SIMULATION OF SUGARCANE YIELD AT THE SCALE OF A MILL SUPPLY AREA

TG LUMSDEN, NL LECLER AND RE SCHULZE

Department of Agricultural Engineering, University of Natal, Private Bag X01, Scottsville, 3209

Abstract

In order to develop a sugarcane yield forecasting system, a crop yield simulation model is evaluated at the scale of a mill supply area (MSA), using historical climate and yield data. The areal extent of this scale necessitates the development of spatially distributed model input, both for historical model evaluation and for future implementation of a yield forecasting system. Research into the effect of crop cycles on yield, the derivation of input required to run the simulation model at the scale of a MSA and a preliminary comparison of simulated and observed cane yields is presented in this paper. It was confirmed that cycles have an effect on yields at the field scale. A modelling strategy incorporating crop cycles was developed for modelling at the scale of a MSA and a comparison of simulated versus observed yields gave rise to an index of agreement (d) of 0,836 and a Pearson's correlation coefficient (r) of 0,760. The implications of research carried out to date on the subsequent incorporation of seasonal climate forecasts are discussed.

Introduction

In the sugar industry there are numerous benefits that can be derived from having accurate and timely forecasts of cane yield. These benefits include:

- mill operations planning
- improved harvest and haulage scheduling
- tailoring of agronomic practices to suit the climate expected for a given season.

A variety of techniques are available for yield forecasting. These techniques include rules of thumb, neural networks, statistical modelling (including regression modelling), remote sensing and simulation modelling.

Regression models can be simple to develop and give good results for the areas where they are developed. However, application of the equations developed by such techniques to other areas can be problematic, and their development is reliant on long reliable records of crop yields. Remote sensing techniques monitor the condition of crops over wide areas with good spatial and temporal resolution. Regression techniques can be used to relate remote sensing derived vegetation indices to crop yield. Crop yield simulation techniques have the advantage of being suitable for the incorporation of increasingly accurate mid to long range seasonal climate forecasts (Landman, 1997) and have the potential to be used to determine the management strategies suited to particular seasonal conditions.

In order to develop a cane yield forecasting system using crop simulation models and seasonal climate forecasts, the ability of the simulation models to predict historical yields at the scale of a MSA using observed climate and yield data, must first be assessed. The development and assessment of a cane yield simulation model, namely the *ACRU*-Thompson model, using historical data from the Eston mill supply area, is presented in this paper.

Study area

The Eston mill supply area is in the Midlands of KwaZulu-Natal province in South Africa and is located around latitude 29°55'S and longitude 30°30'E. The area is comprised largely of farms which supplied the former Illovo Mill. Yield records are available from the old mill and are used solely for the MSA study. Mean annual precipitation (MAP) in the MSA ranges from 600 to 950 mm, and altitude ranges from 600 to 950 metres. Soils originate mostly from Table Mountain Sandstone (ordinary) and Dwyka tillite and, to a lesser extent, from Table Mountain Sandstone (mistbelt), dolerite and Lower Ecca shale parent materials (Hellmann, 1993).

Methods

Investigation of effect of crop cycles on field scale yields

The harvest season for sugarcane typically extends for nine months of the year, reflecting the many possible combinations of ratooning/planting dates and crop age. A useful classification of the crop cycles practised in the Eston area has been developed by Hellmann (1993), who showed that crop cycles have a significant effect on the yield of variety N12. This has implications for modelling a sugarcane crop as it may be necessary to account for a number of different cycles within the modelling framework. In order to determine a modelling strategy to cater for various crop cycles a study of the effect of crop cycles on yield was conducted using observed field scale yield data from the Field Record System for a representative farm in the Eston MSA. The results of this study are presented in Figure 1.

