Early prediction models for cassava root yield in different water regimes

Alison Borges Vitor[a], Rafael Parreira Diniz[a], Carolina Vianna Morgante[b],
Rafaela Priscila Antônio[b], Eder Jorge de Oliveira[c],*

[a] *Universidade Federal do Recôncavo da Bahia, Campus Cruz das Almas, 44380-000, Cruz das Almas, BA, Brazil*
[b] *Embrapa Semiárido, BR 428, Km 152, Zona Rural, Caixa Postal 23, 56302-970, Petrolina, PE, Brazil*
[c] *Embrapa Mandioca e Fruticultura, Rua da Embrapa, Caixa Postal 007, 44380-000, Cruz das Almas, BA, Brazil*

ARTICLE INFO

ABSTRACT

The development of cassava (*Manihot esculenta* Crantz) varieties with greater tolerance of water deficit depends on optimized phenotyping tools. The objective of this work was to develop early prediction models of final root yield (12 months after planting - MAP) using physiological and agronomic data obtained at 4 MAP under two water regimes. Nine genotypes of cassava were evaluated under two treatments (irrigated and with water deficit), using a complete randomized block design, in a factorial scheme of 2 harvest periods (at 4 and 12 MAP) × 9 genotypes, with four replications. Both treatment groups were irrigated until 3 MAP. After this period, irrigation was interrupted for the water deficit treatment group. Fourteen physiological and agronomic traits were evaluated in all harvest periods. Four prediction models were evaluated: linear regression with stepwise selection (LRSS), linear regression with backward selection (LRBS), Bayesian ridge regression (BRR), and partial least squares (PLS). Most of the models presented a high predictive ability for final root yield (R^2 ranging from 0.83 to 0.91). However, in all prediction scenarios, the PLS model presented a high R^2 (0.84 to 0.91) associated with the lowest root-mean-square error (RMSE) (0.82 to 1.60). Differences in the predictive ability of the models may have occurred due to the relative importance of the early traits. In the case of PLS, the most important traits for the model were stomatal conductance, root yield at 4 MAP, leaf area index and number of roots. Regardless of the water condition, the physiological and agronomic data collected at an early stage could successfully be used to predict the final root yield with great efficiency. This strategy can reduce the cost of phenotyping, increasing the capacity for analysis and optimization of genetic gains for tolerance to drought in cassava.

1. Introduction

Climate change has triggered several abiotic stresses that can have negative impacts on agricultural production and food security in major food crops, including cassava. Among the abiotic stresses, water deficit promotes morphological, physiological, biochemical and molecular changes in plants that negatively affect growth and productivity in several species. The reduction of the agronomic performance of cassava under water deficit conditions depends on the traits evaluated. In general, the loss estimates are on the order of 26% for dry matter content in the roots, 33–47% for plant height, 37–55% for shoot yield, 31% for harvest index, and 38–87% for fresh root yield (Aina et al., 2007; Okogbenin et al., 2013; Adjebeng-Danquah et al., 2016; Oliveira et al., 2017).

The future challenges to production of cassava as a highly important species to global food security will be shaped by the ability to maintain or even increase cassava's production potential and root quality for diverse consumer markets while tolerating multiple biotic and abiotic stresses. Overall, the cassava root yield growth rate in the last 10 years was negative (-0.35%), although there were significant increases in South America (3.74%) and in Asia (16.07%). On the other hand, as one of the main cassava-producing regions (56% of global production), the African continent has faced strong reductions in fresh root yield (-4.10%), thus contributing to the worldwide yield reduction (FAO, 2019).

Reductions in cassava yield caused by biotic and abiotic stresses may further hamper the growing demand for cassava roots for various industrial uses (Jerumeh and Omonona, 2018). Therefore, the development of cassava varieties that are tolerant of water deficit is an important mitigation measure for this abiotic stress to satisfy the growing demand for cassava.

Although water deficit tolerance is a quantitative trait with complex phenotyping (because it is a multigenic trait with low heritability and high genotype × environment interaction), breeding programs have

---

made important contributions to identifying genotypes that are more responsive to water stress (Oliveira et al., 2017). This has been possible in part because cassava has wide natural variability and a remarkable ability to tolerate and recover from biotic and abiotic stresses (Okogbenin et al., 2013). The main mechanisms associated with the adaptation of cassava to drought combine multiple strategies to avoid dehydration, such as stomatal closure, reduction of the production of several photosynthetic proteins, and foliar abscission (Zhao et al., 2015).

In addition to the understanding of resistance mechanisms and environmental variants, any research program aiming to increase water deficit tolerance should consider plant phenology. Generally, between 3 and 4 months after planting (MAP), intense development of the photosynthetic apparatus and aerial part of the cassava plants is observed. Consequently, a water deficit in this phase causes the greatest losses due to the decrease in the cassava shoots with consequent loss of stand and reductions in number and root yield (, 2007). On the other hand, in terms of phenotyping for drought tolerance, most studies have attempted to correlate morphological and physiological data with the productive potential of the genotypes at the end of the crop cycle. This requires the experiments to be maintained for long periods (12–18 months) in field conditions, leading to high maintenance costs and long cycles for selection and validation of the potential sources of water deficit tolerance.

The search for early selection methods has been a priority in cassava research programs. Therefore, several groups have sought different alternatives to shorten the selection cycle and lower the costs associated with the phenotyping of characteristics that are complex and difficult to measure in the field. Some recent examples in the literature are i) the understanding of the inheritance of some traits for early and indirect selection and the breeding value of the parents (Ojulong et al., 2010; Ceballos et al., 2016; Nduwumuremyi et al., 2018); ii) use of genome-wide association studies and genomic selection (Oliveira et al., 2012; Wolfe et al., 2017; Elias et al., 2018; Kayondo et al., 2018); and iii) use of ground-penetration radar to predict root growth rates by detecting the total biomass of the roots during the growth cycle (Delgado et al., 2017).

Even with the implementation of these methodologies, few studies have been devoted to the creation of yield prediction models based on phenotypic data obtained in the early stages of cassava development. Prediction methods generally allow for an early assessment of crop yield to improve agricultural planning and management, as well as adequate allocation of resources in the field. In breeding programs, the establishment of accurate prediction models can allow the early selection of only the most interesting genotypes, allowing the optimization of resources.

In other crops, the development of prediction models based on morphological traits using artificial neural networks (ANNs) was able to predict forage palm yield with high precision ($R^2 = 0.87$) (Guimarães et al., 2018). Azevedo et al. (2015) also reported a high R^2 (0.92 to 0.96) when evaluating the potential use of prediction models for indirect selection against flowering using six predictors. In maize, Soares et al. (2015) evaluated the performance of artificial neural networks in yield prediction based on morphological variables of the crop. According to these authors, there was a high predictive ability ($R^2 = 1.00$), evidencing the strong correlation between the estimated values and the actual grain yield data obtained in the field experiments.

Prediction models can contribute to better experimental planning and management decision making as well as to optimizing the time required for field evaluation and the resources allocated to research programs. However, according to our knowledge, there are no predictive models of cassava root yield under water deficit conditions. Therefore, the objective of this work was to use agronomic evaluations from early harvests of different genotypes of cassava cultivated under two water conditions in the semiarid region of the Brazilian Northeast for the construction of prediction models for root yield.

2. Material and methods

2.1. Characterization of the experimental area and plant material

The experiments were carried out in the experimental area of Embrapa Semiarido, located in the irrigated perimeter of Bebedouro, Petrolina, Pernambuco State, Brazil (09°09′ S, 40°29′ W and 365 m altitude). The climate of the region is characterized by low rainfall and great thermal amplitude, being classified by the Köppen-Geiger climate classification system as a hot semi-arid climate (BSh). The time frame of the experiments was 12 months (2015–2016), during which a total precipitation of 281.51 mm was registered, with 215 mm distributed in the first three months after planting and 66.51 mm in the following 9 months. The mean temperature ranged from 25.0 to 29.5 °C and the relative air humidity from 46.3 to 72.6%.

Nine genotypes were evaluated, including four landraces (BGM0163, BGM0785, BGM0815, and BGM0279) and five improved varieties (GCP001, 9624-09, Cacau, BRS Formosa, and BRS Dourada). The choice of these genotypes was based on their contrast for water deficit tolerance, considering their productive attributes and tolerance indices (Oliveira et al., 2015, 2017) (Supplement - Table S1).

