Available online at www.sciencedirect.com
SCIENCE DIRECT®
Agricultural and Forest Meteorology 124 (2004) 99–120
AGRICULTURAL AND FOREST METEOROLOGY
www.elsevier.com/locate/agrformet

Design and optimisation of a large-area process-based model for annual crops

A.J. Challinor[a,b,*], T.R. Wheeler[b], P.Q. Craufurd[b], J.M. Slingo[a], D.I.F. Grimes[c]

[a] *CGAM, Department of Meteorology, University of Reading, PO Box 243, Early Gate, Reading RG66BB, UK*
[b] *Department of Agriculture, University of Reading, PO Box 236, Earley Gate, Reading RG6 6AT, UK*
[c] *Department of Meteorology, University of Reading, PO Box 243, Early Gate, Reading RG66BB, UK*

Received 13 May 2003; received in revised form 22 December 2003; accepted 7 January 2004

Abstract

The formulation of a new process-based crop model, the general large-area model (GLAM) for annual crops is presented. The model has been designed to operate on spatial scales commensurate with those of global and regional climate models. It aims to simulate the impact of climate on crop yield. Procedures for model parameter determination and optimisation are described, and demonstrated for the prediction of groundnut (i.e. peanut; *Arachis hypogaea* L.) yields across India for the period 1966–1989. Optimal parameters (e.g. extinction coefficient, transpiration efficiency, rate of change of harvest index) were stable over space and time, provided the estimate of the yield technology trend was based on the full 24-year period. The model has two location-specific parameters, the planting date, and the yield gap parameter. The latter varies spatially and is determined by calibration. The optimal value varies slightly when different input data are used. The model was tested using a historical data set on a 2.5° × 2.5° grid to simulate yields. Three sites are examined in detail-grid cells from Gujarat in the west, Andhra Pradesh towards the south, and Uttar Pradesh in the north. Agreement between observed and modelled yield was variable, with correlation coefficients of 0.74, 0.42 and 0, respectively. Skill was highest where the climate signal was greatest, and correlations were comparable to or greater than correlations with seasonal mean rainfall. Yields from all 35 cells were aggregated to simulate all-India yield. The correlation coefficient between observed and simulated yields was 0.76, and the root mean square error was 8.4% of the mean yield. The model can be easily extended to any annual crop for the investigation of the impacts of climate variability (or change) on crop yield over large areas.

Keywords: Crop simulation models; Spatial scale; General circulation model; Groundnut; Crop yield

1. Introduction

Crop productivity in many environments is highly dependent upon weather and climate. This is particularly so in tropical regions, which are often monsoon environments with low levels of crop management technology. Increased understanding of the impacts of current sub-seasonal and inter-annual climate variability on crop yields would support agricultural

Abbreviations: LAI: leaf area index; GCM: general circulation model; VPD: vapour pressure deficit; RMSE: root mean square error; ICRISAT: International Crops Research Institute for the Semi-Arid Tropics; SLA: specific leaf area; RUE: radiation use efficiency; GLAM: general large-area model for annual crops

* Corresponding author. Tel.: +44-118-3786016; fax: +44-118-378-8316.
E-mail addresses: ajc@met.rdg.ac.uk (A.J. Challinor), t.r.wheeler@reading.ac.uk (T.R. Wheeler).

0168-1923/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.agrformet.2004.01.002

Nomenclature

A	mean albedo of the surface
C_x	dimensionless constant identified by x
C_{YG}	yield gap parameter
e	vapour pressure
e_{sat}	saturation vapour pressure
E	evaporation rate (cm per day)
E_T	normalised transpiration efficiency
$E_{TN,max}$	maximum transpiration efficiency
G	soil heat flux
H_I	harvest index
k	extinction coefficient
k_{DIF}	uptake diffusion coefficient
k_{sat}	saturated hydraulic conductivity of soil
l_v	root length density (by volume)
L	leaf area index
N_{SL}	number of soil layers
P	precipitation
R	runoff
R_N	net surface all-wave radiation
S	soil water stress factor
S_{rad}	surface incoming solar radiation
t	time
t_{TT}	thermal time
t_{TT_i}	thermal duration of development stage i
T	temperature
T_b	base (cardinal) temperature
T_{eff}	effective temperature
T_m	maximum (cardinal) temperature
T_{max}	maximum daily temperature
T_{min}	minimum daily temperature
T_o	optimal (cardinal) temperature
T_T	transpiration rate (cm per day)
T_{Tpot}	potential transpiration
$\bar{T}$	mean daily temperature
V	vapour pressure deficit
V_{EF}	extraction front velocity
W	above-ground biomass
x_{cr}	critical, or threshold, value of parameter x
x^e	energy-limited value of parameter x
Y	yield
z_{ed}	depth of soil from which evaporation occurs
z_{ef}	depth of the root extraction front
z_{max}	depth of soil profile

Greek letters

α	Priestley–Taylor coefficient
γ	ratio of the specific heat of air at constant pressure to λ
θ	soil water content
θ_{dul}	drained upper limit
θ_{ll}	lower limit
θ_{pe}	potentially extractable soil water
θ_{sat}	saturated upper limit
λ	latent heat of vapourisation of water

planning over these timescales. In addition, an understanding of the impacts of current climate variability is a prerequisite for the study of the impacts of climate change (Challinor et al., 2003). These impacts will be seen across many spatial scales, but it is relatively large spatial scales, similar to those of regional and global climate models, that facilitate an integrated modelling approach to the problem. Hence there is a need for process-based crop models that can capture the impact of weather and climate variability on these spatial scales.

Crop models seek to quantify the relationships between crop growth and weather, combined with some of the effects of crop management. These models often perform well at the field scale. For example, the five wheat simulation models compared by Jamieson et al. (1998) provided reasonable predictions of the response of grain yield to water supply. PNUTGRO, the CROPGRO configuration for groundnut (i.e. peanut) has been tested for a plot of unspecified size in Punjab, India (Kaur and Hundal, 1999) and found to predict seed yield to within ±10% of observed values ($r = 0.94$). In order to capture variability at the field scale, such models have a high input data requirement for information on genotype, soil type, slope and so on. Other models (e.g. Parthasarathy et al., 1992) use statistical regressions, and so require less data. However, statistical models are less adaptable to different conditions over both time (e.g. changing CO_2 or temperature) and space (e.g. water- versus radiation-limited growth).

There is an increasing number of studies examining the combination of general circulation models (GCMs) and crop models (Mearns et al., 1999; Hansen and Jones, 2000; IRI, 2000; Tsvetsinskaya et al., 2001a;

Mearns et al., 2001). One of the principle issues in the application of climate forecasts to crop modelling (and indeed agricultural applications in general: Hansen, 2002) is the disparity in spatial scale between forecast and impact assessment. Challinor et al. (2003) set out a methodology for the design of a combined weather and crop yield forecasting system. It begins with an exploration of the spatial scale of the relationship between crop yield and weather. A crop model is then designed to run at the spatial scale at which a relationship between crop yield and weather is observed. A working spatial scale of the order of tens to a few hundred kilometres (comparable to the resolution of regional and general circulation models) was established for the case of groundnut yield in India. The current study presents the design and optimisation of a crop model (the general large-area model for annual crops, GLAM) which operates on these scales. Accordingly, the model is sufficiently process-based to simulate crop productivity over a range of tropical environments, whilst being simple enough to avoid the need for large amounts of location-specific input data or calibration.

There are a variety of approaches to the issues of choice of complexity (empirical versus process-based modelling) and spatial scale for modelling crops over large areas. For example, Brooks et al. (2001) proposed a meta-model of the wheat simulation model SIRIUS for regional impacts assessments based on monthly mean weather input data. A spatial model for simulating wheat crop phenology across Europe was developed by Harrison et al. (2000) in which daily temperature data were derived by interpolation from monthly values. Landau et al. (2000) described a multiple regression model of wheat yield across the UK using weather variables averaged over five development stages. This semi-empirical approach sought to combine the pragmatism of empirical modelling with a simulated phenology. This model captures well the average spatial variability in yield, but cannot be extrapolated outside of current UK climate conditions. Jagtap and Jones (2002) noted the impracticality of modelling yields at the field scale for large area applications. Their study maintained the process-based approach by simulating regional variability in soybean yields using an ensemble of nine CROPGRO-soybean runs (varying planting date and crop variety) on a 0.5° square grid. They used daily weather inputs from a weather generator. Simulated yields had a (spatial) median root mean square error (RMSE) of between 7 and 85%. Finally, more empirical approaches such as the FAO method (Doorenbos and Kassam, 1979) are sometimes employed for large-area crop modelling (Fischer et al., 2002; Martin et al., 2000). Such approaches assume a degree of stationarity in derived crop–weather relationships which are not necessarily valid in future climate scenarios. Further discussion on crop modelling for climate variability applications can be found in Hansen and Jones (2000) and Challinor et al. (2003).

The spatial scale of a mathematically one-dimensional crop model is defined by the level of detail and spatial aggregation of its input data requirements. By aiming for an intermediate level of input data requirement, the spatial scale of GLAM can be varied by varying the spatial scale of the inputs. Note however that different input scales can produce very different simulated yield impacts (Mearns et al., 2001) and that caution is needed where input crop parameters are calibrated to account for this since aggregation error can result (Hansen and Jones, 2000). More complex models such as CROPGRO (Boote and Jones, 1998) include more non-linear processes, so that spatial aggregation of inputs over heterogeneous land is less justifiable. However, if resources permit, remote sensing can be used to identify homogeneous crop regions over which a model such as this can be run (Basso et al., 2001; Jones and Barnes, 2000; Guerif and Duke, 2000). GLAM is intended for large ($\gg$ field scale) spatial scales. The precise spatial scale on which GLAM is run should be determined by the scale of the weather–yield relationship (Challinor et al., 2003).

GLAM employs a process-based approach, in order to maintain as high a domain of spatial and temporal (e.g. climate change) applicability as possible, and to account for the effects of sub-seasonal variability in weather. The objective of the model is to reproduce the impact of weather on observed crop yield. This aim leads to two particular model characteristics: firstly, complexity at a level far-removed from yield-determining processes is omitted (see Sinclair and Seligman, 2000). Hence, for example, photosynthesis is not modelled directly, but is represented by a transpiration efficiency. In general, simple param-

eterisations are favoured over more complex methods. Secondly, of the impacts on yield due to factors other than weather (pests, diseases, management factors, etc.), only two are modelled explicitly: planting date and soil type. The rest, whilst in reality complex and varied, are modelled using a single yield-gap parameter. This allows the model to focus on the impact of weather and climate on the spatio-temporal variability of crop yield. The impact of climate on model yields is two-fold. Firstly, a different climate implies different statistics of daily weather, which may impact upon crop yield. Secondly, a different climate may imply different values of model parameters. For example, the impact of potential future CO_2 concentrations can be modelled using a parameterised effect of CO_2 on transpiration efficiency. For future climates, the model may also need to include the sensitivity of yield to weather extremes such as high temperature (e.g. Wheeler et al., 2000) which may be encountered with increased frequency. In accordance with the guidelines of Sinclair and Seligman (2000) all the parameters, except the yield-gap parameter, can be measured independent of the model formulation.

