INFLUENCE OF CLIMATE DATA QUALITY ON THE ACCURACY OF SOUTH AFRICAN SUGARCANE YIELD ESTIMATES

SITHOLE P[1,2], SINGELS A[1] AND WALKER S[2]

[1]*South African Sugarcane Research Institute, P/Bag X02 Mount Edgecombe, 4300, South Africa*
[2]*University of Free State, Faculty of Natural and Agricultural Sciences, Dept Soil, Crop and Climate Sciences, Box 339, Bloemfontein, 9300, South Africa*
phillemon.sithole@sugar.org.za abraham.singels@sugar.org.za walkers@ufs.ac.za

Abstract

The Canesim model uses climate data to estimate sugarcane production at homogeneous climate zone (HCZ), mill and industry levels in the South African sugarcane belt. This is important for scheduling harvest and mill operations and for marketing purposes. The objective of this study was to investigate the link between the quality of climate data input and the accuracy of yield estimates from two versions of the model, namely a simple version driven primarily by evapotranspiration (ET) and a more sophisticated version driven primarily by solar radiation (Srad). A second objective was to test whether improvements in the estimation of solar radiation from sunshine duration would improve the accuracy of yield estimates.

The quality of Srad and ET data for the period 1980 to 2002 was quantified based on whether it was measured or estimated, and on the proximity of recording site to application site. Two types of Srad estimates were evaluated, namely a previously calibrated industry standard and a method calibrated for cloudiness and site. The accuracy of yield estimates was determined by comparing estimates to detrended actual yield data at mill and industry levels for the period 1980 to 2002.

The study found that the accuracy of ET-based yield estimates was better than that of pre-recalibration Srad-based estimates (industry error of 4.6% compared with 5.6%), corresponding to the differences in quality between ET and Srad data. However, it could not be demonstrated that the quality of climate data influenced the accuracy of yield estimates. The recalibration of Srad estimates reduced the error of Srad-based industry yield estimates by 0.9%. It is recommended that the recalibrated Srad estimations and the two versions of the model be used in combination to improve the accuracy of the Canesim sugarcane production forecasts for the South African sugar industry.

Keywords: yield estimate, climate data, quality, accuracy, solar radiation, evapotranspiration

Introduction

The South African Sugarcane Research Institute (SASRI) uses the Canesim crop forecasting system (Singels and Bezuidenhout, 2002) to produce a sugarcane yield forecast for the South African (SA) sugar industry. The forecast, for the forthcoming milling season, is important for decision making purposes in areas such as cane haulage scheduling, mill operations and marketing of the sugar. The Canesim model uses daily inputs of climate data (past and likely future scenarios) to simulate soil water balance and sugarcane crop growth and development. Two versions of the Canesim model are used. In one version, sugarcane yield is primarily a function of crop evapotranspiration (ET) (Singels *et al.*, 1999), and in the other version sugarcane yield is primarily a function of intercepted solar radiation (Srad) (Singels and Bezuidenhout, 2002).

The model uses climate data recorded at manual and automatic weather stations situated in the SA sugarcane belt. Until about 1997, manual weather stations (MWSs) were used for recording rainfall, wet and dry bulb temperature, sunshine duration and wind run. From these records, Srad was estimated from sunshine duration. On days when no data were recorded (more prevalent in the 1990s when manual recordings over weekends were phased out), the missing climate data were estimated using long-term mean values and on days when only sunshine duration data were missing, Srad was estimated from temperature and rainfall (Hunt *et al.*, 1998) or temperature alone (Clemence, 1992).

For the Srad estimates to be accurate, calibration for local conditions is vital (Allen *et al.*, 1998). Sithole *et al.* (2008) noted significant differences between pyranometer measured Srad and Srad estimates based on sunshine duration. The accuracy of estimates varied from one site to another, with time of year and with different levels of cloud cover. This indicated a need for calibration, for local conditions, of the coefficients in the Angström-Prescott (AP) equation (Prescott, 1940; eq. 7) used for estimating daily Srad from sunshine duration.

Since 1997, most of the MWSs were converted to automatic weather stations (AWSs) and this allowed more variables, including Srad, to be measured directly. AWSs also reduced periods of missing data and thus improved the climate data quality. However, the distribution of the weather stations across the sugarcane belt, as well as at mill and homogeneous climate zone (HCZ) levels (Bezuidenhout and Singels, 2007a), is not uniform (Figure 1) and some HCZs have better AWS coverage than others. This results in differences in the quality of the climate data used for crop forecasting purposes among HCZs and also from one season to another.

Bezuidenhout and Singels (2007b) evaluated the accuracy of yield estimates from the ET version of the Canesim model, and noted differences in the model's predictive accuracy from one mill to another and also from one season to another at mill and industry levels. The aim of this investigation was to establish how the quality of the ET and Srad data (as determined by whether it was measured or estimated and proximity of site of measurement to the area of crop growth) influences the accuracy of the estimated sugarcane yields from the two versions of the model. A second objective was to determine the impact of improvements in Srad estimation on the accuracy of yield estimates. This information will indicate the most appropriate modification to the methodology for forecasting SA sugarcane production.

Methods

The Canesim crop forecasting system and climate data were used to estimate sugarcane yields for the period 1980-2002 for 15 mill areas, covering 48 homogeneous climate zones (HCZs) in the South African sugarcane belt (see Figure 1 for spatial distribution of HCZs and weather stations). The two versions of the Canesim model, one driven primarily by crop evapotranspiration (ET) and the other driven primarily by intercepted solar radiation (Srad), were used. The sources (weather stations) of climate data used for the different HCZs, were taken from Bezuidenhout (2005).

Yield estimates were compared to detrended annual actual yields obtained from Bezuidenhout and Singels (2007b). The accuracy of yield estimates were then related to the quality of weather input data.

