Mechanistic crop modelling and AI for ideotype optimization: Crop-scale advances to enhance yield and water use efficiency

Edgar S. Correa [1,2,3*]

1 UMR AGAP Institut, Univ. Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France
2 CIRAD, UMR AGAP Institut, F-34398, Montpellier, France
3 Pontificia Universidad Javeriana, School of Engineering, Bogota Cra. 7 No. 40-62, 110311 Bogota, Colombia

* e_correa@javeriana.edu.co

Abstract

Advancing our understanding and optimization of crop–environment interactions is essential for enhancing food production while conserving resources. Mechanistic crop models (MCMs) simulate crop growth by integrating physiological processes such as photosynthesis, biomass allocation, and phenological development. When combined with satellite-derived environmental data, these models support large-scale assessment of yield potential and water use efficiency, particularly under drought-prone conditions.

In this study, we evaluate the CERES-Rice model to define an ideotype that maximizes grain productivity while optimizing water use. A global sensitivity analysis using the Morris method identifies the most influential genetic-based parameters affecting traits such as biomass accumulation, grain yield, tiller number, anthesis timing, and maturity. These parameters were then optimized using a Genetic Algorithm (GA), simulating 1,884 virtual cultivars across 5,692 runs, guided by an integrated Harvest Index–Water Use Efficiency (HI-WUE) metric.

The resulting ideotype exhibits favorable traits including optimized phenological timing, and improved grain and water conversion efficiencies across four contrasting environments. To evaluate its genetic feasibility, the optimized parameter profile was compared to 21 field-characterized rice cultivars (indica, japonica, and hybrids) using multidimensional similarity metrics, including Euclidean, Mahalanobis, and Cosine distances.

This integrative framework demonstrates how combining MCMs with AI-based optimization and genotypic comparison can guide ideotype design and accelerate cultivar improvement. The findings highlight actionable physiological traits and genetic targets for breeding climate-resilient, high-efficiency rice varieties.

Author Summary

Efforts to improve crop productivity and conserve water resources are a key motivation that guides modern agricultural strategies. This study investigates how AI-driven advancements and mechanistic crop modelling can be combined to enhance

drought-adaptation strategies in rainfed systems, with an emphasis on capturing genotype-by-environment interactions at regional scales.

We introduce an integrated framework that couples the CERES-Rice crop model with genetic algorithm to simulate a wide range of virtual cultivars under diverse environmental conditions. Using the proposed HI-WUE metric, combining harvest index and water use efficiency, this framework identifies ideotypes optimized for resource efficiency. A multidimensional similarity analysis quantifies the genetic distance between computationally optimized ideotypes and field-characterized cultivars, revealing key trait alignments and supporting physiologically grounded, genetically feasible breeding recommendations.

Central to this approach is the model's ability to comprehensively reproduce the behavior of individual varieties, relying on iterative calibration and validation processes. This process ensures robust ideotype predictions and demonstrates how the co-development of crop models and phenotyping platforms, through collaboration between biologists and modelers, can translate these predictions into practical breeding outcomes for climate-resilient crop varieties. While mechanistic models may not yet fully capture genetic complexity at the gene network or 3D architectural levels, their capacity to represent functional genetic diversity remains a promising asset. Future work will likely focus on enhancing these models to better integrate the complex dynamics and underlying mechanisms of plant stress responses, further strengthening their predictive power for breeding outcomes.

Introduction

To address the challenges of modern agriculture, breeders must rapidly phenotype large populations and accurately target the most promising progeny [1–3]. As global food security demands rise, increasing the efficiency and precision of crop breeding has become essential. Recent advances in artificial intelligence (AI), robotics, and multispectral imaging have driven progress in smart agriculture, enabling high-throughput, non-invasive phenotyping and data-driven decision-making [4–11]. However, accurately predicting crop performance under diverse and variable conditions remains a critical challenge due to complex genotype-by-environment (G×E) interactions [12, 13].

Mechanistic crop models (MCMs) provide a structured framework to simulate plant development as a dynamic system, integrating physiological processes such as photosynthesis, biomass allocation, phenology, and soil–plant–atmosphere interactions [14, 15]. Widely used MCMs such as DSSAT, APSIM, STICS, and AquaCrop simulate crop responses under different environmental and management scenarios [16–18]. These models typically consist of five core modules, assimilation, water/nutrient balance, biomass partitioning, grain formation, and phenology, each influenced by genotype and environmental drivers (Fig 1).

While MCMs have played a pivotal role in climate adaptation and agronomic research [19, 20], their integration with genetic information remains underdeveloped. There is a pressing need to connect canopy-level physiological simulations with underlying genetic architectures to enable in silico ideotype design [21]. Ideotypes are conceptual models of crop types that combine favorable traits to optimize performance under specific environmental constraints [2, 22, 23]. However, breeding such ideotypes remains constrained by long generation cycles, resource-intensive field evaluations, and G×E-induced trait variability [24–26].

State-of-the-art research on ideotype optimization with crop modelling has identified morphological traits such as leaf area, shoot diameter, internode length, and branching

angle as proxies for carbon assimilation efficiency [24, 27]. Additional traits such as root depth, leaf senescence sensitivity, and vegetative cycle duration also influence resource-use efficiency and yield potential [24, 28]. However, these traits often exhibit strong G×E dependencies, limiting the stability of trait-based selection across sites and seasons.

To overcome these limitations, computational frameworks integrating MCMs with AI-based optimization have emerged as promising tools to simulate G×E interactions and design virtual cultivars tailored to specific environments [27, 29]. These models enable exploration of trait synergies that maximize yield and resource efficiency, offering a more predictive and scalable approach to ideotype design.

In this study, we explore the potential of the CERES-Rice model to support ideotype design under varying environmental scenarios, with an emphasis on physiologically grounded and genetically attainable traits. Unlike previous approaches centered primarily on yield-driven morphological optimization, our framework introduces a dual-performance metric, integrating Harvest Index (HI) and Water Use Efficiency (WUE), to quantify grain conversion efficiency relative to water availability. This HI-WUE index addresses the dual imperatives of improving productivity and conserving limited water resources.

A global sensitivity analysis using the Morris method identifies the most influential genotype-based parameters, including phenological stages (P1, P2O, PHINT), assimilate partitioning, and growth traits. These parameters are optimized via a Genetic Algorithm (GA), simulating 1,884 virtual cultivars across 5,692 model runs spanning four distinct environments.

We then compare the optimized ideotype against 21 field-characterized rice cultivars, spanning indica, japonica, and hybrid groups, using multidimensional similarity metrics (Euclidean, Mahalanobis, and Cosine distances) [30, 31]. This comparative analysis

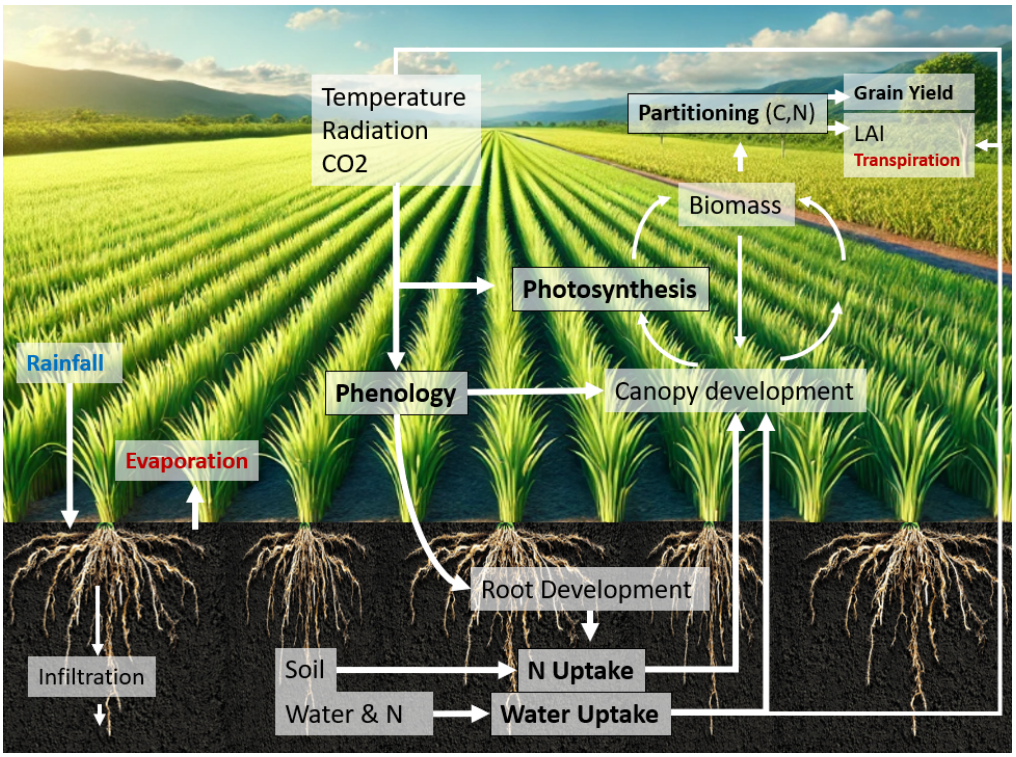


Fig 1. General structure of mechanistic crop models and their core components.

Table 1. Soil, climate, and crop parameters used in the crop growth model

Environment		Cultivar
Soil Parameters	**Climate Parameters**	**Crop Observations**
Clay content (g/kg) - SLCL	Minimum Temperature (°C)	Biomass (kg/ha)
Silt content (g/kg) - SLSI	Maximum Temperature (°C)	Grain Yield (kg/ha)
Soil Sand (g/kg) - Soil Sand	Solar Radiation (MJ/m^2/day)	Number of Grains (#/m^2)
Bulk density (cg/cm^3) - SBDM	Wind Speed (m/s)	Number of Tillers (#/m^2)
Coarse fraction (cm^3/dm^3) - SLCF	Relative Humidity (%)	Anthesis (days)
Soil Nitrogen (cg/kg) - SLNI	Precipitation (mm/day)	Maturity (days)
Total Organic Carbon (dg/kg) - SLOC		
pH (unitless) - SLHW		

bridges theoretical design with real-world breeding, identifying genotypes closest to the virtual ideotype and highlighting targets for future crossing strategies.

By aligning physiological optimization with genetic feasibility, this study proposes a reproducible, data-driven framework for ideotype design. The integration of MCMs, AI-driven optimization, and genotypic similarity metrics provides a powerful foundation for accelerating the development of high-performing, climate-resilient cultivars.

Materials and methods

Database

This research is based on rainfed field experiments conducted between 2012 and 2014 along the west coast of Africa, covering the Casamance and Eastern Senegal, including Ziguinchor, Sédhiou, Kolda [32]. This study examines the crop's response to environmental factors by integrating soil properties and climatic variables with a comprehensive database that includes growth responses such as yield and biomass measurements across 21 varieties.

