Research

Yield prediction of experimental plots based on the harvest of specific fruit clusters for selection of fresh market tomato hybrids

Paulo TP Silva [1][iD]; **Gustavo E Oliveira** [1][iD]; **Paulo R Peloia** [1][iD]; **Regis C Carvalho** [2][iD]; **Flávia Maria A Gonçalves**[2][iD]

[1]Syngenta Proteção de Cultivos, Holambra-SP, Brasil; paulo.silva-2@syngenta.com; gustavo.oliveira@syngenta.com; paulo.peloia@syngenta.com; [2]Universidade Federal de Lavras, Departamento de Biologia (UFLA), Lavras-MG, Brasil; regisccarvalho@hotmail.com; avelar@ufla.br

ABSTRACT

The objective of this study was to propose models based on the harvest of specific fruit clusters to estimate the plot yield of trials containing fresh market tomato hybrids. Three experiments were conducted at the Experimental Station of Syngenta Crop Protection in the municipality of Holambra, state of São Paulo, Brazil, in 2016 and 2017. The experimental design was randomized complete blocks with 12 genotypes (experiments 1 and 3) and 13 genotypes (experiment 2), with 4 replicates and 10 plants per plot. Multiple linear regression models were fitted (stepwise method) with experiment 1 data (cross-validation), the best models were selected (higher adjusted R^2) and then tested with experiments 2 and 3 data by mean of absolute percentage error (MAPE), for the traits mean weight of marketable fruit per plant A (MWA), AA (MWAA), AAA (MWAAA), AA and AAA (MW23A), and the ratio between the market weight and the number of marketable fruits (MFW). The models with four fruit clusters (2, 3, 6 and 9) showed the best balance between prediction capacity and the number of fruit clusters to harvest. The traits MWAA, MW23A and MFW generated reliable predictions, with MAPE approximately 5%. The multiple linear regression can be used to estimate the plot yield what, in the last instance, contributes to the reduction of the costs to conduct fresh market tomato hybrids.

Keywords: *Solanum lycopersicum*, tomato breeding, linear regression.

RESUMO

Predição de produtividade de parcelas experimentais com base na avaliação de pencas específicas para seleção de híbridos de tomate de mesa

O objetivo deste estudo foi propor modelos baseados na avaliação de colheita de pencas de frutos específicas para estimar a produtividade de parcelas de experimentos com híbridos de tomate de mesa. Foram realizados três experimentos, em 2016 e 2017, na Estação Experimental da Syngenta Proteção de Cultivos, Holambra-SP. Os experimentos foram implantados no delineamento em blocos casualizados com 12 genótipos (experimentos 1 e 3) e 13 genótipos (experimento 2), com quatro repetições e parcela de dez plantas. O ajuste de modelos de regressão linear múltipla (método *stepwise*) foi realizado com base no experimento 1 (validação cruzada), no qual os melhores modelos (maior R^2 ajustado) foram selecionados e posteriormente testados com os dados dos experimentos 2 e 3, para cada uma das variáveis-resposta: peso comercial dos frutos tipo AAA (MWAAA), AA (MWAA), A (MWA), AA+AAA (MW23A) e peso médio de frutos comerciais (MFW), por meio do parâmetro erro absoluto percentual médio (MAPE). Os modelos com quatro pencas (2, 3, 6 e 9) apresentaram melhor balanço entre capacidade preditiva e número de pencas a serem colhidas. As variáveis MWAA, MW23A e MFW apresentaram estimativas confiáveis, uma vez que MAPE foi em torno de 5%. O modelo de regressão linear múltipla pode ser usado para estimativa de produtividade de parcelas o que, em última instância, contribui para a redução dos custos para condução de experimentos com híbridos de tomate de mesa.

Palavras-chave: *Solanum lycopersicum*, melhoramento do tomateiro, regressão linear.

The tomato is the most popular vegetable because of its ease of use and versatility. It is also an economically important crop due to the social value of tomato production, which directly and indirectly creates jobs (Souza *et al.*, 2010; Alvarenga, 2013; Stajcic´ *et al.*, 2015).

Approximately 30 national and international tomato seed companies conduct research and breeding programs in search of cultivars that are better adapted to local conditions and have higher productivity, resistance to diseases, and quality fruits with a good color, flavor, and long shelf life (Mohamed *et al.*, 2012; Treichel, 2016; Rothan *et al.*, 2019). Of the several tomato types in the market, Redondo or Salada Longa Vida are the most marketed (Treichel, 2016).

Tomato hybrid development experiments are expensive and time-consuming because they require about seven visits for fruit harvest and classification, during the harvest period that may last about 60 days. Therefore, developing a method that allows the tomato yield prediction without requiring this frequency of evaluations

will decrease travel and labor costs, and make the process more efficient. In tomato, there is a huge difference of earliness among the different genotypes, which is the reason why this study did not consider evaluating yield per week because the genotypes with shorter cycle would have more yield collected, it means, more harvested clusters than the ones with longer cycle. Therefore, this research hypothesizes that a model based on the harvest data of specific fruit clusters predicts the tomato plot yield without compromising the experiment's capability of differentiating genotypes.

