Scientific Papers. Series B, Horticulture. Vol. LXVIII, No. 1, 2024
Print ISSN 2285-5653, CD-ROM ISSN 2285-5661, Online ISSN 2286-1580, ISSN-L 2285-5653

RELATIONSHIPS BETWEEN RGB COLOR CHARACTERISTICS AND TOMATO FRUIT QUALITY

Mima TODOROVA[1], Stefka ATANASSOVA[2], Dimitar YORGOV[2], Petya VELEVA[2], Dimitar ZHELYAZKOV[3], Antonya PETROVA[2], Velika BOTEVA[4]

[1]Trakia University, Faculty of Agriculture, Department of Plant Breeding, 6015, Stara Zagora, Bulgaria
[2]Trakia University, Faculty of Agriculture, Department of Agricultural Engineering, 6015, Stara Zagora, Bulgaria
[3]Trakia University, Faculty of Agriculture, student, 6015, Stara Zagora, Bulgaria
[4]Maritsa Vegetable Crops Research Institute, 32 Brezovsko shosse St, Plovdiv, Bulgaria

Corresponding author email: mima.todorova@trakia-uni.bg

Abstract

The present study aimed to establish the relationship between the characteristics of RGB images of tomato fruits of three cultivars and important indicators of the quality of these fruits. In the study, 30 fruits from three cultivars at different ripening stages were used. The digital images of each tomato from four sides were obtained using a document camera. The same fruits were analyzed for dry matter, total dyes, vitamin C, titr. organic acids, beta carotene and lycopene content. The data of each image's G, R, and B channels were extracted and averaged. The average data of grayscale were taken as well. Based on the color characteristics and the chemical analyses, a procedure of descriptive statistics, correlation, and regression analysis was performed. The image procedure of all obtained data was performed using Jupyter notebook. The highest correlation coefficient (-0.860) was obtained between the Gray channel and total dye content, followed by the strong correlation between the Green channel and lycopene content (-0.820). And the models with the highest predictive ability were the models for total dye by Gray channel values (R^2 = 0.815) and lycopene by Green channel (R^2 = 0.761).

Key words: *tomato, images, color, Jupyter notebook, regression.*

INTRODUCTION

The tomato (*Solanum lycopersicum*) one of the economically important agricultural crops that are eaten fresh or processed. Colour or pigment changes during tomato ripening are characterized by loss of chlorophyll and rapid accumulation of carotenoids, especially lycopene. Colour is an important quality indicator of fruits and vegetables. Given the consumption of fresh fruit and the short shelf-life, there is a growing need for new, alternative technologies to assess the quality of such fruit and the possibility of using it in processing for fruit juices, jams, etc. Alternative technologies with the potential to sort fruit by appearance, texture, taste, nutritional value would provide higher quality fruit, increase consumer confidence and satisfaction, and increase the competitiveness and profitability of the agricultural and fruit industry (ElMasry et al., 2006., Hasanzadeh et al., 2022). In recent years, successful studies have been conducted to determine quality parameters of tomatoes at different maturity levels using portable spectrometers in the visible and near infrared range (Huang et al., 2018; Lu et al., 2016; Skolik et al., 2019).

The aim of the present study was to establish the relationship between the characteristics of RGB images of tomato fruits of three cultivars and important indicators of the quality of these fruits with the possibility to develop regression models for prediction.

MATERIALS AND METHODS

Tomato samples

In the paper is presented analysis of tomato fruits from three famous cultivars: Manusa, Mirsini and Red Bounty, grown in the greenhouse production base in Plovdiv,

Bulgaria. From each variety, 10 fruits of different maturity levels were taken. A document-camera is used for image data acquisition. A database of 120 digital images of the tomato fruits (40 from each variety) is formed. The digital images are acquired in the main color space RGB.
The dry matter, %, titratable organic acids, % calculated as citric acid after titration with 0.1 N NaOH, total pigments, mg %, vitamin C mg % by Tilman's reaction with 2,6-ichlorophenolindophenol, total dyes, mg %, lycopene, mg % (Manuelyan, 1991), and beta carotene, mg % content were measured on each fruit.