Soil Data: Soil properties were analyzed through laboratory tests at a depth of 30 cm. In contrast, the satellite-based SoilGrids database provides soil characteristics at three depth levels (15 cm, 30 cm, and 100 cm), which are used as input variables for the crop growth model [33]. The properties include soil texture, essential nutrients, and other critical parameters, as outlined in Table 1. Additionally, the crop growth model requires soil hydraulic properties such as the permanent wilting point (SLLL), field capacity (SDUL), saturation (SSAT), and saturated hydraulic conductivity (SSKS), which are derived using Saxton and Rawls' pedotransfer functions [34–37].

Climate Data: The crops and the mechanistic crop model representation (MCM) is driven by critical inputs such as minimum and maximum temperature (°C), solar radiation (MJ/m^2/day), wind speed (m/s), relative humidity (%), and precipitation (mm/day), all of which are derived from the rainfed rice study [32]. For regional-scale applications, these meteorological variables are sourced from the NASA POWER dataset [38].

Crop Observations: To study crop growth in response to the environment (GxE), it is crucial to understand the response of each genotype to environmental conditions. The rainfed rice study [32] analyzes 21 varieties and reports the crop's response in terms of variables such as organic matter accumulation including grain yield, biomass, number of grains, and number of tillers, as well as phenology variables like anthesis and maturity.

Environments

A study conducted in Casamance and Eastern Senegal identified twelve distinctive environments using Principal Component Analysis (PCA) and a Gaussian Mixture Model (GMM) in three dimensions, considering all soil and meteorological variables [39]. Fig 2 shows that 89% of the region is represented by four environments, covering two soil types and two climatic zones. The convex hull is the polygon that encloses all given points in an environment, and the closest point to the centroid is selected as the representative sample for that environment.

Soil type 1, predominantly located in the eastern part of the region, is characterized by higher pH (5.91), bulk density (1.50 g/cm^3), silt content (30.53 g/kg), and clay content (26.33 g/kg). It also presents intermediate hydraulic properties related to water retention, with SLLL (0.16 cm^3/cm^3), SDUL (0.29 cm^3/cm^3), SSAT (0.44 cm^3/cm^3), and a moderate saturated hydraulic conductivity (SSKS, 0.75 cm/h). In contrast, soil type 2 is classified as sandy (56.59 g/kg), with higher saturated hydraulic conductivity (SSKS, 1.09 cm/h) and a slightly lower pH (5.87). Due to its soil composition, soil type 2 exhibits low water retention capacity, which facilitates rapid water infiltration and drainage when saturated.

Climate Type 1 predominates in the southern region and is characterized by higher relative humidity (80.33%), lower average solar radiation (19.16 MJ/m^2/day) with lower variability (3.44 MJ/m^2/day), and moderate temperatures, including minimum (22.19°C), maximum (30.60°C), and average (26.39°C) values, along with greater accumulated precipitation (917 mm). Climate Type 2 is mainly located in the northern part of the study area and is distinguished by 44% lower accumulated rainfall (513.24 mm) and 18% lower average relative humidity (67.98%), along with higher average solar radiation (19.62 MJ/m^2/day), and higher minimum (22.51°C), maximum (33.04°C), and average (27.78°C) temperatures compared to the southern climate.

Genetic crop parameters of CERES-Rice model

The genetic-based crop growth parameters established in the MCM represent the genetic traits of the crop variety that govern its growth and development. These parameters are intrinsic to each variety and determine its physiological responses to environmental factors [40, 41].

Plant growth and development are grounded in the accumulation of temperature, quantified through Growing Degree Days (GDD). This metric, widely used in mechanistic crop models (MCMs), represents the thermal time required for plants to progress through phenological stages. GDD is determined by calculating the difference between the daily mean temperature and a base temperature, below which growth is inhibited [42–45]. Accumulating these daily GDD values throughout the growing period provides a temperature-driven measure of physiological progress, supporting dynamic

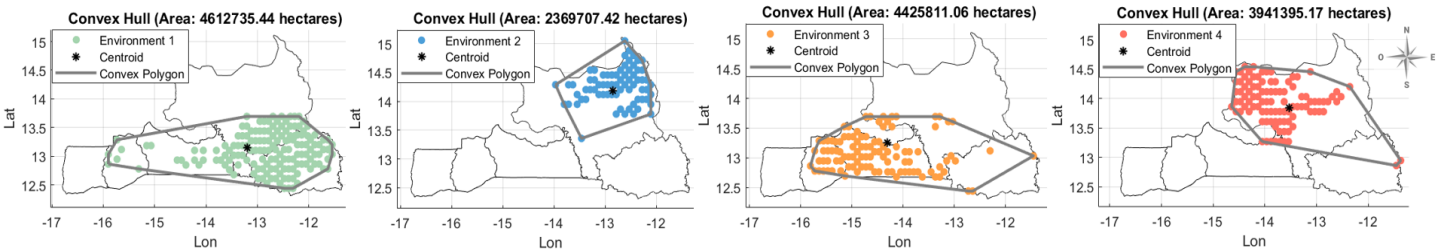


Fig 2. Environmental categories in the Casamance and Eastern Senegal region (adapted from Correa et al., 2025, under CC BY license) [39].

growth processes across diverse crop varieties.

Fig 3 illustrates key crop parameters that describe the growth and development of the plant. These parameters are interrelated, collectively reflecting the crop's growth behavior. The determination of these parameters is inherently complex due to the extensive number of possible value combinations and their strong interdependencies. Traditionally, this process has relied on manual calibration or empirical methods [46, 47].

Parameters P1 and P5 are expressed in Growing Degree Days (GDD) and govern the duration of the vegetative stage and the period from grain filling to physiological maturity, respectively. These two parameters have a significant impact on the overall duration of the crop cycle, from seed emergence to physiological maturity.

P1 strongly influences the flowering event (anthesis) and determines the length of the vegetative stage. A longer vegetative stage promotes greater biomass accumulation, more robust stems and a higher number of spikelets. During this phase of plant development, three parameters play a critical role in determining plant yield: (i) PHINT, the phyllochron expressed in GDD, is a measure of the rate of leaf emergence and influences biomass production [48, 49]; (ii) G1, which defines the number of spikelets; and (iii) G3, the tillering coefficient. P5 significantly affects the grain filling stage. In this phase, two parameters influence grain yield: (i) P2R, which defines a delay in panicle initiation, expressed in GDD, and (ii) G2, which defines grain weight in grams.

Finally, P20 represents the critical photoperiod (the longest day length), influencing the rate of development [50, 51].

Spikelet sterility in rice is influenced by temperature conditions, particularly through three key parameters: THOT (heat-induced spikelet sterility), TCLDF (spikelet sterility caused by low temperatures), and TCLDP (delay in panicle initiation under low-temperature stress). Low temperatures between 15–19°C during the panicle development phase result in sterile pollen formation, leading to increased spikelet sterility [52–55]. Similarly, exposure to temperatures exceeding 35–36°C during flowering impairs pollen viability and disrupts fertilization, ultimately reducing grain yield [56–59]. TCLDF is defined at 15°C and THOT at 35°C. Additionally, TCLDP is also defined by default at 15°C.

Sensitivity Analysis of CERES-Rice Model

The Morris method is a global sensitivity analysis technique used to identify influential parameters in multivariate models, such as mechanistic crop growth models. The method proposes a sampling strategy based on random selection of one factor per time to assess parameter sensitivity while ensuring computational efficiency [60, 61].

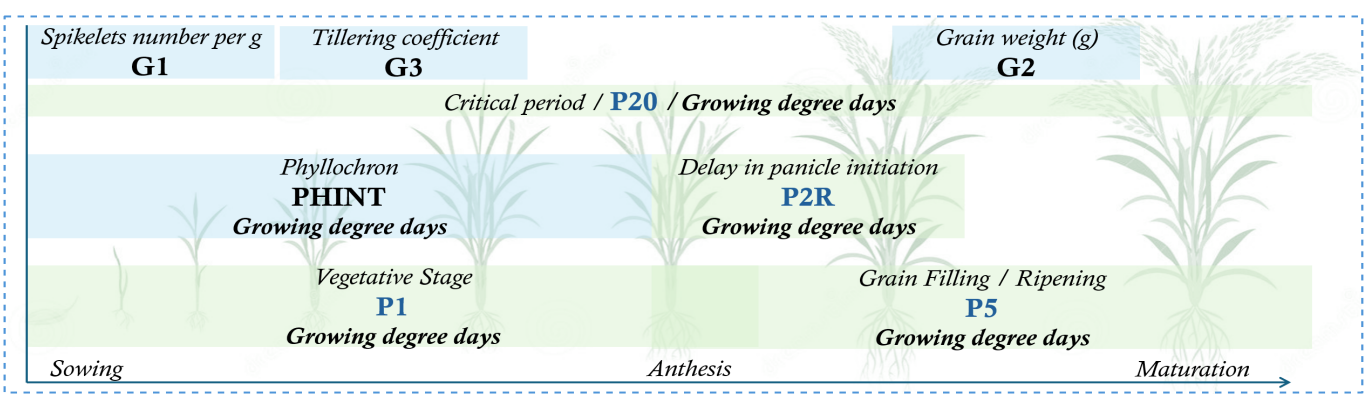


Fig 3. Genetic-based crop parameters of mechanistic crop growth model CERES-Rice.

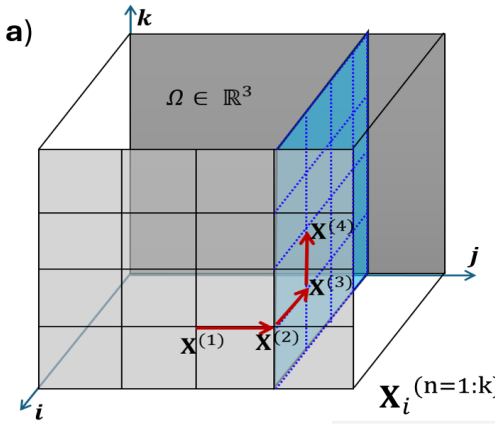
In order to estimate the sensitivity measures, the algorithm 1 focuses on sampling a number r of elementary effects from each F_i. Since each elementary effect requires two sample points, the simplest design would require $2r$ sample points per input, resulting in a total of $2rk$ model executions, where k is the number of input factors.

Morris proposed a more efficient design, where r trajectories of $k+1$ points are generated in the input space. Each trajectory provides k elementary effects (one per input factor or crop parameter), resulting in a total of $r(k+1)$ sample points. The procedure for generating each trajectory is as follows: a "base" value x^* for the vector X is randomly selected within the p-level grid Ω. x^* is not part of the trajectory, but it serves as the reference point for generating all trajectory points. The first trajectory point, $x^{(1)}$, is generated by increasing one or more components of x^* by Δ, ensuring that $x^{(1)}$ remains in Ω. The second trajectory point, $x^{(2)}$, is generated by modifying $x^{(1)}$ such that one of its components (say, the i-th component) is adjusted by $\pm\Delta$, i.e., $X_i^{(2)} = X_i^{(1)} \pm \Delta$, where i is randomly selected from the set $\{1, 2, \ldots, k\}$. The third point, $x^{(3)}$, is generated by modifying only one component j (where $j \neq i$) of $x^{(2)}$, ensuring that each trajectory point differs from the previous one by exactly one component. This process continues until $x^{(k+1)}$, which closes the trajectory. The design ensures that two consecutive points differ in exactly one component, and each component i of the base vector x^* is selected at least once for adjustment by Δ. An example of a trajectory for $k = 3$ is illustrated in Fig 4-a. Fig 4-b presents the variation (or step) for each crop parameter within its respective lower and upper bounds.