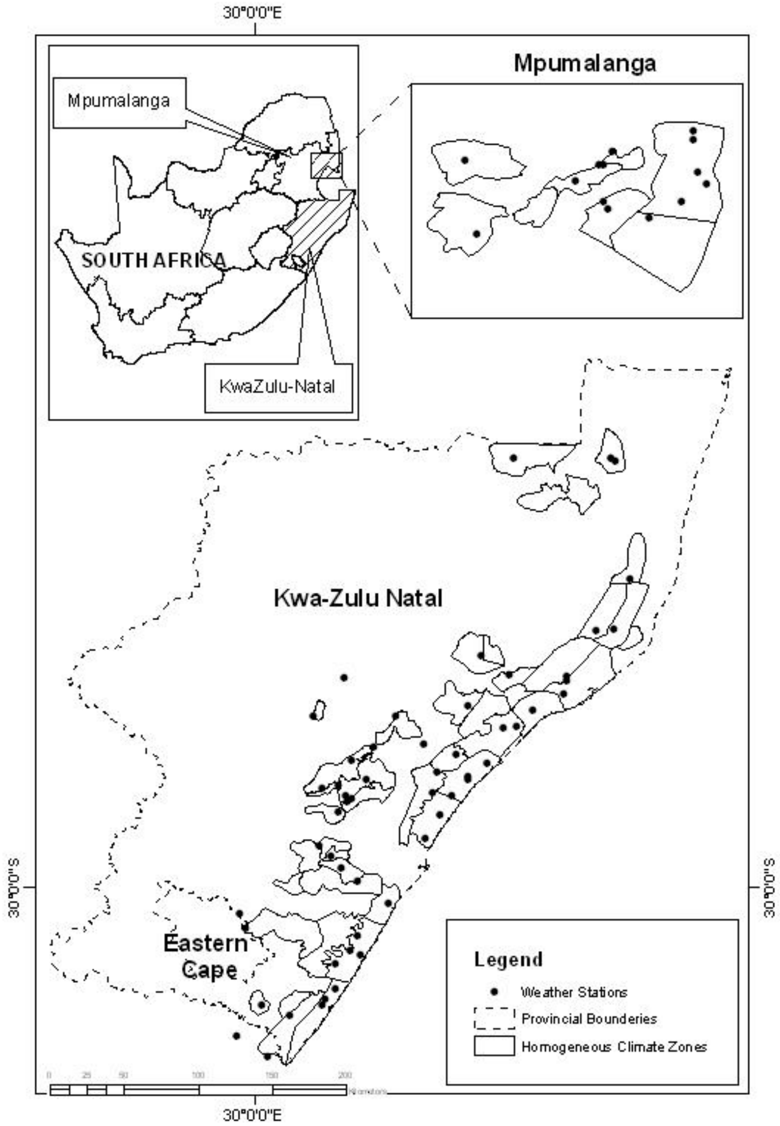

Figure 1. A map of the South African sugar belt showing the homogeneous climate zones (from Bezuidenhout, 2005) and the location of the weather stations.

Climate data quality

A quality index was formulated to rate the quality of daily Srad, maximum and minimum air temperature (Tmax and Tmin), relative humidity (RH) and wind speed (U) data in each HCZ. The quality rating was based on the method of measurement or estimation and spatial proximity of the weather station to the HCZ in question (Tables 1 and 2). The rating was devised to give a maximum rating of 10 to direct measurements within the HCZ. The quality index for Srad (Table 1) was devised to decrease by a single unit from one level to the next. The lowest level had a value of four, given where Srad data were estimated from long term mean values or from temperature. The quality index of the other variables (Table 2) was devised to cover the same range as that for Srad.

**Table 1. Data quality index for solar radiation.
Data quality improves with increasing index value.**

Index value	Solar radiation data source and estimation method
10	Measured by pyranometer at site in HCZ
9	Measured by pyranometer at site outside HCZ but in same mill area
8	Estimated from sunshine duration recorded at site within the HCZ
7	Estimated from sunshine duration recorded at site outside HCZ but in same mill area
6	Measured by pyranometer at site outside mill area
5	Estimated from sunshine duration record at site outside mill area
4	Estimated from long-term mean/temperature

Table 2. Data quality index for maximum and minimum air temperature, relative humidity and wind speed. The data quality improves with increasing index value.

Index value	Data source and estimation method
10	Measured at site in HCZ
9	Measured at site outside HCZ but in same mill area
8	Estimated from closely related parameter recorded at same site
6	Measured at site outside mill area
4	Estimated from long term mean or other related parameter

The daily values of the data quality index for Srad, Tmax, Tmin, RH and U in a given HCZ, (as rated in Tables 1 and 2) were used to calculate the monthly and seasonal average values of data quality for the respective variables for the HCZ. The seasonal data quality index, representing the whole crop growing period up to harvest, at HCZ level, was then used to established the mill average data quality index for the respective mill areas.

Monthly HCZ data quality index ($QI_{m,HCZ}$) was calculated as:

$$\mathrm{QI}_{\mathrm{m,HCZ}} = \frac{1}{\mathrm{d}} \sum_{i=1}^{\mathrm{d}} \mathrm{QI}_{i,\mathrm{HCZ}} \qquad (1)$$

where d is the total number of days for calender month, m, and $QI_{i,HCZ}$ is the HCZ data quality index of the climatic variable Srad, Tmax, Tmin, RH or U on i^{th} day of calendar month, m.

The $QI_{m,HCZ}$ was then used to calculate average data quality index for the growing season, s, for the HCZ ($QI_{s,HCZ}$) as:

$$QI_{s,HCZ} = \frac{1}{age}\sum_{j=1}^{age} QI_{j,HCZ} \qquad (2)$$

where age is the harvesting age of the crop (in months) and $QI_{j,HCZ}$ is the average HCZ quality index in the j^{th} month of the growing season, s (corresponding to the relevant $QI_{m,HCZ}$).

