

# Sensitivity analysis and auto-calibration of ORYZA2000 using simulation-optimization framework

B. Soundharajan · K. P. Sudheer

Received: 18 October 2010/Revised: 23 July 2011/Accepted: 11 September 2011/Published online: 26 October 2011  
© Springer-Verlag 2011

**Abstract** There has been an increasing interest to employ crop growth simulation models for taking decision on irrigation water management. The effectiveness of such decisions mainly lies on the efficiency of the model in simulating the crop growth and the yield, which are influenced by the value of the parameters of the model. Therefore, calibration of such models is necessary before it can be employed for any application. This study proposes an auto-calibration procedure for ORYZA2000, a rice crop growth simulation model, for its application in South India. The data employed for calibration is taken from a field experiment conducted for 2 years in an experimental farm in South India. The ORYZA2000 model was integrated within Genetic Algorithm optimizer, which calls the simulator during each generation to evaluate the objective function. The auto-calibrated model was tested for its performance using a validation data set from the same experimental data. The results showed that the calibrated ORYZA2000 model is capable of simulating the full irrigation and water stress condition of rice crop effectively, and can be used to develop deficit irrigation management schedules.

**Keywords** Rice crop · Crop growth model · ORYZA2000 · Genetic algorithm · Parameter estimation

## Introduction

Agriculture is the world's largest user of water in terms of irrigated volume. The growing competition for water

between the agricultural and non-agricultural sectors has increased the concern for sustainability of the irrigated agricultural system. The need for increasing agricultural production demands an increase in the cropped area regardless of the availability of water resources for irrigation; thereby make the water a limiting factor in many of the irrigation projects. The great challenge for the coming decades will be the task of increasing food production with less water, particularly in countries with limited water and land resources. Therefore, improved water management practices, especially methods to improve the water use efficiency, are warranted in the irrigation sector.

The per capita water availability in India is declining continuously, and is likely to reach the stress/scarcity levels in some regions within the next few years. This has led to injudicious abstraction of surface and ground water resulting in several problems including rapidly declining water table levels and salt water intrusion in coastal areas. The increased frequency of extreme events (especially drought) may further lead to unavailability of water to meet current irrigation demands. The strategy for water allocation and use of irrigation water in the south Indian systems has been neither demand based nor supply based (Mohanakrishnan 1990). Seasonal and in-seasonal allocations depend on the resources available, crop seasons, and priorities for water allocations prevailing in the particular location of the system. The irrigation engineers make the allocation decisions according to the 'water duty' specified at the heads of the main and branch/distributary canals in the command area (Pundarikanthan and Santhi 1996). While this kind of water allocation is ideal in situations where sufficient amount of water is available, the current water availability in the country, which is a deficit situation, warrants an optimal water management strategy to improve the current food production.

B. Soundharajan · K. P. Sudheer (✉)  
Indian Institute of Technology Madras, Chennai 600036, India  
e-mail: sudheer@iitm.ac.in

Optimal irrigation scheduling under deficit conditions is highly complex since it depends on the interaction of physical constraints of the irrigation system, soil moisture availability at the time of irrigation, growth stage of the crop, effect of previous and subsequent irrigations on crop growth and yield, and nature of weather conditions (Soundharajan and Sudheer 2009). Traditionally, yield reduction models based upon evapotranspiration (ET) ratios (e.g., Doorenbos and Kassam 1979; Jensen 1968) have been employed by many researchers for deficit irrigation management (Vedula and Mujumdar 1992; Kumar et al. 2006). However, these models have two major limitations: (i) they cannot provide crop yield in absolute terms and (ii) they do not have endogenous optimization capacity (Brumelow and Georgakakos 2007). Yet another concern is that most of the ET ratio based yield reduction models consider crop yield reduction as a linear function of crop ET within a stage of growth. Researchers have integrated optimization schemes with ET ratio based yield reduction models to arrive at optimal deficit irrigation schedules (Rao et al. 1988; Paul et al. 2000; Prasad et al. 2006). These methods employ crop water production functions on different crop growth stages, which can optimize the total water requirement for different stages of crop growth. Therefore, identification of the timing of irrigation application becomes difficult.

Recent developments in simulation of bio-physical models have given the opportunities for simulating the crop growth and development in the field conditions. Adequately calibrated and validated crop growth simulation models provide a systems approach and a fast alternative method for developing and evaluating agronomic practices that can utilize technological advances in limited irrigation agriculture (Saseendran et al. 2008). A few researchers have employed crop growth simulation models for irrigation scheduling (Rao and Rees 1992; Talpaz and Mjelde 1988; Soundharajan and Sudheer 2009) and the results were encouraging. Nonetheless, the accuracy of process oriented crop growth models depends on conceptual representation of physiological processes and parameter values used in the mathematical representation (Zhai et al. 2004).

Many of the currently used crop growth models are highly complex and are generally characterized by a multitude of parameters (Varella et al. 2010). Due to the variability in agro-climatic zones and the specific cultivars, the value of many of these parameters will not be exactly known. Further, many of them may not be directly measurable. Therefore, in most cases model calibration is necessary. Model calibration helps reduce the parameter uncertainty, which in turn reduces the uncertainty in the simulated results. During a model calibration, selected parameters are allowed to vary within predefined bounds,

until a sufficient correspondence between the model outputs and actual measurements are obtained. The actual measurements for calibration of crop growth models come from the field level experiments. However, generally the experimental data may not be long enough, for accurate estimation of model parameters, because experimentation on crop systems is necessarily lengthy and expensive in terms of land, equipment, and manpower. In addition, when the number of parameters in a model is large (either due to large number of sub-processes being considered or due to the model structure itself) the calibration process becomes complex and computationally extensive (Cibin et al. 2010). In such cases, sensitivity analysis (SA) is helpful to identify and rank parameters that have significant impact on specific model outputs of interest (Saltelli et al. 2000). In general, SA is employed prior to the calibration process in order to identify a candidate set of important factors for calibration so that complexity of calibration process can be reduced.

The objective of the current study is twofolds: (i) to carry out global SA (GSA) of a rice crop growth simulation model, ORYZA2000 (Bouman et al. 2001) and (ii) to develop and demonstrate an auto-calibration procedure for the ORYZA2000 model for applications in South India. The parameters of ORYZA2000 model are optimized and validated using the data collected from field experiments conducted in three seasons in an agricultural research farm. The remainder of the article is organized as follows. Following this introduction, a brief description about the models and optimization algorithm are presented. Also, the method of SA employed in this study is described in detail. Subsequently the experimental set up and the field data employed in this study are discussed, followed by the discussion on results of the SA and the auto-calibration.

## Materials and methods

In this study, the optimal values for the parameters of ORYZA2000 model were estimated by integrating the model within an optimization algorithm, thereby facilitating an auto-calibration framework. The framework used the data collected from field experiments conducted in three seasons and different treatments. The details of the field experiments, crop growth model, optimizer, and the auto-calibration framework are discussed in the following sections.

