Ideotype map research based on a crop model in the context of a climatic gradient

Diariétou Sambakhé, Eric Gozé, Jean-Noël Bacro, Michael Dingkuhn, Myriam Adam, Malick Ndiaye, Bertrand Müller, Lauriane Rouan

Ecological Modelling 498 (2024) 110840

Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel

Ideotype map research based on a crop model in the context of a climatic gradient

Diariétou Sambakhé [a,b,d], Eric Gozé [b,c], Jean-Noël Bacro [d], Michael Dingkuhn [e,f], Myriam Adam [f,g,i], Malick Ndiaye [j], Bertrand Muller [a,f,h], Lauriane Rouan [e,f,*]

[a] Institut Sénégalais de Recherches Agricoles, Centre d'Étude Régional pour l'Amélioration de l'Adaptation à la Sécheresse (CERAAS), BP 3320 Thiès Escale Thiès, Senegal
[b] CIRAD - French Agricultural Research Centre for International Development, UPR AIDA, F-34398, Montpellier, France
[c] AIDA, CIRAD, Université de Montpellier, Montpellier, France
[d] Montpellier University, Institut Montpelliérain Alexander Grothendieck (IMAG), 34000, France
[e] CIRAD - French Agricultural Research Centre for International Development, UMR AGAP Institut, F-34398 Montpellier, France
[f] AGAP Institut, CIRAD, INRAE, Institut Agro, Université de Montpellier, Montpellier, France
[g] CIRAD - French Agricultural Research Centre for International Development, UMR AGAP Institut, Bobo Dioulasso, Burkina Faso
[h] CIRAD - French Agricultural Research Centre for International Development, UMR AGAP Institut, BP 3320 Thiès Escale Thiès, Senegal
[i] Institut National de l'Environnement et de Recherches Agricoles (INERA), Burkina Faso
[j] ISRA/Centre National de Recherches Agronomiques, BP53-Bambey, Senegal

ARTICLE INFO

Keywords:
Experimentation
Crop model
Varietal parameters
Input distribution
Stochastic model
Conditional optimization
Noisy function
PEQI criterion

ABSTRACT

Due to increasing climate uncertainties, optimizing plant traits is essential for sustainable agriculture. This article presents an approach that combines advanced modelling techniques to identify optimal plant traits under various agro-environmental conditions. By integrating a crop model, a climate generator, and our PEQI algorithm (Profile Expected Quantile Improvement), our method aims to create ideotype maps tailored to specific regions.

We use the SAMARA model (Simulator of crop trait Assembly, MAnagement Response, and Adaptation), calibrated with trials carried in Sahel on a set of local varieties, to simulate crop growth in diverse environments. The PEQI algorithm adjusts varietal parameters to maximize expected yield, defining precise selection objectives known as ideotypes, which are particularly important in regions with irregular rainfall patterns like the Sahel.

With the PEQI algorithm based on a kriging metamodel, we ensure effective adaptation to spatially variable environments. By leveraging a climate generator to simulate meteorological variability, our integrated approach optimizes crop yields in regions such as Senegal, southern Mali, Burkina Faso, and Guinea-Bissau. The outcome is an ideotype map for sorghum, providing breeders with a robust decision-support tool to enhance crop performance amidst climate uncertainty.

1. Introduction

As the world's population grows, there is a constant need to increase agricultural production. In the Sahel, where malnutrition is the highest in the world (OCDE, 2014), the saturation of arable land means that new land is rare or fallow periods are shortened. To meet the growing demands of a burgeoning population, production per unit of cultivated land must be increased to raise crop yields.

Increasing yields requires improving crop and their growing environment, which is modified by cultural practices (i.e. tillage, fertilisation, weed management, etc.). For any given crop, the best variety is not necessarily the same in different environments. Therefore, most cultivated crop species exist in the form of numerous varieties, adapted to both the environment they occupy and the use for which they are intended. In temperate climates, for reasons of both marketing convenience and geographical structuring of the environment, the environment is divided into mega-environments, which are assumed to be homogeneous. In the Sahel, the definition of mega-environments is questionable because the environment is strongly constrained by the availability of water in the soil. Indeed, there is a continuous gradient from the Sahara in the north to the so-called Sudan climate zone in the south, where water is more abundant for sorghum cultivation (see Fig. 1;

* Corresponding author.
E-mail address: lauriane.rouan@cirad.fr (L. Rouan).

https://doi.org/10.1016/j.ecolmodel.2024.110840
Received 18 December 2023; Received in revised form 25 July 2024; Accepted 9 August 2024
Available online 12 September 2024

(Fensholt et al., 2013)).

Historically, it was farmers who practised plant breeding. They saved the best of their production as seed and exchanged genetic resources between families through alliances (Labeyrie et al., 2016). Since the scientific discovery of heredity (Yule, 1902), genetic improvement specialists have gradually replaced farmers in this task, leading to significant increases in cereal yields. Genetic improvement involves several phases defining the specifications, searching for donors with the traits of interest, carrying out crosses that randomly combine the different alleles of the genes that determine these traits in the offspring, and selecting the best offspring. The evaluation phase of the resulting varieties is challenging because the environment to which the varieties must adapt changes every year. The best variety in one year may not be the best in the next year or in future years; this is known as genotype by year interaction. While this interaction makes it difficult to release varieties in temperate countries, it makes it even more difficult to evaluate varieties in the Sahel, where the rainy seasons vary greatly from year to year (Sultan et al., 2005). In addition, the irregularity and spatial variability of rainfall events greatly affect crop production, especially in rainfed areas. In this context, the expected yield of a given variety becomes difficult to estimate and a large number of observations are required for accuracy. It is even more so for yield variance or for quantiles, such as the yield exceeded 8 years out of 10. It is therefore difficult to select the most stable varieties at a limited cost. A less empirical alternative would be to identify plant adaptation traits that would allow prediction of plant responses to environmental variation. For example, a variety with a deep-rooting system is more likely than others to withstand prolonged drought. However, direct observation of drought resistance can be subject to the randomness of water shortages, whereas root length measurements are reproducible, provided the conditions under which they are taken are well defined.

The a priori approach to constructing a variety by drawing up a list of desirable characteristics according to the knowledge we have of the environment and the functioning of the plant is the concept of an ideotype. Donald (1968) introduced this concept and constructed an ideotype of cereal by pure reasoning. However, the functioning of a plant is complex, so the search for ideotypes by simple reasoning by experts is necessarily subjective. The mathematical formalisation of physiological phenomena and the computer simulation of the resulting equations to predict the response of the plant to its environment is the subject of various types of crop models (Landsberg and de Wit, 1980; Muller and Martre, 2019).