The seasonal HCZ data quality for the different zones within a mill area was weighted by area, to calculate the seasonal mill data quality index ($QI_{s,M}$) as:

$$QI_{s,M} = \sum_{k=1}^{z} \frac{a_{HCZk}}{A} \times QI_{s,HCZk} \qquad (3)$$

where z is the total number of HCZs in the mill, a_{HCZk} is the area of the k^{th} HCZ (ha) in the mill area, A is total mill area (ha) and $QI_{s,HCZk}$ is the average seasonal data quality index for the k^{th} HCZ in the mill area. All area figures were obtained from the SASRI Canesim crop forecasting system.

Crop ET values used in the model were based on the Penman-Monteith method (Monteith, 1965), as modified by McGlinchey and Inman-Bamber (1996) to calculate reference sugarcane ET (assumes full canopy cover and no water stress) from Srad, Tmax, Tmin, RH and U. The quality index of ET data was therefore determined from quality of these underlying variables, weighted according to their relative direct effect on ET.

The direct effect of each variable on ET was determined by setting each input variable at industry long-term mean (LTM) values, independently adjusting only the variable whose relative effect was being tested at four levels namely 50, 75, 125 and 150% of the LTM. These levels were adequate to represent the range of conditions normally experienced in the sugarcane belt and thus excluded extreme conditions. The absolute change in ET (from the LTM baseline) was noted at each of the four levels for each input variable. Tmax and Tmin were tested simultaneously since their mean (T) is used in ET calculation; hence the relative effects of four variables Srad, T, RH and U were calculated.

The relative effect (RE) of each variable (v) on ET at each of the four levels (L) was calculated as:

$$RE_{v,L} = \left(\frac{\Delta ET_{v,L}}{\sum_{v=1}^{v=4} \Delta ET_{v,L}}\right) \qquad (4)$$

where $RE_{v,L}$ is the relative effect of variable v at level L, $\Delta ET_{v,L}$ is the absolute change in ET (from the baseline value) due to a change in variable v from the LTM value to the value at level L.

The overall relative effect (RE_v) of each variable on ET was calculated as:

$$RE_v = \frac{1}{4}\sum_{L=1}^{4} RE_{v,L} \qquad (5)$$

The calculated values (Table 3) compared closely with those of El-Bably (2003) who used path coefficient analysis. However, the method of ET calculation used by El-Bably (2003), which included sunshine duration and cloudiness, was different from the one used in this study and that might explain some of the differences in relative effects, especially for temperature and wind speed.

Table 3. The overall percentage of relative effect (RE) of climatic variables on evapotranspiration. Values in brackets are from El-Bably (2003).

Climatic variable	RE (%)
Solar radiation	32.6 (31.3)
Temperature	32.3 (24.3)
Relative humidity	22.9 (27.2)
Wind speed	12.2 (4.1)

The RE values for each variable were then used to calculate the seasonal ET data quality index for a given season and mill (QI_{ET}) from the respective data quality indices of the contributing variables (Srad, T, RH and U).

$$\mathrm{QI_{ET}} = \sum_{v=1}^{v=4} \mathrm{RE_v} \times \mathrm{QI_{v,s,M}} \qquad (6)$$

where $QI_{v,s,M}$ is the seasonal mill data quality index for variable v.

Recalibration of solar radiation estimates

The observations by Sithole *et al.* (2008) indicated a need for a recalibration of the coefficients in the AP equation (Prescott, 1940; eq. 7) for estimating Srad. For this purpose, daily measurements of sunshine duration and solar radiation were made concurrently for a period of 12 months at five sites in the South African sugarcane growing belt (Table 4). Pyranometers (LI-200S, Li-Cor, Lincoln, Nebraska, USA), connected to CR 10x or CR1000 data loggers (Campbell Scientific Inc, Logan, Utah, USA), were used for measuring daily incoming solar radiation at a height of 2 m above the ground surface. Daily sunshine duration was measured by Campbell-Stokes sunshine recorders (Nagretti & Zambra, London, UK), situated a few metres from the pyranometer, at a height of 1.5 m above ground surface. The surface was covered with short green grass, except at the Komatipoort site where irrigation was not possible and the grass cover was often dry.

Daily estimates of Srad were then calculated from the sunshine duration data using eq. 7 with AP coefficients taken as *a*=0.29 cos latitude (in radians) and *b*=0.52 after Glover and McCulloch (1958).

$$\mathrm{Srad} = \mathrm{R_a}\left[a + b\left({}^{\mathrm{n}}/_{\mathrm{N}}\right)\right] \qquad (7)$$

where Srad is the solar radiation (MJ/m^2/day), Ra is extraterrestrial solar radiation (MJ/m^2/day) calculated according to Allen *et al*. (1998), n and N are the actual sunshine duration (hours) recorded by the sunshine recorder and the maximum possible daily sunshine duration (hours) respectively, and *a* and *b* are empirically determined coefficients. The method was selected for its simplicity and the few input variables required.

Table 4. Details of solar radiation and sunshine duration recordings for the recalibration of equation 7.

Location	Latitude	Longitude	Elevation (m)	Period
Empangeni	28° 43' S	31° 53' E	102	Feb 08-Jan 09
Komatipoort	25° 33' S	31° 57' E	170	Mar 08-Feb 09
Mount Edgecombe	29° 42' S	31° 2' E	96	Nov 07-Oct 08
Pongola	27° 24' S	31° 35' E	308	Feb 08-Dec 08
Wartburg	29° 25' S	30° 31' E	1014	Apr 08-Mar 09

New AP coefficients for the five different sites, times of year and levels of cloudiness were established by regressing daily Srad/R_a values against n/N values. The accuracy of estimates was quantified using the coefficient of distribution (R^2), mean bias error (MBE) and root mean square error (RMSE), the latter two expressed as a percentage of the mean observed values.