The yield of an experimental plot is given by the sum of the yield of individual fruit clusters, usually 12. Therefore, due to this linear relationship between fruit cluster yield and plot yield, and the influence of multiple fruit clusters in the plot yield, the multiple linear regression model has been chosen to test the hypothesis proposed.

However, in contrast to the hypothesis proposed in the present study, most studies relate yield prediction not only with plant yield factors but also with environmental factors (Higashide. 2009; Qaddoum *et al.*, 2013; Hussain & Hatibaruah, 2015).

The yield of fresh tomatoes is split into different classes according to the quality. Therefore, the main challenge of this study is to select one group of fruit clusters, which will be the parameters of different models, each one to a class of fruits, that maximizes the overall predictive capability.

Thus, this study proposes multiple linear regression models based on harvest data of specific fruit clusters to predict the plot yield of experiments with fresh market tomato hybrids.

MATERIAL AND METHODS

Three experiments were conducted at the Experimental Station of Syngenta Crop Protection in the municipality of Holambra, state of São Paulo, Brazil (22°38'58"S; 47°05'34"W; 600 m altitude).

Experimental conduction

Experiment 1 was conducted between February 16th and August 17th, 2016; experiment 2, from January 23rd to June 27th, 2017, and experiment 3, from March 10th to August 21st, 2017. All under field conditions.

The seedlings were produced in greenhouses at the Holambra experimental station in 200-cell polystyrene trays. Coconut fiber (type 11, Amafibra) was used as substrate, with the following characteristics: pH 5.8, electrical conductivity = 1.1 dS/m, density = 89 kg/m³ and moisture retention capacity = 308 mL/L. The irrigation was performed by micro-sprinkler systems, and fertilization with 0.5 g/L calcium nitrate, 0.4 g/L potassium nitrate, 0.6 g/L monopotassium phosphate (MKP), and 0.2 g/L magnesium sulfate was applied three times per week. Preventive plant health management was performed with the application of fungicides, insecticides, and acaricides registered for tomatoes. Weed control between tomato beds was performed by mechanical weeding as needed.

Seedlings were transplanted into the experimental area, 30 days after sowing, when the seedlings presented three to four leaves.

One plant was planted per hole, and axillary buds were removed, leaving only two stems per plant. Plants were grown until the development of the 12th fruit cluster. Then, tip pruning was performed, which is a common practice in fresh market tomato production. Tip pruning consists of eliminating the terminal bud of the tomato stems to stop the vertical growth of the plant (Alvarenga, 2013).

Top-dressing fertilizations were performed twice a week, beginning 25 days after setting the experiment in the final location. A drip fertigation system was used. Fertilization was divided into three stages based on the crop developmental stage (Alvarenga, 2013). From weeks 3 to 5 after seedling transplantation (WAT), 1.67 kg of calcium nitrate, 1.25 kg of MKP, and 0.535 kg of potassium nitrate were applied. In the second stage from 6 to 11 WAT, 2.1 kg of calcium nitrate, 1.73 kg of MPK, and 0.945 kg of potassium nitrate were applied. In the third stage from 12 to 17 WAT, 1.1 kg of calcium nitrate and 1.74 kg of potassium nitrate

were applied.

Fruit clusters were numbered from 1 to 12 as follows: the first and second fruit clusters in the main stem were numbered 1 and 2, respectively, the first cluster of the secondary stem was numbered 3, and so forth, in a zigzagging pattern up to the 12th cluster.

The fruit clusters were labeled and harvested separately depending on the plant development and fruit ripening stage. The data were expressed as the mean value of the cluster per plot.

Evaluated traits

Tomato fruits were classified using the system typically adopted in the tomato production chain for group Salada Longa Vida (PBMH, 2003), which classifies fruits into three categories according to their transverse diameters: AAA (>85 mm), AA (65 to 85 mm) and A (<65 mm).

Fruit classification was performed using a wooden box with two sieves to separate the fruits. The following traits were evaluated during harvest: mean weight of marketable fruit A (MWA), mean weight of marketable fruit AA (MWAA), mean weight of marketable fruit AAA (MWAAA), mean weight of marketable fruit AA and AAA (MW23A), all in kg/plant; and, the ratio between the market weight and the number of marketable fruits (MFW), expressed in grams/fruit.

Experimental design

A randomized complete block design with four replicates in linear plots with 10 plants each was used. Spacing was 1.5 m between rows and 0.5 m between plants.

For experiment 1, 12 genotypes with indeterminate growth were evaluated, nine tested hybrids (214103, 214312, 214320, 413498, 413504, 413518, 413524, 413669, and 213071), and three commercial hybrids (Paronset, Compack, and Itaipava) from the salad group, which were used as controls.

