# 1 Supplementary Material

**Edgar S. Correa, Francisco C. Calderon, Julian D. Colorado**

## 1.1 CERES-Rice crop parameters

**Table S1** Crop parameters in CERES-Rice model and empirical calculation of Nerica 4.

| Parameter | Range | Value | Description |
|---|---|---|---|
| P1 | 150–800 | 520 | Time period (expressed as growing degree days) |
| P2R | 5–300 | 10 | Panicle initiation (expressed as GDD in °C-d) |
| P5 | 150–850 | 550 | Time period from beginning of grain filling (°C-d) |
| P2O | 11–13 | 13 | The longest day length (in hours) |
| G1 | 38–540 | 75 | Potential spikelet number coefficient |
| G2 | 0.015–0.03 | 0.03 | Single grain weight (g) |
| G3 | 0.7–1.97 | 0.6 | Tillering coefficient (scalar value) |
| PHINT | 55–90 | 60 | Phyllochron Interval (°C-d) |
| THOT | 25–34 | 35 | Spikelet sterility is affected by high temperature |
| TCLDP | 12–18 | 15 | Panicle initiation is delayed by low temperature |
| TCLDF | 10–20 | 15 | Spikelet sterility is affected by low temperature |

## 1.2 Genetic Algorithm optimization

The Genetic Algorithm (GA) is an optimization method based on artificial intelligence (AI), specifically within the subcategory of evolutionary computation. We employed it to calibrate the CERES-Rice crop model by optimizing its parameters through the evolution of a population of candidate solutions across multiple generations. The algorithm starts by initializing a population of individuals, each representing a potential parameter set. It evaluates the fitness of these individuals by comparing their performance with observed values using a fitness function. The selection process favors the best-performing individuals, which are then used to generate the next generation through crossover and mutation operations. Mutation introduces diversity into the population, preventing premature convergence.

The GA iterates over a specified number of generations, during which the algorithm continually refines the population by selecting parents, performing crossover, and applying mutation, aiming to minimize the error between the model's predictions and observed values. Key metrics such as the minimum error (dist_hist), mean error across the population (dist_hist_mean), and the best genotype's error (hist_BEST) are tracked throughout the optimization process to monitor progress and convergence.

The algorithm uses the following parameters: Num_Pop, the number of generations, is set to 15; Num_ind, the initial number of individuals in the population, is set to 10; Thr_mut, the mutation probability, is set to 0.7. At the end of the process, the best set of parameters (ind_best.mat) is saved, and the error trends over iterations are visualized and logged, allowing for further analysis and refinement.

Each individual in the GA represents a potential solution, corresponding to a specific combination of crop parameters for the CERES-Rice crop growth model. The individual is defined as $X_i = \{P1, P20, P2R, P5, G1, G2, G3, PHINT\}$, where each parameter takes a value within a predefined range, as shown in Table S1.

**Operators in Genetic Algorithms** To calibrate the model, we apply a Genetic Algorithm (GA) for optimization, using the $\mathbf{R^2}$-NRMSE Integrated Index (RII) as the fitness function to guide the selection, recombination, and mutation processes. The GA is employed to tune model parameters iteratively, ensuring a balance between model accuracy and genotype variability.

### *Fitness Metric: $R^2$-NRMSE Integrated Index (RII)*

To guide and assess the performance and accuracy of the model for each combination of crop parameter solutions, we introduce a fitness metric, the $\mathbf{R^2}$-NRMSE Integrated Index (RII), defined in (1), where $\mathbf{R^2}$ is given by (2) and NRMSE is given by (3).

$$\mathrm{RII} = \mid R^2 - 1 \mid + \mathrm{NRMSE} \tag{1}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{2}$$

$$\mathrm{NRMSE} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}}{\max(y_i) - \min(y_i)} \tag{3}$$

This metric combines the coefficient of determination (R2) and the normalized root mean square error (NRMSE), two widely recognized indicators of model fit and error. The goal of the metric is to optimize for: (i) NRMSE approaching zero, which indicates a low error between the predicted and observed values, and (ii) $\mathbf{R^2}$ approaching one, which reflects a high proportion of variance explained by the model. Thus, as the RII metric approaches zero, the model's performance improves, reducing both estimation error (NRMSE $\approx 0$) and achieving a model fit ($\mathbf{R^2} \approx 1$).

This optimization incorporates the integrated index RII for Grain Yield (GY), Biomass (BSS), Number of Grains (NG), Number of Tillers (NT), and phenological stages, including anthesis (ANT) and maturity (MAT). The Cost Function (FC) integrates all these variables, as defined in (4). Grain Yield and Biomass are given double priority relative to the other variables.

$$\mathrm{FC} = \left(\mathrm{GY}.\frac{2}{8} + \mathrm{BSS}.\frac{2}{8} + \mathrm{NG}.\frac{1}{8} + \mathrm{NT}.\frac{1}{8} + \mathrm{ANT}.\frac{1}{8} + \mathrm{MAT}.\frac{1}{8}\right) \tag{4}$$

Each individual in the GA represents a potential solution, corresponding to a specific combination of crop parameters. The fitness of each individual is defined as shown in equation (5).

$$\mathrm{FIT\_function} = 1 - \left(\frac{\mathrm{FC}(i)}{\sum_{j=1}^{n}\mathrm{FC}(j)}\right) \tag{5}$$

### *Selection Operator*

Roulette Wheel Selection, or fitness proportional selection, is a probabilistic method in genetic algorithms. It starts by calculating the fitness of each individual in the population, which reflects how well they solve the optimization problem. The total fitness $F_{\text{total}}$ of the population is calculated by adding the fitness values of all individuals $f_i$.

