Jour. Ind. Soc. Ag. Statistics
Vol. XXXVII, No. 2 (1985), pp. 184-191

SIMPLE MODELS FOR PREDICTING SORGHUM GRAIN YIELD USING ENVIRONMENTAL FACTORS

A. K. S. HUDA, S. M. VIRMANI and J. G. SEKARAN
ICRISAT, Patancheru, A. P., 502324

(Received: December, 1983)

SUMMARY

Data on sorghum growth and yield, soil water at critical growth stages and daily climatic variables were collected from 9 locations in India during 1979-82. Regression models that included one or more of the indepnedent variables namely soil water at planting (*SW*), rainfall,mean temperature, solar radiation, evapotranspiration (*ET*) for the whole growing season and for three growth stages were developed from 48 data sets. Stepwise regression technique and Mallow's *Cp* criterion were also utilized to develop models. Results showed that rainfall, mean temperature and their product for three growth stages together explained 68% yield variation. *SW*, rainfall×mean temperature in *GS*2, rainfall × *ET* in *GS*2 and *GS*3, and rainfall × solar radiation in *GS*2 explained 73% yield variation. These two models when tested with 11 independent data sets, the former model explained only 36% yield variation while the later explained 59% yield variation.

Keywords: *Sorghum bicolor* (*L.*) Moench; Regression model; crop-simulation-model; Collaborative multilocation experiments; Phenology.

Introduction

Regression-type crop yield prediction models have been utilized by many researchers to predict crop yield using environmental factors. Gangopadhyaya and Sarkar [6] applied the curvilinear correlation technique to study the effect of weather factors such as rainfall, maximum and minimum temperature on thegrowth and yield of sugarcane. Multiple regression analyses were used by Thompson [12] to determine the influence of selected weather factors in the production of maize in the USA. Brown and Vanderlip [1] utilized stepwise multiple regression models to investigate the relationships between winter wheat yields and monthly weather variables at several locations in Kansas State, USA. Using Fisher's [5] orthogonal polynominal technique, Huda *et al.* [8] studied the relationship between weekly weather variables such as rainfall, humidity, temperature, and rice production in the Tarai regions of India. All of these researchers agreed that weather data when included for specific growth periods are better than monthly data to explain variations in crop yield.

The location-specificity problem of regression type yield prediction models is well recognized. Problems associated with the use of climate-crop-yield models in an operational system were described by LeDuc *et al.* [9]. Crop-simulation models with sound physical and physiological bases are better research tools to quantify the effects of environmental factors on crop growth and development. However, the relatively simpler data requirements of the regression type model make it simple to use for large-scale yield prediction. To predict wheat grain yields on a large scale without the benefit of direct measurements of plant characteristics, mathematical models were developed by Feyerherm and Paulsen [4]. These models require only a modest amount of historical data for application on a real-time basis to geographical regions other than the one where the models were developed. In this paper, an attempt has been made to develop regression-type yield prediction models using different environmental variables collected from collaborative multilocation sorghum modeling experiments coordinated by the authors. These experiments were aimed at developing a better understanding of the physical and physiological processes involved in sorghum production.

Data and Models

Data for this study were obtained from collaborative sorghum modeling experiments conducted under adequate management practice at nine locations in India (11-31°N) during 1979-82 to quantify the effects of environmental factors on sorghum growth and development. The locations were Ludhiana, Hissar, Delhi, Parbhani, Rahuri, Pune, Solapur, Patancheru and Coimbatore. Data on soil, crop, weather and management factors were collected from the experiments from planting to maturity. Data included daily rainfall, solar radiation, open-pan evaporation, maximum and minimum temperature, soil water at planting and at other critical growth stages (panicle initiation,anthesis and physiological maturity), leaf area, phenological data such as dates of emergence, panicle initiation and physiological maturity, total dry matter and dry weights of leaf, culm, and head + grain, There were 59 data sets in total from different locations, and seasons. From these data sets, 48 were randomly selected to develop the relations between environmental factors and sorghum grain yield, referred to as models in this paper; the remaining 11 were utilized as independent data sets for testing the models.

