Evaluating a Cassava Crop Growth Model by Optimizing Genotypic-Specific Parameters Using Multi-environment Trial Breeding Data

Pamelas M. Okoma[1], Siraj S. Kayondo[2], Ismail Y. Rabbi[2], Patricia L. Moreno-Cadena[3,4], Gerrit Hoogenboom[5, 6] and Jean-Luc Jannink[1,7*]

[1] Plant Breeding and Genetics Section, School of Integrative Plant Science, College of Agriculture and Life Sciences, Cornell University, NY, USA

[2] International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria

[3] Plant and Soil Sciences, College of Agriculture Food and Environment, University of Kentucky, KY, USA

[4] Alliance Bioversity and CIAT, Nairobi, Kenya

[5] Global Food Systems Institute, University of Florida, Gainesville, FL, USA

[6] Department of Agricultural and Biological Engineering, University of Florida, Gainesville, FL, USA

[7] Robert W. Holley Center for Agriculture and Health, USDA-ARS, Ithaca, NY, United States

*** Correspondence:**
Corresponding Author
jeanluc.jannink@usda.gov

Keywords: Cassava, Crop Growth Model, Nigeria, Calibration, General Likelihood Uncertainty Estimation

Abstract

Cassava (*Manihot esculenta* Crantz) is a critical food security crop for sub-Saharan Africa. Efforts to improve cassava through breeding have expanded over the past decade. At the same time, crop growth models (CGM) are becoming common place in breeding efforts to expand the inference of evaluations of breeding germplasm to environments that have not been tested and to prepare for breeding for adaptation to future climates. We parameterized a CGM, the CROPGRO-MANIHOT-Cassava model in the DSSAT family of models, using data on 67 clones from the International Institute of Tropical Agriculture cassava breeding program evaluated from 2017 to 2020 and over eight locations in Nigeria using trial and error parameter adjustments and the General Likelihood Uncertainty Estimation method. Our objectives were to assess the feasibility of this large-scale calibration in the context of a cassava breeding program and to identify systematic biases of the model. For each cultivar we calculated the Pearson

correlation between model prediction and observation across the environments, as well as root mean squared error and d statistics. As a result of calibration, the correlation coefficient increased from –0.03 to +0.08, the RMSE dropped from 21 t ha^{-1} to 5 t ha^{-1} while d increased from 0.23 to 0.44. We found that the model underestimated root yield in dry environments (low precipitation and high temperature) and overestimated root yield in wet environments (high precipitation and low temperature). Our experience suggests both that CGM calibration could become a routine component of the cassava breeding data analysis cycle and that there are opportunities for model improvement.

Introduction

Crop Growth Models (CGMs) now have a long history in the study of agronomic management practices (Jones et al. 2003; Holzworth et al. 2018; Jones et al. 2017; Tsuji et al. 1998). CGMs simulate plant phenology and growth by integrating ecophysiological functions with weather data, usually on a daily time step, tracking photosynthesis and partitioning as well as resource extraction from the environment. Once calibrated, a CGM can be used to study such questions as planting dates or fertilizer application (Banterng et al. 2009), conservation tillage (Corbeels et al. 2016), crop rotations (Mohanty et al. 2012), or even intercropping (Berghuijs et al. 2021). An advantage of studying such issues with a model is that the system response can be predicted over a long timeframe or for many different location conditions representing a breadth of environments that would be impossible to evaluate empirically. Even for well-calibrated models, results cannot be taken at face value but should suggest the important factors to explore with field experiments.

Expanding the use of CGMs from agronomic questions to plant breeding has been more recent (Baenziger et al. 2004; Technow et al. 2015; Messina et al. 2018; White et al. 1996). Plant breeders base selection decisions on predictions of future performance of breeding lines, whether those predictions come from prior observations of the lines themselves or of close relatives whose genetic relatedness can be inferred from DNA variants. In either case, the predictions cannot be extrapolated to environments or management practices that have not been observed. Thus, a motivation to parameterize CGMs for breeding lines would be to enable predictions in environments where they have not been tested. Applications of such predictions could be estimating the stability of a line's performance by modeling outcomes over many years or locations (for example, aiding such efforts as Lozada and Carter 2020; Olivoto et al. 2021), identifying specific environments that may be most promising for specific breeding lines (as Chai, et al. 2022 have suggested occurs in modern breeding, for example, Phoncharoen, et al. 2021) or for identifying optimal locations for a multi-location trial evaluation network (Putto et al. 2008). These models may therefore contribute to the development of efficient breeding strategies through decision support in breeding programs. A number of these applications are reviewed in the context of breeding to adapt to climate change in (Ramirez-Villegas et al. 2020).

