HortScience 45(4):684–686. 2010.

Combining Linear Regression and Machine Learning Approaches to Identify Consensus Variables Related to Optimum Sweetpotato Transplanting Date

Arthur Villordon[1]
LSU AgCenter Sweet Potato Research Station, 130 Sweet Potato Road, Chase, LA 71324

Christopher Clark
LSU AgCenter Department of Plant Pathology and Crop Physiology, Baton Rouge, LA 70803

Tara Smith
LSU AgCenter Sweet Potato Research Station, 130 Sweet Potato Road, Chase, LA 71324

Don Ferrin
LSU AgCenter Department of Plant Pathology and Crop Physiology, Baton Rouge, LA 70803

Don LaBonte
LSU AgCenter School for Plant, Environmental, and Soil Sciences, Baton Rouge, LA 70803

Additional index words. Ipomoea batatas, growing degree days, adventitious roots, storage root initiation

Abstract. **Forward and stepwise regression methods identified variables related to the influence of transplanting date on yield of U.S. #1 sweetpotatoes. The variables were mean minimum soil temperature 5 days after transplanting (DAT), wind direction at transplanting, and accumulated heat units (growing degree-days) 5 DAT. Machine learning techniques identified the same variables using leave-one-out and k-fold cross-validation methods. Growers and crop consultants, in collaboration with knowledge workers, can use this information in conjunction with public and subscription-based weather forecasts to further optimize transplanting date determination and for making risk-averse decisions. These results help to underscore the importance of consistent transplant establishment as one of the determinants of storage root yield in sweetpotatoes.**

Decision-making in regard to scheduling transplanting date is one of the least studied aspects of sweetpotato production. It is well documented that transplanting dates can potentially influence storage root yield (Edmond and Ammerman, 1971). Recently, we have documented that up to 85% of adventitious roots extant at 5 to 7 d after transplanting (DAT) have the potential to become storage roots (Villordon et al., 2009b). The uniform and consistent initiation of adventitious roots has been shown to be a critical step in the determination of final yield (Kokubu, 1973; Togari, 1950; Villordon et al., 2009b). Thus, it is important to identify agrometeorological and management variables that exert influence on this specific stage (5 to 7 DAT) to optimize decision-making. In Louisiana, a calendar-based system is used for recommending transplanting dates, i.e., 15 Apr. to 30 June for south Louisiana and 1 May to 30 June for north Louisiana (Boudreaux, 2005). In North Carolina, the recommendations include a provision for soil temperature to be at least 18 °C at a depth of 10 cm for 4 consecutive days before transplanting (North Carolina Sweet Potato Commission, 2009). Some commercial growers in Louisiana are known to temporarily stop transplanting operations if there is a prevailing "northeast wind," which is typically a cold, dry wind that predisposes transplants to desiccation (Cannon, personal communication). Precise information on the relative importance and interactions of agrometeorological variables on transplant establishment and subsequent storage root yield will potentially benefit researchers, growers, and crop consultants.

The objective of this study was to identify consensus variables at transplant time that were related to U.S. #1 yield outcome using least squares-based linear regression and machine learning approaches. Machine learning generally refers to the class of computational methods for deriving insightful knowledge (including heuristics, strategies, or structure) from data, observations, or past solutions (Shaw, 1993). Models derived from machine learning approaches are also referred to as adaptive models and are characterized by learning by example to solve problems. Adaptive modeling techniques are increasingly being used in areas where there is little or incomplete understanding of the problem to be solved but where training data are available (Park et al., 2005).