### Crop growth simulation model—ORYZA2000

ORYZA2000 is an eco-physiological crop growth model that simulates the growth, development, and water balance of rice in situations of potential, water-limited and

nitrogen-limited conditions on a daily basis. While there are a few other crop growth models for rice that are available (e.g., RICEMODE (McMennary and O'Toole 1985), WOFOST (Boogaard et al. 1998)), the ORYZA2000 has been extensively used and tested for its efficiency in water limited conditions, and the results were encouraging (Belder et al. 2004, 2007; Feng et al. 2007; Arora 2006), and therefore is considered in the current study.

In ORYZA2000 model, several modules such as for aboveground crop growth, ET, nitrogen dynamics, and soil–water balance, etc., are combined to simulate the crop growth under different production conditions. The water dynamics are simulated by a one-dimensional multi-layer soil water balance module, which can simulate soil water balance for different growing conditions. The model follows a daily computation scheme for the rate of dry matter production of the plant organs and phenological development. By integrating these rates over time, dry matter production and development stage are simulated throughout the growing season (Bouman and van Laar 2006). Daily dry matter production is related to net radiation, temperature, and leaf area index (LAI). The carbohydrates produced during the crop growth are shared among roots, leaves, stems, and panicles using partitioning factors along the development stages, based on daily heat units and photoperiod. From flowering onwards, leaf loss rate is simulated from an experimentally derived loss rate factor, which is a function of development stage and green leaf biomass (BM). A detailed explanation of the model along with the program source code is given in Bouman et al. (2001) (also available at [www.knowledgebank.irri.org/oryzabeta](http://www.knowledgebank.irri.org/oryzabeta)). The ORYZA2000 model assumes that the crop is well protected against diseases, pests, and weeds, and consequently the model does not consider the yield reduction due to these factors.

In order to simulate the crop growth, the model requires inputs of management practices, soil properties, and weather data in addition to crop parameters. Except the crop parameters, all of these inputs can be directly obtained from field experiments. The required management practices are crop variety, spacing or plant population, transplanting depth, nursery duration, and fertilizer and irrigation application. Soil properties required are volumetric soil water content at saturation, field capacity (FC) and wilting point and corresponding soil water potential, depth of puddled soil, and saturated hydraulic conductivity of the soil. The weather data include the rainfall and temperature during the growing season. The crop parameters include phenological development parameters and many other parameters related to the process of crop growth, and most of them can be obtained from literature. However, the cultivar specific parameters such as development rates, partitioning factors, relative leaf growth rate,

specific leaf area, and leaf death rate are to be calibrated using experimental data (Bouman et al. 2001).

The parameters of the ORYZA2000 that are to be calibrated, specifically for the variety and environment under consideration, were identified from Bouman et al. (2001). A list of these parameters along with their recommended range of values is presented in Table 1. These parameters take different values at different development stage (DVS). In Table 1, the DVS is represented in terms of fraction of heat unit along the crop growth, where a value of DVS equal to 2.0 represents maturity. The parameters RGRLMX and RGRLMN are maximum and minimum relative leaf growth rate, calculated from the daily increase in temperature sum, which controls the leaf area growth. SLATB is the specific leaf area, calculated as a function of development stage (DVS), which in turn is used to calculate the leaf growth during linear phase. During the linear phase of the leaf growth (up to DVS = 0.65), there is a fixed relation between leaf weight and LAI. LRSTR is the stem reserve available for growth after the respiration and growth BM losses. FLVTB, FSTTB, and FSOTB are fraction of shoot BM partitioned to leaves, stems and storage organ, respectively. Till panicle initiation (DVS = 0.65), BM is partitioned to leaves and stems only, as there is no storage organ component during this phase of crop growth. During the growth stage between panicle initiation and flowering, all the three parameters are active so that the BM is partitioned into all the three components. After flowering, the total BM (TBM) produced by the plant is allocated to storage organ only (Bouman et al. 2001). The relative death rate of leaves (DRLVT) is calculated from the weight of the green leaves as a function of DVS, which affects the LAI after flowering. Except RGRLMX and RGRLMN, other parameters are highly nonlinear. Bouman et al. (2001) have recommended values for these parameters for two different rice varieties: IR72 and IR64, and suggested that the values of parameters for any other rice variety would be very close to these recommended values, however, need to be estimated. As there were 18 parameters for the model to be estimated, we performed a SA in order to minimize the complexity in calibration and to reduce the uncertainty in the estimated values of parameters. The Sobol's sensitivity method (Sobol 1993) was employed in the current study for this purpose, and the parameters were optimized using Genetic Algorithm (GA).

#### Sobol's SA

Sobol's method (Sobol 1993) is a variance based GSA method in which the total output variance produced by any model within an ensemble is decomposed into variance caused by each parameter of the model. The method is described below following Tang et al. (2007).

**Table 1** ORYZA2000 parameters that influence the crop growth and their recommended range

Parameter	DVS	Unit	Range		Process represented
			Min	Max	
RGRLMX		°C day <sup>-1</sup>	0.0001	0.09	Max relative growth rate of leaf area
RGRLMN		°C day <sup>-1</sup>	0.0001	0.09	Min relative growth rate of leaf area
SLATB	0.00	ha kg <sup>-1</sup>	0.0001	0.09	Parameters of function to calculate specific leaf area
	0.16	ha kg <sup>-1</sup>	0.0001	0.09	
	0.33	ha kg <sup>-1</sup>	0.0001	0.09	
	0.65	ha kg <sup>-1</sup>	0.0001	0.09	
	0.79	ha kg <sup>-1</sup>	0.0001	0.09	
	2.10	ha kg <sup>-1</sup>	0.0001	0.09	
	2.50	ha kg <sup>-1</sup>	0.0001	0.09	
LRSTR		–	0.01	0.99	Fraction of allocated stem reserve that is available for growth
FLVTB	0.00	–	0.01	0.99	Fraction of shoot dry matter allocated to leaves at corresponding development stages of crop (DVS)
	0.50	–	0.01	0.99	
	0.75	–	0.01	0.99	
FSTTB	0.00	–	0.01	0.99	Fraction of shoot dry matter allocated to stems at corresponding DVS
	0.50	–	0.01	0.99	
	0.75	–	0.01	0.99	
	1.00	–	0.01	0.99	
FSOTB	1.00	–	0.01	0.99	Fraction of shoot dry matter allocated to storage organs at corresponding DVS
DRLVT	1.00	day <sup>-1</sup>	0.001	0.09	Leaf death coefficient as function DVS
	1.60	day <sup>-1</sup>	0.001	0.09	
	2.10	day <sup>-1</sup>	0.001	0.09	
	2.50	day <sup>-1</sup>	0.001	0.09	

Consider a generic model described by:

$$\mathbf{y} = f(\mathbf{x}|\boldsymbol{\theta}) \quad (1)$$

where  $f(\cdot)$  is the function described by the model,  $\mathbf{y}$  is the output from the model (crop yield in this study) corresponding to the inputs  $\mathbf{x}$  and  $\boldsymbol{\theta}$  is the vector of parameters of the model. Sobol's variance decomposition is:

$$D(y) = \sum_i D_i + \sum_{i<j} D_{ij} + \sum_{i<j<k} D_{ijk} + D_{12\dots m} \quad (2)$$

where  $D(y)$  is the total variance of the output of the model,  $D_i$  is the measure of individual variance due to the  $i$ th parameter,  $D_{ij}$  is the variance induced due to the interaction between  $i$ th parameter and  $j$ th parameter, and  $m$  is the total number of parameters. For this study, the primary interest was to get each parameters' individual contribution (first order indices) to the output and the total contribution (total order) to the output. The first order and total order Sobol's sensitivity indices are defined as:

$$\text{First order index: } S_i = D_i/D(y) \quad (3)$$

$$\text{Total order index: } S_{Ti} = 1 - (D_{\sim i}/D(y)) \quad (4)$$

where  $S_i$  refers to the sensitivity of  $i$ th parameter to the model output,  $S_{Ti}$  refers the total order sensitivity that is the sum of independent and interactive effects of  $i$ th parameter to the output, and  $D_{\sim i}$  is the average variance resulting from all the parameters, except  $i$ th parameter.

