Pattern Recognition and Estimating Cultivar Coefficients of a Crop Simulation Model: Old Approach, New Application

Mohammad Bannayan and Gerrit Hoogenboom

Department of Biological and Agricultural Engineering, The University of Georgia, Griffin, GA 30223, USA
Bannayan@uga.edu

Introduction

• The introduction of a new cultivar in a process-based crop simulation model requires the estimation of cultivar coefficients that define its growth and development characteristics. An accurate estimation of these coefficients requires replicated field experiments that, in many cases, are not available to crop model users.

• The objective of this study was to employ a pattern recognition approach to estimate cultivar coefficients from a minimum set of experimental data for use with a crop simulation model (Fig. 1). The pattern recognition approach is based on similarity measures. Its main goal is to classify groups of data or patterns based on either a priori knowledge or on statistical information extracted from the patterns.

Materials & Methods

• The approach employed in this study was based on a few features of maize crop growth and development, including anthesis and harvest maturity dates, maximum leaf area index (LAImax), final above ground biomass, and grain yield, and these characteristics were used as the features vector. To construct the feature database, 18540 hypothetical cultivars were constructed by combining different values of the six cultivar coefficients of the Cropping System Model (CSM)-CERES-Maize model (Hoogenboom et al., 1994; Jones et al., 2003). Patterns of observed crop data were matched with the feature database to estimate the most similar or "*best match*".

• Euclidean distance (1), d_j and Cosine relation (2) were used as the measures of similarity.

$$d_j = \sqrt{\left[\sum_{j=1}^{d} W_j (V_{ij} - V_{mj})^2\right]} \quad (1)$$

$$-\cos(\angle(x,y)) = -\frac{\sum_{i=1}^{5} x_i y_i}{\sqrt{\sum_{i=1}^{5} x_i^2 \sum_{i=1}^{5} y_i^2}} \quad (2)$$

where d_j, is Euclidean distances, V_j is the jth component of vectors of observed and feature and W_j are weights. The cosine function calculates similarity of two vectors of data (x and y), containing the crop characters, thus as the two data segments become more similar, their cosine similarity approaches 1.0 and their distances approaches 0.0. In this equation, i is the number of crop characters which in our case $i = 5$. Therefore the cosine measure was used as a distance:

D (x, y) = 1- (cosine function)

• The coefficients of *best match* cultivar will be used as the new cultivar coefficients.

$$d = 1 - \frac{\sum_{i=1}^{n} (p_i - o_i)^2}{\sum_{i=1}^{n} \left(|p_i - \bar{o}| + |o_i - \bar{o}|\right)^2}, \quad 0 \le d \le 1 \quad (3)$$

• The index of agreement, d (3) (Fig. 3) was used to evaluate the accuracy of the estimations. In equation 3, O_i and P_i are as estimated and observed values and O-bar is the average of the observed data.

Results

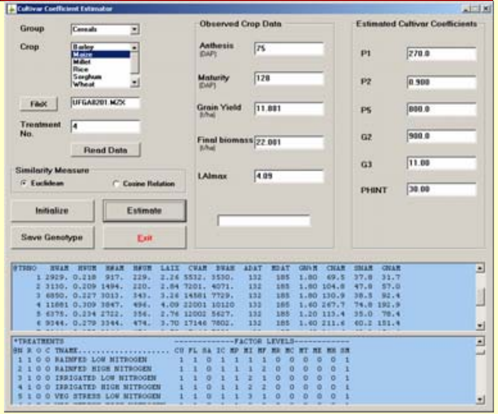
• To evaluate our approach, we used 29 different maize cultivars as reported from field experiments that were available within the DSSAT cultivar database for maize. The CSM-CERES-Maize model simulation of growth and development of the 29 maize cultivars for three different sites using the simulation results based on the *best match* coefficients and based on the original reported coefficients are shown in Figure 2.

• In order to determine if further adjustment of the cultivar coefficients obtained with our approach are needed , we evaluated the performance of three cultivars of which two of them were not used in development and evaluation of this tool. The time series comparison of the original and *best match* aboveground biomass, grain yield and leaf area index (Fig. 3) showed a good similarity for all three variables.

Conclusion

• The reasonable agreement between crop growth and development simulated based on the reported original cultivar coefficients and based on the *best match* coefficients, showed the reliability of pattern recognition approach for determining the coefficients for any maize cultivar with the minimum available crop data.

Fig. 1. Interface of the Cultivar Coefficient Estimator



Acknowledgement

This work was conducted under the auspices of the Southeast Climate Consortium (SECC; secc.coaps.fsu.edu) and supported by a partnership with the United States Department of Agriculture-Risk Management Agency (USDA-RMA), by grants from the US National Oceanic and Atmospheric Administration-Office of Global Programs (NOAA-OGP) and USDA-Cooperative State Research, Education and Extension Services (USDA-CSREES) and by State and Federal funds allocated to Georgia Agricultural Experiment Station Hatch project GEO01654.

References

Hoogenboom, G., Jones, J.W., Wilkens, P.W., Porter, C.H., Batchelor, W.D., Hunt, L.A., Boote, K.J., Singh, U., Uryasev, O., Bowen, W.T., Gijsman, A.J., du Toit, A.S., White, J.W., and Tsuji, G.Y., 2004. Decision Support System for Agrotechnology Transfer Version 4.0 [CD-ROM]. University of Hawaii, Honolulu, Hawaii.

Jones, J.W., G. Hoogenboom, C.H. Porter, K.J. Boote, W.D. Batchelor, L.A. Hunt, P.W. Wilkens, U. Singh, A.J. Gijsman, and J.T. Ritchie. 2003. DSSAT cropping system model. Eur. J. Agron. 18: 235–265.

Fig. 2. Comparison of anthesis, maturity, grain yield, final biomass and LAImax of 29 different cultivars for Gainesville, Florida, Ames, Iowa, and Spain. (DAP: days after planting)

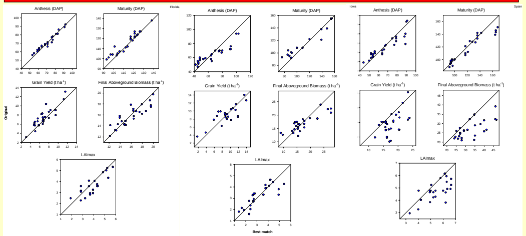


Fig. 3. Comparison of observed leaf area index (LAI), aboveground biomass, and grain weight of three different cultivars against simulated using original and *best match* cultivar coefficients.