Forecasting Grain Sorghum Yields Using Simulated Weather Data and Updating Techniques

G. F. Arkin, S. J. Maas, C. W. Richardson
MEMBER ASAE MEMBER ASAE

ABSTRACT

A methodology was developed by which SORGF, a grain sorghum growth-simulation model, could be used to forecast crop status during the growing season. The methodology utilizes simulated weather data, generated by a Markov chain model, as input to the grain sorghum model. The modeled status of sorghum plants may be updated at any time during the growing season with actual plant status observed in the field. Application of the methodology was demonstrated by forecasting date of physiological maturity (PM) and head dry weight at PM for grain sorghum crops grown in 10 fields in Central Texas during 1976.

INTRODUCTION

Yield forecasts for the major agricultural crops are generally produced using between-year crop yield models. These models are based on an established relationship between yield and weather obtained by statistically regressing historical yield information onto observed climatic data. Such models are location specific. Formulation of a between-year crop yield model for a location requires enougn concurrent records of yield and weather data to quantify the effects of year-to-year variations in weather during the growing season. Yield and weather data over many years also contain the effects of improved farm-management practices and possibly long-term climatic trends that must be removed before the year-to-year variations in climate and yield data can be obtained.

In recent years, within-year growth-simulation models have been developed for a number of agricultural crops (Arkin et al., 1979). These models generally consist of one or more submodels that translate environmental variables (atmospheric and soil) over an incremental time period into changes in plant parameter values. Growth-simulation models need not be location specific and do not require a base of climate and yield information.

This paper describes a methodology for using a within-year growth-simulation model, SORGF, for forecasting grain sorghum yield on an individual-field basis. SORGF simulates the daily accumulation of dry matter in the roots, leaves, stem, head, and grain of a sorghum plant as a function of daily weather data (Arkin et al., 1976; Vanderlip and Arkin, 1977; Maas and Arkin, 1978). The model includes submodels for determining photosynthetically-active radiation intercepted by the plant canopy (Arkin et al., 1978b) and water in the soil available to the root system (Ritchie, 1972). A sensitivity analysis of SORGF was performed by Maas and Arkin (1980).

In an earlier study Arkin et al. (1978a) using SORGF predicted grain sorghum yield based on conditional probability functions. The forecasting methodology presented here extends the flexibility of the previous method, overcoming previous restrictions on updating real-time model inputs of weather and crop status.

ASPECTS OF YIELD FORECASTING

The yield forecasting technique described in this paper in unique in its use of weather and crop data inputs, namely,

1 the use of simulated daily weather data allows yields to be computed for any desired number of seasons;

2 model inaccuracies propagated through the simulation can be corrected by updating the modeled growth and development during the season with observed plant parameter values.

Simulated Weather Data

The execution of SORGF requires the input of values of insolation, rainfall, and maximum and minimum air temperature for each day of the growing season. The fact that current weather prediction techniques apparently exhibit no skill for forecast periods greater than 1 month (Ramage, 1978) effectively precludes their use as a source of data inputs for the yield model. An alternate source of daily weather data is climatic records. By executing SORGF using a series of seasonal sequences of daily weather, we can determine certain statistical parameters relevant to yield that have predictive significance for any given season. In general, the stability of these parameter values increases as the number of observations increases. Thus, it might be desirable to simulate crop growth and yield at a location for more individual seasons than historic weather records would normally allow.

In recent years, Markov chain modeling has been used as a means of generating sequences of simulated weather data that closely resemble actual sequences of observed weather at a location. Crank (1977) utilized a Markov chain model and empirically derived functions to generate daily values of rainfall, sunshine, and maximum and minimum air temperature for Des Moines, Iowa, during June, July and August. This methodology

Article was submitted for publication in April 1979; reviewed and approved for publication by the Structures and Environment Division of ASAE in January 1980.

Contribution from the Texas Agricultural Experiment Station (TAES), Texas A&M University, College Station, in cooperation with Science and Education Administration, Agricultural Research, U.S. Department of Agriculture (USDA-SEA-AR), Temple, TX.

The authors are: G. F. ARKIN, Assistant Professor, and S. J. MAAS, Research Scientist, TAES; and C. W. RICHARDSON, Agricultural Engineer, USDA-SEA-AR, Temple, TX.

