1.1. Correlation analysis and heat map of data

The results of the correlation analysis are presented in Fig. 4a. The diameter of the circle and the circle's colour are influenced by the correlation coefficients; the blue colour significates a positive correlation, while the red colour represents the negative correlation. The circle's size significates the absolute value of the obtained correlation coefficient. The highest positive correlations were found between γ-tocopherol and total tocopherol content (r = 0.88; $p \leq 0.05$) and α-tocopherol and total tocopherol content (r = 0.74; $p \leq 0.05$). The strongest negative correlation was observed between C18:1 and C22:1 content (r=-0.92; $p \leq 0.05$).

The heat map of fatty acids content and tocopherols data is presented in Fig. 4b. The first hierarchical cluster contained C22:1, C:16:0 and C18:0, while other cluster comprised C18:1, C18:2, C18:3 and tocopherols. The order of the variables were re-ordered according to the hierarchical clustering result, putting similar variables close to each other. The colour scheme was applied for the visualization of the data and to simplify the recognition of variable's belonging to a specific cluster.

Fig. 4.

3.2.ANN model

The developed optimal neural network model showed adequate generalization capabilities for the modelling of experimental results: fatty acids profile (C16:0; C18:0; C18:1; C18:2; C18:3 and C22:1), α-tocopherol, γ-tocopherol, and total tocopherols, according to the year of production and genotype. The optimum number of neurons in the hidden layer of ANN model was 10 (network MLP 44-10-9) (Table 2), while the r^2 values were: 0.813; 0.763; 0.906; 0.866; 0.828; 0.958; 0.854; 0.860 and 0.884, accordingly, during the training cycle for output variables. The obtained r^2 values during the testing cycle were: 0.712; 0.587; 0.860; 0.718; 0.772; 0.946; 0.655; 0.641 and 0.639 for C16:0; C18:0; C18:1; C18:2; C18:3, C22:1, α-tocopherol, γ-tocopherol, and total tocopherols contents modelling.

Table 2

The developed ANN model for fatty acids profile, α-tocopherol, γ-tocopherol and total tocopherols modelling was consisted of 117 weights-bias coefficients showing the high nonlinearity of the system (Chattopadhyay & Rangarajan, 2014; Montgomery, 1984). Supplementary Table 2 presents the elements of matrix W_1 and vector B_1, while Supplementary Table 3 presents the elements of matrix W_2 and vector B_2, which were derived during the ANN model development, using Equation 1. The goodness of fit between experimental and model-calculated results, were shown in Table 3.

Table 3

The results obtained from database were fitted to the developed ANN model. reduced chi-square (χ^2), root mean square error (RMSE), mean bias error (MBE), mean percentage error

(MPE), sum of squared errors (SSE), coefficient of determination (r^2) and average absolute relative deviation (AARD) were calculated statistical parameters applied for determination of fitting quality between database and the developed model. The particularly high values of r^2 and low χ^2, RMSE, MBE, MPE,SSE and AARD suggested adequate fit (Tables 3 and 4). The ANN model showed better fit to C22:1, α-tocopherol, γ-tocopherol and total tocopherols content data, according to relatively low χ^2, RMSE, MBE, MPE, SSE and AARD, as well as the high r^2 values (Table 3).

The ANN models satisfactorily modelled experimental variables for various process variables. For the ANN model, the model calculated fatty acids profile, α-tocopherol, γ-tocopherol and total tocopherols content were not too close to the experimental values in most cases, in terms of r^2 values, while the sum of squares (*SOS*) values acquired using the ANN model was of the same order of magnitude as experimental errors for outputs mentioned in the literature (Doumpos & Zopounidis, 2011; Kollo & von Rosen, 2005).

The efficiency of the ANN model in modelling fatty acids profile, α-tocopherol, γ-tocopherol, and total tocopherol content is graphically illustrated by scatter plots (Fig. 5). In most scatter plots, data are distributed with large dispersion, indicating low prediction accuracy.

Fig. 5

The subsampling testing was included in this investigation using Statistica's software routine, applied in the same MLP environment. More than 2000 subsampling was tested, and the results of the training and testing cycles were presented in Fig. 6.

Fig. 6

The training curve presented in Fig. 6 showed that the loss of the model follows a descending trend, for training and testing curves. The gap between these curves is referred as the "generalization gap". A plot of learning curves shows a good fit, knowing that the training loss decreases to a point of stability, and that the plot of validation loss decreases to a point of stability and has a small gap with the training loss.

3.2.1. Global sensitivity

Within this chapter, the investigation of factor's impacts (such as production year and winter rapeseed genotype) on the fatty acids content (C16:0; C18:0; C18:1; C18:2; C18:3 and C22:1) and α-tocopherol, γ-tocopherol and total tocopherols content were performed, according to the results of the developed ANN model. As illustrated in Fig. 7, most intensive positive influences for the C16:0 content were observed for: genotype 3 [NS-H-R 3] (+7.39%), genotype 5 [Slavica] (+6.59%) and year 2018 (6.37%), while the most prominent negative influence was observed for genotype 34 [NS-L-44] (-7.39%). The most intensive positive impact for the C18:0 content was observed for year 2018 (8.62%) and for genotype 26 [NS-L-102] (6.07%), whilst the strongest negative influence was noticed for genotype 35 [NS-L-45] (-5.32%). The strongest negative influence for the C18:1 content was recorded for genotype 26 (-16.23%), while the strongest positive influence was observed for genotype 8 [Zlatna] (+6.75%). The strongest negative influence for the C18:2 content was also noticed for genotype 26 (-9.79%), while the strongest positive influence was observed for genotype 30 [NS-L-137] (+5.59%). The strongest negative influence for the C18:3 content was also noticed for genotype 6 [Valeska tamna] (-8.75%) and genotype 16 [NS-Kata] (-8.43%), while the strongest positive influence was also observed for genotype 30 (+5.77%). The strongest positive influence for the C22:1 content was noticed for genotype 26 (+16.54%).

The strongest negative influence for the α-tocopherol content was noticed for genotype 16 (-5.82%) and genotype 20 [NS-L-33] (-5.23%), while the strongest positive influence was observed for genotype 6 (+8.54%), genotype 14 [Valesca] (+7.86%) and genotype 28 [NS-L-32] (+7.34%), during year 2015 (+4.64%). The strongest negative influence for the γ-tocopherol content was noticed for genotype 6 (-9.30%) and genotype 16 (-7.74%), while the strongest positive influence was observed for genotype 36 [NS-L-46] (+6.31%)%), during year 2015 (+6.34%). The strongest negative influence for the γ-tocopherol content was noticed for genotype 16 (-8.58%), while the strongest positive influence was also observed for genotype 14 (+5.99%), genotype 22 [Svetlana] (+5.12%), genotype 28 (+5.48%) and genotype 36 (+5.03%), during year 2015 (+6.99%).

Fig. 7

3.3.RFR model

The developed optimal random forest models demonstrated slight better modelling capabilities of the fatty acids and tocopherols, according to the year of production and genotype, which could be realised by Table 3 and Table 4. The number of trees for RFR models were: 920, 1000, 1000, 1000, 1000, 120, 760 and 1000, respectively to acquire the highest values of r^2 (throughout the training cycle r^2 for output variables were: 0.989; 0.989; 0.986; 0.807; 0.823; 0.707; 0.631; 0.671 and 0.654, respectively), Table 4.

Table 4

The potential of the RFR model to predict fatty acids profile (C16:0; C18:0; C18:1; C18:2; C18:3 and C22:1), α-tocopherol, γ-tocopherol and total tocopherols is shown in Fig. 8.

Fig. 8

4. Discussion

4.1.Rapeseed seed quality

Rapeseed oil is a source of essential fatty acids, linoleic (omega-6) and alpha-linolenic acid (omega-3), which the human body cannot create, but need to be ingested through food. The range of oleic and palmitic fatty acids in rapeseed and canola quality seed in Matthaus et al. (2016) was similar to our results. The average content of C18:1 in cultivar Express was around 4% lower compared to the values of the same cultivar grown in the region of Southeast Anatolia (Turkey) (Ozturk et al., 2019). Different environmental conditions affect plant growth and development and reflect on seed quality, mainly in terms of oil and fatty acid content.

The main source of vitamin E in the human diet are vegetable oils, which contain 200–1,000 mg of tocopherols per kilogram of oil (Grilo et al., 2014). Similar to the results of other authors (Matthaus et al., 2016; Siger et al., 2015; Wang et al., 2012), alpha- and gamma-tocopherols were dominant forms of tocopherols in the analysed oil samples. In Grilo et al. (2014), the average concentration of gamma- and alpha-tocopherols in Brazilian canola samples had similar values 122 mg/kg, and 120 mg/kg respectively. On the other hand, Matthaus et al. (2016) determined lower alpha- (13-40%) and gamma-tocopherol content (34-51%). The authors have also detected beta-tocopherol in traces and a small amount of delta-tocopherol.

4.2.Mathematical modelling

Determination of fatty acids and tocopherol content in the traditional way by chromatography is time-consuming and expensive for a large number of samples as is the case in most breeding

programmes. Apart from non-destructive, indirect methods, the use of accurate modelling statistics can significantly enhance this process and provide reliable results. In study of Niedbała and Piekutowska (2018), ANN model with low mean absolute percentage error of 2.81% was built for the quality prediction of potato tuber based on meteorological and fertilizer data. Aćimović et al. (2022) developed simple regression model based on weather data (temperature and precipitations). They succeeded in forecasting the content of active compounds and hydrolate composition in Lavandula essential oil. Random forest regression is also used as modelling tool for the determination of water quality (Islam Khan et al., 2021). Recently, Rajković et al. (2022) used year of production and genotype data as inputs for RFR and ANN models to model the seed yield, oil and protein yield, oil and protein content, and 1000 seed weight in rapeseed. In this study, the RFR model had a slightly better modelling abilities with high values of coefficient of determination in comparison with ANN. Regarding quality traits, ANN model had higher accuracy in modelling oil and protein content (Rajković et al., 2022) than fatty acids and tocopherols in this research.

Both tested models in this study have the potential to be applied to other field crops to determine their seed quality, still RFR model gave slightly superior modelling quality. The best fit of modelled to measured traits in obtained ANN model was observed for C22:1 content (r^2 = 0.952), while the RFR model gave higher r^2 values, the best fit was for C16:0 and C18:0 (r^2 = 0.989) and C18:1 content (r^2 = 0.986). The coefficients of determination for erucic acid content were 0.925 for ANN model, and 0.707 for RFR model. The "goodness of fit" tests for the developed ANN and RFR models also indicate that the RFR model is more favorable in modelling fatty acids (C16:0; C18:0; C18:1; and C18:3), while tocopherols content (α-, γ- and total-tocopherols) was better modelled using ANN model.