Eight models are reported in this paper (Table 1). Models 1 to 4 were

TABLE 1—DESCRIPTION OF MODELS WITH THEIR COEFFICIENT OF DETERMINATION (R^2) AND ROOT MEAN SQUARE ERROR (RMSE)

Number	*Model*	R^2	*RMSE (kg/ha)*
1	2	*3*	*4*
1	$Y = 1913.71 + 5.1664\,SW - 1.1777\,R1$ $(2.3551)^{*} \quad (0.9515)$ $+ 5.0313\,R2 + 4.4864\,R3$ $(1.4638) \quad (1.1812)$	0.53	748
2	$Y = 478.83 - 1.5737\,ET1 + 13.82\,ET2 + 18.426\,ET3$ $(4.1570) \quad (5.0584) \quad (3.2959)$	0.52	758
3	$Y = -1017.1 - 37.9141\,R1 + 31.3804\,R2 + 33.6311\,R3$ $(13.9815) \quad (13.4277) \quad (23.8986)$ $-145.42\,T1 + 185.993\,T2 + 129.522\,T3 +$ $(104.375) \quad (154.548) \quad (109.582)$ $1.1928\,R1T1 - 1.069\,R2T2 - 1.143\,R3T3$ $(0.4630) \quad (0.5319) \quad (0.9382)$	0.68	635
4	$Y = -6015.56 - 3.1597\,R1 + 7.3877\,R2 + 5.3287\,R3$ $(1.5694) \quad (1.8105) \quad (1.6564)$ $+ 115.088\,T1 + 36.6705\,T2 - 152.018\,T3$ $(134.3910) \quad (143.8940) \quad (138.8490)$ $- 3.4349\,SR1 + 10.3843\,SR2 + 11.7183\,SR3$ $(4.0591) \quad (4.4873) \quad (6.4925)$	0.66	642
5	$Y = -455.3 + 5.3326\,SW - 2.9756\,TR - 0.11\,R2ET2$ $(2.0405) \quad (1.0463) \quad (0.0348)$ $+ 0.07194\,R3ET3 + 0.063\,R2SR2 + 0.1848\,T2SR2$ $(0.0111) \quad (0.0114) \quad (0.0747)$	0.76	540
6	$Y = 561.39 - 0.241\,SWTR - 0.0701\,R2ET2$ $(0.1245) \quad (0.0403)$ $+ 0.062\,R3ET3 + 0.0443\,R2SR2 + 0.1925\,T2SR2$ $(0.0130) \quad (0.0127) \quad (0.0896)$	0.64	656

Table 1 (*Contd. on page* 187)

Table 1 (*Contd. from page* 186)

1	2	3	4.
7	$Y = -2996.65 + 45.2459\,R3 + 5.1302\,SR3 + 0.5130\,R1T1$ (13.7164) (2.2664) (0.2010) $-1.9849\,R3T3 - 0.0782\,R2ET2 + 0.1011\,R3ET3$ (0.5631) (0.0314) (0.0339) $-0.0420\,R1SR1 + 0.0418\,R2SR2 + 0.2952\,T2SR2$ (0.0137) (0.0105) (0.7662)	0.81	474
8	$Y = 782.11 + 7.8741\,SW - 0.2403\,R2T2 - 0.1232\,R2ET2$ (1.9345) (0.132) (0.0354) $+ 0.0539\,R3ET3 + 0.0719\,R2SR2$ (0.009) (0.1322)	0.73	572

*Figure in parenthesis refers to standard errors.