As for many developments in breeding, attempts to integrate CGMs into breeding have been spurred on by the plummeting cost of DNA markers that has made it possible to associate traits with alleles or haplotypes at specific genomic segments. A challenge to the use of CGMs in breeding is the effort and cost to estimate many CGM parameters for individual breeding lines, as exemplified by early efforts (e.g., White and Hoogenboom 1996; Hoogenboom and White 2003). In particular, CGMs require many parameters that are not routinely measured in breeding programs so that they come at an extra cost to the program, which is prohibitive. A work around

to this problem is to seek to estimate parameter values from data that *are* routinely collected in breeding programs. Estimation procedures essentially fit parameter values so that CGM predictions match as closely as possible to observed breeding data. One such method is General Likelihood Uncertainty Estimation (GLUE) that has been incorporated in the Decision Support System for Agrotechnology Transfer (DSSAT) (He et al. 2010; Hoogenboom et al. 2021; Ferreira et al., 2024). To extend this methodology to cassava (Manihot esculenta Crantz), the initial step involves the development of a robust growth model.

The worldwide importance of cassava is undeniable, especially in sub-Saharan Africa where cassava roots are a valuable source of calories (Adebayo 2023; Chiaka et al. 2022). Numerous efforts to improve cassava breeding to deliver greater genetic gain are underway. Such efforts span a wide range of technologies from on farm testing (Nanyonjo et al. 2024), to genomic selection (Wolfe et al. 2017; Ceballos et al. 2021), to developing cassava hybrids (Zhang et al. 2024). Integrating crop growth models into these efforts could prove beneficial for optimizing field evaluations, matching variety candidates to target environments, as well as for experimenting with new management options. A number of cassava CGMs have been reviewed (Phoncharoen, et al. 2021, Moreno-Cadena et al. 2021). An advantage of the CSM-MANIHOT-Cassava model described there is that is has been integrated within the DSSAT Crop Modeling Ecosystem in DSSAT v. 4.8 (Hoogenboom et al. 2019) for which there is also an R interface (Alderman 2020). The DSSAT system in turn incorporates model calibration functionality, including GLUE.

The MANIHOT-Cassava model was developed from the CSM-CROPSIM model, itself a development of the GUMCAS model (Moreno-Cadena et al. 2020). While (Moreno-Cadena et al. 2020) only verbally describes the model updates, comprehensive details including equations are available in (Moreno-Cadena 2021). Plant development is a function of accumulated thermal time with different cardinal temperatures for the processes of forking and leaf growth. Daily photosynthetic assimilates are obtained by multiplying the solar radiation intercepted by the photosynthetically active radiation use efficiency (PARUE). The model uses a spill-over strategy for biomass allocation: assimilates are allocated by prioritizing the growth of aboveground biomass and fibrous roots according to their demand and the remainder is assigned to the storage roots. Drought stress dynamics are influenced by soil moisture content, impacting germination, branching, leaf appearance, and photosynthesis. Nitrogen limitations are also factored in, though they necessitate further refinement. The current model version requires an array of parameters at the species, ecotype, and cultivar levels, which define stress thresholds, senescence processes, Vapor Pressure Deficit (VPD) effects, root water uptake, and other physiological traits like leaf morphology and growth characteristics. As a DSSAT crop growth model, MANIHOT-Cassava requires minimum data sets including daily weather data, soil surface and profile characteristics, crop management, initial conditions, and crop measurements (Hoogenboom et al. 2012, 2019).