The variance terms of the Eqs. 3–5— $D$ ,  $D_i$ , and  $D_{\sim i}$ , are calculated by numerical integration in Monte Carlo approximation framework (Sobol 1993, 2001; Tang et al. 2007). The total variance  $D$  is the statistical variance of the output across the simulations. The Monte Carlo approximation for the variance terms are:

$$\hat{f}_0 = \frac{1}{n} \sum_{s=1}^n f(\theta_s) \quad (5)$$

$$\hat{D} = \frac{1}{n} \sum_{s=1}^n f^2(\theta_s) - \hat{f}_0^2 \quad (6)$$

$$\hat{D}_i = \frac{1}{n} \sum_{s=1}^n f(\theta_s^a) f(\theta_{(-i)s}^b, \theta_{is}^a) - \hat{f}_0^2 \quad (7)$$

$$\hat{D}_{-i} = \frac{1}{n} \sum_{s=1}^n f(\theta_s^a) f(\theta_{(-i)s}^a, \theta_{is}^b) - \hat{f}_0^2 \quad (8)$$

where  $n$  defines the Monte Carlo sample size,  $\theta_s$  represent the sampled individual in the unit hypercube, and (a) and (b) are two different sets of samples. Parameters from the sample set (a) denoted as  $\theta_s^a$  and  $\theta_{is}^b$  denotes  $i$ th parameter is taken from sample (b).  $\theta_{(-i)s}^a$  denote all the parameters from sample set (a) are taken except  $i$ th parameter. The Eqs. 5–8 provide a way to compute the first order and total order sensitivity of each parameter of the model.

### Implementation of SA

The Eqs. 5–8 depict the original Monte Carlo approximation formulae for estimating the terms in the decomposition of total variance. However, a robust computation strategy proposed in Liburne et al. (2006) for the Sobol’s method is applied in this study for computing the variance terms  $D$ ,  $D_i$ , and  $D_{\sim i}$ , which is as follows:

- Choose a base sample dimension (2000 in this study) and generate sample using some sampling technique (Latin Hypercube sampling in this study).
- Split the sampled parameter set into two equal matrices (A and B) (see Fig. 1).
- Derive two matrices  $C_i$  and  $D_i$  by swapping  $i$ th columns of A and B (see Fig. 1).
- Perform the Monte Carlo simulation of the model using all the samples in all four matrices (A, B, C, and D), and compute the model performance index for each simulation.
- Calculate sensitivity following Eqs. 9–11.

According to Liburne et al. (2006), the first order and total order equations are:

$$A = \begin{bmatrix} \theta_1^1 & \dots & \theta_l^1 & \dots & \theta_k^1 \\ \theta_1^2 & \dots & \theta_l^2 & \dots & \theta_k^2 \\ \vdots & & & & \\ \theta_1^{\frac{N}{2}} & \dots & \theta_l^{\frac{N}{2}} & \dots & \theta_k^{\frac{N}{2}} \end{bmatrix} \quad B = \begin{bmatrix} \theta_1^{\frac{N}{2}+1} & \dots & \theta_l^{\frac{N}{2}+1} & \dots & \theta_k^{\frac{N}{2}+1} \\ \theta_1^{\frac{N}{2}+2} & \dots & \theta_l^{\frac{N}{2}+2} & \dots & \theta_k^{\frac{N}{2}+2} \\ \vdots & & & & \\ \theta_1^N & \dots & \theta_l^N & \dots & \theta_k^N \end{bmatrix}$$

$$D_i = \begin{bmatrix} \theta_1^1 & \dots & \theta_l^{\frac{N}{2}+1} & \dots & \theta_k^1 \\ \theta_1^2 & \dots & \theta_l^{\frac{N}{2}+2} & \dots & \theta_k^2 \\ \vdots & & & & \\ \theta_1^{\frac{N}{2}} & \dots & \theta_l^N & \dots & \theta_k^{\frac{N}{2}} \end{bmatrix} \quad C_i = \begin{bmatrix} \theta_1^{\frac{N}{2}+1} & \dots & \theta_l^1 & \dots & \theta_k^{\frac{N}{2}+1} \\ \theta_1^{\frac{N}{2}+2} & \dots & \theta_l^2 & \dots & \theta_k^{\frac{N}{2}+2} \\ \vdots & & & & \\ \theta_1^N & \dots & \theta_l^{\frac{N}{2}} & \dots & \theta_k^N \end{bmatrix}$$

**Fig. 1** Illustration of parameter matrix formulation in Sobol’s method. The matrices A and B are the base matrices of sampled parameters. C and D are derived from A and B by swapping the column of first parameter

$$S_i = \frac{\hat{D}_i}{\hat{D}} = \frac{Y_{(A)}Y_{(Ci)} - f_0^2}{Y_{(A)}Y_{(A)} - f_0^2} \tag{9}$$

$$ST_i = \frac{\hat{D}_{\sim i}}{\hat{D}} = 1 - \frac{Y_{(A)}Y_{(Di)} - f_0^2}{Y_{(A)}Y_{(A)} - f_0^2} \tag{10}$$

where, the mean

$$f_0 = \frac{1}{n} \sum_{s=1}^n Y_A^s \tag{11}$$

Note that the  $S_i$  and  $S_{Ti}$  can be computed in eight different ways (Liburne et al. 2006) using Eqs. 9 and 10 by interchanging the simulation matrix A, B, C, and D in Eqs. 9 and 10. The average value of these eight ensembles of sensitivity indices is considered to be the representative value of sensitivity for the parameter.

The total number of simulations required for this computation is  $((k + 1) \times n)$ , where  $k$  is the number of parameters and  $n$  is the sample size. Consequently in the current study 38,000 model simulations are performed for the SA (2,000 samples each for 18 parameters along with 2,000 base simulations).

### Optimizer—GA

As discussed earlier, the calibration of the model is a highly complex, non-linear optimization problem. The objective of the optimizer in the current study is to identify the optimal combination of parameters of the model that closely match the simulated and measured crop yield. The major concern here is that the objective function of minimizing the error between the simulated and measured crop yield is not a direct function of the decision variables (model parameters in this case). Therefore, despite the existence of a large number of traditional non-linear programming techniques for solving this kind of optimization problem, a search based optimizer is appropriate. In the current study, we employed GA (Holland 1975; Goldberg 1989; Michalewicz 1992) as the optimizer because of its various advantages, which includes their potential to search the solution from a population of points (not a single point), use objective function information itself but not any derivatives, and use probabilistic transitions rules but not deterministic rules. GA has found a large number of applications in complex optimization problems in various branches of science and engineering (Kohler 1990; Bickel and Bickel 1990; Suckley 1991; Cook and Wolfe 1991). GA was used for parameter estimation of bio-physical models (Bulatewicz et al. 2009) and the results were encouraging.