The Relative Sensitivity Index (RSI) quantifies the relative variation in the output variables of the crop growth model, including biomass, grain yield, number of grains, number of tillers, anthesis, and maturity, in response to changes in input crop parameters. By comparing absolute changes in these variables to their maximum observed absolute change, RSI provides a standardized measure of sensitivity. This allows for a precise evaluation of how variations in input crop parameters influence crop

Algorithm 1 Morris Method for Sensitivity Analysis of CERES-Rice Model

Input: Number of input factors $k = 11$ (representing the crop parameters of the CERES-Rice model), number of trajectories r, grid size p, step size Δ
Output: Set of $r(k+1)$ sample points for sensitivity analysis
Initialize: Randomly select a base value x^* for the input vector X (the crop parameters) within the p-level grid Ω
for each trajectory $t = 1$ to r **do**
    Generate the first trajectory point $x^{(1)}$ by increasing one or more components of x^* by Δ
    for each point $i = 2$ to $k+1$ **do**
        if point $i = 2$ **then**
            Modify $x^{(1)}$ by adjusting the i-th component of $x^{(1)}$ by $\pm\Delta$
        else
            Modify $x^{(i)}$ by adjusting one component (different from previous) of $x^{(i-1)}$ by $\pm\Delta$
        end if
    end for
    Store trajectory points: $x^{(1)}, x^{(2)}, \ldots, x^{(k+1)}$ (corresponding to 11 crop parameters of the CERES-Rice model)
end for
Return: Set of $r(k+1)$ sample points representing the sensitivity analysis for the 11 crop parameters of the CERES-Rice model



b)

Δ

Parameter	Δ1	Δ2	Δ3	Δ4	Δ5	Δ6
P1	150.0	280.0	410.0	540.0	670.0	800.0
P2O	11.0	11.4	11.8	12.2	12.6	13.0
P2R	5.0	64.0	123.0	182.0	241.0	300.0
P5	150.0	290.0	430.0	570.0	710.0	850.0
G1	50.0	55.0	60.0	65.0	70.0	75.0
G2	0.015	0.018	0.021	0.024	0.027	0.030
G3	0.700	0.954	1.208	1.462	1.716	1.970
PHINT	55.0	62.0	69.0	76.0	83.0	90.0
THOT	25.0	26.8	28.6	30.4	32.2	34.0
TCLDP	12.0	13.2	14.4	15.6	16.8	18.0
TCLDF	10.0	12.0	14.0	16.0	18.0	20.0

For ***CERESRice*** $\Omega \in \mathbb{R}^{k=11}$

$\mathbf{X}_i^{(n=1:k)} = \{$P1, P2O, P2R, P5, G1, G2, G3, PHINT, THOT, TCLDP, TCLDF$\}$

Fig 4. Morris method for global sensitivity analysis in crop growth model CERES-Rice. a) Sampling strategy based on random selection of one factor per time for $\mathbb{R}^{k=3}$. b) Step size (Δ) per crop parameter in $\mathbb{R}^{k=11}$

growth and development, making it a valuable tool for sensitivity analysis in crop modeling and agronomic research.

Mathematically, if ΔX_i represents the absolute change between consecutive values of the variable X, it is defined as: $\Delta X_i = \mid X_{i+1} - X_i \mid$. The mean absolute change is given by $\mu = \frac{1}{n}\sum_{i=1}^{n} \mid \Delta X_i \mid$, where $\mid \Delta X_i \mid$ represents the absolute value of the change. Thus, the proposed metric, the Relative Sensitivity Index (RSI), is defined as Eq 1

$$\text{RSI} = \frac{1}{n}\sum_{i=1}^{n} \frac{\mid \Delta X_i \mid}{\max(\mid \Delta X_i \mid)}. \tag{1}$$

Genetic Algorithm optimization

The Genetic Algorithm (GA) is an artificial intelligence approach based on heuristic search for optimization problems. It is used with the CERES-Rice crop growth model to characterize the ideotype that optimizes grain conversion efficiency and water use efficiency by optimizing genetic crop growth parameters through the evolution of a population of candidate solutions over multiple generations for each environment. The algorithm starts by initializing a population of individuals, each representing a potential set of genetic parameters for the crop growth model. It evaluates the fitness of these individuals by assessing the harvest index (HI) and water use efficiency (WUE), which are integrated into the 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 Genetic Algorithm for ideotyping in the CERES-Rice crop model is outlined in Algorithm 2. It iterates over multiple generations, optimizing genetic crop parameters to enhance resource efficiency and yield performance across different environments. Throughout this process, the algorithm refines the population by selecting parents, applying crossover, and introducing mutations to evolve toward the optimal ideotype. Key performance metrics, including maximum fitness, mean population fitness, and the best genotype's fitness, are continuously tracked to assess progress and convergence.

The algorithm uses the following parameters: the number of generations, set to 20; the initial population size, set to 15; and the mutation probability, set to 0.7.

Each individual in the GA represents a potential solution, corresponding to a

Algorithm 2 Genetic Algorithm for ideotyping in CERES-Rice

```
Initialize population P with Num_ind individuals
Set global best ind_best and min_dist_global to infinity
for generation = 1 to Num_P do
    Evaluate fitness of each individual in P
    for each individual i in P do
        if fitness of individual i is better than current best then
            Update ind_best min_dist_global
        end if
    end for
    P ← Select_parents(P)
    P ← Crossover_parents(P)
    P ← Mutation(P, Thr_mut)
    Track and plot dist_hist and dist_hist_mean
end for
Save ind_best and plot final fitness results
```

specific combination of genetic crop parameters for the CERES-Rice crop model. The individual is defined as $X_i = \{P1, P5, P2R, PHINT, P20, G1, G2, G3\}$, where each parameter takes a value within a predefined range, as shown in Fig 4-b.

Operators in Genetic Algorithms

To maximize crop yield, grain conversion efficiency, and water use efficiency, a genetic algorithm (GA) is implemented to iteratively optimize the model parameters. The fitness function guides the processes of selection, recombination, and mutation, ensuring convergence toward parameter sets that balance agronomic performance and genotypic diversity.

1. **Fitness Metric: HI-WUE Integrated Index**

   To guide ideotype optimization and evaluate each set of genetic crop parameter combinations, an integrated efficiency index (HI-WUE) is employed as the fitness metric. This index combines two critical agronomic performance indicators: the Harvest Index (HI) and Water Use Efficiency (WUE), as defined in Eq 2. The HI quantifies the efficiency of assimilate partitioning by representing the ratio of grain yield to total biomass. In contrast, the WUE reflects the efficiency of water utilization, expressed as the amount of grain produced per unit of accumulated evapotranspiration. The HI-WUE index enables an integrated assessment of genotypic performance in terms of both yield potential and resource use efficiency.

   $$\text{HI-WUE} = HI + \text{WUE} * 0.1 \tag{2}$$

   This metric combines the harvest index (HI) and water use efficiency (WUE), two widely recognized indicators of drought adaptation. The goal of the metric is to optimize: (i) HI, ranging from 0 to 1, representing grain conversion efficiency, and (ii) WUE, spanning from 2 to 15 kg/mm for adapted varieties [62], which reflects high performance under water limited conditions.

2. **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_{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_{total}, as shown in Eq 3.

$$P_i = \frac{f_i}{\sum_{i=1}^{n} f_i} \tag{3}$$

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 P_i. The individual P_i is selected if the random number r lies within the cumulative probability interval $[C_{i-1}, C_i)$, as defined in Eq 4.

$$C_{i-1} \leq r < C_i, \quad \text{where } C_0 = 0 \text{ and } C_n = 1 \tag{4}$$

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.

3. **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 Eq 5.

$$O^i = \alpha P_1^i + (1 - \alpha) P_2^i, \quad \forall i \in \{1, 2, \ldots, n\} \tag{5}$$

where α 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.

Ideotype similarities with field-characterized cultivars

Fig. 5 presents 21 distinct rice cultivars from a rainfed experiment conducted in the region [32], classified into three genetic groups: indica, japonica, and hybrid. Each cultivar is characterized by a set of genetic and growth parameters. The ideotype, derived through metaheuristic optimization using a Genetic Algorithm (GA), is then compared against these 21 cultivars to identify genotypic similarities and inform the selection of varieties that simultaneously optimize productivity, grain conversion efficiency, and water use efficiency. The similarity between the ideotypes and the field-characterized varieties is assessed using five metrics: Euclidean, Manhattan, Minkowski, Cosine, and Mahalanobis.

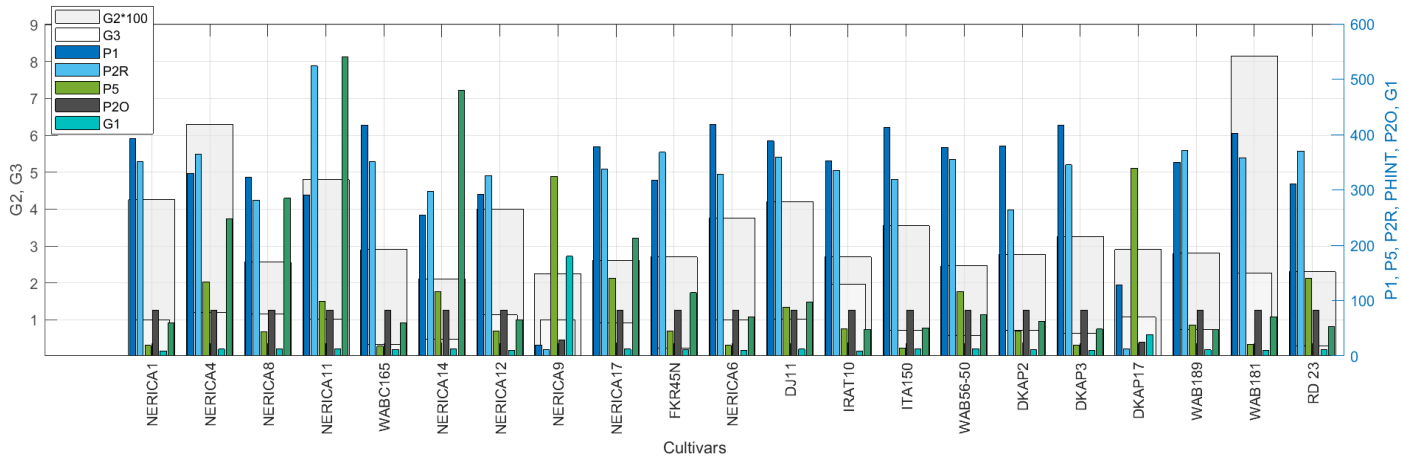


Fig 5. Genetic crop growth parameters characterized with rainfeed field experiment for 21 rice varieties.