Proposed models fit in this background as non-destructive, cost-effective, and environment-friendly. For the best results, it is crucial to train the network with adequate high-quality input

488 data. The modelling of not only fatty acids and tocopherols but also glucosinolates, phenols,
489 and other quality traits can be improved with more supporting input data from the field, e.g.,
490 cultivation practice, soil properties, and genotype such as pedigree information, presence of
491 specific genes, etc.
492
493 **5. Conclusion**
494
495 Apart from yield and oil content, fatty acids and tocopherol content are very important
496 characteristics of rapeseed seed quality. Results of present study indicated strong negative
497 correlation between oleic and erucic acids. This study is the first report of modelling fatty acids
498 and tocopherols of rapeseed oil with the aid of ANN and RFR. Accomplished results are
499 encouraging as the concept of modelling seed quality based on environmental and cultivar data
500 are proved to be effective. Impact of genotype and year on each fatty acid and tocopherol
501 component was assessed in sensitivity analysis. The results of this study disclose that the fatty
502 acids (C16:0; C18:0; C18:1; C18:2; C18:3 and C22:1) and tocopherols content (α-, γ- and total-
503 tocopherols) can be modelled, based on the year of production and genotype, according to
504 relatively low χ^2, RMSE, MBE, MPE, SSE and AARD, as well as the high r^2 values. We
505 demonstrated that the artificial neural network and random forest models are adequate for the
506 modelling of output variables, yet the modelling with RFR model provides slightly better
507 performance. Proposed models can provide relevant results and reduce the costs of standard
508 chromatographic analysis.
509

510 **Acknowledgments**
511

This work was carried out as a part of the activities of the Centre of Excellence for Innovations in Breeding of Climate Resilient Crops–Climate Crops, Institute of Field and Vegetable Crops, Novi Sad, Serbia.

This research was supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia [grant numbers 451-03-68/2022-14/ 200032, 451-03-9/2021-14/200051, and 451-03-9/2021-14/200134].

References

Aćimović, M., Lončar, B., Stanković Jeremić, J., Cvetković, M., Pezo, L., Pezo, M., Todosijević, M., Tešević, V. (2022). Weather Conditions Influence on Lavandin Essential Oil and Hydrolate Quality. *Horticulturae, 8(4),* 281. https://doi.org/10.3390/horticulturae8040281

Adjonu, R., Zhou, Z., Prenzler, P. D., Ayton, J., Blanchard, C. L. (2019). Different Processing Practices and the Frying Life of Refined Canola Oil. *Foods, 8(11),* 527. https://doi.org/10.3390/foods8110527

Agatov, I. (2019). Artificial Neural Networks (ANNs) as a Novel Modeling Technique in Tribology, *Frontiers of Mechanical Engineering*, *5*, 30. https://doi.org/10.3389/fmech.2019.00030

Anitescu, C., Atroshchenko, E., Alajlan, N., Rabczuk, T. (2019). Artificial Neural Network Methods for the Solution of Second Order Boundary Value Problems. *Computers, Materials & Continua, 59(1),* 345-359. https://doi.org/10.32604/cmc.2019.06641.

AOCS Official Methods Ce 1–62, Fatty Acid Composition by Gas Chromatography. American Oil Chemists' Society, IL, USA

Bakre, S. M., Gadmale, D. K., Toche, R. B., Gaikwad, V. B. (2015). Rapid determination of alpha tocopherol in olive oil adulterated with sunflower oil by reversed phase high-performance liquid chromatography. *Journal of Food Science and Technology, 52(5),* 3093-8. doi: 10.1007/s13197-014-1309-7.

Basir, M. S., Chowdhury, M., Islam, M. N., Ashik-E-Rabbania, M. (2021). Artificial neural network model in predicting yield of mechanically transplanted rice from transplanting parameters in Bangladesh, *Journal of Agriculture and Food Research*, *5*, 100186. https://doi.org/10.1016/j.jafr.2021.100186

Baux, A., Colbach, N., Allirand, J. M., Jullien, A., Ney, B., Pellet, D. (2013). Insights into temperature effects on the fatty acid composition of oilseed rape varieties. *European Journal of Agronomy*, *49*, 12–19. https://doi.org/10.1016/j.eja.2013.03.001

Benos, L., Tagarakis, A. C., Dolias, G., Berruto, R., Kateris, D., Bochtis, D. (2021). Machine Learning in Agriculture: A Comprehensive Updated Review. *Sensors, 21(11),* 3758. https://doi.org/10.3390/s21113758

Berrueta, L. A., Alonso-Salces, R. M., Héberger, K. (2007). Supervised pattern recognition in food analysis. *Journal of Chromatography A*, *1158(1–2),* 196-214. https://doi.org/10.1016/j.chroma.2007.05.024

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32. https://doi.org/10.1023/A:1010933404324

Campbell, M. T., Hu, H., Yeats, T. H., Brzozowski, L. J., Caffe-Treml, M., Gutiérrez, L., Smith, K.P., Sorrells, M. E., Gore, M. A., Jannink, J.-L. (2021). Improving Genomic Prediction for Seed Quality Traits in Oat *(Avena sativa* L.) Using Trait-Specific Relationship Matrices. *Frontiers in Genetics*, *12*, 643733. https://doi.org/10.3389/fgene.2021.643733

559 Chattopadhyay, P. B., & Rangarajan, R. (2014). Application of ANN in sketching spatial
560 nonlinearity of unconfined aquifer in agricultural basin. *Agricultural Water Management*,
561 *133*, 81-91. https://doi.org/10.1016/j.agwat.2013.11.007
562 Dehghani, A. A., Mohammadi, Z. B., Maghsoudlou, Y., Mahoonak, A. S. (2012). Intelligent
563 Estimation of the Canola Oil Stability Using Artificial Neural Networks. *Food and*
564 *Bioprocess Technology*, *5*, 533–540. https://doi.org/10.1007/s11947-009-0314-8
565 Doumpos, M., & Zopounidis, C. (2011). Preference disaggregation and statistical learning for
566 multicriteria decision support: A review. *European Journal of Operational Research*,
567 *209(3)*, 203-214. https://doi.org/10.1016/j.ejor.2010.05.029
568 Dutta, J., & Roy, S. (2022). OccupancySense: Context-based indoor occupancy detection &
569 prediction using CatBoost model. *Applied Soft Computing, 119,* 108536.
570 https://doi.org/10.1016/j.asoc.2022.108536
571 Gharechaei, N., Paknejad, F., Shirani Rad, A. H., Tohidloo, G., Jabbari, H. (2019). Change in
572 oil fatty acids composition of winter oilseed rape genotypes under drought stress and
573 different temperature regimes. *Plant, Soil and Environment*, *65*, 503–507.
574 https://doi.org/10.17221/519/2019-PSE
575 Gonzalez-Fernandez, I., Iglesias-Otero, M. A., Esteki, M., Moldes, O. A., Mejuto, J. C., Simal-
576 Gandara, J. (2018). A critical review on the use of artificial neural networks in olive oil
577 production, characterization and authentication. *Critical Reviews in Food Science and*
578 *Nutrition*, *59(12)*, 1913–1926. https://doi.org/10.1080/10408398.2018.1433628
579 Grilo, E. C., Costa, P. N., Gurgel, C. S. S., Beserra, A. F. de L., Almeida, F. N. de S.,
580 Dimenstein, R. (2014). Alpha-tocopherol and gamma-tocopherol concentration in
581 vegetable oils. *Food Science and Technology*, *34(2)*, 379–385.
582 https://doi.org/10.1590/S0101-20612014005000031

583 Gruszka, J., & Kruk, J. (2007). RP-LC for determination of plastochromanol, tocotrienols and
584 tocopherols in plant oils. *Chromatographia*, *66*, 909–913.
585 Imahara, H., Minami, E., Saka, S. (2006). Thermodynamic study on cloud point of biodiesel
586 with its fatty acid composition. *Fuel*, *85(12-13)*, 1666–1670.
587 https://doi.org/10.1016/j.fuel.2006.03.003
588 Iniyan, S., Jebakumar, R., Mangalraj, P., Mohit, M., Nanda, A. (2020). Plant Disease
589 Identification and Detection Using Support Vector Machines and Artificial Neural
590 Networks, In: S. Dash, C. Lakshmi, S. Das & B. Panigrahi (Eds.), *Artificial Intelligence*
591 *and Evolutionary Computations in Engineering Systems*. *Advances in Intelligent Systems*
592 *and Computing* 1056, (pp. 15–27). Springer, Singapore. https://doi.org/10.1007/978-981-
593 15-0199-9_2
594 Islam Khan, S. I., Islam, N., Uddin, J., Islam, S., Nasir, M. K. (2021). Water quality prediction
595 and classification based on principal component regression and gradient boosting
596 classifier approach, *Journal of King Saud University - Computer and Information*
597 *Sciences, in press*. https://doi.org/10.1016/j.jksuci.2021.06.003
598 Jawad, M., Ghulam-e, M., Muhammad, A. (2022). Accurate estimation of tool wear levels
599 during milling, drilling and turning operations by designing novel hyperparameter tuned
600 models based on LightGBM and stacking, *Measurement*, *190*, 110722.
601 https://doi.org/10.1016/j.measurement.2022.110722
602 Khanal, S., Fulton, J., Klopfenstein, A., Douridas, N., Shearer, S. (2018). Integration of high
603 resolution remotely sensed data and machine learning techniques for spatial prediction of
604 soil properties and corn yield. *Computers and Electronics in Agriculture*, *153*, 213–225.
605 https://doi.org/10.1016/j.compag.2018.07.016
606 Kleijnen, J. P. C. (2018). Design and Analysis of Simulation Experiments. In: J. Pilz, D. Rasch,
607 V.Melas, & K. Moder, (Eds.), *Statistics and Simulation*. IWS 2015, (pp. 3-22). Springer