Acknowledgement: Funding for this study was provided by the Economics, Statistics and Cooperatives Service, United States Department of Agriculture, and the Texas Agricultural Experiment Station, Texas A&M University System.

TABLE 1. COMPARISON OF SIMULATED AND OBSERVED WEATHER DATA FOR RIESEL, TEXAS

Variable and units	Statistic*	Average absolute difference†
Daily maximum temperature, °C	MEAN	0.9
	SD	0.9
Daily minimum temperature, °C	MEAN	0.9
	SD	0.6
Daily rainfall, cm	MEAN	0.1
	SD	0.4
Daily solar radiation, Langleys	MEAN	39
	SD	29

*MEAN = Mean for a given day over length of record;
SD = Standard deviation among values for a given day over length of record.
†Average difference between daily Mean or SD values over the period March 1 through August 31.

was used to generate simulated weather data for March through August for Bell County in central Texas. The only significant modification to the original technique was a provision allowing daily insolation to be output in terms of langleys per day, instead of minutes of sunshine per day. We used 38 yr of precipitation data obtained at Riesel, Texas (96°53′W, 31°29′N) to formulate the Markov probabilities required to generate daily rainfall occurrences during each month of the growing season. We also used these data to define empirical distribution functions for each month used in determining the amount of rainfall per occurrence. We constructed an empirical relationship that determined today's predicted maximum temperature from today's predicted precipitation and yesterday's observed maximum temperatuare, using 37 yr of temperature and precipitation data from the Riesel, Texas location. We also derived a function relating daily maximum and minimum air temperature from this set of data. An empirical relationship for determining insolation from today's predicted precipitation and maximum temperature was developed from 8 yr of data at Riesel.

Simulated weather data generated using this technique are similar to actual weather data observed in the Bell County area. The average (over the period from March 1 through August 31) of the absolute differences between the means and standard deviations for 50 and 37 yr, respectively, of simulated and observed (at Riesel, Texas) daily values of maximum and minimum temperature, rain and solar radiation is presented in Table 1. These results indicate that the mean daily difference between simulated and observed maximum or minimum temperature is less than 1 °C. The standard deviation values also differ by less than 1 °C, indicating that the year-to-year variability in temperatures is similar for both simulated and observed weather data. Mean and standard deviation values derived from simulated and observed rainfall data have an average difference of only a few millimeters per day, while similar statistics for simulated and observed solar radiation exhibit an average difference of less than 40 langleys per day. Thus, results of crop modeling efforts utilizing this simulated weather data should be applicable to the Bell County area.

Model Updating Technique

For each day from plant emergence through physiological maturity that values of the climatic variables are input, SORGF produces a description of the modeled sorghum plant in terms of number of leaves, areas of leaves, and weights of plant organs (roots, leaves, stalk, head, and grain). A feature built into SORGF allows the simulation to be updated during the growing season by re-assigning plant parameter values based on field observations. This method for updating plant parameters has been termed "feedback". By using feedback, the model simulation may be started on any day after emergence of the sorghum crop. Execution of the model then proceeds normally through the remainder of the growing season to maturity.

Updating the progress of a modeled crop during the growing season with observed plant status represents a distinct advantage in forecasting crop yield. If, for example, the model output indicates that sorghum plants should have 12 leaves 40 days after emergence (DAE), and actual field measurements show that sorghum plants have only eight leaves, then through feedback the model simulation may be re-started on day 40 (the feedback date) with modeled plants having eight leaves. Thus the model implicitly accounts for whatever environmental factor caused the difference between modeled and observed leaf number before the feedback date. In some cases, reassignment of a plant parameter based on field data may dictate that the values of other associated plant parameters be likewise adjusted in the model. Although feedback does not guarantee that additional differences between modeled and observed parameters will not occur after the feedback date, it does eliminate the contribution to the total forecast error incurred before the feedback date.

Forecast Methodology

Making a yield forecast using SORGF involves simulating crop growth and development for many years. A distribution of yields may be constructed from the simulation results, from which the following can be determined:

1 the probability that a certain yield value might occur,
2 the most likely occurring yield,
3 the greatest and smallest occurring yield,
4 the probabilities that the yield may be greater or smaller than a particular value,
5 the average yield value expected over many years, and
6 the expected year-to-year variability in yields over many years.