These metrics provide a quantitative measure of dissimilarity, where lower values indicate greater similarity, and higher values reflect increased difference. The Euclidean distance measures direct separation between points, following the Pythagorean theorem. The Manhattan distance sums displacements along Cartesian axes, emulating grid-based movement. The Minkowski distance generalizes both through a control parameter (p), here set to 3. The cosine distance assesses angular similarity between vectors, disregarding magnitude, whereas the Mahalanobis distance accounts for correlations and data dispersion, making it crucial in multivariate analysis.

The similarity index for each virtual cultivar is calculated as the average of the similarities obtained from multiple distance metrics. Each metric contributes its own similarity value, which is inverted from its corresponding distance, as defined in Eq 6.

$$\text{Avg Similarity} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{1+\text{distance}_i}\right) \tag{6}$$

where n represents the number of distance metrics considered, and distance_i is the distance value associated with the i-th metric. This index reflects the degree of proximity between the virtual cultivar and its nearest neighbors based on various characteristics.

To facilitate comparison between different sets of cultivars, the similarity index is normalized by adjusting its scale so that the values are proportional and lie within the interval $[0, 1]$. The normalization is defined as follows:

$$\text{Normalized Similarity} = 1 - \frac{\text{Avg Similarity}}{\max(\text{Avg Similarity})} \tag{7}$$

This allows for uniform interpretation and comparison of similarity values, regardless of the metrics or cultivar set used. Additionally, *frequency* refers to the number of times a cultivar appears in the set of closest cultivars to the virtual variety, based on all distance metrics. This indicator quantifies the prevalence of the cultivar within the nearest results, providing insight into its commonality and relevance within the search space.

Algorithm 3 presents the structured approach for identifying the most genetically related cultivars. Together, the similarity index and frequency serve as complementary metrics. The joint representation of both metrics allows for a comprehensive evaluation

Algorithm 3 Identification of the most similar cultivars

Input: Virtual cultivar dataset V, real cultivar dataset R, distance metrics D
Output: Frequency analysis and similarity index visualization
▷ **Step 1: Compute distance and find closest cultivars**
for each metric $d \in D$ **do**
  for each virtual cultivar $v \in V$ **do**
    Compute distances: $dist[v, r] \leftarrow d(v, r), \quad \forall r \in R$
    Sort R by $dist[v, r]$ in ascending order
    Select the top-k closest cultivars and store in $T[v, d]$
  end for
end for
▷ **Step 2: Aggregate closest cultivars for frequency analysis**
for each virtual cultivar $v \in V$ **do**
  Concatenate $T[v, :]$ into a single list of cultivars
  Compute frequency of each cultivar
  Sort cultivars by frequency in descending order
end for
▷ **Step 3: Compute normalized similarity index**
for each virtual cultivar $v \in V$ **do**
  for each real cultivar $r \in T[v, :]$ **do**
    Compute similarity: $\text{similarity}[v, r] = 1 - \frac{\text{avg_distance}[v,r]}{\max(\text{avg_distance})}$
  end for
  Normalize similarity scores relative to the highest value
end for
End Algorithm

of both the degree of similarity and the relevance of cultivars within the studied population.

Results

This section presents the key findings of the optimization of the CERES-Rice crop model. Sensitivity analysis quantifies the impact of genetic-based crop parameters on biomass accumulation, grain yield, grain number, tiller count, and critical phenological stages, including anthesis and maturity. The ideotyping optimization process determines the optimal combination of genetic parameters, while similarity analysis, based on field-characterized cultivars, provides targeted recommendations for selecting the most promising progeny and potential genetic modifications to enhance resilience across diverse environments in the region.

Sensitivity analysis of genetic parameters

The sensitivity analysis assess the impact of 11 crop parameters on six output variables of the CERES-Rice model. Fig 6 illustrates μ, the mean absolute change, in units of each variable. These output variables include biomass (kg/ha), grain yield (kg/ha), number of grains (grains/ha), number of tillers (tillers/ha), anthesis (days after sowing), and maturity (days after sowing). The Relative Sensitivity Index (RSI) was used to quantify the contribution of each crop parameter to the variation observed in the model's outputs.

Biomass is influenced by several key parameters. The vegetative stage in growing

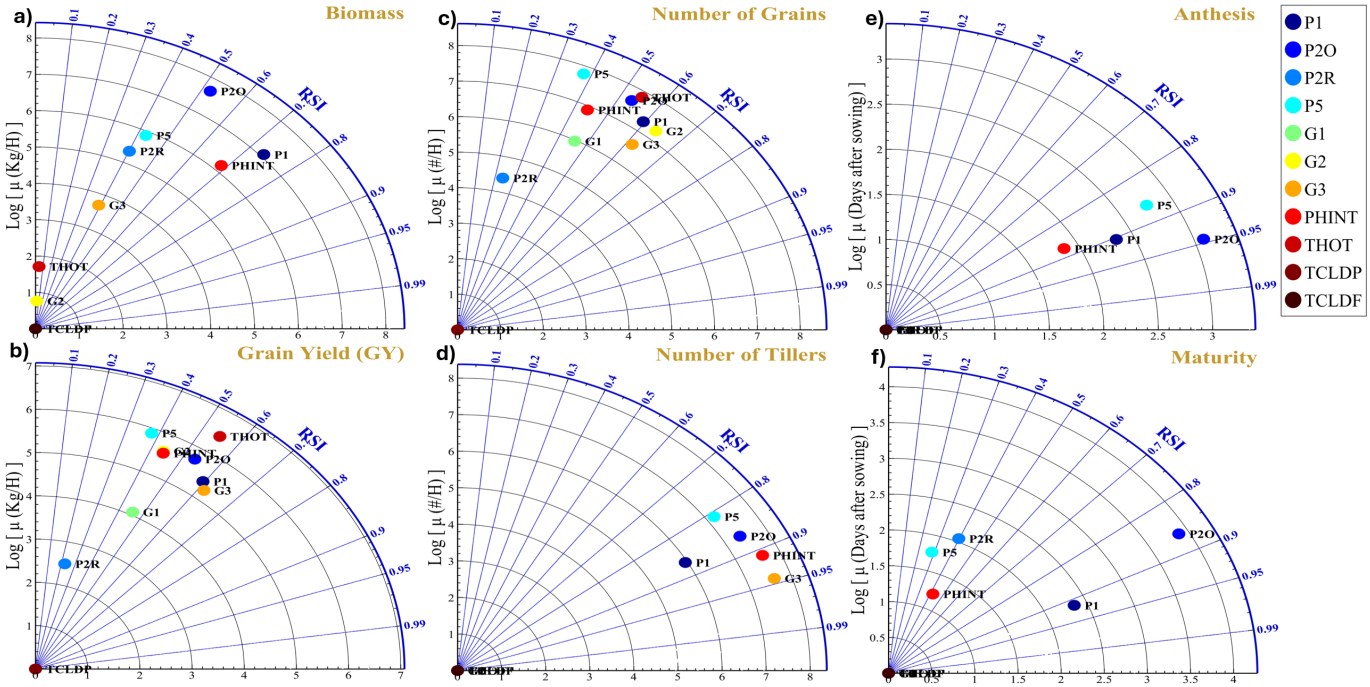


Fig 6. Relative Sensitivity Index (RSI) for CERES-Rice crop growth model. a) Biomass(Kg/H), b) Grain Yield(Kg/H), c) Number of Grains(#/H), d) Number of Tillers(#/H), e) Anthesis(Days after sowing), f) Anthesis(Days after sowing).

degree days (P1) has an RSI of 0.73, followed by phyllochron (PHINT) with an RSI of 0.68 and critical period in growing degree days (P2O) with an RSI of 0.52. Other parameters that contribute to biomass variability include grain filling in growing degree days (P5) with an RSI of 0.42, delay in panicle initiation (P2R) with an RSI of 0.40, and the tillering coefficient (G3) with an RSI of 0.39. Parameters such as heat-induced spikelet sterility (THOT) (RSI = 0.047), grain weight (G2) (RSI = 0.039), and spikelet number per gram (G1) (RSI = 0.03) have minimal impact on biomass.

Grain yield is influenced by the tillering coefficient (G3) with an RSI of 0.61 and the vegetative stage in growing degree days (P1) with an RSI of 0.59. Other significant contributors include heat-induced spikelet sterility (THOT) with an RSI of 0.54, critical period in growing degree days (P2O) with an RSI of 0.53, and spikelets per gram (G1) with an RSI of 0.45. Additionally, phyllochron (PHINT) (RSI = 0.44), grain weight (G2) (RSI = 0.43), grain filling in growing degree days (P5) (RSI = 0.37), and delay in panicle initiation (P2R) (RSI = 0.22) also influence grain yield, though to a lesser extent. The TCLDF parameter (RSI = 0.01) has a minimal effect on grain yield.

The number of grains is primarily influenced by grain weight (G2) with an RSI of 0.63, tillering coefficient (G3) with an RSI of 0.61, and the vegetative stage in growing degree days (P1) with an RSI of 0.59. Heat-induced spikelet sterility (THOT) with an RSI of 0.54 and critical period in growing degree days (P2O) with an RSI of 0.53 also have substantial effects. Other influential parameters include spikelet number per gram (G1) with an RSI of 0.45, phyllochron (PHINT) with an RSI of 0.44, grain filling in growing degree days (P5) with an RSI of 0.37, and delay in panicle initiation (P2R) with an RSI of 0.23. The TCLDF parameter (RSI = 0.03) contributes minimally to the number of grains.