Following the calibrations at the five sites, the AP coefficients were accordingly adjusted for different levels of cloud cover and for the different HCZs according to their proximity to the five sites of calibration.

Evaluating accuracy of Canesim yield estimations

In this study the same methods of determining the accuracy of yield estimates as explained in Bezuidenhout and Singels (2007b) were followed. They referred to the yield estimation error as 'system error' (ε_{sys}), defined as the root mean (for the 22 seasons from 1980 to 2002) square error (difference between bias corrected yield estimate and actual yields for each year at mill and industry levels), expressed as a percentage of the mean actual yield.

In order to determine the accuracy of yield estimates in a given season, the difference between the bias corrected yield estimate and the actual yield was calculated for that season at mill and industry levels and expressed as a percentage of the actual yield. This was called the 'yield deviation' and was used to relate the variability of the model accuracy to the variability of ET and Srad data quality over time.

The accuracy of Canesim yield estimates after the recalibration of AP coefficients was also determined.

Results and Discussion

Results are presented and discussed, addressing (i) the quality of climate input data, (ii) the recalibration of solar radiation estimations, and (iii) the impact of these on the accuracy of yield estimates.

Climate data quality

Quality of ET (QI_{ET}) and Srad (QI_{Srad}) data for the different mills varied little over time from 1980 to 1997 (Figure 2). From 1997, MWSs were being replaced by AWSs, a transition that affected data quality negatively for many mills. Initial teething problems with newly deployed AWSs, such as power supply problems due to theft of solar panels, sometimes led to faulty or missing data. After the transition, data quality improved for the Amatikulu, Eston, Maidstone, Malelane, Noodsberg and Union Co-op mill areas. For Entumeni and Sezela data quality remained constant from 1980 to 2002, because the same MWSs were used throughout the period with minimum disruptions to data recordings. In Umfolozi data quality declined sharply from 1992, when the recording of sunshine duration ceased at the only available MWS in the mill area. Data quality remained low until 1996 when it improved due to the deployment of AWSs in the area. Data quality in Felixton was low from 1980 to 1997 because climate data were mostly estimated from weather stations in neighbouring mill areas until the introduction of an AWS in the mill area in 1998. For Darnall and Gledhow, the premature closure of some MWSs before the deployment of AWSs led to reduced data quality in the post 1997 period.

Industry average QI_{ET} and QI_{Srad} varied little until about 1997, when QI_{ET} declined slightly and QI_{Srad} increased gradually, due to the increased incidence of direct measurement of Srad through AWS pyranometers.

QI_{ET} was consistently higher than QI_{Srad} for all mills and seasons, expect for Komati in 2002. This was because direct measurements of temperature, wind and humidity used for calculating ET were available, while Srad was mostly derived from sunshine duration. Furthermore, in some cases, sunshine duration recorders were stolen and never replaced thereafter, causing frequent missing data for this variable.

At HCZ level (not shown), 32 out of the 48 had LTM data QI values of seven or better for ET and Srad, indicating acceptable data quality. Eight HCZ had LTM QI values of 5 or below, indicating poor data quality. These zones fell in the Amatikulu, Noodsberg and Eston mill areas (two HCZs each), and the Gledhow and Umzimkulu mill areas (one HCZ each).

Recalibrated solar radiation estimates

The pre- and post-recalibration Angström-Prescott (AP) coefficients were significantly different between cloudy and clear-sky conditions at all sites (Table 5). The values of *a* were lower for cloudy days than for clear days while the reverse was true for *b* values. The AP coefficients also differed significantly between some sites. Differences in the coefficients between summer and winter were not significant at any of the sites.

Hussain *et al*. (1999) and Rehman and Halawani (1997) have shown that that AP coefficients are affected by site specific factors such as latitude, altitude and mean solar elevation angle, as well as by atmospheric conditions (e.g. type and thickness of clouds and humidity). It can therefore be expected that the coefficients would have different values for the different sites and levels of cloudiness investigated here.