A crop model describes the functioning of a soil-plant system and a plant's response to its environment. It is based on mathematical equations involving environmental and plant variables. Some models are sufficiently generic to cover multiple varieties or even species (Asseng, van Keulen and Stol, 2000). So-called variety parameters can be used to modify the plant's response to the environment in the model.

A crop model can thus be considered a relevant means of defining the ideal plant in a target environment, based on the parameters that allow it to best respond to the growth conditions of that environment. For example, if the goal is to achieve a high yield in a given environment, one would seek the values of the variety parameters that maximize yield under these conditions. For the same set of variety parameters, any change in environmental conditions should alter the model's results. Therefore, for future conditions with a known probability distribution, one can try to maximize yield expectation, minimize its variance, or maximize a trade-off, which could be a production quantile. However, the estimation of these moments or quantiles as a function of varietal parameters is, in practice, based on a limited number of simulations and is therefore subject to error.

To determine a gradient of the best varieties across a gradient of target environments, conditional optimization based on the environment must be considered. From a methodological point of view, this involves finding, for each location u, i.e., each pixel on a map, the variety that maximizes yield. In other words, performing optimization of the crop model parameters for each point on a map. This is a time-consuming task. In this paper, we present a less time-consuming methodological approach. If it is assumed that, with slow spatial variations in soil and climate, the optimal yield at a location ($u+\delta$) close to (u) is similar to the optimal yield at (u), it is possible to fit a response surface to account for this dependency. The response surface is an approximation of the crop model called a metamodel. Among the possible metamodels, the Gaussian process, and with it kriging as a prediction method, has already yielded remarkable results in optimization (Bonte et al., 2008; Janusevskis and Le Riche, 2013; Queipo et al., 2005). Slow variations in space are then seen as realizations of a random field with correlations between neighbors (Cressie, 1991; Chilès & Delfiner, 2012).

In this paper, we will show how a conditional optimization method based on this type of metamodel and the PEQI criterion (Profile Expected Quantile Improvement) we developed (Sambakhé et al, 2019) can be used to establish a map of ideotypes without repeating optimizations from scratch in neighbouring locations. To illustrate our approach, we will focus on defining sorghum ideotypes, one of the most consumed cereals in the Sahel region, and show how to obtain a map of sorghum ideotypes suited to the current and future Sahelian climate using a crop model. Specifically, for this proof of concept with user-chosen parameters, we will demonstrate how to obtain maps of optimal values which maximize the expected yield of the crop.

2. Material and methods

2.1. Multi-local experiment for model calibration and evaluation

2.1.1. Experimental sites

The SAMARA model (Dingkuhn et al., 2011) was calibrated and evaluated on the basis of experimental data collected during four consecutive rainy seasons (2013, 2014, 2015 and 2016) in three

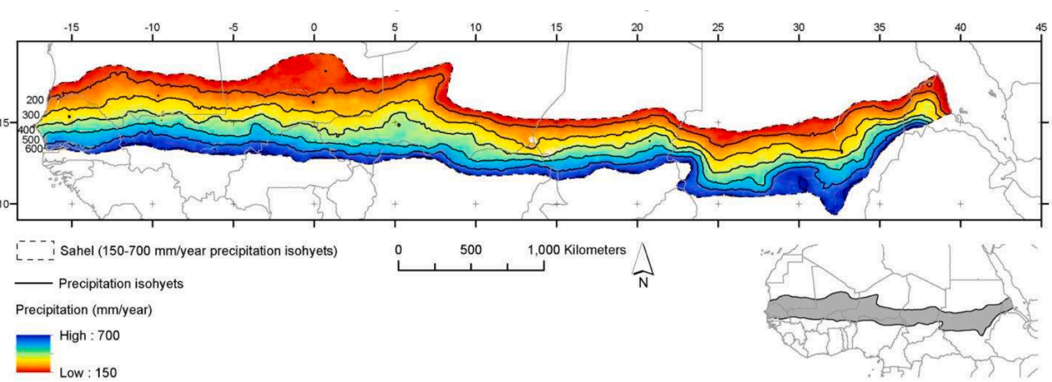

Fig. 1. North to south gradient of annual rainfall in the Sahel (Fensholt et al., 2013).

agro-ecological zones, namely Bambey (northern peanut basin), Nioro du Rip (southern peanut basin) and Sinthiou Malem (eastern Senegal). These areas have different soil and climatic conditions. The characteristics of these different areas are shown in Table 1.

Bambey Station is subject to a typical Sahelian climate, characterised by a long dry season of 8 to 9 months and a rainy season of 3 to 4 months. Rainfall varies greatly from year to year. The predominant soils are sandy with a very low water retention capacity of 80 to 100 $mm.m^{-1}$ (Affholder, 1997). The Nioro du Rip and Sinthiou Malem stations are located at the interface between the Sahelian and Sudanian zones. They benefit from a rainy season of 4 to 5 months, which is wetter than in Bambey, but still characterised by strong interannual variability. The soils are still predominantly sandy, but have a slightly higher clay and silt content and a water retention capacity of 90 to 120 $mm.m^{-1}$ (Baret, 1980).

2.1.2. Plant material

The plant material consists of ten varieties from different regions of West and Central Africa. Each variety is known to perform well in its area of distribution. They were selected to provide a contrasting sample in terms of cycle length (each adapted to its target region), architecture (height, stem diameter in particular), structural characteristics (lignin, cellulose), grain production and biomass (Table 2).

2.1.3. Measured variables

In the trials conducted, observations were made on flowering and maturity dates, number of emerged leaves, leaf area index (LAI), biomass and plant height dynamics, yields (grain and biomass) and their components. Total above-ground biomass and its distribution among stems, leaves and panicles, leaf area index and number of emerged leaves were monitored every week after flowering (see (Ndiaye et al., 2021) for more details).

The main characteristics of these trials (soil type, sowing date, standard daily climatic data - minimum, maximum and average temperatures, minimum, maximum and average relative humidity, global radiation, average wind speeds at 2 m above the ground and rainfall heights) were also measured.