GLAM aims to combine the benefits of more empirical modelling methods such as Fischer et al. (2002) and Landau et al. (2000) (low input data requirement, validity over large areas) with the benefits of a process-based approach (the potential to capture variability due to different sub-seasonal weather patterns, and hence increased validity under future climates). The model has been developed such that different crop-specific parameter values, and in some cases equations, allow simple and transparent operation across many annual crops. This approach is similar to that of Wang et al. (2002), who developed a generic crop modelling template for more physiologically and spatially detailed applications.

GLAM is optimised and tested here for groundnut (i.e. peanut; *Arachis hypogaea* L.) in India. Extension to other crops can proceed along similar lines to the calibrations described in Section 3.1. The model results presented here are large-scale deterministic hindcasts carried out in order to look at the stability of optimal parameters over time, space and input data (Section 3) and to obtain an estimate of model consistency and skill (Section 4).

2. Model description

The model formulation is described below, using the following notation: capital subscripts form part of the identification of a variable type (e.g. harvest index, H_I) whereas lower case subscripts distinguish within variable types (e.g. daily maximum and minimum temperatures T_{max} and T_{min}). Empirically determined coefficients are denoted by C_x in the dimensionless case, and K_x for other variables, where x identifies the particular coefficient. Subscript T is used for the two transpiration-related quantities (transpiration efficiency, E_T, and transpiration rate T_T) in order to distinguish them from temperature (T) and evaporation rate (E). Values of all the parameters, together with sources, are listed in Appendix A. In total, the model has forty parameters, five of which vary spatially, and an additional twenty of which are crop-specific.

2.1. Growth and development

The crop is planted either on a specified date, or on the first day that the soil moisture exceeds a given fraction (C_{sow}) of the maximum available soil water. If this threshold is not reached within a given time limit (30 days has been used in this study) then the crop is planted regardless. Emergence (the first day on which LAI, leaf area index, becomes non-zero) occurs t_{em} days after sowing ($t_{em} = 8$ for groundnut). The thermal time elapsed within a given development stage is given by

$$t_{TT} = \int_{t_i}^{T} (T_{eff} - T_b)\, dt \quad (1)$$

where t is the time, T_b is the base temperature, below which development ceases, and i is the development stage number, equal to 0 between sowing and flowering, 1 between flowering and pod-filling, 2 between pod-initiation and maximum LAI and 3 between maximum LAI and maturity. Development stage i starts at time t_i and is completed after a specified duration t_{TT_i} has elapsed and harvest is at maturity.

The effective temperature, T_{eff} is defined as follows using the cardinal temperatures T_b, T_o and T_m, where the subscripts denote base, optimum and maximum

temperatures, respectively.

$$T_{\text{eff}} = \begin{cases} \bar{T} & T_{\text{b}} \le \bar{T} \le T_{\text{o}} \\ T_{\text{o}} - (T_{\text{o}} - T_{\text{b}}) \left(\dfrac{\bar{T} - T_{\text{o}}}{T_{\text{m}} - T_{\text{o}}} \right) & T_{\text{o}} < \bar{T} < T_{\text{m}} \\ T_{\text{b}} & \bar{T} \ge T_{\text{m}}, \bar{T} < T_{\text{b}} \end{cases} \quad (2)$$

The mean daily temperature, $\bar{T}$ is taken either directly from measurements, or as the average of T_{min} and T_{max}.

The growth of the crop leaf area is determined as follows:

$$\frac{\partial L}{\partial t} = \begin{cases} \left(\dfrac{\partial L}{\partial t} \right)_{\text{max}} C_{\text{YG}} \min \left(\dfrac{S}{S_{\text{cr}}}, 1 \right) & i < 3 \\ 0 & i = 3 \end{cases} \quad (3)$$

where L is the effective LAI (see below), $(\partial L/\partial t)_{\text{max}}$ is a prescribed constant and

$$S = \frac{T_{\text{T}}}{T_{\text{Tpot}}} \quad (4)$$

is the soil water stress factor, which begins to affect growth at values less than the critical threshold value S_{cr}. T_{T} and T_{Tpot} are the rates of transpiration and potential transpiration, respectively. C_{YG} is the yield gap parameter, used to reduce LAI from the physical value to an effective value which accounts for the mean effects of pests, diseases and non-optimal management. Yield potential (maximum obtainable yields as determined by weather and crop) may be simulated by setting $C_{\text{YG}} = 1$. This approach to the yield gap is similar to that of Jagtap and Jones (2002), whose yield correction factor is applied *after* the crop simulation. The roots grow according to the following equations:

$$\begin{aligned} &\frac{\partial l_{\text{v}}(z=0)}{\partial L} = \text{prescribed constant}, \\ &V_{\text{EF}} = \text{prescribed constant}, \\ &l_{\text{v}}(z = z_{\text{ef}}) = \text{prescribed constant} \end{aligned} \quad (5)$$

where l_{v} is the root length density by volume, z denotes depth into the soil, z_{ef} is the depth of the root extraction front and V_{EF} is the extraction front velocity. Above-ground biomass (W) is determined by a separate prognostic equation:

$$\frac{\partial W}{\partial t} = T_{\text{T}} \min \left(\frac{E_{\text{T}}}{V}, E_{\text{TN,max}} \right) \quad (6)$$

where V is the vapour pressure deficit (VPD $= e_{\text{sat}}(\bar{T}) - e$, where e is the vapour pressure), E_{T} is the normalised transpiration efficiency in Pa (i.e. $V \times$ transpiration efficiency in g kg^{-1}), and $E_{\text{TN,max}}$ is the maximum transpiration efficiency in g kg^{-1}. In the case of studies in possible future climates, E_{T} would be increased from current values as a simulated response to higher CO_2 levels.

Where humidity measurements are not used, V is parameterised as

$$V = C_V [e_{\text{sat}}(T_{\text{max}}) - e_{\text{sat}}(T_{\text{min}})] \quad (7)$$

where C_V is a constant (Tanner and Sinclair, 1983). The saturation vapour pressure deficit at temperature T, $e_{\text{sat}}(T)$, is determined after Bolton (1980). Yield is determined using harvest index, H_{I}. For $i < 2$ the yield component is zero, and for $i \ge 2$

$$\frac{\partial H_{\text{I}}}{\partial t} = \text{prescribed constant}, \qquad Y = H_{\text{I}} W \quad (8)$$

2.2. Water balance equations

The drained upper limit, lower limit and saturation limit of the soil (θ_{dul}, θ_{ll} and θ_{sat}, respectively) are assumed to be constant throughout the profile, which is of depth z_{max} (set equal to the maximum attainable rooting depth). The soil is split into N_{SL} soil layers, each with a value of $l_{\text{v}}(z)$ determined by Eq. (5), and a value of the volumetric soil water content, θ, determined at each time step (1 day) by carrying out the following steps in order: firstly, rainfall runoff occurs according to the US Soil Conservation Service method (USDA-SCS, 1964; Choudhury and DiGirolamo, 1998):

$$R = \frac{P^2}{P + S} \quad (9)$$

where R is the runoff, P is the precipitation and S is the amount of water that can soak into the soil. The latter is set equal to the saturated hydraulic conductivity of the soil, k_{sat}. Interception by the leaf canopy is ignored since it is a small term made hard to estimate by uncertainty in rainfall data and canopy architecture.

The second step in the soil water balance sequence is drainage, which occurs according to the scheme of Suleiman (1999):

$$\frac{\partial\theta}{\partial t} = -FD(\theta_s - \theta_{dul}),$$
$$D = C_{d1}\theta_{dul}^2 + C_{d2}\theta_{dul} + C_{d3},$$
$$F = 1 - \frac{\ln(Q_i + 1)}{\ln(k_{sat} + 1)}, \qquad k_{sat} = K_{ks}\left(\frac{\theta_{sat} - \theta_{dul}}{\theta_{dul}}\right)^2 \tag{10}$$

where FD is the drainage rate, the factor F accounting for simultaneous inflow from the layer above, θ_s is the initial value of θ, Q_i is the incoming water flux from the layer above ($P - R$ in the case of the uppermost layer). C_{d1}, C_{d2}, C_{d3} and K_{ks} are empirical constants.

Finally, water is extracted over the depth z_{ed} by evaporation, and over the root-zone depth by roots, according to the amount of transpiration (Eq. (19)). Note that z_{ed} must be a multiple of the soil layer depth (two in the case of the current study). The optimal N_{SL} was found by plotting yield against N_{SL} for a given z_{ed} over a number of environments (not shown). For $N_{SL} > 25$ variation in yield was small (< 1%), therefore $N_{SL} = 25$ was adopted for all model runs.

2.3. Evaporation and transpiration

Transpiration (T_T) and evaporation rates (E) are determined by considering separately the limitations imposed by plant/soil structure, energy availability, and water availability. Potential values of E and T_T are defined as being limited by only the first two of these constraints. The physiologically limited transpiration is modelled using an empirical relationship based on the data of Azam-Ali (1984):

$$T_{Tpot}^p = \begin{cases} T_{Tmax}\left(1 - \dfrac{L_{cr} - L}{L_{cr}}\right) & L < L_{cr} \\ T_{Tpot}^p = T_{Tmax} & L \geq L_{cr} \end{cases} \tag{11}$$

where L_{cr} is a critical threshold value of L and T_{Tmax} is the maximum possible potential transpiration rate.

The energy-limited evaporation and transpiration rates (E^e and T_T^e, respectively) are defined to be consistent with the Priestley–Taylor equation (Priestly and Taylor, 1972) so that potential evapotranspiration rate is given by

$$E_{pot}^T = E^e + T_T^e = \frac{\alpha}{\lambda}\frac{\Delta(R_N - G)}{\Delta + \gamma} \tag{12}$$

where R_N is the net all-wave radiation, G is the soil heat flux, λ is the latent heat of vapourisation of water, $\Delta = \partial e_{sat}/\partial T$, determined after Bolton (1980), and γ is the ratio of the specific heat of air at constant pressure to the latent heat of vapourisation of water. The Priestley–Taylor coefficient (α) is parameterised after Jury and Tanner (1975) as a function of VPD, a constant reference value of VPD (V_{ref}) (Steiner et al., 1991) and a pre-correction value $\alpha_0 = 1.26$ (Priestly and Taylor, 1972):

$$\alpha = 1 + (\alpha_0 - 1)\frac{V}{V_{ref}} \tag{13}$$

This method was chosen because it takes some account of aerodynamic (i.e. advective) effects on the exchange of water vapour from the surface without the need for windspeed data.

The net radiation is estimated from the solar radiation using

$$R_N = (1 - A)S_{rad} \tag{14}$$

where S_{rad} is the incoming short-wave radiation and A is the mean albedo of the surface. This assumes that net long-wave radiation is zero, which is very reasonable under monsoon (i.e. cloudy) conditions. During monsoon breaks, and during the dry season, clear skies mean that R_N can be over-estimated by up to $100\,W\,m^{-2}$. However, sensitivity studies have shown that this has an insignificant effect on the model simulations.