The yield data were averaged for the farm over the period 1986 to 1993. The first letter of the cycle type indicates whether the cycle is a plant (P) or ratoon (R) crop, the second letter indicates the season of planting or ratooning, this being either spring/summer (S) or winter (W), the third letter indicates season of harvest, this being spring/summer (S) or winter (W) and the final 1S or 2S indicates whether the crop grew through one or two summers. Curves are plotted for data averaged over all years and also for the individual years of 1987, 1991 and 1992. The curves indicate that cycles do affect yield and that this trend is evident even though there are a number of varieties present. The curves for the individual years follow the trend of the average curve for all years, although their absolute values differ substantially in some cases. The curve for 1991 tends to be higher than the average curve, for 1992 lower than the average and for 1987 the curve is very close to the average. Climate is the major factor that varies from year to year and it is evident that both cycles and climate (during the cycle under consideration) contribute to variability of cane yields. The influence of climate on yield implies that care must be taken when developing climate data sets for MSA modelling. The effect of cycles on yield would indicate that they should be taken into account when modelling. A strategy for modelling the MSA was proposed whereby each farm would be modelled under all cycles as defined by Hellmann (1993) for the Eston MSA. The yields simulated for each cycle would then be weighted according to the district average proportion that a particular cycle contributes to total production, giving rise to an average yield for each farm. Farm averages could then be averaged to obtain an MSA average yield. The district average proportions that the various cycles contribute to production (by area), were estimated by analysing field data from 11 farms in the district. The proportions calculated may be found in Table 1.

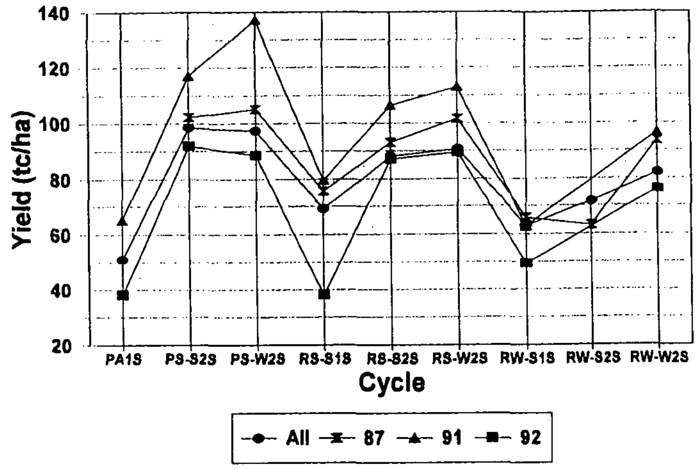

Figure 1. Mean observed yield according to crop cycle for all and select years between 1986 and 1993.

Table 1. District average proportions of production (by area) according to crop cycle (%).

PA1S	PA2S	PS-W2S	PS-S2S	RS-S1S	RW-S1S	RS-W2S	RS-S2S	RW-W2S	RW-S2S	RW-W1S
1,69	0,08	5,30	11,38	6,65	8,23	22,78	17,70	20,00	5,41	0,79

Thompson (1976) based ACRU cane yield simulation model

The *ACRU* agrohydrological modelling system (Schulze, 1995) contains yield simulation modules for a variety of crops. The existing sugarcane yield module is based on the Thompson (1976) relationship (Equation 1) and in *ACRU* it was developed to estimate sugarcane yields assuming a 12 month growing cycle (July to June). In order to cater for various growing cycles and planting dates that occur in the Eston study area, the yield model was modified by incorporating a thermally driven crop coefficient based on the relationship given in Equation 2 which was derived by Hughes (1992). The rate of development of the crop potential evapotranspiration is thus related to accumulated temperature and the consequences of varying planting dates and inter- and intra-seasonal temperature differences on the rate of crop development and potential water use are therefore represented. The limitation on the length of the growing season has also been removed so that growing seasons of various lengths can be simulated.

In the *ACRU*-Thompson model:

$$Y = 9{,}53\,(AET_{sum}/100) - 2{,}36 \qquad (1)$$

where Y = sugarcane yield (t/ha)

AET_{sum} = accumulated growing season total evaporation (actual evapotranspiration) in mm, determined using the 'dryland' water budget in *ACRU*.

In order to estimate total evaporation, a reference evaporation (e.g. Class A pan) is used together with the crop coefficient, which is estimated as:

$$K_c = 0{,}297 + (1{,}32\text{x}10^{-6}\text{x}GD_a^{\,2}) - (6{,}83\text{x}10^{-10}\text{x}GD_a^{\,3}) - K_{red} \qquad (2)$$

$$K_{red} = 0{,}05 + (1{,}32\text{x}10^{-6}\text{x}GD_r^{\,2}) - (6{,}83\text{x}10^{-10}\text{x}GD_r^{\,3})$$

where
K_c = sugarcane crop coefficient
GD_a = accumulated degree days since planting (°C day)
GD_r = accumulated degree days since initiation of ripening (°C day).

$$\text{Degree day} = ((T_{max}+T_{min})/2) - 12 \ (\text{°C day}) \qquad (3)$$

where T_{max} = daily maximum temperature (°C)
T_{min} = daily minimum temperature (°C).

