

Estimating solar radiation for crop modeling using temperature data from urban and rural stations

T. Mavromatis^{1,*}, S. S. Jagtap²

¹Department of Meteorology–Climatology, School of Geology, Aristotle University of Thessaloniki, Thessaloniki 54124, Greece

²Department of Agricultural and Biological Engineering, University of Florida, Gainesville, Florida 32611, USA

ABSTRACT: Lack of site-specific global solar radiation (GSR) data is a significant impediment for most crop model applications. First, several empirical methods for estimating GSR from observed maximum and minimum temperature were evaluated using data from urban (National Climatic Data Center [NCDC] data set) and rural (Florida Automated Weather Network [FAWN] data set) sites in Florida. Next, the spatial structure of empirical model parameters was investigated and the possibility of employing spatially interpolated coefficients developed at urban or rural sites to predict solar radiation at rural locations, where most crops are grown and often crop modeling is required to address various issues, was assessed. Finally, the effects of observed and estimated GSR on simulated potential yield and evapotranspiration with CROPGRO-Peanut and CERES-Maize models at 6 crop–location combinations were evaluated. The model developed by Donatelli and Bellocchi (DB) achieved the most accurate estimations of GSR across Florida. The quality of those predictions varied with environment setting (urban-rural) and latitude. RMSE (root mean square error) between observed and fitted daily GSR were in the range of 3.1–4.1 and 3.2–4.9 MJ m⁻² d⁻¹ for FAWN and NCDC evaluation sub-sets, respectively. Radiation for FAWN sites was efficiently estimated by the coefficients interpolated from rural sites but was substantially overestimated by 17.6%, on monthly basis, when urban sites supplied the model empirical parameters. NCDC sites were site specific and suitable sources of solar radiation at urban environments only. Input of estimated with the DB model GSR into a crop model would be a realistic option at sites where only temperature and precipitation are available.

KEY WORDS: Solar radiation estimation · Air temperature · RadEst · Crop simulation models · Urban–rural stations · Florida

— Resale or republication not permitted without written consent of the publisher —

1. INTRODUCTION

The increasing interest in using crop models as research tools for annual yield estimation (Jagtap & Jones 2002), for tailoring crop management practices to expected weather (Mavromatis et al. 2002), or for yield forecasting (Jagtap et al. 2002), has led to an urgent requirement of daily weather data (i.e. global solar radiation (GSR), maximum (T_{\max}) and minimum (T_{\min}) air temperatures, and rainfall). Globally, the number of ground-level weather stations recording GSR is limited compared to the number recording air temperature and precipitation (McVicar & Jupp 1999,

Thornton & Running 1999). In Florida, nearly all of the 88 weather stations have long-term records of temperatures and rainfall, but only 7 stations have long-term records of GSR.

A number of formulae and methods have been developed to estimate GSR at instrumental sites, where it is not measured, based on other commonly measured meteorological variables. Available methods include stochastic weather generators and empirically derived relationships. Empirical approaches require either the development of equations that relate solar transmissivity to daily observations of maximum and minimum air temperature using site-specific empirical

*Email: thmavrom@geo.auth.gr

parameters (Donatelli et al. 2003) or conversion from sunshine duration (Rivington et al. 2002). Stochastic weather generators may be used for risk analysis but not for model validation and simulation analysis for a specific period as the models may not generate the data to match the actual weather at a particular time of interest (Hayhoe 2000, Liu & Scott 2001, Mavromatis & Hansen 2001). Therefore, it is not surprising that much effort has been directed at empirical methods relying on more readily available data such as temperature only (Goodwin et al. 1999, Mahmood & Hubbard 2002) or both temperature and rainfall (Hunt et al. 1998, Liu & Scott 2001). Empirical methods were found to provide more accurate estimates than weather generators (Hayhoe 1998).

To our knowledge, very few studies have addressed the impact of the environment setting of the station (inland–coastal, urban–rural) on model predictability (Liu & Scott 2001). Two data sets from Florida were used in this study, one from urban sites and another from rural, in order to investigate the predictability between rural and urban sites. Most urban stations in Florida are at airports, as their primary duty is to serve the aviation industry. Rural stations, on the other hand, are located in the midst of agricultural areas. This study was initiated to estimate daily GSR at sites where maximum and minimum temperature and rainfall data were available for crop growth simulation.

The objectives for this study were 3 fold. First, daily solar radiation was related to T_{\max} and T_{\min} at 2 subsets of geographically dispersed urban and rural sites in Florida where temperature data is available using several empirical models. Afterwards, the spatial structure of model parameters was evaluated and the possibility of using coefficients developed at rural or urban stations to predict the solar radiation at rural locations, where most crops are grown and often crop modeling is required to address various issues, was explored. Finally, the effects of measured and estimated GSR on simulated potential yield and evapotranspiration (PET) with CERES-Maize (Ritchie et al. 1998) and CROPGRO-Peanut (Boote et al. 1998) crop models at 6 crop–location combinations were investigated.

2. MATERIALS AND METHODS

2.1. Models for estimation of solar radiation from daily air temperature

Four radiation models were evaluated, contained within the RadEst global solar radiation estimation tool (beta v 3.0) (Donatelli et al. 2003), including the Bristow–Campbell (BC, Bristow & Campbell 1984),

Campbell–Donatelli (CD, Donatelli et al. 2003), Donatelli–Bellocchi (DB, Donatelli & Bellocchi 2000, Bellocchi et al. 2002), and the modular model DCBB (MDCBB, hereafter) to estimate daily solar radiation. Daily radiation (RadEst_i) is calculated as the product of the daily atmospheric solar radiation transmissivity (tt_i) and potential radiation (PotRad_i). Potential radiation outside the earth's atmosphere was expressed quantitatively as a function of latitude and day of year. In these models, tt_i is estimated using Eqs. (1) to (4), representing Models BC, CD, DB and MDCBB, respectively:

$$tt_i = \tau \left[1 - \exp\left(\frac{-b\Delta T_i^C}{\Delta T_{\text{month}}}\right) \right] \quad (1)$$

$$tt_i = \tau [1 - \exp(-b\{0.017 \times \exp[\exp(-0.053 + T_{\text{avg}})]\} \Delta T_i^2 [f_1(T_{\min})])] \quad (2)$$

$$tt_i = \tau [1 + f_2(i)] \left[1 - \exp\left(\frac{-b\Delta T_i^2}{\Delta T_{\text{week}}}\right) \right] \quad (3)$$