2.2. Modelling yield as a response to climate and crop parameters

This study aims to define a map of sorghum ideotypes using a crop model. Crop models differ in the formalisms they use to represent the soil-plant system in interaction with climate. We used the SAMARA model (Dingkuhn et al., 2011), which is a crop simulation model that operates on a daily time step basis. As a crop model requires some soil and meteorological inputs to produce a yield prediction, ideotyping for a given climate requires that the (joint) probability distribution of these meteorological inputs is known or at least sampled. While this is the case for a limited number of so-called synoptic meteorological stations, these stations are irregularly distributed and scarce in developing countries, so the meteorological data collected in the current climate are not sufficient to produce an ideotype map. However, using a stochastic model, the probability distribution of the inputs can be summarised by a few parameters that can be interpolated between synoptic stations, thus predicting these distributions at all points. An analytical mapping of this distribution of weather inputs of a crop model to that of its outputs is not feasible, so the Monte Carlo method is used by generating equiprobable weather sequences under a given climate model. The climate model is then represented as a weather generator. Moreover, meteorological records are not (yet) available under the future climate: to obtain samples, one also needs a weather generator. Therefore, for this study, we coupled the SAMARA model with the MarkSim stochastic weather generator (Peter G. Jones and Thornton, 2000).

Table 1
Mean rainfall and temperature at the three experimental sites during the rainy season.

Locations	Positions	Soil type	rain	Tmin	Tmax
Bambey	14°42'N 16°28'W	Sandy soil	500	20	37
Nioro du Rip	13°45'N 15°45'W	Sandy soil	600	22	36
Sinthiou Malem	13°49'N 13°55'W	Loamy sandy soil	700	22	36

2.2.1. Overview of the SAMARA crop model for sorghum growth and yield simulation

Crop models use almost the same concepts to represent the mechanisms of plant growth and development. However, they differ in the way physiological assumptions made, the environmental factors considered, and how they are formalised. The SAMARA model, developed to study plant concepts *in silico* under different climatic environments, different soil types and different cultural practices (Dingkuhn et al., 2011; Kumar et al., 2016, & 2017), is well suited to find trade-offs between plant traits in order to optimise an agronomic outcome such as grain yield. It is based on a carbon and water balance. The capture of light and its transformation by photosynthesis is the source of carbohydrates. These carbohydrates are distributed between different sinks: the reserves (i.e. sugar in the stems, starch in the grains), the constitution of the structures (i.e. stem, roots, leaves) and their respiration. Carbohydrates are distributed to each sink in proportion to its strength. In SAMARA, growth can be source or sink limited, governed by an index of internal competition. Under structural sink limitation, excess assimilates are stored as reserves and can also cause feedback inhibition of photosynthesis. The rate of photosynthesis depends, among other factors, on the availability of water to the roots.

In our study, system water inflow is limited to rainfall, while water outflow is simulated as the sum of runoff, soil surface evaporation, plant transpiration and deep drainage, which occurs when the water holding capacity of the soil is exceeded. As in its simpler precursor model SARRA-H, frequently used for climate impact studies in the Sahel (Baron et al., 2005; Sultan et al., 2005), rooting depth is limited by a dynamic soil wetting front. For monsoonal sorghum in West Africa, flowering should ideally take place a month before the rain stops, to minimize grain diseases and losses to birds and other pests, and furthermore, to receive maximal solar radiation during grain filling (Kouressy et al., 2008a). In this case, grain filling takes place mainly by exploiting the soil's residual water reserves. Flowering date is simulated using thermal-time budgets and photoperiod sensitivity for varieties that are sensitive to it. The grains receive the product of current photosynthesis and, if insufficient, sugar reserves accumulated in the stems. During this period, there is a trade-off between terminal leaf senescence (that relocates C and N to grains) and stay-green (which sustains photosynthesis). In SAMARA, leaf senescence is driven by C competition among organs, attenuated by a genotypic parameter. Depending on the available reserves and the number of grains ready to receive them, the filling of the grains will be more or less complete.

SAMARA employs a vast array of 83 crop parameters (see https://umr-agap.cirad.fr/en/research/scientific-teams/samara-model for a complete description), categorized into several functional groups including seed, leaf, internode and panicle properties, phenology and photoperiodism, root growth, tillering, light extinction and conversion, water relations, maintain respiration and thermal stress. By incorporating such a diverse range of parameters, the model can replicate the various stages and processes involved in sorghum growth.

2.2.2. Description of MarkSim stochastic climate generator

The MarkSim stochastic climate generator was developed in the 1990s to simulate weather series from known sources of weather data from around the world (P. G. Jones and Thornton, 1993; Peter G. Jones and Thornton, 2013). The basic MarkSim algorithm is a daily precipitation simulator. Precipitation is modelled using a two-step process: the

Table 2
Characteristics of the ten varieties used in the multi-location trial. Photoperiod sensitivity is categorized as low ($0 < K \leq 0.3$), moderate ($0.3 < K \leq 0.6$), and high ($0.6 < K \leq 0.9$), with photoperiod sensitivity K defined as the (reduction in cycle length)/(delay in sowing date) as per Sawadogo et al. (2022). Cycle length is measured from sowing to the physiological maturity of the grain, indicated by the grain hilum turning black. Isohyet values indicate the range of annual rainfall (measured in millimeters) to which each sorghum genotype is best adapted. Plant height is measured under optimal growing conditions from the ground to the top of the panicle. Potential yield refers to the yield obtained under optimal growing conditions.

Genotype	Type	photoperiod sensitivity	Cycle length	Isohyet	goal	plant height	Potential yield	other	origin
Fadda	Guinea (Hybrid)	moderate	128 days	700-1000 mm	Grain-biomass	2-3m	4.5t/ha	Tolerant of: mould, anthracnose	Mali, breeding IER/ ICRISAT, pedigree
Nielni	Caudatum (Hybrid)	low	115 days	700-800 mm	Grain	3m	4t/ha	Tolerant of: mould, anthracnose	Mali, breeding IER/ ICRISAT
IS15401	Guinea	high	115 days	900-1200 mm	Biomass	4-4.5m	2t/ha	Resistant to: mould, striga and midge	Cameroun, breeding IER/ICRISAT
Pablo	Guinea (Hybrid)	moderate	125 days	700-1000 mm	Biomass	4m	4t/ha	Tolerant of: mould, anthracnose	Mali, breeding IER/ ICRISAT
CSM6	Guinea	Low	90 days	600-1000 mm	Grain	4m	2tha	Tolerant of: anthracnose and insects	Mali, traditional variety
SK5912	Caudatum	High	170 days	700-900 mm	Biomass	2m	2.5-3.5t/ha	Tolerant of: mould, anthracnose	Nigeria
Grinkan	Caudatum	No	90 days	500-800 mm	Grain-biomass	1.2m	4t/ha	Resistant to: midge, insects	Mali, breeding ICRISAT
Soumba	Caudatum	Low	115 days	600-1000 mm	Grain-biomass	2.5m	2.5t/ha	Tolerant of: anthracnose, insect and striga	Mali
621B	Caudatum	No	105 days	600-900 mm	Grain	1.75m	2.5-3t/ha	Mould resistant	Senegal, breeding ISRA
F2-20	Caudatum	Low	110 days	600-900 mm	Grain	2.1m	3-5.3t/ha	Resistant to: mould and striga	Senegal, breeding ISRA

first determines whether a given day is rainy or not, and the second determines the amount of rain.