In experiment 2, 13 different genotypes (114572, 114594, 150029, 214452, 414929, 411451, TL130060, 412029, TL145408, Comodo, Dylla, Paronset and Compack) were used. And 12 groups were used in experiment 3 (all genotypes used in experiment 2 except

hybrid TL145408). Only the Paronset and Compack hybrids were common to all three experiments.

Yield prediction

The yield prediction method for plots of tomato hybrids experiments (Figure 1) consisted of fitting multiple linear regression models to the results from experiment 1 (cross-validation dataset with 5 folds) and selection of the group of fruit clusters which are part of the best model for each evaluated trait. For each one of these groups, we created five models (one for each trait) and their overall performance was performed with the data from experiments 2 and 3 (test dataset). Finally, we identified a group of fruit clusters with the best predictive capacity.

The models were fitted for each response variable studied (MWA, MWAA, MWAAA, MW23A, and MFW) through the stepwise method with forward selection of terms, considering the maximum adjusted coefficient of determination (R^2 adj) as criteria to enter the independent variables to the model (Equation 1).

In cultivars with indeterminate growth in which the plants have two stems, the first bud immediately below the first cluster should be selected as the second stem. After bud removal, new inflorescences are emitted every two or three leaves depending on the cultivar and the environmental conditions (Alvarenga, 2013). In the field, the first cluster of the secondary stem begins developing slightly after the second cluster of the main stem. Consequently, fruit ripening and the harvest period for those two clusters are almost concomitant. This characteristic is very important when creating the model because it allows travel costs to be reduced because data for the two clusters can be collected in a single visit. For this reason, the models were always initiated with the selection of these two clusters (C2 and C3) (equation 1)

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_k X_{ki} + e_i$$

where Y is the dependent variable (yield) for the ith plot, X_{1i}, X_{2i}, ..., X_{ki} are the independent variables (fruit cluster) for the ith plot, β_0, β_1, β_2, ..., β_k are the regression parameters and e_i is

the residual error for the ith plot.

The selection of the group of fruit clusters, which are part of the best model for each evaluated trait, was based on the visual analysis (Le Clerg, 1967) of the chart R^2 adj by the number of fruit clusters. Despite the higher the number of terms in the model the higher is the predictive capacity, this relation is not linear. In general, there is a decrescent increment in the predictive capacity as the model size increases up to a point where the prediction gain is practically null, the beginning of a plateau. The selected group was the one in the plateau with fewer components.

For each group of selected fruit clusters, five models were fitted, one for each evaluated trait. A multicollinearity check for each one of these models was performed by the variance inflation factor (VIF), considering values lower than 5.0 as acceptable (Marquardt, 1980). The minimum value that VIF can assume (absence of collinearity) is 1.0.

In order to evaluate the predictive performance of these groups of clusters, the mean absolute percentage error (MAPE) was calculated (equation 2) with the test dataset.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - y_i'|}{y_i} \times 100 \ (2)$$

Where n is the number of plots Y_i', is the predicted value of the ith plot and Y_i is the observed value for the ith plot.

The MAPE was chosen to evaluate the prediction performance of the models because it measures the prediction error, which is easily interpretable (difference between observed and predicted value), in a standardized way (percentage) because the traits are on different measurement scales.

The MAPE of groups of fruit clusters was compared by Kruskal-Wallis test followed by Steel-Dwaas pairwise comparisons test, both at 5% significance.

Aiming to illustrate the predictive performance of the best group of fruit from a practical perspective of a breeder point of view, the analysis of variance followed by the Scott-Knott grouping test at 5% significance level was performed with observed and predicted data of the test dataset.

All analyses were performed using

JMP® version 15.2 (SAS Institute Inc., Cary, NC, USA, 1989-2020) except for the Scott-Knott grouping test, which was performed using R (R Development Core Team, 2016).

RESULTS AND DISCUSSION

The adjusted coefficient of determination (R^2 adj) increased with the increment of the number of clusters added to the model for each response variable (Figure 2). The fruit clusters C2, C3, C6, and C9 were selected in the case of variables MWA, MWAA, and MW23A, with R^2 adj equals 0.90, 0.92, and 0.92, respectively. In the case of MWAAA, the best model (R^2 adj = 0.93) was the one with three parameters (C2, C3, and C6). And for MFW, C2, C3, C5, C6, C7 and C10 (R^2 adj = 0.91).

The difference between the curve of training and validation datasets when the plateau was reached, was minimal for all variables studied except for MFW (Figure 2). This small difference means there is no overfitting which makes the model flexible enough to be applied to other datasets different from the one used to train the model. In the case of MFW, both curves got closer (R^2 adj 0.97 and 0.91 for training and validation, respectively) from the 6th parameter onwards (best model selected), which should not cause any decrease in the performance of the model.