Genetic Algorithm is a random search optimization algorithm inspired by biological evolution that provides a robust method for searching of the optimum solution to complex problems. In a GA, the solution set is represented

by a population of strings, which comprises of a number of blocks each representing the individual decision variables of the problem. Strings are processed and combined according to their fitness (objective function value evaluated using the components in the string), in order to generate new strings that contain the best features of two parent strings. Strings with the highest fitness have the greatest chance of contributing to future generations, similar to the process of natural selection. Initially GA suggests a set of candidate solutions to the problems, evaluates the fitness function that is to be optimized, and arrives at the optimal solution by the genetic operations in subsequent generations. A detailed description about the GA is beyond the scope of this article, and the readers are referred to Goldberg (1989) and Michalewicz (1992).

### Field experiments

The ORYZA2000 model was calibrated using the data from field experiments conducted at Tamil Nadu Agricultural University, Coimbatore, India during 2 years (1999 and 2000). The experiments were laid out in a split plot design with three replications of three different water applications by growing medium duration rice variety. The experiments were continued in three consecutive seasons (June–October 1999 (kharif season), September 1999–February 2000 (rabi season); June–October 2000) in the 2 years of study (Luikham 2001). The water applications considered for the experiment were (i) application of 5 cm irrigation water depth as and when the standing water has disappeared—no deficit condition (IR1) (ii) application of

5 cm irrigation water depth 1 day after the standing water has disappeared (IR2), and (iii) application of 5 cm irrigation water depth 3 days after the standing water disappeared (IR3). Details of the experiments and crop period are presented in Table 2, which also includes the information about rainfall during the crop growth period.

The nutrient supply for the crop was done at full recommended levels as per the Crop Production Manual for the area (TNAU 1994) in order to insure that the crop will not have any nutrient deficiency during the experiment. The major properties of the soil in which the crop was grown is presented in Table 3. Soil information collected during the experiment were fraction of sand, silt, and clay; textural class; organic matter (%); soil pH; electrical conductivity ( $\text{dS m}^{-1}$ ); volumetric water content at FC; and permanent wilting point (PWP), and infiltration rate. Using these informations, the soil properties such as saturated hydraulic conductivity, volumetric water content at saturation, and soil moisture tension at different moisture levels were determined using the pedo-transfer function proposed by Saxton and Rawls (2006). It may be noted that the computed values of moisture content at FC and PWP were closely matching with the measured values.

During the experiments, the dates of sowing, emergence, transplanting, active tillering (AT), panicle initiation, flowering, and physiological maturity were recorded in each experimental plot. In order to determine the total crop BM and LAI at different stages of crop growth, crop samples were collected at AT, panicle initiation, flowering, and maturity. At the time of harvest, yield components were measured in terms of total crop yield, weight of 1,000

**Table 2** Details of experiments on Rice (*Oryza sativa*)

Location	Coimbatore (11°N 77°E)—altitude 426.7 m		
Cropping season	Kharif 1999 (June–October)	Rabi 1999 (October–February)	Kharif 2000 (June–October)
Water application	IR1, IR2, IR3	IR1, IR2, IR3	IR1, IR2, IR3
Nitrogen splits	N1, N2, N3	N1, N2, N3	N1, N2, N3
Cultivar	ADTRH 1	CORH 2	ADTRH 1
Sowing date	June 10	Sep 30	June 9
Transplanting date	July 6	Oct 29	July 4
Panicle initiation date	Aug 10	Dec 8	Aug 4
Flowering date	Sep 4	Jan 3	Sep 7
Maturity date	Oct 10	Feb 8	Oct 6
Duration (days)	120	130	118
Seedlings per hill	1	1	1
Plant population (hills $\text{m}^{-2}$ )	50	50	50
Rainfall (mm)	118.4	141	310.8

IR1 application of 5 cm irrigation water depth as and when the standing water has disappeared—no deficit condition, IR2 application of 5 cm irrigation water depth 1 day after the standing water has disappeared, IR3 application of 5 cm irrigation water depth 3 days after the standing water is disappeared, N1 7DAT—25.5 kg; 21DAT—49.5 kg; PI—49.5 kg, and FF—25.5 kg, N2 Basal—75 kg; AT—37.5 kg, and PI—37.5 kg, N3 Basal—37.5 kg; AT—37.5 kg; PI—37.5 kg; and FF—37.5 kg, DAT days after transplanting, PI panicle initiation, FF fifty percent flowering, Basal 1 day before transplanting, AT active tillering)

**Table 3** Soil Properties of the experimental farm

Parameter	Value
Soil textural class	Clay loam
Clay:silt:sand (%)	39.5:18.5:40.6
$\theta_{wp}(\psi_{wp})$	0.236 (1463 kPa)
$\theta_{fc}(\psi_{fc})$	0.357 (33 kPa)
$\theta_s(\psi_s)$	0.441 (5 kPa)
$K_s$	1.51 mm h <sup>-1</sup>

$\theta$  Volumetric soil water content (fraction);  $\psi$  soil water potential (kPa);  $w_p$ ,  $f_c$ , and  $s$  are wilting point, field capacity, and saturation, respectively; and  $K_s$  saturated hydraulic conductivity