Modelling the light interception using the Beer–Bougert equation (see, e.g. Arya, 1988) then gives

$$E^e = (1 - C_G)E_{max}^T\,e^{-kL} \tag{15}$$

and

$$T_T^e = E_{max}^T(1 - e^{-kL}) \tag{16}$$

and where C_G is the constant in the equation for the soil heat flux, $G = C_G R_N\,e^{-kL}$ (Choudhury et al., 1987) and k is the extinction coefficient. E_{max}^T, the maximum possible energy-limited evapotranspiration is given by setting $G = 0$ in Eq. (12).

The potential (energy and soil-structure limited) daily evaporation is modelled after Cooper et al. (1983):

$$E^{s}_{pot} = \frac{E^{e}}{t_R} \quad (17)$$

where t_R is the number of days since the daily total rainfall was greater than a threshold value of P_{cr}. This threshold value was chosen to be just above zero (1 mm) to avoid unrealistically high potential evaporation just after low rainfall.

The potential (energy and physiology limited) transpiration rate is given by

$$T_{Tpot} = \min(T^{p}_{Tpot}, T^{e}_{T}) \quad (18)$$

Now that energetic and structural constraints have been modelled, it only remains to account for limited water availability. This is done by partitioning the available water according to demand where necessary, so that

$$T_T = T_{Tpot} \quad \text{and} \quad E = E_{pot} \quad \text{for } \theta_{pe} \geq E^{T}_{pot},$$
$$T_T = \theta_{pe}\frac{T^{e}_{T}}{T^{e}_{T} + E^{e}} \quad \text{and} \quad E = \theta_{pe}\frac{E^{e}}{T^{e}_{T} + E^{e}} \quad \text{for } \theta_{pe} < E^{T}_{pot} \quad (19)$$

where θ_{pe} is the potentially extractable soil water. This is given by using the parameterisation of Passioura (1983):

$$\theta_{pe} = \int_0^{z_{max}} \theta_{cr}(1 - e^{k_{DIF} l_v(z) t_e(z)})\, dz \quad (20)$$

where $t_e(z)$ is the time of first root uptake in layer z, and k_{DIF} is the uptake diffusion coefficient. $t_e(z)$ is determined initially as the time of arrival of the extraction front, and subsequently as the time at which the equation is re-initiated as a result of wetting such that $\theta(z) > \theta_{cr}$, where

$$\theta_{cr} = \theta_{rll} + C_\theta(\theta_{dul} - \theta_{rll}) \quad (21)$$

(Allen et al., 1998) and C_θ is a constant.

3. Calibration and parameter optimisation

3.1. Method

The model has been calibrated and tested for groundnut in India, using thirty-five 2.5° × 2.5° grid cells. Both Rabi (irrigated November sowing, March–April harvest) and Kharif (monsoon season; June or July sowing, September–December harvest) seasons were simulated. The Rabi season crop was simulated with water non-limiting ($S = 1$; see Eq. (4)). The current study focusses mainly on three grid cells, identified by the (majority) geographical State in which they lie: (i) Andhra Pradesh (AP), an inland cell towards the south of India, (ii) Gujarat (GJ), the westernmost cell, and (iii) Uttar Pradesh (UP) in the northeast of India. These were chosen for their geographical and meteorological diversity, and for the range of Rabi:Kharif growing area ratios (Fig. 1). The States of Gujarat and Andhra Pradesh together accounted for 44% of India's total groundnut production over the study period (1966–1989).

Model parameters (see Appendix A) were determined from the literature where possible, with a preference for the most widely grown cultivar in southern India, TMV-2. The parameters which were not taken from the literature were determined as follows: the maximum LAI growth rate, $(\partial L/\partial t)_{max}$, was allowed a broad range corresponding to the range of possible maximal (i.e. non-water limited) LAI values for a groundnut crop planted across a range of sowing densities. The thermal durations, t_{TT_0}–t_{TT_3}, were obtained by calculating the durations of the corresponding periods in a series of runs of CROPGRO-groundnut (Boote and Jones, 1998) carried out using TMV-2 coefficients over various years and locations in India. The maximum transpiration efficiency, $E_{TN,max}$, was given a broad range which included values low enough to give biomass results comparable to and less than parallel CROPGRO runs, and high enough to not limit E_T at all. The critical soil water stress factor, S_{cr} (see Eq. (3)), was given the broad range 0.5–1, and found not to greatly affect results.

In addition to the above global (site-independent) parameters, there are two model parameters which vary spatially. These are the yield gap parameter, C_{YG}, given the broad range of 0.1–1.0, and the planting date. The latter was derived for rainfed crops using a threshold soil moisture parameter (C_{sow}) as described in Section 2.1. The value of C_{sow} had little impact upon the results over the range tested (0.3–0.7) and the central value of 0.5 (similar to the value of 0.6 used by Gadgil et al. (1999)) was adopted. For irrigated crops, the planting date was set equal to the earliest sowing

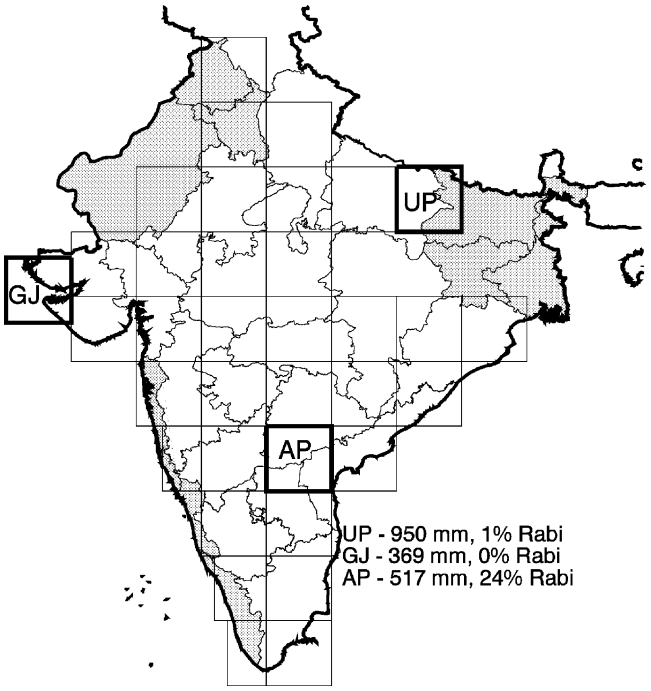

Fig. 1. Location of the three grid cells studied (AP in Andhra Pradesh, GJ in Gujarat, and UP in Uttar Pradesh). The legend shows June–September rainfall and the fraction of total production grown in the Rabi season. Also shown are the remaining 32 grid cells used in the all-India (ALLIN) runs, and the subdivisional boundaries. Grey shading indicates subdivisions which had a mean groundnut-growing area of less than 150 km^2 over the period 1966–1989. See Section 3.2 for data sources.

date (the irrigated runs were not sensitive to planting date). The earliest planting date was taken from Reddy (1988), and the soil moisture on this date (θ_s) was chosen to be equal to the drained lower limit (θ_{ll}). The lack of separate Kharif and Rabi yield data means that the Kharif and Rabi runs had to be calibrated jointly; a single set of crop parameters, including C_{YG}, was derived for each grid cell. Whilst being a simplification, this was considered more realistic than not simulating the Rabi season crop at all.

3.2. Datasets

The daily weather inputs for this study were one of two possible sets: (i) MMT—maximum and minimum temperature over 24 h on a macro-regional scale (six regions across India; see Pant and Kumar, 1997) with Eq. (7) to determine the humidity, and $\bar{T} = (T_{max} + T_{min})/2$; and (ii) VAP—vapour pressure deficit and mean temperature on a 0.5° × 0.5° grid (New et al., 2000). Both sets use daily rainfall for June–September on a 2.5° × 2.5° grid from the Indian Institute of Tropical Meteorology (http://www.tropmet.res.in/), and radiation on a 0.5° × 0.5° grid (New et al., 1999). All inputs except rainfall are monthly means, which were interpolated linearly to daily values. As a result, inputs other than rainfall will have a reduced standard deviation in comparison to observations. All data have interannual variability with the exception of surface incoming solar radiation (S_{rad}), which is a 1966–1990 climatology. All the inputs were (linearly) spatially interpolated (T_{max} and T_{min}) or averaged (S_{rad}, $\bar{T}$, VPD) onto the rainfall grid, and the model was run on this grid.

Soil hydrological properties were derived from FAO/UNESCO (1974). The textural categories were used to assign each 0.5° × 0.5° grid square to one of seven classes, five ranging in texture from sand to clay, with the addition of lithosol and organic categories (Tate, 2001). Values of the soil parameters (θ_{sat}, θ_{ll} and θ_{dul}) for each run were obtained by averaging onto the model grid. To account for the uncertainty in the representation of soil hydrological parameters

over such large areas using these parameters, three possible sets of the soil parameters were defined, corresponding to mean, upper, and lower values of the maximum available soil water within each of the seven classes. These three sub-classes are referred to as course, medium and fine. The soil parameters are listed in Appendix A.

Yearly district-level groundnut production and irrigation level (fraction of total growing area irrigated) data for the period 1966–1995 have been compiled by the Socio-economic Policy Division of the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, India, from yearly agricultural bulletins (Agricultural Situation in India, Department of Agriculture, Government of India). Districts range in size from less than 1 to 46,800 km^2 (average 8300 km^2), although there are only two districts which are less than 690 km^2 in area. AP has five groundnut-growing districts, and GJ and UP each have one. The data set contains growing-area and production by weight. Yield data on the model grid were calculated from the district data. Linearly detrending the yield data (after upscaling to the model grid) produced data at the production technology level of the start year, and these data were used in the model calibration and evaluation procedure. District-level yields within AP in any 1 year varied by over 200% on five occasions. Because of this high level of variability, district-level yields were not used in the validation procedure. However, higher resolution rainfall data would allow a study of subgrid variability in the future.

All the results presented, unless otherwise stated, have the yield calculated from the two seasons (Rabi and Kharif, see Section 3.1) summed using the district-level irrigation data upscaled to the model grid. The statistics of the weather and groundnut production in the three chosen grid cells over the study period are presented in Table 1. Further information on the statistics of groundnut production, and its relationship to rainfall in India, can be found in Challinor et al. (2003). All three cells have predominantly silt loam or clay loam sands.

3.3. Determination and stability of optimal parameters

The GLAM model seeks to maintain a broad range of applicability without the need for heavy local fine-tuning. Hence the optimal value of global parameters should remain constant over space and time. The yield gap parameter (C_{YG}) should be stable over time for each site, and together with local soil parameters and weather data will provide the spatio-temporal variability in yield. Stability of parameters implies that optimal parameters are independent of the subset of data used to determine them; hence this analysis is an alternative method of model evaluation to methods separating data into calibration and testing subsets.

The optimal parameter set was determined by first examining one grid cell, using one set of input data (MMT; see Section 3.2). GJ was chosen, as the crop there is 100% rainfed. In order to examine the consistency of model parameters across space, all model parameters were varied in turn across the ranges listed in Appendix A. The optimal set was defined as that which minimised the RMSE. This procedure was repeated a number of times to ensure the local minimum in RMSE had been found. Although no systematic study was undertaken, there was no evidence for multiple minima.

