Prediction of cotton yield in Kenya

Josphat Igadwa Mwasiagi[*‡], **Xiu Bao Huang**[*] and **Xin Hou Wang**[*]

COTTON YIELD IS ONE OF THE INDICATORS for describing agricultural efficiency from different resource management methods in the cotton-growing industry. Selected cotton-growing cost factors were used to design an artificial neural network model to predict cotton yield in Kenya. This neural network model was able to predict cotton yield with a satisfactory performance error of 0.204 kg/ha and a regression correlation coefficient between network output and actual yield of 0.945.

Introduction

Cotton can be grown in all the eight provinces of Kenya and especially in the semi-arid regions where few other commercial crops are viable. The Kenyan cotton-processing industry has an installed capacity of over 120 000 bales of cotton lint per year, which is able to meet half of the national demand for cotton products. Kenya has consistently produced less than 30 000 bales of cotton lint annually since 1990, so the country is a net importer of the commodity, causing adverse strain on its foreign exchange reserves.

Crop yield is one of the indicators for describing agricultural efficiency arising from different resource management methods. The need to monitor cotton yield in Kenya is essential, given that cotton-growing and yields have undergone an unprecedented decline in the last ten years. Cotton yield forecasts, a few weeks before harvest time, can be of strategic importance for advance trade planning, in order to pre-empt national cotton lint demand, and rationalize cotton lint ginning and marketing. An effective cotton yield model can also be useful for analysis of factors that affect cotton agriculture in a region. The main factors that influence cotton yield are government policies, crop husbandry methods and cotton price.[1–3] Crop husbandry methods such as land preparation, the type and quantity of fertilizers used, and weed and pest control methods, have been reported to have a primary influence on cotton quality and yield.[4–6]

Cotton yield can be predicted using satellite forecasting.[7,8] This method makes use of the latest technologies in surveillance and data collection, and is set to replace traditional means of crop prediction. While prediction of cotton yield using satellite forecasting is gaining popularity, especially in the main cotton-producing countries, there are other, less-affluent countries where the cotton-growing industry struggles to survive, and a cheaper yet effective forecasting method is still needed. Such a method will serve as an interim measure for a struggling cotton-growing industry to rationalize itself and achieve a requisite level of competitiveness.

Kenya, being mainly an agricultural economy, has a well-established ministry of agriculture, spreading its operations over the entire country. The district agricultural officers (DAOs), at district level, are responsible for collection of specific data pertaining to all agricultural crops grown in their area. These data are archived at district, provincial and national levels. We have undertaken a study of the data for cotton growing from 1996 to 2003. The cost factors that influence cotton growth, together with corresponding yields have been used to predict cotton yield by means of an artificial neural network (ANN) model.

Materials and methods

Cotton-growing data were accessed from district agricultural officers in all the cotton-growing regions in Kenya. Data volumes from a given cotton region correlated with the amount of cotton grown in that region. This ensured that the data in hand were statistically representative of the local industry. There are three cotton-growing regions in the country: eastern/central (e/c), coastal (c) and western (w). Their respective percentages of cotton lint production for the eight years from 1996 to 2003 were 54%, 15% and 31%. Corresponding records of data collected from the eastern/central, coastal and western regions were 39, 11 and 22, respectively. The DAOs normally collect cost factor data, crop yields and prices that were paid for all the crops grown in a given district. We identified a total of ten cost factors within the archived cotton-growing data. These cost factors are ploughing (plo), harrowing (hr), seed cost (sd), fertilizer cost (ft), planting (pla), weeding (we), pesticide cost (pes), harvesting (ha), transportation to the ginnery (tr), as well as other miscellaneous expenses (ot). These factors, together with the related crop yields, form the main core of our data, from which we correlated production costs with crop yields using our ANN yield prediction model. The aim in designing the model was to create a predictive tool for cotton yield before harvest-time. Seven cost factors—ploughing, harrowing, seeds, fertilizers, planting, weeding and pesticides—were selected, being data that a DAO collects from the farmers prior to harvest. The input and output vectors were 72 × 7 and 72 × 1, respectively. A back-propagation feed-forward ANN, with one hidden layer was used, and the number of neurons in the hidden layer varied from 2 to 20 in steps of two. The performance of the ANN was assessed using root mean squared error (RMSE),[9] defined in Equation (1), where N is the number of input–output pairs, y is the cotton yield and $\hat{y}$ is the ANN output (predicted cotton yield). RMSE has the same units as the factor being predicted.

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum^{N}(y_i - \hat{y}_i)} \,. \qquad (1)$$

*Department of Textile Engineering, Donghua University, Shanghai, China.
‡Author for correspondence. Address for correspondence: Department of Textile Engineering, Moi University, P.O. Box 3908, Eldoret, Kenya. E-mail: igadwa@yahoo.com

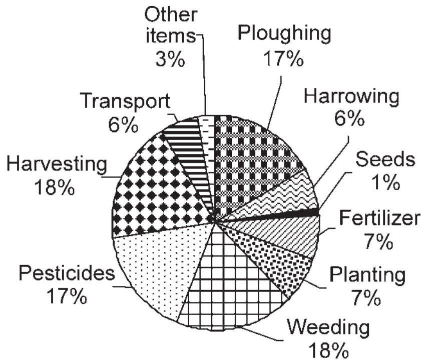

Fig. 1. Contribution of cost factors to the cost of growing cotton lint in Kenya.

Whereas RMSE can be used to measure the performance of an ANN, it is also useful to investigate the network response in more detail.[10] One option is to perform a regression analysis between the network response ($\hat{y}$) and the corresponding targets (y). In this investigation we have performed a regression between y and ($\hat{y}$) and report the correlation coefficient (R-value) between the outputs and targets.