$$tt_i = \tau [1 + f_2(i)] \left[1 - \exp\left(\frac{-b\Delta T_i^2 f_1(T_{\min})}{\Delta T_{\text{avg}}}\right) \right] \quad (4)$$

where: τ = clear sky transmissivity, $\Delta T = (T_{\max i} - (T_{\min i} + T_{\min i+1})/2)$ (with i representing here day, week or month, as appropriate), b and C are model specific parameters, $T_{\text{avg}} = (T_{\max i} + T_{\min i})/2$, ΔT_{avg} = mobile weekly ΔT (MDCBB_m, hereafter) or fixed monthly ΔT (MDCBB_f, hereafter), $f_1(T_{\min}) = \exp(T_{\min i}/Tnc)$ and $f_2(i) = C1 \times [\sin(i \times C2 \times \pi/180) + \cos(i \times f_3(C2) \times \pi/180)]$. $T_{\max i}$ and $T_{\min i}$ are daily air maximum and minimum temperature (°C), respectively, for Day i . In the equation defining $f_1(T_{\min})$, Tnc is a model specific parameter. In the equation defining $f_2(i)$, $f_3(C2) = 1 - 1.9 \times C3 + 3.83 \times C3^2$, where $C3 = C2 - \text{integer}(C2)$. $C1$ and $C2$ are model specific parameters.

The BC model exploits the relationship between diurnal air temperature range and irradiance load to estimate the daily flux of incoming solar irradiance. This model has been used in numerous studies, and improvements have been developed over the last years (Goodin et al. 1999, McVicar & Jupp 1999). The CD model accounts for situations in which the night air temperature cooling is less than the corresponding clear day. It also accounts for the date using the air average temperature. The Campbell–Donatelli model has been implemented in the weather generators MarkSim (Jones & Thornton 2000) and ClimGen (Castellvi & Stockle 2002). The DB model differs from the previous ones by accounting for the effect of the seasonal variation of the τ and ΔT using trigonometric functions. The MDCBB model includes features of the previous 3 models allowing switching features on/off.

More specifically, the trigonometric functions that allow accounting for seasonal variation of both τ and ΔT can be excluded by setting $C1 = 0$. Two versions of the model were evaluated. In the first one, the ΔT was estimated weekly (as the average over 7 d around the current day) (MDCBB_m). In the second version the ΔT was fixed (average monthly ΔT) (MDCBB_f).

2.2. Data sets

Two data sets, one from urban, predominantly coastal sites and another from rural, primarily inland sites, were used to determine spatial-temporal stability and accuracy of model parameters (Fig. 1). Ten yr of daily weather data from National Climatic Data Center (NCDC) (1978–1987) for 6 urban sites (Tallahassee, Miami, Palm Beach, Tampa, Daytona, and Jacksonville) along the coast of Florida (Table 1, Fig. 1), and Gainesville were used. Gainesville was considered to represent urban environment because it covered the same period and for a better spatial coverage. The second data set included fifteen rural stations, spanning from 1998 through 2000, from the Florida Automated Weather Network (FAWN) which was installed in late 1997 (Table 2, Fig. 1). The urban data set includes stations with a population density of the surrounding area of more than 60000 (Census 2001). On the other hand, the majority of the FAWN stations represents regions with less 40000 population. Most importantly, most NCDC stations are located at the airports as their primary duty is to serve the aviation industry. In contrast, the rural data set includes stations placed in predominantly rural environments surrounded by cultivated areas with the task for providing growers with real time weather information such as freeze warning and irrigation advice. All stations from both data sets were low-level altitude stations (<100 m). Only 4 out of the 15

Table 1. National Climatic Data Center (NCDC) weather stations (1978 to 1987) used in the study. The model parameters for the Donatelli–Bellocchi method are also presented. *b*: temperature range factor; *C1*: seasonal variation magnitude factor; *C2*: seasonal variation profile factor

Location	Latitude (°N)	Longitude (°W)	<i>b</i>	<i>C1</i>	<i>C2</i>
Jacksonville	30.30	-81.42	0.121	-0.081	1.435
Tallahassee	30.23	-84.22	0.109	-0.067	1.329
Gainesville	29.63	-82.37	0.095	-0.035	1.534
Daytona	29.11	-81.30	0.185	-0.067	1.513
Tampa	27.58	-82.32	0.167	-0.053	1.513
Palm Beach	26.41	-80.60	0.221	-0.037	1.614
Miami	25.48	-80.16	0.219	0.019	1.001

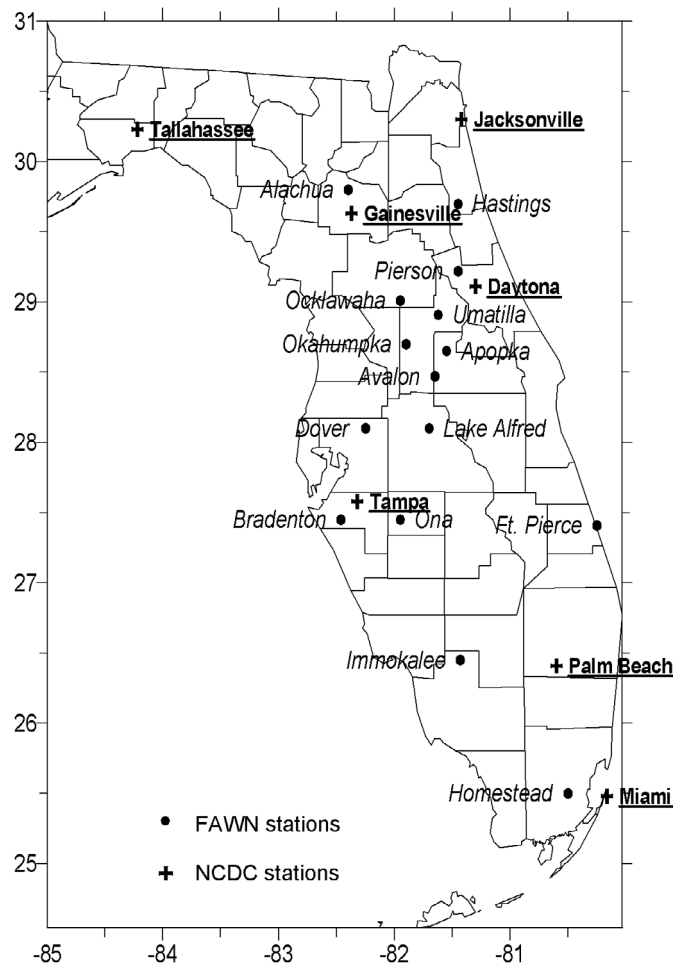


Fig. 1. Locations of National Climatic Data Center (NCDC) (underlined) and Florida Automated Weather Network (FAWN) stations in Florida, USA