The yield gap parameters (C_{YG}) for the other two grid cells (UP and AP) were then determined, again by

Table 1
Statistics of groundnut production and weather in the three study cells

Cell code	A_{cell}/A_{AI} (%)	A^{Rabi}/A_{cell}		JJAS ppn (mm)		$\bar{T}$ (°C)		S_{rad} (MJ m^{-2} per day)	
		Mean	σ	Mean	σ	JJAS	DJFM	JJAS	DJFM
AP	6.4	24	6	517	159	31.0	24.4	15.2	18.9
GJ	5.1	0	0	369	181	30.1	23.5	17.5	17.3
UP	0.1	1	1	950	182	30.6	20.5	15.6	14.6

A_{cell} is the area used for groundnut cultivation in grid cell, and A_{AI} is the total all-India groundnut area. Superscript Rabi indicates Rabi growing season only. JJAS and DJFM are June–September and December–March, respectively. $\bar{T}$ is the time-mean temperature from the MMT (maximum and minimum daily temperature) data, ppn is the precipitation, and S_{rad} is the incoming solar radiation.

minimising the RMSE. In order to examine the spatial variability of the optimal parameter set, the RMSE was examined across parameter space for these two cells. The optimal parameter set was found not to vary appreciably across the three cells, despite differences in the climate and ratio of Rabi:Kharif. The optimal value of C_{YG} for the remaining 32 groundnut-growing grid cells (Fig. 1) was determined using this optimal parameter set. The RMSE of all 35 cells taken together (referred to as ALLIN) was measured and found to produce optimal parameters consistent with those of individual cells.

A similar procedure was carried out using VAP data. Different optimal values of C_{YG} were found (Fig. 2a) whilst the optimal values of the other parameters remained the same. When the C_{YG} values from the MMT run were used, different optimal values of other parameters resulted, and these optimal values varied spatially. Hence model calibration can compensate to some degree for differences in input weather data and produce a parameter set which is internally consistent across space. When the model is calibrated with, for example, GCM output, a different set of C_{YG} is likely to result.

The consistency of model parameters across time is also important. For the most part there is little to no difference between optimal values of C_{YG} when three different time periods are used (Fig. 2). The exception is GJ, where the optimal value varied between 0.3 and 0.9 for the three time periods. However, the RMSE for the time period 1966–1989 varied by less than 4% when using these three C_{YG} values. Yields in GJ are often heavily rainfall-limited, so that the simulated yields are relatively insensitive to C_{YG}. Because of this insensitivity to the time window used for optimisation, the C_{YG} set obtained from optimisation over 1966–1989 was defined as the calibrated set of values (rather than performing cross-validation).

Transpiration efficiency, E_T, is the key parameter which links the water budget of the system to biomass. The stability of the optimal transpiration efficiency across space and input data is illustrated for the Medium soil subclass in Fig. 3. Use of the coarse and fine subclasses produced very similar results (the same optimal parameter values and e.g. $r^2 > 0.9997$ for all three pairs of time series for GJ) and so were not plotted. The extinction coefficient, k, is a key parameter in determining the energy available for transpiration, and its stability across space and input data is presented for the medium soil subclass in Fig. 4. Coarse and fine subclasses again produced very similar results and so are not plotted. The optimal values of all the model parameters were found to be within the ranges suggested by the literature (Appendix A). The optimal transpiration efficiency, E_T (Fig. 3) is nearer to the values observed by Kakani (2001) (e.g. 1.3 Pa) in a field experiment in Hyderabad, India, than the values (up to 4 Pa) reported for experiments in Queensland, Australia (Chapman et al., 1993; Wright et al., 1988). These latter values were for fertilised crop, and included root biomass in the calculation, which may explain why they are higher.

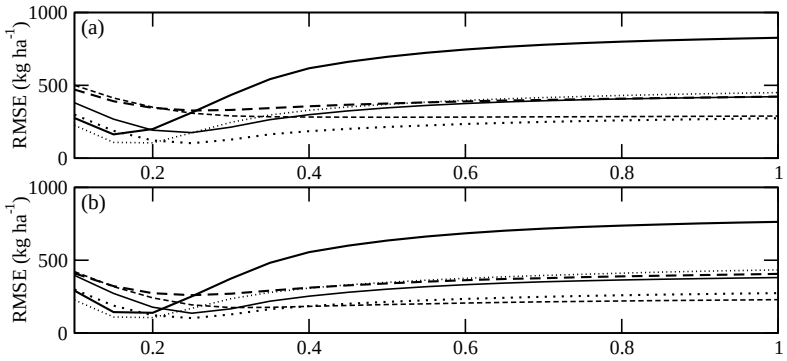

Fig. 2. Root mean square error as a function of the yield gap parameter, C_{YG}, for three grid cells—UP (solid line), GJ (dashed line) and AP (dotted line)—and using two time periods—(a) 1966–1989 and (b) 1966–1977. Using the 1978–1989 time period (not shown) produces very similar results. The MMT (thin lines) run uses input maximum and minimum daily temperature, and the VAP (thick lines) run uses input vapour pressure deficit.

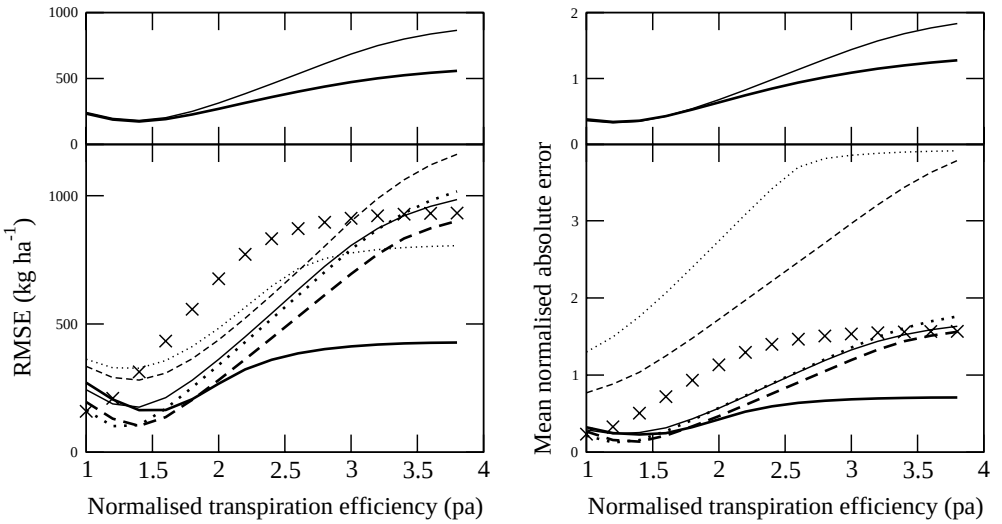

Fig. 3. Two measures of model errors for a range of values of transpiration efficiency. Lower panels: UP (solid line), GJ (dashed line) and AP (dotted line) grid cells for both the MMT (thin lines) and VAP (thick lines) runs. Also shown is the VAP UP run with MMT-optimal values of C_{YG} (crosses). Upper panels: ALLIN runs for MMT data (thin lines) and VAP data (thick lines).

Fig. 5 shows the variation in $(\partial H_I/\partial t)$ (a key model parameter in linking biomass to yield) when using different sections of the available time period. In addition, two detrending methods are used. The optimal value is stable over time if the baseline period is kept constant (i.e. a single linear detrend over 1966–1989). However, if two piece-wise regressions are used, the optimal values are not constant. This effect can be seen with other parameters and other subdivisions also (not shown). It highlights a difficulty with making a system such as this operational; separating the trend from the variability is not trivial, particularly

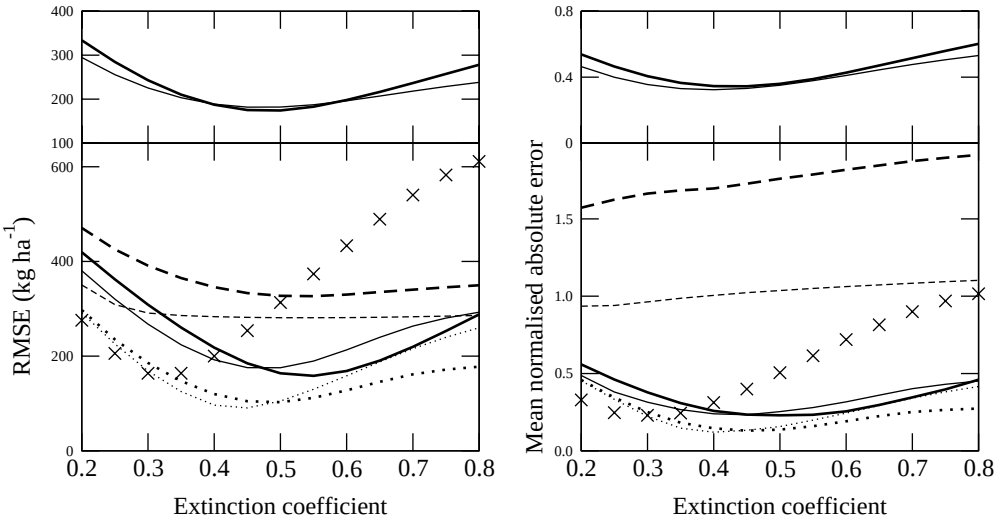

Fig. 4. Two measures of model errors for a range of values of the extinction coefficient (k). Lower panels: UP (solid line), GJ (dashed line) and AP (dotted line) grid cells for both MMT (thin lines) and VAP (thick lines) runs. Also shown is the VAP UP run with MMT-optimal values of C_{YG} (crosses). Upper panels: ALLIN runs for MMT data (thin lines) and VAP data (thick lines).

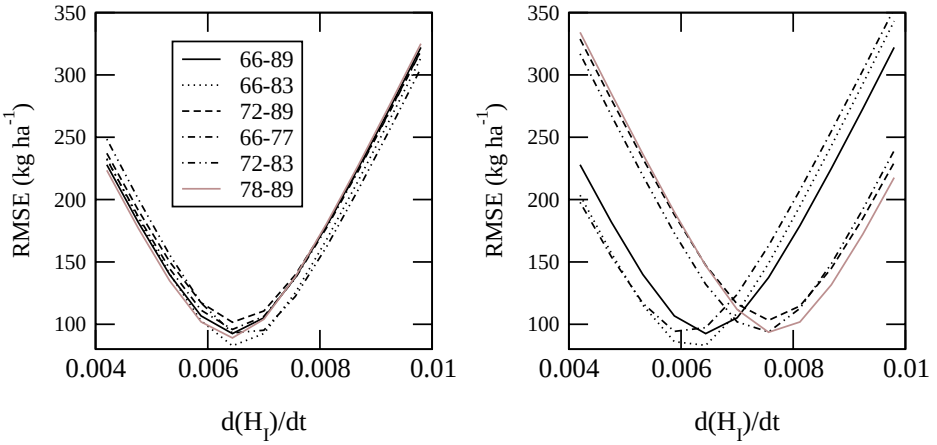

Fig. 5. Root mean square error for the period YR_s–YR_e for the AP grid cell using baseline period 1966–1989 (left panel) and two-piece linear regression (1966–YR_s), (YR_s–YR_e). YR_s and YR_e are given in the legend.

when one considers the potential effects of climate change.