Where Y = Observed grain yield (kg/ha)
SW = Available soil water (mm) at planting
TR = Total rainfall (mm) for the whole crop growing season
$R1$ = Total rainfall (mm) during $GS1$
$R2$ = Total rainfall (mm) during $GS2$
$R3$ = Total rainfall (mm) during $GS3$
ET = Total evapotranspiration (mm) for the whole crop growing season
$ET1$ = Total evapotranspiration (mm) during $GS1$
$ET2$ = Total evapotranspiration (mm) during $GS2$
$ET3$ = Total evapotranspiration (mm) during $GS3$
$T1$ = Mean temperature in °C during $GS1$
$T2$ = Mean temperature in °C during $GS2$
$T3$ = Mean temperature in °C during $GS3$
$SR1$ = Mean solar radiation (ly/day) during $GS1$
$SR2$ = Mean solar radiation (ly/day) during $GS2$
$SR3$ = Mean solar radiation (ly/day) during $GS3$

developed by selecting one or more independent variables, namely, soil water at planting (SW), rainfall (R), mean temperature (T), open-pan evaporation (E), solar radiation (SR), estimated evapotranspiration (ET) (computed after Ritchie [11]) for the whole growing season and for the three growth stages defined by Eastin [3]. These growth stages are from emergence to panicle initiation ($GS1$), from panicle initiation to anthesis ($GS2$), and from anthesis to physiological maturity ($GS3$). Several combinations of these variables—product and inverse relations—were made. A stepwise regression technique was used, and models 5 and 6 were selected based on R^2 values.

Models 7 and 8 were selected using the Mallow's (Daniel and Wood, [2]) Cp criterion. This criterion is defined as $Cp = (RSS/s^2) - (N - 2p)$ where RSS = residual sum of square for the best subset being tested; p = number of variables in the subset (including the intercept, if any); s^2 = residual mean square based on the regression using all independent variables. Neter and Wasserman [10] described that the model which has the smallest *Cp* value should be selected. Thus model 7 was selected because it had the lowest *Cp* value of 0.2. Neter and Wasserman [10] further described that the bias component of the models should also be examined to select the best model. This can be done by plotting the *Cp* values for all possible models against the number of parameters (p) used in the respective models. The model which has little bias component tends to fall near the line $Cp = p$. Those models with substantial bias will tend to fall above this line. Thus the relationship between *Cp* and p was plotted in Figure 1. Model 8 was selected as it had 6 parameters (p) and was close to the line $Cp = P$ with 4.9 as *Cp* value.