For MWA (smaller fruits), there was an increment on R^2 adj, from 0.23 to 0.84, when fruit clusters in the upper part of the plant (C9) were added to the model with clusters in lower positions only (C2 and C3) (Figure 2). In contrast, the variable MWAAA (larger fruits) was best estimated using combinations of fruit clusters in the lower part of the plant (C2 and C3), with R^2 adj equal to 0.86. These findings are in accordance with results of other studies (Streck *et al.*, 1996; Azevedo *et al.*, 2010), who observed that fruits from upper inflorescences were small in greenhouse crops, and also with those who reported that fruit size was closely related to several parameters, namely the position of the fruit in the cluster and the position of the cluster in the plant (Kinet & Peet,

1997; Wamser *et al.*, 2012).

The check for multicollinearity shows that all 15 models generated with the three groups of fruit clusters selected are acceptable (VIF <5) (Table 1). As an expected trend, the higher the number of terms in the model, the higher the VIF values. Despite getting ripe concomitantly, there is no evidence of collinearity for fruit clusters C2 and C3 in any model, once the values are around 2.0.

The analysis of MAPE mean across all traits shows that the group of fruit clusters C2, C3, and C6 has an overall lower predictive capacity (16.48%) compared to the other two ones (C2, C3, C6 and C9: 14.16%; C2, C3, C5,
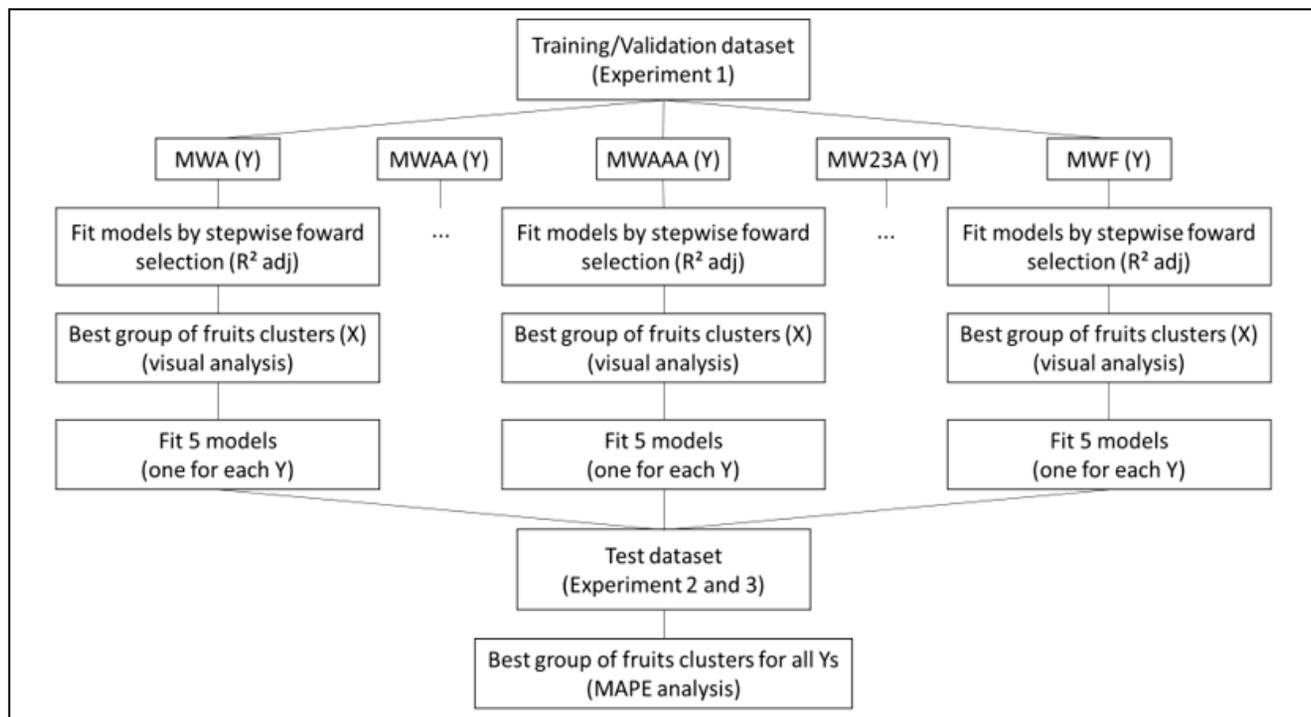
Figure 1. Methodology to fit multiple linear regressions to predict the yield of plots of tomato hybrids experiments. Y = response variable (yield); MWA, MWAA, MWAAA, MW23A and MFW = weight of marketable fruits per plant A, AA, AAAA, AA and AAA and mean weight of marketable fruits, respectively; X = independent variable (fruit cluster); R² adj = adjusted coefficent of determination; MAPE = mean absolute percentage error. Holambra, Syngenta Proteção de Cultivos, 2016-2017.