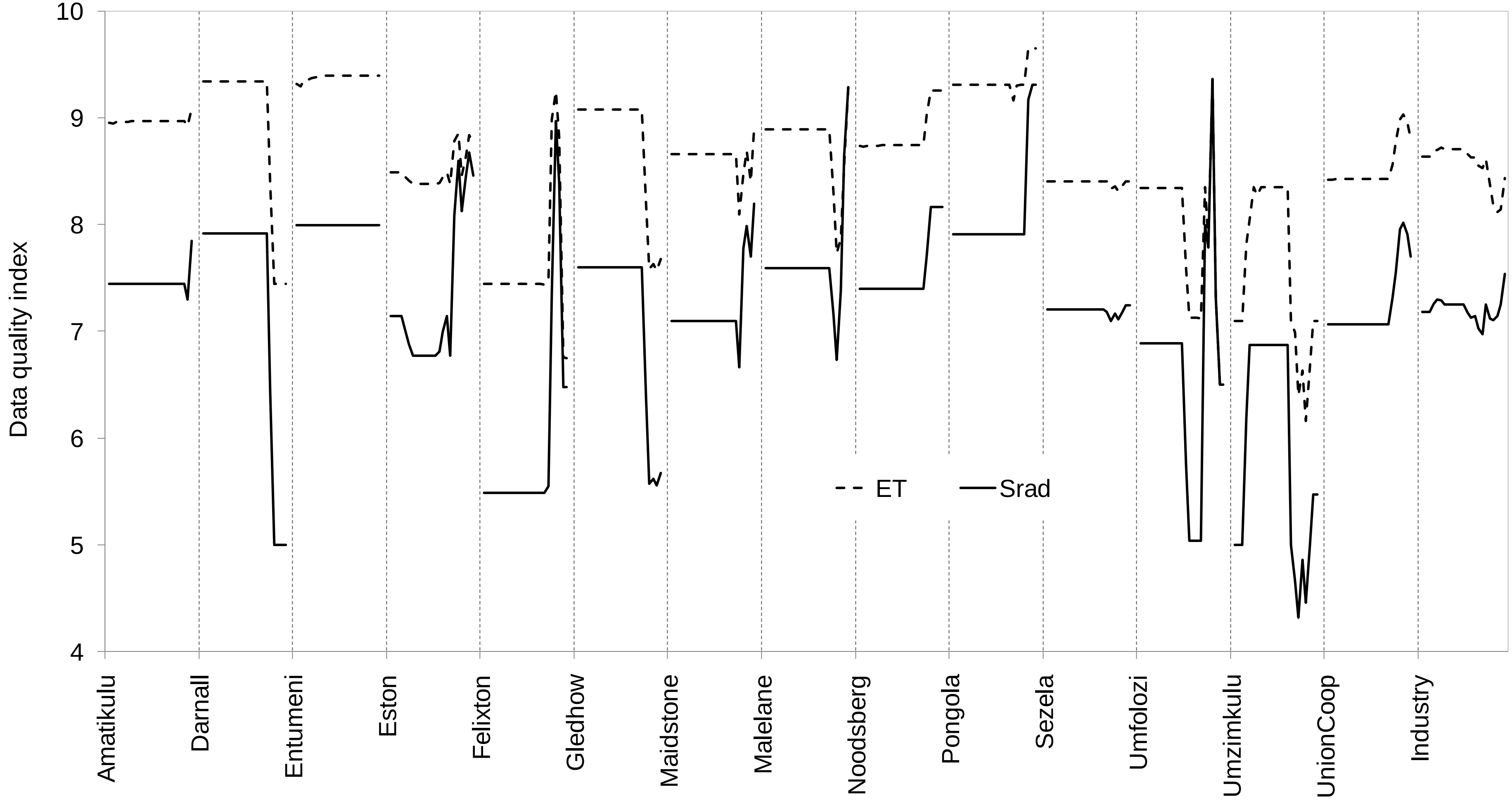

Figure 2. Year to year variation in the data quality indices of evapotranspiration (QI_ET) and solar radiation (QI_Srad) from different mills. For each mill the data for the period 1980-2002 are shown. (Data for the Komati mill are not shown because the data series started only in 1994.)

Table 5. Pre- and post-calibration AP coefficients and the associated accuracy of solar radiation (Srad) estimates under cloudy (n/N ≤ 0.25) and clear sky conditions (n/N > 0.25). Accuracy is quantified in terms of R^2 obtained by regressing estimated Srad against measured Srad, mean bias error (MBE) and root mean square error (RMSE). Common subscripts denote non-significance of differences in *a* and *b* values between sites.

Site	Cloudy conditions					Clear sky conditions				
	a	*b*	R^2	MBE (%)	RMSE (%)	*a*	*b*	R^2	MBE (%)	RMSE (%)
Pre-calibration										
Empangeni	0.254	0.52	0.58	22.9	31.2	0.254	0.52	0.94	5.9	9.1
Komatipoort	0.262	0.52	0.42	21.1	37.2	0.262	0.52	0.86	7.2	12.8
Mt Edgecombe	0.252	0.52	0.66	27.3	40.2	0.252	0.52	0.91	9.2	13.5
Pongola	0.257	0.52	0.41	27.3	44.5	0.257	0.52	0.89	6.8	12.1
Wartburg	0.253	0.52	0.67	40.1	49.1	0.253	0.52	0.93	4.4	8.5
Pooled (5 sites)	–	–	0.56	27.5	41.0	–	–	0.91	6.8	11.4
Post calibration										
Empangeni	0.187_r	0.728_u	0.64	0.4	23.5	0.228_r	0.508_u	0.94	0.0	6.8
Komatipoort	0.203_s	0.553_v	0.39	-1.7	31.6	0.272_s	0.444_v	0.86	-0.4	10.4
Mt Edgecombe	0.158_t	0.865_x	0.70	-1.0	27.4	0.270_s	0.406_x	0.93	-1.3	10.8
Pongola	0.183_r	0.644_z	0.42	0.4	34.7	0.298_t	0.394_x	0.90	3.6	11.9
Wartburg	0.151_t	0.834_x	0.77	0.1	23.9	0.227_r	0.518_u	0.93	-0.2	7.3
Pooled (5 sites)	0.174	0.738	0.60	-0.4	30.4	0.259	0.454	0.91	0.4	9.7

When the calibrated AP coefficients were applied, there were improvements in estimation accuracy for all sites (MBE and RMSE in Table 5), as could be expected. This was largely due to the removal of the bias for cloudy day estimates using pre-calibrated coefficients. Slight improvements in accuracy were also noted at some sites under clear sky conditions, a sign that the pre-calibration AP coefficient values may be good enough on days with limited or no cloud cover.

The *a* and *b* values derived by Glover and McCulloch (1958) were based on monthly mean data, possibly underestimating the effect of cloudiness on cloudy days. This could explain the overestimation of Srad under cloudy conditions observed here. It is therefore preferable to use daily data for calibration of AP coefficients.

There was a strong correlation between the calibrated AP coefficients for cloudy conditions and latitude of the site with R^2 values of 0.77 and 0.98 for *a* and *b* respectively. The correlation was weak for clear conditions, with R^2 values of 0.245 and 0.077 for *a* and *b* respectively. Glover and McCulloch (1958) also found a relationship between coefficient *a* and latitude (but not between coefficient *b* and latitude) although they did not distinguish between cloudy and clear sky conditions. This relationship as well as that between AP coefficients and other factors (such as altitude, elevation and mean humidity), will be investigated further in the hope of obtaining a more generic and widely applicable method for estimating Srad on cloudy and/or clear days.