Color components of tomato friuts images

The software Jupyter notebook was used to process the images. Jupyter Notebook - browser based front-end originally developed for the programming languages Julia, Python. (Palkovits, 2020) Each image was cropped and the average values of R channel, G channel, B channel are taken through the corresponding code:

```
avgR = np.mean(mirsini1A_crop[:,:,0])
avgG = np.mean(mirsini1A_crop[:,:,1])
avgB = np.mean(mirsini1A_crop[:,:,2])

r,g,b = mirsini1A_crop [:,:,0], mirsini1A_crop[:,:,1], mirsini1A_crop[:,:,2]
```

Each cropped image was represented in grayscale format by multiplying the values of the three channels by the corresponding coefficients via code:

```
mirsini1A_crop_gray = 0.299 * r + 0.587 * g + 0.144 * b
mirsini1A_crop_gray.mean()
```

Tomato fruits images from Mirsini cultivar (a), Manusa cultivar (b) and Red Bounty cultivar (c) are presented in Figure 1 after image processing.

Statistical analysis

Correlation analysis was used to examine the strength of the relationships between the examined chemical components of tomato fruits from different varieties (Manusa, Mirsini, Red Bounty) and the color components from their images.

Regression analysis was applied to calculate the predictive models defining relations between color components and chemical traits of tomato fruits. Two regression models were compared - Linear (L), (1) and Quadratic (Q), (2), expressed with the equations:

$$Y = b_0 x + b_1 \quad (1)$$
$$Y = b_0 + b_1 x + b_2 x^2 \quad (2)$$

where: Y are the observed parameters (Dry matter, Vitamin C, Titr. organic acids, Total dyes, Lycopene, and Beta carotene); x is the fixed factor (color components extracted from the tomato images); b_0; b_1; b_2 are the model coefficients. The models' estimation was done by comparing the parameters Coefficient of determination (R^2), Standard error of the estimation (SEE), and Standard error of predicted values (SEPv). The analysis and processing of the experimental data was done with the IBM® SPSS® Statistics 26.0 software at the significance level α=0.05

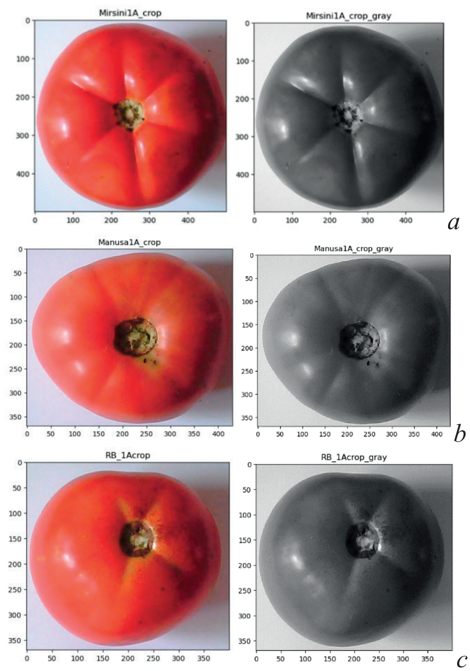

Figure 1. Tomato fruits from Mirsini var (a), Manusa var. (b) and Red Bounty var. (c)

RESULTS AND DISCUSSIONS

Correlation analysis

Descriptive statistics of average values of R, G, B, Gray channels and the studied chemical parameters of all tomato fruits are presented in Table 1.

Table 1. Descriptive statistics of color characteristics and the studied chemical parameters in all tomato fruits

	min.	max.	mean	SD
R mean	109.70	226.05	183.67	13.96
G mean	83.36	145.38	107.52	14.44
B mean	83.61	197.51	100.97	14.99
Gray mean	113.82	172.16	132.93	10.39
Dry matter, %	3.50	4.70	3.93	0.28
Vitamin C, mg %	13.10	31.00	20.38	4.01
Titr. organic acids, %	0.23	0.45	0.32	0.05
Total dyes, mg %	1.25	4.43	2.58	0.89
Lycopene, mg %	0.36	3.90	2.09	0.96
Beta carotene, mg %	0.00	2.43	0.31	0.42

Table 2 presents the results of the bivariate correlation analysis regarding the strength of the relationship between color components extracted from the tomato images and the investigated chemical components.

As can be seen from the table, in all three tomato cultivars studied, the correlation between the G (green) and Grey channel of the images with the chemical components studied was the strongest.