Table 1. Variable inflation factor (VIF) for the best fruit cluster models for the variables mean weight per plant of marketable fruits AAA (MWAAA), AA (MWAA), A (MWA), AA and AAA (MW23A) and mean weight of marketable fruits per plant (MFW) for training and validation data set. Holambra, Syngenta Proteção de Cultivos, 2016-2017.

Model	Term	MWA	MWAA	MWAAA	MW23A	MFW
	C2	1.1669	1.6862	1.7415	1.4419	1.1137
C2, C3, C6	C3	1.0100	2.2367	1.9587	1.9210	1.1753
	C6	1.1768	1.6289	1.4957	1.6074	1.2634
	C2	1.1735	1.7298	1.8633	1.4814	1.1142
C2, C3, C6, C9	C3	1.0141	2.2399	2.6321	1.9490	1.2384
	C6	1.7441	2.0246	1.9372	1.9311	2.5209
	C9	1.6344	1.6138	2.2587	1.6345	2.5538
	C2	1.2765	1.7350	1.9027	1.5067	1.2345
	C3	1.2814	2.5186	3.5977	2.3702	1.2074
C2, C3, C5, C6, C7, C10	C5	2.4823	2.2099	2.9203	2.6031	3.5653
	C6	2.5905	2.7431	1.7396	2.7282	3.1243
	C7	3.3434	2.3012	2.6161	2.4193	3.4819
	C10	2.5228	1.4154	1.0976	1.4208	1.6344

C2 to C10 = identification of the position of fruit clusters on the tomato plant. The second fruit cluster in the main stem was numbered 2; the first cluster of the secondary stem was numbered 3, and so forth, in a zigzagging pattern up.

C6, C7 and C10: 11,77%) (Table 2). This behavior is also valid for the isolated analysis of variable MWA. The performance of the three models was similar in the case of MWAA, MWAAA and MFW. There was no evidence to

statistically differentiate the models with four and six fruit clusters for any variable or the overall mean.

A fresh market tomato hybrids experiment requires, on average, seven visits during the harvest period. If the

number of harvest trips can be reduced through the usage of a predictive model, for example, one visit to harvest clusters 2 and 3 and two or three to harvest the other two fruit clusters, the total number of trial visits and their cost can be significantly reduced. Therefore, the proposed multiple linear regression models are consistent with the aims of the study and hybrid development.

Considering the balance between operational efficiency (number of fruit clusters in the model) and the predictive capability (MAPE), the group of fruit clusters selected, as the best for plot yield prediction of fresh market tomato hybrids experiments, was C2, C3, C6 and C9 and the regression parameters for the prediction at plot level for each trait evaluated are available in Table 3. In essence, only these four fruit clusters of the ten plants/plot can be assessed and data feed into the regression models to predict the total yield of each plot.

The grouping (Scott-Knott test) was

Table 2. Mean absolute percentage error (MAPE) for the best fruit cluster models for the variables mean weight per plant of marketable fruits AAA (MWAAA), AA (MWAA), A (MWA), AA and AAA (MW23A) and mean weight of marketable fruits per plant (MFW) for test data set. Holambra, Syngenta Proteção de Cultivos, 2016-2017.

Variable	C2, C3, C6	C2, C3, C6, C9	C2, C3, C5, C6, C7, C10
MWA	25.00 b	18.69 a	14.47 a
MWAA	7.29 a	5.34 a	5.54 a
MWAAA	39.88 a	38.99 a	30.90 a
MW23A	6.84 b	4.28 a	5.09 ab
MFW	3.41 a	3.50 a	2.86 a
Mean	16.48 b	14.16 a	11.77 a

