Proposal Category: _X__Research ___Outreach

Proposal Status: _X_ New Proposal ___Previously funded by SRSFC; has been previously funded for ___ years

Title:

Prediction of fruit firmness across multiple southern highbush blueberry cultivars using sugars and acids

Name, Mailing and Email Address of Principal Investigator(s):

Savithri Nambeesan
Associate Research Scientist/ Faculty
University of Georgia/ CAES
Department of Horticulture
1111 Plant Science Building
Athens, GA 30602
sunamb@uga.edu
(706) 542-0777

Lynne Seymour
Associate Professor
Department of Statistics
310 Herty Dr.
University of Georgia
Athens, GA 30602
seymour@uga.edu
(706) 542-3307

Objective:

The main objective of the proposal is to predict fruit firmness (compression and puncture) during postharvest storage using fruit metabolic traits (sugars and acids) in southern highbush cultivars.

Justification and Description:

Blueberries are a perishable commodity and deteriorate quickly after harvest. Therefore, the shelf-life of blueberry is relatively short, ranging from 1-6 weeks, depending on the cultivar, harvesting, handling, and storage method. The main causes of decreased fruit quality during postharvest storage are water loss, increase in fruit softening, and decay caused by postharvest pathogens (Li et al., 2011; Mehra et al., 2013; Paniagua et al., 2013). Although breeding efforts to improve fruit quality are concentrated towards increasing fruit firmness, currently there is no information on the role of total soluble solids (TSS), titratable acidity (TA) and other quantities of metabolites (individual sugars and acids) on predicting fruit quality. If we are able to identify the metabolic traits that are key determinants of fruit firmness, this information can be used in breeding efforts to select for cultivars with improved fruit quality attributes. This is especially important because wholesale buyers and consumers pay attention to the appearance and firmness of fruits, which are major factors associated with fruit quality (NCSU Extension [Boyette et al.]; Maclean and Nesmith, 2011). Approaches to identify key metabolites influencing fruit quality and improved agronomic traits have previously been applied to tomato (Gómez-Romero et al., 2010), peach (Lombardo et al., 2011), grapes (Degu et al., 2014), and strawberry (Zhang et al., 2011).

In Georgia, southern highbush (species complex between *Vaccinium corymbosum* L. and *V. darrowii* Camp.) and rabbiteye (*V. virgatum* Aiton) are commonly grown blueberries. Southern highbush blueberry fruit ripen early and growers get a premium price for these fruit. However, rabbiteye blueberries are native to the southeastern United States and therefore make up a significant portion of the blueberry industry in this region. Therefore, in our initial analysis we attempted to determine if fruit chemical traits such as TSS and TA predict fruit firmness in both these types of blueberries (see preliminary data presented below). Since TSS and TA are composite measurements of multiple sugars and acids, separating their individual components will provide better resolution and more information of predictors of fruit quality. For example, a previous study in blueberries indicated a positive correlation between fruit firmness and quinic acid, and a negative correlation between fruit firmness and shikimic acid (Montecchiarini et al. 2018). However, this study was performed using only three southern highbush blueberry cultivars at immature green and ripe stages. In this study, we plan to evaluate additional stages during postharvest storage in multiple southern highbush blueberry cultivars. Since this analysis is time consuming and expensive, here we propose to perform metabolite analysis to assay for individual sugars and acids only for southern highbush blueberry.