Limits K_c $\leq$ 1 for plant crop
$\leq$ 0,96 for first ratoon crop
$\leq$ 0,92 for second and subsequent ratoons
$\geq$ 0,5 during ripening
GD_a $\leq$ 1 300 (to prevent –ve values).

The *ACRU*-Thompson cane yield model combines aspects of simulation and regression modelling.

Preparation of climate input information for MSA modelling

Climate information required for the *ACRU*-Thompson model includes daily rainfall and temperature. For the period under consideration (1988 to 1995) there are 13 rainfall and 12 temperature stations reporting on a daily basis in and around the MSA. The location of these climate stations is indicated on a map of the Eston MSA in Figure 2. There is also a location map of the Eston MSA indicating its position within South Africa. The stations indicated as having only temperature also have rainfall records. However, their rainfall data were not used as they were not considered as representative as the data at other rainfall reporting stations.

The mapped rainfall stations were used to derive daily rainfall data sets for the 93 farms considered in the MSA. When deriving these data sets any missing days or periods of missing data, had to be patched or infilled to ensure completeness of record. The method employed was the inverse distance weighting technique, which makes use of data from a number of surrounding stations and weights these data according to the inverse of the square of the distance of the surrounding stations to the station whose record is being patched. There is also account taken of the relative MAP differences between the stations. The equation used is as follows:

$$r_s = \frac{\frac{R_d r_1}{R_1 d_1^2} + \frac{R_d r_2}{R_2 d_2^2} + \frac{R_d r_3}{R_3 d_3^2} + \ldots}{\frac{1}{d_1^2} + \frac{1}{d_2^2} + \frac{1}{d_3^2} + \ldots} \qquad (4)$$

where

r_s = synthesised rainfall (mm) of the station being patched for a specific day

$r_{1,2,3}$ = actual rainfall (mm) recorded at surrounding stations

$d_{1,2,3}$ = distance (degrees decimal) of surrounding stations from the station being patched

$R_{1,2,3}$ = MAP (mm) at the grid points of the surrounding stations

R_d = MAP (mm) at the grid point of the patch station.

The grid point MAP values are taken from a national grid coverage developed by Dent *et al.* (1989). This coverage has a resolution of one minute by one minute of a degree latitude and longitude. Once the rainfall station values were patched, each farm in the MSA was associated with a rainfall station that is most likely to represent the rainfall occurring on that farm. These stations are referred to as the 'driver stations' for those farms. Their daily data are multiplied by the ratio of the