Optimisation was also carried out for the five district-level yield time series within AP. In crop modelling terms, the spatial scale of districts is still considered to be 'large area'. The lack of smaller-scale weather data prevented an assessment of model output at this scale. However, the optimal value of global parameters remained invariant, so long as newly calibrated district values of C_{YG} were used.

4. Model evaluation

The performance of the model was evaluated using two methods. First, consistency checks: output diagnostic variables such as specific leaf area and radiation use efficiency were calculated (Section 4.1). Second, measured and modelled yield time series were compared across three different regions for different input data, using statistical weather–yield correlations as a benchmark (Section 4.2). All quoted correlations are for a 24-year time series, so that correlations above 0.41 are significant at the 5% level, and correlations above 0.52 are significant at the 1% level.

4.1. Assessment of model internal consistency

Some tests were carried out on the output from the model for all grid cells in India (ALLIN) using the optimal parameter set. Final harvest indices ranged from 0.34 to 0.44, and the final LAI for most runs was in the region 0.5–1.5 for Kharif runs, and 0.5–2.5 for irrigated runs.

In order to assess specific leaf area (SLA, the ratio of leaf area to dry leaf weight), MMT runs were carried out with $C_{YG} = 1$, since this gives an actual rather than effective LAI (Section 2.1), which is consistent with the biomass calculation. SLA was estimated using above-ground biomass and LAI; this results in a slight underestimation, since roots and stems are included in the calculation. The mean SLA for all the rainfed Kharif runs was 214 $cm^2\ g^{-1}$, with a standard deviation of 66 $cm^2\ g^{-1}$. The Rabi runs gave the slightly higher mean value of 227 $cm^2\ g^{-1}$, with a standard deviation of 25 $cm^2\ g^{-1}$. These values are comparable to measured values reported by Nautiyal et al. (2002) (144–241 $cm^2\ g^{-1}$) and Nigam et al. (2001) (123–206 $cm^2\ g^{-1}$).

End-of-season root length per unit leaf area had a mean (24 years) value of 79 cm^{-1} for the MMT Kharif runs. This is larger than the range of values ($\sim$10–60 cm^{-1}) published in Monteith et al. (1997) for various field experiments (see also Rao et al., 1989). However, the ALLIN runs with $C_{YG} = 1$, produced a more realistic average of 57 cm^{-1}, due to the higher LAI values achieved.

The radiation use efficiency (RUE, the ratio of above-ground dry weight to radiation intercepted) of irrigated MMT Kharif runs for UP, AP and GJ was

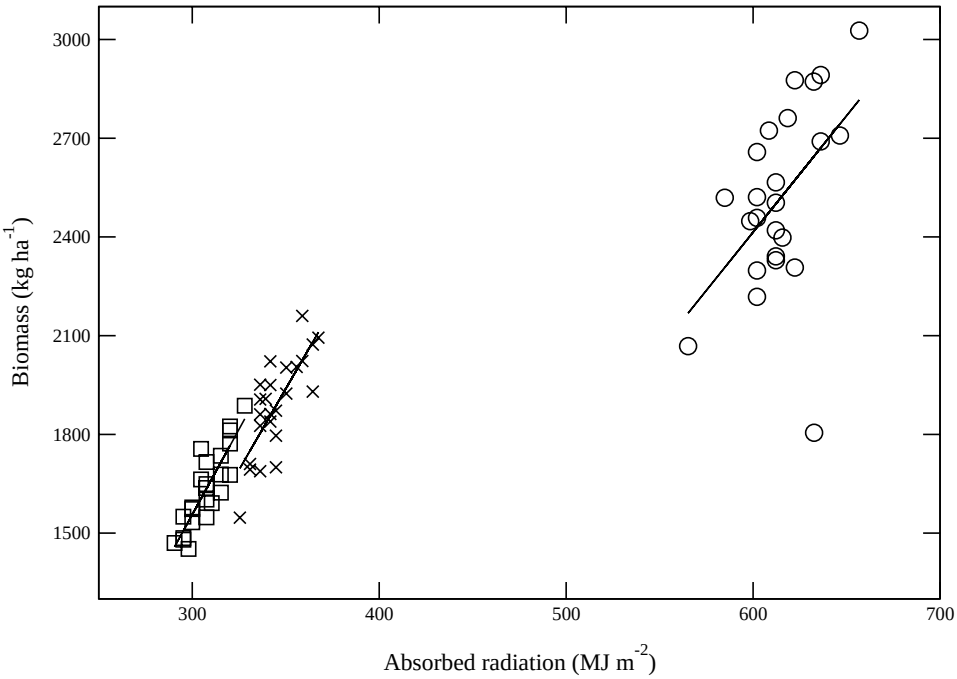

Fig. 6. End-of-season above-ground biomass vs. cumulative absorbed radiation for UP (crosses), GJ (circles) and AP (squares) grid cells. The regression lines shown lead to radiation use efficiency values of 0.99, 0.71 and 1.00 g MJ^{-1}, respectively. The correlation coefficients for the regressions are 0.77, 0.49 and 0.86, respectively. Note that intercepted (as opposed to absorbed) radiation is not explicitly modelled. The total (soil + crop) albedo is 0.2.

0.99, 1.00 and 0.71, respectively (Fig. 6). These values agree with the reported values of 0.74 g MJ^{-1} (Azam-Ali, 1998) and 1.0 g MJ^{-1} (used by Hammer et al. (1995) as a model parameter and based on the observed values of Bell et al. (1992) and Bennett et al. (1993)). RUE in the rainfed case is lower (0.83, 0.63 and 0.40, respectively).

4.2. Assessment of model skill

This section compares model output (using optimised model parameters) to measured yields. Firstly, model skill at the grid scale is assessed, together with the ability of GLAM to capture the effects of sub-seasonal weather variability on crop yield. Secondly, an assessment of model skill using larger spatial scales (up to all-India) is presented.

Differences between the yields simulated using the three different soil subclasses were relatively small (less than 5% change in RMSE, and no statistically significant differences in the correlation coefficient). Hence, in all cases, the medium soil subclass is used.

4.2.1. Model skill at the grid scale

Model skill can be assessed both in absolute terms and relative to statistical relationships. As well as mean skill it is important to consider the marginal benefit of the inclusion of the Rabi season yields, as well as variability in skill over time and across different environments (e.g. water-limited versus well-watered). It is also important to examine the skill of the model in translating sub-seasonal weather variability into crop yield impacts, since this is one of the objectives of GLAM. This section deals briefly with these issues.

The mean RMSE for all 35 grid cells was 214 kg ha^{-1} (with standard deviation 77 kg ha^{-1}) for the MMT run and 202 kg ha^{-1} (with standard deviation 51 kg ha^{-1}) for the VAP data. Most of the correlation between data and model comes from the rainfed Kharif runs, although for four cases the cor-

relation was higher with the inclusion of the Rabi season. Where irrigation is significant ($\gtrsim$50%), the inclusion of the Rabi fraction is essential (in minimising RMSE) as it removes the summer climate signal from the yields. For AP, the RMSE was similar whether the Rabi fraction was included or not (as long as the model was re-calibrated on the Kharif runs only). The Rabi yields are generally lower than the Kharif yields, since Rabi VPD is higher (see Eq. (6)). This is contrary to observations from southern India (Virmani and Shurpali, 1999) which show that Rabi yields tend to be higher than Kharif by a factor of approximately 1.4–2. Errors in the model simulation of Rabi yields may result from the use of an imposed humidity which fails to account for the modification of local humidity by crop evapotranspiration. Another potential source of error arises from the vapour pressure data: this is the mean of one or several daily measurements (New et al., 1999), which may or may not be consistent with the daily average temperature. Hence the non-linear relationship between temperature and e_{sat} can lead to errors in the estimation of VPD. Fixing the Rabi-season VPD such that yields are 1.4–2 times the Kharif yields and recalibrating C_{YG} did not lower the RMSE. Hence the inclusion of Rabi simulations where irrigation is not significant does not improve agreement with data.

The three observed and simulated yield time series are presented in Fig. 7. GJ has RMSE 281 kg ha^{-1} (44% of the mean detrended observed yields, and 68% of the observed interannual standard deviation, σ_y) and correlation coefficient $r = 0.74$. AP has RMSE 105 kg ha^{-1} (18% of the mean, and 106% of σ_y) and correlation coefficient $r = 0.42$. UP has RMSE 176 kg ha^{-1} (27% of the mean, 111% of σ_y) and correlation coefficient $r = 0.00$. These differences in r reflect differences in the strength of the climate signal. Higher correlations are associated with higher standard deviation, so that the impact of increased climate variability is a stronger and hence more predictable impact upon yield. This means that predictability is greatest where human vulnerability to yield variability is greatest.

When the skill of the model was evaluated for six different time periods (1966–1989, 1966–1983, 1972–1989, 1966–1977, 1972–1983 and 1978–1989) no statistically significant differences in correlation were found. In GJ, any time period including the years 1983–1989 produced a substantially higher

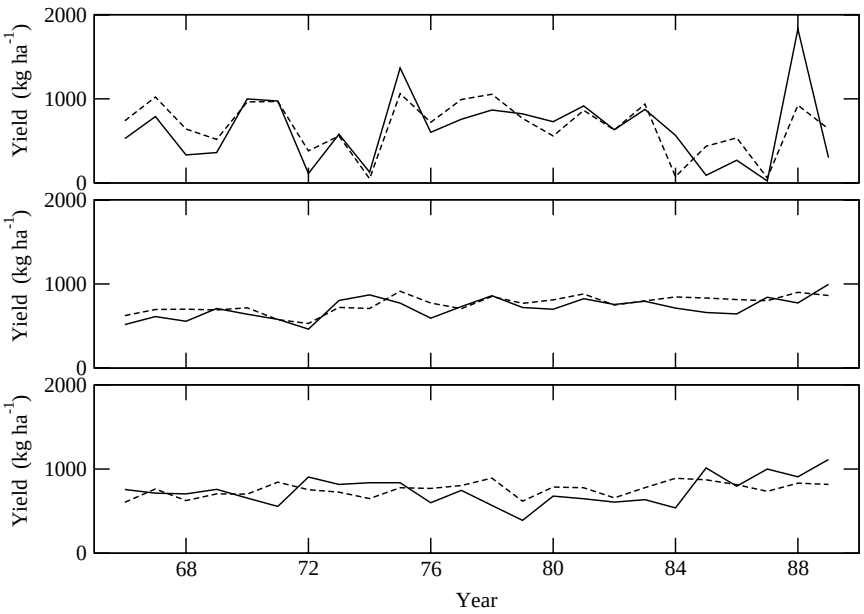

Fig. 7. Time series of observed (solid) and simulated (dashed) yields for the three grid cells—GJ (upper panel), AP (middle panel) and UP (lower panel). MMT data (daily maximum and minimum temperature) were the model inputs for these simulations.