The number of tillers is influenced by the tillering coefficient (G3) with an RSI of 0.94, followed by phyllochron (PHINT) with an RSI of 0.91 and the vegetative stage in growing degree days (P1) with an RSI of 0.86. The critical period in growing degree

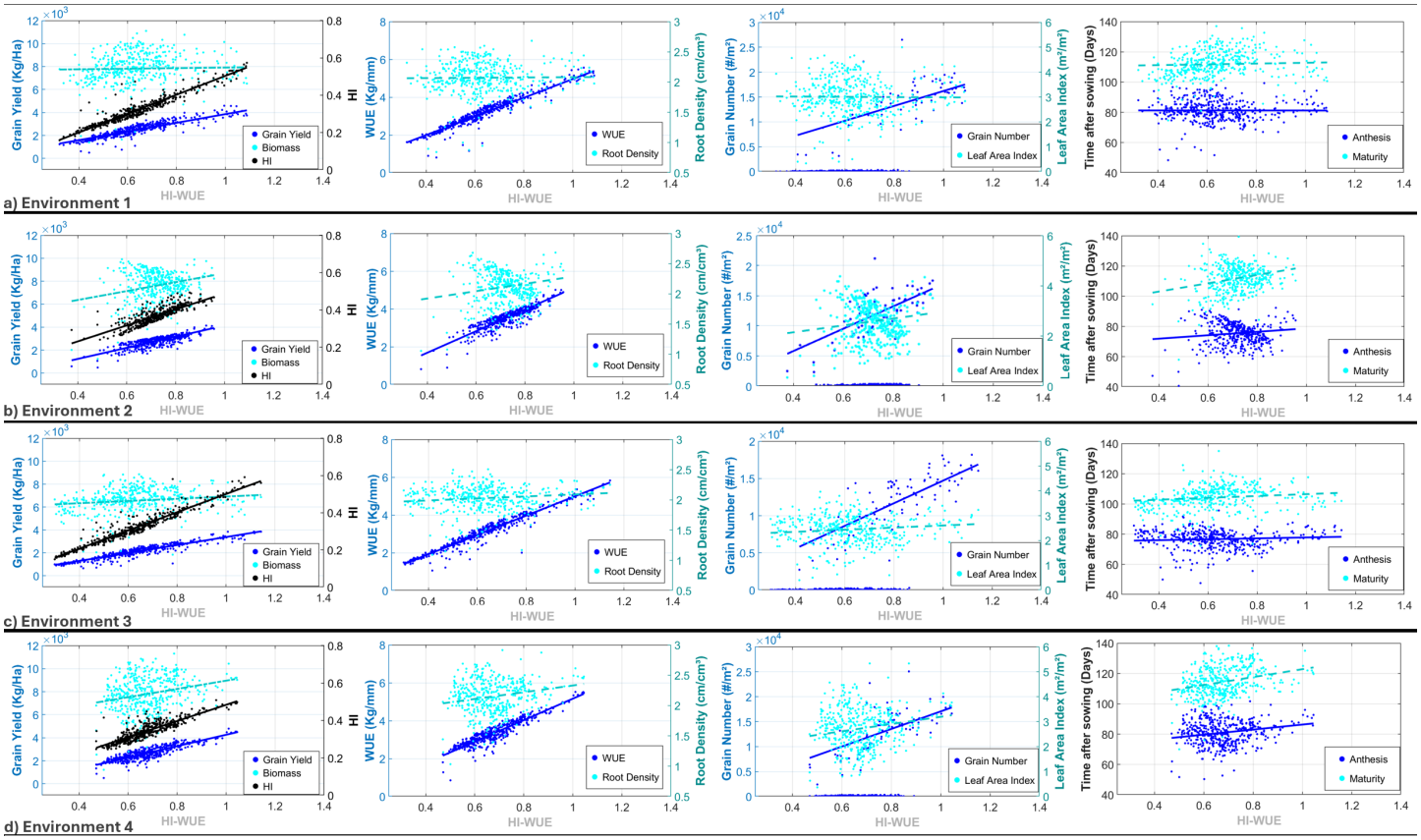


Fig 7. Ideotype construction for four environments based on the variables: Harvest index, water use efficiency, root density, grain number, leaf area index, and phenology (anthesis and maturity).

days (P2O) (RSI = 0.86) and grain filling in growing degree days (P5) (RSI = 0.81) also contribute. The delay in panicle initiation (P2R) has a minor effect on the number of tillers, with an RSI of 0.1.

Anthesis, or the flowering time, is influenced by the critical period in growing degree days (P2O) with an RSI of 0.94, followed by the vegetative stage in growing degree days (P1) with an RSI of 0.90, phyllochron (PHINT) with an RSI of 0.87, and grain filling in growing degree days (P5) with an RSI of 0.86.

Maturity, the time to harvest, is influenced by the vegetative stage in growing degree days (P1) with an RSI of 0.91, followed by the critical period in growing degree days (P2O) with an RSI of 0.86. Phyllochron (PHINT) contributes with an RSI of 0.42, while delay in panicle initiation (P2R) (RSI = 0.39) and grain filling in growing degree days (P5) (RSI = 0.28) have a smaller effect on the timing of maturity.

Ideotype through AG optimization

The genetic algorithm executes the CERES-Rice model 5,692 times, evaluating 1,884 virtual cultivars to characterize the ideotype represented by the optimal combination of crop growth parameters. Each virtual cultivar exhibits critical morphological traits such as root density, number of grain, and leaf area index (LAI), physiological traits that include anthesis and maturity, and other vital agronomic traits such as grain conversion efficiency (HI) and water use efficiency (WUE) in four main environments in the region.

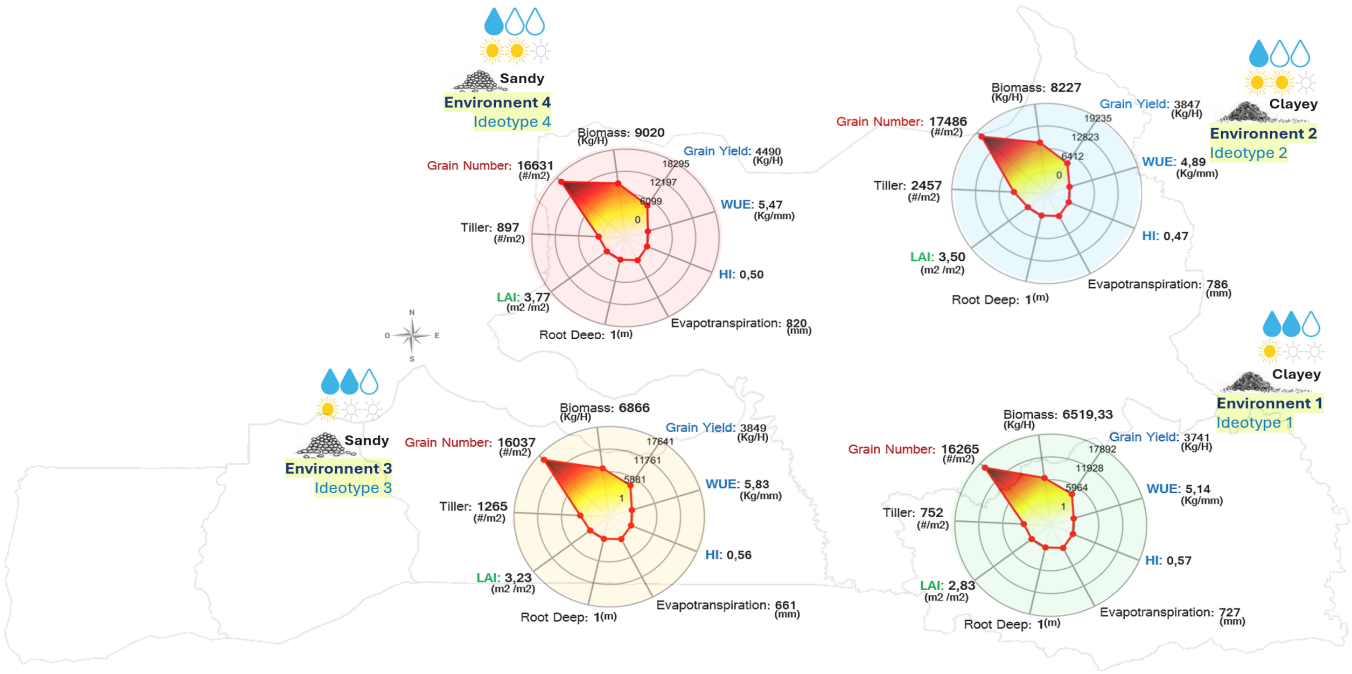


Fig 8. Optimized genotypes in varying Environments: Clayey in the East and Sandy in the West with high (917 mm) and low (513 mm) Precipitation.

Fig 7 illustrates the trends of all virtual cultivars, each represented by a individual point. The independent axis across all environments corresponds to the HI-WUE index. HI-WUE demonstrates strong Pearson correlations with crop yield (0.90), harvest index (HI) (0.92), and water use efficiency (WUE) (0.94). These correlations highlight that the ideotype represents the combination of genetic crop parameters that maximizes HI-WUE. The optimal HI-WUE values for each environment are as follows: 1.08 for Environment 1, 0.95 for Environment 2, 1.14 for Environment 3, and 1.04 for Environment 4.

The harvest index (HI) values for each environment are 0.57, 0.46, 0.56, and 0.49, with corresponding grain yields of 3741, 3847, 3849, and 4490 kg/ha, respectively. The water use efficiency (WUE) follows a similar trend, with values of 5.14, 4.89, 5.83, and 5.47 kg/mm across environments.

The root density (cm/cm^3) shows a Pearson correlation of 0.159 and a Spearman correlation of 0.06, indicating a weak relationship, with values of 1.97, 2.38, 2.25, and 2.47. Similarly, the leaf area index (LAI) (mm^2/mm^2) presents a Pearson correlation of 0.132 and a Spearman correlation of 0.021, which confirming a weak association, with values of 2.83, 3.5, 3.23, and 3.76 across environments.

In contrast, the grain number (#/m^2) exhibits a Pearson correlation of 0.66, with values of 16265, 17486, 16038, and 16632. The phenological cycle, measured in days to anthesis and maturity, with 80 and 100 in Environment 1, 84 and 110 in Environment 2, 82 and 102 in Environment 3, and 92 and 119 in Environment 4.

Fig. 8 presents the performance of the optimized ideotype in different environments for key agronomic variables: biomass, grain yield, grain number, tiller number, leaf area index, root density, root depth, harvest index, evapotranspiration and water use efficiency. Biomass decreased by 23.9% and grain yield by 14.27% in sandy soils with high saturated hydraulic conductivity (1.1 cm h^{-1}), and by 20.76% and 2.75%, respectively, in clay soils with low saturated hydraulic conductivity (0.75 cm h^{-1}),

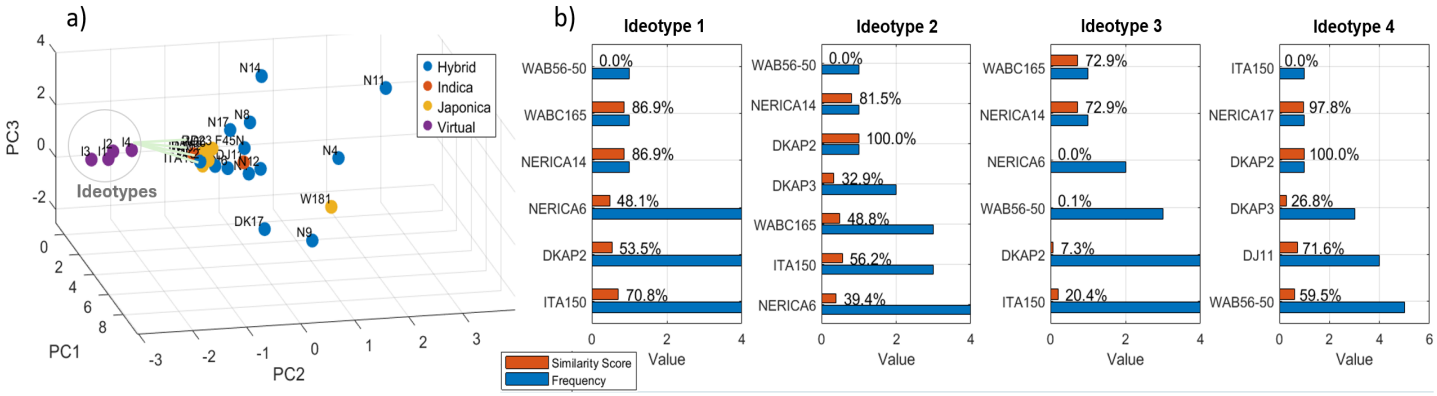


Fig 9. Genetic crop growth parameters characterized with rainfeed field experiment for 21 rice varieties.

compared to the same soil types in the northern region. The harvest index (HI) remained stable across environments at 0.50 ± 0.07, while the average grain number was $16{,}605 \pm 881$.