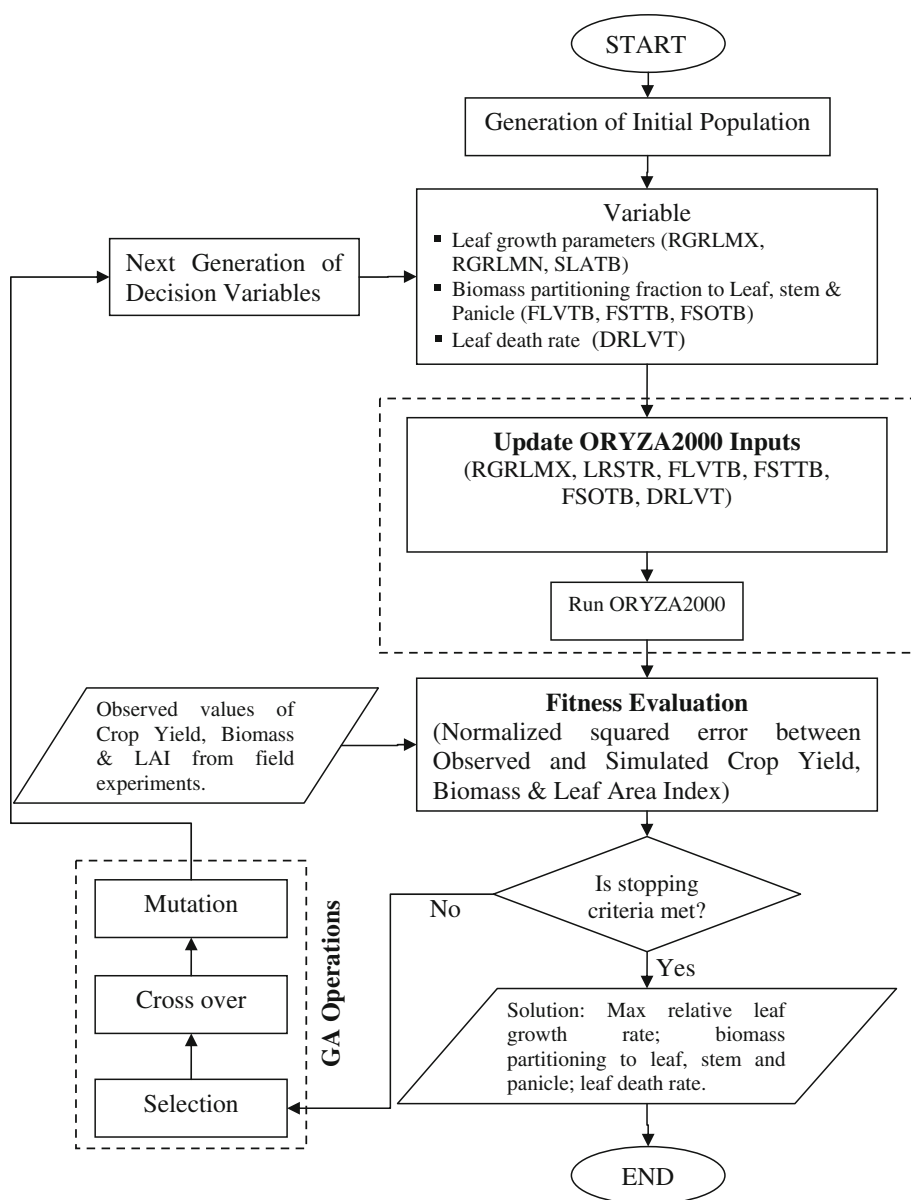
grains and the straw weight. During the period of experiment, the climatic parameters such as values of minimum and maximum temperature, minimum and maximum

relative humidity, sunshine hours, wind speed, and rainfall on each day were recorded.

Auto-calibration of ORYZA2000

The procedure of auto-calibration of the model is presented in Fig. 2 in the form of a flow chart. Initially, the GA generates candidate models for the decision variables (parameters) from the feasible region. Using these values for the decision variables, the ORYZA2000 model simulates the crop growth and yield. The model is provided with the soil properties observed during the experiment. The simulated values are used in evaluating the fitness function, based on which the GA develops the next generation candidates. The optimization of fitness function is continued till the maximum number of generation is reached.

**Fig. 2** Flow chart of the auto-calibration in simulation-optimization framework



The average daily percolation rate was fine tuned after the calibration of the model in order to represent the actual field conditions by considering the measured values of water balance components in the experimental field. The contribution of water through capillary pores to the crop root zone was not considered, since the water table of the experimental field is significantly deep. The effective rainfall during the period of crop growth was computed using the procedure outlined by Bouman et al. (2001), in which the any amount of rainfall above the field bund height is considered to be not supplementing the irrigation.

## Results and discussions

### Sensitivity of ORYZA2000 model parameters

The Sobol's SA suggested that nine parameters out of the total 18 parameters of ORYZA2000 were sensitive for the cultivars considered in this study. Table 4 presents the ranking of parameters in their order of sensitivity according to the value of Sobol's first order index. The most sensitive parameter is RGRLMX, which influence the leaf growth of the crop. The FLVTB, a parameter that facilitates the transfer of shoot BM to leaf growth in the model

**Table 4** Sobol's sensitivity indices for ORYZA2000 model parameters

Parameter	DVS	Rank	Sobol's first order index
RGRLMX		1	0.6687
RGRLMN		10	0
SLATB	0.00	10	0
	0.16	10	0
	0.33	10	0
	0.65	10	0
	0.79	10	0
	2.10	10	0
	2.50	10	0
LRSTR		6	0.0261
FLVTB	0.00	9	0.0181
	0.50	2	0.0435
	0.75	5	0.0284
FSTTB	0.00		
	0.50		
	0.75		
	1.00	8	0.0237
FSOTB	1.00		
DRLVT	1.00	3	0.0415
	1.60	4	0.0324
	2.10	7	0.0252
	2.50	10	0

simulation, is found to be the next sensitive one at a growth stage corresponding to 0.50 DVS, which represents an active leaf growing stage where more leaf production occurs. As the death of leaves take place after flowering, the parameter DRLVT (leaf death rate) at grain filling stage and later (corresponding to 1.00 DVS and above) is also found to be sensitive. The parameters LRSTR and FSTTB (at growth stage corresponding to 1.00 DVS, which is grain filling stage) are found to be important in the crop simulation as these parameters influence the transfer of BM from the stem to grain during the filling process in the model simulation. It is noted that some of the parameters are not sensitive at all for these cultivars as their sensitivity index was zero. The results of the SA suggest that calibration should mostly be focused on estimating the values of the first nine parameters, and therefore only these parameters are considered for auto-calibration. For the non-sensitive parameters, the values corresponding to cultivar IR72 was used, as suggested by Bouman et al. (2001).

### Parameter estimation using auto-calibration procedure

The data from the experiments described above corresponding to maximum production condition (no deficit condition—IR1) was used for estimation of ORYZA2000 model parameters (9 sets of experiments). The remaining 18 sets of experimental data were used for assessing the performance of the model (validation). As mentioned earlier, 9 model parameters, viz. BM partitioning factors (total 4 parameters, stem reserve factor (1 parameter), relative leaf growth rate (1 parameter), leaf death rate (total 3 parameters), are estimated during calibration of the model. The measured data during the field experiments were LAI and TBM at four different stages of growth and the crop yield at the time of maturity, and were used for computing the fitness function by the GA during the auto-calibration procedure.

The selection of fitness function in any optimization problem is crucial in identifying the appropriate values for the decision variables. In the current study, the decision variables were the parameters of the ORYZA2000 model. The model output is in terms of LAI, TBM at different stages of growth, and the total yield of the crop at the end of growing season. Typically the model parameters are usually estimated by solving the minimization problem, where the sum of squared error between the simulated and measured values of the model output is minimized. It may be noted that the three outputs from the model are different in nature in terms of their unit of expression and the magnitude of values. Therefore, a proper combining mechanism has to be selected for considering the error minimization of all the three outputs together. As suggested by Wallach et al. (2001), we used the following fitness function to be evaluated by GA:



**Table 5** Auto-calibrated values for ORYZA2000 model parameters

Parameter	DVS	Value
RGRLMX		0.0104
LRSTR		0.755
FLVTB	0.00	0.799
	0.50	0.759
	0.75	0.100
FSTTB	1.00	0.598
DRLVT	1.00	0.005
	1.60	0.015
	2.10	0.059

$$\sum_{j=1}^n \sum_{i=1}^m \left( \frac{\text{LAI}_{ij}^o - \text{LAI}_{ij}^s}{\text{LAI}_{ij}^o} \right)^2 + \sum_{j=1}^n \sum_{i=1}^m \left( \frac{\text{TBM}_{ij}^o - \text{TBM}_{ij}^s}{\text{TBM}_{ij}^o} \right)^2 + \sum_{j=1}^n \left( \frac{Y_j^o - Y_j^s}{Y_j^o} \right)^2 \quad (12)$$

where, LAI—leaf area index; TBM—total BM (kg ha<sup>-1</sup>); Y—yield (kg ha<sup>-1</sup>); with subscript variables ‘o’ and ‘s’ corresponding to the observed and simulated values, respectively; ‘i’ refers to the stage of crop growth, ‘j’ refers to the season number of the experimental data.