In the absence of a generic method for estimating Srad, it was decided to (1) derive AP coefficients that could be applied across the entire industry by calibrating using data pooled from all calibration sites, and (2) allocate AP coefficients for each HCZ according to its proximity to the five calibration sites. The impact of both these methods of estimating Srad on the accuracy of yield estimates was assessed.

Table 6. Recalibrated values of coefficients *a* and *b* to be used for estimating Srad in different homogeneous climate zones (HCZs).

HCZ	Site of calibration	Clear		Cloudy	
		a	*b*	*a*	*B*
All	All 5 sites pooled	0.259	0.454	0.174	0.738
1-8	Komatipoort	0.272	0.444	0.203	0.553
9-11	Pongola	0.298	0.394	0.183	0.644
12-23, 25	Empangeni	0.228	0.508	0.187	0.728
24, 26, 30, 32, 34-40	Mount Edgecombe	0.227	0.518	0.151	0.834
29, 31, 33, 41-48	Wartburg	0.270	0.406	0.158	0.865

Accuracy of Canesim yield estimates in relation to climatic data quality

Generally, yield deviations varied significantly between seasons, while the quality of ET and Srad data remained unchanged (for example Entumeni for the entire period and Eston and Maidstone up to 1997) (Figure 3). After 1997, the introduction of AWSs improved the data quality for Eston and Maidstone, although yield deviations did not decrease as might have been expected. At an industry level, there was less variability in the yield deviations with time than at mill level. Again, the variability in yield estimation accuracy did not relate to variability in the data quality indices.

These results suggest that factors which affect yield, other than climate, varied significantly with time and space. These factors may include pest, disease and weed pressures, soil fertility, and crop damage from extreme events such as frost or storms. The quality of rainfall and soil input data could also have played a role.

The LTM accuracy of yield estimations indicated that yield estimates based on ET (Y_{ET}) had a lower ε_{sys} at 11 of the 15 mills when compared with yield estimates based on Srad (Y_{Srad}) (Table 7). At the industry level Y_{ET} was more accurate than Y_{Srad} with ε_{sys} of 4.63% compared with 5.58% for Y_{Srad}.

The observation that Y_{ET} was generally more accurate than Y_{Srad} agrees with the observation that QI_{ET} was higher than QI_{Srad}. However, there seems to be no strong link between the QI_{ET} and QI_{Srad} on one hand, and the accuracy of their respective yield estimates on the other (Table 7). Two examples to illustrate the poor correlation between QI and yield estimation accuracy are Umzimkulu (with poor data quality but relatively good yield estimation accuracy) and Entumeni (with good data quality but poor yield estimation accuracy). A regression analysis between pre-calibration ε_{sys} and QI at mill level produced R^2 values of 0.123 and 0.038 for Y_{ET} and Y_{Srad} respectively, confirming the poor link between the two entities.