Cassava genotypes were evaluated using two treatments, i.e., irrigated and water deficit conditions. The treatment groups were harvested twice, the first four months after planting (MAP) and the second, at 12 MAP. The experiment was carried out in three-factor factorial experiment with four replications in a randomized complete block design. The factors were two treatments (irrigated and water deficit conditions), two harvest times (4 and 12 MAP), and nine cassava genotypes. The experimental plots were composed of two lines of five plants, with spacing of 0.90 m between rows and 0.80 m between plants.

The two treatments were set up on the same day and irrigated for the first 3 MAP. After this period, irrigation ceased for the water deficit treatment, and no irrigation was applied until harvest. In the other treatment, the irrigation levels were maintained throughout crop development until the harvest. A drip irrigation system localized with one lateral line per row of plants at intervals of 0.20 m with a flow rate of approximately ($0.86 L h^{-1}$) was used based on the crop's evapotranspiration. During the experiments, climatic variations were monitored by means of a meteorological station installed near the experimental area. Daily data on rainfall, temperature, air humidity and evapotranspiration were obtained, considering precipitation and evapotranspiration measurements from previous days.

2.2. Evaluations of physiological traits

The following physiological traits were evaluated at each harvesting period. 1) Leaf temperature (TempoPo) and stomatal conductance (StoCond - $mmol kg^{-1}$) were determined using the SC-1 leaf porometer. The evaluations were carried out on completely expanded leaves, located in the apical part of the plant and exposed to solar radiation, between 10:00 am and 12:00 pm. Three random plants were evaluated in each plot. 2) Foliar temperature was measured with an infrared camera E6, FLIR (TempInfr - °C), in completely expanded leaves. 3) Leaf area index (LAI) was evaluated using an Accupar Sphygmometer, which measures the photosynthetically active radiation (PAR) under and above the plant canopy. This assessment was carried out in all plots on the same day between 9:00 am and 12:00 pm. 4) Leaf osmotic potential (LOP - $mmol kg^{-1}$) was measured by collecting the first fully expanded leaf with an east orientation and then storing it at - 80 °C. Then, small leaf segments (5 cm) were dry macerated, and the leaf sap was centrifuged at 10,000 x g for 10 min at 4 °C. The resulting supernatant was used to determine osmolality (c) using a vapor pressure osmometer (Vapro 5600). The osmotic potential was obtained from the osmolality ($mmol kg^{-1}$) of leaf tissue sap based on the Van't Hoff equation (Souza et al., 2012), $\Psi = -R x T x C$, where Ψ = osmotic potential of the solution

(kPa), R = universal gas constant ($8.2\ L.kPa.ºK^{-1}.mol^{-1}$), T = absolute solution temperature (°K), and C = concentration of solutes in the solution (mol L^{-1}).

2.3. Evaluation of agronomic traits

Trials were planted during the rainy season, and harvesting took place in two rounds, namely 4 and 12 months after planting (MAP). In both harvest time, the storage roots were hand-uprooted by carefully removing the top soil, and the roots were then pulled out and used for the agronomic evaluations. Data for both treatments and both harvesting times, for the following traits, were collected from each plot: 1) plant height (PH, in m), measured from the soil level to the first apical leaf; 2) shoot yield (ShY, in t ha^{-1}), obtained by weighing the aerial parts of all plants of the plot, after root removal, and therefore including leaves and stems; 3) stem diameter (DiSt, in mm), measured using a digital caliper 0.20 m above the ground; 4) number of roots per plant (NRP), determined from the average obtained by counting the roots of five plants per plot; 5) fresh root yield (FRY, in t ha^{-1}), obtained by weighing all roots of the plot using a suspended digital scale; 6) length of roots (RoLe, in cm); 7) root diameter (RoDi, in mm), measured using a digital caliper. For these two traits, we used five marketable roots selected for each plot; and 8) dry matter content of roots (DMC, in%), measured by drying a homogeneous sample with 500 g of roots in a greenhouse with forced air circulation at 65 °C until reaching constant weight.

Leaf retention (LeRet) was measured using a scale considering the presence and distribution of leaves throughout the plant: 1 = leaves present basically on the top, covering less than 10% of the plant height; 2 = leaves present at the top and middle, covering between 11 and 25% of the plant height; 3 = leaves present at the top and middle, covering between 26 and 50% of the plant height; 4 = leaves present at the top, middle and bottom, covering between 51 and 75% of the plant height; 5 = leaves present on almost the entire plant, covering more than 75% of the plant height.

2.4. Data analysis

The physiological and agronomic variables were submitted to analysis of variance (ANOVA) using the *easyanova* (Arnhold, 2013) and *lattice* (Sarkar, 2008) packages from R software v.3.3 (R Core Team, 2017). The best linear unbiased prediction (BLUP) for each trait was performed using the linear mixed models approach. The effects of replicates, harvest times and genotypes were assumed random, whereas water treatment was considered as a fixed effect. The mathematical model used was $y = Xr + Zg + \varepsilon$, in which y is the vector of the data; r is the repeating effect vector (assumed to be corrected) added to the overall mean; g is the vector of genotypic effects (assumed to be random); ε is a vector of errors and residues (random); X is the incidence matrix for fixed effects; and Z is the incidence matrix for random effects.

2.5. Prediction models

Morphological and physiological data from the harvest at 4 MAP were used to predict the fresh root yield at 12 MAP. The physiological and morphological data from the harvest at 4 MAP from the irrigated treatment were used in the prediction of FRY in the irrigated (IT_FI) and water deficit (IT_WD) experiments. Likewise, the 4 MAP data from the experiments under water deficit were used for the prediction of FRY in the irrigated (WDT_FI) and water deficit (WDT_WD) conditions. Six prediction models were tested to verify the accuracy of prediction of FRY in treatment groups submitted to different water stresses.

2.5.1. Linear regression with stepwise selection (LRSS)
LRSS consists of iteratively adding and removing predictors in the predictive model to find the subset of variables in the dataset that results in the best performance model, that is, the model with the lowest prediction error (Gareth et al., 2014). In stepwise selection, variables are added sequentially, and those with the highest contribution are retained in the model. Then the variables that are not able to improve the fit of the model are discarded.

2.5.2. Linear regression with backward selection (LRBS)
In backward selection, the model starts with all of the predictors (complete model), and at each stage, the variables that contribute least to the model are iteratively removed. Discarding of variables ends when all remaining variables are statistically significant. The LRSS and LRBS models were implemented in the *caret* package of R software v.3.3 (R Core Team, 2017), having as an adjustment parameter the *nvmax* function that specifies the maximum number of predictors to be incorporated into the model. The *train* function provides an easy workflow for performing stepwise and backward selections using the *leaps* and *MASS* packages from R software v.3.3 (R Core Team, 2017).

2.5.3. Bayesian ridge regression (BRR)
Bayesian regression techniques are used to include regularization parameters when obtaining estimates, which are defined based on available data. This can be done by introducing non-informative priors in the hyper parameters of the model. The L2 regularization used in the ridge regression finds a maximum a posteriori estimate under a Gaussian prior for the parameters ω with precision λ^{-1}, estimated from the data as a random variable. The complete probabilistic model is as follows: $p(y|X, \omega, \alpha) = \mathcal{N}(y|X\omega, \alpha)$, assuming y to have a Gaussian distribution around $X\omega$, while α is treated as a random variable that must be estimated from the data. Bayesian ridge regression estimates a probabilistic regression model in which the priori of the parameter ω is given by a spherical Gaussian: $p(\omega|\lambda) = \mathcal{N}(\omega|0, \lambda^{-1}I_p)$. The priors for α and ω are chosen to be gamma distributions.

2.5.4. Partial least squares (PLS)
The least squares solution for multiple linear regression, $Y = XB + \varepsilon$, is given by $B = (X^TX)^{-1}X^TY$. However, the problems in this case are that 1) X^TX is singular because the number of variables (columns) in X exceeds the number of objects (lines) and 2) there is multicollinearity in the model. In this case, PLS bypasses these problems by decomposing X into orthogonal T-scores and P-loads, being $X = TP$, in addition to regressing Y not on X itself but on the first a column of the T-scores. The PLS aims to incorporate information about X and Y in the definition of scores and loads (Wehrens and Mevik, 2007).