Preliminary Results: Previously we collected information on fruit quality attributes in southern highbush and rabbiteye blueberry cultivars from 2015-2018. Briefly, ripe fruit were collected from the Alapaha Research farm (UGA) and other commercial farms. These fruits were sorted to remove defective fruit and then stored in clamshells in a walk-in cooler at 4° C under high relative humidity. Fruit quality measurements were performed at regular intervals (up to 6-7 weeks). Fruit quality measurements included visual quality inspection of bruised fruit. Using this parameter an index of percent healthy fruit was developed, which is defect free fruit over total number of fruits assayed. Other measurements included, compression (fruit firmness), puncture (skin toughness), TSS, TA, pH and fruit weight. This information was used to develop the regression model for predicting fruit firmness (compression and puncture) using the fruit's physical traits. Below we describe how fruit compression and fruit puncture were predicted. We utilized stepwise linear regression model for the prediction of fruit firmness using the physical traits parameter as described below:

$$y_i=\beta_0 + \beta_1 x_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \dots\dots\dots + \beta_p X_{ip} + \varepsilon,$$

where y_i = dependent variable, X_i = independent variable, β_0 = intercept, i = number of observations, β_p = slope coefficients, and ε = model error.

Our model suggested that fruit compression is associated with the puncture values, healthy berries (%), TA, fruit weight, TSS, and pH (Table 1). Similarly, fruit puncture is significantly associated with compression, TA, Type (SHB), and percent healthy berries (Table 1). From these predictions it was not surprising that compression and puncture values were associated with each other. Compression measures fruit firmness (skin and flesh firmness) and puncture is specific to skin toughness. Thus, a positive association between them was expected (Table 1). TSS is negatively associated with fruit compression. If fruit contain higher sugars (and soluble solids), it may lead to increased water uptake and cell turgor and influence cell wall properties such that fruit firmness is lowered. However, this is speculative and warrants further investigation. Surprisingly, compression was negatively associated with TA, whereas puncture is positively associated with it.

As mentioned earlier, TA is a composite measurement of several acids such as malic, citric and quinic acid. Individual acids may allow for better separation of effects on fruit firmness and skin toughness. Therefore, in this proposed work we performed follow-up studies to explore association between fruit firmness and skin toughness with individual sugars and acids using several machine learning methods (regression) as described below.

Table 1. Model predicting the fruit firmness (compression Left, and puncture: Right) by stepwise linear regression model. Significantly associated variables are presented below.

Compression		
Fruit quality traits	Estimate	*P-value*
(Intercept)	0.2029	<0.0001
Puncture	0.7554	<0.0001
Healthy berry (%)	0.0005	<0.0001
TA	-0.0307	0.0115
Fruit weight	0.0129	<0.0001
TSS	-0.0032	<0.0001
pH	-0.0275	<0.0001

Puncture		
Fruit quality traits	Estimate	*P*-value
(Intercept)	0.0478	<0.0001
Compression	0.3939	<0.0001
TA	0.0653	<0.0001
Type (SHB)	0.0165	<0.0001
Healthy berry (%)	-0.0002	<0.0001

Significance

Southern highbush and rabbiteye are the major types of blueberry type grown in Georgia and the southeastern US. Variations in fruit physical and chemical traits (metabolites) are found among cultivars. It is hypothesized that fruit firmness is associated with these physical and chemical traits. Validation and usage of various machine learning methods (regression model) can help predict fruit firmness. This can help identify major physical and chemical traits that are associated with the fruit's firmness. In the future, this information can be used to incorporate into the southeastern blueberry breeding program and help select new cultivars. Overall, this knowledge can benefit the blueberry grower, consumers, and the industry by increasing blueberry fruit firmness and shelf life.

Description of Procedures:

For the proposed work as described previously, we measured individual sugars and acids in multiple southern highbush type blueberries.

Table 2. Fruit material for which data have been collected.