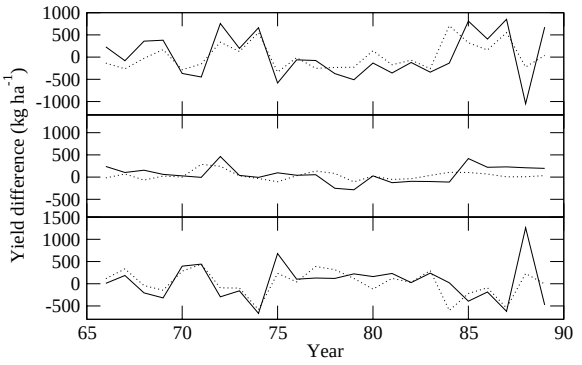

Fig. 8. Observed (solid lines) and simulated (dotted lines) values of the difference between yield at two grid cells: UP minus GJ (upper panel), UP minus AP (middle panel), and GJ minus AP (lower panel). MMT data were the model inputs for these simulations.

RMSE; e.g. 1966–1977 had RMSE $= 196\,\mathrm{kg\,ha^{-1}}$, and 1978–1989 had RMSE $= 346\,\mathrm{kg\,ha^{-1}}$.

As a simple test of spatial variability, Fig. 8 shows the skill of the model in reproducing yield differences between the three sites. Since GJ has much greater variability in yield than the other two sites, this signal dominates two of the difference plots. The correlation coefficient for each of the difference pairs (observed and simulated yields) is 0.69 (UP minus GJ), 0.30 (UP minus AP) and 0.67 (GJ minus AP). The difference is of the correct sign in 20, 16 and 22 cases out of 23, respectively.

A statistical model does not resolve biophysical processes and may not retain applicability under changing climates. However, a comparison between observed rainfall–yield correlations and model correlations provides a useful benchmark for the assessment of skill of the GLAM formulation. If the model can perform comparably to statistics using simple interpolated monthly means (except for rainfall, which is daily) then this is encouraging: using daily data, the model should perform better still. A Taylor diagram (Taylor, 2001) summarises and compares the performance of the model runs for the three sites using VAP and MMT data (Fig. 9). This diagram allows the simultaneous comparison of modelled and measured standard deviation, correlation and RMSE. Although GJ has the highest absolute RMSE, this diagram shows that when observed variability is taken into account, it is clearly the cell with the most skill. The model correlations are comparable to or better than those obtained from correlating June–September rainfall and yield. In two of the three cases the MMT runs perform better than the VAP runs. In all three cases the standard deviation of yield is lower than that observed.

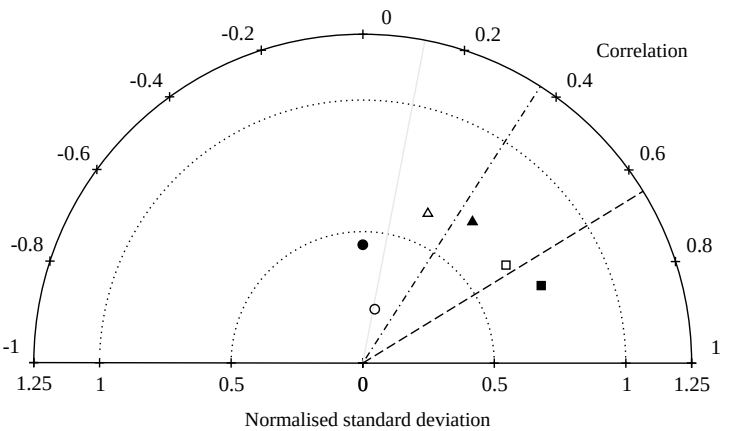

Fig. 9. Taylor diagram of 1966–1989 model output statistics using both MMT (filled) and VAP (open) data, for the UP (circles), GJ (squares) and AP (triangles) grid cells. Standard deviation has been normalised with respect to observations, so that the distance from any point to the point (1, 1) is proportional to the centred (i.e. non-systematic) RMSE in units of observed standard deviation. Also shown are the observed correlations between June and September rainfall and observed yield for UP (solid grey line), GJ (dashed line) and AP (dot-dashed line). A perfectly optimised statistical model based on rainfall alone would be represented by the perpendicular bisector of the point (1, 1) and the correlation line. (See Taylor, 2001, for the full theory behind this diagram.)

A more realistic standard deviation may be achieved by changing the soil type and/or sowing criteria, but only in the situations were the correlation coefficient is high will this lower the RMSE. Perhaps more importantly, higher temporal resolution in the weather inputs, and use of solar radiation data with inter-annual variability may result in greater variability in model output.

UP was the only site where rainfall was not often limiting yields. Here modelled yields were well-correlated with duration ($r = 0.82$) and VPD ($r = -0.68$). Measured yields in UP showed no correlation with VPD, and some correlation with modelled duration ($r = 0.43$). In order to attempt to capitalise on the RUE relationship which was explored in the last section, the solar radiation parameterisation of Ramkrishnan and Ritchie (2000), which uses T_{max} and T_{min} as a proxy for cloud cover, was used in place of the CRU data. It was thought that this may improve correlations in well-watered conditions, where radiation is the limiting factor. Using a re-calibrated value of $C_{YG} = 0.15$, the RMSE for the MMT run was slightly lower than with the CRU data (167 and 176 kg ha^{-1}, respectively). This indicates that the use of solar radiation data with a higher temporal resolution could improve model skill. Where such data are not available, the approach of Ramkrishnan and Ritchie (2000) may prove useful.

The impact of sub-seasonal weather variability on crop yields can be significant (Gadgil et al., 2002; Hansen and Jones, 2000). As an illustration of the importance of capturing the effects of sub-seasonal variability, consider the observed rainfall in GJ in the years 1975 and 1981 (Fig. 10). The total rainfall during the simulation period was similar in each case (394 and 389 mm, respectively), but because of the different sub-seasonal distributions, modelled yields in each case were different (1059 and 844 kg ha^{-1}, respectively). This difference is reflected in the observed yields (1360 and 901 kg ha^{-1}, respectively).

4.2.2. Model skill at larger scales

When the model output is aggregated over more that one grid cell (2.5° square), the RMSE generally falls. For example, aggregating GJ with the neighbouring grid cell to the southwest gives an RMSE of 232 kg ha^{-1}, compared with 281 and 256 for the two cells taken individually. The correlation between observed and simulated yields remains high ($r = 0.80$ compared with 0.74 and 0.69 for the individual cells). Part of the reason for the fall in RMSE upon aggregation is associated with a reduction of observed stan-

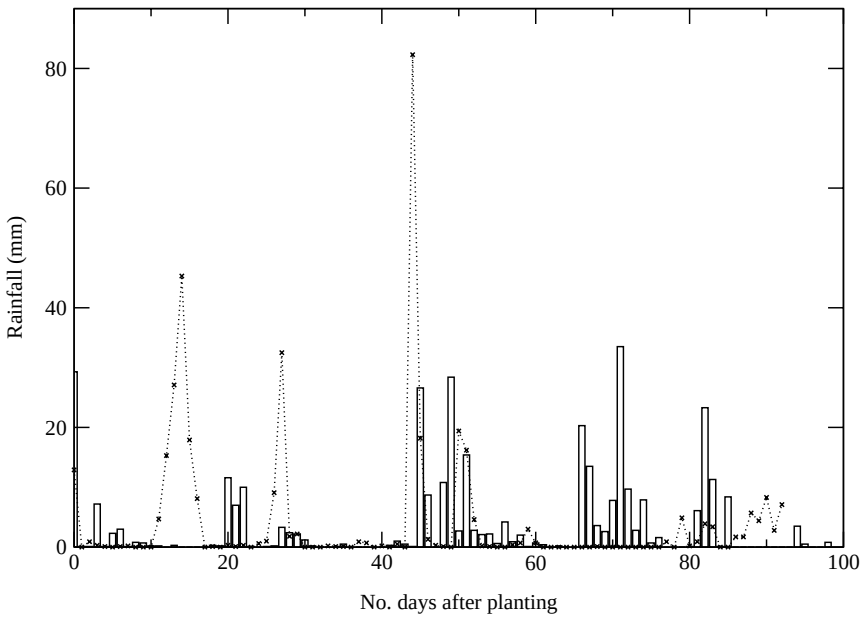

Fig. 10. Rainfall time series for grid cell GJ for 1975 (bars) and 1981 (crosses).

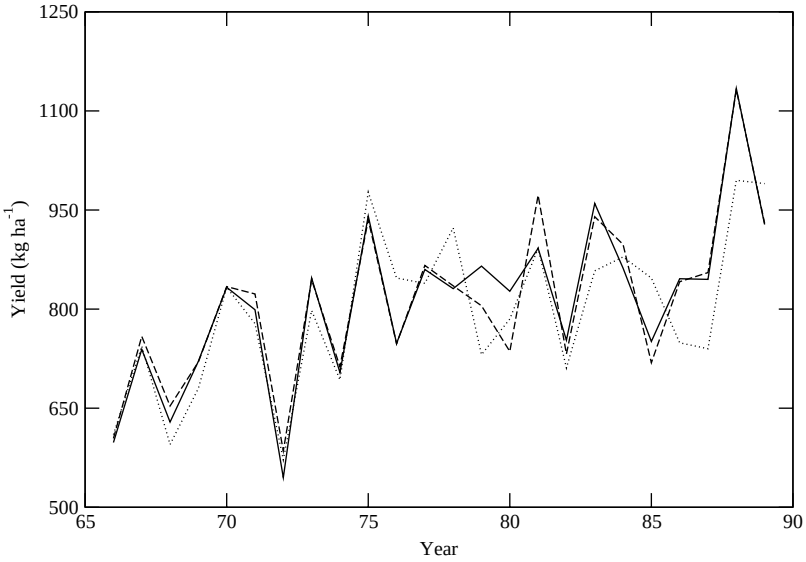

Fig. 11. Observed all-India yields from the Food and Agriculture Organisation (dashed line) and as calculated from ICRISAT data (solid line) together with GLAM–MMT all-India yields (dotted line). The latter were formed by adding the observed grid-scale trend to the modelled yields and aggregating (since simulation was for yields at 1966 levels).

dard deviation (which is 305 and 146 kg ha^{-1} for the individual sites, and 282 for the aggregated data).

Aggregating over all the grid cells allows a hindcast of the all-India yields. For this simulation, the values of C_{YG} were determined by two-fold cross-validation—i.e. the values determined by RMSE-minimisation for the first half of the time series (1966–1977) were used to determine yields in the second half (1978–1989) and vice versa. This method was introduced here since for some cells optimal values of C_{YG} were different in the first and second half of the time series. These small changes were significant in only a few cases. For example, for the MMT runs, use of C_{YG} values determined from the first or second half of the data set resulted in RMSE increases of over 5% in only one (first half) or two (second half) cells.