For a particular set of initial field conditions the stability of these statistical parameters increases as the number of simulated growing seasons are increased. Based on this consideration, 50 yr of simulated weather data were used to make the yield forecasts described in this paper. Additional simulations result in only a small increase in statistical parameter stability.

The methodology for forecasting grain sorghum yield is outlined in Fig. 1. Initially, cultural practices and field characteristics, like row spacing, geographical latitude, and soil water content, must be specified in the model, whether or not feedback is used. If the model simulation begins at planting, then the feedback capability of SORGF is bypassed, and plant parameter values are set at zero. If sorghum crop simulation begins after

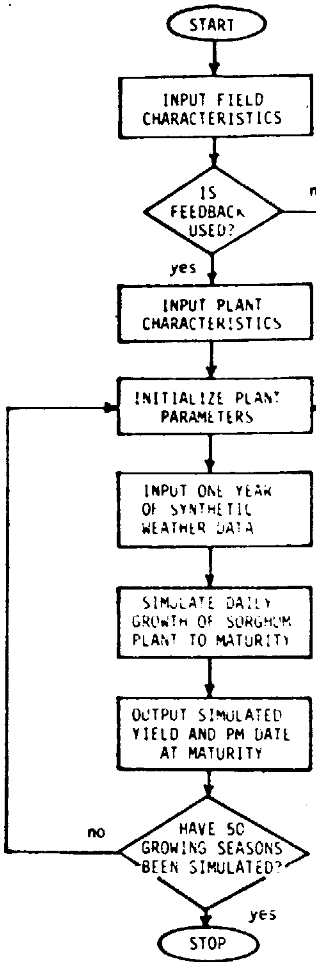

FIG. 1 Flow chart of activities in the yield forecasting technique.

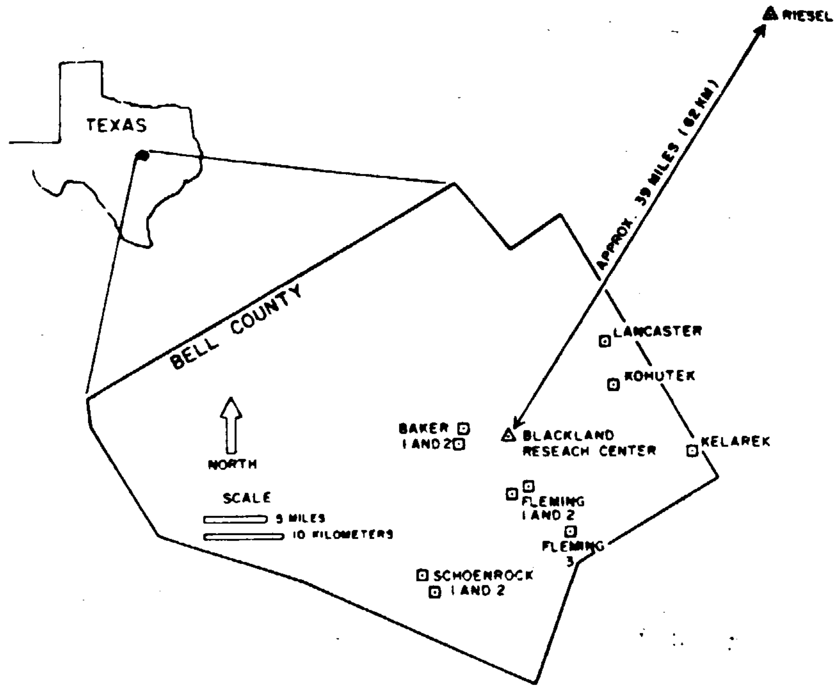

FIG. 2 Location of field measurement sites (denoted by squares and designator names) in the 1976 field study. Triangles indicate the Riesel and Blackland Research Center locations where insolation and temperature data were obtained.

emergence, then plant characteristics observed on that feedback date are used to initialize plant parameters in the model. These values of the field and plant parameters at this point in the forecast sequence become the initial conditions for all the yearly crop growth and yield simulations. In each of these yearly simulations, a set of simulated weather data for the growing season is input. Based on the initial conditions and the weather data, the model simulates the growth and development of the sorghum crop to maturity. When maturity is reached, a value for crop yield is output. This sequence of operations was repeated for all 50 yr for which simulated weather data were specified in this study.