The calibrated values of parameters are presented in Table 5. Note that the phenological development parameters were directly estimated from effective temperature and observed phenology (dates of transplanting, panicle initiation, flowering and maturity) following Bouman et al. (2001). It may be noted that the calibrated values of parameters are lying in between the recommended value of these parameters for rice varieties IR72 and IR64 (Bouman et al. 2001).

The effectiveness of the model simulation during calibration and validation are presented in Table 6 in terms of the performance indices. The performance index used in this study is NRMSE, which is a ratio of the root mean square error (RMSE) of the estimation and the observed average value of the variable. Ideal value of this index is zero, and any value close to zero indicates a good model performance (less residual in model output). It can be

**Table 6** Performance measures of simulations of ORYZA2000 using auto-calibrated values of parameters (for full season)

Parameters	Calibration			Validation		
	Mean	RMSE	NRMSE	Mean	RMSE	NRMSE
LAI	3.32	0.398	0.12	2.91	1.05	0.360
BM (kg ha <sup>-1</sup> )	9078	812	0.09	6295	2331	0.370
Y (kg ha <sup>-1</sup> )	7816	285	0.036	6273	287	0.045

observed from Table 6 that the ORYZA2000 model is able to simulate the total crop yield (Y) reasonably well, as the NRMSE during calibration as well as validation are close to zero. It may be noted that the RMSE reported in Table 6 is computed from a total 18 experiment data. A lower value of RMSE, equal to 287 kg.ha<sup>-1</sup> during the validation of the model, indicates that the parameters that are estimated by the auto-calibration framework are able to simulate the crop growth reasonable well in other situations than it is calibrated for. It can also be observed from Table 6 that the estimated parameters of ORYZA2000 is effective in simulating the LAI and the TBM during the validation of the model, though the NRMSE is slightly on the higher side compared to that corresponding to the yield.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (x_i^o - x_i^s)^2}{n}} \quad (13)$$

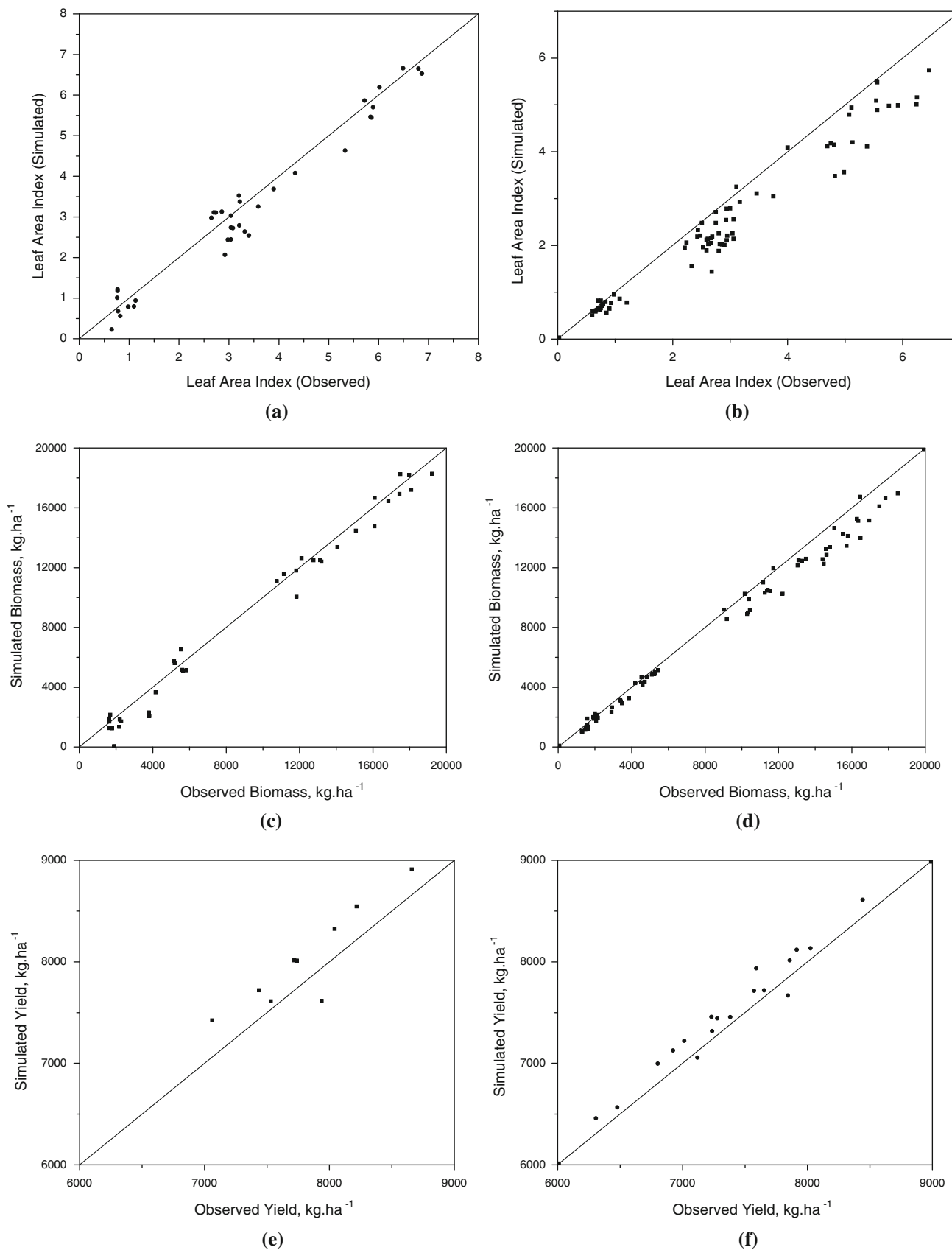
$$\text{NRMSE} = \frac{\sqrt{\sum_{i=1}^n (x_i^o - x_i^s)/n}}{\sum_{i=1}^n x_i^o/n} \quad (14)$$

where  $x_i^o$ —observed parameter ‘i’;  $x_i^s$ —simulated parameter ‘i’;  $n$ —number of parameters.

A scatter plot for comparison of the observed and simulated values of LAI, TBM and the total yield are presented in Fig. 3a–f for visual examination. It may be noted that the data points are less scattered are very close the ideal line of 45°. Nonetheless, the LAI (Fig. 3b) and BM (Fig. 3d) shows a higher scatter in validation compared to the others. This can be plausibly attributed to the error in simulation at various stages of the crop growth, which can be better understood with the data presented in Table 7, which depicts the performance measures of the model at various stages of growth.

Note from Table 7 that the RMSE value of LAI at flowering and maturity in the validation data are close to each other (1.355 and 1.384, respectively), indicating the simulation of LAI by the model at these two stages are similar. This result may plausibly be due to the model assumption that the entire amount of BM is sent to grain filling after the flowering, which may not be actually the case in the field. Therefore, the simulated value of LAI shows much deviation from the measured, which attributes to the scattering in the Fig. 3b.