Cultivar	**2015**	**2017**
Suziblue	x	x
Rebel	x	x
Farthing		x
Emerald		x
Miss Lilly		x
Miss Alice Mae		x
Miss Jackie		x

Fruit physical traits (compression, puncture, TSS, TA, pH, and weight at various postharvest stages were measured in two years as presented above (Table 2). Fruits samples were already collected and stored at -80 ℃. We have already performed metabolite analyses using gas chromatography on two cultivars (Suziblue and Rebel) in 2015 and 2017. Having two years of data for two cultivars will help to determine variability across years. For the proposed work, metabolite analysis was performed on five more cultivars: Farthing, Emerald, Miss Lilly, Miss Alice Mae, and Miss Jackie. Including five more cultivars will give us the statistical power to make useful predictions. Due to handling of multiple samples from various cultivars, fruit measurement time-points were slightly different among cultivars. This should not affect chemical analysis since differences of 1-3 days is unlikely to dramatically affect metabolite composition and prediction models. Metabolite analyses were performed on samples collected during postharvest (PH) storage at < 1 week, approximately 2 weeks, and at three weeks of storage. Four replicates were used at each time point for each cultivar.

Metabolite profiling

Identification of compounds was performed using gas chromatography-mass spectrometry (GCMS) equipped with a 5973 quadrupole mass spectrometer detector (Agilent Technologies 6890N Network GC system) and an HP-5 fused capillary column (J&W Scientific, Fulsom, CA, USA) was employed. The quantification of compounds was performed by using the GC-flame ionization detector GC-FID (GC-2014; Shimadzu, Japan). The method set up is similar for GCMS and GC-FID as described below. The extraction protocol was performed according to Chapman and Horvat (Chapman Jr & Horvat, 1989) with some modifications. Around 100-150 mg of frozen grounded samples was extracted with 100% methanol, followed by centrifugation at 22,000 *g* for 30 minutes. After that, 100 µL of supernatant was transferred into a GC-vial. Supernatants were evaporated under nitrogen gas at 45 °C. 50 µL of methoxyamine-HCl (20 mg metoxyamine in 1 ml pyridine) was added to each sample and heated at 50 °C for 30 minutes to make the oxime derivatives. Finally, derivatization of compounds was performed by adding 100 µL of N-Methyl-N-(trimethylsilyl) trifluoroacetamide (MSTFA) + 1% TMCS (trimethylchlorosilane) and heating at 50 °C for 30 minutes. In the GC-FID, helium was used as a carrier gas. The initial temperature of the oven was set up at 120 °C for 1 minute, then 4 °C per minute ramped to 180 ℃, 0.5 min at 180 °C, 0.5 ℃ per minute ramped to 185 ℃, 0.5 min at 185 ℃, 1 ℃ per minute ramped to 210 ℃, 0.5 minutes at 210 ℃, 10 °C per minute ramped to 260 °C, and finally held for 12 minutes at 260 °C. A standard solution was prepared for each of the identified metabolites. The standards were extracted and derivatized as described for the fruit samples. Standard curves were generated individually for each metabolite and used for the quantification.

Statistical analysis:

In order to predict fruit firmness using the metabolite data, we looked at the variance inflation factors (VIP) value to see if we have multicollinearity problems in our data. After that, we analyzed our data using LASSO (Least absolute shrinkage and selection operator).

LASSO regression: The LASSO regression model was used to predict compression or puncture as a dependent variable and fruit metabolite as an independent variable. Fruit metabolite measurements during the PH storage in 2015 and 2017 and their respective compression and puncture at the same time points were taken for the analysis. LASSO finds the optimal regression coefficients by minimizing the function given as below (James et al., 2013).

$$J = \mathrm{RSS} + \lambda \sum_{j=1}^{p} |\beta_j|, \text{ and } \quad RSS = \sum_{i=1}^{n} \left(Y_i - \beta_0 - \sum_{j=1}^{p} \beta_j \, x_{ij}\right)$$

Where, J is the objective function to be minimized, RSS is the residual sum of squared, $\lambda \sum_{j=1}^{p} |\beta_j|$ is the regularization component and λ is the penalty term, p is the number of predictor variable, Y_i is the target variable for the i^{th} data, β_0 is the intercept, β_j is the coefficient for the j^{th} variable, and x_{ij} is the value for the j^{th} predicted variable for the i^{th} data.