608 Proceedings in Mathematics & Statistics, 231, Springer, Cham.
609 https://doi.org/10.1007/978-3-319-76035-3_1
610 Kollo, T., & von Rosen, D. (2005). Advanced Multivariate Statistics with Matrices, in
611 Hazewinkel, M. (Ed.), *Mathematics and Its Applications*. *Vol. 579*, (pp. 1–485). Springer:
612 Dordrecht, The Netherlands. https://doi.org/10.1007/1-4020-3419-9
613 Koprna, R., Nerušıl, P., Kolovrat, O., Kučera, V., Kohoutek, A. (2006). Estimation of fatty acid
614 content in intact seeds of oilseed rape *(Brassica napus* L.) lines using near-infrared
615 spectroscopy. *Czech Journal of Genetics and Plant Breeding*, *42(4)*, 132–136.
616 https://doi.org/10.17221/3643-CJGPB
617 Kravić, S., Suturović, Z., Švarc-Gajić, J., Stojanović, Z., Pucarević, M. (2010). Determination
618 of trans fatty acids in foodstuffs by gas chromatography-mass spectrometry after
619 simultaneous microwave assisted extraction-esterification. *Journal of the Serbian*
620 *Chemical Society*, *75(6)*, 803–812. https://doi.org/10.2298/JSC090717051K
621 Kujawa, S., & Niedbała, G. (2021). Artificial Neural Networks in Agriculture. *Agriculture, 11*,
622 497. https://doi.org/10.3390/agriculture1106097
623 Lazzez, A., Perri, E., Caravita, M.A., Khlif, M., Cossentini, M. (2008). Influence of olive
624 maturity stage and geographical origin on some minor components in virgin olive oil of
625 the Chemlali variety. *Journal of Agricultural and Food Chemistry*, *53*, 982-988.
626 https://doi.org/10.1021/jf0722147
627 Lebold, K. M., & Traber, M. G. (2014). Interactions between α-tocopherol, polyunsaturated
628 fatty acids, and lipoxygenases during embryogenesis. *Free Radical Biology and*
629 *Medicine*, *66*, 13–19. https://doi.org/10.1016/j.freeradbiomed.2013.07.039
630 Ma, H., Ding, F., Wang, Y. (2022). A novel multi-innovation gradient support vector machine
631 regression method, *ISA Transactions, in press*.
632 https://doi.org/10.1016/j.isatra.2022.03.006

633 Matthaus, B., Özcan, M. M., Al Juhaimi, F. (2016). Some rape/canola seed oils: Fatty acid
634 composition and tocopherols. *Zeitschrift für Naturforschung C, 71(3-4)*, 73–77,
635 https://doi.org/10.1515/znc-2016-0003
636 Montgomery, D. C. (1984). Design and Analysis of Experiments (2nd ed). New York, USA:
637 John Wiley and Sons Inc. https://doi.org/10.1002/qre.4680030319
638 Niedbała, G., Piekutowska, M. (2018). Application of artificial neural networks for the
639 prediction of quality characteristics of potato tubers – Innovator variety. *Journal of*
640 *Research and Applications in Agricultural Engineering*, *63(4)*, 132-138.
641 https://www.pimr.eu/wp-content/uploads/2019/05/2018_4_GNMP.pdf
642 Niedbała, G., Piekutowska, M., Weres, J., Korzeniewicz, R., Witaszek, K., Adamski, M.,
643 Pilarski, K. Czechowska-Kosacka, A., Krysztofiak-Kaniewska, A. (2019). Application of
644 Artificial Neural Networks for Yield Modeling of Winter Rapeseed Based on Combined
645 Quantitative and Qualitative Data. *Agronomy*, *9*, 781.
646 https://doi.org/10.3390/agronomy9120781
647 Niedbała, G., Kurasiak-Popowska, D., Stuper-Szablewska, K., Nawracała, J. (2020).
648 Application of Artificial Neural Networks to Analyze the Concentration of Ferulic Acid,
649 Deoxynivalenol, and Nivalenol in Winter Wheat Grain. *Agriculture*, *10(4)*, 127.
650 https://doi.org/10.3390/agriculture10040127
651 Opinion of the Scientific Panel on Contaminants in the Food Chain on a request from the
652 European Commission on glucosinolates as undesirable substances in animal feed
653 Glucosinolates as undesirable substances in animal feed - Scientific Opinion of the Panel
654 on Contaminants in the Food Chain, (2008). *EFSA Journal*, *590*, 1-76.
655 https://doi.org/10.2903/j.efsa.2008.590
656 Ozturk, F. (2019). Evaluation of three canola *(Brassica napus* L.) cultivars for yield and some
657 quality parameters under the environmental condition of southeastern Anatolia, Turkey.

Applied Ecology and Environmental Research, *17*, 2167-2177. http://dx.doi.org/10.15666/aeer/1702_21672177

Pezo, L., Ćurčić, B. Lj., Filipović, V. S., Nićetin, M. R., Koprivica, G. B., Mišljenović, N. M., Lević, Lj. B. (2013). Artificial neural network model of pork meat cubes osmotic dehydratation. *Chemical Industry*, *67*, 465-475. https://doi.org/10.2298/HEMIND120529082P

Puntarić, E., Pezo, L., Zgorelec, Ž., Gunjača, J., Kučić Grgić, D., Voća, N. (2022). Prediction of the Production of Separated Municipal Solid Waste by Artificial Neural Networks in Croatia and the European Union. *Sustainability*, *14*, 10133. https://doi.org/10.3390/su141610133

Rajković, D., Marjanović Jeromela, A., Pezo, L., Lončar, B., Zanetti, F., Monti, A., Kondić Špika, A. (2022). Yield and Quality Prediction of Winter Rapeseed—Artificial Neural Network and Random Forest Models. *Agronomy*, *12(1)*, 58. https://doi.org/10.3390/agronomy12010058

Rasaei, Z., & Bogaert, P. (2019). Spatial filtering and Bayesian data fusion for mapping soil properties: A case study combining legacy and remotely sensed data in Iran. *Geoderma*, *344*, 50–62. https://doi.org/10.1016/j.geoderma.2019.02.031

Samaniego, E., Anitescu, C., Goswami, S., Nguyen-Thanh, V.M., Guo, H., Hamdia, K., Zhuang, X., Rabczuk, T. (2020). An energy approach to the solution of partial differential equations in computational mechanics via machine learning: Concepts, implementation and applications. *Computer Methods in Applied Mechanics and Engineering*, *362*, 112790. https://doi.org/10.1016/j.cma.2019.112790.

Shirani Rad, A. H., Bitarafan, Z., Rahmani, F., Taherkhani, T., Moradi Aghdam, A., Nasresfahani, S. (2014). Effects of planting date on spring rapeseed (*Brassica napus* L.)

682 cultivars under different irrigation regimes. *Turkish Journal of Field Crops, 19(2)*, 153-
683 157. http://www.field-crops.org/assets/pdf/product5496955d5a795.pdf
684 Schulte, L. R., Ballard, T., Samarakoon, T., Yao, L., Vadlani, P., Staggenborg, S., Rezac, M.
685 (2013). Increased growing temperature reduces content of polyunsaturated fatty acids in
686 four oilseed crops. *Industrial Crops and Products*, *51*, 212–219.
687 https://doi.org/10.1016/j.indcrop.2013.08.075
688 Siger, A., Michalak, M., Cegielska-Taras, T., Szała, L., Lembicz, J., Nogala-Kałucka, M.
689 (2015). Genotype and environment effects on tocopherol and plastochromanol-8 contents
690 of winter oilseed rape doubled haploid lines derived from F1 plants of the cross between
691 yellow and black seeds. *Industrial Crops and Products*, *65*, 134–141.
692 https://doi.org/10.1016/j.indcrop.2014.12.006
693 Su, J., Wang, Y., Niu, X., Shaa, S., Yu, J. (2022). Prediction of ground surface settlement by
694 shield tunneling using XGBoost and Bayesian Optimization. *Engineering Applications of*
695 *Artificial Inteligence*, *114*, 105020. https://doi.org/10.1016/j.engappai.2022.105020
696 Vetter, W., Darwisch, V., Lehnert, K. (2020). Erucic acid in *Brassicaceae* and salmon – An
697 evaluation of the new proposed limits of erucic acid in food. *Nutrition and Food Science*
698 *Journal*, *19*, 9–15. https://doi.org/10.1016/j.nfs.2020.03.002
699 Wang, X., Zhang, C., Li, L., Fritsche, S., Endrigkeit, J., Zhang, W., Long, Y., Jung, C., Meng,
700 J. (2012). Unraveling the genetic basis of seed tocopherol content and composition in
701 rapeseed (*Brassica napus* L.). *PLoS one*, *7(11)*, 1–15.
702 https://doi.org/10.1371/journal.pone.0050038
703 Wang, C-, Peng, G., De Baets, B. (2022). Embedding metric learning into an extreme learning
704 machine for scene recognition, *Expert Systems with Applications*, *203*, 117505.
705 https://doi.org/10.1016/j.eswa.2022.117505

Yang, J., Xu, Y. (2021). Prediction of fruit quality based on the RGB values of time–temperature indicator. *Journal of Food Science*, *86(3)*, 932-941. https://doi.org/10.1111/1750-3841.15518

Yoon, Y., Swales, G., Margavio, T. M. (1993). A Comparison of Discriminant Analysis versus Artificial Neural Networks. *Journal of the Operational Research Society, 44(1)*, 51–60. https://doi.org/10.1057/jors.1993.6

Yu, F., Feng, S., Du, W., Wang, D., Guo, Z., Xing, S., Jin, Z., Cao, Y., Xu, T. (2020). A Study of Nitrogen Deficiency Inversion in Rice Leaves Based on the Hyperspectral Reflectance Differential. *Frontiers in Plant Science*, *11*, 573272. https://doi.org/10.3389/fpls.2020.573272

Zhang, L., Yang, L., Ma, T., Shen, F., Cai, Y., Zhou, C. (2021). A self-training semi supervised machine learning method for predictive mapping of soil classes with limited sample data. *Geoderma, 384*, 114809. https://doi.org/10.1016/j.geoderma.2020.114809

Figure captions

Fig. 1. Average monthly **a)** temperatures, **b)** precipitations and **c)** sunshine hours in period 2014-2018

Months are written in Roman numerals starting from August (VIII); MYA multi year average

Fig. 2. Bar plots showing proportion of individual fatty acids in analyzed genotypes in period 2015-2018 [**(a)** 2015, **(b)** 2016, **(c)** 2017, **(d)** 2018]. C16:0 palmitic acid; C18:0 stearic acid; C18:1 oleic acid; C18:2 linoleic acid; C18:3 linolenic acid; C22:1 erucic acid

Fig. 3. Bar plots showing proportion of alpha (AT) and gamma tocopherols (GT) in analyzed genotypes in period 2015-2018 [**(a)** 2015, **(b)** 2016, **(c)** 2017, **(d)** 2018].

Fig. 4. Correlation heatmap (a) with hierarchical clustering (b) of fatty acids and tocopherols. Coefficients of correlation are presented on side panel (a), positive correlations are labeled with blue while negative correlations are labeled with red colour. Colour intensity indicates strength of correlation (a, b). C16:0 palmitic acid; C18:0 stearic acid; C18:1 oleic acid; C18:2 linoleic acid; C18:3 linolenic acid; C22:1 erucic acid

Fig. 5. Comparison between experimentally obtained and ANN model predicted values of (**a**) C16:0, (**b**) C18:0, (**c**) C18:1, (**d**) C18:2, (**e**) C18:3, (**f**) C22:1, (**g**) α-tocopherol, (**h**) γ-tocopherol and (**i**) total tocopherols

Fig. 6. Training graph for MLP 44-10-9 network

Fig. 7. The relative importance of the input variables on outputs, determined using Yoon interpretation method. Genotype number is presented in Supplementary Table 1.