Materials and Methods

Modeling data set. A recently concluded multistate research project has generated a database of storage root yield, soil analysis, chemical use, and agrometeorological data. A portion of this database was recently used in the development of a growing degree-day model (Villordon et al., 2009a). For this study, we used a subset of data consisting of yield data from replicated plots established on farms in Louisiana from 2004 to 2006 (n = 63). This data set was merged with location-specific agrometeorological data obtained from the Louisiana Agriclimatic Information Network (Louisiana Agriclimatic Information, 2009) and will be referred to as the modeling data set (MDS). The following stations were used: R & D Research Farm (Port Barre; long. 30°39′ N, lat. 91°59′ W), Sweet Potato Research Station (Chase; long. 32°5′ N, lat. 91°42′ W), and University of Louisiana at Monroe (Monroe; long. 32°30′ N, lat. 92°7′ W). Daily agrometeorological variables included maximum and minimum air temperature, maximum and minimum soil temperature (10.2-cm depth), solar radiation, relative humidity, wind (direction and speed), and rainfall. Means (total for rainfall) for all variables were calculated for 5 d before and after transplanting (DAT). Accumulated heat units, expressed as growing degree-days (base = 15.5 °C, ceiling = 32.2 °C), were used to adjust for differences in the length of growing periods (Riha et al., 1996). The model for calculating growing degree-days in Louisiana was previously described (Villordon et al., 2009a). A separate growing degree-day model was derived for transplants 5 DAT

Received for publication 16 Oct. 2009. Accepted for publication 26 Jan. 2010.
Portions of this paper were supported by USDA, CSREES, RAMP Grant Award No. 370831201-02106 "Development of Grower Decision-making Tools to Reduce Risk and Enhance Sustainability of Southern Sweetpotato Pest Management Systems."
Approved for publication by the director of the Louisiana Agricultural Experiment Station as manuscript No. 2009-260-3842.
Mention of trademark, proprietary product or method, and vendor does not imply endorsement by the Louisiana State University AgCenter nor its approval to the exclusion of other suitable products or vendors.
[1]To whom reprint requests should be addressed; e-mail avillordon@agcenter.lsu.edu.

(base = 18.3 °C, ceiling = 32.2 °C). This was derived by calculating growing degree-days using various base and ceiling temperatures after 5 d and finding the best linear fit to U.S. #1 yield ($t \cdot ha^{-1}$). A similar approach was used to estimate soil heat units. Representations of air and soil temperature as air and soil heat units, respectively, have been routinely used in crop modeling (Pale et al., 2003; Riha et al., 1996). Descriptive statistics of agrometeorological variables used in subsequent modeling experiments are shown in Table 1.

Statistical and machine learning analyses. The MDS was subjected to normality analysis and collinearity diagnostics using SPSS Version 15 (SPSS, Chicago, IL). Outliers were removed after residual analysis (within ± 2 SDs) and Cook's distance (>1). After removing outliers, this modeling data set (n = 60) was used for least squares-based linear regression analysis (forward and stepwise selection; SPSS Version 15). The predictor variables included wind (speed and direction), air temperature (growing degree-days), soil temperature (maximum and minimum, accumulated heat units), and relative humidity. The MDS was subjected to machine learning experiments using Waikato Environment for Knowledge Analysis (WEKA, Version 3.5.8; The University of Waikato, 2009). WEKA is free software available under the GNU General Public License. WEKA was run in the "Experimenter" mode that enabled simultaneous comparisons of models trained and validated on the data sets. The following algorithms were used: linear regression (attribute selection method = Greedy method, value of Ridge parameter = 1.0E-8), multilayer perceptron (backpropagation neural network classifier), Gaussian radial basis function network, support vector machine, various decision tree procedures (decision stump, M5P, and REPtree), and metaclassifiers. The metaclassifiers included VOTE (class for combining models), BAGGING (class for bagging a model to reduce variance), and ENSEMBLE (combines several models using the ensemble selection method). Except where otherwise indicated, default parameter settings were used and model performance was estimated using the following cross-validation methods: leave-one-out and k-fold (k = 10, repetitions = 10). Model performance was measured using relative mean square error (RMSE).

Table 1. Descriptive statistics of some agrometeorological variables 5 d after transplanting sweetpotatoes in Louisiana.