The highest correlation coefficient (-0.860) was obtained between the Gray channel and total dye, mg % content for the Manusa cultivar, followed by the strong correlation between the G channel and lycopene content, mg % in the Mirsini cultivar (-0.820) and between the Gray colour channel and lycopene content in the Manusa cultivar (-0.798).

Table 2. Bivariate correlation between colour characteristics and examined components of tomato fruits

	Dry matter (%)	Vit C (mg %)	Titr. organic acids (%)	Total dyes (mg %)	Lycopene (mg %)	Beta carotene (mg %)
			Manusa			
R	0,435	-0,008	-0,216	-0,599**	-0,634**	0,500*
G	0,096	-0,236	-0,176	**-0,774 ** **	**-0,732 ** **	0,179
B	0,309	-0,070	-0,302	-0,751**	-0,668**	0,318
Gray	0,210	-0,127	-0,282	**-0,860** **	**-0,798** **	0,314
			Mirsini			
R	-0,254	-0,305	-0,125	0,109	0,234	-0,365
G	-0,031	0,450*	0,398	**-0,774** **	**-0,820** **	0,664**
B	0,013	-0,015	0,323	-0,421	-0,435	0,270
Gray	-0,073	0,345	0,324	-0,645**	-0,753**	0,605**
			Red Bounty			
R	0,050	-0,503*	-0,017	0,218	0,246	-0,361
G	0,517*	0,684**	-0,456*	**-0,692** **	**-0,691** **	0,450*
B	0,485*	0,402	-0,496*	-0,397	-0,394	0,237
Gray	0,542*	0,532*	-0,476*	-0,620**	-0,611**	0,338

*Correlation is significant at the 0.05 level
**Correlation is significant at the 0.01 level

In the our study, a moderate correlation was also found between the G channel and total dye content for the three cultivars (-0.774; -0.774 and -0.692), and between the G channel values and lycopene content in cultivar Red Bounty (-0.691). Ropelewska et al. (2022) reported similar results, with a much higher degree of correlation between the green channel and lycopene content. The authors indicated a strong negative correlation, with values of r = -0.99. The authors also constructed calibration models with a high degree of accuracy for the determination of lycopene content based on the colour characteristics obtained from the G channel. What is interesting in their study is that they used tomato varieties of different colour and shape (yellow, orange, pink, red, including cherry tomatoes). These results are a prerequisite for creating predictive regression models for the investigated chemical components of the three tomato cultivars based on the green and gray components of their colour images.

Regression analysis

Linear and nonlinear regression equations were constructed based on data from laboratory analyses of the chemical components examined in the three tomato cultivars and the green and gray components of their colour images.

Dry matter (%), Vitamin C (mg %) and Titr. organic acids (%)

Table 3 presents the regression models estimated to predict the dry matter content of the studied tomato cultivars. For the cultivars Manusa and Mirsini, both linear and non-linear models were statistically insignificant and not suitable for predicting dry matter content. Only the models developed for the cultivar Red Bounty were statistically significant, with the highest coefficient of determination (R^2 = 0.323) being the quadratic model constructed for the green colour component. The coefficient of determination (R^2) indicated that 32.3% of the variation in dry matter in the Red Bounty cultivar was due to variation in the G values. Furthermore, SEE = 0.152 and SEP = 0.040 of this model are the lowest, which is a confirmation that it could be be used to predict the dry matter content in Red Bounty cultivar.

Table 3. Models summary and parameters estimation showing the relations between Dry matter (%) and Green and Gray channel values for the investigated tomato fruits

Parameter estimates		Equations	R^2	SEE	SEPV	Sig.
		Manusa				
Green	L	$DM = 0{,}001\,x + 3{,}688$	0,009	0,188	0,057	0,686
	Q	$DM = 1{,}803 + 0{,}038x - 0{,}0001x^2$	0,036	0,173	0,050	0,732
Mean gray	L	$DM = 0{,}004\,x + 3{,}311$	0,044	0,186	0,066	0,374
	Q	$DM = -5{,}447 + 0{,}139x - 0{,}0005x^2$	0,122	0,175	0,062	0,331
		Mirsini				
Green	L	$DM = -0{,}0007\,x + 4{,}204$	0,001	0,333	0,103	0,897
	Q	$DM = 1{,}626 + 0{,}048x - 0{,}0002x^2$	0,021	0,336	0,110	0,836
Mean gray	L	$DM = -0{,}003\,x + 4{,}474$	0,005	0,333	0,101	0,760
	Q	$DM = -8{,}950 + 0{,}205x - 0{,}0008x^2$	0,058	0,301	0,098	0,603
		Red Bounty				
Green	L	$DM = 0{,}007\,x + 3{,}067$	0,268	0,162	0,049	0,019
	Q	$DM = 0{,}237 + 0{,}058x - 0{,}0002x^2$	**0,323**	**0,152**	**0,040**	**0,036**
Mean gray	L	$DM = 0{,}012\,x + 2{,}275$	0,293	0,159	0,048	0,014
	Q	$DM = -4{,}174 + 0{,}108x - 0{,}0003x^2$	0,313	0,160	0,042	0,041