Fig. 8. Comparison between experimentally obtained and RFR model predicted values of (**a**) C16:0, (**b**) C18:0, (**c**) C18:1, (**d**) C18:2, (**e**) C18:3, (**f**) C22:1, (**g**) α-tocopherol, (**h**) γ-tocopherol and (**i**) total tocopherols

Figure 1

Click here to access/download;Figure;Fig 1.tif

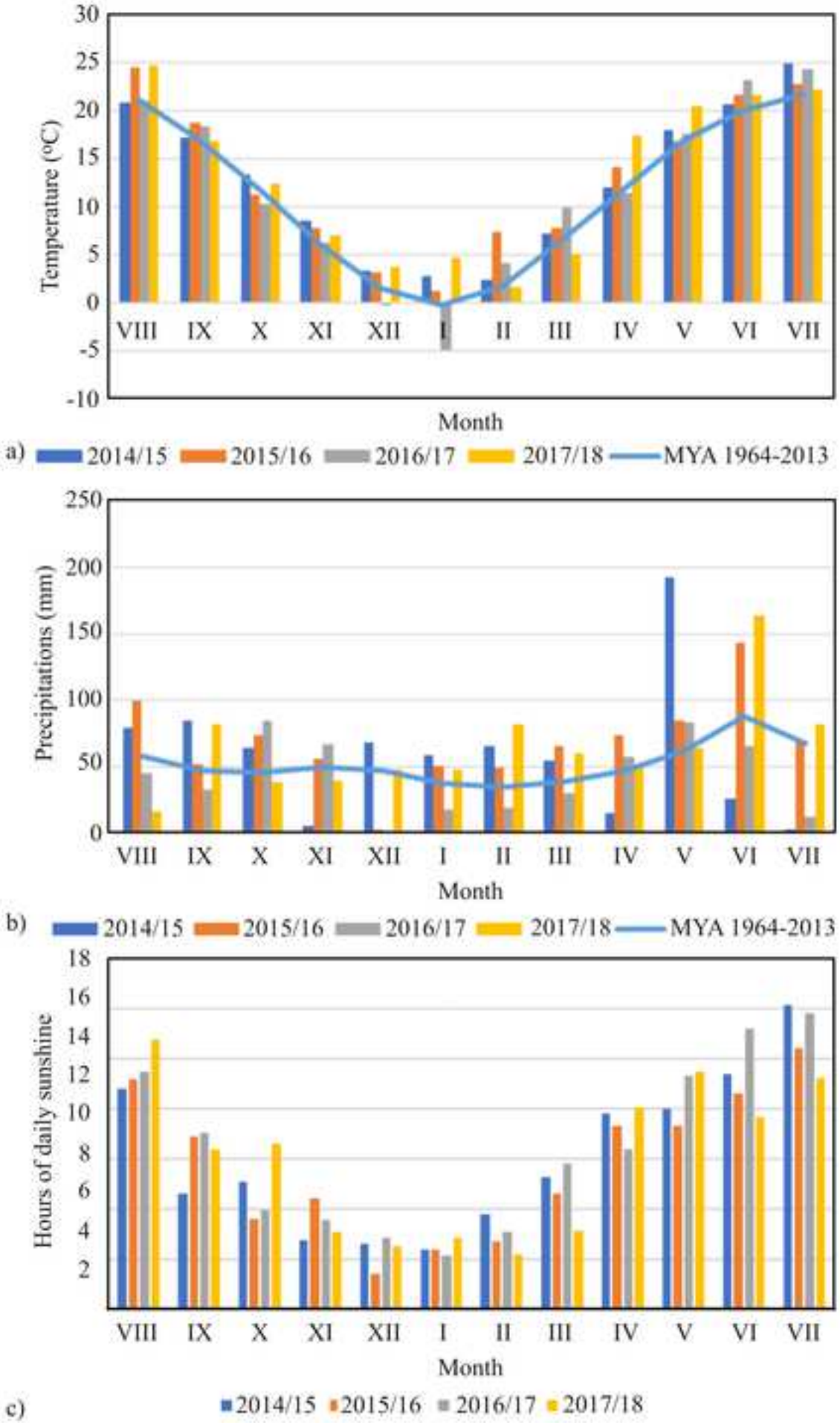


Figure 2

Click here to access/download;Figure;Fig 2 Rev 1.TIF

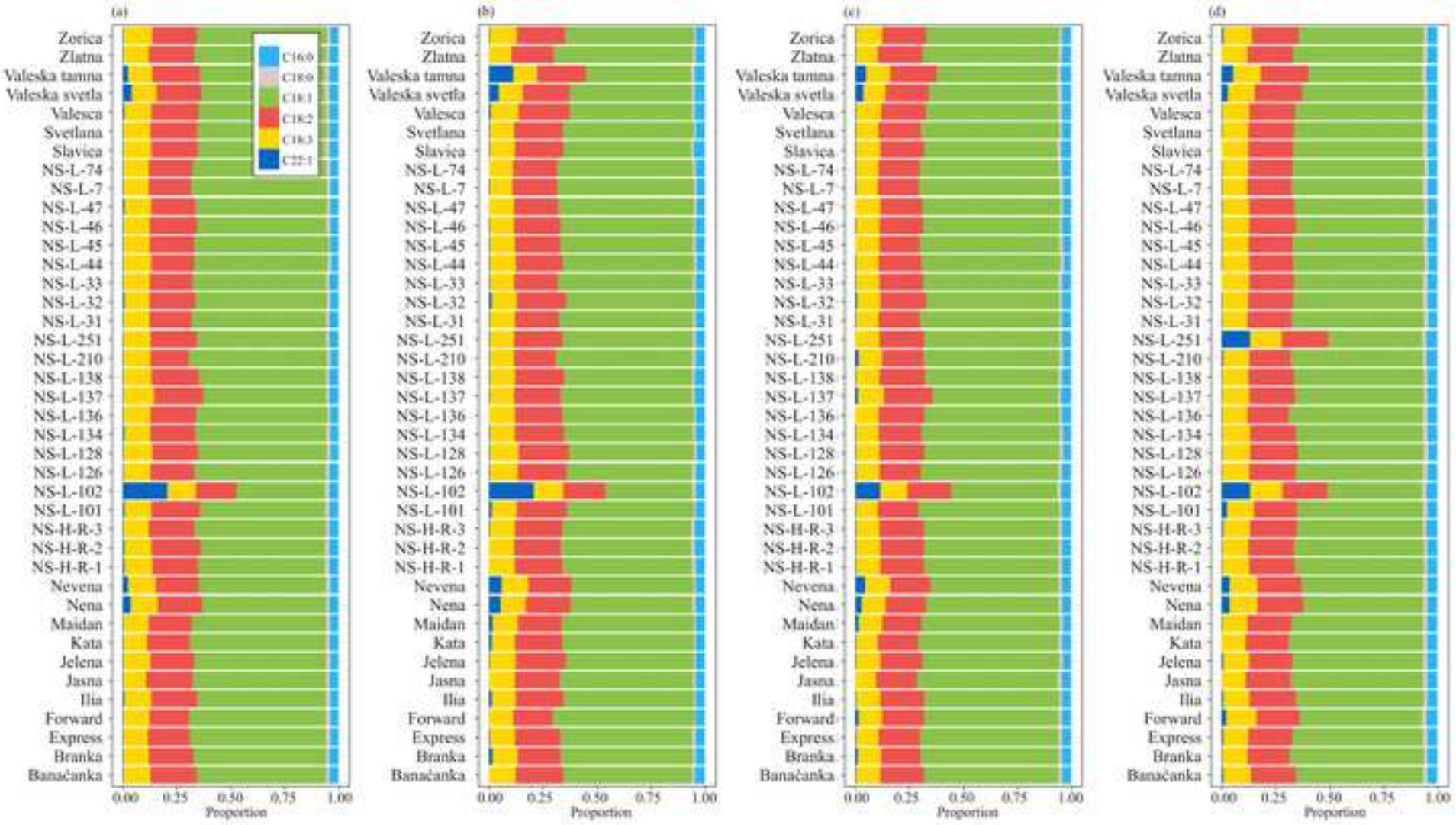


Figure 3

Click here to access/download;Figure;Fig 3 Rev 1.TIF

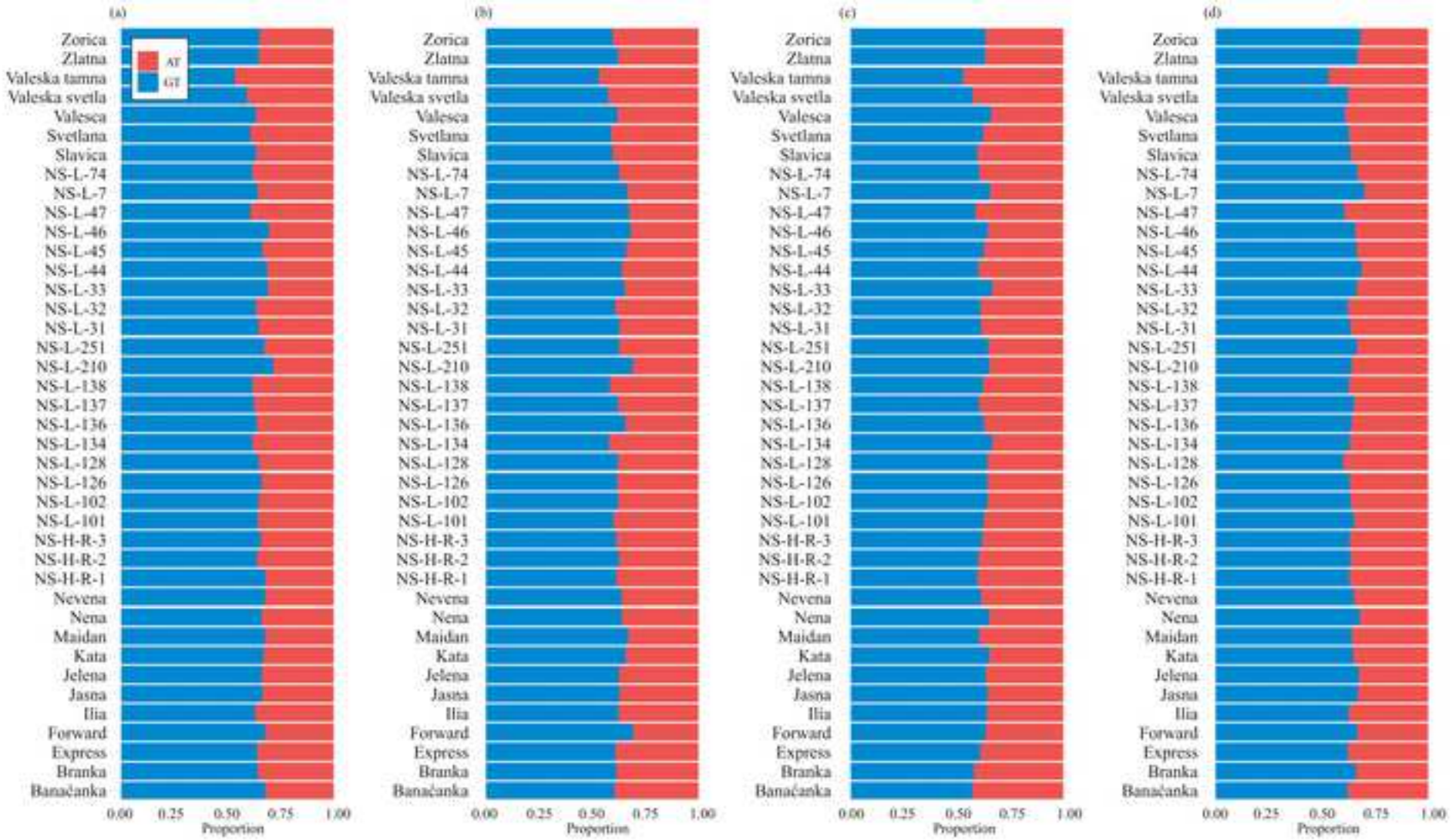
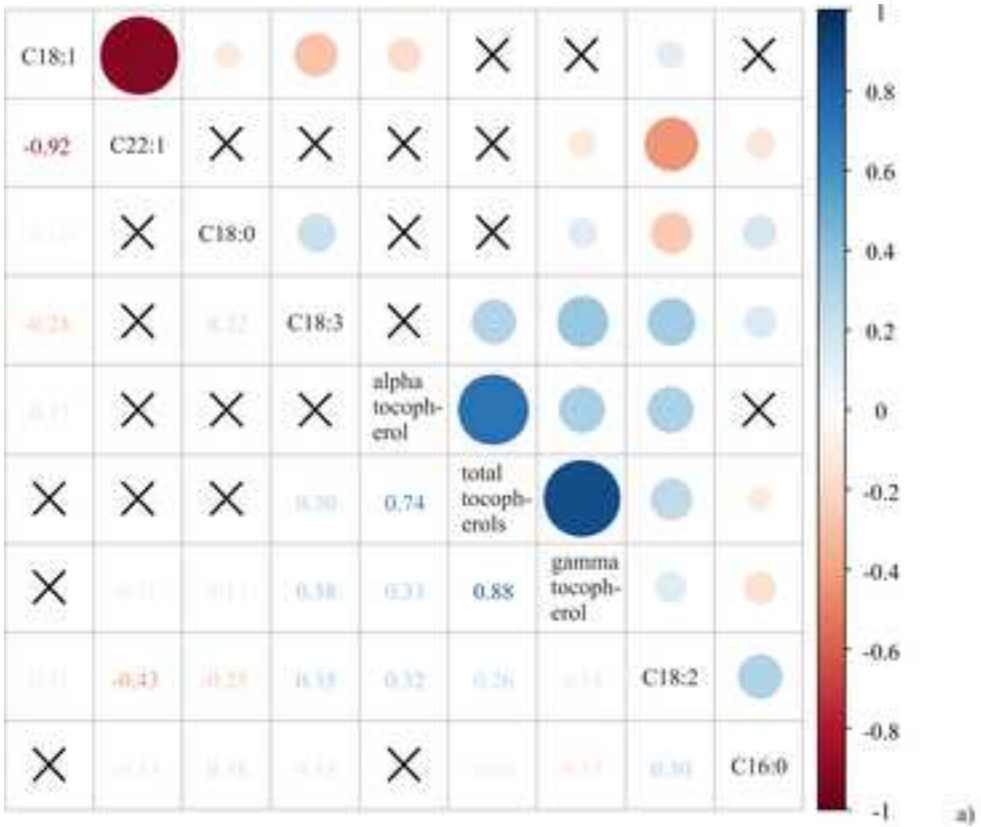