DEMONSTRATION OF METHODOLOGY

In 1976 the growth and development of grain sorghum in ten fields in Bell County, Texas (Fig. 2) were observed throughout the growing season. Data obtained from these observations were used to demonstrate the application of the forecasting techniques.

Acquisition of Feedback Data

At the start of the growing season, each field was characterized by soil type and soil water was measured gravimetrically. Management practices, including fertilizer applications, sowing rates, row spacings and row orientations, were recorded at planting. After emergence, measurement of plant characteristics coincident with important phenological events, i.e., growing point differentiation (GPD), half-bloom (HB), and physiological maturity (PM) was scheduled. Direct observations of GPD were not obtained in the field, but we inferred GPD occurred midway between complete expansion of the fifth leaf (counting from the leaf nearest the ground) and appearance of the flag leaf in the whorl (Vanderlip et al., 1977).

We selected two sampling sites in each field to reduce the effects of possible within-field variations. On each sampling date, all plants within a 0.5- to 1-m row portion in each site were cut at ground level. Numbers and areas of leaves were measured and plant parts were separated and dried at 65 °C to determine weights of leaves, stalks, and heads. Representative plant characteristics for each field were obtained by averaging all plants sampled within both sites of each sampling date.

Sets of feedback data were constructed based on representative plant characteristics from each of the ten fields. Some feedback data (leaf number, area, and weight; stalk and head weight; and the dates of emergence, GPD, HB and PM) could be obtained directly from the field measurements. However, some feedback data (root weight and soil water on the feedback date) were not measured. Values of these parameters were estimated by simulating growth and development of sorghum in each field up to the feedback date using SORGF and observed weather data for 1976. Rainfall data for these simulations were recorded at each of the ten fields, although rainfall for the Kohutek and Lancaster fields during a portion of May and June was estimated from amounts received at nearby fields. Daily maximum and minimum temperatures were measured at Blackland Research Center and assumed representative for the ten fields sampled in Bell County. Daily insolation data were obtained at Riesel and also assumed representative for these fields.

TABLE 2. CROP STATUS FORECAST FOR THE BAKER 1 FIELD, WITH AND WITHOUT FEEDBACK INPUT

Sampling time§	Parameters and units	Parameter values: Forecasted using: No feedback	Feedback at GPD§	Feedback at HB§	Observed
GPD	Julian date GPD	111			123
	Leaf area/plant, cm^2	425	Simulation		372
(51 DAE*)	LAI‡	0.8	initiated		0.8
	Plant dry weight†, g	2.0			2.4
	Head dry weight, g	0.0			0.0
HB	Julian date HB	141	153		158
	Leaf area/plant, cm^2	1807	1737	Simulation	1526
(86 DAE*)	LAI‡	3.3	3.2	initiated	2.0
	Plant dry weight†, g	24.4	26.6		28.1
	Head dry weight, g	4.9	5.6		3.7
PM	Julian date PM	169	183	192	195
	Leaf area/plant, cm^2	1618	1532	1337	873
(123 DAE*)	LAI‡	2.9	2.8	2.4	1.4
	Plant dry weight†, g	54.5	59.5	58.4	50.7
	Head dry weight, g	33.8	37.4	34.6	35.7

*DAE = Days after emergence
†All above-ground plant parts
‡Leaf Area Index = (leaf area/plant) / (ground area/plant)
§GPD = Growing point differentiation; HB = Half-Bloom; PM = Physiological maturity

Execution of Model

To demonstrate the yield forecasting technique, SORGF was executed according to Fig. 1 using 50 yr of simulated weather data. For each of the 10 fields, executions of the model were initiated on the dates of planting, GPD and HB. Those executions initiated on the planting date did not require feedback information. Model executions initiated on GPD and HB utilized feedback information obtained from the field study. Use of simulated weather data resulted in 50 forecasts of yield (head weight) and physiological maturity date for each of the three initiation dates at each of the ten locations—a total of 1500 executions of SORGF.

RESULTS AND DISCUSSION

A sample of detailed forecast results for one field is presented in Table 2 to show how feedback during the growing season can affect model predictions. The table contains the results of forecasts initiated at planting (requiring no feedback), GPD and HB. Forecasted parameter values are averages of the results of the 50 simulated growing seasons and may be compared with their respective values observed in the field.