We used the R package 'glmnet' to fit the linear regression with LASSO. The train and test data sets were divided into 80% and 20%, respectively, and the parameter was optimized using 10-fold cross-validation. The optimal λ values were selected based on the minimum mean squared error (MSE) via cross-validation. Finally, the coefficients of the most important variables, variable importance, and predicted vs. observed values were identified and presented.

Result and discussion*:*

In the metabolite data, we encountered issues of multicollinearity, particularly concerning fructose, glucose, and sucrose, which had variance inflation factors of 247, 277, and 25, respectively (Table 3). In such a scenario, the SLR model does not perform well due to the collinearity issues, that can lead to overfitting and a loss of robustness in the models (Saranwong et al., 2001; Næs et al., 2002; Nicolai et al., 2007). Hence, we employed the LASSO regression model to predict fruit firmness using the metabolite data. The LASSO regression model reduces the coefficients of less important variables in the dataset to zero, retaining only the most important variables for model prediction (Ljubobratović et al., 2022). This approach helps address the issue of multicollinearity and enhances the model's performance and interpretability.

In this work, we identified several metabolites that were either positively or negatively associated with fruit compression and puncture predictions, as presented in Table 4. Notably, serine, quinate, citrate, and myo-inositol exhibited positive associations with fruit compression. Conversely, sucrose, fructose, xylose, shikimate, succinate, glutamate, threonine, aspartate, anthocyanin, and malate were negatively associated with fruit compression. Similarly, for puncture prediction, sugars like fructose and xylose showed negative associations, while acids like quinate and citrate were positively associated (Table 4). These findings suggest that cultivars with higher citrate and quinate concentration along with a reduction in malate levels, are associated with higher fruit firmness. In blueberries, a higher concentration of citrate and phenylalanine in ripe fruit is positively correlated with fruit firmness, while xylose, leucine, and shikimate show a negative correlation (Montecchiarini et al., 2018).

Additionally, decrease in sugar concentration may enhance fruit firmness. An increase in sugar import in the fruit from source tissue increases water influx during fruit ripening. If an increase in water uptake can increase cell turgor and associated decline in fruit firmness due to pressure exerted on the cell wall will warrant further investigation. However, it is essential to consider that high sugars and low acids in ripe fruit play a significant role in contributing to the flavor of the fruits.

The variable importance plot provides insights into the contribution of each variable to the model's prediction. In the compression prediction, serine made the most significant contribution, followed by quinate, shikimate, sucrose, and succinate (Figure 1). On the other hand, aspartate, citrate, and malate made relatively lesser contributions in this context. Similarly, for the puncture prediction, fructose was the most influential variable, followed by glutamine, quinate, and myo-inositol. In contrast, citrate, glycerate, and glutamate were among the variables that made the least contribution (Figure 1). Lastly, the prediction versus observed plot demonstrated the model's accuracy, yielding an R^2 value of 0.70 for the compression prediction and 0.60 for the puncture prediction (Figure 1A-B), showing the model's effectiveness in explaining the observed data.

Table 3: Variance inflation factors (VIF) of depended variables

Dependent variables	VIF
Stages	6.5
Succinate	4.3
Glycerate	3.7
Serine	8.1
Threonine	7.6
Malate	7.2
Aspartate	8.8
Glutamate	2.6
Xylose	2.1
Glutamine	7.3
Shikimate	2.6
Citrate	4.4
Quinate	2.5
Fructose	246.9
Glucose	277
Myo-inisitol	6
Sucrose	25.2
Anthocyanin	3.6

Table 4: Prediction of Fruit Firmness (Compression & Puncture) using LASSO Regression Model with metabolites during postharvest storage. We conducted a study to predict fruit firmness, both in terms of compression and puncture, utilizing a LASSO regression model. The model incorporated various primary and secondary metabolites as predictor variables.