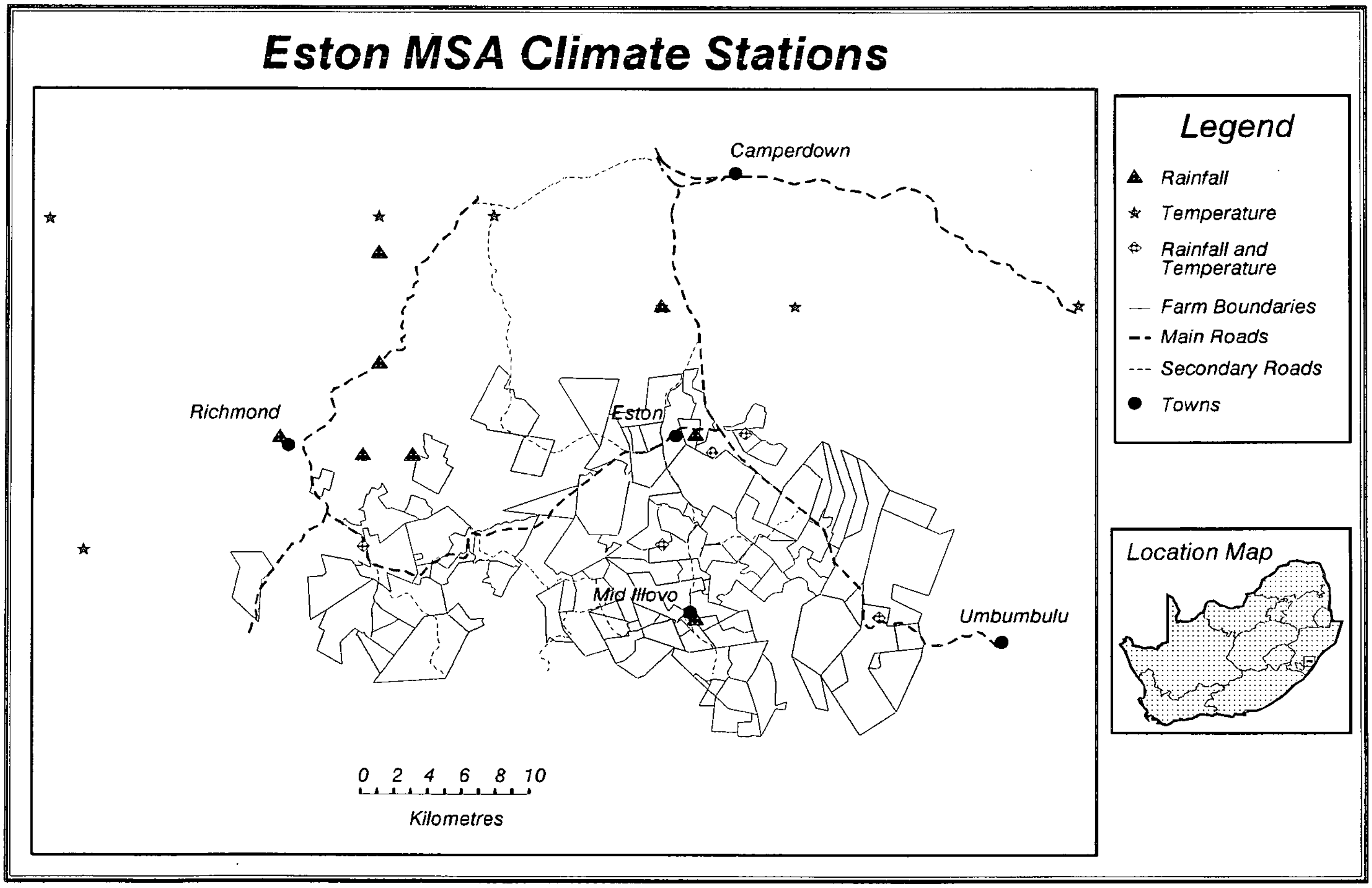

Figure 2. Daily reporting climate stations falling within and around the Eston MSA over the period 1988 to 1995 and location map of the Eston MSA within South Africa.

average of the grid point MAP of the farm to the grid point MAP at the driver station, in order to develop the farm data sets.

To estimate daily maximum and minimum temperature at each farm, the missing daily values at surrounding temperature stations were first patched using a 'nearest neighbour' technique, where the nearest neighbour station was selected by a weighting according to distance and altitude difference ([1]personal communication). From these patched temperature station records the temperatures were estimated at the farms by application of regional lapse rates. The regional lapse rates have been determined for each month of the year and for maximum and minimum temperatures separately. Verification tests have proved this method to give very realistic daily temperature estimates ([1]personal communication).

Climate data were sourced from the Computing Centre for Water Research, the Institute for Soil, Climate and Water (ISCW) and the South African Sugar Association Experiment Station (SASEX) databases.

Preparation of soil input information for MSA modelling

Two sources of soils information are envisaged for use in deriving soil input parameters for the crop simulation model:

- soil parent material maps which are widely available for areas in the sugar industry
- Land Type information which may be purchased for most of South Africa from the ISCW.

The results presented in this paper were derived using the Land Type information. Land Types are units of homogeneous land use potential and their data relate to, among other variables, soil form and series as well as soil depth. Schulze *et al.*, (1985) have developed generic model inputs for all soil forms and series found in the South African classification system, based on clay distribution models and texture classes. The soil parameters for all the soil series falling within a particular Land Type may rapidly be area and depth weighted via the AUTOSOIL (Pike and Schulze, 1995) computer program to give averaged values for each of the Land Types falling within the MSA. The proportions of Land Types occurring per farm which must previously have been determined, are then used by the computer program to calculate area weighted soil model input parameters for each farm. There are 60 Land Types present in the rectangular area that contains the MSA and stretches 5-10 km around it.