* Level of significance $p < 0.05$; R^2 - Coefficient of determination; SEE - Standard Error of the Estimate; SEPV - Standard Error of Predicted Value

Table 4 presents the regression models estimated to predict the vitamin C, mg % content of the studied tomato varieties. For the Manusa the regression models were statistically insignificant and not suitable for predicting vitamin C content. For the Mirsini only the models based on the G values were statistically significant. In these models, the quadratic and linear models have very similar coefficient of determination values (0.204 and 0.202), i.e. about 20% of the variation in vitamin C content in Mirsini var. is due to variation in the green colour channel. However, the quadratic model has lower values of both the SEE = 3.768 and the SEP = 1.385, making it more suitable than the linear model for predicting the amount of vitamin C in Mirsini var. For the cultivar Red Bounty, all fitted models were statistically significant. Again, the quadratic model based on the G colour component had the highest coefficient of determination (R^2 = 0.481), i.e. 48.1% of the variation in vitamin C in Red Bounty could be explained by variation in G values. Also, SEE = 1.287 and SEP = = 0.402 of this model are the lowest, i.e. it can be used to predict the vitamin C content in Red Bounty.

Table 4. Models summary and parameters estimation showing the relations between Vitamin C (mg %) and Green or Gray channel values for the investigated tomato varieties

Parameter estimates		Equations	R^2	SEE	SEPv	Sig.
		Manusa				
Green	L	$Vit\ C = 0{,}077\ x + 28{,}108$	0,056	4,396	1,344	0,317
	Q	$Vit\ C = 38{,}616 - 0{,}282\ x + 0{,}0008\ x^2$	0,057	4,376	1,365	0,607
Mean gray	L	$Vit\ C = -0{,}056\ x + 27{,}352$	0,016	4,487	1,365	0,594
	Q	$Vit\ C = -132{,}679 + 2{,}415\ x - 0{,}009\ x^2$	0,061	4,542	1,322	0,584
		Mirsini				
Green	L	$Vit\ C = 0{,}155\ x + 5{,}199$	0,202	4,573	1,406	0,047
	Q	$Vit\ C = 15{,}793 - 0{,}045\ x - 0{,}0009x^2$	**0,204**	**3,768**	**1,385**	0,044
Mean gray	L	$Vit\ C = 0{,}191\ x - 3{,}390$	0,119	4,806	1,459	0,136
	Q	$Vit\ C = 87{,}360 - 1{,}215x - 0{,}005x^2$	0,129	4,274	1,472	0,309
		Red Bounty				
Green	L	$Vit\ C = 0{,}091\ x + 9{,}829$	0,468	1,342	0,407	0,001
	Q	$Vit\ C = -3{,}35 + 0{,}326x - 0{,}001x^2$	**0,481**	**1,287**	**0,402**	0,004
Mean gray	L	$Vit\ C = 0{,}113\ x + 4{,}753$	0,283	1,558	0, 474	0,016
	Q	$Vit\ C = 87{,}987 - 1{,}125x - 0{,}005x^2$	0,318	1,543	0,479	0,039

* Level of significance $p < 0.05$; R^2 - Coefficient of determination; SEE - Standard Error of the Estimate; SEPV - Standard Error of Predicted Value

Similar to the models developed to predict dry matter content, only the models developed to predict titratable organic acid content of Red Bounty were statistically significant. The model with the highest coefficient of determination (R^2 = 0.405) was the quadratic model based on Green colour channel. The coefficient of determination indicates that 40.5% of the variation in titratable organic acid content in the cultivar Red Bounty depends on variation in the green colour component. Both the SEE (0.019) and SEP (0.003) of this model are very low, i.e., the model is suitable for predicting the content of titratable organic acids in Red Bounty cultivar.