The first step is a third-order Markov chain: the probability that day j of month i will be rainy depends on the rainfall of the previous three days. The probability of rain is defined by a probit model:

$$P(W / D_1 D_2 D_3) = \phi^{-1}\left(b_i + a_{j-1} d_1 + a_{j-2} d_2 + a_{j-3} d_3\right)$$

where the probit function ϕ^{-1} is the inverse of the distribution function of the standard normal distribution, b_i is the monthly baseline probit of a wet day after three consecutive dry days, a_j is an indicator equal to 1 if it rained on day j and 0 otherwise, and d_k is a persistence parameter.

The probability of a wet day is thus specified by 15 parameters, including the baseline probit, b_i, for each month and three persistence parameters, d_1, d_2 and d_3, which do not vary from one month to another.

In the second step, rainfall is generated using the truncated gamma distribution, restricted to values greater than or equal to 0.1 mm, to determine the daily amount of rainfall on days where rainfall has been decided in the first step. The maximum likelihood method is used to estimate the mean and parameters of this distribution for each month, giving a total of 24 additional parameters. To generate rainfall records, the monthly base probabilities (probability of rain after 3 consecutive dry days) are interpolated to daily probabilities using the 12-point Fourier transform (Peter G. Jones and Thornton, 2000). Similarly, the monthly mean and shape parameters of the gamma distribution of precipitation amounts are interpolated, also using the 12-point Fourier transform, and downscaled to daily values. MarkSim allows the simulation of daily maximum and minimum temperatures and daily solar radiation, based on the methods used in (Richardson, 1981). Maximum temperature, minimum temperature and solar radiation are considered as continuous multivariate stochastic processes with daily means and standard deviations conditioned on the wet or dry state of the day.

Markov chain parameters were estimated at 9,200 weather stations where daily data were available, and the results were classified into 720 climate groups(Peter G. Jones and Thornton, 2013). For each group, a regression of these parameters was estimated on monthly averages of total precipitation and daily minimum and maximum temperatures. These monthly means are available for almost all land areas. They were interpolated on a grid of 30 arc-second steps, corresponding to a ground distance of 1.1 km at the equator and decreasing with the cosine of latitude - such data are available in the WorldClim database.

The authors of the generator (Peter G. Jones and Thornton, 2013) found that this 3-day dependence model with seasonal coefficients was not sufficient to reproduce the variance in total annual precipitation. To recover this variability, they modulated the transition probabilities of the Markov chain from one simulated year to the next, following a distribution whose variance is the variance of the estimate of these parameters.

2.2.3. Coupling the SAMARA and MARKSIM models

For our ideotype search, the SAMARA model was set up with a sowing strategy and a vector of soil and crop management inputs set to fixed value t, with a v vector of parameters that vary according to variety, and coupled with the Marksim weather generator to feed it with weather data simulated over a large time period for any location u. For each combination (u,v), we have been able to estimate an expected yield by the average of 99 simulated yields at location u with parameter set v (Fig. 2).

In this study, a simplified case was considered in which the soil characteristics were set to identical non-limiting values. In the same way, the technical itineraries were assumed to be identical regardless of the location, except for the sowing date, which was determined for each location and year according to a fixed sowing strategy. Following the recommendations of (Balme et al., 2005), a cumulative rainfall of 20 mm over three days is required before sowing sorghum. We increased this threshold to 30 mm to match farmers' perceptions of sufficient rainfall. For a given set of fixed soil and crop management inputs, the model outputs then depend only on the sorghum variety and the meteorological environment (i.e. weather/climatic variability).

For each location u and variety v, we assumed mean simulated yields of the form

$$\widetilde{y}(u,v) = y(u,v) + \varepsilon \tag{1}$$

where ε represents observation noise, assumed to be the realisation of a random variable, and $y(u,v)$ represents the yield expectation for variety v in environment u. In the following, we also assumed that the observation noise is normally distributed, centred, and independent from one run to another, $\varepsilon \sim \mathcal{N}(0,\tau^2)$.

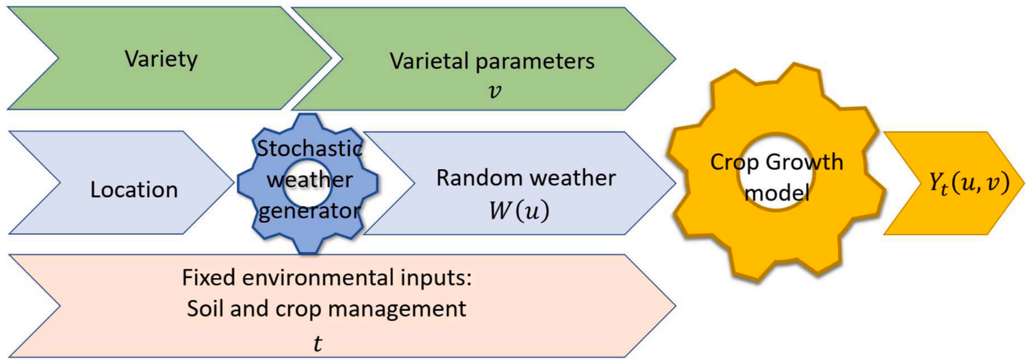

Fig. 2. Coupling of the stochastic weather simulator Marksim and the crop growth model Samara to simulate at any location u an expected yield $Y_t(u, v)$ of a variety v under fixed environmental inputs t.

2.3. Searching for a map of optimal varieties

2.3.1. Choice of varietal parameters for optimization

As many of crop models, the SAMARA model uses a large number of varietal parameters. Most of these parameters are not good candidates for ideotypes search using an optimisation algorithm, either because they are not variable within the species or because the output of interest (i.e., grain yield) is not sensitive to them in the target environment. Another rejection criterion is the parameter whose output is known to be a monotonic function, which is the case for photosynthetic efficiency. Increasing it always increases the output, so there is no need to use an optimisation algorithm. Among other possibilities, for our study we chose the following parameters to characterize our ideotypes:

- Maximum root length (RootFrontMax): Investing in long roots consumes energy to the detriment of the grain. On the other hand, long roots allow the crop to survive days without rain (dry spells) and can therefore save production. It is therefore an ideal parameter to find a compromise.
- Length of vegetative cycle (sdjbvp): Lengthening the vegetative cycle allows more leaf production and accumulates more biomass, leading to better grain production. However, a long vegetative cycle could be detrimental to grain production if the rainy season is shorter than expected, leading to drought stress during the grain filling period. We focused here on photoperiod insensitive varieties.
- Leaf death sensitivity coefficient (leaf.death): This parameter regulates leaf death throughout the cycle based on inter-organ competition, but impacts mostly during the grain filling phase when panicles are the dominant sink. A value of 0 simulates a "stay green" sorghum genotype that keeps its leaves green at maturity. This genotype will be able to photosynthesise and thus make better use of the water reserves in the soil. However, green leaves transpire and bind carbon reserves. There is therefore a trade-off between the consumption of carbon reserves and the expectation that the water reserves will be converted into carbohydrates.
- Reserve capacity in the stems (stem reserve): At the end of the rainy season, under unfavourable conditions, sorghum can mobilise sugar reserves accumulated in the stem (more precisely in the internodes). However, these reserves can prove useless when conditions are favourable.