The parent material derived soil model inputs can be derived in a similar way except that the soil form and series are initially predicted from various appropriate factors as opposed to the Land Type based technique where soil form and series have been mapped. The advantage of the parent material based technique is that parent material maps are widely available in the sugar industry. It is also believed that there is a strong correlation between parent material and cane yield ([2]personal communication). The factors used in predicting soil form, series and depth include parent material, MAP, slope and slope position. A set of working rules for predicting soil type was developed by soil experts and the local extension officer from SASEX ([2]personal communication). The working rules will be incorporated into a geographical information system (GIS) along with gridded coverages of parent material, MAP, slope and slope position. The coverage of slope position remains to be developed, and once completed will allow for the coverage of soil form, series and depth to be predicted. The final model inputs will then be generated as was described earlier. The development of two methodologies for generating soil model inputs should allow in future for an interesting comparison in terms of effort required and resulting accuracy in yield simulations.

Observed cane yield database

Two cane yield databases were considered as sources for model verification. These were the mill database and the South African Sugar Association (SASA) Annual Survey of Area Under Cane and Cane Production. The two databases are very similar, the difference being that the Mill reviews the areas estimated by farmers and adjusts them if necessary. The tonnes cane weighed at the mill are recorded in both databases. Records are kept by the mill only for total farm production and not for individual fields. The SASA database was selected for use in model verification. This is because the database is consistently maintained for large areas of the sugar industry and has a format that is more compatible for incorporation into a GIS. The data have been incorporated into a GIS making use of a relational database concept for convenience of use and storage. The system allows for easy viewing, mapping and querying of data. Data for an eight year period between 1988 and 1995 have been obtained, as they were most readily available for this period. Some data are, however, missing or have suspect values.

MSA modelling strategy

In order to model the 93 farms in the MSA over eight years and for a number of cycles, a large number of model runs was required. It was decided that separate runs would be done for each harvest season as the cycles being modelled range in length from one to two years. Running sequences of years would not prove practical in this situation. A computer program was developed that facilitates multiple runs of the *ACRU*-Thompson model. It requires a file detailing all the runs to be made as well as a soil inputs file and a planting details file. The program sequentially generates the model input menu required to run the model. Soil and planting details are read in from the accompanying files. After each menu has been generated the model is run and the yield output appended to a results file. All 11 cycles practised in the district were modelled. A separate program had been written previously in the field scale study which weights yields from the various cycles to give a mean farm yield. These cycles are weighted according to the mean proportions that each cycle contributes by area to the MSA production (Table 1).

[1]RE Schulze and M Maharaj, Dept of Agricultural Engineering, University of Natal, 1997

[2]QV Mann, JH Meyer and DB Hellmann, SASA Experiment Station, Mount Edgecombe, 1997

The weighting program allows for any number of the 11 cycles to be considered, simply by giving a weighting of zero to those which will not be considered. The proportions of the remaining cycles are then recalculated. A flexible system that can account for various combinations of cycles is thus maintained.

Results

Preparation of model inputs

Climate data files were developed for all 93 farms in the MSA. Climate data obtained from the various sources had to be checked first to ensure that data did not fall outside reasonable ranges. In some cases days or months were missing and had to be inserted in the records with a corresponding missing data code. Up to six rainfall stations were required in some cases to patch rainfall records completely. Where necessary, stations outside of the MSA were used for this purpose. The seven driver stations were then associated with surrounding farms and their data adjusted. Temperature data were patched without difficulty.

The generation of *ACRU* soil inputs from Land Type information was completed for the 93 farms. The grid coverage of slope position required for development of parent material derived soil inputs, will be developed shortly.

MSA model runs

Results of modelling the MSA using the *ACRU*-Thompson model are preliminary. They have not been analysed beyond a simple comparison of simulated and observed MSA average cane yield over eight years. This comparison is illustrated in Figure 3.