FAWN sites had 3 yr of data. A 2 yr period was available for each of the remaining 11 rural stations. Dividing the data into days with and without rain did not substantially improve correlation and RMSE estimates (Boisvert et al. 1990). The seasonal subdivision of the data from 10 mid-continental locations in Kansas did not have any significant positive effect in the accuracy of radiation estimates as well (Goodwin et al. 1999). On the other hand, season influenced the model estimates over the Northern Great Plains (Mahmood & Hubbard 2002). The use of separate equations based on seasonal criteria requires an abrupt and physically unrealistic transition when the threshold dates/periods are reached. We therefore concluded that a single annual equation was preferable. Consequently, NCDC and Gainesville data sets were divided randomly into a calibration set for fitting parameters and a validation data set. The first 2 yr from 8 FAWN sites were used for fitting model

Table 2. FAWN weather stations used in the study. Calibration period and model parameters for the Donatelli–Bellocchi method are also presented. Explanations for model parameters (b , $C1$, $C2$) as in Table 1

Location	Latitude (°N)	Longitude (°W)	Data record	b	$C1$	$C2$
Calibration sub-set						
Pierson	29.22	-81.45	1999–2000	0.085	0.040	0.951
Umatilla	28.91	-81.62	1999–2000	0.085	0.015	1.093
Apopka	28.65	-81.55	1998–1999	0.107	0.015	1.791
Lake Alfred	28.10	-81.70	1998–1999	0.112	0.043	1.371
Ona	27.45	-81.95	1999–2000	0.103	-0.040	1.550
Ft. Pierce	27.41	-80.25	1999–2000	0.122	0.029	1.235
Immokalee	26.45	-81.43	1998–1999	0.103	0.029	1.405
Homestead	25.50	-80.50	1998–1999	0.105	0.059	1.401
Spatial evaluation sub-set						
Alachua	29.80	-82.40	2000			
Hastings	29.70	-81.45	2000			
Ocklawaha	29.01	-81.95	1999–2000			
Okahumpka	28.70	-81.90	1999–2000			
Avalon	28.47	-81.65	1999–2000			
Dover	28.10	-82.25	1999–2000			
Bradenton	27.45	-82.46	1999–2000			

Table 3. Crop simulation scenarios used to evaluate simulated potential yield and PET (potential yield and evapotranspiration). Mean bias errors for yield (MBE_Y) and PET (MBE_{PET}) are also given. Sources for crop simulation are: Maize: Florida Agricultural Statistics Service (2002); peanut: Mavromatis et al. (2002)

Code	Station	Year	Crop	Planting date	Cultivar	MBE_Y (%)	MBE_{PET} (%)
A	Alachua	2000	Maize	15 Apr	McCurdy	19.1	5.5
B	Hastings	2000	Maize	15 Apr	McCurdy	-6.6	-10.7
C	Bradenton	1999	Maize	15 Apr	McCurdy	26.4	11.5
D	Bradenton	2000	Maize	15 Apr	McCurdy	-7.6	-5.6
E	Ocklawaha	1999	Maize	15 Apr	McCurdy	-1.4	-1.0
F	Ocklawaha	2000	Maize	15 Apr	McCurdy	2.8	1.1
G	Alachua	2000	Peanut	15 Apr	Florunner	2.5	4.8
H	Ocklawaha	1999	Peanut	15 Apr	Florunner	-2.7	-2.0
I	Ocklawaha	2000	Peanut	15 Apr	Florunner	1.7	-0.9

parameters (Table 2). Sites–years for temporal validation were: Apopka 2000, Lake Alfred 2000, Homestead 2000, and Immokalee 2000. The station–years employed for spatial evaluation are presented in Table 2.

2.3. Model parameter fitting and spatial interpolation

Site-specific empirical parameters were developed for each model by using all the data available at calibration sites (C_s) where both air temperature and GSR measurements were available. The model parameters were fitted one at a time by employing an optimization technique that minimized different objective functions

using an iterative procedure. A clear and detailed description of the steps required to fit the parameters can be found in the manual provided with the software package and need not be repeated here. After parameter fitting, solar radiation was estimated with each model for all calibration site–years. The best model was selected by comparing the goodness of fit between measured and predicted radiation data.

Due to a strong spatial variability in input data, model parameters at the calibration sites were treated as spatial variables denoted by their geographic coordinates. This spatial variability was modeled using an inverse distance weighted method using spatial analysis in ARC-VIEW (v. 3.0) (ESRI 1996). The Inverse Distance Weighted (IDW) interpolator assumes that each input point has a local influence that diminishes with distance. It weights the points closer to the processing cell greater than those farther away. A specified number of points, or optionally all points within a specified radius, can be used to determine the output value for each location. The resulting 2-dimensional variogram was used to obtain parameter values at the evaluation sites (E_s). The model was then applied to the interpolated values of the empirical parameters and to measured T_{max} and T_{min} to calculate daily solar radiation at E_s .

2.4. Crop simulation

Solar radiation provides the energy for photosynthesis, carbohydrate partitioning and biomass growth of the individual plant components (Boote & Loomis 1991). Two crop simulation models, the CERES (Ritchie et al. 1998) for maize, and the CROPGRO (Boote et al. 1998) for peanut, were employed to simulate crop responses to historic and estimated solar radiation. Potential yield and PET were chosen among the simulation output since they represent non-water limiting situations affected by GSR and temperatures only. The Ritchie modification of Priestley–Taylor approach was adopted for estimating PET. Inputs to these models include daily weather data (minimum and maximum temperature, precipitation and solar irradiance), soil

properties, initial soil water content, cultivar characteristics, planting date, and N fertilizer management (Hunt & Boote 1998). The study considered 9 crop–site–years combinations. Representative cultivars and planting dates from published studies or researchers familiar with each region were used (Table 3). Soil water balance and N dynamics were disabled for these simulations since models assumed potential production. Crop simulations were repeated using estimated daily solar radiation for the same number of years as the observed sequences.

2.5. Analyses

The performance of the empirical models was compared with the help of descriptive statistics that indicate the degree of consistency between results based on monthly and daily observed and estimated solar irradiance. The goodness of fit was assessed by: the slope, intercept and coefficient of determination (R^2) of the regression line between estimated and measured values, mean bias error (MBE) expressed on percent basis and, the root mean squared error (RMSE). Independent *t*-tests (at $\alpha = 0.05$) were used to test the hypothesis of equal central tendency when the assumptions of normality and equality of variances held.