Means in the same line followed by different letters are significantly different from each other. Kruskall-Wallis test (p≤0.05) followed by Steel-Dwaas test (p≤0.05). C2 to C10 = identification of the position of fruit clusters on the tomato plant. The second fruit cluster in the main stem was numbered 2; the first cluster of the secondary stem was numbered 3, and so forth, in a zigzagging pattern up.
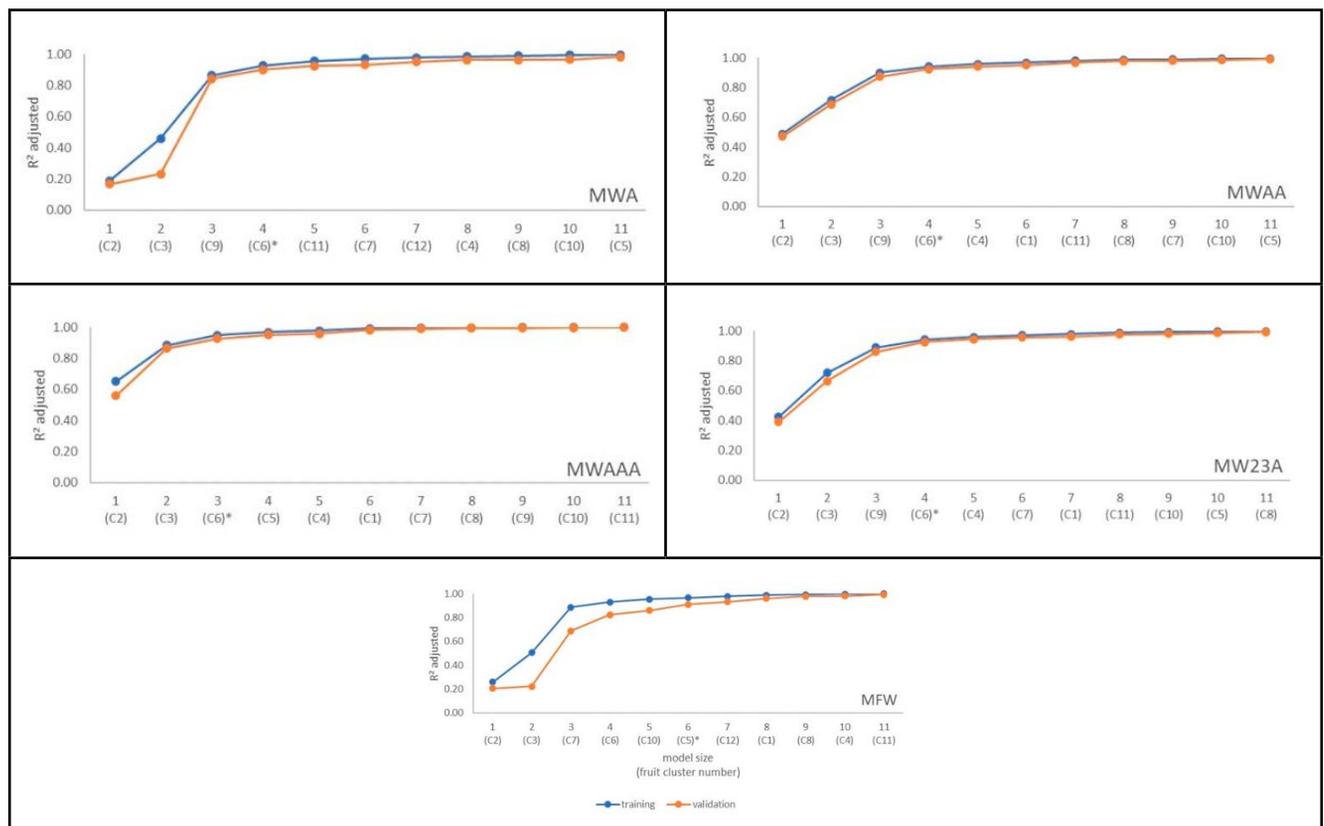
Figure 2. Adjusted coefficient of determination (R2 adjusted) as a function of the fruit clusters added to the model for the response variable weight of marketable fruits per plant AAA (MWAAA), AA (MWAA), A (MWA), AA and AAA (MW23A) and mean weight of marketable fruits (MFW) for training and validation data set. *best model selected through visual analysis. Holambra, Syngenta Proteção de Cultivos, 2016-2017.
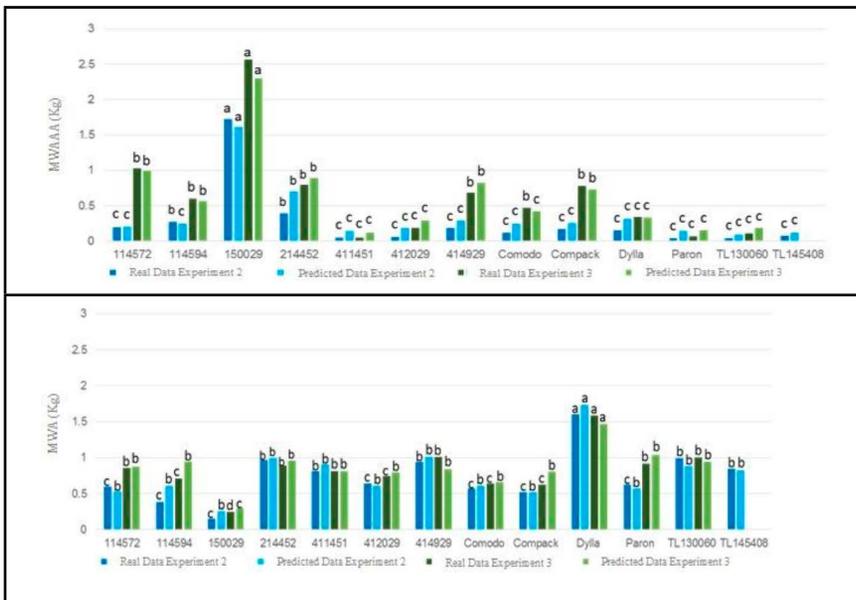
Figure 3. Observed and predicted mean weight per plant of marketable fruits AAA (MWAAA) and A (MWA) for experiments 2 and 3. Means followed by different letters within the same bar color for each variable differ significantly from each other, Scott-Knott test, p≤0.05. Holambra, Syngenta Proteção de Cultivos, 2016-2017.