Fruit compression Prediction		Fruit puncture Prediction	
Fruit chemical traits	Estimate	Fruit chemical traits	Estimate
(Intercept)	0.2501	(Intercept)	0.1569
Serine	0.0244	Fructose	-0.0091
Quinate	0.0209	Glutamine	0.006
Shikimate	-0.0176	Quinate	0.0054
Sucrose	-0.0143	Myo-inisitol	-0.0052
Succinate	-0.0115	Shikimate	-0.0035
Xylose	-0.0076	Xylose	-0.0031
Glutamate	-0.0072	Malate	-0.0028
Threonine	-0.0066	Glutamate	-0.0009
Fructose	-0.005	Glycerate	0.0009
Anthocyanin	-0.0046	Citrate	0.0006
Myo-inisitol	0.0034	R^2	0.60
Malate	-0.0008	RMSE	0.0719
Citrate	0.0007		
Aspartate	-0.0001		
R^2	0.70		
RMSE	0.0212		

R^2= coefficient of determination and RMSE=root mean square error determines from the respective test samples of compression and puncture.

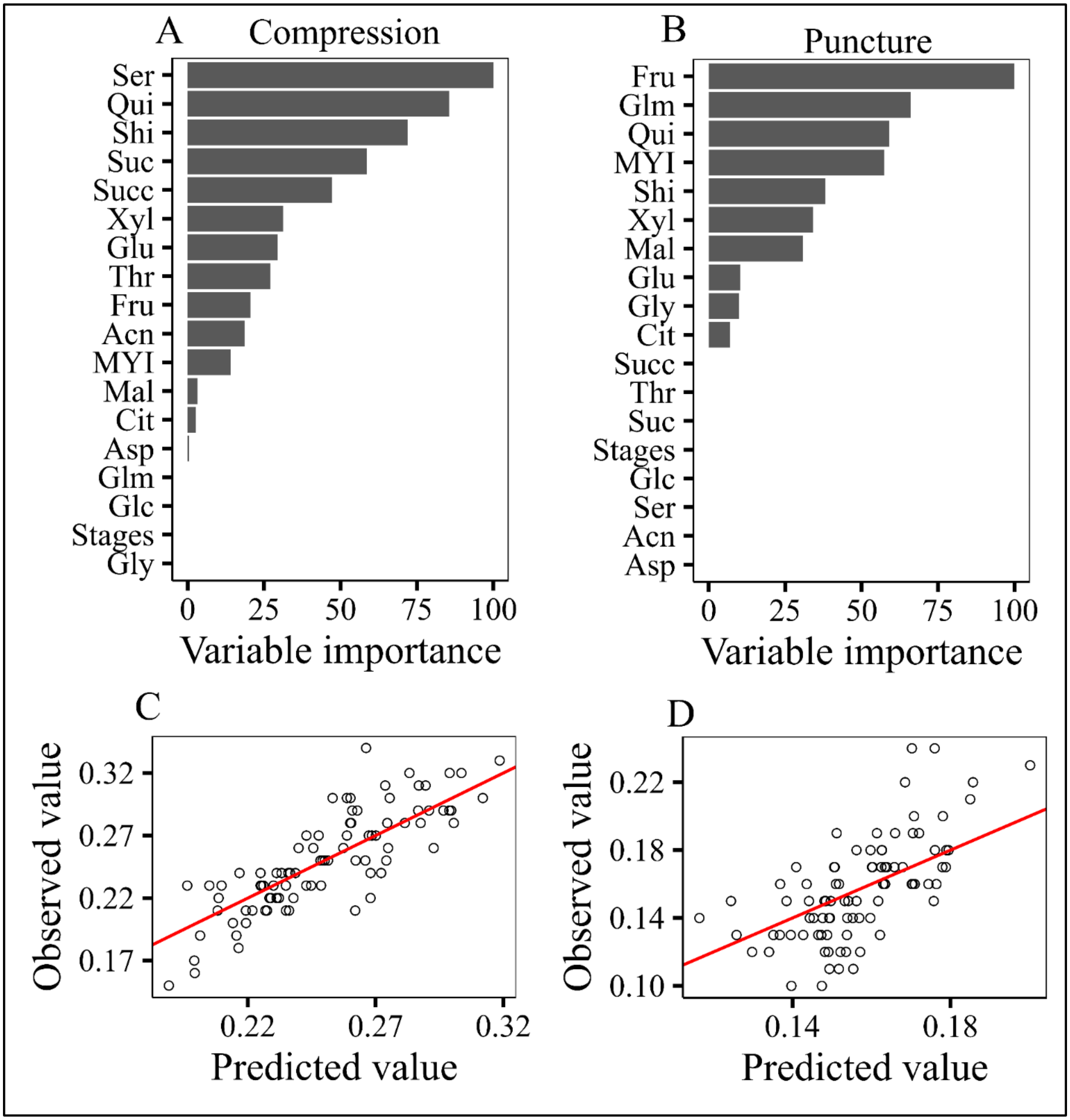

Figure 1: Variable importance (A, B) and Predicted Vs Observed (C, D) plot during the prediction of compression (A, C) and puncture (B, D).

References:

Boyette, M.D., Estes, E., et al. (1993) Postharvest cooling and handling of blueberries. https://content.ces.ncsu.edu/postharvest-cooling-and-handling-of-blueberries

Chapman Jr, G. W., & Horvat, R. J. (1989). Determination of nonvolatile acids and sugars from fruits and sweet potato extracts by capillary GLC and GLC/MS. *Journal of agricultural and food chemistry*, *37*(4), 947-950.

Degu, A., Hochberg, U., et al. (2014) Metabolite and transcript profiling of berry skin during fruit development elucidates differential regulation between Cabernet Sauvignon and Shiraz cultivars at branching points in the polyphenol pathway. *BMC Plant Biol.* 14: 188.

Gómez-Romero, M., Segura-Carretero, A., et al. (2010) Metabolite profiling and quantification of phenolic compounds in methanol extracts of tomato fruit. *Phytochem.* 71: 1848–64.