Lycopene (mg %), Total dyes (mg %), and Beta carotene (mg %)

Figure 2 visualizes the estimated regression models for predicting Lycopene content, mg% in the studied tomato cultivars. Similar to the models for predicting Total dyes content, the estimated models for all three cultivars were statistically significant, with higher coefficients of determination and lower errors obtained for all quadratic models compared to the linear models. The quadratic models for tomato Manusa var. developed with the Gray channel data were characterized by a high coefficient of determination (0.761), i.e. 76.1% of the variation in Lycopene content in tomatoes of this cultivar could be explained by variation in the Gray values. The errors of the model created with the Gray component data (SEE = 0.300; SEPv = 0.087) were lower than those of the model estimated with the G channel.

For the other two cultivars Mirsini and Red Bounty, the quadratic models estimated on the basis of the Green colour component were more suitable for predicting the Lycopene content. Their coefficients of determination (0.675 and 0.485) are higher than the coefficients of the models for predicting Lycopene content obtained with the gray component. The errors of the prediction models for lycopene content in Mirsini tomatoes were SEE = 0.561 and SEPv = 0.173, and in Red Bounty tomatoes SEE = 0.580 and SEPv = 0.173, being lower in comparison with the errors of the quadratic models obtained with the gray colour component data. There are various successful studies related to the color used to measure lycopene (Villaseñor-Aguilar et al., 2021), in this respect the results obtained in our study were expected.

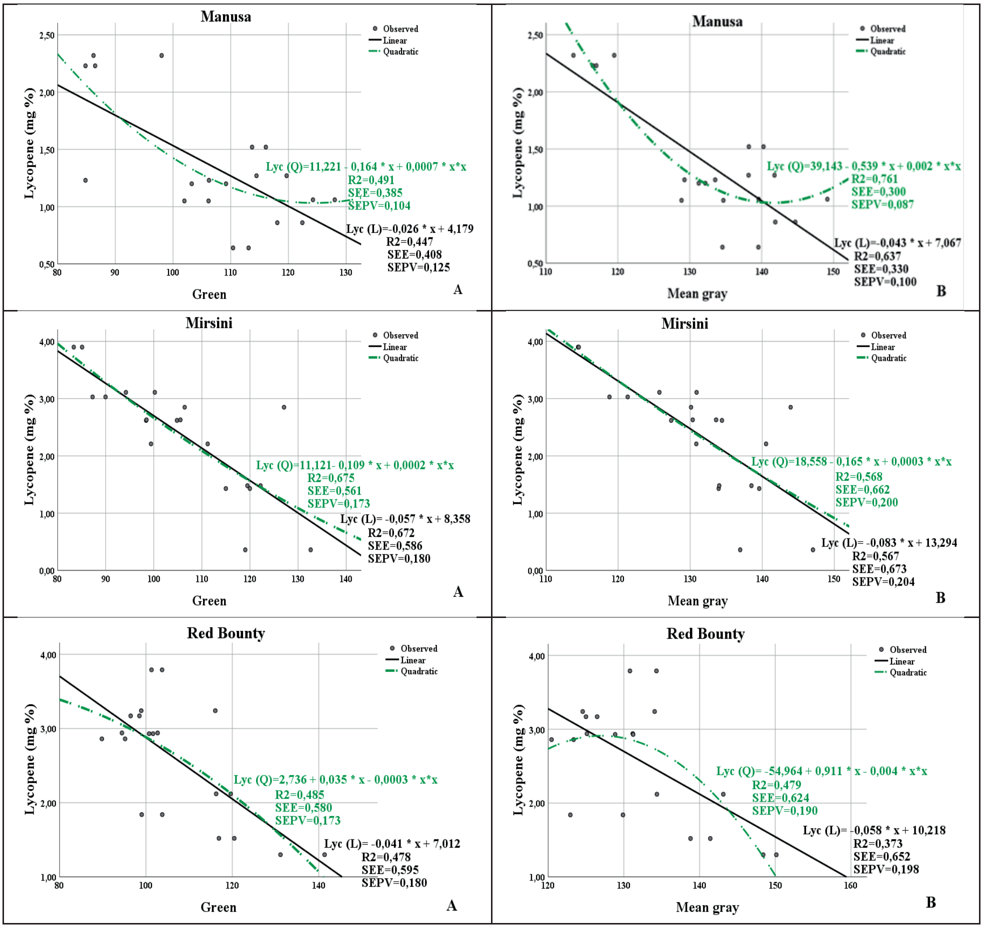

Figure 2. Estimation of the regression models' curves showing the relation between the content of Lycopene (mg %) and Green or Gray colour components for the investigated tomato varieties