The applicability of the estimated GSR to crop models was evaluated by computing the RMSE, the mean absolute error (MAE), 2 statistics proposed by Jorgensen et al. (1986), the General Standard Deviation $GSD = \frac{RMSE}{Mean} \times 100$ and the General Absolute Standard Deviation $GASD = \frac{MAE}{|Mean|} \times 100$ and, the index of agreement *d* (Willmott & Wicks 1980) between crop simulation output derived from measured and estimated GSR.

3. RESULTS

3.1. Urban sites

Over 90% of the total variance in measured monthly radiation of the calibration sub-set of urban sites was accounted for by models (Table 4). The best overall fit, in terms of the summary statistics, was reached using MDCBB_m. The CD approach produced the weakest agreement in terms of RMSE and R^2 for the calibration and the independent set of 5 yr. On the other hand, DB model, which was among the best choices during parameterization stage, maintained the best fit of GSR in terms of slope and intercept of the regression line. In

Table 4. Comparison of model simulations of solar radiation ($\text{MJ m}^{-2} \text{d}^{-1}$), on a monthly basis, for the 7 sites of NCDC data set over calibration (1978, 1979, 1980, 1982 and 1986) and evaluation (1981, 1983, 1984, 1985 and 1987) period. The observed values of solar radiation (Obs.) are also given. n: product of 12 monthly values at 7 sites over 5 yr; R^2 : slope; Int.: coefficient of determination, slope and intercept, respectively, of the regression between predicted and observed values; RMSE: root mean squared error; MBE: mean bias error; BC: the Bristow–Campbell model (Bristow & Campbell 1984); CD: the Campbell–Donatelli model (Donatelli et al. 2003); DB: the Donatelli–Bellocchi model (Donatelli & Bellocchi 2000, Bellocchi et al. 2002); MDCBB_m: the modular model with mobile weekly temperature range ΔT (Donatelli et al. 2003); MDCBB_f: the modular model with fixed monthly temperature range ΔT (Donatelli et al. 2003)

	n ^s	Mean	Slope	Int.	RMSE	R ²
Calibration						
Obs.	420	16.88				
Model		MBE (%)				
BC	420	-2.43	1.08	-1.79	1.34	0.926
CD	420	-1.01	0.98	0.18	1.41	0.906
DB	420	-1.18	1.01	-0.38	1.24	0.928
MDCBB _m	420	-0.18	1.00	-0.03	1.24	0.926
MDCBB _f	420	-1.24	0.99	0.03	1.26	0.926
Evaluation						
Obs.	420	17.01				
Model		MBE (%)				
BC	420	-3.41	1.05	-1.51	1.43	0.916
CD	420	-3.17	0.93	0.61	1.67	0.879
DB	420	-2.35	0.97	0.06	1.41	0.909
MDCBB _m	420	-1.06	0.97	0.36	1.37	0.909
MDCBB _f	420	-2.12	0.95	0.47	1.41	0.909

conclusion, Donatelli–Bellocchi's equation can be the most suitable substitute for measured solar radiation to urban environments for which T_{\max} and T_{\min} are available. Furthermore, determining 3 (*b*, *C1*, and *C2*) instead of 4 (*b*, *C1*, *C2* and *Tnc*) empirical parameters is a significant advantage over MDCBB_m approach when simplifications in computations are considered.

Since crop models require daily input, DB equation should also be evaluated on daily basis. The spatial implementation of DB model for the validation data demonstrated a regional trend (Table 5). It performed better at the northern than at the southern locations. RMSE of daily predictions were 3 to 4 times higher than the respective monthly values. Comparable ranges of RMSE values were 2.3–3.9 $\text{MJ m}^{-2} \text{d}^{-1}$ in Northern Italy (Bechini et al. 2000), 3.4–4.1 $\text{MJ m}^{-2} \text{d}^{-1}$ in Canada (Hunt et al. 1998) and 2.6–4.7 $\text{MJ m}^{-2} \text{d}^{-1}$ in Kansas (Goodin et al. 1999). The model generally overestimated GSR in winter and underestimated in summer, as evidenced by the slopes (< 1) of the regression lines between measured and predicted data (Table 5) (Goodin et al. 1999, Bechini et al. 2000, Mahmood &

Table 5. Spatial performance the Donatelli–Bellocchi model, on a daily basis, for the 7 sites of NCDC data set over the evaluation period (1981, 1983, 1984, 1985 and 1987). n: product of 365 daily values over 5 yr. For all other abbreviations, see Table 4

Station	n	Latitude (° N)	MBE (%)	Slope	Int. (MJ m ⁻² d ⁻¹)	RMSE (MJ m ⁻² d ⁻¹)	R ²
Jacksonville	1825	30.30	0.00	0.83	2.9	3.40	0.720
Tallahassee	1825	30.23	-2.37	0.84	2.4	3.18	0.750
Gainesville	1825	29.63	0.00	0.65	5.6	3.81	0.650
Daytona	1825	29.11	-4.52	0.75	5.1	4.22	0.620
Tampa	1825	27.58	-1.67	0.77	4.5	3.17	0.710
Palm Beach	1825	26.41	1.69	0.70	5.7	4.38	0.480
Miami	1825	25.48	-9.58	0.56	9.0	4.92	0.370

Hubbard 2002). These biases were potentially associated with local-scale advection, frontal movements, and the regression approach adopted. Improperly installed and poorly calibrated instrumentation can be partially responsible for these errors.

When the empirical parameters were interpolated in space towards the locations of FAWN data set and daily GSR was estimated with the DB equation, monthly radiation estimates consistently overpredicted the historic values at rural sites by 17.6%, on average (Fig. 2). MBE increased as latitude decreased and varied from 4.8% for Alachua to 28.7% for Homestead. *t*-tests showed significant differences at 7 out of the 15 FAWN locations. Furthermore, RMSE increased by about 3 times over compared with that found when the model parameters used where they derived from, and ranged from 1.2 MJ m⁻² d⁻¹ for Alachua to 5.8 MJ m⁻²

d⁻¹ for Homestead. Possible reasons include: (1) the spatial variation of parameters for FAWN sites is more complicated than the one found using the 7 urban sites; (2) regression between (ΔT) and GSR do not stand because of the differences in site instrumentation and, (3) maritime–inland effects on the corresponding coastal and inland stations were assumed negligible, as they were not taken into account in the interpolation procedure. Furthermore, while T_{\max} and T_{\min} in FAWN data set (1998 to 2000) were higher and lower, respectively, compared with those from urban-sites (1978 to 1987), similar radiation average values were found at both data sets. The proximity of most NCDC stations to the coast and airports, where daytime temperatures are depressed, can justify the above-mentioned temperature contrasts. In general, airports are located around large cities where the conditions are not very representative for the main agricultural production regions, due to buildings, runways and other developmental structures normally found there (Hoogenboom 2000). In the night, heat accumulated in concrete buildings is released making nights warmer in their vicinity. Apparently these same reasons do not affect radiation measured at rural stations which are located in predominantly rural areas surrounded by cultivated areas that show large differences in day and night temperatures.