The selection probability $P_i$ for each individual $i$ is then computed by dividing the fitness of the individual $f_i$ by the total fitness $F_{\text{total}}$, as shown in equation (6).

$$P_i = \frac{f_i}{\sum_{i=1}^{n} f_i} \tag{6}$$

Next, a random number $r \in [0,1]$ is generated, and the cumulative selection probability $C_i$ is computed by summing the selection probabilities up to individual $i$, as expressed in equation (7). The individual $i$ is selected if the random number $r$ lies within the cumulative probability interval $[C_{i-1}, C_i)$, as defined in Equation (8).

$$C_i = \sum_{j=1}^{i} P_j \tag{7}$$

$$C_{i-1} \leq r < C_i, \quad \text{where } C_0 = 0 \text{ and } C_n = 1 \tag{8}$$

This process ensures that individuals with higher fitness have a greater chance of selection, while still allowing for genetic diversity by giving less fit individuals a small chance to be selected.

### *Recombination and mutation Operators*

Arithmetic Crossover is a recombination operator. The primary objective of this operator is to combine two parent solutions to produce offspring solutions by generating new combinations of their values. Unlike traditional crossover methods that exchange segments or genes, arithmetic crossover blends the parent solutions through a weighted average. Given two parent solutions, $P_1 = \{P_1^1, P_1^2, \ldots, P_1^n\}$ and $P_2 = \{P_2^1, P_2^2, \ldots, P_2^n\}$, where each $P_1^i$ and $P_2^i$ represent the values of the corresponding genes in the solution vectors, arithmetic crossover produces a new offspring $O$ based on a linear combination of the parents' gene values. The new gene values in the offspring are calculated as expressed in equation (9).

$$O^i = \alpha P_1^i + (1-\alpha) P_2^i, \quad \forall i \in \{1, 2, \ldots, n\} \tag{9}$$

where $\alpha$ is a random scalar weight factor in the interval [0, 1] that controls the influence of each parent on the offspring. Finally, a random mutation is applied, prioritizing the crop parameters with the highest sensitivity, based on the sensitivity analysis.

### 1.2.1 Results

The algorithm runs the CERES-Rice model 8280 times for 9 sites of the rainfed experiment in the Nerica 4 cultivar [1]. It evaluates 920 combinations of crop parameters, and the performance of the best candidate is shown in Figure S1.

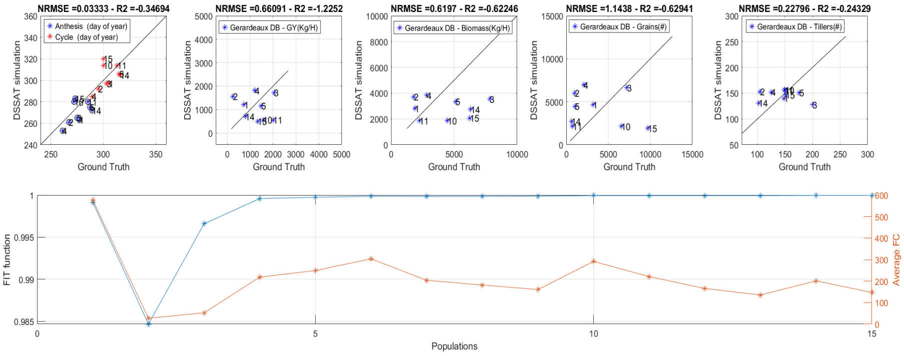

**Fig. S1** GA-based optimization for the calibration of Nerica 4 cultivar.

The variability of the cost function in each generation ensures variability in the exploration of possible solutions. The best combination of crop parameters is shown in Table S2.

The results of the Genetic Algorithm in Table S2 are consistent with the empirical calibration in Table S1. The difference in the GA lies in three crop parameters: PHINT (Phyllochron Interval), where the algorithm defined 63.5 instead of 60; G2 (Single grain weight), where the algorithm defined 0.026 instead of 0.03; and P5 (Time period from the beginning of grain filling), where the algorithm defined 427 instead of 550. The other crop parameters remained unchanged.

**Table S2** GA-driven calculation of crop parameters in the CERES-Rice model.

| **Parameter** | **Value** | **Description** |
|---|---|---|
| P1 | 520 | Time period (expressed as growing degree days) |
| P2R | 10 | Panicle initiation (expressed as GDD in °C-d) |
| P5 | 427 | Time period from beginning of grain filling (°C-d) |
| P2O | 13 | The longest day length (in hours) |
| G1 | 75 | Potential spikelet number coefficient |
| G2 | 0.026 | Single grain weight (g) |
| G3 | 0.6 | Tillering coefficient (scalar value) |
| PHINT | 63.5 | Phyllochron Interval (°C-d) |
| THOT | 35 | Spikelet sterility is affected by high temperature |
| TCLDP | 15 | Panicle initiation is delayed by low temperature |
| TCLDF | 15 | Spikelet sterility is affected by low temperature |

### 1.3 Support data

- https://data.mendeley.com/preview/fwp748vfkx?a=5bffaa91-1590-46b5-9c27-41ecfd82c119
- https://github.com/EdgarStevenC/Crop-Growth-Modelling

## References

[1] Gérardeaux, E., Falconnier, G., Gozé, E., Defrance, D., Kouakou, P.-M., Loison, R., Sultan, B., Affholder, F., Muller, B.: Adapting rainfed rice to climate change: a case study in senegal. Agronomy for Sustainable Development **41**(4), 57 (2021) https://doi.org/10.1007/s13593-021-00710-2