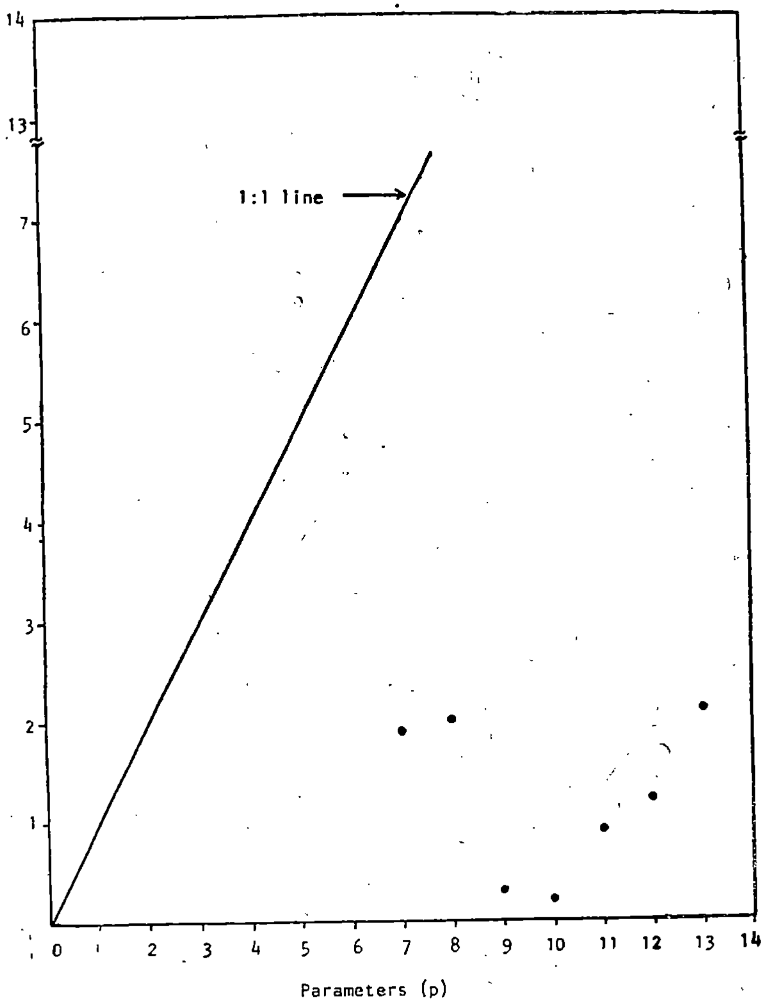

Figure 1. Relationship between *Cp* and parameters (p) used in several possible regression models.

Model Tests

The correlation coefficients between observed and predicted yield using these 8 models (Table 1) for 11 independent data sets were 0.59, 0.46, 0.60, 0.72, 0.72, 0.71, 0.72, and 0.77 respectively. These results show that only five models (4, 5, 6, 7, and 8) could explain the 50% to 59% variability in the yield data; the RMSE values for these models ranged from 851 to 994 kg/ha. The other three models had very low r^2 and high RMSE values. The actual grain yield for these 11 data sets were compared using model 8 simulated values (Fig. 2), as this model provided high r^2 values (0.59) and lowest RMSE values (851 kg/ha) when tested with independent data sets.

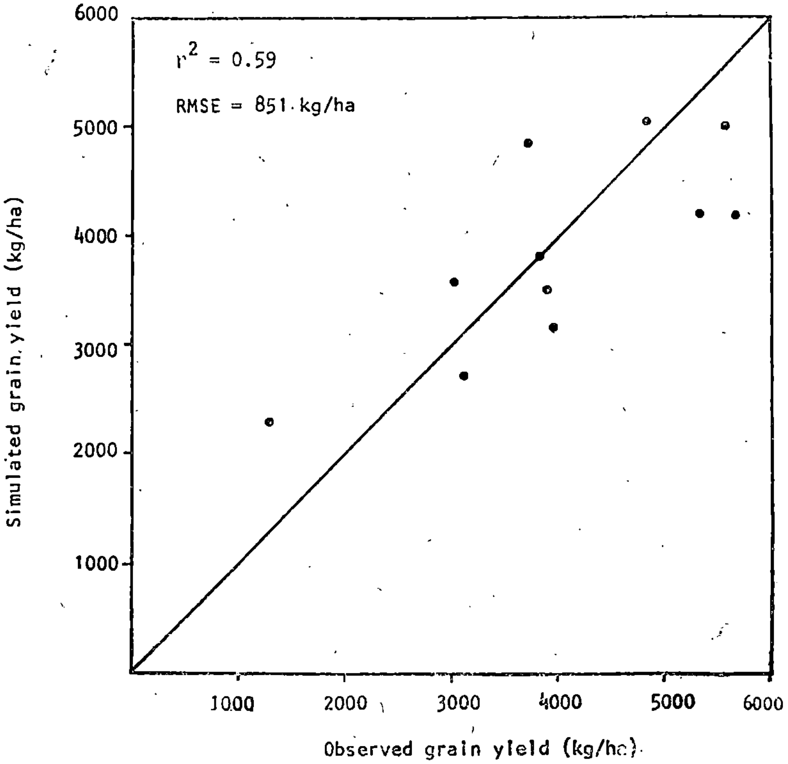

Figure 2. Relationship between observed and simulated grain yield of sorghum (model 8 test results with independent data set).

Discussion

Total rainfall for the entire growing season could only explain 29%

variability, but rainfall for *GS*1, *GS*2 and *GS*3 explained 53% variability. Similarly, evapotranspiration for the three growth stages gave higher R^2 values (0.52) compared to the seasonal total *ET* ($R^2 = 0.37$). Rainfall, mean temperature and their product for three growth stages (model 3) explained 68% variability in yield. Models 3, 4, and 6 gave similar R^2 values (0.64 and 0.68). Model 7 gave the highest (R^2) values (0.81).