Figure 4

Click here to access/download;Figure;Fig. 4 Correlation.tif



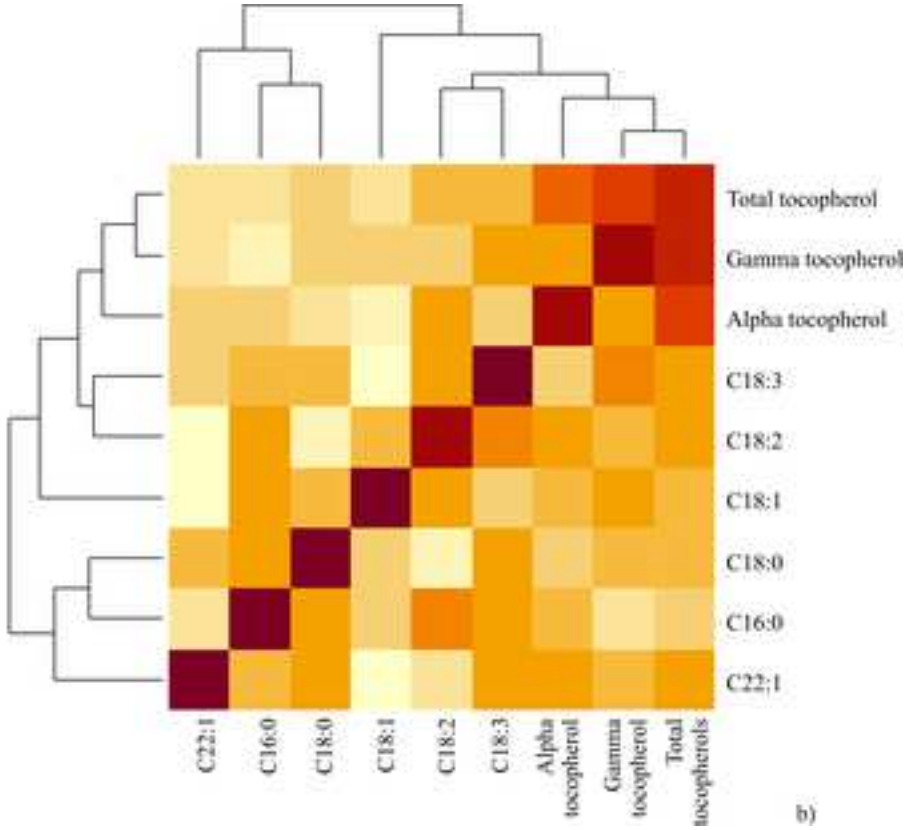


Figure 5

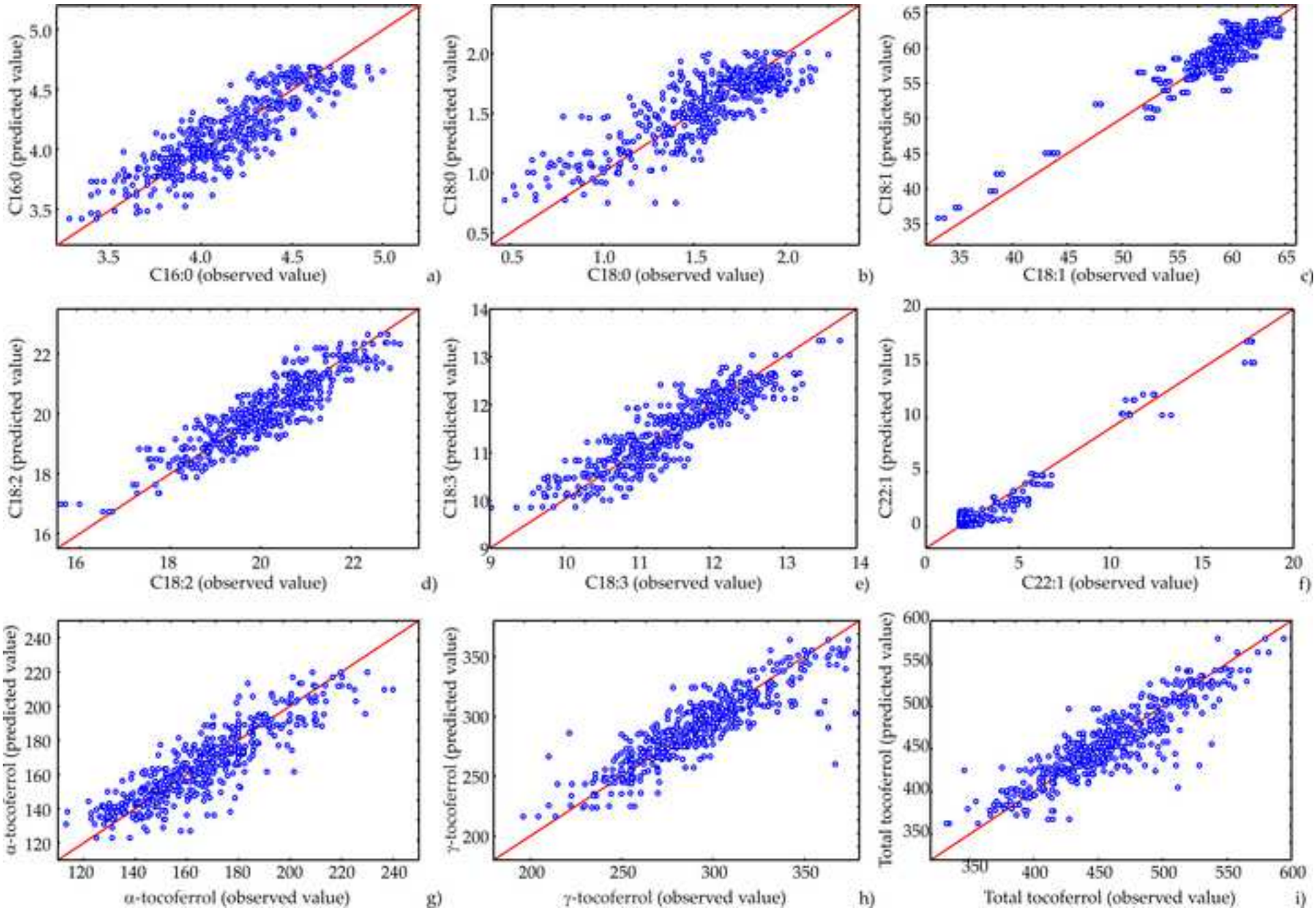


Figure 6

Click here to access/download;Figure;Fig. 6.tif

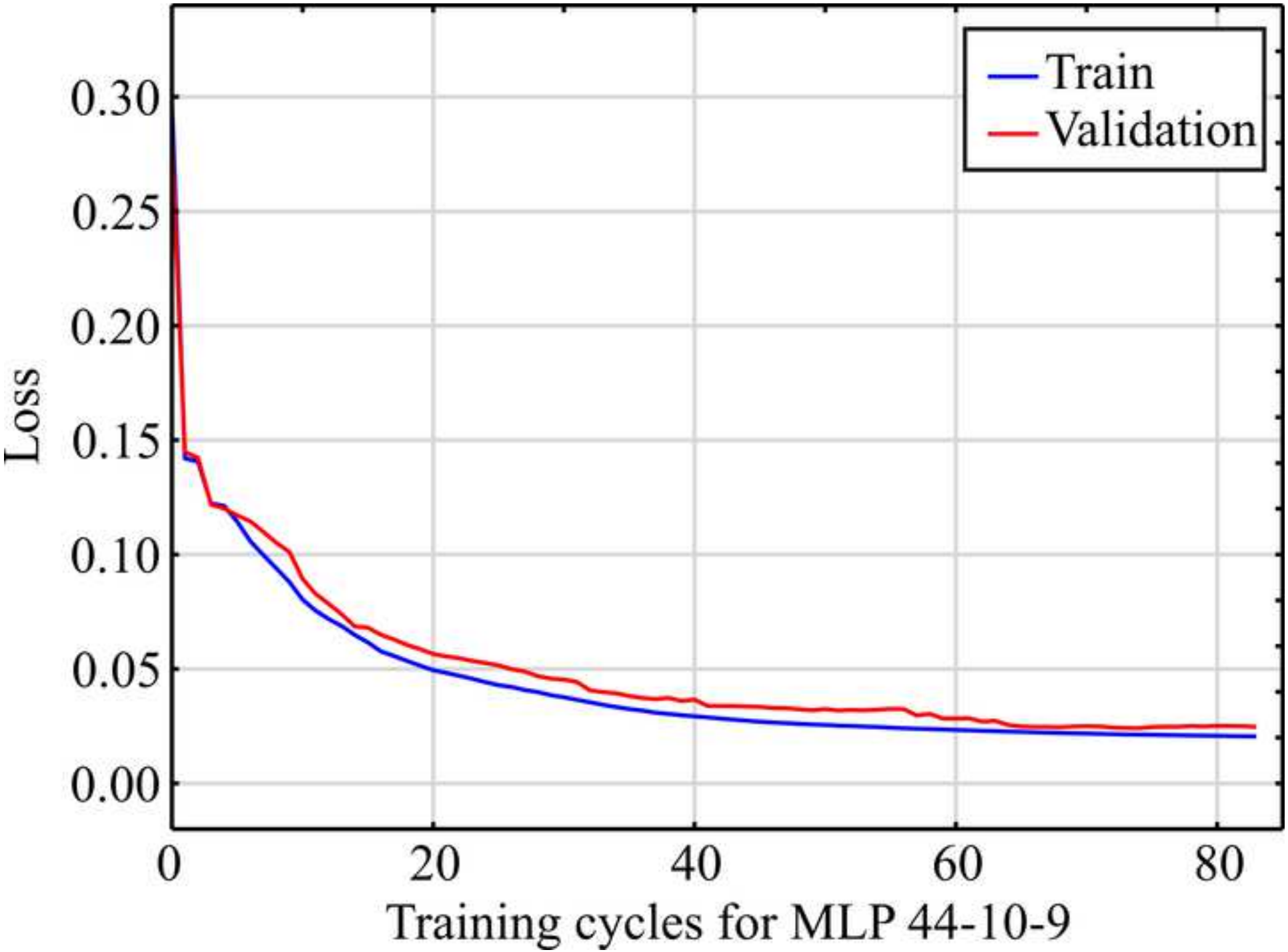


Figure 7

Click here to access/download;Figure;Fig. 7.tif

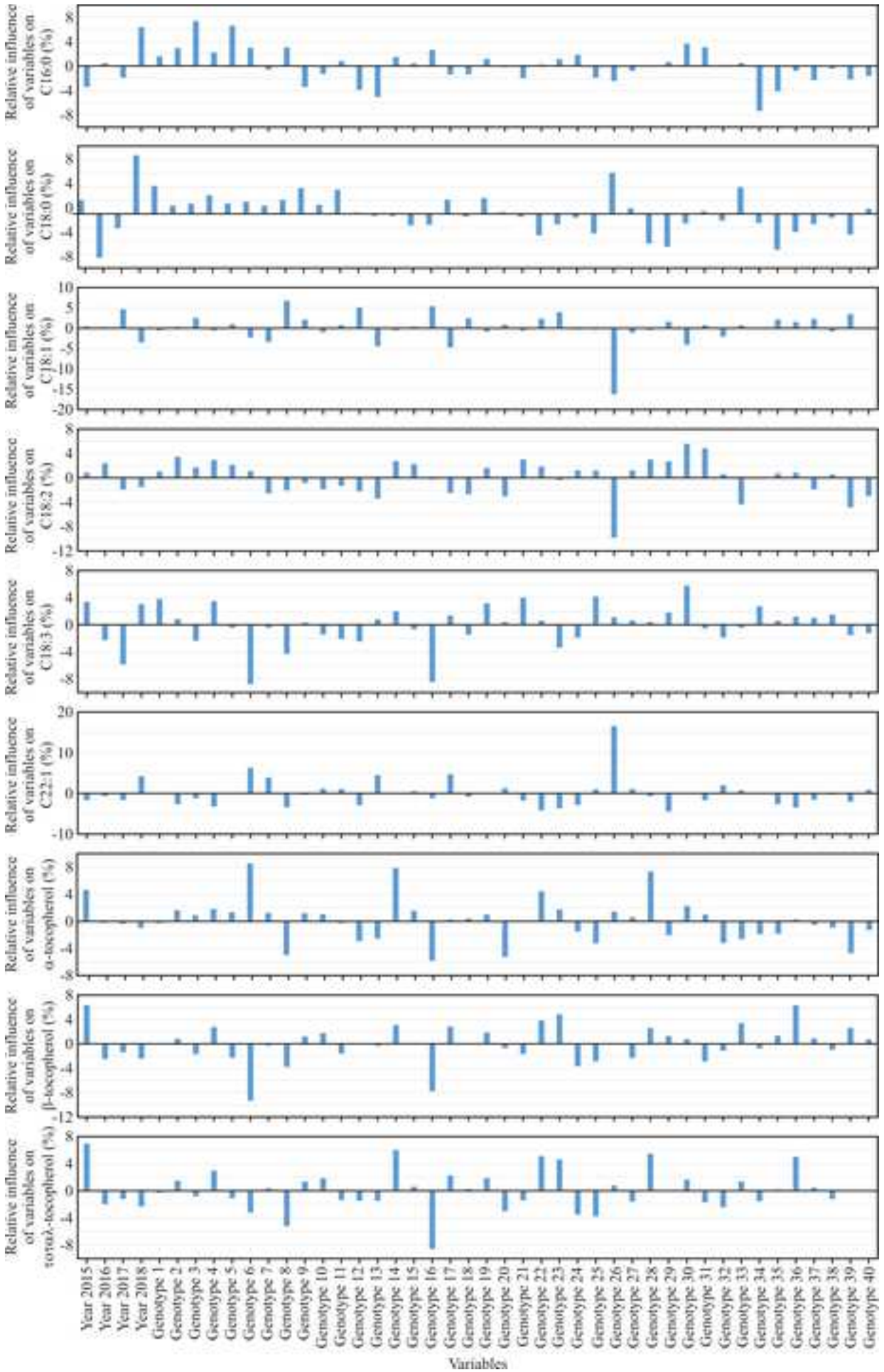


Figure 8

Click here to access/download;Figure;Fig. 8 RFR.tif

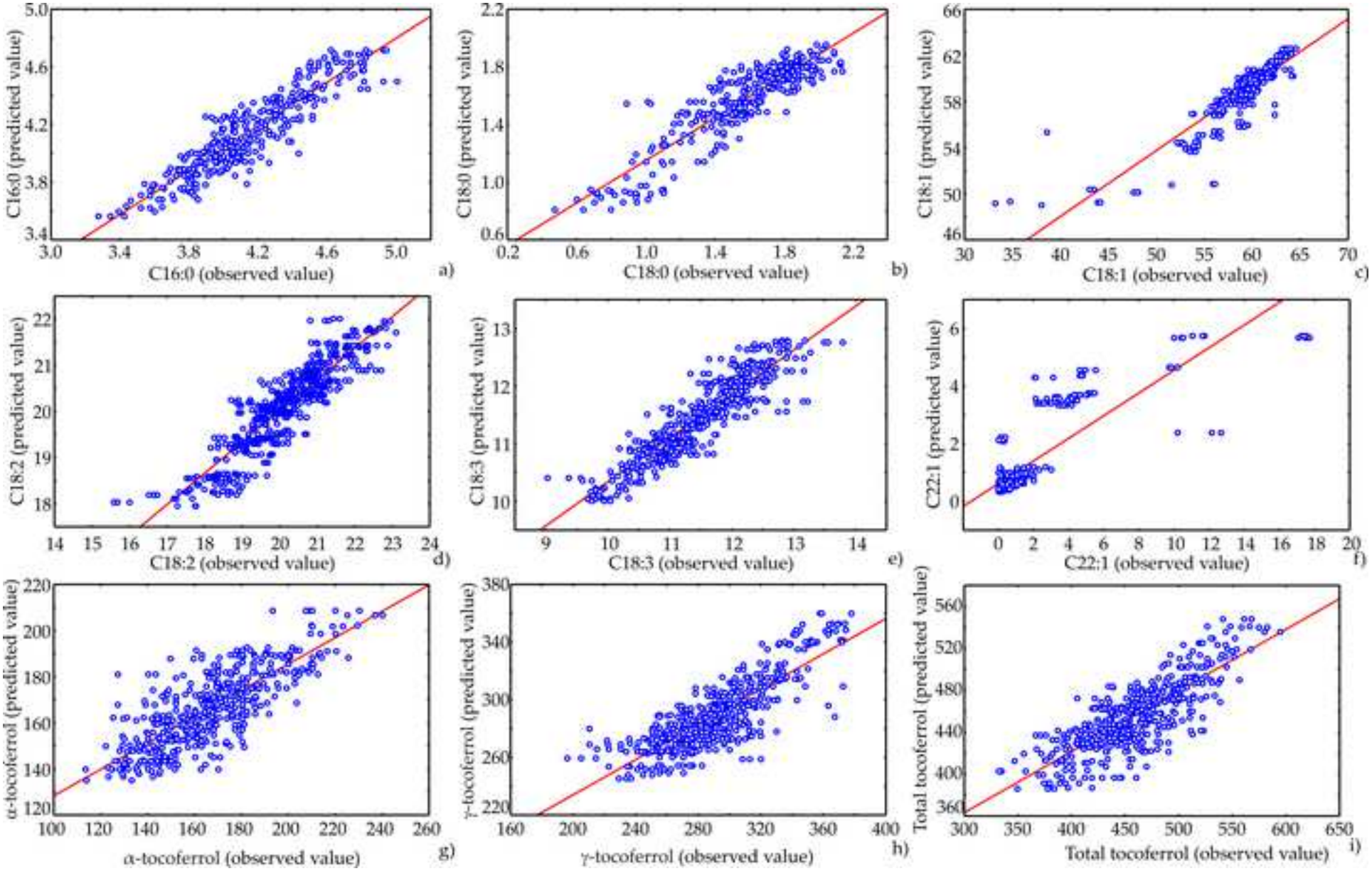


Table 1. Dosage of applied NPK fertilizer

	NPK[1] ratio	Dosage of fertilizer (kg/ha)
2014	1:0.1:0.1	250
2015	09:15:15	450
2016	08:15:15	350
2017	16:16:16	350

[1]NPK nitrogen, phosphorus, potassium

Table 2. Artificial neural network model summary (performance and errors), for training, testing, and validation cycles

Network name	Performance		Error		Training algorithm	Error function	Hidden activation	Output activation
	Train.	Test.	Train.	Test.				
MLP 44-10-9	0.858	0.722	239.273	698.266	BFGS 63	SOS	Tanh	Logistic

*Performance term represents the coefficients of determination, while error terms indicate a lack of data for the ANN model, Train. stands for training, Test. for testing, BFGS is Broyden–Fletcher–Goldfarb–Shanno algorithm, Tanh - hyperbolic tangent function, Logistic – logistic function.