At GPD, comparison of observed and forecasted plant parameters in Table 2 indicates that the length of the vegetative period was underestimated by 12 days using the model. This hastened phenological development resulted in a greater leaf area development for the modeled plant. However, the modeled plant had a shorter period to accumulate dry matter, resulting in a smaller plant dry weight. No significant accumulation of dry matter by the head had occurred by GPD.

At HB, the model simulations initiated at planting were on the average 17 days ahead of plant development observed in the field, while the error in phenology was reduced to 5 days by feedback at GPD. Continued rapid development of the modeled plant resulted in a considerable overestimate of leaf area and leaf area index (LAI) at HB, although feedback at GPD did reduce the magnitude of the error. Feedback at GPD resulted in greater accumulation of plant and head dry weights, as compared to the forecasts initiated at planting. Both forecasts slightly overestimated head dry weight at HB, although the error in either case is less than 2 g.

At PM, model simulations initiated at planting, GPD and HB were on the average 26, 12 and 3 days, respectively, ahead of observed plant development. Neither of the three forecasts could account for the observed magnitude of leaf area senescence between HB and PM, although the use of feedback at GPD and HB did lead to greater reductions in leaf area as compared to the forecast made at planting. The additional leaf area ascribed to the modeled plants accounts for the overestimates of LAI and plant dry weight at PM shown in the table. A small improvement in accuracy is exhibited for the forecasts of head dry weight using feedback, although all forecasts are within 2 g of the observed value of head weight at PM.

Average forecast dates of PM and head weight per plant at PM, obtained from the 50 growing seasons simulated for each of the ten fields is summarized in Tables 3 and 4, respectively. The results of forecasts initiated at planting (using no feedback), GPD and HB,

TABLE 3. SUMMARY OF PM DATE FORECASTS FOR TEN FIELDS IN BELL COUNTY, TEXAS, 1976

Field designator	Forecasted using: No feedback Mean	No feedback SD*	Feedback at GPD Mean	Feedback at GPD SD*	Feedback at HB Mean	Feedback at HB SD*	Observed
Baker 1	169†	6	183†	4	192†	0	195†
Baker 2	162	6	174	3	188	0	195
Fleming 1	194	6	201	4	211	0	216
Fleming 2	168	6	190	3	194	0	196
Fleming 3	208	12	227	3	215	0	216
Kelarek	170	6	192	3	208	0	216
Kohutek	172	6	200	3	207	0	217
Lancaster	185	6	201	3	208	0	217
Schoenrock 1	176	6	194	3	190	0	195
Schoenrock 2	168	6	185	4	194	0	195
Average	177	7	195	3	201	0	206
RMSE (days)	29		13		5		

*Standard deviation, days.
†Units = Julian date.

TABLE 4. SUMMARY OF HEAD WEIGHT FORECASTS (g/plant) FOR TEN FIELDS IN BELL COUNTY, TEXAS, 1976

Field designator	Forecasted using:						Observed
	No feedback		Feedback at GPD		Feedback at HB		
	Mean	SD*	Mean	SD*	Mean	SD*	
Baker 1	33.8	3.6	37.4	4.5	34.6	4.9	35.7
Baker 2	23.4	4.5	26.8	2.7	24.0	3.0	26.1
Fleming 1	25.9	2.9	29.0	3.5	25.3	1.8	27.1
Fleming 2	25.3	2.9	29.0	3.5	25.3	1.8	37.1
Fleming 3	52.4	12.6	49.1	15.2	35.3	10.7	30.2
Kelarek	29.4	3.2	33.3	4.3	21.1	7.3	23.8
Kohutek	26.2	2.9	32.4	6.4	24.4	7.3	25.0
Lancaster	33.8	5.0	33.7	7.2	26.3	8.1	36.3
Schoenrock 1	32.2	3.9	35.9	5.0	23.2	4.2	24.7
Schoenrock 2	31.1	3.5	34.4	3.3	27.6	4.0	27.6
Average	31.4	4.6	34.1	5.5	27.2	5.7	28.5
RMSE (grams)	5.2		6.1		2.6		

*Standard deviation, grams.

along with values of PM date and head weight observed for the ten fields is given in these tables.