2.6. Evaluation of models

The *train* function of the *caret* package (Kuhn, 2008) from R software v. 3.4.3 (R Core Team, 2017) was used to evaluate the effect of model tuning parameters on performance using resampling methods; to choose the optimal model across these parameters and; to estimate model performance from a training set. The prediction models and the values of the method argument, as well as the complexity parameters used by train function were: 1) Linear Regression with Backwards Selection (LRBS), using *leapBackward* method with maximum number of predictors (nvmax) as tuning parameters; 2) Neural Network (NeNet), using *neuralnet* method with tuning parameters of number of Hidden Units in Layer 1, 2, and 3; 3) Partial Least Squares (PLS), using *kernelpls* method with number of components (ncomp) as tuning parameters; 4) Bayesian Ridge Regression (BRR) using bridge method with no tuning parameters; 5) Support Vector Machines with Radial Basis Function Kernel (SVM) using *svmRadial* method with sigma and cost as tuning parameters; 6) Linear Regression with Stepwise Selection (LRSS) with *leapSeq* method and maximum number of predictors (nvmax) as tuning parameters.

To define the tuning parameter values and to validate the quality of the model fits, a cross-validation with 10 folds was performed in four replicates. The efficiency of the models was evaluated by 1) root-mean-square error (RMSE), which is the measure of the mean magnitude of the estimated errors, i.e., $RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}$, where $\hat{y}_i$ and y_i are the estimated and observed (measured) values, respectively, and n is the number of observations, and 2) coefficient of determination (R^2), which describes the proportion of the total variance in the data observed according to the model: $R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$ where $\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$ corresponds to the explained variation, and $\sum_{i=1}^{n}(y_i - \bar{y})^2$ corresponds to the variation not explained by the model.

The function *varImp* of the *caret* package (Kuhn, 2008) was used as a generic method for calculating variable importance for objects produced by *train* methods. All the prediction models used have a built-in variable importance score. The functions have an ancillary method that returns a vector indicating which predictors were used in the final model. This built–in feature selection typically fit the outcome and the predictor, and a parameter estimation are used to optimize the error rates. Usually, the *varImp* function automatically scales the importance scores to be between 0 and 100.

3. Results

3.1. Variation of physiological and agronomic traits

There were significant differences in the physiological traits between treatments (irrigated and water deficit) except for leaf osmotic potential (LOP) and foliar temperature measured by infrared (TempInfr). There were also significant differences among cassava genotypes for all traits except leaf temperature measured by the porometer (TempPo) (Table 1). In addition, there were some interactions between water treatment × genotype for TempPo, TempInfr and leaf area index (LAI); between water treatment × harvest period for all traits except LOP; and between genotype × harvest period for all traits except TempPo. These significant interactions indicate that the effects of these factors on these physiological traits cannot be analyzed individually.

In general, greater variation in stomatal conductance and leaf area index were observed between the two water treatments and the two

Table 1
Summary of the analysis of variance for five physiological traits obtained from the evaluation of nine cassava genotypes in irrigated and water deficit conditions, harvested at 4 and 12 months after planting (MAP). Petrolina, Pernambuco State, Brazil (2017).

Source of variation	DF[1]	Mean square				
		StoCond	TempPo	TempInfr	LAI	LOP
Water treatment (WT)	1	572.65**	28.28**	8.38	475.11**	0.39
Genotype (G)	8	4.79**	1.22	10.02**	5.00**	3.45**
Harvest period (HP)	1	162.16**	374.99**	232.68**	219.62**	31.70**
Block (B)	3	2.67	12.14**	8.15**	1.13	2.14
WT x G	8	2.60	1.22**	5.85**	8.10**	1.60
WT x HP	1	49.26**	51.66**	57.76**	31.02**	0.37
G x HP	8	15.18**	1.87	6.33**	5.76**	2.84**
WT x G x HP	8	5.00**	2.14	1.57	2.01	4.17
Residue	103	14696.03	4158.37	19.87	10.18	4158.37
Coeficiente of variation (%)		20.75%	2.41%	3.82%	22.59%	10.75%
General mean		526.30	37.50	37.00	4.50	599.80

1DF: degree of freedom; **, *: Significant at 1 and 5% of probability by the F test, respectively; StoCond: stomatal conductance; TempPo: Leaf temperature measured by porometer (°C); TempInfr: leaf temperature measured by infrared camera (°C); LAI: leaf area index; LOP: osmotic leaf potential (mmol kg^{-1}).

harvest periods (Supplement – Figure S1). In general, stomatal conductance was higher in irrigated compared to water deficit treatment samples. Another important observation is the greater discrepancy of stomatal conductance between irrigated and water deficit experiments at 12 MAP. At 4 MAP, the cassava genotypes under water deficit presented stomatal conductance between 100 and 800 mmol kg^{-1}, while at 12 MAP in the same treatment, the values did not exceed 375 mmol kg^{-1} (Supplement – Figure S1). The same trend was observed for the leaf area index, in which plants at 4 MAP under water deficit presented values above 3.00, while at 12 MAP, the plants in the same treatment presented values below 1.00, possibly as a result of severe water stress.

The foliar temperature was higher at 12 MAP in comparison to 4 MAP regardless of the measurement method. Possibly, this result is related to a higher environment temperature and lower leaf area index at the end of the crop cycle. Another important observation refers to the greater variation of leaf temperature evaluated by the infrared camera in comparison to the porometer (Supplement – Figure S1). In relation to the leaf osmotic potential, there was a great variation within the cassava genotypes evaluated at the different harvest periods, although at 12 MAP, the within-genotype differences between the irrigated and the water deficit experiments were not pronounced except for the BGM0279 and Cacau genotypes.

Regarding the ANOVA of the agronomic data (Table 2), except for number of roots per plant (NRP), all traits differed significantly between treatments (irrigated and water deficit). The effects of cassava genotypes and harvest period were also significant for all agronomic traits, indicating the presence of genetic variability in these genotypes for productive performance under water deficit conditions, besides the effect of water stress in these genotypes as a function of the time of the plants in the field. There were also significant interactions between water treatment × genotype for plant height (PH), shoot yield (ShY), fresh root yield (FRY), and dry matter content in roots (DMC); between water treatment × harvest period for PH, ShY, stem diameter (DiSt), FRY, root length (RoLe) and root diameter (RoDi), and DMC; and between genotype × harvest period for ShY, leaf retention (LeRet), FRY, RoLe, RoDi, and DMC. As observed for the physiological traits, with the exception of NRP, the understanding of the effect of the variables under analysis must be carried out jointly to verify the effects of the interactions on the productive performance of the cassava varieties in the northeastern semi-arid region.

For the agronomic traits of the above-ground part of the plant, the greatest variation between the water treatments occurred at 12 MAP, when the genotypes exhibited maximum productive and growth capacity, especially in the irrigated treatment, for all traits except foliar retention (Supplement – Figure S2). In general, the low difference in leaf retention at the genotypic level at 12 MAP between irrigated and water deficit experiments is due to the natural fall of cassava leaves after reaching physiological maturity.

Of all of the agronomic traits of the root, the highest variations at 12 MAP occurred for yield, diameter and dry matter content (Supplement – Figure S3). Independently of the water regime, a higher number of roots was counted at 4 MAP in comparison with 12 MAP, which can be explained by the fact that many of the roots that form in the initial phase of cassava development are not tuberized to become reserve roots.

Regarding the fresh root yield, there was not a great increase from 4 to 12 MAP, due to the growth and physiological maturation of the plants in the experiments with water deficit. Indeed, this also generally occurred in the irrigated experiment for most genotypes (except BGM0785) (Supplement – Figure S3). Another important observation is that the root length remained practically unchanged from 4 to 12 MAP. The opposite was observed for the root diameter. These results are consistent with the fact that at 4 MAP, the roots are practically formed and from this phase onward, the main changes are related to root tuberization.

Table 2

Summary of the analysis of variance for nine agronomic traits obtained from the evaluation of nine cassava genotypes under irrigated and water deficit conditions, harvested at 4 and 12 months after planting (MAP). Petrolina, Pernambuco State, Brazil (2017).

Source of variation	DF[1]	Mean square								
		PH	ShY	LeRet	DiSt	NRP	FRY	RoLe	RoDi	DMC
Water treatment (WT)	1	407.57**	1207.35**	26.84**	184.71**	10.85	667.15**	1115.61**	6907.81**	349.19**
Genotype (G)	8	21.79**	62.40**	6.64**	14.25**	24.56**	110.02**	189.76**	508.75**	62.85**
Harvest period (HP)	1	1478.19**	64.89**	977.69**	191.96**	243.83**	491.64**	23.07	18821.56**	518.13**
Block (B)	3	3.23	2.17	0.63	3.32	1.08	1.87	6.20	25.81	1.57
WT x G	8	4.96**	15.60**	3.57	0.63	1.27	35.02**	19.19	34.48	25.72**
WT x HP	1	53.72**	82.79**	2.58	48.97**	0.005	375.79**	222.63**	5232.55**	882.01**
G x HP	8	3.16	10.85**	7.83**	2.87	2.75	42.28**	32.74**	135.32**	30.52**
WT x G x HP	8	4.88**	5.78**	0.72	0.49	3.63	33.03**	25.27*	44.24	11.23**
Residue	103	0.16	9.58	0.21	5.07	0.24	2.10	7.47	15.54	1.75
Coefficiente of variation (%)		6.97%	13.11%	14.96%	8.78%	18.15%	26.25%	11.80%	15.54%	4.72%
General mean		1.8	23.6	3.1	25.7	8.18	5.5	23.15	29.14	27.76

1DF: degree of freedom; **, *: significant at 1 and 5% probability by the F test, respectively; PH: plant height; ShY: shoot yield; LeRet: leaf retention; DiSt: stem diameter; NRP: number of roots per plant; FRY: fresh root yield; RoLe: root length; RoDi: root diameter; DMC: dry matter content.