Table 3. The "goodness of fit" tests for the developed ANN model

Output variable	χ^2	RMSE	MBE	MPE	SSE	AARD	r^2
C16:0	0.027	0.163	0.012	3.187	12.763	126.917	0.769
C18:0	0.036	0.188	-0.010	11.015	16.955	92.873	0.693
C18:1	2.707	1.630	-0.189	2.180	1.3×10^3	6.9×10^2	0.887
C18:2	0.315	0.556	0.031	2.239	147.802	234.491	0.812
C18:3	0.135	0.365	0.010	2.583	63.739	268.693	0.806
C22:1	0.355	0.590	0.040	354.471	166.430	238.442	0.952
α-tocopherol	1.3×10^2	11.335	0.882	5.262	6.1×10^4	7.1×10^3	0.763
γ-tocopherol	2.6×10^2	15.834	0.948	3.900	1.2×10^5	9400	0.770
Total tocopherols	4.7×10^2	21.518	1.874	3.452	2.2×10^5	7900	0.783

χ^2 – reduced chi-square; RMSE - root mean square error; MBE – mean bias error; MPE – mean percentage error; SSE – sum of squared errors; AARD – absolute average relative deviation; r^2 – coefficient of determination; C16:0 palmitic acid; C18:0 stearic acid; C18:1 oleic acid; C18:2 linoleic acid; C18:3 linolenic acid; C22:1 erucic acid; α-T alpha-tocopherol; γ-T gamma-tocopherol.

Table 4. The "goodness of fit" tests for the developed RFR model

Output variable	χ^2	RMSE	MBE	MPE	SSE	AARD	r^2
C16:0	770.384	27.494	3.697	4.028	356291.711	5431.315	0.989
C18:0	770.386	27.494	3.695	7.890	356296.887	5432.765	0.989
C18:1	773.680	27.553	3.735	3.704	3.6×10^5	5.8×10^3	0.986
C18:2	0.356	0.591	0.021	2.267	167.321	237.465	0.807
C18:3	0.126	0.352	0.014	2.300	59.416	140.972	0.823
C22:1	3.105	1.746	0.129	536.263	1454.465	305.157	0.707
α-tocopherol	2.0×10^2	14.122	0.089	6.767	9.6×10^4	1.5×10^4	0.631
γ-tocopherol	3.7×10^2	19.078	1.572	5.115	1.7×10^5	1.4×10^4	0.671
Total tocopherols	7.7×10^2	27.434	2.685	4.740	3.6×10^5	1.0×10^4	0.654

χ^2 – reduced chi-square; RMSE - root mean square error; MBE – mean bias error; MPE – mean percentage error; SSE – sum of squared errors; AARD – absolute average relative deviation; r^2 – coefficient of determination.

Declaration of interests

☒The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Dragana Rajković Conceptualization, Methodology, Formal analysis, Investigation, Writing - Original Draft, Writing - Review & Editing, Visualization **Ana Marjanović Jeromela** Resources, Writing - Review & Editing, Supervision, Project administration, Funding acquisition **Lato Pezo** Conceptualization, Methodology, Software, Validation, Formal analysis, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization **Biljana Ćurčić** Methodology, Software, Validation, Formal analysis, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization **Nada Grahovac** Methodology, Formal analysis, Writing - Review & Editing **Ankica Kondić Špika** Writing - Review & Editing

Supplementary Table 1. Mean genotypic values for fatty acids (%) and tocopherols content (mg/kg) during four-year period.

No.	Genotype	C18:1	C18:2	C18:3	C16:0	C18:0	C22:1	α-T	γ-T	Total-T
1	NS-H-R-1	59.43 no	20.80 opq	11.99 opq	4.21 mnopq	1.56 fghijk	0.31 abcde	164.31 ijkl	282.85 fghi	447.16 defgh
2	NS-H-R-2	59.08 klm	21.13 rs	11.53 ijkl	4.40 rs	1.62 hijk	0.41 abcdef	172.82 mno	291.49 ijkl	464.31 jklm
3	NS-H-R-3	59.02 kl	20.77 op	11.37 fghi	4.46 s	1.54 cdefghijk	0.52 defg	179.20 nop	305.80 mn	485.00 opq
4	Banaćanka	58.73 ij	20.79 pq	12.01 opq	4.33 qrs	1.62 ghijk	0.38 abcdef	183.82 pqr	303.71 mn	487.53 q
5	Slavica	59.07 klm	20.90 op	11.83 mno	4.57 t	1.60 ghijk	0.26 abcd	187.54 qr	301.65 lmn	489.19 q
6	Valeska tamna	49.72 b	20.39 jkl	10.54 a	4.35 qrs	1.63 hijk	5.50 n	218.51 s	249.95 a	468.46 klmn
7	Valeska svetla	54.45 c	19.92 hi	10.73 bc	4.23 klmnopq	1.61 ghijk	3.59 l	182.85 pqr	267.86 bc	450.71 efghij
8	Zlatna	61.13 v	20.30 jk	10.65 ab	4.30 opqr	1.57 fghijk	0.23 abc	155.35 efgh	279.86 efgh	435.21 cde
9	NS-L-74	61.43 wx	19.63 f	11.05 d	3.95 bcde	1.66 jk	0.23 abc	163.69 ijkl	278.84 defgh	442.53 defg
10	Branka	60.64 st	19.02 c	11.27 efg	4.16 hijklmn	1.65 jk	0.84 hi	165.97 jklm	279.46 defgh	445.43 defg
11	Express	60.94 u	19.39 e	11.12 de	4.20 klmnop	1.80 l	0.41 abcdef	165.27 jklm	271.38 cde	436.65 cdef
12	NS-L-7	61.90 y	19.40 e	10.85 c	3.89 bc	1.59 fghijk	0.31 abcde	137.64 a	277.84 cdefgh	415.49 b
13	Nevena	54.69 d	18.48 b	11.69 klm	3.94 bcde	1.53 cdefghijk	3.88 m	145.77 b	268.67 bcd	414.43 b
14	Valesca	57.63 e	21.28 tu	12.09 qr	4.29 opqr	1.54 cdefghijk	0.60 fgh	188.76 r	325.42 o	514.18 r
15	Ilia	59.28 mn	20.72 nop	11.50 hijk	4.07 efghij	1.48 bcdefgh	0.72 ghi	172.73 mno	295.44 jklm	468.18 klmn
16	Kata	61.50 x	19.68 fg	10.54 a	4.24 lmnopq	1.56 bcdefghij	0.44 bcdef	135.93 a	261.13 b	397.06 a

17	Nena	54.76 d	19.28 de	11.45 ghij	4.11 fghijk	1.67 k	3.72 lm	154.23 defg	297.68 klmn	451.91 fghij
18	NS-L-31	61.37 wx	19.34 e	11.28 efg	4.10 fghijk	1.54 cdefghijk	0.28 abcde	165.36 jklm	286.24 ghij	451.59 fghij
19	NS-L-126	59.71 pq	20.52 lm	12.08 pqr	4.25 mnopq	1.64 ijk	0.18 ab	172.22 mn	307.33 n	479.55 mnopq
20	NS-L-33	60.58 tu	20.03 i	11.43 ghij	4.26 nopq	1.61 ghijk	0.26 abcd	147.47 bcd	301.09 lmn	448.57 defghi
21	NS-L-128	58.49 gh	21.45 tu	12.34 st	3.94 bcde	1.49 bcdefghi	0.33 abcde	167.45 klm	287.85 hijk	455.30 ghijk
22	Svetlana	60.27 s	20.72 nop	11.49 hij	4.18 hijklmn	1.44 bcdef	0.24 abc	189.74 r	300.31 lmn	490.05 q
23	Jasna	61.48 x	20.29 jk	10.49 a	4.32 pqr	1.57 efghijk	0.13 a	163.30 ijkl	308.56 n	471.86 lmnop
24	NS-L-101	58.34 g	20.48 klm	11.26 efg	4.14 fghijklmn	1.53 bcdefghijk	1.10 k	173.12 mno	297.75 klmn	470.88 klmno
25	Zorica	58.57 hi	20.67 mno	12.42 t	4.06 efghi	1.41 bcd	0.54 efg	153.16 cde	281.43 efghi	434.59 cd
26	NS-L-102	37.59 a	17.08 a	11.70 lm	3.85 b	1.66 jk	14.13 o	161.50 ghijk	284.75 ghij	446.25 defg
27	NS-L-134	59.52 op	20.50 lm	11.46 ghij	4.03 defg	1.56 efghijk	0.50 cdefg	170.58 lm	285.26 ghij	455.84 ghijk
28	NS-L-32	59.13 lm	20.99 qr	11.21 def	4.06 efghi	1.40 bc	0.72 ghi	183.98 pqr	303.13 mn	487.11 pq
29	NS-L-136	60.51 t	20.55 lmn	11.42 ghij	4.20 jklmnop	1.24 a	0.14 a	165.53 jklm	297.44 klmn	462.97 ijkl
30	NS-L-137	57.95 f	21.59 u	12.20 rs	4.41 rs	1.42 bcde	0.61 fgh	180.16 opq	303.40 mn	483.56 nopq
31	NS-L-138	58.88 jk	21.30 qr	11.55 ijkl	4.30 opqr	1.56 defghijk	0.40 abcdef	172.68 mno	276.13 cdefg	448.81 defghij
32	NS-L-251	54.73 d	20.59 mno	11.84 mno	4.15 hijklmn	1.56 defghijk	3.06 k	162.45 hijk	307.71 n	470.15 klmno
33	NS-L-210	61.24 vw	18.37 b	11.32 fgh	4.19 ijklmno	1.63 hijk	0.73 ghi	153.63 def	323.33 o	476.96 lmnopq
34	NS-L-^{44}	60.^{83} tu	20.^{04} i	11.90 nop	3.62 a	1.50 bcdefghi	0.21 ab	146.22 bc	277.82 cdefgh	424.04 bc