Figure 3 visualizes the estimated regression models to predict the total dyes, mg% content in the studied tomato fruits. The resulting regression models for all three cultivars are statistically significant, but the quadratic equations have higher coefficients of determination and lower errors compared to the linear models. The quadratic regression models for tomato cultivar Manusa, developed with the data on the Gray component, had a very high coefficient of determination (0.815), i.e. 81.5% of the variation in total dyes, mg % content in this cultivar could be explained by variation in the Gray colour component. Accordingly, the errors of the model generated with the gray component data (SEE = 0.225; SEPv = 0.043) were lower than those of the model estimated with the green colour component.

For the Mirsini cultivar, the quadratic models estimated with the G colour values have higher values of both the coefficient of determination (0.611) and the corresponding errors (SEE = 0.524; SEPv = 0.157) when compared to the quadratic models obtained with the gray component data. The predictive models obtained for the total dyes, mg% content of Red Bounty had the same coefficient of determination values (0.488), but the quadratic model obtained with the green colour component data had lower errors (SEE = 0.547; SEPv = 0.170) compared to that obtained with the gray colour component, i.e. it was more suitable for predicting courts.

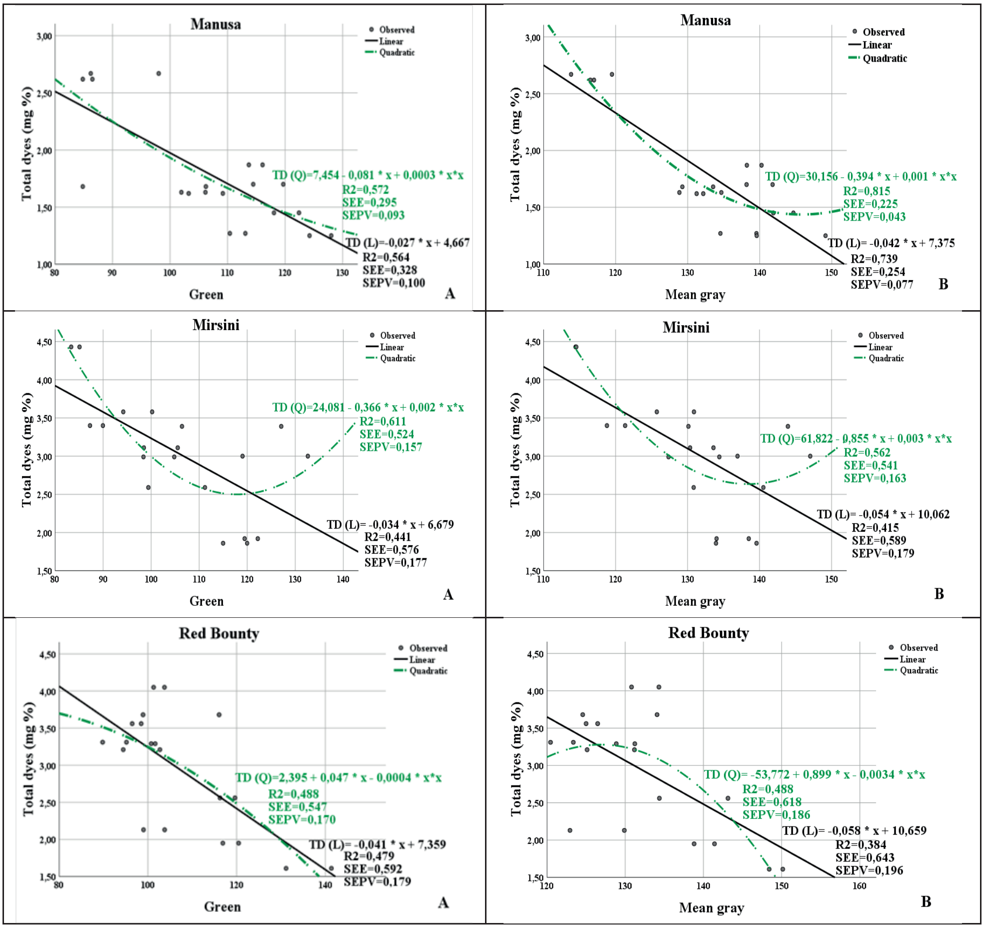

Figure 3. Estimation of the regression models' curves showing the relation between the content of Total dyes (mg %) and Green or Mean gray color components for the investigated tomato varieties