The all-India simulations are presented in Fig. 11. The correlation coefficient between observed and modelled yields is 0.76 for both data-sets, and the normalised standard deviation (simulated/observed) is 1.03 when ICRISAT observed yields are used and 0.99 when FAO data are used. The RMSE with respect the ICRISAT data is 69 kg ha^{-1}, and with respect to the FAO data it is 67 kg ha^{-1} (less than 9% of the mean detrended observed yields, and less than 74% of observed standard deviation). That the all-India results are better than the individual cell output is partly due to a cancellation of errors in simulated standard deviation of yield between the northwest and the southern peninsula: under-prediction in the GJ region, and over-prediction south of AP. Regional variation in model errors can be attributed to a number of factors, which fit into three broad categories: (i) weather and yield data inaccuracy, (ii) imperfect model response to different weather regimes, and (iii) temporal variability in management, pests and diseases, which is not accounted for by the model.

5. Summary and conclusions

The general large-area model for annual crops has been presented and optimised for the case of groundnut in India. The model is relatively simple, having 40 parameters, 5 of which vary spatially (3 soils parameters, planting date, and the yield gap parameter), 20 of which are crop-specific. When run in deterministic hindcast mode, model parameters are relatively stable over space and time. This stability breaks down when shorter time periods are used as the baseline for removal of the technology trend. In the context of the

prediction of inter-annual variability in crop yield under changing climates, this presents a challenge: the long-term (∼20 years) time trend must be well estimated.

GLAM output has been shown to have skill comparable to or greater than the skill of a statistical model. The skill varies with the strength of the climate signal and should be considered relative to the skill of other modelling options: high resolution, input-heavy modelling was not an option for our application. Aggregated all-India yields were simulated with an RMSE of less than 9%. The character of the results was independent of the two types of input data used. The model has the potential to resolve yield response to sub-seasonal climate variability, and a simple case study has explored this for the case of rainfall.

The model is suitable for use with large-scale input data such as gridded climate model output and re-analysis data, where an exploitable correlation exists between weather and yield on the scale in question (Challinor et al., 2003). The model can be used where more spatially detailed modelling followed by aggregation is not plausible (due to, for example, resource limitations or input data availability/quality). It could be used in the kind of hybrid modelling approach described by Hansen and Jones (2000).

Further development to include interactive coupling with a GCM (see, e.g. Tsvetsinskaya et al., 2001a,b) may reduce model errors by allowing the crop to modify its own environment through surface VPD feedbacks. This may be particularly important for irrigated crops grown during the dry season, when observed large-scale VPD may be higher than that found in the crop canopy.

This study is a first attempt to model yields over large areas using minimal data input and calibration. GLAM succeeds in its aim of capturing climate-induced variability; use of higher-frequency input climate data may improve results further by increasing modelled interannual standard deviations. The pragmatic use of process-based modelling techniques enables assessments of crop productivity under future climates which are less dependent upon statistical crop–weather relationships derived under the current climate.

Acknowledgements

This work has been funded by a fellowship from the Research Endowment Trust Fund of The University of Reading, UK. The authors wish to thank the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) and the Indian Institute for Tropical Meteorology (IITM) for the crop productivity and weather data, as well as Emma Tate and the Centre for Ecology and Hydrology for discussions and processing of the soils data. Model development and evaluation was aided by discussions with Michael Dingkuhn, Christian Baron and Benoit Sarr from CIRAD (Centre de coopération internationale en recherche agronomique pour le développement).

Appendix A. Values of parameters used in the model

The constants and parameters used in the model are listed below. Values shown outside brackets correspond to the listed source, whilst values in square brackets indicate the calibrated value (where this is different to the cited value). All parameters are global except for C_{YG}, which is determined locally

Parameter	Used in	Value(s)	Source
Growth and development			
C_{sow}	Section 2.1	0.5	Section 3.1
$\left(\frac{\partial L}{\partial t}\right)_{max}$	Eq. (3)	0.01–0.10 [0.1] per day	Section 3.1
S_{cr}	Eq. (3)	0.5–1.0 [0.7]	Section 3.1
$\frac{\partial l_v(z=0)}{\partial L}$	Eq. (5)	1	Simmonds and Azam-Ali (1989)
V_{EF}	Eq. (5)	1–2 [1] cm per day	Matthews et al. (1988) and Wright and Rao (1994)

Parameter	Used in	Value(s)	Source
$l_v(z = z_{ef})$	Eq. (5)	0.3	Simmonds and Azam-Ali (1989)
E_T	Eq. (6)	1.3–4.5 [1.4] Pa	Hammer et al. (1995), Chapman et al. (1993), Wright et al. (1988) and Kakani (2001)
$E_{TN,max}$	Eq. (6)	1.5–5 [3] g kg^{-1}	Section 3.1
$\frac{\partial H_I}{\partial t}$	Eq. (8)	0.007 per day	Hammer et al. (1995) and Wheeler et al. (1997)
t_{em}	Section 2.1	8 days	Section 2.1
T_b	Eq. (2)	8–12 [10] K	Ong (1986) and Mohamed (1984)
T_o	Eq. (2)	28–37 [28] K	Ong (1986) and Mohamed (1984)
T_m	Eq. (2)	40–50 [50] K	Ong (1986) and Mohamed (1984)
t_{TT_0}	Eq. (1)	350–400 [375] K day	Section 3.1
t_{TT_1}	Eq. (1)	310–330 [320] K day	Section 3.1
t_{TT_2}	Eq. (1)	200–300 [250] K day	Section 3.1
t_{TT_3}	Eq. (1)	500–750 [620] K day	Section 3.1
Evaporation and transpiration			
L_{cr}	Eq. (11)	0.7	Azam-Ali (1984)
T_{Tmax}	Eq. (11)	0.15–0.4 [0.30] cm per day	Azam-Ali (1984)
α_0	Eq. (13)	1.26	Priestly and Taylor (1972)
V_{ref}	Eq. (13)	1 kPa	Steiner et al. (1991)
A	Section 2.3	0.2	Monteith and Unsworth (1990)
C_G	Eq. (15)	0.4	Choudhury et al. (1987)
k	Eq. (15)	0.2–0.8 [0.5]	Choudhury et al. (1987) and Hammer et al. (1995)
P_{cr}	Eq. (17)	0.1 cm	Section 2.3
k_{DIF}	Eq. (20)	0.06 cm^2 per day	Dardanelli et al. (1997)
C_θ	Eq. (21)	0.5	Allen et al. (1998)
Soil submodel and miscellaneous			
z_{max}	Section 2.2	210 cm	Section 2.2
N_{SL}	Section 2.2	25	Section 2.2
C_{d1}	Eq. (10)	2.96 per day	Suleiman (1999)
C_{d2}	Eq. (10)	−2.62 per day	Suleiman (1999)
C_{d3}	Eq. (10)	0.85 per day	Suleiman (1999)
K_{ks}	Eq. (10)	37 cm per day	Suleiman and Ritchie (2001) and Suleiman (1999)
θ_s	Eqs. (10)	θ_{ll}	Section 3.1
z_{ed}	Section 2.2	16.8 cm	Section 2.2
C_V	Eq. (7)	0.7	Tanner and Sinclair (1983)
C_{YG}	Eq. (3)	0.1–1.0	Section 3.1

Values of the soil parameters used in the model are listed below. Each soil class (listed by row) has three sub-classes (coarse, medium and fine) which correspond to different water holding capacities

	Coarse			Medium			Fine		
	θ_{ll}	θ_{dul}	θ_{sat}	θ_{ll}	θ_{dul}	θ_{sat}	θ_{ll}	θ_{dul}	θ_{sat}
Silt loam	0.10	0.23	0.46	0.10	0.26	0.46	0.10	0.30	0.46
Clay loam	0.13	0.26	0.48	0.13	0.28	0.48	0.13	0.31	0.48

References

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop evapotranspiration. Guidelines for computing crop water requirements. FAO Irrigation and Drainage 56. FAO, Viale delle Terme di Caracalla, Rome, Italy.

Arya, S.P., 1988. Introduction to Micrometeorology. Academic Press, San Diego, CA.

Azam-Ali, S.N., 1984. Environmental and physiological control of transpiration by groundnut crops. Agric. For. Meteorol. 33, 129–140.

Azam-Ali, S.N., 1998. Population, growth and water use of groundnut maintained on stored water. III. Dry matter, water use and light interception. Exp. Agric. 25, 77–86.

Basso, B., Ritchie, J.T., Pierce, F.J., Braga, R.P., Jones, J.W., 2001. Spatial validation of crop models for precision agriculture. Agric. Syst. 68, 97–112.

Bell, M.J., Wright, G.C., Hammer, G.L., 1992. Night temperature effects radiation use efficiency in peanuts. Crop Sci. 32, 1329–1335.

Bennett, J.M., Sinclair, T.R., Ma, L., Boote, K.J., 1993. Single leaf carbon exchange and canopy radiation use efficiency of four peanut cultivars. Peanut Sci. 20, 1–5.

Bolton, D., 1980. The computation of equivalent potential temperature. Monthly Weather Rev. 108, 1046–1053.

Boote, K.J., Jones, J.W., 1998. Simulation of crop growth: CROPGRO model. In: Peart, R.M., Curry, R.B. (Eds.), Agricultural Systems Modeling and Simulation. Marcel Dekker, New York, Chapter 18, pp. 651–692.

Brooks, R.J., Semanov, M.A., Jamieson, P.D., 2001. Simplifying sirus: sensitivity analysis and development of a meta-model for wheat yield prediction. Eur. J. Agron. 14, 43–60.

Challinor, A.J., Slingo, J.M., Wheeler, T.R., Craufurd, P.Q., Grimes, D.I.F., 2003. Towards a combined seasonal weather and crop productivity forecasting system: determination of the spatial correlation scale. J. Appl. Meteorol. 42, 175–192.

Chapman, S.C., Ludlow, M.M., Blamey, F.P.C., Fischer, K.S., 1993. Effect of drought during early reproductive development on growth of groundnut (*Arachis hypogaea* L.). I. Utilization of radiation and water during drought. Field Crops Res. 32, 193–210.

Choudhury, B.J., DiGirolamo, N.E., 1998. A biophysical process-based estimate of global land surface evaporation using satellite and ancillary data. I. Model description and comparison with observations. J. Hydrol. 205, 164–185.

Choudhury, B.J., Idso, S.B., Reginato, R.J., 1987. Analysis of an empirical model for soil heat flux under a growing wheat crop for estimating evaporation by an infrared temperature based energy balance equation. Agric. For. Meteorol. 39 (4), 283–297.

Cooper, P.J.M., Keatinge, J.D.H., Hughes, G., 1983. Crop evapotranspiration—a technique for calculation of its components by field measurements. Field Crops Res. 7, 299–312.

Dardanelli, J.L., Bachmeier, O.A., Sereno, R., Gil, R., 1997. Rooting depth and soil water extraction patterns of different crops in a silty loam Haplustoll. Field Crops Res. 54, 29–38.

Doorenbos, J., Kassam, A.H., 1979. Yield response to water. FAO Irrigation and Drainage 33. FAO, Viale delle Terme di Caracalla, Rome, Italy.

FAO/UNESCO, 1974. FAO/UNESCO soil map of the world, 1:5,000,000, ten volumes.