3.2. Efficiency of prediction models

When the physiological and agronomic data from the irrigated treatment group harvested at 4 MAP were used as training data to predict the FRY at 12 MAP under the same conditions (IT_FI), we found a very high predictive ability with R^2 ranging from 0.85 (LRSS) to 0.90 (LRBS), whereas the RMSE ranged from 1.52 (PLS) to 2.12 (LRSS) (Fig. 1). The LRBS and PLS models exhibited the highest R^2 (0.90 and 0.87, respectively) and lower RMSE (1.98 and 1.52, respectively) when the FRY predictions were performed in the irrigated experiments at 12 MAP. In this prediction condition, the LRSS method presented a greater variation of R^2 and RMSE in the cross-validation.

Similarly, when the physiological and agronomic data from the water deficit treatment group harvested at 4 MAP were used as training data to predict the FRY at 12 MAP under water deficit conditions (WDT_WD), a high predictive ability was also verified, especially for BRR and PLS models, based on the high R^2 (0.90 and 0.86, respectively) and lower RMSE (0.96 and 0.82, respectively) (Fig. 1). Therefore, regardless of water treatment, the physiological and agronomic data collected early could be used to predict the final FRY with great efficiency. Although all prediction models resulted in high R^2, the cross-validation information for the LRSS model indicated a greater variation in its prediction.

The predictive ability was still quite high for most of the models in the cross predictions, that is, when using the physiological and agronomic data from the irrigated treatment harvested at 4 MAP for FRY predictions at 12 MAP in the water deficit treatment (IT_WD) and vice versa (WDT_FI). However, in these cases, the variation in R^2 was higher

(0.83 to 0.88) for some models, such as LRSS and LRBS (Fig. 1). The BRR and PLS models showed high R^2 (0.84 and 0.85, respectively) and lower RMSE (1.01 and 0.82, respectively) in the IT_WD scheme. In the WDT_FI scheme, the BRR and PLS models was also quite efficient for the predictions and yielded low RMSE.

Although the LRBS, BRR and PLS models presented high predictive abilities in the IT_FI scheme, it was verified that the PLS model was more reliable compared to other prediction models due to its low RMSE, regardless the scenario of prediction (Supplement – Figure S4). Similarly, for the WDT_WD scheme, the PLS and BRR model yielded a better prediction accuracy compared to LRBS and LRSS.

In the IT_WD cross-prediction, again, PLS was accurate for FRY prediction (similarly to LRSS, LRBS, and BRR models); however, with the lowest RMSE. For the other cross-prediction (WDT_FI), we also found a higher accuracy of PLS compared to LRSS and BRR models as well as the low RMSE values (Supplement – Figure S4). Regarding the predicted and observed values (Fig. 2), the BRR, and PLS models had low standard deviation of the residuals in comparison with the set of validations performed for all training conditions. This differed from the LRSS and LRBS models, which presented high data variation in all prediction scenarios.

3.3. Key predictive traits

Regardless of the water regime, the most important traits (relative importance over 80%) for the FRY predictions at 12 MAP were StoCond, FRY and LAI when the PLS model was applied using the physiological and agronomic data from the irrigated treatment (4 MAP)
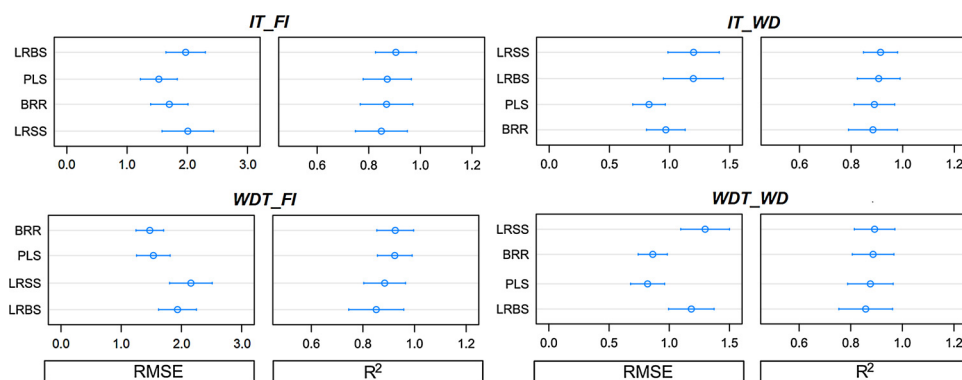

Fig. 1. Performance of the prediction models: linear regression with stepwise selection (LRSS), linear regression with backward selection (LRBS), neural network (NeNet), vector support (SVM), bayesian ridge regression (BRR), and partial least squares (PLS), evaluated by the coefficient of determination (R^2) and root-mean square error (RMSE). IT_FI and IT_WD schemes: models trained using physiological and agronomic data obtained from the irrigated treatment group harvested at 4 MAP for prediction of the fresh root yield of the irrigated and water deficit groups at 12 MAP, respectively. WDT_FI and WDT_WD: trained using the physiological and agronomic data obtained from the water deficit treatment group harvested at 4 MAP for prediction of the fresh root yield of the irrigated and water deficit groups at 12 MAP, respectively.
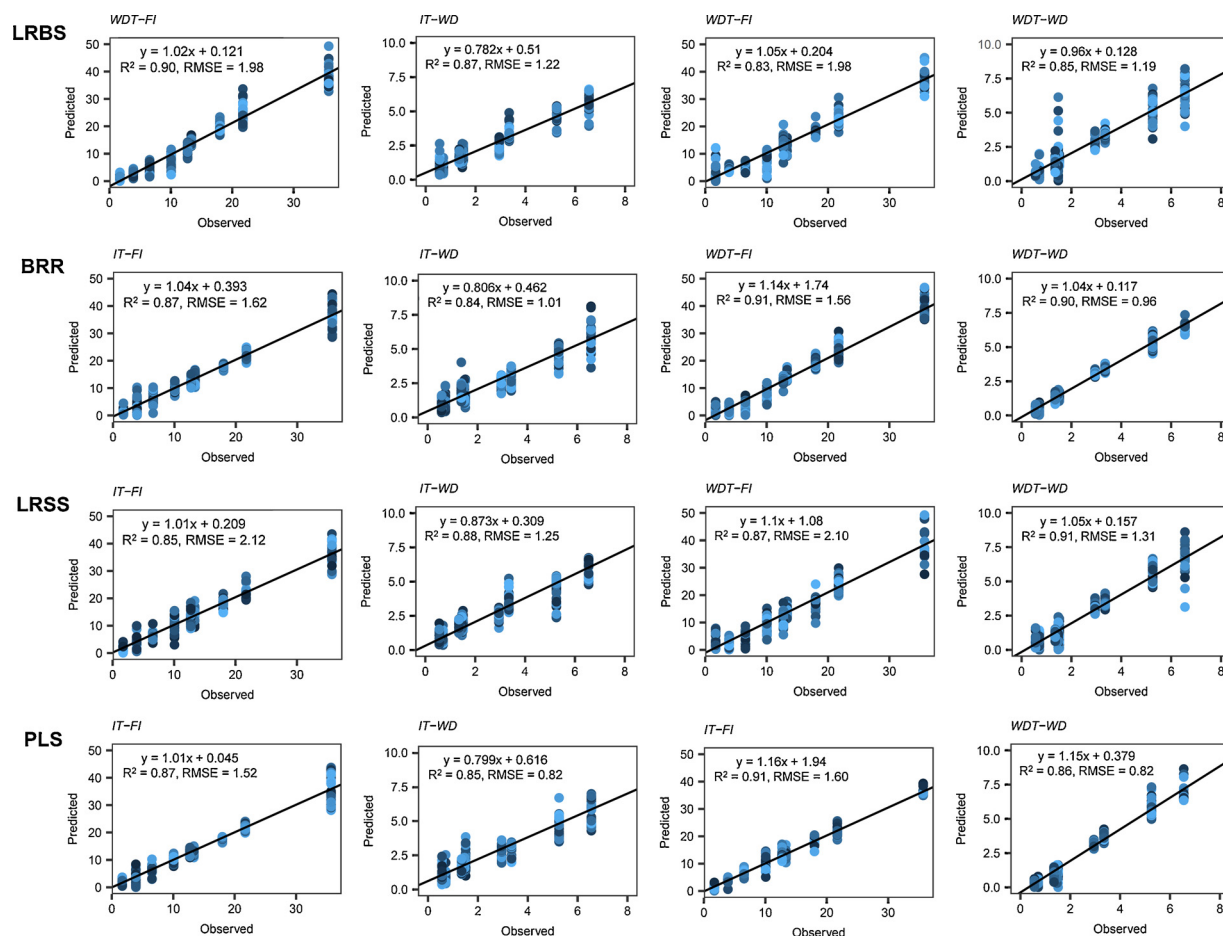
Fig. 2. Relationship between observed and predicted for fresh root yield at 12 months after planting (MAP. The prediction was performed based on six different models using cross-validation. IT_FI and IT_WD schemes: models trained using physiological and agronomic data obtained from the irrigated treatment harvested at 4 MAP used to predict the fresh root yield of the irrigated and water deficit groups at 12 MAP, respectively. WDT_FI and WDT_WD: models trained using the physiological and agronomic data obtained from the water deficit treatment harvested at 4 MAP used to predict the fresh root yield of the irrigated and water deficit groups at 12 MAP, respectively. Different colors represent each fold of the cross-validation. The numeric data included in the graphs represent: linear equations (y), coefficient of determination (R^2), and root mean square error (RMSE). The line represents the 1:1 isoline.