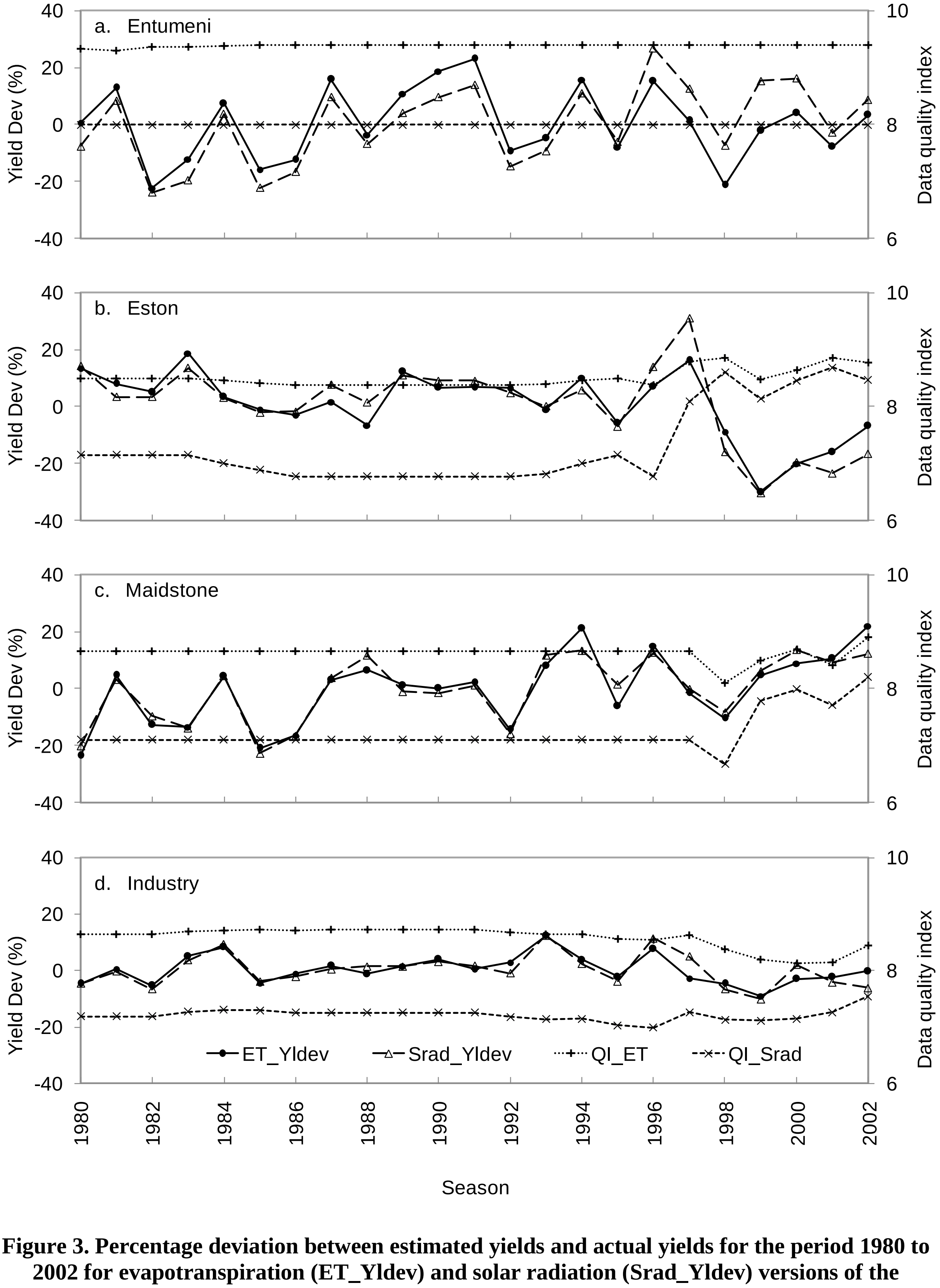

Figure 3. Percentage deviation between estimated yields and actual yields for the period 1980 to 2002 for evapotranspiration (ET_Yldev) and solar radiation (Srad_Yldev) versions of the Canesim model for three mill areas and for the industry. The data quality indices for ET (QI_ET) and Srad (QI_Srad) over the same period are also shown.

Table 7. Average data quality (expressed as the quality index) for evapotranspiration (ET) and solar radiation (Srad) and the accuracy of ET based (Y_{ET}) and Srad based (Y_{Srad}) yield estimates. Pre-cal and Post-cal values are, respectively, values before and after recalibration of Angström-Prescott coefficients.

Mill area	Data quality		Yield estimation error (%)					
			Y_{ET}			Y_{Srad}		
	ET	Srad	Pre-cal	Post-cal		Pre-cal	Post-cal	
				Pooled	By site		Pooled	By site
Amatikulu	9.0	7.5	7.7	7.9	7.9	6.8	7.3	7.3
Darnall	9.0	7.3	9.3	8.6	8.4	8.6	8.8	8.5
Entumeni	9.4	8.0	13.4	12.9	13.0	13.8	13.6	13.6
Eston	8.5	7.3	11.8	12.5	12.5	14.5	10.3	10.4
Felixton	7.6	5.9	12.8	13.3	13.3	11.1	11.1	11.1
Gledhow	8.8	7.2	11.5	10.6	10.3	11.9	11.9	11.7
Komati*	7.3	6.7	11.1	7.2	7.4	11.8	8.1	8.5
Maidstone	8.6	7.2	11.9	11.4	11.3	10.8	10.2	10.0
Malalane	8.8	7.7	10.6	10.6	10.7	11.7	11	11.2
Noodsberg	8.8	7.6	10.2	10.4	10.4	11.8	11.5	11.6
Pongola	9.4	8.1	7.4	7.1	7.2	7.6	7.5	7.5
Sezela	8.4	7.2	11.3	10.4	10.2	12.5	12.4	12.3
Umfolozi	7.9	6.7	12.9	12.9	13.0	14.3	13.4	13.6
Umzimkulu	7.6	5.9	10	10.2	10.2	11.5	11.3	11.3
Union Coop	8.5	7.2	9.3	9.3	9.3	9.9	9.3	9.3
Averages	8.6	7.2	10.8	10.4	10.3	10.9	10.4	10.4
Industry			**4.63**	**4.66**	**4.65**	**5.58**	**4.71**	**4.66**

*Data is from 1994 to 2002

Results suggest that spatial and temporal variations in yield estimation accuracy at mill level are influenced more by factors other than climate (ET and Srad) data quality. Possible factors are the quality of other model input data such as rainfall, soil property, crop management and irrigation data. Yield estimates also do not account for sub-optimal agronomic conditions such as weed competition, pest and disease damage, low nutritional status, ratoon yield decline and damage from extreme events such as frost and storms. Another reason for the poor link between ε_{sys} and QI could be the relatively coarse spatial resolution used (15 mills). An investigation at HCZ level (48 zones) could reveal more information because less spatial variation would be masked. However, actual yield was not available at this level for most of the industry.

Climate data from AWSs have the potential to improve the accuracy of estimated yields compared with manually recorded data, because they record Srad directly. However, that could not be verified in this study because transformed actual yield data at mill level was not available beyond 2002, when the drive to deploy AWSs gained momentum. (The number of AWSs increased from 15 in 2002 to 45 in 2010.) Transforming the yield data to uniform conditions is necessary to detrend the data (Bezuidenhout and Singels, 2007b) to account for various underlying spatial and temporal trends; for example, due to changes over time in irrigated areas and harvest age. A post-2002 analysis will provide a better basis for assessing the impact of improved climate data quality through the deployment of AWSs, on the accuracy of yield estimates.

Effect of recalibrated solar radiation estimates
Recalibration of AP coefficients had an overall positive effect on Y_{Srad}, but not on Y_{ET} (Table 7). The Y_{Srad} showed reduced yield estimation error (ε_{sys}) at 11 of the 15 mills, while the ET version showed reduced ε_{sys} at only seven mills. The best improvement in accuracy was for the Eston mill, where the Y_{Srad} error was reduced by 4.2%, followed by Komati with a reduction of 3.6%. The Y_{ET} showed the best improvement in accuracy for the Komati mill (3.9%). At the industry level, ε_{sys} increased slightly for Y_{ET} while it was reduced by 0.9% for Y_{Srad}, almost eliminating the differences in accuracy between the two types of yield estimates.