According to the Beta carotene, mg % prediction the results of the regression analysis show that the estimated models for the Manusa and Red Bounty cultivars were statistically insignificant.

All the models developed for the cultivar Mirsini were statistically significant, but as with the other chemical components studied, the one with the highest coefficient of determination (R^2 = 0.617) was the quadratic model constructed from the green colour component data.

The errors of this model are the lowest (SEE = 0.040 and SEPV = 0.012, respectively), i.e. the model is suitable for predicting the Beta carotene, mg % content in the Mirsini cultivar.

CONCLUSIONS

An experiment was conducted with tomato fruits of three different cultivars (Manusa, Mirsini and Red Bounty) to evaluate the relationship between their major chemical elements and colour components obtained from their images.

Using bivariate correlation analysis, it was found that the strongest correlation was between the Green and Gray colour components of the images with the chemical elements studied.

Regression models were created to predict the investigated chemical elements for the three tomato cultivars from the Green and Gray colour component data of their images.

It was found that the models with the highest predictive ability were the models for total dye, mg % and lycopene, mg % contents for all three tomato varieties.

This study complements previously known methods for non-destructive evaluation of major components of tomato fruit composition and highlights the potential of regression models to be integrated into automated tomato fruit harvesting, sorting, processing systems.

ACKNOWLEDGEMENTS

This work was supported by the Bulgarian Ministry of Education and Science under the National Research Programme "Smart crop production" approved by Decision of the Ministry Council № 866/26.11.2020 г.

REFERENCES

ElMasry, G., Wang,N., ElSayed, A., Ngadi, M. (2007). Hyperspectral imaging for nondestructive determination of some quality attributes for strawberry. *Journal of Food Engineering* 81, 98–107

Hasanzadeh, B., Abbaspour-Gilandeh, Y., Soltani-Nazarloo, A., Hernández-Hernández, M., Gallardo-Bernal, I., Hernández-Hernández, J.L. (2022) Non-Destructive Detection of Fruit Quality Parameters Using Hyperspectral Imaging, Multiple Regression Analysis and Artificial Intelligence. *Horticulturae* 8,598. https://doi.org/10.3390/horticulturae8070598

Huang, Y., Lu, R., Chen, K. (2018). Prediction of firmness parameters of tomatoes by portable visible and near-infrared spectroscopy. *Journal of Food Engineering* 222, 185-198.

Lu, H., Wang, F., Liu, X., Wu Y. 2016. Rapid Assessment of Tomato Ripeness Using Visible/Near-Infrared Spectroscopy and Machine Vision. *Food Analytical Methods, 10 (6)*

Manuelyan, H. (1991). Express methods for assessing the carotenoid composition of tomato fruits", in Genetic Improvement of Tomato, G. Kallo, Ed. Spring –Velag, 1991, pp. 193-195

Ropelewska, E.,Szwejda-Grzybowska, J. (2022) Relationship of Textures from Tomato Fruit Images Acquired Using a Digital Camera and Lycopene Content Determined by High-Performance Liquid Chromatography. *Agriculture* 12, 1495.

Skolik, P., Morais, C., Martin, F., & McAinsh M. (2019). Determination of developmental and ripening stages of whole tomato fruit using portable infrared spectroscopy and Chemometrics. *BMC Plant Biology* 19, 236

Villaseñor-Aguilar, M.-J., Padilla-Medina, J.-A., Botello-Álvarez, J.-E, Bravo-Sánchez, M.-G., Prado-Olivares, J., Espinosa-Calderon, A., Barranco-Gutiérrez, A.-I. (2021) Current Status of Optical Systems for Measuring Lycopene Content in Fruits: *Review. Appl. Sci*. 2021, 11, 9332. https://doi.org/10.3390/app11199332