Li, C., Luo, J., et al. (2011) A novel instrument to delineate varietal and harvest effects on blueberry fruit texture during storage. *J. Sci. Food Agric.* 91: 1653–1658.

Lombardo, V., Osorio, S., et al. (2011) Metabolic profiling during peach fruit development and ripening reveals the metabolic networks that underpin each developmental stage. *Plant Physiol.* 157: 1696–710.

James G, Witten D, Hastie T, Tibshirani R (2013) An introduction to statistical learning, Vol 112. Springer.

Ljubobratović D, Vuković M, Brkić Bakarić M, Jemrić T, Matetić M (2022) Assessment of various machine learning models for peach maturity prediction using non-destructive sensor data. Sensors 22: 5791.

Maclean, D.D., Nesmith, D.S. (2011) Rabbiteye blueberry postharvest fruit and stimulation of ethylene production by 1-methylcyclopropene. *HortSci.* 46: 1278–1281.

Montecchiarini, M. L., Bello, F., Rivadeneira, M. F., Vazquez, D., Podesta, F. E., & Tripodi, K. E. J. (2018). Metabolic and physiologic profile during the fruit ripening of three blueberries highbush (Vaccinium corymbosum) cultivars. *Journal of Berry Research*, *8*(3), 177-192.

Mehra, L.K., MacLean, D.D., et al. (2013) Postharvest disease development on southern highbush blueberry fruit in relation to berry flesh type and harvest method. Plant Dis. 97: 213–221.

Paniagua, A.C., East A.R., et al. (2013) Moisture loss is the major cause of firmness change during postharvest storage of blueberry. Postharvest Biol. Technol. 79: 13-19.

Zhang, J., Wang, X., et al. (2011) Metabolic profiling of strawberry (Fragaria x ananassa Duch.) during fruit development and maturation. *J. Exp. Bot.* 62: 1103–1118.