for training (Fig. 3). In addition to the StoCond, FRY and LAI, the other prediction models indicated TempPo, NRP, RoDi, DMC, and TempInfr as traits of high importance under these conditions of prediction (Fig. 3).

The traits FRY, RoDi, and StoCond presented high relative importance in the PLS model, independent of the water regime used for validation, when the predictions of FRY at 12 MAP used the physiological and agronomic data obtained from the water deficit treatment (4 MAP) as the training set. However, NRP was also important in the PLS model in the FRY predictions for the irrigated treatment. Similar to when the irrigated treatment data were used for training, the StoCond and FRY traits kept their relative importance in the other prediction models, while the NRP trait presented greater relative importance in the water deficit treatment used for training (Fig. 3).

Regardless of the water regime used for validation, and considering the model of higher predictive ability (PLS), the StoCond, FRY, and NRP traits collected at 4 MAP were the most important for predicting the fresh root yield at 12 MAP. On the other hand, the LAI and NRP traits presented high relative importance in the validation of the FRY in the irrigated and water deficit treatments, respectively. Therefore, in addition to its high predictive ability, the PLS model indicated that of the 14 traits evaluated, four of them (StoCond, FRY, LAI, and NRP) were most important in the early evaluations to predict future FRY.

For the remaining prediction models, independent of the water regime used for validation, the LAI, StoCond, RoDi, and DMC traits were

common for the different FRY predictions at 12 MAP. On the other hand, TempPo and NRP were unique to the prediction models using the irrigated treatment group as the training population, whereas FRY and TempInfr were exclusive to the models based on the water deficit treatment group (Fig. 3). Thus, for the LRSS, LRBS, NeNet, SVM, and BRR prediction models, seven of the 14 physiological and agronomic traits (StoCond, LAI, RoDi, DMC, NRP, FRY, TempInfr, and TempPo) were the most important.

In general, regardless of the prediction model, or the training and validation populations used, the ShY, DiSt, LOP, PH, and LeRet traits presented the lowest relative importance for FRY prediction at 12 MAP in both the irrigated and water deficit treatments.

4. Discussion

4.1. Variation of physiological and agronomic traits

Cassava plants subjected to water stress tend to decrease their physiological activity, culminating with root and shoot yield reductions (El-Sharkawy, 2006). In general, a water deficit for 30 days did not drastically influence the traits TempPo, TempIfr, or LOP. However, there was a greater within-genotype variation in StoCond and LAI between the water regimes. At 12 MAP, the same trend was observed in comparison with the 4 MAP data. This shows that even in the first month of water stress, the plants display physiological responses that
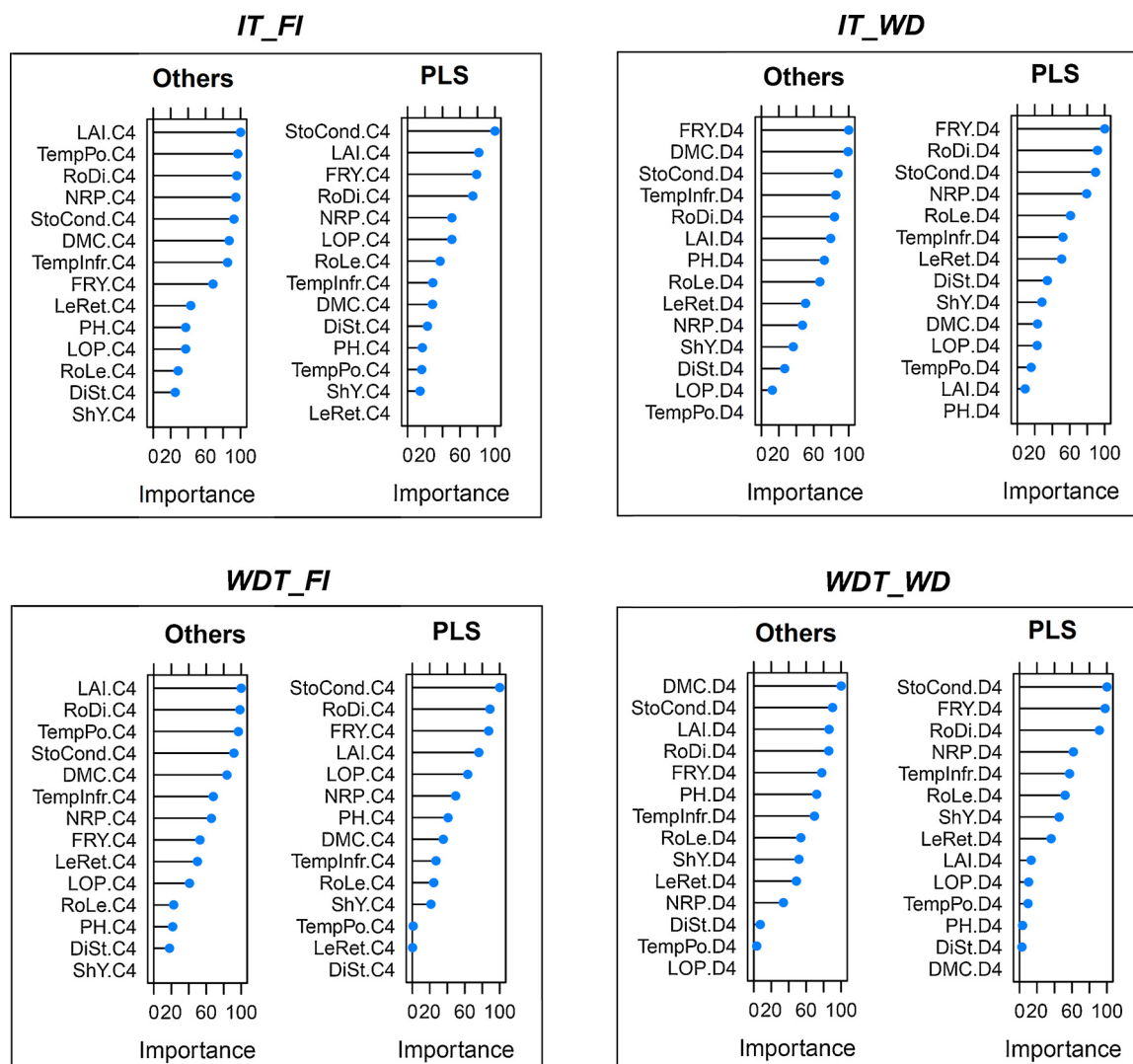
Fig. 3. Relative importance of some physiological and agronomic traits in the prediction models: linear regression with stepwise selection (LRSS), linear regression with backward selection (LRBS), neural network (NeNet), vector support (SVM), bayesian ridge regression (BRR), and partial least squares (PLS). IT_FI and IT_WD schemes: models trained using physiological and agronomic data obtained from the irrigated treatment group harvested at 4 MAP for prediction of the fresh root yield of the irrigated and water deficit groups at 12 MAP, respectively. WDT_FI and WDT_WD: trained using the physiological and agronomic data obtained from the water deficit treatment group harvested at 4 MAP for prediction of the fresh root yield of the irrigated and water deficit groups at 12 MAP, respectively. StoCond: stomatal conductance; TempPo: foliar temperature measured by porometer (°C); TempInfr: foliar temperature measured by infrared camera (°C); LAI: leaf area index; LOP: leaf osmotic potential (mmol kg^{-1}); PH: plant height; ShY: shoot yield; LeRet: leaf retention; DiSt: stem diameter; NRP: number of roots per plant; FRY: fresh root yield; RoLe: root length; RoDi: root diameter; DMC: dry matter content of the roots.