35	NS-L-45	61.23 ^{vw}	19.78 ^{gh}	11.59 ^{jkl}	4.01 ^{cdef}	1.38 ^{b}	0.25 ^{abcd}	142.35 ^{ab}	272.57 ^{cdef}	414.92 ^{b}
36	NS-L-46	59.96 ^{r}	20.89 ^{pq}	11.44 ^{ghij}	4.04 ^{defgh}	1.45 ^{bcdefg}	0.29 ^{abcde}	157.27 ^{efghi}	324.23 ^{o}	481.50 ^{nopq}
37	NS-L-47	60.86 ^{u}	19.89 ^{hi}	11.80 ^{mn}	3.93 ^{bcd}	1.48 ^{bcdefgh}	0.25 ^{abcd}	178.81 ^{nop}	291.34 ^{ijkl}	470.16 ^{klmno}
38	Jelena	59.85 ^{qr}	20.25 ^{j}	11.80 ^{mn}	4.12 ^{fghijklm}	1.53 ^{bcdefghijk}	0.41 ^{abcdef}	161.05 ^{fghijk}	300.88 ^{lmn}	461.93 ^{hijkl}
39	Forward	60.20 ^{s}	18.37 ^{b}	11.57 ^{ijkl}	4.13 ^{fghijklmn}	1.42 ^{bcde}	0.95 ^{jk}	158.63 ^{efghij}	318.59 ^{o}	477.22 ^{lmnopq}
40	Maidan	60.79 ^{tu}	19.12 ^{cd}	11.32 ^{fgh}	4.11 ^{fghijkl}	1.54 ^{cdefghijk}	0.86 ^{hij}	155.48 ^{efgh}	287.81 ^{hijk}	443.29 ^{defg}
	Mean	58.64	20.12	11.48	4.15	1.55	1.22	166.41	291.60	458.01
	Maximum	64.58	22.88	13.60	4.91	2.09	17.41	233.88	369.42	572.28
	Minimum	33.53	15.75	9.31	3.34	0.61	0.05	121.83	204.87	341.21

Different letters indicate that means are significantly different from each other ($p < 0.05$). Fatty acids order is from highest to lowest content. They are reported as percent of total fatty acids in the sample. C18:1 oleic acid; C18:2 linoleic acid; C18:3 linolenic acid; C16:0 palmitic acid; C18:0 stearic acid; C22:1 – erucic acid; α-T alpha-tocopherol; γ-T gamma-tocopherol; Total-T total tocopherols

Supplementary Table 2. Elements of matrix W_1 and vector B_1 (presented in the bias row)

Variable	1	2	3	4	5	6	7	8	9	10
YEAR(2015)	-0.066	-0.225	0.482	0.925	0.679	-0.670	-0.067	0.145	2.104	-1.243
YEAR(2016)	0.894	0.099	0.319	-0.385	-1.113	-0.131	-1.553	-1.095	-1.795	0.899
YEAR(2017)	0.030	0.483	-0.041	1.005	-0.192	-2.079	0.697	-0.789	0.945	-0.501
YEAR(2018)	-1.421	0.488	-0.730	-1.455	0.726	3.150	1.607	1.534	0.075	0.636
GENOTYPE(1)	-0.724	0.047	0.233	-0.811	0.522	1.133	0.724	0.725	0.002	-0.461
GENOTYPE(10)	0.239	-0.240	-0.275	1.024	0.511	-0.065	-0.488	0.016	0.623	0.414
GENOTYPE(11)	-0.461	0.496	0.103	0.517	0.385	0.499	0.345	0.073	0.358	0.070
GENOTYPE(12)	0.345	-0.333	0.875	0.471	0.948	-0.562	0.212	-0.570	0.458	0.468
GENOTYPE(13)	1.435	-0.853	-0.223	0.930	0.129	-0.004	-0.308	0.256	0.160	-0.070
GENOTYPE(14)	-0.409	0.483	0.842	-0.473	-1.274	0.640	-0.001	-1.599	0.266	-0.672
GENOTYPE(15)	0.256	1.082	0.279	0.482	-0.391	0.188	-0.939	0.354	0.886	-0.170
GENOTYPE(16)	-0.083	1.484	-0.198	0.494	0.398	-0.294	-0.541	0.555	-0.411	1.231
GENOTYPE(17)	0.250	0.038	0.041	1.112	-0.640	0.870	-0.350	0.367	0.700	0.028
GENOTYPE(18)	0.336	-0.635	0.048	-0.705	0.522	0.135	0.904	-1.058	0.289	0.877
GENOTYPE(19)	-0.649	0.323	0.425	-0.145	-0.217	0.648	0.090	0.477	0.233	-0.610
GENOTYPE(2)	-0.365	0.095	-0.855	-0.074	1.670	0.147	-1.222	1.489	0.326	0.492
GENOTYPE(20)	-0.399	0.419	0.072	-0.383	-0.845	-0.043	1.789	0.688	0.693	-0.340
GENOTYPE(21)	0.307	-0.839	0.872	-0.826	-0.039	-0.164	-0.122	-0.745	-1.844	-1.157
GENOTYPE(22)	0.576	-0.782	-0.801	-0.895	1.669	-0.222	-0.878	0.032	0.992	1.385
GENOTYPE(23)	-0.192	0.085	0.079	0.518	0.599	0.078	-1.295	-0.320	0.811	2.115
GENOTYPE(24)	0.125	-1.528	-0.507	-0.692	1.049	-0.135	-1.222	-0.874	-3.601	1.374
GENOTYPE(25)	0.585	0.195	1.033	-1.409	-1.308	0.481	0.974	-0.171	-0.550	-1.078
GENOTYPE(26)	1.007	-0.815	-2.860	2.002	-0.790	0.585	1.634	1.271	1.672	-1.140
GENOTYPE(27)	0.044	0.102	0.886	0.107	-0.743	0.265	-0.123	-0.912	-1.200	-0.894
GENOTYPE(28)	0.772	-0.023	-0.076	-0.401	-0.281	-0.021	-1.140	-1.107	0.503	0.308
GENOTYPE(29)	0.262	-0.692	-0.337	-1.051	0.317	-0.739	-0.794	0.502	-0.953	0.693
GENOTYPE(3)	-1.855	1.539	-0.570	-0.693	-0.691	0.124	1.052	0.555	0.218	-0.108
GENOTYPE(30)	-0.101	0.010	0.501	-1.383	-0.923	1.848	-1.535	0.098	-1.970	0.276
GENOTYPE(31)	-0.034	0.803	-0.335	-0.029	1.285	0.576	-1.710	1.242	-0.165	0.359
GENOTYPE(32)	-0.680	1.148	0.362	1.936	-2.425	-1.712	-0.283	0.326	-0.797	-1.971
GENOTYPE(33)	-1.094	0.192	-0.009	0.592	-0.270	0.114	1.434	0.561	1.205	0.155
GENOTYPE(34)	1.895	-0.873	0.792	0.346	1.309	-0.028	-0.858	0.161	0.501	-0.370
GENOTYPE(35)	1.082	-0.507	0.458	0.027	-0.027	-1.182	-0.651	-0.216	0.139	-0.003
GENOTYPE(36)	-0.314	-0.174	0.277	0.536	-0.526	-1.110	-0.624	0.155	0.579	0.059
GENOTYPE(37)	0.240	-0.685	-0.051	-0.908	0.167	-0.816	1.593	-0.503	0.687	-0.344
GENOTYPE(38)	0.043	-0.506	0.411	-0.522	-0.518	0.003	0.140	-0.621	-1.402	-0.320
GENOTYPE(39)	-0.080	-0.622	-0.394	-0.266	-0.431	-1.680	1.870	-0.064	0.937	0.331
GENOTYPE(4)	-0.865	-0.611	-0.497	-0.331	1.192	-0.162	-0.392	0.918	-0.400	-0.306
GENOTYPE(40)	0.121	-0.325	-0.090	0.464	0.065	-0.246	0.495	-0.187	0.378	0.234
GENOTYPE(5)	-1.535	1.309	-0.357	-1.072	-0.838	0.545	1.120	0.354	-0.110	-0.414

GENOTYPE(6)	0.465	1.625	-1.021	1.302	0.039	0.094	-0.659	-1.377	0.122	-0.700
GENOTYPE(7)	0.318	-0.157	-0.990	0.427	0.052	-0.120	0.633	0.301	0.851	-0.330
GENOTYPE(8)	-1.441	0.606	0.308	-0.763	0.071	-0.293	1.943	-0.265	-0.477	0.423
GENOTYPE(9)	0.081	0.018	1.619	0.728	0.532	0.876	-0.133	-0.957	0.539	-0.151
Bias	-0.563	0.897	0.037	0.113	0.104	0.352	0.733	-0.185	1.333	-0.233

Supplementary Table 3. Elements of matrix W_2 and vector B_2 (presented in the bias column)

	1	2	3	4	5	6	7	8	9	10	Bias
C16:0	-0.822	0.529	-0.683	-0.371	-0.049	0.157	-0.296	-0.149	-0.123	0.058	-0.386
C18:0	-0.509	-0.168	-0.079	0.591	0.426	0.633	0.318	-0.063	-0.114	-0.301	0.236
C18:1	-0.215	0.971	1.159	-0.496	0.734	-1.116	0.294	-0.121	0.044	0.809	0.815
C18:2	-0.485	0.422	0.035	-1.107	0.200	-0.188	-1.294	-0.115	0.090	-0.769	0.378
C18:3	-0.311	-1.190	0.309	-0.595	-0.399	0.409	-0.497	0.472	0.407	-0.673	0.554
C22:1	1.425	0.158	-1.569	1.498	-0.743	1.803	0.918	-0.097	-0.324	-0.629	-3.300
α-tocopherol	-1.067	-0.364	-1.391	-0.639	0.081	0.681	-1.541	-1.827	1.145	-1.189	-0.713
γ-tocopherol	-1.494	-2.191	0.364	-0.266	-0.967	-0.028	-1.102	0.063	1.461	0.104	0.732
Total tocopherols	-1.459	-1.647	-0.326	-0.450	-0.639	0.242	-1.383	-0.688	1.471	-0.388	0.057

C16:0 palmitic acid; C18:0 stearic acid; C18:1 oleic acid; C18:2 linoleic acid; C18:3 linolenic acid; C22:1 – erucic acid; α-tocopherol alpha-tocopherol; γ-tocopherol gamma-tocopherol