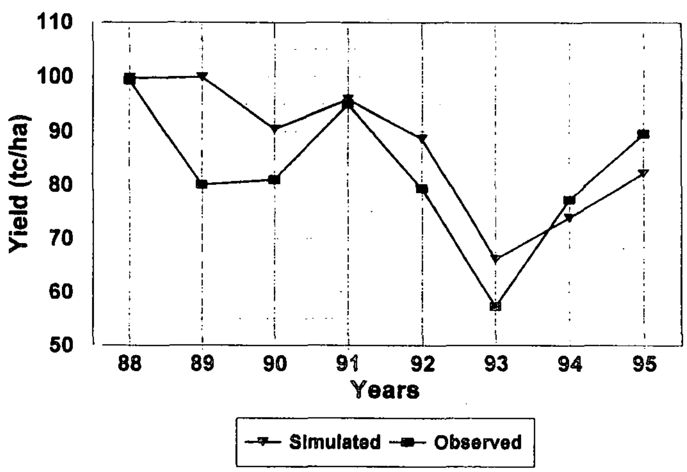

Figure 3. Simulated (*ACRU*-Thompson model) and observed average cane yields for the Eston mill supply area, 1988-1995.

The indication from the graph is that the model has over-simulated slightly in most years except for 1989 where the oversimulation is more marked and for 1994 and 1995 where the model has undersimulated. It should be noted that a number of observed data points were missing or had to be rejected on the grounds of being poor data in the earlier years of simulation. Corresponding simulated data were rejected when averaging model output to ensure an equal number of simulated and observed data points. The average difference between simulated and observed yields can be expressed through the root mean square error (RMSE) and was found to be 9,498 tc/ha. The square of the RMSE, mean square error (MSE), is thus 90,206. The systematic and unsystematic proportions of MSE can be calculated with systematic error reflecting the differences between simulated and observed values and unsystematic error reflecting variation within simulated yields. The proportion of systematic error was calculated as 0,38 while the unsystematic error proportion was 0,62. This represents a fairly large systematic error proportion and would be as a result of the general over-simulation. The Willmott (1981) index of agreement (d) between simulated and observed data is 0,836 (d=1 would indicate perfect agreement and d=0, no agreement). The RMSE, systematic and unsystematic proportions of MSE and d, quantify the differences between simulated and observed values. The Pearson correlation coefficient (r) reflects the degree of association between simulated and observed data. The conservation of upward and downward trends between simulated and observed data is thus indicated by the coefficient. The index, like d, can lie in value between zero and one, and its value was computed as being 0,760 indicating a fair correlation.

Discussion

Preliminary *ACRU*-Thompson model results are encouraging although further analysis is required to verify the suitability of the model for forecasting purposes. The general over-simulation by the model may be ascribed to management considerations. The Thompson (1976) model was originally developed using data from experimental plots where management of weeds, soils, drainage and other factors would have been thorough. In practise a farmer cannot usually attain these levels of management consistently and must also contend with factors such as lodging, frost and losses of cane between the field and mill. If it were pertinent management factors could be applied to simulated yields to improve results. Management factors have been suggested by Smith (1992) and may be a useful guide in this regard. Values of 1,0, 0,9, 0,8, 0,7 and 0,6 were suggested for experimental, excellent, very good, good and average management conditions respectively. A further source of error that could explain the systematic error in simulation, is the quality of the observed yield data. The mass of cane is accurately measured at the mill; however, the area under cane is estimated by the individual farmers. These estimates are known to be a source of inaccuracy in the observed yield data.

Further analysis of the results presented in this paper, could include graphing scatter plots of simulated versus observed yield at farm scale. This would indicate the accuracy of the model at this scale and would give greater confidence in the averaged MSA results. This would need to be done over all the years to check the model over different inter-annual climate patterns. Another analysis would be to calculate residuals between simulated and observed yield at farm scale and to map these to indicate any particular farms or subareas where simulations are poor. If there are such areas then investigations would have to be made to find the cause of these poor simulations. Plotting of observed yields over a number of years and the resulting trends may help to eliminate instances where poor observed data are the reason for simulations not matching the observed data. A further useful analysis would be to check how many cycles need to be simulated in order to accurately predict the farm and MSA yield. All 11 cycles have been simulated in this exercise, however, simulation of a fewer number of cycles might be adequate particularly as some of the cycles contribute a very small proportion to overall production. Some of the cycles also result in very similar yields (Figure 1) and the number of cycles could thus be reduced. It is intuitively more likely that at MSA scale the influence of cycles will be less pronounced.