The model performed well in simulating cassava development and growth, including forking dates, leaves and stem biomass, and the yield and yield components (Moreno-Cadena et al. 2020; Phoncharoen, et al. 2021). A previous study assessed the MANIHOT-Cassava model as a breeding tool for identifying the most suitable production environments for potential varieties in Thailand (Phoncharoen, et al. 2021). Nevertheless, any model is a work in progress with the potential to better track responses of many growth processes to external conditions and attempting to use more parsimonious parameterizations. Our objectives were to 1) assess the feasibility of calibrating the model based on observations made by the International Institute of

Tropical Agriculture (IITA) breeding program using GLUE and across a range of elite Nigerian cassava germplasm and growing environments and 2) to identify any systematic biases in model predictions after that calibration, to suggest further improvements to the model. The experimental data included 67 clones evaluated across 11 locations in three years, thus representing a large dataset for model assessment. This report presents the first calibration of the CSM-MANIHOT-Cassava model in Nigeria, a country with a wide range of environmental conditions for cassava cultivation.

Materials and Methods

Model calibration

Experimental design and fields management

Phenotypic data were collected for 67 clones from the Uniform Yield Trials of IITA, which are the most advanced evaluation stage of the cassava breeding program prior to variety release decisions. The clones were derived from crosses within IITA's elite population as part of the genomic recurrent breeding program. They underwent four stages of screening for diseases, vigor, agronomic, and quality traits in earlier trials of the breeding program, showing resistance to diseases and potential for high yield and dry matter content. To standardize the model terminology, we call clone a "cultivar".

The trials were targeted across 11 IITA testing stations in Nigeria (Figure 1) over three growing seasons (2017-2018, 2018-2019, and 2019-2020). We considered each year by location combination to represent one environment. The fields selected for the trials had been fallow, with no crops grown in the previous three years. Before planting, the selected field were mowed, ploughed, harrowed, and ridged. The planting date varied from April to August depending on the rainy season in each location (Table 1). Cultivars were planted in a randomized complete block design with 3 replications. Each plot consisted of 6 rows of 5.6 m (seven plants) with an inter-row spacing of 1 m and intra-row spacing of 0.8 m (plot size = 28 m^2). The trials were managed without pesticides, irrigation, or fertilizer.

Preliminary analysis of each trial was conducted to determine the relative strength of the genetic signal its data. The analysis was done as described by (Bakare et al. 2022) who worked on the same data. A linear mixed-effects model was fitted using the lmer function from the lme4 package in R 4.1.0 (R Core Team 2022). The model formula was:

Trait ~ PropHar + (1|repInTrial) + (1|blockInRep) + (1|germplasmName)

- Trait = fresh storage root weight or fresh above ground biomass as responses variables,
- PropHar = proportion of plants harvested per plot (fixed effect),
- repInTrial = number of replications nested within trial (random effect),
- blockInRep = block nested within replication (random effect),
- germplasmName = cultivar's name (random effect).

The reliability of the model was extracted using the r.squaredGLMM function from MuMIn package in R. The broad-sense heritability and the reliability for both traits (fresh storage root weight or fresh above ground biomass) were calculated by trial. Trials were filtered to retain those

with H^2 over 0.25 and reliability over 40% for at least one trait (Supplementary Figure S1). A total of 28 trials out of 38 trials were retained after filtering, included eight locations and all three years. These trials spanned 16 environments.

Soil and weather

For each location, specific soil and weather data were used. Soil profile characteristics were obtained from the Harvard Dataverse soils data Repository for crops (Han et al. 2019) according to the GPS coordinates for each location. The soil ID code and properties are given in Supplementary Table S1. The considered parameters included: Soil Lower limit of plant extractable soil water (SLLL), Soil Drained upper limit (SDUL), Saturated upper limit (SSAT), Soil root growth factor (SRGF), Saturated hydraulic conductivity (SSKS), Soil bulk density (SBDM g/cm^3), Soil organic carbon (SLOC), Soil clay (SLCL), Soil silt (SLSI), Soil total nitrogen (SLNI), Soil pH in water (SLHW) and Soil cation exchange capacity (SCEC).

Daily weather data were collected from the National Aeronautics and Space Administration Prediction of Worldwide Energy Resource project *(https://power.larc.nasa.gov/data-access-viewer/)*. The data set for the growing season, summarized in the Table 1, included: minimum temperature (TMIN), maximum temperature (TMAX), precipitation (RAIN), relative humidity (RH2M), solar radiation (SRAD) and wind speed (WIND).