The effect of feedback in improving forecasts of PM date as the growing season progresses is evident in comparisons of forecasted and observed date of PM in Table 3. The root-mean-square error (RMSE) between predicted and observed PM dates for the ten fields was reduced from nearly a month for forecasts without feedback to less than 1 wk for forecasts initiated at HB. The standard deviation among the 50 forecasts of PM date also decreased as the forecast date was delayed from planting to GPD and HB. Since the phenological development in SORGF is determined for a given field by average daily temperatures (Maas and Arkin, 1978a), errors in phenology tend to be accumulated at a relatively constant rate through the growing season (c.f., the errors in the predictions of GPD, HB and PM using no feedback in Table 1). Initializing the 50 simulations through feedback during the growing season reduces the length of the period over which errors in phenology may accrue, thus reducing the total error in PM date predicted at the end of each simulated growing season.

The effects of feedback are less evident in the averages of predicted head weights at PM shown in Table 4. Unlike PM date, no significant reduction in the RMSE between predicted and observed values of head weight is achieved until feedback is used at HB. Maas and Arkin (1979) demonstrated that, of all the environmental variables, SORGF was most sensitive to changes in soil water in simulating yield. Most of the response of the model to changes in soil water was found to occur during the period from HB to PM. Whereas error was accumulated at a relatively constant rate over the entire growing season in the prediction of PM date, most of the error in predicting head weight was accumulated after HB when the model was most sensitive to changes in soil water. Thus a consistent decrease in the standard deviation among head weight forecasts for a field is not evident in Table 4, since all the forecasts were initiated at or before HB.

CONCLUSIONS

Several conclusions concerning this forecasting methodology may be drawn from the preceeding analysis, vis.,

1 The use of feedback may not eliminate errors in forecast parameter values, but feedback generally does result in predicted parameter values closer to their respective observed values.

2 For forecast parameter values that accumulate error at approximately a constant rate over the growing season, feedback at any time during the growing season generally results in greater overall forecast accuracy and less variability among forecasts for individual years.

3 For forecast parameter values that are sensitive to environmental conditions during only a portion of the growing season, feedback prior to the period of sensitivity may not improve overall forecast accuracy or reduce variability among forecasts for individual years.

Aside from being used strictly as a yield forecasting technique, this methodology could be useful in decision making processes involving irrigation scheduling, supplemental fertilizer application and pest management programs.

References

1 Arkin, G. F., R. L. Vanderlip, and J. T. Ritchie. 1976. A dynamic grain sorghum growth model. TRANSACTIONS of the ASAE 19(4):622-626, 630.

2 Arkin, G. F., C. W. Richardson, and S. J. Maas. 1978a. Forecasting grain sorghum yields using probability functions. TRANSACTIONS of the ASAE 21(5):874-877, 880.

3 Arkin, G. F., J. T. Ritchie, and S. J. Maas. 1978b. A model for calculating light interception by a grain sorghum canopy. TRANSACTIONS of the ASAE 21(2):303-308.

4 Arkin, G. F., C. L. Wiegand and H. Huddleston. 1979. The future role of a crop model in large area yield estimating. Proc. of the Crop Modeling Workshop, Columbia, Missouri, October 3-5, 1977.

5 Crank, K. 1977. Simulating daily weather variables. Research and Development Branch, Research Division, Economics, Statistics, and Cooperatives Services, USDA, Washington, DC.

6 Maas, S. J. and G. F. Arkin. 1978. User's guide to SORGF: A dynamic grain sorghum growth model with feedback capacity. Research Center Program and Model Documentation No. 78-1 (Blackland Research Center at Temple), Texas Agricultural Experiment Station, College Station, TX.

7 Maas, S. J. and G. F. Arkin. 1980. Sensitivity analysis of SORGF, a grain sorghum model. TRANSACTIONS of the ASAE 23(3):671-675 (this issue).

8 Ramage, C. S. 1978. Further outlook—hazy. Bull. Amer. Meteorol. Soc. 59(1):18-21.

9 Ritchie, J. T. 1972. Model for predicting evaporation from a row crop with incomplete cover. Water Resources Research 8(5):1204-1213.

10 Vanderlip, R. L. and G. F. Arkin. 1977. Simulating accumulation and distribution of dry matter in grain sorghum. Agron. J. 69:917-923.