It can also be noted from Table 7 that the NRMSE for BM simulation at the AT stage (1.85) is much higher than other growth stages (all are less than unity). It should be noted that the AT of rice crops takes place after transplantation, and the transplanted crop in the actual field may take some time to stabilize and establish the growth (transplantation shock). Since the model assumes continuous growth of the crop even after transplantation, the transplantation shock may not be appropriately simulated



**Fig. 3** Scatter plot of simulated and measured values of LAI (**a** calibration, **b** validation), BM (**c** calibration, **d** validation), Yield (**e** calibration, **f** validation)

**Table 7** Performance measures of simulations of ORYZA2000 using auto-calibrated values of parameters (stage wise)

Parameters	Calibration				Validation			
	AT	PI	F	M	AT	PI	F	M
<b>Mean</b>								
LAI	0.86 (0.82)	3.08 (2.94)	6.09 (5.90)	3.27 (3.01)	0.73 (0.52)	2.70 (2.56)	5.38 (4.35)	2.82 (1.70)
BM (kg ha <sup>-1</sup> )	1878 (1469)	4973 (4592)	12316 (11992)	17147 (16801)	603 (1652)	2618 (4255)	8954 (11295)	14494 (15137)
Y (kg ha <sup>-1</sup> )				7816 (8019)				6273 (6332)
<b>RMSE</b>								
LAI	0.308	0.410	0.342	0.503	0.306	0.752	1.355	1.384
BM (kg ha <sup>-1</sup> )	762	941	770	759	1118	2093	2722	2128
Y (kg ha <sup>-1</sup> )				285				287
<b>NRMSE</b>								
LAI	0.357	0.133	0.056	0.153	0.416	0.278	0.252	0.490
BM (kg ha <sup>-1</sup> )	0.405	0.189	0.062	0.044	1.85	0.80	0.30	0.14
Y (kg ha <sup>-1</sup> )				0.036				0.04

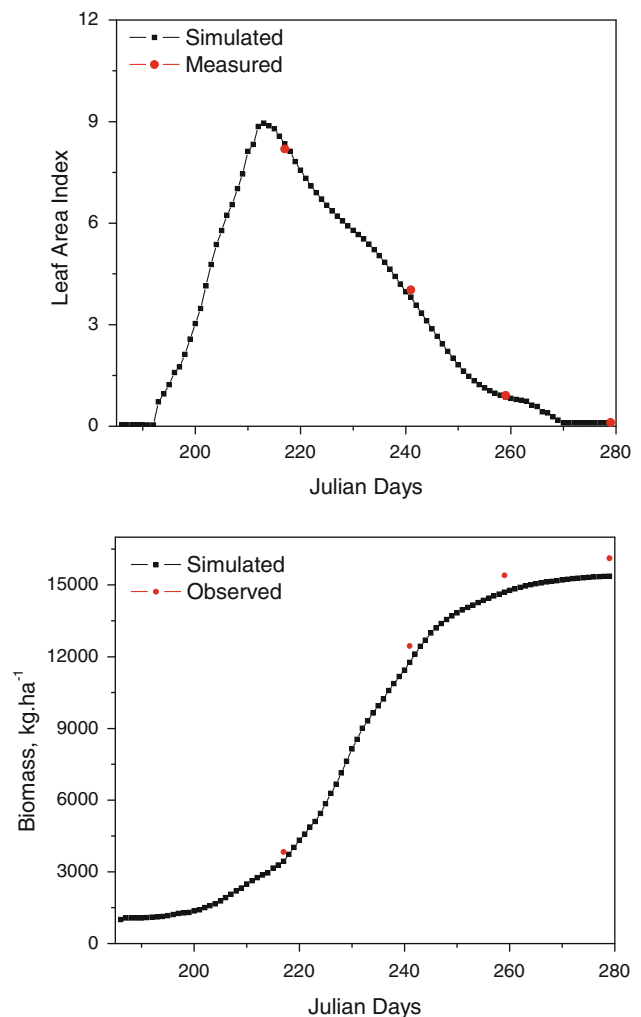
Mean the values in the bracket are the average of the simulated values and those outside the brackets are the average of the observed values, AT active tillering, PI panicle initiation, F flowering, M maturity

by the model. This difference in growth rate (between actual and simulated) may be the plausible reason for high NRMSE value for BM at the AT stage. This may also create a bias in simulation of BM along the crop growth, which may be the plausible reason for the difference between simulated and measured BM at different stages of growth (Fig. 3d).

From Fig. 3e, f, it can be observed that the simulated yield is slightly overestimated compared to the measured yield during the experiments. This can be plausibly attributed to the time duration for the flowering to occur in the actual field. It may be noted that it may take a few weeks for all the plants to complete the flowering, while the model considers that the flowering of all plants take place simultaneously. Since the BM is completely used for grain filling by the model after flowering, the model may output a higher yield.

Figure 4 depicts the daily simulation of LAI and TBM along the crop growth for a typical experiment set up used for validation of the model. The observed values of these variables at different stages of growth are also presented in the Fig. 4, for visual comparison. It is noted from Fig. 4 that the behavior of simulation of LAI and BM accumulation along the crop growth is very well represented in the model as is evidence by the close values of these variables at different growth stages.

It is worth mentioning that simulated yield, LAI and BM using the default values of the parameters (those recommended by Bouman et al. (2001)) for IR72 and IR64 varieties of rice showed large deviation from those simulated by the model using the estimated parameters, though the values are close to default values. The yield simulated by the model, when IR72 parameters are used, is



**Fig. 4** Simulated and measured LAI and BM along the crop growing period (Kharif 1999, Irrigation Treatment IR2)

8103.8 kg ha<sup>-1</sup>, and that corresponding to IR64 parameters is 5703.9 kg ha<sup>-1</sup> against a measured yield of 8,660 kg ha<sup>-1</sup>. Note that the calibrated model simulated a yield of 8,910 kg ha<sup>-1</sup>, which is sufficiently close to the measured value. Therefore, it can be suggested that a calibration of the model parameters is essential when a different variety of rice and cultivar are used in the field.

From the foregoing discussions, it is evident that the parameters of the ORYZA2000 models can be effectively estimated for any rice variety using field experimental data in a simulation-optimization framework. The estimated parameters of the model show good generalization property, and therefore can be used in developing appropriate irrigation schedules for the crop.

### Summary and conclusions

In the current study, an auto-calibration framework is proposed to estimate the optimal parameter values of ORYZA2000 model, a rice crop growth simulation model, using GA. The crop growth simulation model is integrated within the GA, and field experimental data has been employed to calibrate the model using the simulation-optimization framework. Prior to the calibration of the model, a SA was performed to prune the number of parameters that are to be calibrated.

The results of the study indicate that the calibrated ORYZA2000 model was able to effectively simulate the crop growth under full irrigation and water deficit conditions. During the validation of the model, it is observed that the simulated yield closely matches with the measured yield under different experiment treatment. The model was found to be efficiently simulating the BM production at various stages of crop growth. It is noted that while there are 18 parameters for the model, some of them are not very sensitive to the final yield and can assume the values recommended by Bouman et al. (2001). The results also suggest that calibration of the model is necessary as the simulated values of yield, LAI and BM are showing large deviation from those values simulated using default values of the parameters. Overall, the results of the study are highly encouraging, and the calibrated model could be used for developing optimal irrigation schedule for rice crop under various levels of water deficit condition.