The average leaf area index (LAI) was 3.63 m^2 m^{-2} in the north, and 16.6% lower in the south (3.03 m^2 m^{-2}). Root depth remained constant at 1 m across regions; however, root density was 2.42 cm cm^{-3} in the north and 13% lower in the south (2.11 cm cm^{-3}).

Evapotranspiration decreased by 19.39% (661 mm) and water use efficiency (WUE) increased by 6.58% (5.83 Kg/mm) in sandy soils, compared to the northern region. Similarly, in clay soils, evapotranspiration decreased by 7.5% (727 mm) and WUE increased by 5.11% (5.14 Kg/mm).

Similarity and affinity of optimized Ideotypes

Each variety and ideotype is characterized by a set of genetically-driven crop growth parameters: $P1, P5, P2R, PHINT, P2O, G1, G2, G3$. Fig. 9-a presents a 3D visualization of the Principal Component Analysis (PCA) conducted on these genetic parameters, which accounts for 86.05% of the total variance. The top four ideotypes, identified as the highest-performing, are highlighted in purple. Varieties based on experimental data are classified into three distinct genetic groups: hybrids (blue), indica (orange), and japonica (yellow). Fig. 9-b depicts the similarity scores and frequencies of these varieties across all distance metrics, illustrating the alignment of genetic traits.

A single variety, **DKAP2**, appears consistently across all four ideotypes, with an average similarity of 65.20% and an average frequency of 3. Three varieties are shared among three ideotypes: **NERICA 14** (80.43% similarity, frequency 1), **WABC165** (69.53% similarity, frequency 1.3), and **ITA150** (49.13% similarity, frequency 4). Two varieties are common to two ideotypes: **NERICA 6** (43.75% similarity, frequency 4) and **WAB56** (29.80% similarity, frequency 4). Finally, **NERICA 17** (97.80% similarity, frequency 1) and **DJ11** (71.60% similarity, frequency 4) were found in only one ideotype each.

Crop parameter optimization relative to current high-performing cultivars

Fig. 10 presents the genetic crop parameters of the eight high-performing cultivars most similar to ideotypes optimized for grain yield and water use efficiency.

The parameter **P1**, representing the thermal time (GDD) from seedling emergence to panicle initiation, averages **400.44 GDD** in the current cultivars. To align with the ideotypes, this phase should be extended by **29.0%**, reaching **564.54 GDD**. This

adjustment would delay floral transition, allowing for increased vegetative growth and potentially greater biomass accumulation prior to reproduction.

In contrast, the grain-filling duration, governed by **P5**, currently averages **328.17 GDD** in the cultivars. Ideotypes require a shorter but more efficient grain-filling phase of **257.89 GDD**, reflecting a **27.2% reduction**. Such optimization suggests faster resource translocation to grains while maintaining yield potential.

The delay in panicle initiation (**P2R**) in existing cultivars is **69.45 GDD**, significantly shorter than the ideotypic value of **155.90 GDD**. A **55.5% increase** in this parameter would synchronize reproductive development with optimal environmental windows, enhancing stress resilience.

G2, the grain weight parameter, currently averages **2.93 g** in the selected cultivars. However, ideotype optimization suggests a reduction to **2.41 g** (**21.7% decrease**), which may be balanced by an increase in spikelet number and overall yield efficiency.

Photoperiod sensitivity (**P2O**) is nearly stable between cultivars (**11.37 GDD**) and ideotypes (**11.70 GDD**), requiring only a minor increase of **2.8%**. This stability ensures adaptability across varying daylengths with minimal adjustment.

G1, denoting spikelets per gram of biomass, is slightly higher in cultivars (**56.63**) compared to ideotypes (**59.24**), indicating a **4.4% increase** is needed to match the ideotypic configuration for optimal sink development.

Tillering potential (**G3**) shows minor difference, with current cultivars at **1.00** and ideotypes at **0.99**, suggesting negligible adjustment.

Lastly, the phyllochron interval (**PHINT**), controlling leaf appearance rate, is shorter in cultivars (**63.17 GDD**) than in ideotypes (**72.36 GDD**), implying that a **12.7% increase** would moderate leaf emergence, potentially improving canopy structure and photosynthetic balance.

Discussion

The sensitivity analysis demonstrates the distinct influences of genetic parameters on the crop growth dynamics modeled by CERES-Rice. Among all variables, P1 (vegetative duration) and PHINT (phyllochron) emerge as the most influential across multiple output indicators, underscoring their determinant role in biomass accumulation, phenological timing, and reproductive development.

P1 (RSI = 0.73) and PHINT (RSI = 0.68) are the most influential parameters for

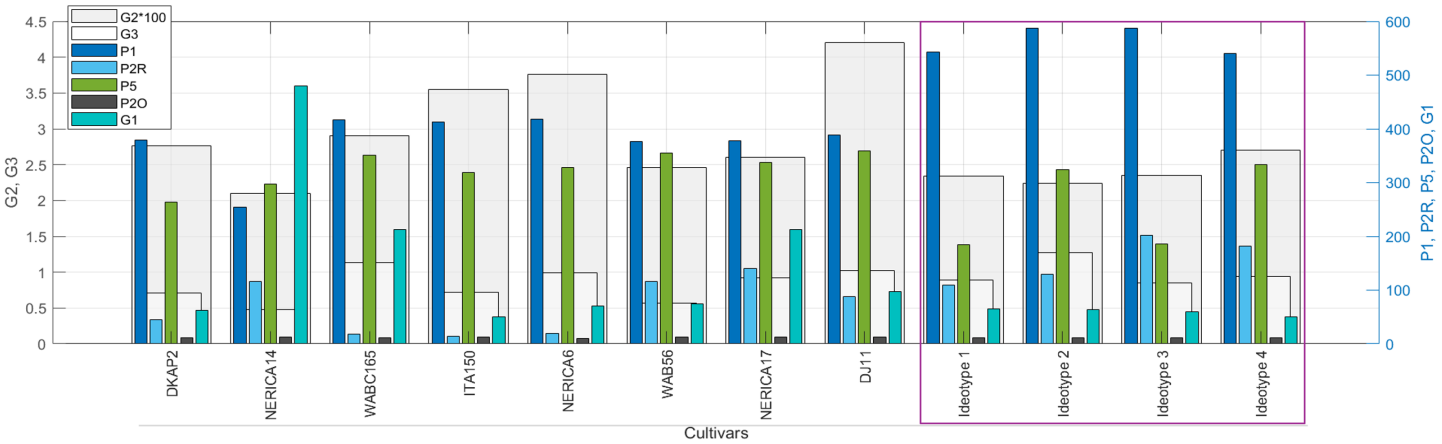


Fig 10. Genetic crop growth parameters of ideotypes and field-characterized varieties with the highest similarity and affinity.

biomass accumulation, highlighting the importance of vegetative duration and leaf emergence rate in biomass partitioning. Grain yield is more sensitive to G3 (tillering coefficient) and P1, with substantial contributions from THOT (heat-induced sterility) and P2O (photoperiod sensitivity). These results indicate the relevance of tillering and reproductive stability in yield formation.

The number of grains is mainly influenced by G2 (grain weight) and G3, while THOT and P2O also play significant roles. This indicates a complex interaction between the plant's ability to allocate assimilated resources to grain development (sink strength), the proportion of spikelets that successfully develop into grains (spikelet survival), and the timing of key developmental stages. For tiller number, G3 (RSI = 0.94) is the dominant parameter, with PHINT and P1 also contributing, reflecting the key role of plant architecture in determining yield potential.

The timing of key phenological stages, including anthesis and maturity, is most affected by P2O, P1, and PHINT. This reflects the close genetic control over development, which is essential for synchronizing growth with favorable environmental conditions, particularly in stress-prone environments.

The analysis highlights the essential role of early vegetative traits in determining plant potential [63, 64] and reproductive stress tolerance [65], which are crucial for yield. Strategic optimization of these traits offers a pathway to boost productivity, particularly under diverse climatic conditions.

The AG optimization of the CERES-Rice model evaluated 1,884 virtual cultivars, identifying an optimal ideotype that combines key agronomic traits. Strong correlations between the HI-WUE index and yield-related traits (crop yield: 0.90, HI: 0.92, WUE: 0.94) highlight the relevance of optimizing both HI and WUE to enhance productivity across diverse environmental conditions.

This work expands current perspective of ideotype optimization by focusing on crop-scale mechanistic modeling by prioritizing functional agronomic parameters such as grain conversion efficiency and water use efficiency as key indicators of performance. The integrated HI-WUE index employed here offers a effective metric for balancing yield and water use, which is critical for improving sustainability in the face of climate variability. The optimization results expose that HI-WUE values span a range from 0.95 to 1.14 across different environments, illustrating the potential benefit for environment-specific adjustments to achieve optimal performance.

The higher solar radiation and lower humidity in the northern region were more favorable for enhanced photosynthetic efficiency, as reflected in the higher root density and leaf area index (LAI), leading to better overall performance [66]. In contrast, the southern region's higher humidity and lower solar radiation resulted in a 16.6% reduction in LAI and weaker root development, limiting biomass and yield. This finding underscores the critical role of genotype-by-environment (G×E) interactions with regional climate characteristics in determining crop performance, highlighting the challenges they pose for standardizing effective breeding strategies [24].

The ideotype's response to water stress, particularly in the southern region, highlighted its capacity for water use efficiency (WUE) plasticity. In sandy soils, WUE improved by 6.58%, with a corresponding reduction in evapotranspiration (ET) (19.39%), suggesting a more efficient use of limited water. In clay soils, though ET decreased by 7.5%, WUE still improved by 5.11%, pointing to the ideotype's ability to conserve water across diverse soil conditions. However, while this water conservation is crucial in drought-prone areas, it comes at the cost of biomass production, particularly in the southern region, where the reduced solar radiation further constrained photosynthetic capacity. Nevertheless, this dynamic distribution may vary depending on planting dates aligned with the region's seasonal rainfall patterns [67].