Figure 4. Observed and predicted mean weight per plant of marketable fruits AA (MWAA), AA +AAA (MW23A) and mean weight of marketable fruits (MFW) for experiments 2 and 3. Means followed by different letters within the same bar color for each variable differ significantly from each other, Scott-Knott test, p≤0.05. Holambra, Syngenta Proteção de Cultivos, 2016-2017.

practically the same for the observed and predicted MWAAA values for the tested hybrids and only differed for hybrid 114594 in experiment 2 and Comodo in experiment 3 (Figure 3). A difference in the groups was observed between the observed and predicted MWA for genotypes 114594, 150029, 412029, Comodo, and Compack in experiments 2 and 3. This was expected due to the size of MAPE, 18.69 and 38.99% for MWA and MWAAA, respectively (Table 2). However, this finding does not affect decision-making because the best hybrid for each experiment remains the same. Nevertheless, it might not be the same behavior observed when the model was applied to other experiments.

For MWAA and MW23A, the groups of observed data and predicted plot values were the same (Figure 3). This outcome was expected because the errors associated with the value estimation were approximately 5% (Table 2).

Despite the low predictive performance of the model when applied to variables MWA and MWAAA, according to CEPEA (2017), larger fruits classified as AA and AAA are more important than those classified as A for the grower to obtain a return over investment. For this reason, the models can still be considered useful for predicting the yield of plots.

For the variable MFW (MAPE equals to 3.50%), considering experiment 2, no difference in the groups was observed (Figure 4). For experiment 3, the estimated values for hybrids 214452 and 414929 were underestimated. Although MFW is not important for growers because fruits are marketed by size and not by their mean weight, it is relevant for hybrid characterization given that fruit size can be used to infer fruit weight.

Of the variables studied, reliable estimates were obtained for MWAA and MW23A, as indicated by the low MAPE (Table 2), 5.34 and 4.28%, respectively, and by their similar groupings according to the Scott-Knott test (Figure 4). Although MFW did not present the same grouping order according to the Scott-Knott test (Figure 4), this variable could also be used as a reliable deciding factor
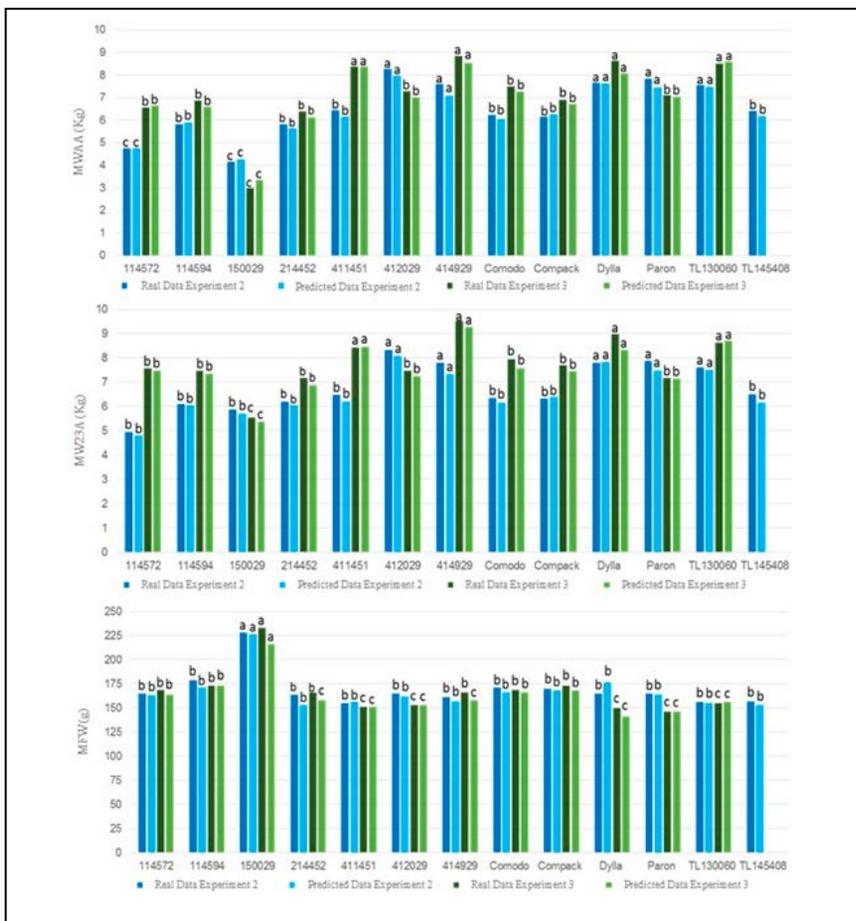
Table 3. Coefficients of the linear equations (Y = I + aC2 + bC3 + cC6 + dC9) to predict, at plot level, the variables mean weight per plant of marketable fruits AAA (MWAAA), AA (MWAA), A (MWA), AA and AAA (MW23A) and mean weight of marketable fruits (MFW). Holambra, Syngenta Proteção de Cultivos, 2016-2017.