Fischer, G., Shah, M., van Velthuizen, H., 2002. Climate change and agricultural vulnerability. Technical Report. International Institute for Applied Systems Analysis. Available at http://www.iiasa.ac.at/Research/LUC/.

Gadgil, S., Rao, P.R.S., Rao, K.N., 2002. Use of climate information for farm-level decision making: rainfed groundnut in southern India. Agric. Syst. 74, 431–457.

Gadgil, S., Rao, P.R.S., Sridhar, S., 1999. Modelling impact of climate variability on rainfed groundnut. Curr. Sci. 76 (4), 557–569.

Guerif, M., Duke, C.L., 2000. Adjustment procedures of a crop model to the site specific characteristics of soil and crop using remote sensing data assimilation. Agric. Ecosyst. Environ. 81 (1), 57–69.

Hammer, G.L., Sinclair, T.R., Boote, K.J., Wright, G.C., Meinke, H., Bell, M.J., 1995. A peanut simulation model. I. Model development and testing. Agron. J. 87, 1085–1093.

Hansen, J.W., 2002. Realizing the potential benefits of climate prediction to agriculture: issues, approaches, challenges. Agric. Syst. 74, 309–330.

Hansen, J.W., Jones, J.W., 2000. Scaling-up crop models for climatic variability applications. Agric. Syst. 65, 43–72.

Harrison, P.A., Porter, J.R., Downing, T.E., 2000. Scaling-up the AFRCWHEAT2 model to assess phenological development for wheat in Europe. Agric. For. Meteorol. 101, 167–186.

IRI, 2000. Summary report of the workshop: linking climate prediction model out-put with crop model requirements. Technical Report IRI-CW/00/2. International Research Insti-

tute for Climate Prediction, Palisades, New York. Available from http://iri.columbia.edu/outreach/publication/irireport/workshop2000.pdf.

Jagtap, S.S., Jones, J.W., 2002. Adaptation and evaluation of the CROPGRO-soybean model to predict regional yield and production. Agric. Ecosyst. Environ. 93, 73–85.

Jamieson, P.D., Porter, J.R., Goudriann, J., Ritchie, J.T., van Keulen, H., Stol, W., 1998. A comparison of the models AFRCWHEAT2, CERES-Wheat, Sirius, SUCROS2 and SWHEAT with measurements from wheat grown under drought. Field Crops Res. 55, 23–44.

Jones, D., Barnes, E.M., 2000. Fuzzy composite programming to combine remote sensing and crop models for decision support in precision crop management. Agric. Syst. 65 (3), 137–158.

Jury, W.A., Tanner, C.B., 1975. Advection modification of the Priestly and Taylor evapotranspiration formula. Agron. J. 67 (6), 840–842.

Kakani, V.G., 2001. Quantifying the effects of high temperature and water stress in Groundnut. Ph.D. Thesis. University of Reading, UK.

Kaur, P., Hundal, S.S., 1999. Forecasting growth and yield of groundnut (*Arachis hypogaea*) with a dynamic simulation model 'PNUTGRO' under Punjab conditions. J. Agric. Sci. 133, 167–173.

Landau, S., Mitchell, R.A.C., Barnett, V., Colls, J.J., Craigon, J., Payne, R.W., 2000. A parsimonious, multiple-regression model of wheat yield response to environment. Agric. For. Meteorol. 101, 151–166.

Martin, R.V., Washington, R., Downing, T.E., 2000. Seasonal maize forecasting for South Africa and Zimbabwe derived from an agroclimatological model. J. Appl. Meteorol. 39 (9), 1473–1479.

Matthews, R.B., Harris, D., Rao, R.C.N., Williams, J.H., Wadia, K.D.R., 1988. The physiological basis for yield differences between four genotypes of groundnut (*Arachis hypogaea*) in response to drought. I. Dry matter production and water use. Exp. Agric. 24, 191–202.

Mearns, L.O., Easterling, W., Hays, C., Marx, D., 2001. Comparison of agricultural impacts of climate change calculated from the high and low resolution climate change scenarios. Part I. The uncertainty due to spatial scale. Clim. Change 51, 131–172.

Mearns, L.O., Mavromatis, T., Tsvetsinskaya, E., 1999. Comparative response of EPIC and CERES crop models to high and low spatial resolution climate change scenarios. J. Geophys. Res. 104 (D6), 6623–6646.

Mohamed, H.A., 1984. Varietal differences in the temperature responses of germination and crop establishment. Ph.D. Thesis. University of Nottingham, UK.

Monteith, J., Gallagher, J.N., Gregory, P.J., et al., 1997. Microclimatology in tropical agriculture. Final Report Research Schemes R3208 and R3819. The University of Nottingham. Funded by the Overseas Development Agency, Eland house, Stag Place, London.

Monteith, J.L., Unsworth, M.H., 1990. Principles of Environmental Physics, 2nd ed. Edward Arnold, London, 291 pp.

Nautiyal, P.C., Rachaputi, N.R., Joshi, Y.C., 2002. Moisture-deficit-induced changes in leaf-water content, leaf carbon exchange rate and biomass production in groundnut cultivars differing in specific leaf area. Field Crops Res. 74, 67–79.

New, M., Hulme, M., Jones, P., 1999. Representing twentieth-century space–time climate variability. Part I. Development of a 1961–1990 mean monthly terrestrial climatology. J. Clim. 12, 829–856.

New, M., Hulme, M., Jones, P., 2000. Representing twentieth-century space–time climate variability. Part II. Development of 1901–1996 monthly grids of terrestrial surface climate. J. Clim. 13, 2217–2238.

Nigam, S.N., Upadhyaya, H.D., Chandra, S., Rao, R.C.N., Wright, G.C., Reddy, A.G.S., 2001. Gene effects for specific leaf area and harvest index in three crosses of groundnut (*Arachis hypogaea*). Ann. Appl. Biol. 139 (3), 301–306.

Ong, C.K., 1986. Agroclimatological factors affecting phenology of groundnut. In: Proceedings of the International Symposium on Agrometeorolgy of Groundnut, ICRISAT Sahelian Center, Niamey, Niger, pp. 115–125.

Pant, G.B., Kumar, K.R., 1997. Climates of South Asia. Wiley, Chichester, 320 pp.

Parthasarathy, B., Kumar, K.R., Munot, A.A., 1992. Forecast of rainy-season foodgrain production based on monsoon rainfall. Indian J. Agric. Sci. 62, 1–8.

Passioura, J.B., 1983. Roots and drought resistance. Agric. Water Manage. 7, 265–280.

Priestly, C.H.B., Taylor, R.J., 1972. On the assessment of surface heat flux and evaporation using large-scale parameters. Monthly Weather Rev. 100 (2), 81–92.

Ramkrishnan, P., Ritchie, J., 2000. Estimation of solar radiation for India. In: Poster Presented in ASA Meeting, Minneapolis. Available at http://nowlin.css.msu.edu/ConfNov2000/posters/POSTEROK.PDF.

Rao, R.C.N., Simmonds, L.P., Azam-Ali, S.N., Williams, J.H., 1989. Population, growth and water use of groundnut maintained on stored water. I. Root and shoot growth. Exp. Agric. 25, 51–61.

Reddy, P.S. (Ed.), 1988. Groundnut. Indian Council of Agricultural Research, Krishi Anusandhan Bhavan, Pusa, New Delhi, India.

Simmonds, L.P., Azam-Ali, S.N., 1989. Population, growth and water use of groundnut maintained on stored water. IV. The influence of population on water supply and demand. Exp. Agric. 25, 87–98.

Sinclair, T.R., Seligman, N., 2000. Criteria for publishing papers on crop modelling. Field Crops Res. 68, 165–172.

Steiner, J.L., Howell, T.A., Schneider, A.D., 1991. Lysimetric evaluation of daily potential evapotranspiration models for grain sorghum. Agron. J. 83, 240–247.

Suleiman, A.A., 1999. Assessing and modeling the spatial variability of soil water redistribution and wheat yield along a sloping landscape. Ph.D. Thesis. Crop and Soil Sciences Department, Michigan State University.

Suleiman, A.A., Ritchie, J.T., 2001. Estimating saturated hydraulic conductivity from soil porosity. Transactions of the ASAE 44 (2), 235–239; MAR–APR 2001.

Tanner, C.B., Sinclair, T.R., 1983. Efficient water use in crop production: Research or research? In: Taylor, H.M., Jordan, W.R., Sinclair, T.R. (Eds.), Limitations to Efficient Water Use

in Crop Production. ASA, CSSA, and SSA, Madison, WI, pp. 1–27.

Tate, E., 2001. Personal Communication. Centre for Ecology and Hydrology, Maclean Building, Crowmarsh Gifford, Wallingford, Oxfordshire, UK.

Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. J. Geophys. Res. 106 (D7), 7183–7192.

Tsvetsinskaya, E.A., Mearns, L.O., Easterling, W.E., 2001a. Investigating the effect of seasonal plant growth and development in three-dimensional atmospheric simulations. Part I. Simulation of surface fluxes over the growing season. J. Clim. 14 (5), 692–709.

Tsvetsinskaya, E.A., Mearns, L.O., Easterling, W.E., 2001b. Investigating the effect of seasonal plant growth and development in three-dimensional atmospheric simulations. Part II. Atmospheric response to crop growth and development. J. Clim. 14 (5), 711–729.

USDA-SCS, 1964. National Engineering Handbook, Section 4, Part I, Watershed Planning. Soil Conservation Service, US Department of Agriculture, Washington, DC.

Virmani, S.M., Shurpali, N.J., 1999. Climate prediction for sustainable production of rainfed groundnuts in SAT. Crop establishment risks in the Anantapur region. Technical Manual No. 4. International Crop Research Institute for the Semi-Arid Tropics, Patancheru, Andhra Pradesh, India.

Wang, E., Robertson, M.J., Hammer, G.L., Carberry, P.S., Holzworth, D., Meinke, H., Chapman, J.N.G., Chapman, J.N.G., Hargreaves, J.N.G., Huth, N.I., McLean, G., 2002. Development of a generic crop model template in the cropping system model APSIM. Eur. J. Agron. 18, 121–140.

Wheeler, T.R., Chatzialioglou, A., Craufurd, P.Q., Ellis, R.H., Summerfield, R.J., 1997. Dry matter partitioning in groundnut exposed to high temperature stress. Crop Sci. 37, 1507–1513.

Wheeler, T.R., Craufurd, P.Q., Ellis, R.H., Porter, J.R., Prasad, P.V.V., 2000. Temperature variability and the annual yield of crops. Agric. Ecosyst. Environ. 82, 159–167.

Wright, G.C., Hubick, K.T., Farquhar, G.D., 1988. Discrimination in carbon isotopes of leaves correlates with water-use efficiency of field-grown peanut cultivars. Aust. J. Plant Physiol. 15, 815–825.

Wright, G.C., Rao, R.C.N., 1994. Groundnut water relations. In: Smartt, J. (Ed.), The Groundnut Crop: A Scientific Basis for Improvement. Chapman & Hall, London, pp. 281–335.