Prior to optimization, on a set of 5 localities evenly distributed along a north-south gradient, we verified that a change in these parameters modified the distribution of simulated yields at least in some locations (results not shown). In other words, yields were sensitive to these parameters, but not in every location.

The other parameters were set to the values estimated for the non-photoperiodic variety 621B described in Table 2. Some parameters, such as those related to potential panicle size, were set at the upper limit of the variation interval specified in the Samara reference manual (available at https://umr-agap.cirad.fr/en/research/scientific-teams/samara-model), so as not to limit potential yield when exploring the parameter space during optimisation.

2.3.2. Optimization method to determine a sorghum ideotype map

Our objective was to find a map of optimal parameter values that would allow each location to maximise an expectation or quantile of production. To reduce the computational time, which can be long (if we compute the optimum of each location), we used a metamodel-based optimization algorithm (see Victor Soares do Amaral et al., 2022 for a review). At its core, this type of algorithm leverages the concept of metamodels, which are surrogate models created to approximate the behaviour of the actual objective function, in our case the expected yield. The process starts by sampling a limited number of points from the parameters search space, evaluating the objective function at these points, and then fitting a metamodel. Once the metamodel is constructed, it replaces the computationally expensive objective function, drastically reducing the number of actual evaluations required. The algorithm then iteratively refines the metamodel and explores the parameters search space using intelligent strategies like genetic algorithms to identify promising areas to evaluate the true objective function. By effectively balancing exploration and exploitation, the metamodel-based optimization algorithm converges towards optimal solutions with significantly fewer evaluations, making it a powerful and efficient tool for solving real-world optimization problems in various domains.

In this study we used a method based on a kriging metamodel and a specific optimization criterion called Profile Expected Quantile Improvement (PEQI) developed by Sambakhé et al. (2019) for the conditional optimisation of a noisy function. This optimization algorithm consisted of the following steps:

1. Design an initial plan: an initial latin hypersquare experimental design of 100 points $D = \{(u_1, v_1), \dots, (u_{100}, v_{100})\}$ was created using the optimumLHS function from the lhs R package.
2. Construct the metamodel: the crop model was evaluated at each point in the experimental design D, and the variance of the noise ε was estimated (see Eq. (1)). This information was used to construct a kriging metamodel.
3. Search for optimal sampling points: the algorithm aimed to find a new point (u^*, v^*) that maximized the Profile Expected Quantile Improvement (PEQI) criterion. This involved searching for the optimal combination of location u and varietal parameters v by considering a random grid of environments G and all possible varieties V.
4. Calculate the function output: the expected yield for the selected point (u^*, v^*) was calculated by averaging 99 simulated yields for that particular location with the given parameter set.

5. Update the experimental design and metamodel: the new point (u^*, v^*) was added to the experimental design, and the kriging metamodel was updated using the additional data point.

Steps 3, 4, and 5 were repeated iteratively for a predetermined number of 200 iterations. The final metamodel was then used to determine the optimal v (ideotype) at each node u of a regular grid in order to produce an ideotype map.

The optimisation algorithm is implemented in R (Version 3.4.0, R Core Team, 2017) using DiceOptim (Picheny et al, 2021), DiceKriging (Roustant et al., 2012), rgenoud and lhs (Carnell, 2022) packages. The algorithm produces a grid of geographic coordinates and a set of parameters corresponding to the ideotype for each point on the grid. ArcGIS Pro software (Esri, 2020) was used to produce the Ideotype maps. This software has several interpolation methods for predicting cell values in a raster from a limited number of sample points.

2.4. Comparison of sorghum ideotypes with a reference sorghum variety

In order to assess the improvement in expected yield resulting from the conditional optimization, we generated maps to compare the expected yields of both the ideotype and real-world plant varieties within the context of the Sahel. As real-world counterpart we chose the 621B variety previously used as starting point of our optimisation process. We generated the yield maps using ArcGIS Pro.

3. Results

3.1. Sorghum ideotype map for Sahelian region

We generated a sorghum ideotype map for part of the Sahel countries, including Senegal and parts of Mauritania, Guinea-Bissau, Guinea, Mali, Niger, and Burkina Faso (Fig. 3). The area chosen is much wider in latitude than the sorghum's range.

For cycle length, the results of the optimisation were consistent with the breeders' varietal profiling. At low latitudes, where rainfall is regular, long duration sorghum ideotypes are better adapted. On the other hand, at high latitudes where rainfall is scarce, short duration sorghum ideotypes give better yields (see Fig. 3-A).

As far as root length is concerned, sorghum ideotypes with short roots are only adapted in the south-western part of the map (Guinea-Bissau and Guinea). The transition is very rapid northwards, where the long-root trait is best throughout Senegal. At about 10 degrees west longitude (i.e. Mali), the optimum variety has medium length roots (about 1000 mm; see Fig. 3-B). The interpretation of the results is less obvious for the optimal values of the parameters leaf.death and stem.reserve (see Fig. 3-C, Fig. 3-D). In some environments, a sensitivity analysis (not shown) reveals that these parameters have less influence on yield compared to crop cycle and root length. However, in the western part of the map and between 13°N and 16°N there is a trend for yield improvement with sorghum ideotypes with high stem reserves.

In Mauritania and north part of West Mali, around 16° N and 5, 8 and 9° W, the optimum values of the leaf mortality and stem reserve optimum coefficients are unstable, as they vary much within a small area. Again, these two parameters have a low influence on expected yield in these areas (sensitivity analysis not shown).