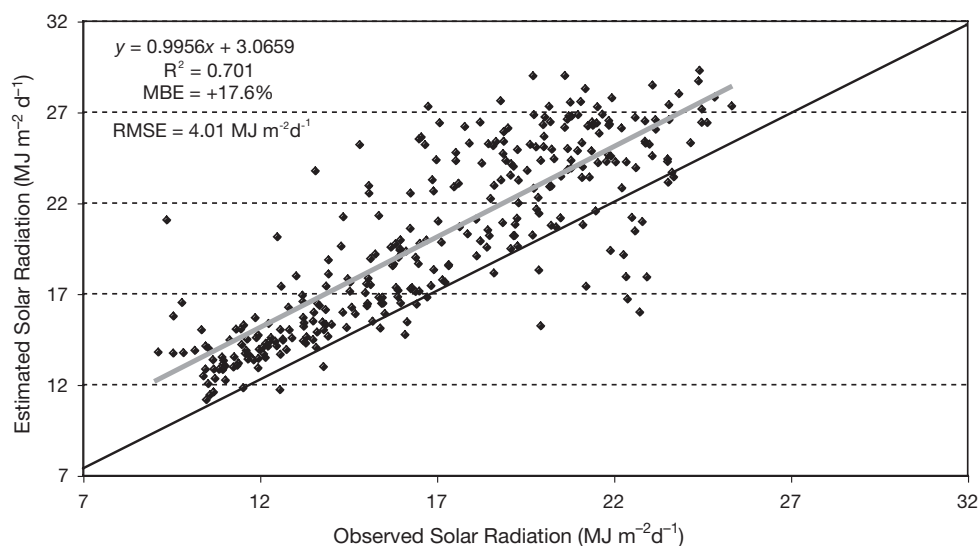


Fig. 2. Comparison of monthly observed and estimated with Donatelli–Bellocchi (DB) model solar radiation at FAWN sites (32 station–years) with interpolated model parameters from NCDC data set (N is 376 instead of 384 due to 6 mo of missing data). RMSE: root mean squared error; MBE: mean bias error

3.2. Rural sites

The fitted models resulted in remarkably good agreement (Table 6) between the estimated and the observed monthly radiation of 16 site-years of the FAWN sub-set (Table 2). The MDCBB_m model achieved the best fit showing the highest R², and the lowest RMSE. Similar to the performance in urban sites, Donatelli-Bellocchi's equation, one of the best choices during parameterization stage, produced also one of the highest correlations and the smallest RMSE for the independent set of the 4 site-years (Table 6). However, during both development and evaluation stage, it tends to under-predict irradiance in the higher range of measured values and over-predict in the lower. When the spatial performance of DB was tested on daily basis (not shown), the poorest agreement between observed and estimated radiation was found for Homestead 2000, where the model showed the lowest R² (0.48) and largest deviations (RMSE was 3.88 MJ m⁻² d⁻¹). It performed much better at the 3 northern site-years where the respective estimates of R² and RMSE ranged from 0.68 to 0.70 and from 3.13 to 3.39 MJ m⁻² d⁻¹, respectively.

The empirical parameters were then interpolated towards the 7 locations of FAWN sub-set reserved for spatial evaluation (see Section 2.2) and daily GSR was estimated using the DB model. The regression analysis between the monthly measured and calculated solar radiation for the 12 site-years of the FAWN sub-set demonstrated a satisfactory accuracy and consistency, as evidenced by the high coefficient of determination (R² = 0.91) and a slope and an intercept close to unity

Table 6. Comparison of model simulations of solar radiation (MJ m⁻² d⁻¹) for the FAWN data set over 16 and 4 calibration and validation site-years, respectively. The observed values of solar radiation (Obs.) are also given. n: product of 12 monthly values over 16 for calibration and 4 for validation sites-years, respectively. For other abbreviations, see Table 4

	n	Mean	Slope	Int.	RMSE	R ²
Calibration						
Obs.	192	16.91				
Model		MBE (%)				
BC	192	-0.06	0.914	1.45	1.12	0.925
CD	192	1.54	0.990	0.43	1.39	0.898
DB	192	0.47	0.888	1.97	1.12	0.927
MDCBB _m	192	0.18	0.883	1.99	1.09	0.932
MDCBB _f	192	0.00	0.872	2.21	1.14	0.926
Evaluation in time						
Obs.	48	17.45				
Model		MBE (%)				
BC	48	-0.06	0.922	1.35	1.09	0.921
CD	48	6.25	0.958	1.83	1.69	0.893
DB	48	0.34	0.873	2.27	1.00	0.939
MDCBB _m	48	-0.57	0.859	2.37	1.00	0.941
MDCBB _f	48	-0.34	0.865	2.29	1.00	0.939

and zero, respectively (Fig. 3). Donatelli-Bellocchi's approach resulted in slight overestimations of the recorded monthly values in the lowest range and in more substantial under-predictions in the higher. Possible reasons of this behavior were given in Section 3.1. On a daily basis, mean bias errors ranged from -2.8 to +3.6% with the exception of Hastings 2000, where the model significantly underestimated the measured mean by 9.4% (Table 7). RMSE ranged from 3.1 at