will be maintained throughout the period of water restriction. In the first months of stress, cassava can reduce the StoCond by up to 43%, as this mechanism allows the retention of water for maintenance of basic cellular activities up to 12 MAP (El-Sharkawy, 2006). However, the reduction of the stomatal opening may decrease the amount of CO_2 in the leaf, with direct effects on photosynthesis and growth (Alves, 2002; El-Sharkawy, 2007).

With respect to the shoot traits, the imposition of water stress on the plants affected the data at 4 MAP, except for leaf retention, which was not influenced. Water is needed to promote nutrient transport in regions of cell growth and multiplication (Payvandi et al., 2014). Therefore, reduced plant growth is a short-term consequence of water deficit. On the other hand, leaf abscission caused by stress occurs when water stress continues for more than 150 days, explaining the low variation in this trait between the two water treatments at 4 MAP. At 12 MAP, leaf retention was also similar between the two treatments due to the natural fall of leaves of cassava plants at this age. According to Alves (2002), foliar abscission occurs from 40 to 210 days depending on

the severity of the water stress and the susceptibility of the genotype.

The low variation of the traits related to the roots at 4 MAP is possibly due to the fact that, independently of the genotype, the development of cassava roots occurs in the first 3–4 months, with intense production of fibrous roots for water and nutrient extraction for growth (Alves, 2006). In contrast, at 12 MAP, the plant response to water deficit is genotype-dependent, resulting in large variations in root yield and other related traits.

4.2. Prediction of fresh root yield in semi-arid regions

The production capacity of cassava in marginal and unfavorable environments makes this species very important as a food security crop in areas where cassava is a staple food, as in Sub-Saharan Africa and some regions of the Brazilian northeast (Okogbenin et al., 2013; Oliveira et al., 2017). Therefore, evaluation of cassava root yield is a key feature for their cultivation in semi-arid regions, although conventional procedures for estimating cassava root yield are time-

consuming and expensive, especially when field experiments are conducted until the end of the crop cycle.

In this work, physiological and agronomic data were collected at an early stage (4 MAP) from treatment groups kept under irrigated and water deficit conditions for the prediction of FRY at the end of the production cycle (12 MAP). Most of the prediction models presented high predictive ability (R^2 ranging from 0.83 to 0.91, and RMSE ranging from 0.82 to 2.12).

Among the models, PLS exhibited the highest predictive ability and the lowest RMSE in all prediction scenarios for FRY (R^2 ranging from 0.85 to 0.91 and RMSE ranging from 0.82 to 1.60). Similar results have also been observed in other species; the PLS model presented high prediction of maize grain yield ($R^2 = 0.99$ and RMSE = 17.73) using phenological traits (Shaibu et al., 2015). According to these authors, the high ability of the PLS model, using phenological traits evaluated before harvest, to predict the grain yield of maize may reduce the time required for the development of maize varieties tolerant to water deficit.

Even using other data types (i.e., hyperspectral fluorescence imaging), the PLS model has been very accurate in the prediction of the effects of water stress on soybean cultivars (Mo et al., 2015). The results of the cross-validation of PLS showed high predictive ability ($R^2 = 0.973$ and 0.969 when hyperspectral images were taken at intervals between 8 and 6 days, respectively) in the development phase of the plants. In contrast, there is usually no consensus on the best prediction method for application in all situations (species, phenotypic data type, growing environment, etc.). For example, Zaefizadeh et al. (2011) evaluated the impact of different barley genotypes and genotype × environment interactions on crop yield and found that the neural network model presented higher predictive ability compared to linear regression methods.

The evaluation of physiological and agronomic data collected in the early stages of cassava development for prediction of FRY at the end of the cycle can be a very useful tool for the breeder because it greatly reduces the effort required to obtain the phenotypic data. In this case, even if the estimate of root yield is not highly accurate, the ranking and separation of the genotypes by best and worst performance already has a great impact on reducing the work and costs associated with phenotyping for tolerance to drought. In addition, with the screening of germplasm and segregation of populations performed every four months, it will be possible to triple the annual genotype evaluation capacity in field conditions and obviously contribute to a reduction in the breeding cycle of the species.

In wheat, Garriga et al. (2017) carried out the phenotyping of several wheat genotypes, using remote and proximal sensing techniques, for various agronomic and physiological traits associated with grain yield and drought adaptation. Those authors demonstrated that 1) the traits with the greatest predictive ability were grain yield and isotopic carbon discrimination ($\Delta^{13}C$), in both irrigated and rainfed experiments, and 2) for these traits, the partial least squares discriminant analysis (PLS-DA) method presented excellent performance in the classification of wheat genotypes into two classes, i.e., lower (80%) and higher agronomic performance (20% elite). Therefore, this study corroborates the possibility of using predictive models of phenotypic data (physiological and agronomic) evaluated in early stages to predict and classify the behavior and ranking of cassava genotypes based on productive attributes under water deficit conditions.

Overall, although the prediction of fresh root yield in cassava at the end of the cycle still remains a high-cost process, there are certainly several advantages compared to the traditional harvests and annual assessments. In addition, the risk of errors in selecting the best genotypes is minimized by the fact that screening of the genetic material occurs in its adaptation environment itself.

4.3. Contributions of physiological and agronomic traits to the prediction of fresh root yield

Some of the differences in the predictive ability may have occurred due to the relative importance of the different physiological and agronomic traits included in the early evaluation. For example, in the case of PLS, the StoCond and FRY traits evaluated at 4 MAP were the most important for the prediction of FRY at 12 MAP, independent of the water regime (irrigated or water deficit). On the other hand, the LAI trait presented high relative importance in the validation of predicted FRY in the irrigated treatment, while NRP was important for prediction in the water deficit treatment. Thus, in the PLS model, only four traits (StoCond, FRY, LAI and NRP) explained most of the data variance in the early assessments for FRY predictions.

Indeed, stomatal conductance and CO_2 photosynthetic assimilation are physiological attributes of extreme importance when cassava plants are submitted to water stress. Strong reductions occur in these characteristics, leading to carbon depletion, which is an additional challenge to the survival of cassava plants during long periods of drought (Duque, 2012). However, other mechanisms help cassava to survive in these situations. The carbohydrate reserves stored in the stem and petioles are remobilized during water stress, providing a source of carbohydrates for continuing metabolic activity (Duque and Setter, 2005). Although stomatal closure during water stress may result in higher plant canopy temperatures, the TempPo and TempInfr traits were less significant in the PLS model. However, cassava presents great sensitivity in stomatal responses, since stomatal closure may occur under high evaporative requirements, even under irrigated conditions, and at moderate levels of water scarcity (El-Sharkawy, 2006).

The other two agronomic traits identified by the PLS model as of high relative importance for prediction were FRY and NRP at 4 MAP. Root productivity at early stages of development is expected to reflect the productive potential in the final cycle of the crop, especially in the cassava, because root formation and the onset of tuberization occur up to four months after planting (Alves, 2006). From this period onward, basically only the filling of the roots takes place. Therefore, the productive potential is defined by the 4 MAP stage in the development of cassava plants. Additionally, several authors have demonstrated that the number of roots in cassava has a positive correlation with root yield (Silva et al., 2016), and therefore has a high potential for indirect selection of root yield.