Results and discussion

Cost factors. The contribution of the cost factors to the cost of growing cotton lint in Kenya during the study period is given in Fig. 1. The four biggest cost items are weeding (we), harvesting (ha), ploughing (plo) and pesticides (pes). Ploughing and harrowing are land preparation, their combined contribution being 23%, making land preparation the largest cost component in the Kenyan cotton-growing industry.

The cost factors for the three cotton-growing regions in Kenya are given in Fig. 2. From the yields, illustrated in Fig. 3, it can be inferred that the eastern/central region is comparatively better at optimizing cost factors to generate higher yields. The western region has low costs, but with low cotton yield, whereas the coastal region features as a highly variable average performer, when compared with the two other regions.

Prediction of cotton yield. Table 1 gives the performance of the cotton yield prediction ANN model. The optimum ANN has 12 neurons and can predict cotton yield with a favourable error (RMSE) of 0.204 kg/ha. The regression correlation coefficient between the predicted and actual yield is 0.945. This deems the model to be a reliable tool for the purposes that have been discussed.

Table 1. Performance of cotton yield prediction algorithm.

No. of neurons	RMSE	R-value
2	0.437	0.859
4	0.368	0.884
6	0.327	0.899
8	0.321	0.905
10	0.286	0.917
12	0.204	0.945
14	0.188	0.908
16	0.199	0.886
18	0.117	0.878
20	0.118	0.868

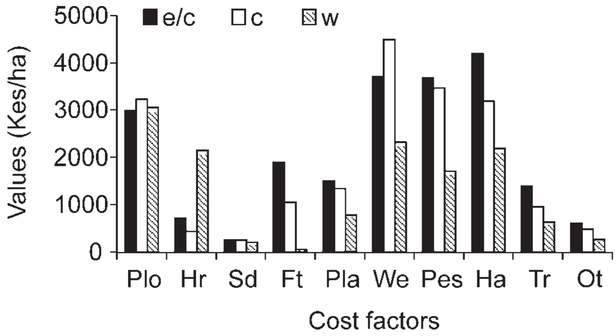

Fig. 2. Cost factors for cotton-growing in Kenya (1US$ = 70 Kes). Key: Plo, ploughing; Hr, harrowing; Sd, cost of seed; Ft, cost of fertilizer; Pla, planting; We, weeding; Pes, pesticide cost; Ha, harvesting; Tr, transportation; Ot, other.

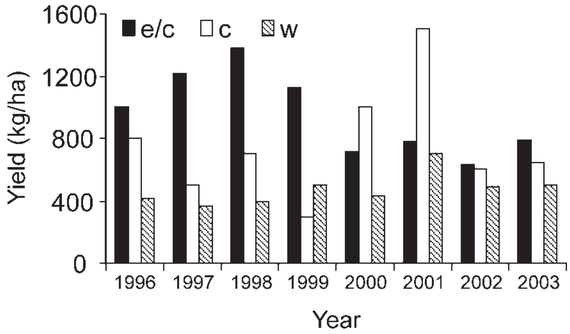

Fig. 3. Cotton yield in Kenya between 1996 and 2003, according to the three main growing regions: e/c, eastern/central; c, coastal; and w, western.

We acknowledge financial assistance from the University Research Fund of Moi University, Eldoret, Kenya, and the China Scholarship Council, Beijing.

1. Bradow J.M., Bauer P.J., Hinojosa O. and Sassenrath-Cole G. (1997). Quantitation of cotton fibre-quality variations arising from boll and plant growth environments. *Eur. J. Agron.* **6**, 191–204.
2. Gillen N.P. (2004). The responsibility of governments for the enforcement of arbitral awards; Cotton: Review of the world situation. *ICAC* **58**(2), 14–15.
3. Rico M.R. (2002). Impact of low prices on production practices and fiber quality – A farmer's perspective. In *Proc. 61st ICAC,* Cairo, pp. 3–7.
4. Read J.J, Reddy K.R. and Jenkins J.N. (2006). Yield and fiber quality of upland cotton as influenced by nitrogen and potassium nutrition. *Eur. J. Agron.* **24**, 282–290.
5. Dadari S.A. and Kuchinda N.C. (2004). Evaluation of some pre- and post-emergence weed control measures on rain-fed cotton (*Gossypium hirsutum* L) in Nigerian savannah. *Crop Protection* **23**, 457–461.
6. Silvie P., Deguine J.P, Nibouche S. Michel B. and Vaissayre M. (2001). Potential of threshold-based interventions for cotton pest control by small farmers in West Africa. *Crop Protection* **20**, 297–301.
7. Bastiaanssen W.G.M. and Ali S. (2003). A new crop yield forecasting model based on satellite measurements applied across the Indus Basin. *Agric. Ecosyst. Environ.* **94**, 321–340.
8. Dalezios N.R., Domenikiotis C., Loukas A., Tzortzios S.T. and Kalaitzidis C. (2001). Cotton yield estimation based on NOAA/AVHkR produced NDVI. *Phys. Chem. Earth* (*B*) **26**(3), 247–251.
9. Ham F.M. and Kostanic I. (2003). In *Principles of Neurocomputing for Science & Engineering,* pp. 132–136. China Machine Press, Beijing.
10. Demuth H., Beale M. and Hagan M. (2005). *Neural Network Toolbox User Guide,* version 4, chap. 5, pp. 59–63. The MathWorks Inc., Natick, MA.