The differences in response to the recalibration of AP coefficients by the two versions of the model can be explained by the fact that Y_{Srad} is more directly affected by Srad data than Y_{ET}. Recalibration improved only the quality of Srad data. The quality of other climatic variables influencing ET data quality was not improved, and these can mitigate the effects of improved Srad data quality.

Site-specific calibrations of AP coefficients resulted in slightly more accurate industry yield estimates than the pooled calibrations. At mill level, no calibration method demonstrated consistent superiority over the other in terms of yield estimation accuracy. The relative simplicity of pooled calibration makes it an attractive option for operational implementation.

Conclusions

The main findings from this investigation were:

- There was little variability in mill average climate data quality except for the period when manually operated weather stations were replaced with automatic weather stations (1997-2002). For the majority of the mill areas, data quality was eventually improved after a troublesome transitional period.

- The quality of evapotranspiration data was consistently better than that of solar radiation for all mill areas, as solar radiation was mostly estimated and not directly measured. Other climate variables affecting evapotranspiration were directly measured.

- The accuracy of evapotranspiration-based yield estimates was better than that of pre-calibrated radiation-based estimates for 11 of the 15 mills. At an industry level, the former was also more accurate than the latter (error of 4.63% compared with 5.58%). Although good quality climate data is vital for accurate yield estimates, a strong link between data quality and yield estimation accuracy at a mill level could not be established, presumably because of the stronger influence of other factors such as variation in agronomic conditions (not taken into account by the forecasting system) and the quality of rainfall and soil input data.

- The accuracy of solar radiation estimates from sunshine duration was improved significantly by recalibrating the Angström-Prescott equation for cloudiness and site. The recalibration also improved the accuracy of radiation-based yield estimates for 11 mills, while yield estimates based on evapotranspiration improved for only seven mills. At the industry level, there was an improvement in the accuracy of radiation-based yield estimates, with the estimation error decreasing from 5.58% to 4.66%. Accuracy of evapotranspiration-based yield estimates was not improved.

For more accurate sugarcane production forecasts, it is recommended that the recalibrated solar radiation estimations be used. Radiation-based yield estimates could then be used for those mills where these estimates were more accurate than those based on evapotranspiration, such as Eston, Maidstone and Felixton. The combined use of the two versions should result in improved accuracy of the Canesim sugarcane production forecasts for South Africa.

Acknowledgements

The authors would like to acknowledge immense assistance of Aresti Paraskevopoulos in running and troubleshooting the Canesim crop forecasting system, Jarman Chetty for processing sunshine cards, Joe Govender for setting up sunshine recorders and Rod Harding for programming assistance.

REFERENCES

Allen RG, Pereira LS, Raes D and Smith M (1998). Crop evapotranspiration - Guidelines for computing crop water requirements. FAO Irrigation and Drainage Paper 56. FAO, Rome. 301 pp.

Bezuidenhout CN (2005). Development and evaluation of model-based operational yield forecasts in the South African sugar industry. PhD thesis, School of Bioresources Engineering and Environmental Hydrology, University of KwaZulu-Natal, Pietermaritzburg, South Africa. 137 pp.

Bezuidenhout CN and Singels A (2007a). Operational forecasting of South African sugarcane production: Part 1 - System description. *Agric Systems* 92: 23-38.

Bezuidenhout CN and Singels A (2007b). Operational forecasting of South African sugarcane production: Part 2 - System evaluation. *Agric Systems* 92: 39-51.

Clemence BSE (1992). An attempt at estimating solar radiation at South African sites which measure air temperature only. *S Afr J Plant Soil* 9: 40-42.

El-Bably AZ (2003). Estimation of evapotranspiration using statistical model. pp 441-449 In: Camarda D and Grassini L (Eds), *Local Resources and Global Trades: Environments and Agriculture in the Mediterranean Region*. Bari: CIHEAM-IAMB.

Hunt LA, Kuchar L and Swanton CJ (1998). Estimation of solar radiation for use in crop modeling. *Agric Forest Meteorol*, 91: 293-300.

Hussain M, Rahman L and Rahman MM (1999). Techniques to obtain improved predictions of global radiation from sunshine duration. *Renew Energy* 18: 263-75.

Glover J and McCulloch JSG (1958). The empirical relationship between radiation and hours of sunshine. *Q J R Meteorol Soc* 84: 172-175.

McGlinchey MG and Inman-Bamber NG (1996). Predicting sugarcane water use with the Penman-Monteith equation. pp 592-597 In: Camp CR, Sadler EJ and Yoder RE (Eds), *Proceedings of the*

International Conference on Evapotranspiration and Irrigation Scheduling. American Society of Agricultural Engineers, San Antonio.

Monteith JL (1965). Evaporation and the environment. pp 205-234 In: Fogg GE (Ed), *The State and Movement of Water in Living Organisms*. XIXth Symposium, Society of Experimental Biology, Swansea, Cambridge University Press.

Prescott JA (1940). Evapotranspiration from water surface in relation to solar radiation. *Trans R Soc Sci Aust* 64: 114-118.

Rehman S and Halawani TO (1997). Global solar radiation estimation. *Renewable Energy* 12: 369-385.

Singels A and Bezuidenhout CN (2002). A new method of simulating dry matter partitioning in the Canegro sugarcane model. *Field Crops Res* 78: 151-164.

Singels A, Bezuidenhout CN and Schmidt EJ (1999). Evaluating strategies for scheduling supplementary irrigation of sugarcane in South Africa. *Proc Aust Soc Sug Cane Technol* 21: 219-226.

Sithole P, Singels A and Walker S (2008). Accuracy of solar radiation estimates in the South African sugar belt. *Proc S Afr Sug Technol Ass* 81: 358-361.