For the prediction models LRSS, LRBS, and BRR, the LAI, StoCond, RoDi, and DMC traits presented greater relative importance (> 80%) in both treatments. In addition, TempPo and NRP were more important in the irrigated treatment training data, while FRY and TempInfr presented high importance when using the water deficit treatment group as the training population. According to , 2012Duque (2012), the canopy temperature is the result of evaporative cooling of the leaves due to the opening of the stomata. Thus, the temperature differences in the leaves of plants subjected to drought stress provide an idea of the genotypic differences in stomatal conductance. In general, lower canopy temperatures in plants subjected to water stress may indicate that soil moisture is still available. Therefore, the differences in leaf temperature captured by the TempPo and TempInfr measurements, as a result of water stress, explained a lot of the data variation in the prediction models LRSS, LRBS, and BRR.

In our study, the ShY, DiSt, LOP, PH, and LeRet traits presented the lowest relative importance for FRY prediction for both the irrigated and water deficit treatments. These observations contradict several reports in the literature that demonstrate that the first responses to water deficit are reductions in the growth and expansion of leaves and shoots associated with stomatal closure and foliar abscission. This causes the plants to conserve available water and limit the respiratory and carbohydrate demand during stress (El-Sharkawy, 2006; Setter and Fregene, 2007).

Another trait affected by water stress was plant height, whose

negative correlation with fresh root yield under water stress conditions can be an advantage for favoring root growth at the expense of above-ground development (Duke, 2012). On the other hand, according to the latter author, leaf retention was not significantly associated with root yield, whereas Lenis et al. (2006) have reported that leaf retention is a key to achieving high root yield. In view of these reports, it is possible to speculate that the low genetic variability of the cassava genotypes used in this study may have contributed to the low relative importance of ShY, DiSt, LOP, PH, and LeRet traits for FRY predictions at 12 MAP. Also, after a month of severe water stress (third to fourth MAP), all genotypes under water stress maintained a small number of leaves in the shoot apex, which probably provided a sufficient photosynthetic rate for plant survival under these conditions.

4.4. Perspectives for selecting cassava genotypes with water stress tolerance

The effects of climate change on climate patterns and the occurrence of unexpected events are affecting productivity gains in several plant species (Hernández-Barrera et al., 2017). This relatively recent scenario has challenged cassava breeders to accelerate the development and recommendation of new varieties with high yield and root quality as well as tolerance of more complex cropping environment conditions involving high temperatures and long drought periods, especially in semi-arid regions.

Genomic research has produced a lot of information on the genetics of various plant species, with increasingly important developments in the generation of large amounts of low-cost molecular information associated with large platforms for data storage and analysis. However, this molecular information far outweighs the current capacity of plant phenotyping (Yang et al., 2014), especially under field conditions. This "bottleneck of phenotyping" has limited the ability to understand how expressed phenotypes correlate with genetic and environmental factors, restricting progress in understanding complex traits such as drought tolerance in genetic breeding programs (Großkinsky et al., 2019; Ubbens and Stavness, 2017). Therefore, constant improvements in methodologies for phenotyping and plant selection under water stress should be pursued to keep pace with genomic evolution and, at the same time, continue to yield practical results in increasing crop yield.

It is agreed in the literature that cassava tolerance of water deficit is due to several physiological mechanisms that allow the species to withstand long periods of drought. In general, the main mechanisms involve 1) stomatal closure of the leaves to avoid damage to the photosynthetic system and maintain photosynthetic activity when water is supplied; 2) drastic reduction in leaf growth; 3) abscission of older leaves and maintenance of leaves at the apex of the plant to retain a minimum of active photosynthesis; and 4) root development in deeper soil areas for water and nutrient extraction (El-Sharkawy, 2004; Lenis et al., 2006; Duke, 2012; Okogbenin et al., 2013; Mulualem and Bekeko, 2015).

Considering the mechanisms mentioned above, several physiological and agronomic traits have been used in the evaluation of germplasm for water deficit tolerance. However, the adoption of varieties that are tolerant of water stress depends on the interest and willingness of farmers to test and continue to use these varieties. Thus, the selection criteria for new varieties should take into account characteristics that effectively meet the farmer's needs and expectations. This is necessary because the adoption of improved varieties increases when differences in yield or even in root quality between improved varieties and landraces (traditionally cultivated) are perceived by farmers. Therefore, the traits used in the present work are those considered most important for future recommendations of new varieties for semi-arid cultivation conditions, since they involve productive attributes of both roots and shoots that are sought by farmers when adopting cassava varieties.

Although our data come from only one year of cultivation, the results of this proof of concept were sufficient to determine the potential utility of this approach for early prediction of FRY in breeding programs

or even research programs that simply aim to select landraces with high water deficit tolerance. In other species, such as maize, even using smaller numbers of genotypes (6), there was a high grain yield prediction ($R^2 = 0.99$) using phenological traits as predictors (Shaibu et al., 2015). Therefore, even though further studies with a greater number of cassava genotypes and additional years of cultivation are required for methodology validation, our results suggest the possibility of identifying elite genotypes with higher yield potential under water deficit conditions by using phenotypic data collected at 4 MAP. In addition, the evaluation of physiological and agronomic traits of major relative importance, such as StoCond, FRY, LAI, and NRP, may contribute to greater efficiency of the breeding programs by accelerating the selection and release of elite varieties with better adaptation to environmental conditions in areas prone to drought.

The strategy proposed in this study can reduce the cost of phenotyping for tolerance of water deficits in cassava, considering several aspects: 1) the amount of biomass (shoot and roots) produced at 4 MAP is much lower than that produced at 12 MAP. Therefore, the time required for phenotyping is much lower. 2) Phenotyping for tolerance of water deficits at 4 MAP would allow at least three annual evaluation cycles in the same area, contributing to the optimization of the area used and to the rapid screening of germplasm and segregant populations. 3) The costs for setting up and managing the field trials would be lower because trials require only 4 months instead of 12 months. 4) The cassava breeding cycle for tolerance of water deficits would be shortened. 5) The chances of adoption of new cassava varieties would be maximized due to the fact that the selection for tolerance of water deficits is being carried out at the adaptation site itself.

In addition, although the proposed methodology does not involve high-performance phenotyping strategies, it is possible to associate a wide range of technologies to obtain non-destructive biometric information. For example, thermal, colored, fluorescence, near infrared, and hyperspectral images can be utilized to increase the capacity for analysis and optimization of phenotype processes for tolerance to drought in cassava (Rousseau et al., 2019; Mo et al., 2015).

Acknowledgments

Parte inferior do formulário

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.fcr.2019.05.017.

References

Adjebeng-Danquah, J., Gracen, V.E., Offei, S.K., Asante, I.K., Manu-Aduening, J., 2016. Genetic variability in storage root bulking of cassava genotypes under irrigation and no irrigation. Agric. Food Secur. 5, 1–12.

Aina, O.O., Dixon, A.G.O., Akinrinde, E.A., 2007. Effect of soil moisture stress on growth and yield of cassava in Nigeria. Pak. J. Biol. Sci. 10, 3085–3090.

Alves, A.A.C., 2006. Fisiologia da mandioca. In: Souza, L.S., Farias, A.R., Mattos, P.L.P., Fukuda, W.M.G. (Eds.), Aspectos socioeconômicos e agronômicos da mandioca. Embrapa Mandioca e Fruticultura Tropical, Cruz das Almas, pp. 138–169.

Arnhold, E., 2013. Package in the R environment for analysis of variance and complementary analyses. Braz. J. Vet. Res. Anim. Sci. 50, 488–492.

Azevedo, A.M., Andrade Júnior, V.C., Pedrosa, C.E., Oliveira, C.M., Dornas, M.F.S., Cruz, C.D., Valadares, N.R., 2015. Application of artificial neural networks in indirect selection: a case study on the breeding of lettuce. Bragantia 74, 387–393.

Ceballos, H., Pérez, J.C., Barandica, O.J., Lenis, J.I., Morante, N., Calle, F., Pino, L., Hershey, C.H., 2016. Cassava breeding I: the value of breeding value. Front. Plant Sci. 7, 1227.

Delgado, A., Hays, D.B., Bruton, R.K., Ceballos, H., Novo, A., Boi, E., Selvaraj, M.G., 2017. Ground penetrating radar: a case study for estimating root bulking rate in cassava

(*Manihot esculenta* Crantz). Plant Methods 13, 1–11.

Duque, L., 2012. Cassava Drought Tolerance Mechanisms Re-visited: Evaluation of Drought Tolerance in Contrasting Cassava Genotypes Under Water Stressed Environments. PhD thesis. Cornell University.

Duque, L.O., Setter, T.L., 2005. Response of Cassava (*Manihot Esculenta* Crantz) to Terminal Water Stress: ABA, Sugar and Starch accumulation/partitioning and Root Growth Under Different Water Regime Treatments. Interdrought II, Rome, Italy, pp. 509.