Conclusions

The focus of this paper has been to describe the methodologies that are being used to generate model inputs for simulation of sugarcane yields at the farm and MSA scale. These inputs relate to climate and soils and their generation has been completed successfully. Preliminary results have also been presented of *ACRU*-Thompson model runs. These results are encouraging, although areas where further analyses would be required have been identified. A systematic oversimulation by the model has been detected and the use of management factors has been suggested as a possible solution to improving simulations. It would appear that the *ACRU*-Thompson model, when using input data generated as in this study, may be adequate for simulating historical cane yields and thus be a suitable model for combining with seasonal climate forecasts for the purposes of yield forecasting. The *ACRU*-Thompson model may not be the only suitable model and investigations should be expanded to include the use of currently employed industry models as well as the more complex CANEGRO/DSSAT growth model ([3]personal communication). The accuracies of these models and any benefits they may have should be weighed up against each other. The methodology of using soil parent materials to derive soil input parameters should be evaluated in future, especially if modelling is carried out at a similar scale in other cane growing areas.

It is important that any model used is able to accurately represent historic yields, as the incorporation of climate forecasts brings in a further element of uncertainty. Yield forecasts cannot be expected to be accurate if the yield model is incapable of accurately simulating historic yields. When performing historical model verifications, the accuracy of the observed yield database should also be questioned. It is known that estimates of area under cane are the main source of inaccuracies in the observed yield database used in this study.

[3]NG Inman-Bamber, CSIRO, Davies Lab, Queensland, Australia; and GA Kiker, Dept of Agricultural Engineering, University of Natal, 1997

Acknowledgements

The authors wish to thank the following institutions and individuals:

The SASA Experiment Station for financial support of this collaborative research project and for supplying a variety of data

Mr DB Hellmann, SASEX Midlands South Extension Officer for many useful discussions, during which he made available his extensive knowledge of the Eston MSA

The Computing Centre for Water Research, for making available climate data and assisting in formatting of data

The ISCW for the sale of climate and Land Type data

Messrs QV Mann, JH Meyer and DB Hellmann for developing the working rules for soil type and depth prediction from soil parent material and other factors

The Eston mill for supplying an electronic version of the SASA observed yield database

Mrs M Maharaj for performing temperature patching

Mr GA Kiker for crop modelling related discussions

Messrs B Howe and M Horan for assistance with GIS.

REFERENCES

Dent, MC, Lynch, SD and Schulze, RE (1989). Mapping mean annual and other rainfall statistics over southern Africa. Water Research Commission, Pretoria, Report 109/1/89. 198 pp.

Hellmann, DB (1993). The use of FRS data to interpret the effect of different growth cycles on the yield performance of variety N12. *Proc S Afr Sug Technol Ass* 67: 88-93.

Hughes, AD (1992). Sugarcane yield simulation with the ACRU modelling system. MSc Thesis. Department of Agricultural Engineering, University of Natal, Pietermaritzburg.

Landman, WA (1997). Bulletin No. 97/08. Bulletin Research Group for Seasonal Climate Studies, South African Weather Bureau, Pretoria.

Pike, A and Schulze, RE (1995). AUTOSOIL Version 3: A Soils Decision Support System for South African Soils. Department of Agricultural Engineering, University of Natal, Pietermaritzburg.

Smith, JMB (1992). Sugar yield estimation. In: Crop, Pasture and Timber Yield Estimate Index. Cedara Report No. N/A/94/4.

Schulze, RE (1995). Hydrology and Agrohydrology: A Text to Accompany the ACRU 3.00 Agrohydrological Modelling System. Water Research Commission, Pretoria, Report TT69/95.

Schulze, RE, Hutson, JL and Cass, A (1985). Hydrological characteristics and properties of soils in southern Africa II: Soil water retention models. *Water SA* 11: 129-136.

Thompson, GD (1976). Water use by sugarcane. *S Afr Sug J* 60: 593-600 and 627-635.

Willmott, CJ (1981). On the validation of models. *Phys Geog* 2(2): 184-194.