Model 8 simulates 79 kg of additional grain yield if the available soil water at planting exceeds by 10 mm from its mean values of 92 mm. This is expected as soil moisture at planting helps in crop establishment and particularly for the postrainy season crop when sorghum is mainly raised on residual moisture stored in the soil profile. Reduction in grain yield due to additional rainfall in *GS*2 increases with the increase in temperature and evapotranspiration rate during this period. However, the beneficial effect of additional rainfall in the grain-filling period increases with increased *ET* rate during *GS*3. Probably these results have some basis, for example, in *GS*2 high *ET* means a higher vegetative growth rate and, thus, with additional rainfall, the crop tends to add vegetative parts causing a negative effect on grain yield. On the other hand, additional rainfall in *GS*3 along with high *ET* rate provides more green leaf and also probably a longer grain filling period thus resulting in increased grain yield.

Models 4 and 6 have lower R^2 values compared with that of models 5, 7, and 8 (Table 1) but all these models except model 8 provided similar R^2 values (0.50 to 0.53) when tested with independent data sets. Model 8 had R^2 value of 0.59. These results are in agreement with that of Gardner (7) that the performance of a prediction equation is better in the sample from which it is obtained than in a second independent sample. Comparing the data requirements of these models, we would suggest that model 4 could be used for prediction purposes as it requires only data on rainfall, temperature and solar radiation. The other four models need information on soil water and *ET*, which in turn requires leaf area data. If these data are available, model 8 should be used for yield prediction.

Conclusion

One of the eight models developed in this study (model 4) that requires rainfall, temperature and solar radiation data is recommended for yield prediction purposes in areas where growth simulation type models can not be applied because of data limitations. If data on soil moisture and evapotranspiration are available in addition to rainfall, temperature and solar radiation, model 8 can be used for yield prediction.

REFERENCES

[1] Brown, J. T. and Vanderlip, R. L. (1969): Relations of Palmer drought variables to yield of wheat, *Agronomy Journal* **61**: 95-98.

[2] Daniel, C. and Wood, F. S. (1971): *Fitting Equations to Data*, Wiley, New York.

[3] Eastin, J. D. (1971): Photosynthesis and translocation in relation to plant development. *In* N. G. P. Rao and L. R. House (Eds.), *Sorghum in the Seventies*, Oxford and IBH Publishing Co., Bombay, India.

[4] Feyerherm, A. M. and Paulsen, G. M. (1981): Development of a wheat yield prediction model, *Agronomy Journal* **73**: 277-282.

[5] Fisher, R. A. (1924): The influence of rainfall on the yield of wheat at Rothamsted, *Royal Society of Philosophical Transactions Series B* **213**: 89-142, London.

[6] Gangopadhyaya, M. and Sarkar, R. P. (1964): Curvilinear study on the effect of weather on growth of sugarcane, *Indian Journal of Meteorology and Geophysics* **15**: 215-226.

[7] Gardner, M. J. (1972): On using an estimated regression line in a second samples. *Biometrika* **59**; 263-274.

[8] Huda, A. K. S., Ghildyal, B. P., Tomar, V. S. and Jain, R. C. (1975) : Contribution of climatic variables in predicting rice yield, *Agricultural Meteorology* **15**: 71-86.

[9] LeDuc, S., Sakamoto, C., Strommen, N. and Steyeart, L. [1980] : Some problems associated with using climate-crop-yield models in an operational system : an overview. *Biometeorology* **7**; Part 2 : 104-114, supplement of *International Journal of Biometeorology* **24**.

[10] Neter, J. and Wasserman, W. (1974) : *Applied Linear Statistical Models : Regression, Analysis of Variance, and Experimental Designs*, Richard D. Irwin, Inc., Homewood, Illinois, 60430, USA.

[11] Ritchie, J. T. (1972) : Model for predicting evaporation from a row crop with incomplete cover, *Water Resources Research* **8** : 1204-1213.

[12] Thompson, L. M. (1969) : Weather and technology in the production of corn in the U. S. corn belt, *Agronomy Journal* **61**: 453-456.