3.2. Comparison of ideotypes' expected yield with that of a reference sorghum variety

Comparison of the average yields simulated using Samara for the ideotypes and for the reference variety 621B (Fig. 4) shows a great improvement in the expected yield, which varies according to latitude. Average yield improvement is around 800 kg/ha near Dakar (14.44°N 17.27°W), 1200 kg near Ouahigouya in Burkina Faso (13.35°N 2.25°W) and 1600 kg at the Mali-Guinea-Senegal triple border (12.41°N 11.38°W), but is very low at the southern border of Mauritania (15.50°N 8.12°W), which remains unsuitable for Sorghum cultivation. These

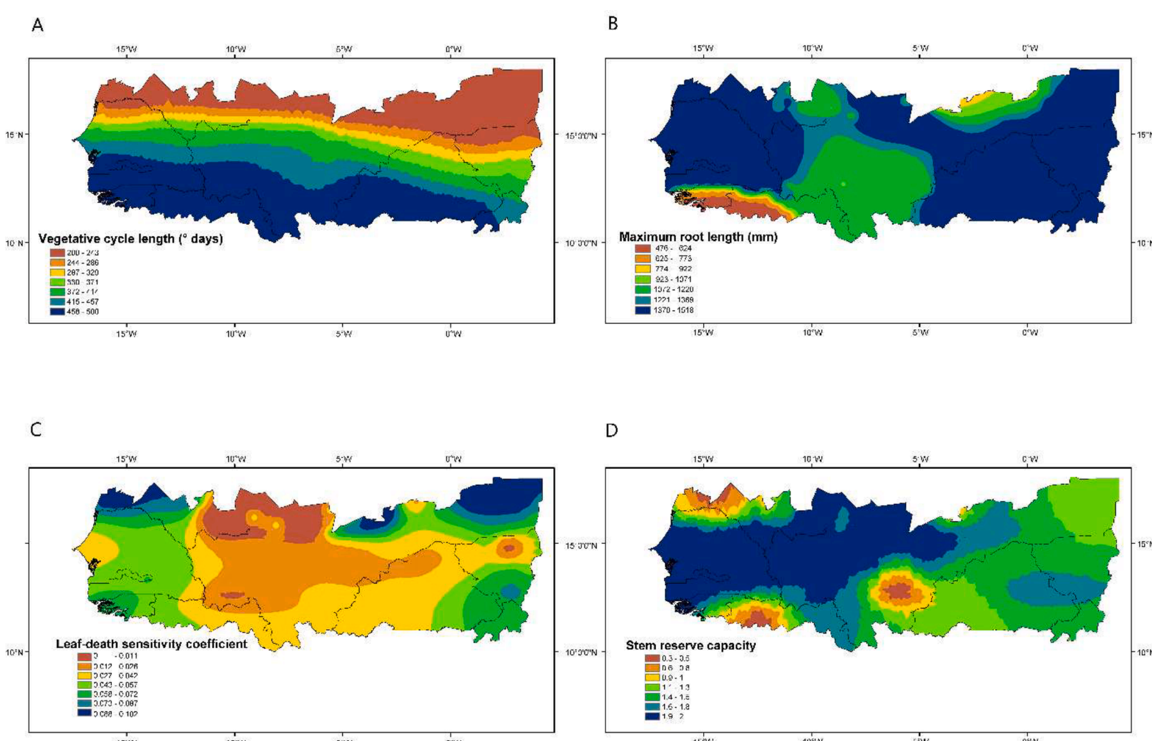

Fig. 3. Sorghum ideotype varies continuously with climate gradient as described by the maps of optimal values of four model varietal parameters. (A) Map of optimal values of cycle length, (B) Map of optimal values of root length, (C) Map of optimal values of leaf mortality, (D) Map of optimal values of stem reserve capacity. The optimal parameters are those that maximise expected yield.

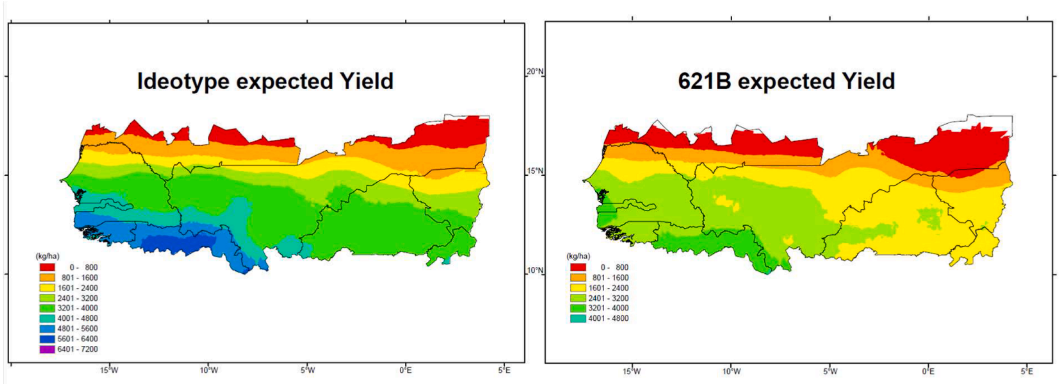

Fig. 4. Comparison of ideotypes' expected yield with that of the 621B reference sorghum variety.

yields are potential yields with no limiting factors other than radiation and water supply.

4. Discussion

The overall objective of this study was to find sorghum ideotypes adapted to the present and future Sahelian climate. In this study, we chose to find these ideotypes among non-photoperiodic varieties, by optimizing four parameters: length of the vegetative phase, maximum root length, potential stem reserve and leaf senescence. The model's optimization outputs for cycle length were consistent with its matching with useful rainy season. For root length, high yields were associated with root lengths of about 1 m and above, except in the south-western part of the map (in Guinea-Bissau) where short roots were found to be favourable. This result seems to be consistent with the results of (Chen et al., 2020), who showed that under controlled progressive drought conditions, deep roots of sorghum play an important role in yield production, making it a trait of interest for varietal improvement in a drought context. Assefa et al. (2010) also identified root depth and density as important drought adaptation mechanisms in sorghum.

The result of the model optimisation also showed that the sensitivity of leaf senescence to the water stress parameter had a positive effect on yield only in the most stressful environments. This is in accordance with the findings of Kouressy et al (2008) who state that as yield is sink limited, stay-green introduction would not theoretically improve yield, except when expressed before flowering. Without working specifically on leaf sensitivity to water stress, but more globally on the stay green trait, (Borrell et al., 2014) showed that this is an important adaptive trait to end-of-cycle water stress in sorghum. In fact, they showed that the stay green loci are also associated with a reduction in canopy size at flowering and hence water use before anthesis, which in the case of post-flowering water stress increases water availability during grain filling and hence grain yield. However, there have been no multi-site trials to show whether the beneficial effects of the stay green trait are limited to the most stressful environments.