The comparative analysis between optimized ideotypes and existing cultivars reveals a significant genetic-based gap, underscoring the need for targeted breeding interventions. The multidimensional similarity scores clearly demonstrate that none of the current cultivars fully align with the optimized genetic profiles. However, some cultivars, such as NERICA 14 (80.43% similarity) and DKAP2 (65.20% similarity), exhibit partial affinity. This partial alignment suggests promising starting points for breeding programs, while also confirming that the current genotypes do not fully meet the ideotypic ideal. These results are consistent with the analysis of upland rice genotype-environment interactions in Senegal, which showed that NERICA 14 achieved the highest yield across the environments [68].

The proposed genetic reconfiguration, specifically the 29% extension of the vegetative phase (P1) and the 27.2% reduction in the grain-filling period (P5), represents a strategic decoupling of vegetative growth and reproductive development to enhance carbon assimilation and translocation efficiency. This adjustment supports greater biomass accumulation without compromising yield, aligning with the ideotype's objective to optimize both harvest index and water use efficiency (WUE). Additionally, the substantial delay in panicle initiation (P2R, +55.5%) and the moderate extension of the phyllochron interval (PHINT, +12.7%) are critical for aligning crop development with environmental conditions, thereby enhancing resilience under increasingly variable climate regimes.

These findings support the predictive potential of mechanistic models and AI-based optimization strategies in defining physiologically grounded ideotypes. The integration of data-driven methods provides breeders with structured guidance for selection, including specific targets such as vegetative growth determinants or optimal grain weight (G2), which can directly inform parental selection and crossing schemes. Furthermore, the multidimensional similarity analysis establishes a methodology for identifying elite candidates for crossing or introgression.

Conclusion

This study demonstrates the progressive potential of integrating mechanistic crop models with AI-driven advancements to define physiologically grounded, high-efficiency ideotypes.

The sensitivity analysis exposes a non-uniform pattern of genetic control, where developmental timing traits, particularly P1, P2O, and PHINT, manifest as systemic regulators of yield-related outcomes. Their consistent influence on both productivity and phenological traits underscores the importance of coordinating growth phases rather than focusing exclusively on isolated morphological improvements. This suggests that ideotype design should prioritize the coordination of morphological traits with their temporal dynamics.

The proposed ideotype, defined by the integrated HI-WUE index, demonstrates significant potential for optimizing both grain yield and water use efficiency. The model's ability to adapt to varying environmental conditions, such as solar radiation and humidity, reinforces the importance of genotype-by-environment interactions in determining crop performance. Optimizing these traits offers a pathway to enhance productivity while conserving resources in diverse climatic contexts.

The integration of distance metrics, frequency calculations, and similarity indices constitutes a robust framework for cultivar evaluation and comparison. This methodology not only identifies the cultivars most closely related to virtual varieties (ideotypes) but also effectively quantifies and represents their degree of genetic similarity. Comparing the optimized ideotype with existing cultivars reveals a notable genetic gap.

Although some cultivars, such as NERICA 14 and DKAP2, show partial alignment with the ideal genetic profile, the findings suggest that targeted breeding interventions are essential. These interventions could focus on improving the identified key traits, advancing the development of cultivars more suited to specific environmental conditions.

This study provide insights regarding the potential of mechanistic models and AI-driven approaches in crop breeding, offering a more efficient, data-driven alternative to traditional trial-and-error methods. By integrating mechanistic models with optimization techniques, future breeding programs can refine ideotype design, facilitating the development of high-performing, climate-resilient crop varieties.

The framework defines optimal genetic-based and physiological traits for rice ideotypes across diverse environmental conditions. Notably, the integration of genetic optimization with field validation ensures that these ideotypes are both theoretically optimal and practically viable for agricultural implementation.

This study reinforces the critical role of physiologically grounded crop parameters, such as leaf emergence rate, grain filling duration, and biomass partitioning, as actionable targets in cultivar selection and ideotype design. These parameters, though modeled at the canopy or organ scale, are underpinned by identifiable genetic elements at the level of genes, loci, and allelic variants. Bridging these physiological traits with their genetic architecture enables breeders not only to simulate ideal phenotypes but also to refine selection based on a cultivar's genetic capacity to express high-performing traits under specific environmental conditions.

The successful application of this approach is grounded in the model's ability to comprehensively reproduce the behavior of individual varieties, as evaluated through calibration and validation metrics. In this study, a crop yield prediction efficacy of 0.66 was achieved, as reported by Gerardeaux et al. (2021), using rice as the target species. While the results are crop-specific, the underlying methodology can be generalized to other cereal crops and modeling frameworks following a similar approach. While mechanistic models may not yet fully capture genetic complexity at the gene network or 3D architectural levels, their capacity to represent functional genetic diversity remains a promising asset. The iterative co-development of crop models and phenotyping platforms, driven by collaboration between biologists and modelers, will be essential for generating robust ideotype predictions and translating them into practical breeding outcomes. This work contributes to that vision by providing a concrete framework for ideotype optimization that is both genetically informed and environmentally responsive.

Author Contributions

Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing—original draft, Writing—review and editing, Visualization: Edgar S. Correa.

Data and Code Availability

All data and materials used in this study are provided within the manuscript. Additional simulation outputs and environmental input files are publicly available at:

Correa, Edgar S. (2025). *Ideotype Optimization with Crop Modelling: CERES-Rice*. Mendeley Data, V1. `https://doi.org/10.17632/msg7bfg5t6.1`.

All code for model optimization, sensitivity analysis, and figure generation is openly accessible at: `https://github.com/EdgarStevenC/Crop-Growth-Modelling`, under the Creative Commons Attribution 4.0 International License (CC BY 4.0).

References

1. McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H, Sun Q, et al. Genetic Properties of the Maize Nested Association Mapping Population. Science. 2009;325(5941):737–740. doi:10.1126/science.1174320.

2. Araus JL, Cairns JE. Field high-throughput phenotyping: the new crop breeding frontier. Trends in Plant Science. 2014;19(1):52–61. doi:https://doi.org/10.1016/j.tplants.2013.09.008.

3. Farooq MA, Gao S, Hassan MA, Huang Z, Rasheed A, Hearne S, et al. Artificial intelligence in plant breeding. Trends in Genetics. 2024;40(10):891–908. doi:https://doi.org/10.1016/j.tig.2024.07.001.

4. Arvanitis KG, Symeonaki EG. Agriculture 4.0: The Role of Innovative Smart Technologies Towards Sustainable Farm Management. The Open Agriculture Journal. 2020;14:130–135. doi:https://doi.org/10.2174/1874331502014010130.

5. Zhai Z, Martínez JF, Beltran V, Martínez NL. Decision support systems for agriculture 4.0: Survey and challenges. Computers and Electronics in Agriculture. 2020;170:105256. doi:https://doi.org/10.1016/j.compag.2020.105256.

6. Noshita K, Murata H, Kirie S. Model-based plant phenomics on morphological traits using morphometric descriptors. Breeding Science. 2022;72(1):19–30.

7. Colorado JD, Calderon F, Mendez D, Petro E, Rojas JP, Correa ES, et al. A novel NIR-image segmentation method for the precise estimation of above-ground biomass in rice crops. PloS one. 2020;15(10):e0239591.

8. Jimenez-Sierra DA, Correa ES, Benítez-Restrepo HD, Calderon FC, Mondragon IF, Colorado JD. Novel feature-extraction methods for the estimation of above-ground biomass in rice crops. Sensors. 2021;21(13):4369.

9. Correa ES, Calderon F, Colorado JD. GFkuts: a novel multispectral image segmentation method applied to precision agriculture. 2020; p. 1–6. doi:10.1109/OMICAS52284.2020.9535659.

10. Correa ES, Calderon FC, Colorado JD. A Novel Multi-camera Fusion Approach at Plant Scale: From 2D to 3D. SN Computer Science. 2024;5(5). doi:10.1007/s42979-024-02849-7.

11. Correa ES, Parra CA, Vizcaya PR, Calderon FC, Colorado JD. Complex Object Detection Using Light-Field Plenoptic Camera. Communications in Computer and Information Science. 2022;1576 CCIS:119 – 133. doi:10.1007/978-3-031-07005-$1_1$2.

12. Fatichi S, Pappas C, Ivanov VY. Modeling plant–water interactions: an ecohydrological overview from the cell to the global scale. Wiley Interdisciplinary Reviews: Water. 2016;3(3):327 – 368. doi:10.1002/wat2.1125.

13. Wagner GP, Altenberg L. Complex adaptations and the evolution of evolvability. Evolution. 1996;50(3):967 – 976. doi:10.1111/j.1558-5646.1996.tb02339.x.

14. Pasley H, Brown H, Holzworth D, Whish J, Bell L, Huth N. How to build a crop model. A review. Agronomy for Sustainable Development. 2023;43(1). doi:10.1007/s13593-022-00854-9.

15. Yin X, Struik PC. Modelling the crop: from system dynamics to systems biology. Journal of Experimental Botany. 2010;61(8):2171–2183. doi:10.1093/jxb/erp375.

16. Muller B, Martre P. Plant and crop simulation models: powerful tools to link physiology, genetics, and phenomics. JOURNAL OF EXPERIMENTAL BOTANY. 2019;70(9, SI):2339–2344. doi:10.1093/jxb/erz175.

17. Stöckle CO, Donatelli M, Nelson R. CropSyst, a cropping systems simulation model. European Journal of Agronomy. 2003;18(3):289–307. doi:https://doi.org/10.1016/S1161-0301(02)00109-0.

18. Ritchie J, Singh U, Godwin D, Bowen W. Cereal growth, development and yield. Understanding options for agricultural production. 1998; p. 79–98.

19. Boote KJ, Sau Sau F, Porter CH, Dzotsi K, Tollenaar M, Kumudini SV, et al. Testing of crop Models for Accurate Predictions of Evapotranspiration and crop Water Use. 2014;.

20. Amiri E, Rezaei M, Rezaei EE, Bannayan M. Evaluation of Ceres-Rice, Aquacrop and Oryza2000 models in simulation of rice yield response to different irrigation and nitrogen management strategies. Journal of Plant Nutrition. 2014;37(11):1749–1769.

21. Parent B, Tardieu F. Can current crop models be used in the phenotyping era for predicting the genetic variability of yield of plants subjected to drought or high temperature? Journal of Experimental Botany. 2014;65(21):6179 – 6189. doi:10.1093/jxb/eru223.

22. Chardon F, Noël V, Masclaux-Daubresse C. Exploring NUE in crops and in Arabidopsis ideotypes to improve yield and seed quality. Journal of Experimental Botany. 2012;63(9):3401–3412. doi:10.1093/jxb/err353.

23. Redden R. New Approaches for Crop Genetic Adaptation to the Abiotic Stresses Predicted with Climate Change. Agronomy. 2013;3(2):419–432. doi:10.3390/agronomy3020419.

24. Gu J. Ecophysiological Crop Modelling Combined with Genetic Analysis Is a Powerful Tool for Ideotype Design. Agronomy. 2022;12(1). doi:10.3390/agronomy12010215.