Variable	Mean	Minimum	Maximum	SD
MXAIR	32.0	27.6	35.4	1.7
MINAIR	20.6	13.9	23.9	1.9
MXSOIL	34.3	27.1	42.6	3.5
MINSOIL	25.3	20.4	29.5	2.2
RH	74.4	37.5	100	24.9
WINDDIR	168.91	5	359	82.0
WINDSPEED	6.4	1.8	21.1	3.4

Agrometeorological data (2004–2006) were obtained from the Louisiana Agriclimatic Information Network (LAIS) stations in Chase (long. 32°5′ N, lat. 91°42′ W), Monroe (long. 32°30′ N, lat. 92°7′ W), and Port Barre (long. 30°39′ N, lat. 91°59′ W) (Louisiana Agriclimatic Information, 2009).
MXAIR = maximum air temperature (temp; °C); MINAIR = minimum air temp (°C); MXSOIL = maximum soil temp (°C); MINSOIL = minimum soil temp (°C); RH = relative humidity; WINDDIR = wind direction (0 = north, 90 = west, 180 = south, 270 = east); WINDSPEED = $km \cdot hr^{-1}$. Soil temperature measured at 10.2-cm depth.

Results

The least squares-based linear regression method identified the following U.S. #1 ($t \cdot ha^{-1}$) predictor variables: mean minimum soil temperature 5 DAT (MINSOIL5DAT), accumulated heat units of transplants 5 DAT (GDD5DAT), and wind direction at transplanting (WDD0). Growing degree-days (GDD) was used to adjust for differences in the length of the growing period. The predictive equation was: U.S. #1 yield = 0.25*GDD5DAT – 0.16*MINSOIL5DAT + 0.01*GDD + 0.016*WDD0–16.2 (adjusted $R^2 = 0.32$). In general, the predictive performance estimates (RMSE) among machine learning models were comparable, except for multilayer perceptron, which had a significantly higher error rate compared with the baseline model (machine learning linear regression) (Table 2). The machine learning linear regression model was: U.S. #1 yield = 4.33 + 0.03 * WDD0–8.0 * WSD1–0.01 * WDD4 + 0.008 * GDD + 0.18 * GDD5DAT – 0.25 * MINSOIL5DAT in which WSD1 = wind speed 1 d after transplanting. Based on the least squares-based linear regression and the machine learning linear regression results, the consensus variables related to U.S. #1 yield were MINSOIL5DAT, GDD5DAT, and WDD0.

Table 2. Predictive performance of machine learning models that represent relationship of some agrometeorological variables and sweetpotato U.S. #1 yield in Louisiana.

	Cross-validation	
	Leave-one-out	k-fold
LR	4.14 (2.48)	4.67 (1.19)
REPTREE	3.67 (2.45)	4.60 (1.19)
SVM	3.85 (2.91)	5.10 (1.24)
RBF	3.88 (2.57)	4.57 (0.99)
VOTE	4.02 (2.51)	4.53 (1.18)
BAGGING	4.08 (2.61)	4.78 (1.17)
M5P	4.14 (2.70)	4.25 (1.22)
ENSEM	4.24 (3.08)	4.98 (1.55)
DECSTUMP	4.67 (2.49)	4.80 (1.05)
MLP	6.42 (5.65)	7.48* (3.13)

Agrometeorological data represented the period 5 d after transplanting. Values represent root mean square error. Values in parentheses are SDs. For k-fold cross-validation, k = 10, number of repetitions = 10; two-tailed pairwise *t* tests (confidence = 0.05) were performed using machine learning linear regression as the baseline model. *Significant differences from baseline model. Analysis was performed using the Waikato Environment for Knowledge Analysis (WEKA, Version 3.5.8; The University of Waikato, 2009). The following algorithms were used: linear regression (LR; attribute selection method = Greedy method, value of Ridge parameter = 1.0E-8), multilayer perceptron (MLP; backpropagation neural network classifier), Gaussian radial basis function network (RBF), support vector machine (SVM), and various decision tree procedures (DECSTUMP = decision stump, M5P, and REPTREE). The metaclassifiers included VOTE (class for combining models), BAGGING (BAG-LR; class for bagging a model to reduce variance), and ENSEMBLE (ENSEM; combines several model using the ensemble selection method). Descriptions of each algorithm are available in the WEKA manual (The University of Waikato, 2009).