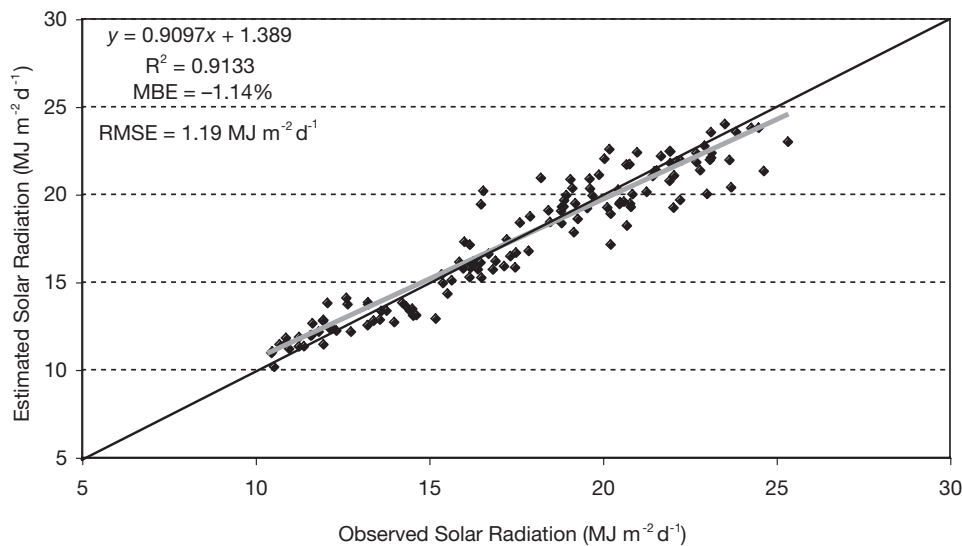


Fig. 3. Comparison of monthly observed and estimated with DB model solar radiation at FAWN sites (12 station-years) with interpolated model parameters from FAWN data set (N is 136 instead of 144 due to 8 mo of missing data). RMSE: root mean squared error; MBE: mean bias error

Table 7. Spatial performance of the DB model on daily basis, using interpolated parameters derived from different FAWN sites, for the 7 sites over 12 evaluation site-years of FAWN sub-set. n: number of years at each site. See Table 4 for other abbreviations

Station	n	Latitude (°N)	MBE (%)	Slope	Int. (MJ m ⁻² d ⁻¹)	RMSE (MJ m ⁻² d ⁻¹)	R ²
Dover	2	28.10	0.6	0.630	6.40	3.70	0.630
Avalon	2	28.47	2.3	0.690	5.70	3.40	0.680
Alachua	1	29.80	3.6	0.780	4.30	3.10	0.750
Hastings	1	29.70	-9.4	0.620	5.00	4.08	0.660
Bradenton	2	27.45	-1.7	0.620	6.50	3.74	0.590
Okahumpka	2	28.70	-2.8	0.670	5.20	3.76	0.660
Ocklawaha	2	29.01	-1.7	0.680	5.30	3.39	0.700

Alachua to 4.1 MJ m⁻² d⁻¹ at Hastings and the models accounted for over 63% of the daily historic GSR variability.

3.3. Spatial structure of model empirical parameters

Model's DB empirical parameters *b*, *C1* and *C2* estimated with NCDC and FAWN data sets were compared with each other against latitude and their spatial attitude was investigated (Fig. 4). Parameter *b* has a more substantial influence on model performance as it modifies ΔT more than *C1* and *C2* (see Eq. 3). The estimates from urban sites reduced in a linear fashion with increasing latitude ($R^2 = 0.78$) and exhibited a spatial trend decreasing from over 0.18 for sites on the coast to lower values at inland sites, reaching a low of 0.095 at Gainesville (Table 1). The respective values from rural sites were much lower, compared with those from NCDC data set, and showed a weak relationship with latitude. The difference in *b* estimates between the 2 data sets declined as latitude increased (Fig. 4a). The *b* value estimated at Gainesville (Table 1, Fig. 4a) is more representative to those derived from rural sites and it would intercept the extension of the regression for FAWN stations line. This suggests that Gainesville appears to be related more to rural than urban locations and our decision to include it with the NCDC data set maybe was an unfortunate one. As expected, the opposite behavior was found for the parameter controlling the seasonal variation magnitude (*C1*). The estimated values developed at rural sites were consistently higher, and positive in their majority except for Ona station, presenting a non-linear association with latitude ($R^2 = 0.05$) (Fig. 4b). In contrast, the resulting values from urban sites, all being negative but Miami, exhibited a much stronger linear relationship ($R^2 = 0.68$) and decreased with increasing latitude. No latitudinal pattern was found for *C2*, the parameter controlling seasonal variation profile, for both data sets

(Fig. 4c). The parameter estimates obtained from FAWN sub-set declined in a non-linear way, as suggested by the very low coefficient of determination (R^2 was equal to 0.09), as latitude increased. The possible causes for the contrasting patterns of empirical parameters *b* and *C1* were mentioned in Section 3.1. It was further assumed that the different time series length of the 2 data sets, in part, might be an additional explanation. In the longer time series of urban sites, notable changes in air temperatures climatology have taken place. These climatic changes inevitably affected the observed relationship between ΔT and daily radiation. However, judging from Gainesville, the environment setting of the station appears to be more important than time series length.

3.4. Crop simulation results

Solar radiation drives potential biomass accumulation and evapotranspiration in both crop models. No significant differences were identified between the simulated potential yield and PET using observed and estimated with DB model solar radiation (Table 8). However, the use of the GSR estimations tended to underpredict the actual year-to-year variability. Based

Table 8. Comparison of (a) potential simulated yield and (b) evapotranspiration (PET) for the 9 crop-site-year scenarios defined in Table 3 (6 for corn and 3 for peanut) with observed (Obs.) and estimated (Est.) solar radiation. RMSE: root mean squared error; MAE: mean absolute error; GSD: general standard deviation; GASD: general absolute standard deviation (see Section 2.5 on how to calculate GSD and GASD); *d*: index of agreement

	Corn		Peanut	
	Obs.	Est.	Obs.	Est.
(a) Yield				
Mean (t ha ⁻¹)	9.12	9.48	6.29	6.32
STD (t ha ⁻¹)	1.41	0.94	0.13	0.08
RMSE (t ha ⁻¹)		1.12		0.15
MAE (t ha ⁻¹)		0.90		0.15
GSD (%)		12.28		2.38
GASD (%)		9.87		2.32
<i>d</i>		0.69		0.00
(b) Pet				
Mean (mm)	592	591	764	76
STD (mm)	43.09	34.68	21.63	22.23
RMSE (mm)		41.30		22.88
MAE (mm)		34.33		19.33
GSD (%)		6.98		2.99
GASD (%)		5.80		2.53
<i>d</i>		0.60		0.58

on the summary statistics of Table 8 and the MBE of Table 3, predictions of simulated results with estimated GSR were more accurate for peanut than for corn suggesting the higher sensitivity of the latter crop to errors in radiation input. The GSD and GASD statistics for corn (peanut) were higher (lower) for yield than for PET. Two crop scenarios, A and C, accounted for most of the error in predicting simulated maize yield (Table 3). Often the effects of errors in radiation in the

performance of a crop model tend to cancel each out despite the non-linearity in the plants response to radiation (Nonhebel 1994). In this study, however, the sign of the simulated yield errors concurred with the respective of PET in 8 out of 9 cases (Table 3). The important difference in the bias of modeled maize yield for scenario A was attributed to the significant overestimation of measured GSR by the estimated radiation particularly in the summer.