Similarly, without specifically addressing stem reserve capacity, (Blum et al., 1997) showed in an experimental study that among two isogenic varieties with different stem sizes, the shorter one with less stem reserve was less resistant to water stress than the taller one, with less desiccation-induced mass transfer from stems to grains.

The ideotype map presented here was produced by optimising average yield according to four parameters of the SAMARA model. The parameters were chosen by expert opinion for a proof of concept because they were known or suspected to be subject to trade-offs. However, other options can be considered for the choice of test parameters as well as the variable to be optimised. For example, photoperiodism parameters and yield free from fungal disease losses can be chosen for optimisation.

Our methodology based on a kriging metamodel and the PEQI criterion allows us to produce, in a reasonable computational time, a map of ideotypes with continuous variation of the parameters, which is better able to follow the continuous variation of the climatic stress.

For this proof of concept, we chose to set the soil parameters at a non-limiting value. This facilitated the presentation and interpretation of optimizations in relation to the climatic gradient alone. To also consider optimization in relation to soil parameters such as useful reserve, if their variations remain slow in space, we can apply exactly the same approach. If there are discontinuous variations between soil categories, separate optimizations for each category should be used.

Depending on the farmer's objectives and risk aversion, a quantile of yield will give a better compromise between its expectation and its variability. Among the plant traits reflected by the model parameters, some are easier to improve genetically than others because they are cheaper to measure or more heritable. A possible improvement to the optimisation algorithm would be to consider the cost of trait improvement, either in the form of a constraint or by including the cost in a multi-criteria optimisation.

Our approach also offers an effective means of addressing the impacts of climate change on sorghum cultivation, as varietal parameters may also be optimized using climate change scenarios. As extreme temperatures, erratic rainfall and other climate-related stressors become increasingly pronounced, our methodology thus provides a proactive framework for improving the climatic resilience of sorghum crops. Using the complex mechanisms encoded in the crop model, this innovative approach not only helps to find the right compromises between varietal traits in the context of changing climatic variables, but also provides agricultural stakeholders with actionable information to optimize sorghum cropping strategies, ensuring food security in the face of a dynamic and uncertain climatic future.

Lastly, it should be noted that the crop parameters we considered here did not cover the full spectrum of adaptive traits. For example, short plant stature and tillering may increase harvest index in favourable environments but less so under drought; likewise, high genotypic panicle sink potential may increase the yield ceiling under favourable conditions but incur unnecessary carbon costs under drought. The present study thus presented a practical approach to quantitative ideotype definition without claiming completeness.

Funding source

We are very grateful to the West Africa Agricultural Productivity Program (WAAPP) that funded this research as part of a Ph.D. thesis grant. Grant number: P09408.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used DeepL write (https://www.deepl.com/write) in order to improve the quality of

English-language writing. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

CRediT authorship contribution statement

Diariétou Sambakhé: Writing – original draft, Methodology, Formal analysis. **Eric Gozé:** Writing – original draft, Supervision, Methodology. **Jean-Noël Bacro:** Writing – review & editing, Supervision, Methodology. **Michael Dingkuhn:** Validation. **Myriam Adam:** Validation. **Malick Ndiaye:** Data curation. **Bertrand Muller:** Validation. **Lauriane Rouan:** Writing – original draft, Supervision, Methodology.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Diarietou Sambakhe reports financial support was provided by West Africa Agricultural Productivity Program. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgements

The authors thank Christophe Pradal for his careful reading of this paper and his constructive advice.

References

Affholder, F., 1997. Empirically modelling the interaction between intensification and climatic risk in semiarid regions. Field Crops Res. 52 (1–2), 79–93. https://doi.org/10.1016/S0378-4290(96)03453-3.

Assefa, Y., Staggenborg, S.A., Prasad, V.P.V., 2010. Grain sorghum water requirement and responses to drought stress: a review. Crop Manag. 9 (1), 1–11. https://doi.org/10.1094/CM-2010-1109-01-RV.

Asseng, S., van Keulen, H., Stol, W., 2000. Performance and application of the APSIM Nwheat model in the Netherlands. Eur. J. Agron. 12 (1), 37–54. https://doi.org/10.1016/S1161-0301(99)00044-1.

Balme, M., Galle, S., Lebel, T., 2005. Démarrage de la saison des pluies au Sahel : variabilité aux échelles hydrologique et agronomique, analysée à partir des données EPSAT-Niger. Sci. Change. Planétaires/Sécheresse 16 (1), 15–22. Retrieved from. http://www.jle.com/download/sec-265216-demarrage_de_la_saison_des_pluies_au_sahel_variabilite_aux_echelles_hydrologique_et_agronomique_analysee_a_partir_des_donnee-cirad-WmshPX8AAQEAAGFyvBsAAAAN-u.pdf.

Baret, F., 1980. Caractérisation Hydrodynamique d'un Sol de la Région de Nioro. Institut Sénégalais de Recherches Agricoles (ISRA), Dakar, Sénégal.

Baron, C, Sultan, B, Balme, M, Sarr, B, Traore, S, Lebel, T, Janicot, S, Dingkuhn, M., 2005. From GCM grid cell to agricultural plot: Scale issues affecting modelling of climate impact. Phil. Trans. R. Soc. B 360, 2095–2108.

Blum, A., Golan, G., Mayer, J., Sinmena, B., 1997. The effect of dwarfing genes on sorghum grain filling from remobilized stem reserves, under stress. Field. Crops. Res. 52 (1–2), 43–54. https://doi.org/10.1016/S0378-4290(96)03462-4.

Bonte, M.H.A., van den Boogaard, A.H., Huétink, J., 2008. An optimisation strategy for industrial metal forming processes. Struct. Multidiscipl. Optimiz. 35 (6), 571–586. https://doi.org/10.1007/s00158-007-0206-3.

Borrell, A.K., Mullet, J.E., George-Jaeggli, B., Van Oosterom, E.J., Hammer, G.L., Klein, P.E., Jordan, D.R., 2014. Drought adaptation of stay-green sorghum is associated with canopy development, leaf anatomy, root growth, and water uptake. J. Exp. Bot. 65 (21), 6251–6263. https://doi.org/10.1093/JXB/ERU232.

Carnell, R. (2022). lhs: Latin Hypercube Samples. R package version 1.1.6. https://CRAN.R-project.org/package=lhs.

Chen, X., Wu, Q., Gao, Y., Zhang, J., Wang, Y., Zhang, R., Huang, R., 2020. The role of deep roots in sorghum yield production under drought conditions. Agronomy 10 (4), 611. https://doi.org/10.3390/AGRONOMY10040611. *2020, Vol. 10, Page 611*.