25. Rasheed A, Hassan MU, Aamer M, Batool M, Fang S, WU Z, et al. A Critical Review on the Improvement of Drought Stress Tolerance in Rice (Oryza sativa L.). Notulae Botanicae Horti Agrobotanici Cluj-Napoca. 2020;48(4):1756 – 1788. doi:10.15835/48412128.

26. Watson A, Ghosh S, Williams MJ, Cuddy WS, Simmonds J, Rey MD, et al. Speed breeding is a powerful tool to accelerate crop research and breeding. NATURE PLANTS. 2018;4(1):23–29. doi:10.1038/s41477-017-0083-8.

27. Picheny V, Casadebaig P, Trépos R, Faivre R, Da Silva D, Vincourt P, et al. Using numerical plant models and phenotypic correlation space to design achievable ideotypes. Plant Cell and Environment. 2017;40(9):1926 – 1939. doi:10.1111/pce.13001.

28. Sambakhé D, Gozé E, Bacro JN, Dingkuhn M, Adam M, Ndiaye M, et al. Ideotype map research based on a crop model in the context of a climatic gradient. Ecological Modelling. 2024;498:110840. doi:https://doi.org/10.1016/j.ecolmodel.2024.110840.

29. Rötter RP, Tao F, Höhn JG, Palosuo T. Use of crop simulation modelling to aid ideotype design of future cereal cultivars. Journal of Experimental Botany. 2015;66(12):3463–3476. doi:10.1093/jxb/erv098.

30. Li Z, Ding Q, Zhang W. A Comparative Study of Different Distances for Similarity Estimation. In: Chen R, editor. Intelligent Computing and Information Science. Berlin, Heidelberg: Springer Berlin Heidelberg; 2011. p. 483–488.

31. Kokare M, Chatterji BN, Biswas PK. Comparison of similarity metrics for texture image retrieval. In: TENCON 2003. Conference on Convergent Technologies for Asia-Pacific Region. vol. 2; 2003. p. 571–575 Vol.2.

32. Gérardeaux E, Falconnier G, Gozé E, Defrance D, Kouakou PM, Loison R, et al. Adapting rainfed rice to climate change: a case study in Senegal. Agronomy for Sustainable Development. 2021;41(4):57. doi:10.1007/s13593-021-00710-2.

33. Hengl T, Mendes de Jesus J, Heuvelink GBM, Ruiperez Gonzalez M, Kilibarda M, Blagotić A, et al.. SoilGrids250m: Global gridded soil information based on machine learning; 2017. Available from: `https://soilgrids.org`.

34. Romero CC, Hoogenboom G, Baigorria GA, Koo J, Gijsman AJ, Wood S. Reanalysis of a global soil database for crop and environmental modeling. Environmental Modelling Software. 2012;35:163–170. doi:https://doi.org/10.1016/j.envsoft.2012.02.018.

35. Alaya I, Masmoudi MM, Lagacherie P, Coulouma G, Jacob F, Ben Mechlia N. Performance of Saxton and Rawls pedotransfer functions for estimating soil water properties in the Cap Bon Region-Northern Tunisia. Water and Land Security in Drylands: Response to Climate Change. 2017; p. 77–85. doi:10.1007/978-3-319-54021-4_8.

36. Sung CTB, Iba J. Accuracy of the Saxton-Rawls method for estimating the soil water characteristics for mineral soils of Malaysia. Pertanika J Trop Agric Sci. 2010;33(2):297–302.

37. Rawls WJ, Brakensiek DL, Saxtonn K. Estimation of soil water properties. Transactions of the ASAE. 1982;25(5):1316–1320.

38. Hodges R, et al.. NASA POWER (Prediction Of Worldwide Energy Resources) Data; 2017.

39. Correa ES, Calderon FC, Colorado JD. AI-driven remote sensing for environmental characterization and rice crop modeling in water-limited regions; 2025. Preprint, Research Square. Available from: `https://doi.org/10.21203/rs.3.rs-6440110/v1`.

40. Boote KJ, Jones JW, Hoogenboom G. Incorporating realistic trait physiology into crop growth models to support genetic improvement. In Silico Plants. 2021;3(1). doi:10.1093/insilicoplants/diab002.

41. Xu L, Henke M, Zhu J, Kurth W, Buck-Sorlin G. A functionalstructural model of rice linking quantitative genetic information with morphological development and physiological processes. Annals of Botany. 2011;107(5):817 – 828. doi:10.1093/aob/mcq264.

42. Russelle MP, Wilhelm WW, Olson RA, Power JF. Growth Analysis Based on Degree Days. Crop Science. 1984;24(1):28–32. doi:10.2135/cropsci1984.0011183X002400010007x.

43. McMaster GS, Wilhelm WW. Growing degree-days: One equation, two interpretations. Agricultural and Forest Meteorology. 1997;87(4):291 – 300. doi:10.1016/S0168-1923(97)00027-0.

44. Fraisse CW, Paula-Moraes SV. Degree-Days: Growing, Heating, and Cooling: ABE381/AE428, rev. 4/2018. EDIS. 2018;2018(2). doi:10.32473/edis-ae428-2018.

45. Zhou G, Wang Q. A new nonlinear method for calculating growing degree days. Scientific Reports. 2018;8(1). doi:10.1038/s41598-018-28392-z.

46. Wallach D, Palosuo T, Thorburn P, Mielenz H, Buis S, Hochman Z, et al. Proposal and extensive test of a calibration protocol for crop phenology models. Agronomy for Sustainable Development. 2023;43(4). doi:10.1007/s13593-023-00900-0.

47. Seidel SJ, Palosuo T, Thorburn P, Wallach D. Towards improved calibration of crop models – Where are we now and where should we go? European Journal of Agronomy. 2018;94:25–35. doi:https://doi.org/10.1016/j.eja.2018.01.006.

48. Cao W, Moss DN. Temperature effect on leaf emergence and phyllochron in wheat and barley. Crop Science. 1989;29(4):1018 – 1021. doi:10.2135/cropsci1989.0011183X002900040038x.

49. Frank AB, Bauer A. Phyllochron Differences in Wheat, Barley, and Forage Grasses. Crop Science. 1995;35(1):cropsci1995.0011183X003500010004x. doi:https://doi.org/10.2135/cropsci1995.0011183X003500010004x.

50. YAN W, WALLACE DH. Simulation and Prediction of Plant Phenology for Five Crops Based on Photoperiod×Temperature Interaction. Annals of Botany. 1998;81(6):705–716. doi:10.1006/anbo.1998.0625.

51. Yang W, Wu T, Zhang X, Song W, Xu C, Sun S, et al. Critical photoperiod measurement of soybean genotypes in different maturity groups. Crop Science. 2019;59(5):2055 – 2061. doi:10.2135/cropsci2019.03.0170.

52. Gunawardena TA, Fukai S, Blamey FPC. Low temperature induced spikelet sterility in rice. II. Effects of panicle and root temperatures. Australian Journal of Agricultural Research. 2003;54(10):947 – 956. doi:10.1071/AR03076.

53. Zeng Y, Zhang Y, Xiang J, Uphoff NT, Pan X, Zhu D. Effects of low temperature stress on spikelet-related parameters during anthesis in indica–Japonica hybrid rice. Frontiers in Plant Science. 2017;8. doi:10.3389/fpls.2017.01350.

54. Shimono H, Okada M, Kanda E, Arakawa I. Low temperature-induced sterility in rice: Evidence for the effects of temperature before panicle initiation. Field Crops Research. 2007;101(2):221–231. doi:https://doi.org/10.1016/j.fcr.2006.11.010.

55. Shimono H, Hasegawa T, Moriyama M, Fujimura S, Nagata T. Modeling spikelet sterility induced by low temperature in rice. Agronomy Journal. 2005;97(6):1524 – 1536. doi:10.2134/agronj2005.0043.

56. Xu Y, Chu C, Yao S. The impact of high-temperature stress on rice: Challenges and solutions. The Crop Journal. 2021;9(5):963–976. doi:https://doi.org/10.1016/j.cj.2021.02.011.

57. Satake T, Yoshida S. High temperature-induced sterility in indica rices at flowering. Japanese Journal of Crop Science. 1978;47(1):6 – 17. doi:10.1626/jcs.47.6.

58. Jagadish S, Craufurd P, Wheeler T. High temperature stress and spikelet fertility in rice (Oryza sativa L.). Journal of Experimental Botany. 2007;58(7):1627–1635. doi:10.1093/jxb/erm003.

59. Ren H, Bao J, Gao Z, Sun D, Zheng S, Bai J. How rice adapts to high temperatures. Frontiers in Plant Science. 2023;14. doi:10.3389/fpls.2023.1137923.

60. Morris MD. Factorial sampling plans for preliminary computational experiments. Technometrics. 1991;33(2):161 – 174. doi:10.1080/00401706.1991.10484804.

61. Pang Z, O'Neill Z, Li Y, Niu F. The role of sensitivity analysis in the building performance analysis: A critical review. Energy and Buildings. 2020;209:109659. doi:https://doi.org/10.1016/j.enbuild.2019.109659.

62. Fukai S, Mitchell J. Factors determining water use efficiency in aerobic rice. Crop and Environment. 2022;1(1):24 – 40. doi:10.1016/j.crope.2022.03.008.

63. Huang M, Zhang R, Chen J, Cao F, Jiang L, Zou Y. Morphological and physiological traits of seeds and seedlings in two rice cultivars with contrasting early vigor. Plant Production Science. 2017;20(1):95 – 101. doi:10.1080/1343943X.2016.1229571.

64. Rebolledo MC, Dingkuhn M, Péré P, Mcnally KL, Luquet D. Developmental Dynamics and Early Growth Vigour in Rice. I. Relationship Between Development Rate (1/Phyllochron) and Growth. Journal of Agronomy and Crop Science. 2012;198(5):374 – 384. doi:10.1111/j.1439-037X.2012.00528.x.

65. Ravikiran KT, Gopala Krishnan S, Abhijith KP, Bollinedi H, Nagarajan M, Vinod KK, et al. Genome-Wide Association Mapping Reveals Novel Putative Gene Candidates Governing Reproductive Stage Heat Stress Tolerance in Rice. Frontiers in Genetics. 2022;13. doi:10.3389/fgene.2022.876522.

66. Huang M, Shan S, Cao F, Zou Y. The solar radiation-related determinants of rice yield variation across a wide range of regions. NJAS - Wageningen Journal of Life Sciences. 2016;78:123 – 128. doi:10.1016/j.njas.2016.05.004.

67. Feng X, Porporato A, Rodriguez-Iturbe I. Changes in rainfall seasonality in the tropics. Nature Climate Change. 2013;3(9):811 – 815. doi:10.1038/nclimate1907.

68. Kanfany G, Ayenan MAT, Zoclanclounon YAB, Kane T, Ndiaye M, Diatta C, et al. Analysis of Genotype-Environment Interaction and Yield Stability of Introduced Upland Rice in the Groundnut Basin Agroclimatic Zone of Senegal. Advances in Agriculture. 2021;2021. doi:10.1155/2021/4156167.