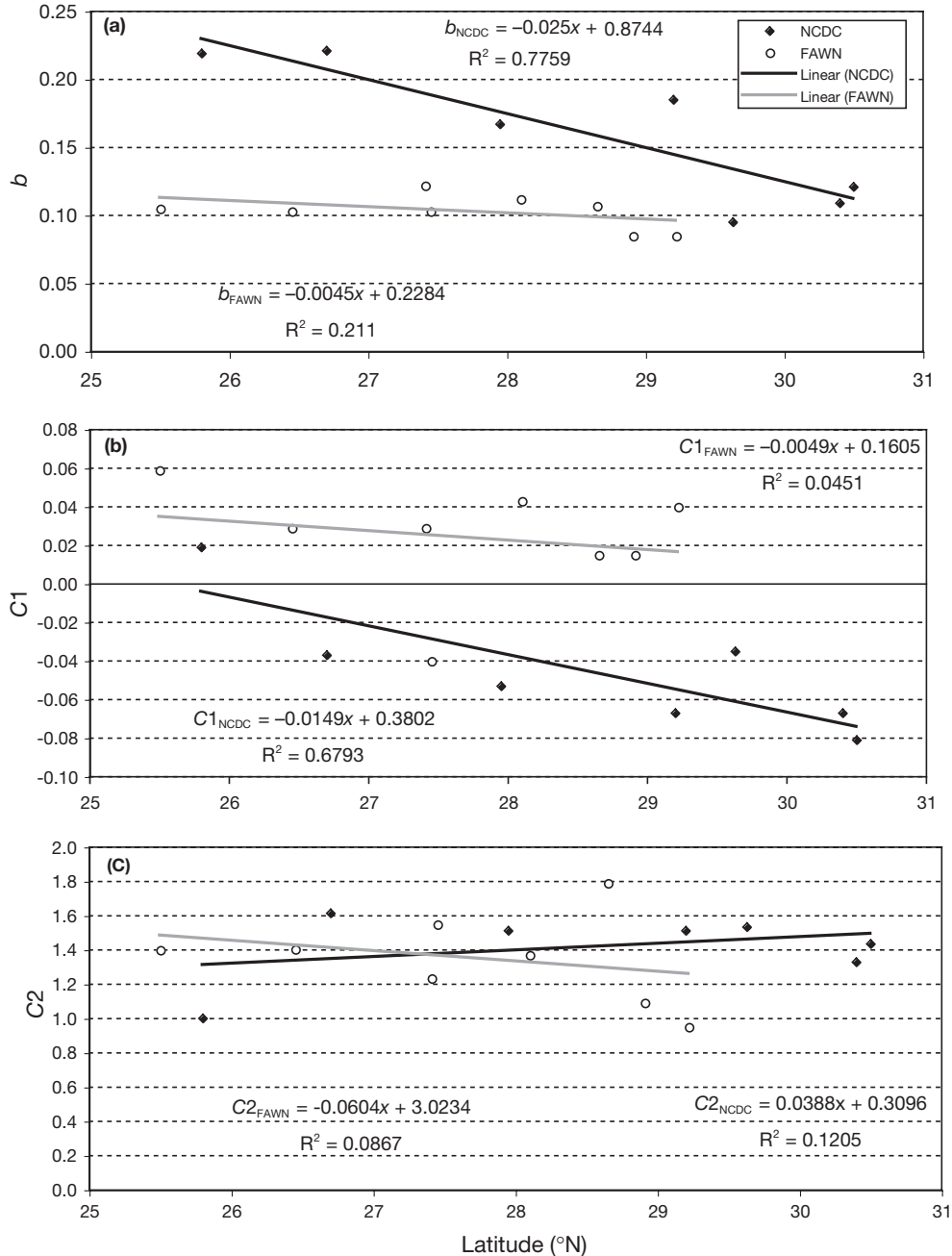


Fig. 4. Comparison of empirical parameters derived from NCDC and FAWN data sets controlling (a) the overall model error (b), (b) the seasonal variation magnitude (C1), and (c) the seasonal variation profile (C2)

In summary, input of estimated with the DB model daily GSR into a crop growth simulation model would be a realistic option at sites where only temperature and rainfall measurements are taken.

4. DISCUSSION AND CONCLUSIONS

This study has evaluated 5 empirically derived models for estimating daily radiation in urban (NCDC data set) and rural (FAWN data set) sites across Florida. Although in many cases not much difference in performance of the models was found, Donatelli–Bellocchi's equation was the best performer. It has shown itself to be a simple, robust and reasonably accurate method for estimating GSR. The quality of those estimates varied with environment (urban–rural) and latitude, yet the errors expressed as RMSE between observed and estimated daily solar radiation were in the range of 3.1–4.1 and 3.2 – 4.9 MJ m⁻² d⁻¹ for FAWN and NCDC evaluation sub-sets, respectively. The model had the tendency to overestimate and under-predict the lower and higher range of the observed distribution. These systematic biases at the extremes of the distribution probably reflected climatic influences on ΔT other than those associated with daily radiation forcing, such as large-scale advection or persistent cloud cover. In addition, forcing factors that were not included may have a role in the type of models used in this study. For example, dust storms, underlying surface conditions, seasonal burning, grassland fires, and pollution from fires may significantly influence recorded radiation (e.g. Thornton & Running 1999). Furthermore, various other atmospheric constituents such as O₂, CO₂, O₃, CH₄, and anthropogenic gases also influence the amount of incident radiation (Rosenberg et al. 1983, Oke 1987).

Radiation for rural sites could be accurately predicted by the coefficients interpolated from rural stations but cannot be adequately predicted by interpolated parameters from urban locations. In contrast, NCDC stations appeared to be site specific and GSR in urban environments can be reasonably predicted from coefficients developed from their own data only. Predicting solar radiation in Australia was also much improved if the coefficients of the models from inland sites were used for inland and those from coastal sites were used for coastal, regardless of the distance between the sites (Liu & Scott 2001). Input of estimated with the DB model GSR into a crop model would be a realistic option at sites where only temperature and rainfall measurements are taken. Another important consideration in this study was the systematic spatial variation of the empirical model coefficients derived from the 2 data sets. The contrasting spatial patterns of

b and $C1$ parameters were mainly attributed to the different effects of urban and rural environmental conditions on temperature range ΔT .