Dingkuhn, M., Soulié, J.C., Lafarge, T., 2011. Samara V2: A cereal crop model to study G x E x M interaction and phenotypic plasticity, and explore ideotypes. In: AgMIP Rice International Workshop, pp. 28–30.

Donald, C. M. (1968). The Design of a Wheat Ideotype. *Finlay, KW and Shepherd*.

Esri, 2020. ArcMap (Version 10.8.1). Environmental Systems Research Institute. https://desktop.arcgis.com/.

Fensholt, R., Rasmussen, K., Kaspersen, P., Huber, S., Horion, S., Swinnen, E., 2013. Assessing land degradation/recovery in the African Sahel from long-term earth observation based primary productivity and precipitation relationships. Remote Sens. (Basel) 5 (2), 664–686. https://doi.org/10.3390/RS5020664. *2013, Vol. 5, Pages 664-686*.

Janusevskis, J., Le Riche, R., 2013. Simultaneous kriging-based estimation and optimization of mean response. J. Glob. Optimiz. 55 (2), 313–336. https://doi.org/10.1007/s10898-011-9836-5.

Jones, P.G., Thornton, P.K., 1993. A rainfall generator for agricultural applications in the tropics. Agric. For. Meteorol. 63 (1–2), 1–19. https://doi.org/10.1016/0168-1923(93)90019-E.

Jones, P.G., Thornton, P.K., 2000. MarkSim: Software to generate daily weather data for Latin America and Africa. Agron. J. 92 (3), 445–453. https://doi.org/10.2134/agronj2000.923445x.

Jones, P.G., Thornton, P.K., 2013. Generating downscaled weather data from a suite of climate models for agricultural modelling applications. Agric. Syst. 114, 1–5. https://doi.org/10.1016/j.agsy.2012.08.002.

Kouressy, M, Dingkuhn, M, Vaksmann, M, Heinemann, A.B., 2008a. Adaptation to diverse semi-arid environments of sorghum genotypes having different plant type and sensitivity to photoperiod. Agric. For. Meteorol. 148, 357–371.

Kouressy, M., Dingkuhn, M., Vaksmann, M., Clément-Vidal, A., Chantereau, J., 2008b. Potential contribution of dwarf and leaf longevity traits to yield improvement in photoperiod sensitive sorghum. Eur. J. Agron. 28 (3), 195–209. https://doi.org/10.1016/J.EJA.2007.07.008.

Kumar, U., Laza, M.R., Soulie, J.-C., Pasco, R., Mendez, K.V.S., Dingkuhn, M., 2016. Phenotypic plasticity of irrigated rice: sequential path analysis of yield components and simulation with the SAMARA crop model. Field. Crops. Res. 193, 164–177.

Kumar, U., Laza, M.R., Soulie, J.-C., Pasco, R., Mendez, K.V.S., Dingkuhn, M., 2017. Analysis and simulation of phenotypic plasticity for traits contributing to yield potential in twelve rice genotypes. Field. Crops. Res. 202, 94–107.

Labeyrie, V., Thomas, M., Muthamia, Z.K., Leclerc, C., 2016. Seed exchange networks, ethnicity, and sorghum diversity. Proc. Ntl. Acad. Sci. 113 (1), 98–103. https://doi.org/10.1073/pnas.1513238112.

Landsberg, J.J., de Wit, C.T., 1980. Simulation of assimilation, respiration and transpiration of crops. J. Ecol. 68 (1), 333. https://doi.org/10.2307/2259277.

Muller, B., Martre, P., 2019. Plant and crop simulation models: powerful tools to link physiology, genetics, and phenomics. J. Exp. Bot. 70 (9), 2339–2344. https://doi.org/10.1093/jxb/erz175.

Ndiaye, M., Muller, B., Ganyo, K.K., Guissé, A., Cissé, N., Adam, M., 2021. Phenotypic plasticity of plant traits contributing to grain and biomass yield of dual-purpose sorghum. Planta 253 (4), 1–14. https://doi.org/10.1007/S00425-021-03599-Z/METRICS.

OCDE, F.A.O., 2014. Perspectives agricoles de l'OCDE et de la FAO 2014-2023. Organisation de Coopération et de Développement Economiques, Paris.

Picheny, V., Ginsbourger, D. & Roustant, O. (2021). DiceOptim: Kriging-based optimization for computer experiments. R package version 2.1.1. https://CRAN.R-project.org/package=DiceOptim.

Queipo, N.V., Haftka, R.T., Shyy, W., Goel, T., Vaidyanathan, R., Kevin Tucker, P., 2005. Surrogate-based analysis and optimization. Prog. Aerosp. Sci. 41 (1), 1–28. https://doi.org/10.1016/j.paerosci.2005.02.001.

R Core Team, 2017. R: A language and environment for statistical computing (Version 3.4.0). R Foundation for Statistical Computing. https://www.R-project.org/.

Richardson, C.W., 1981. Stochastic simulation of daily precipitation, temperature, and solar radiation. Water. Resour. Res. 17 (1), 182–190. https://doi.org/10.1029/WR017i001p00182.

Roustant, O., Ginsbourger, D., Deville, Y., 2012. DiceKriging, DiceOptim: Two R packages for the analysis of computer experiments by kriging-based metamodeling and optimization. J. Stat. Softw. 51 (1), 1–55. URL. https://www.jstatsoft.org/v51/i01/.

Sambakhé, D., Rouan, L., Bacro, J.-N., Gozé, E., 2019. Conditional optimization of a noisy function using a kriging metamodel. J. Glob. Optimiz. 73 (3) https://doi.org/10.1007/s10898-018-0716-0.

Sawadogo, N., Tiendrebéogo, J., Naoura, G., Béré, T.L.K, Tondé, W.H, Sawadogo, B, Bationa-Kando, P, 2022. Photoperiod sensitivity and variability of agromorphological traits and brix content of sweet sorghum cultivated in Burkina Faso under two sowing dates. Adv. Agric. https://doi.org/10.1155/2022/9504150.

Sultan, B., Baron, C., Dingkuhn, M., Sarr, B., Janicot, S., 2005. Agricultural impacts of large-scale variability of the West African monsoon. Agric. For. Meteorol. 128, 93–110.

Victor Soares do Amaral, João, José Arnaldo Barra Montevechi, Rafael de Carvalho Miranda, Wilson Trigueiro de Sousa Junior, 2022. Metamodel-based simulation optimization: A systematic literature review. Simul. Model. Pract. Theory 114, 102403. https://doi.org/10.1016/j.simpat.2021.102403.

Yule, G.U., 1902. Mendel'S laws and their probable relations to intra-racial heredity. New Phytol. 1 (10), 222–238. https://doi.org/10.1111/j.1469-8137.1902.tb07336.x.