Discussion

Based on these results, planning or decisions related to sweetpotato transplanting date should not be based solely on a range of calendar dates. Rather, such decisions should also consider the prevailing agrometeorological conditions. Such conditions can vary widely during the transplanting period in Louisiana (Table 1). A strict calendar-based system potentially ignores agrometeorological factors that can negatively influence transplant establishment and lead to inconsistent yields. Our results in part showed the importance of considering air and soil temperatures in determining optimum transplant conditions. Air temperature, expressed as accumulated heat units, was positively related to U.S. #1 yield based on the least squares-based linear regression model. This is consistent with cumulative research that showed air temperatures below 15 °C suppressed root development (Ravi and Indira, 2009). Low air temperatures (below 15 °C) have been recorded in the first 7 d in May in each of 2004 and 2005 for the Chase location (northeast Louisiana) (Louisiana Agriclimatic Information, 2009). The recommended sweetpotato transplanting dates for this region range from 1 May to 30 June (Boudreaux, 2005). On the other hand, the least squares-based linear regression model showed that minimum soil temperature (5 DAT) was negatively related to U.S. #1 yield. This observation is consistent with cumulative data indicating that soil temperatures between 20 and 30 °C favor storage root formation and growth with night temperatures (minimum) being the most critical (Ravi and Indira, 2009). Eguchi et al. (1994) have previously documented that storage root dry weight was greatest when root temperatures ranged from 24 to 26 °C; storage root dry weight decreased ≈25% and 50% at 28 and 30 °C, respectively. The observed minimum temperatures in the modeling data set varied from 20 to 29 °C (Table 1), underscoring the variability of temperature-related variables during sweetpotato transplanting in Louisiana. Our results also provide preliminary empirical evidence of the validity of growers' practice to temporarily stop transplant operations in recognition of

wind-related variables. Growers have reported inconsistent yields associated with desiccated transplants brought about by "cold air blowing in from the northeast" (Cannon, personal communication). This is consistent with the recognition of wind-related variables in transplanting and establishment of vegetatively propagated crops. For example, Golden et al. (2003) mentioned that wind speeds greater than 16 km·h^{-1} accelerated leaf drying and hindered establishment in bare root strawberry transplants. In newly transplanted bare-root strawberry transplants, extensive overhead irrigation is practiced to prevent desiccation and death of transplants during the plant establishment period (Duval, 2002). Although there are no published reports that specifically examined the relationship between wind direction and speed on sweetpotato transplant survivability, we hypothesize that similar wind speeds and cold air (less than 15 °C) can lead to transplant desiccation that likely interferes with establishment. In this context, we define sweetpotato transplant establishment as the initiation of adventitious roots that can start as early as 3 DAT under field conditions (Villordon et al., 2009c). This period corresponds to Togari's (1950) Period I in the early "tuberous-root thickening stage," defined as the appearance of protoxylem, a prerequisite stage to cambium development and subsequent storage root initiation. More recently, we have confirmed this phenological stage in 'Beauregard' (Villordon et al., 2009a, 2009c). Thus, the presumptive role of agrometeorological conditions 5 DAT appears related to transplant establishment and its direct relationship to the determination of potential yield in sweetpotatoes.

The availability of 10-d weather forecasts and other agrometeorological-related services allows growers, consultants, and knowledge workers to further optimize transplanting date decisions based on predicted U.S. #1 yield. For example, if a grower considers 11 t·ha^{-1} U.S. #1 yield as the breakeven point, then transplanting can be delayed until conditions are more conducive for higher yields. When properly stored, transplants can be held for as long as 5 d without affecting potential storage root yield (Hammett, 1985; Nakatani, 1987).

As a result of limitations of the data set, other variables such as soil moisture were not included in the analysis. This represents a significant limitation of the current study and likely contributed to the relative low coefficient of determination for the least squares-based linear regression model. This trend is also apparent in the relatively high SDS associated with the RMSE estimates of the machine learning models, especially those associated with leave-one-out cross-validation (Table 1). Although leave-one-out cross-validation had the advantage of using as much data as possible for training, the variance of the estimated performance measurement was high because each test set contained only one record (Tan et al., 2005). Follow-up studies that quantify or document the mitigating influence of soil moisture on air and soil heat accumulation will likely improve the predictive abilities of models that focus on this growth period.