Acknowledgements. The authors thank the anonymous referees for their valuable comments and suggestions for improving the presentation of this work.

LITERATURE CITED

- Bechini L, Ducco G, Donatelli M, Stein A (2000) Modeling interpolation and stochastic simulation in space and time of global solar radiation. *Agr Ecosyst Envir* 81:29–42
- Bellocchi G, Acutis M, Filla G, Donatelli M (2002) An indicator of solar radiation model performance based on fuzzy expert system. *Agron J* 94:1222–1233
- Boisvert JB, Hayhoe HN, Dube PA (1990) Improving the estimation of global solar radiation across Canada. *Agric For Meteorol* 52:275–286
- Boote KJ, Loomis RS (1991) Modeling crop photosynthesis—from biochemistry to canopy. In: Boote KJ, Loomis RS (eds) *CSSA Special Publication No 19. Crop Science Society of America, Madison, WI*
- Boote KJ, Jones JW, Hoogenboom G (1998) Simulation of crop growth: CROPGRO Model. In: Peart RM, Curry RB (eds) *Agricultural systems modeling and simulation. Marcel Dekker, New York*, p 651–692
- Bristow K, Campbell GS (1984) On the relationship between incoming solar radiation and daily maximum and minimum temperature. *Agric For Meteorol* 31:159–166
- Castellvi F, Stockle CO (2002) Comparing the performance of WGEN and ClimGen in the generation of temperature and solar radiation. *Trans ASAE* 44:1683–1687
- Census (2001) Redistricting data (Public Law 94–171). Summary file (State of Florida). 2000 Census of population and housing technical documentation. Prepared by the U.S. Census Bureau
- Donatelli M, Bellocchi G (2000) New methods to estimate global solar radiation. In: *Proc 3rd Int Crop Sci Conf. Hamburg, 17–22 August 2000. European Society for Agronomy, Hamburg*, p 186
- Donatelli M, Bellocchi G, Fontana F (2003) RadEst 3: a software to estimate daily radiation data from commonly available meteorological variables. *Eur J Agron* 18:363–367
- ESRI (Environmental Systems Research Institute) (1996) *ArcView, Version 3. ESRI, Redlands, CA*
- Florida Agricultural Statistics Service (2002) *Florida agricultural facts (NASS) 2002. Florida Agricultural Statistics Service, Orlando, FL*
- Goodin DG, Hutchinson JMS, Vanderlip RL, Knapp MC (1999) Estimating solar irradiance for crop modeling using daily air temperature data. *Agric For Meteorol* 91:845–851
- Hayhoe HN (1998) Relationship between weather variables in observed and WXGEN generated data series. *Agric For Meteorol* 90:203–214
- Hayhoe HN (2000) Improvements of stochastic weather data generators for diverse climates. *Clim Res* 14:75–87
- Hoogenboom G (2000) Contribution of agrometeorology to the simulation of crop production and its applications. *Agric For Meteorol* 103:137–157
- Hunt LA, Boote KJ (1998) Data for model operation, calibration, and evaluation. In: Tsuji GY, Hoogenboom G, Thornton PK (eds), *Understanding options for agricultural production. Kluwer Academic Publishers, Dordrecht*, p 9–39

- Hunt LA, Kuchar L, Swanton CJ (1998) Estimation of solar radiation for use in crop modelling. *Agric For Meteorol* 91: 293–300
- Jagtap SS, Jones JW (2002) Adaptation and evaluation of the CROPGRO-soybean model to predict regional yield and production. *Agr Ecosyst Envir* 93:73–85
- Jagtap SS, Jones JW, Hildebrand P, Letson D, O'Brien JJ, Podesta G, Zierden D, Zazueta F (2002) Responding to stakeholder's demands for climate information: from research to applications in Florida. *Agric Syst* 74:415–430
- Jones PG, Thornton PK (2000) MarkSim: software to generate daily weather data for Latin America and Africa. *Agron J* 92:445–453
- Jorgensen SE, Kamp-Nielsen L, Christensen T, Windolf-Nielsen J, Westergaard B (1986) Validation of a prognosis based upon a eutrophication model. *Ecol Model* 35:165–182
- Liu DL, Scott BJ (2001) Estimation of solar radiation in Australia from rainfall and temperature observations. *Agric For Meteorol* 106:41–59
- Mahmood R, Hubbard KG (2002) Effect of time of temperature observation and estimation of daily solar radiation for the Northern Great Plains, USA. *Agron J* 94:723–733
- Mavromatis T, Hansen JW (2001) Interannual variability characteristics and simulated crop responses for 4 stochastic weather generators. *Agric For Meteorol* 10: 283–296
- Mavromatis T, Jagtap SS, Jones JW (2002) El Nino-Southern Oscillation effects on peanut yield and nitrogen leaching. *Clim Res* 22:129–140
- McVicar TR, Jupp DLB (1999) Estimating one-time-of-day meteorological data from standard daily data as inputs to thermal remote sensing based energy balance model. *Agric For Meteorol* 96:219–238
- Nonhebel S (1994) Inaccuracies in weather data and their effects on crop growth simulation results. I. Potential production. *Clim Res* 4:47–60
- Ritchie JT, Singh U, Godwin DC, Bowen WT (1998) Cereal growth, development and yield. In: Tsuji GY, Hoogenboom G, Thornton PK (eds), *Understanding options for agricultural production*. Kluwer Academic Publishers, Dordrecht, p 79–98
- Rivington M, Matthews KB, Buchan K (2002) A comparison of methods for providing solar radiation data to crop models and decision support systems. In: 1st Biennial meeting of the International Environmental Modelling and Software Society, 24–27 June, Lugano, p 193–198
- Thornton PE, Running SW (1999) An improved algorithm for estimating incident daily solar radiation from measurements of temperature, humidity, and precipitation. *Agric For Meteorol* 93:211–228
- Willmott CJ, Wicks DE (1980) An empirical method for the spatial interpolation of monthly precipitation within California. *Phys Geogr* 1:59–73

*Editorial responsibility: Otto Kinne,
Oldendorf/Luhe, Germany*

*Submitted: October 4, 2004; Accepted: July 26, 2005
Proofs received from author(s): August 22, 2005*