### References

- Arora VK (2006) Application of a rice growth and water balance model in an irrigated semi-arid subtropical environment. *Agric Water Manag* 83:51–57
- Belder P, Bouman BAM, Cabangon R, Guoan L, Quilang EJP, Li Y, Spiertz JHJ, Tuong TP (2004) Effect of water-saving irrigation on rice yield and water use in typical lowland conditions in Asia. *Agric Water Manag* 65:193–210
- Belder P, Bouman BAM, Spiertz JHJ (2007) Exploring options for water savings in lowland rice using a modelling approach. *Agric Syst* 92:91–114
- Bickel AS, Bickel RW (1990) Determination of near-optimum use of hospital diagnostic resources using genetic algorithm shell. *Comput Biol Med* 20(1):1–13
- Boogaard HL, van Diepen CA, Rotter RP, Cabrera JMCA, van Laar HH (1998) User's guide for the WOFOST 7.1: crop growth simulation model and WOFOST control center 1.5. Technical Document 52, DLO Winand Staring Centre, Wageningen
- Bouman BAM, Kropff MJ, Tuong TP, Wopereis MCS, ten Berge HFM, van Laar HH (2001) ORYZA2000: modeling lowland rice. International Rice Research Institute, Los Banos
- Bouman BAM, van Laar HH (2006) Description and evaluation of the rice growth model ORYZA2000 under nitrogen-limited conditions. *Agric Syst* 87:249–273
- Brumbelow K, Georgakakos A (2007) Determining crop–water production functions using yield irrigation gradient algorithms. *Agric Water Manag* 87:151–161
- Bulatewicz T, Jin W, Staggenborg S, Lauwo S, Miller M, Das S, Andresen D, Peterson J, Steward DR, Welch SM (2009) Calibration of a crop model to irrigated water use using a genetic algorithm. *Hydrol Earth Syst Sci* 13:1467–1483
- Cibin R, Sudheer KP, Chaubey I (2010) Sensitivity and identifiability of stream flow generation parameters of the SWAT model. *Hydrol Process* 24(9):1133–1148
- Cook DF, Wolfe ML (1991) Genetic algorithm approach to lumber cutting optimization problem. *Cybern Syst* 22(3):357–365
- Doorenbos J, Kassam AH (1979) Yield response to water. Irrigation and drainage paper No. 33. Rome, FAO
- Feng L, Bouman BAM, Tuong TP, Cabangon RJ, Yalong Li, Guoan Lu, Feng Yuehua (2007) Exploring options to grow rice using less water in northern China using a modelling approach I. Field experiments and model evaluation. *Agric Water Manag* 88:1–13
- Goldberg DE (1989) Genetic algorithms in search, optimization and machine learning. Addison-Wesley, Reading
- Holland JH (1975) Adaptation in natural and artificial systems. University of Michigan Press, Ann Arbor
- Jensen ME (1968) Water consumption by agricultural plants. In: Kozlowski TT (ed) Water deficits and plants growth, vol II. Academic Press, New York
- Kohler HM (1990) Adaptive genetic algorithm for the binary perceptron problem. *J Phys A* 23:L1265–L1271
- Kumar DN, Srinivasa Raju K, Ashok B (2006) Optimal reservoir operation for irrigation of multiple crops using genetic algorithms. *J Irrig Drain Eng* 132(2):123–129
- Liburne L, Gatelli D, Tarantola S (2006) Sensitivity analysis on spatial models: new approach. In: Proceedings of the 7th international symposium on spatial accuracy assessment in natural resources and environmental sciences
- Luikham K (2001) Management of irrigation water and advanced methods of nitrogen application on hybrid rice. PhD Thesis, Tamil Nadu Agricultural University
- McMennary J, O'Toole JC (1985) RICEMOD: a physiologically-based rice growth model. IRRI research paper series No. 87, Manila
- Michalewicz Z (1992) Genetic algorithms + data structures = evolution programs. Springer, New York
- Mohanakrishnan A (1990) Selected papers on irrigation. Irrigation Management Training Institute, Trichy
- Paul S, Panda SN, Kumar N (2000) Optimal irrigation allocation: a multilevel approach. *J Irrig Drain Eng* 126(3):149–156

- Prasad AS, Umamahesh NV, Viswanath GK (2006) Optimal irrigation planning under water scarcity. *J Irrig Drain Eng* 132(3):228–237
- Pundarikanthan NV, Santhi C (1996) Irrigation scheduling in a developing country: experiences from Tamil Nadu, India. In: Proceedings of the ICID/FAO workshop on irrigation scheduling: from theory to practice, September 1995, Rome. Water report No. 8, FAO
- Rao NH, Rees DH (1992) Irrigation scheduling of rice with a crop growth simulation model. *Agric Syst* 39:115–132
- Rao NH, Sarma PBS, Chander S (1988) Irrigation scheduling under a limited water supply. *Agric Water Manag* 15:165–175
- Saltelli A, Scott EM, Chan K, Marian S (2000) Sensitivity analysis. Wiley, Chichester
- Saseendran SA, Ahuja LR, Nielsen DC, Trout TJ, Ma L (2008) Use of crop simulation models to evaluate limited irrigation management options for corn in a semiarid environment. *Water Resour Res* 44:1–12
- Saxton KE, Rawls WJ (2006) Soil water characteristic estimates by texture and organic matter for hydrologic solutions. *Soil Sci Soc Am J* 70:1569–1578
- Sobol' IM (1993) Sensitivity estimates for nonlinear mathematical models. *Math Model Comput Exp* 1:404–414
- Sobol' IM (2001) Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math Comp Simul* 55(1):271–280
- Soundharajan B, Sudheer KP (2009) Deficit irrigation management for rice using crop growth simulation model in an optimization framework. *Paddy Water Environ* 7(2):135–149
- Suckley D (1991) Genetic algorithm in the design of FIR filters. *IEEE Proc G Circuits Dev Syst* 138(2):234–238
- Talpez H, Mjelde JW (1988) Crop irrigation scheduling via simulation based experimentation. *West J Agric Econ* 13(2):184–192
- Tang Y, Reed PM, Wagener T, van Werkhoven K (2007) Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation. *Hydrol Earth Syst Sci* 11:793–817
- TNAU (1994) Crop production guide. Tamil Nadu Agricultural University, Coimbatore, pp 1–26
- Varella H, Guerif M, Buis S (2010) Global sensitivity analysis measures the quality of parameter estimation: the case of soil parameters and a crop model. *Environ Model Softw* 25:310–319
- Vedula S, Mujumdar PP (1992) Optimal reservoir operation for irrigation of multiple crops. *Water Resour Res* 28(1):1–9
- Wallach D, Goffinet B, Bergez JE, Debaeke P, Leenhardt D (2001) Parameter estimation for crop models: a new approach and application to a corn model. *Agron J* 93(4):757–766
- Zhai T, Mohtar RH, El-Awar F, Jabre W, Volenc JJ (2004) Parameter estimation for process-oriented crop growth models. *Trans ASAE* 47(6):2109–2119