Conclusion

Statistical and machine learning approaches identified consensus variables 5 DAT that were related to marketable yield. This information can be used in conjunction with public or fee-based weather services that can provide agrometeorological forecasts of up to 10 d. This allows growers, working independently or with consultants, to further optimize decision-making related to transplanting. One of the important limitations of the current model is the lack of validation with soil moisture levels and other variables related to transplant establishment and early development. Follow-up studies that include soil moisture will likely enhance the accuracy of predictive models and further improve decision-making in sweetpotato transplanting.

Literature Cited

Boudreaux, J. 2005. Commercial vegetable production recommendations. Pub. 2433. LSU AgCenter, Baton Rouge, LA.

Duval, J.R. 2002. Use of prohexadione-CA to increase early yield and reduce establishment irrigation of strawberry (*Fragaria ×ananassa*). Proc. Fla. Hort. Soc. 115:220–222.

Edmond, J.B. and G.R. Ammerman. 1971. Sweetpotatoes: Production, processing, marketing. AVI Publ. Co., Westport, CT.

Eguchi, T., M. Kitano, and H. Eguchi. 1994. Effect of root temperature on sink strength of tuberous root in sweetpotato plants (*Ipomoea batatas* Lam.). Biotronics 23:75–80.

Golden, E.A., J.R. Duval, E.E. Albregts, and C.M. Howard. 2003. Intermittent sprinkler irrigation for establishment of bare root strawberry transplants. Document HS947 Florida Cooperative Extension Service, Institute of Food and Agricultural Sciences, Univ., Florida.

Hammett, L.K. 1985. Refrigerated storage influence on sweet potato transplant viability and root yield. HortScience 20: 198–200.

Kokubu, T. 1973. Thremmatological studies on the relationship between the structure of tuberous root and its starch accumulating function in sweet potato varieties. Bull. Fac. Agr. Kagoshima Univ. 22:1–126.

Louisiana Agriclimatic Information. 2009. Louisiana agriclimatic information. 1 Mar. 2009. <http://www.lsuagcenter.com/weather/index.asp>.

Nakatani, M. 1987. Holding of cut-sprouts in sweet potato (*Ipomoea batatas* Lam.). Jpn. J. Crop. Sci. 56:238–243.

North Carolina Sweet Potato Commission. 2009. How to grow commercial sweetpotatoes. 1 Mar. 2009. <http://www.ncsweetpotatoes.com/content/view/53/78/>.

Pale, S., S.C. Mason, and T.D. Galusha. 2003. Planting time for early-season pearl millet and grain sorghum in Nebraska. Agron. J. 95:1047–1053.

Park, S.J., C.S. Hwang, and P.L.G. Vlek. 2005. Comparison of adaptive techniques to predict crop yield response under varying soil and land management conditions. Agr. Syst. 85:59–81.

Ravi, V. and P. Indira. 2009. Crop physiology of sweet potato. Hort. Rev. (Amer. Soc. Hort. Sci.) 23:277–338.

Riha, S.J., D.S. Wilks, and P. Simoens. 1996. Impact of temperature and precipitation variability on crop model predictions. Clim. Change 32:293–311.

Shaw, M.J. 1993. Machine learning methods for intelligent decision support: An introduction. Decis. Support Syst. 10:79–83.

Tan, P.N., M. Steinback, and V. Kumar. 2005. Introduction to data mining. Addison Wesley, Boston, MA.

The University of Waikato. 2009. Waikato Environment for Knowledge Analysis (WEKA). Version 3.5.8. The University of Waikato, Hamilton, New Zealand.

Togari, Y. 1950. A study of tuberous root formation in sweet potato. Bul. Nat. Agr. Expt. Sta. Tokyo 68:1–96.

Villordon, A., C.A. Clark, D. Ferrin, and D. LaBonte. 2009a. Using growing degree days, agrometeorological variables, linear regression, and data mining methods to help improve prediction of sweetpotato harvest date in Louisiana. HortTechnology 19:133–144.

Villordon, A., D. LaBonte, N. Firon, Y. Kfir, E. Pressmn, and A. Schwartz. 2009b. Characterization of adventitious root development in sweetpotato. HortScience 44:651–655.

Villordon, A., D.R. LaBonte, and N. Firon. 2009c. Development of a simple thermal time method for describing the onset of morpho-anatomical features related to sweetpotato storage root formation. Sci. Hort. 121:374–377.