Variable	Intercept (I)	Fruit cluster 2 (a)	Fruit cluster 3 (b)	Fruit cluster 6 (c)	Fruit cluster 9 (d)
MWA	0.1400	3.0647	1.2120	2.6018	3.5739
MWAA	0.8155	1.4808	2.3952	2.4258	3.3510
MWAAA	0.0173	1.7712	2.3144	2.1747	0.5105
MW23A	0.7973	1.4651	2.5861	2.2893	3.3164
MFW	15.6170	0.1107	0.0993	0.3096	0.3829

in tomato hybrid selection because the MAPE that was 3.50%.

REFERENCES

ALVARENGA, MAR. 2013. *Tomate: produção em campo, casa de vegetação e hidroponia.* 2 ed. rev. e ampl. Lavras: Editora UFLA. 457p.

AZEVEDO, VF; ABBOUD, ACS; CARMO, MGF. 2010. Row spacing and pruning regimes on organically grown cherry tomato. *Horticultura Brasileira* 28: 389-394.

CENTRO DE ESTUDOS AVANÇADOS EM ECONOMIA APLICADA - CEPEA. 2017. Especial hortaliças: quanto custa produzir alface, cenoura e tomate no Brasil? Available http://www.hfbrasil.org.br/br/revista/quanto-custa-produzir-hortalicas-no-brasil.aspx. Accessed September 2019.

HIGASHIDE, T. 2009. Prediction of tomato yield on the basis of solar radiation before anthesis under warm greenhouse conditions. *HortScience* 44:1874-1878.

HUSSAIN, SAI; HATIBARUAH, D. 2015. Modeling for prediction of tomato yield and its deviation using artificial neural network. *International Journal of Engineering Trends and Technology* 29: 102-108.

KINET, J; PEET, M. 1997. Tomato. In: WIEN, HC (ed). *The physiology of vegetable crops.* Wallingford, UK: CABI International. p. 207-258.

LE CLERG, EL. 1967. Significance of experimental design in plant breeding. In: FREY, KJ (ed). *Plant breeding symposium.* Ames: Iowa State University. p.243-313.

MARQUARDT, DW. 1980. You should standardize the predictor variables in your regression models. *Journal of the American Statistical Association.* 75:74-103.

MOHAMED, SM; ALI, EE; MOHAMED, TY. 2012. Study of heritability and genetic variability among different plant and fruit characters of tomato (*Solanum lycopersicon* L.). *International Journal of Scientific & Technology Research* 1:55-58.

PROGRAMA BRASILEIRO PARA A MODERNIZAÇÃO DA HORTICULTURA - PBMH. Normas de classificação do tomate. São Paulo: Centro de Qualidade em Horticultura - CQH/CEAGESP, 2003. (Documentos, 26).

QADDOUM, K; HINES, EL; ILIESCU, DD. 2013. Yield prediction for tomato greenhouse using EFuNN. *ISRN Artificial Intelligence* 2013:1-9.

R DEVELOPMENT CORE TEAM. 2016. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing.

ROTHAN, C; DIOUF, I; CAUSSE, M. 2019. Trait discovery and editing in tomato. *The Plant Journal* 97: 73-90.

SAS Institute Inc. JMP®, Version 15, Cary, NC, 1989-2020.

SOUZA, JAR; MOREIRA, DA; FERREIRA, PA; MATOS, AT. 2010. Avaliação de frutos de tomate de mesa produzidos com efluente do tratamento primário da água residuária da suinocultura. *Engenharia na Agricultura* 18: 198-207.

STAJCIC´, S; ´CETKOVIC´, G; CANADANOVIC´-BRUNET, J; DJILAS, S; MANDIC´, A; CETOJEVIC´-SIMIN, D. 2015. Tomato waste: Carotenoids content, antioxidant and cell growth activities. *Food Chemistry* 172: 225-232.

STRECK, NA; BURIOL, GA; SCHNEIDER, FM. 1996. Efeito da densidade de plantas sobre a produtividade do tomateiro cultivado em estufa de plástico. *Pesquisa Agropecuária Brasileira* 31: 105-112.

TREICHEL, M. 2016. Anuário brasileiro do tomate. 2016. Available http://www.editoragazeta.com.br/flip/anuario-tomate-2016/files/assets/basic-html/page4.html. Accessed September 2019.

WAMSER, AF; MUELLER, S; SUZUKI, A; BECKER, WF; SANTOS, JP. 2012. Produtividade de híbridos de tomate submetidos ao cultivo superadensado. *Horticultura Brasileira* 30: 168-174.