Elias, A.A., Rabbi, I., Kulakow, P., Jannink, J.L., 2018. Improving genomic prediction in cassava field experiments by accounting for interplot competition. G3 8, 933–944.

El-Sharkawy, M.A., 2007. Physiological characteristics of cassava tolerance to prolonged drought in the tropics: implications for breeding cultivars adapted to seasonally dry and semiarid environments. Braz. J. Plant Physiol. 19, 257–286.

El-Sharkawy, M.A., 2004. Cassava biology and physiology. Plant Mol. Biol. 56, 481–501.

El-Sharkawy, M.A., 2006. International research on cassava photosynthesis, productivity, ecophysiology, and responses to environmental stresses in the tropics. Photosynthetica 44, 481–512.

FAO Food and Agriculture Organization of the United Nations, database. Accessed on 20 November 2017. http://faostat3.fao.org/analysis/Q/QC/E.

Garriga, M., Romero-Bravo, S., Estrada, F., Escobar, A., Matus, I.A., Del Pozo, A., Astudillo, C.A., Lobos, G.A., 2017. Assessing wheat traits by spectral reflectance: do we really need to focus on predicted trait-values or directly identify the elite genotypes group? Front. Plant Sci. 8, 1–12.

Großkinsky, D.K., Svensgaard, J., Christensen, S., Roitsch, T., 2019. Plant phenomics and the need for physiological phenotyping across scales to narrow the genotype-to-phenotype knowledge gap. J. Exp. Bot. 66, 5429–5440.

Guimarães, B.V.C., Donato, S.L.R., Azevedo, A.M., Aspiazú, I., Silva Junior, A.A., 2018. Prediction of 'Gigante' cactus pear yield by morphological characters and artificial neural networks. Rev. Bras. Eng. Agrã̂â¿â½cola E Ambient. 22, 315–319.

Hernández-Barrera, S., Rodriguez-Puebla, C., Challinor, A.J., 2017. Effects of diurnal temperature range and drought on wheat yield in Spain. Theor. Appl. Climatol. 129, 503–519.

Gareth, J., Witten, D., Hastie, T., TibshiranI, R., 2014. An Introduction to Statistical Learning: With Applications in R. Springer Publishing Company, Incorporated, pp. 426.

Jerumeh, T.R., Omonona, B.T., 2018. Determinants of transition in farm size among cassava-based farmers in Nigeria. Kasetsart J. Soc. Sci. 1–7. https://doi.org/10.1016/j.kjss.2018.02.008. in press.

Kayondo, S.I., Del Carpio, D.P., Lozano, R., Ozimati, A., Wolfe, M., Baguma, I., Gracen, V., Offei, S., Ferguson, M., Kawuki, R., Jannink, J.L., 2018. Genome-wide association mapping and genomic prediction for CBSD resistance in Manihot esculenta. Sci. Rep. 8, 1549.

Kuhn, M., 2008. Caret package. J. Stat. Softw. 28, 1–26.

Lenis, J.I., Calle, F., Jaramillo, G., Perez, J.C., Ceballos, H., Cock, J.H., 2006. Leaf retention and cassava productivity. Field Crops Res. 95, 126–134.

Mo, C., Kim, M.S., Kim, G., Cheong, E.J., Yang, J., Lim, J., 2015. Detecting drought stress in soybean plants using hyperspectral fluorescence imaging. J. Biosyst. Eng. 40, 335–344.

Mulualem, T., Bekeko, Z., 2015. Assessment of conventional breeding on cassava and its physiological adaptive mechanisms: implication for moisture stress. Asian J. Agric. Res. 9, 38–54.

Nduwumuremyi, A., Melis, R., Shanahan, P., Theodore, A., 2018. Genetic inheritance of pulp colour and selected traits of cassava (*Manihot esculenta* Crantz) at early generation selection. J. Sci. Food Agric. 98, 3190–3197.

Ojulong, H., Labuschagne, M.T., Herselman, L., Fregene, M.A., 2010. Yield traits as

selection indices in seedling populations of cassava. Crop. Breed. Appl. Biotechnol. 10, 191–196.

Okogbenin, E., Setter, T.L., Ferguson, M., Mutegi, R., Ceballos, H., Olasanmi, B., Fregene, M., 2013. Phenotypic approaches to drought in cassava: review. Front. Physiol. 4, 1–15.

Oliveira, E.J., Resende, M.D.V., Santos, V.S., Ferreira, C.F., Oliveira, G.A.F., Silva, M.S., Oliveira, L.A., Aguilar-Vildoso, C.I., 2012. Genome-wide selection in cassava. Euphytica 187, 263–276.

Oliveira, E.J., Aidar, S.T., Morgante, C.V., Chaves, A.R.M., Cruz, J.L., Coelho Filho, M.A., 2015. Genetic parameters for drought-tolerance in cassava. Pesquisa Agropecuária Brasileira 50, 233–241.

Oliveira, E.J., Morgante, C.V., Aidar, S.T., Chaves, A.R.M., Antonio, R.P., Cruz, J.L., Coelho Filho, M.A., 2017. Evaluation of cassava germplasm for drought tolerance under field conditions. Euphytica 213, 188.

Core Team, R., 2017. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria accessed on 20 December 2017.

Rousseau, C., Belin, E., Bove, E., Rousseau, D., Fabre, F., Berruyer, R., Guillaumès, J., Manceau, C., Jacques, M.A., Boureau, T., 2019. High throughput quantitative phenotyping of plant resistance using chlorophyll fluorescence image analysis. Plant Methods 9, 17.

Sarkar, D., 2008. Lattice: Multivariate Data Visualization With R. Springer-Verlag, New York, NY ISBN 978-0-387-75968-5. 268.

Setter, T.L., Fregene, M.A., 2007. Recent advances in molecular breeding of cassava for improved drought stress tolerance. In: Jenks, M., Hasegawa, P., Jain, M. (Eds.), Advances in Molecular-Breeding Toward Drought and Salt Tolerant Crops. Springer, Berlin, Germany, pp. 701–711.

Shaibu, A.S., Adnan, A.A., Umar, I.R., 2015. Predicting grain yield of maize using drought tolerance traits. Afr. J. Agric. 10, 3332–3337.

Silva, R.S., Moura, E.F., Farias Neto, J.T., Sampaio, J.E., 2016. Genetic parameters and agronomic evaluation of cassava genotypes. Pesquisa Agropecuária Brasileira 51, 834–841.

Soares, F.C., Robaina, A.D., Peiter, M.X., Russi, J.L., 2015. Predição da produtividade da cultura do milho utilizando rede neural artificial. Ciência Rural 45, 1987–1993.

Souza, E.R., Freire, M.B.G.S., Cunha, K.P.V., Nascimento, C.W.A., Ruiz, H.A., Lins, C.M.T., 2012. Biomass, anatomical change sand osmotic potential in *Atriplex numulária* L indl. cultivated in sodic saline soil under water stress. Environ. Exp. Bot. 82, 20–27.

Ubbens, J.R., Stavness, I., 2017. Deep plant phenomics: a deep learning platform for complex plant phenotyping tasks. Front. Plant Sci. 8, 1190.

Wehrens, R., Mevik, B., 2007. The pls Package: principal component and partial least squares regression in R. J. Stat. Softw. 8, 1–23.

Wolfe, M.D., Del Carpio, D.P., Alabi, O., Ezenwaka, L.C., Ikeogu, U.N., Kayondo, I.S., Lozano, R., Okeke, U.G., Ozimati, A.A., Williams, E., Egesi, C., Kawuki, R.S., Kulakow, P., Rabbi, I.Y., Jannink, J.L., 2017. Prospects for genomic selection in cassava breeding. Plant Genome 10, 1–19.

Yang, W., Guo, Z., Huang, C., Duan, L., Chen, G., Jiang, N., Fang, W., Feng, H., Xie, W., Lian, X., Wang, G., Luo, Q., Zhang, Q., Liu, Q., Xiong, L., 2014. Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. Nat. Commun. 5, 5087.

Zaefizadeh, M., Khayatnezhad, M., Gholamin, R., 2011. Comparison of multiple linear regressions (MLR) and artificial neural network (ANN) in predicting the yield using its components in the hulless barley. Am.-Eur. J. Agric. Environ. Sci. 10, 60–64.

Zhao, P., Liu, P., Shao, J., LI, C., Wang, B., Guo, X., Yan, B., Xia, Y., Peng, M., 2015. Analysis of different strategies adapted by two cassava cultivars in response to drought stress: ensuring survival or continuing growth. J. Exp. Bot. 66, 1477–1488.