The DSSAT Weatherman tool was used to create input weather data files in the model format (Hoogenboom et al. 2019). The initial soil water conditions for each layer were defined by setting available soil water at 100%, the incorporation depth for surface residue at 20 cm and the simulation was done under non-nitrogen-limited and non-phosphorus-limited production conditions.

Plant phenology

Plant phenological and leaf area index (LAI) data were collected only Ibadan in 2021. Five plants per plot, with two plots per cultivar, were assessed at 5 and 6 months after planting as follows. From the first branching level to the top of the plant, one branch was selected to be monitored at each branching point. Plant height at first branching, full plant height (from the bottom to the last leaf), length, and the number of internodes of each section between 2 forking points were measured. Three weeks later, the height and the number of internodes of new growth were measured. From these measurements, the phyllochron or time for node development was calculated. The forking time was estimated as a function of the section height and internode number. In the event of a genotype not showing forks during data collection, the default values provided by the model were used (Moreno-Cadena et al. 2020). The details of these default values can be found at: https://github.com/DSSAT/dssat-csm-os/tree/develop/Data/Genotype.

Table 1. Description of experimental sites in Nigeria

Location	Lat / Long	Soil	YEAR	Name	Planting Date	TMIN (°C)	TMAX (°C)	RAIN (mm)	RH2M (%)	SRAD ($MJ/m^2/day$)	WIND (m/s)
ABUJA	9.07°N, 7.47°E	SCL	2019-2020	Abuja19	June 2019	20.8	30.4	2238	73.2	18.6	1.6
AGO-OWU	7.26°N, 4.32°E	SCL	2017-2018	Ago17	August 2017	21.8	29.2	1479	85.9	16.7	1
AGO-OWU	7.26°N, 4.32°E	SCL	2018-2019	Ago18	June 2018	22.0	28.9	2842	87.1	16.6	0.9
AGO-OWU	7.26°N, 4.32°E	SCL	2019-2020	Ago19	June 2019	21.7	29.1	3152	85.6	16.7	1
IBADAN	7.47°N, 3.93°E	SL	2017-2018	Ibadan17	September 2017	22.0	29.1	2179	86.4	16.5	1.5
IBADAN	7.47°N, 3.93°E	SL	2018-2019	Ibadan18	July 2018	22.2	29.3	2316	86.5	16.9	1.4
IBADAN	7.47°N, 3.93°E	SL	2019-2020	Ibadan19	May 2019	21.8	29.4	2397	85	16.6	1.5
IKENNE	6.87°N, 3.71°E	CL	2017-2018	Ikenne17	August 2017	22.8	29.2	2075	88.2	16.3	1.1
IKENNE	6.87°N, 3.71°E	CL	2018-2019	Ikenne18	June 2018	23.0	29.3	2533	88.4	16.5	1.1
IKENNE	6.87°N, 3.71°E	CL	2019-2020	Ikenne19	June 2019	22.9	29.3	3038	88	16.5	1.1
MOKWA	9.31°N, 5.07°E	SCL	2017-2018	Mokwa17	August 2017	22.2	32.6	1594	67.8	18.3	1.7
MOKWA	9.31°N, 5.07°E	SCL	2018-2019	Mokwa18	January 2018	22.1	32.4	1351	68.5	18.6	1.7
MOKWA	9.31°N, 5.07°E	SCL	2019-2020	Mokwa19	June 19	21.8	32.5	1385	69.1	18.8	1.8
ONNE	4.79°N, 7.17°E	SCL	2018-2019	Onne18	April 2018	23.3	29.6	3319	88.1	14.9	0.9
UBIAJA	6.67°N, 6.39°E	SCL	2019-2020	Ubiaja19	June 2019	22.1	29.3	2725	84.1	16.7	1.6
UMUDIKE	5.47°N, 7.55°E	SCL	2018-2019	Umudike18	July 2018	22.9	29.4	1966	86.3	16.2	1.4

*** Geo Loc = geographical location, SCL = Sandy Clay Loam, SL = Sandy Loam, CL= Clay Loam, Name = Location/year, TMAX = Daily temperature maximum, TMIN = Daily temperature minimum, RAIN = Sum precipitation /year, RH2M = Relative Humidity at 2 m, SRAD = Daily solar radiation, PAR = Daily photosynthetic radiation, WIND = Wind Speed at 2 m. The meteorological data corresponds to the growing season.**

Leaf area index

We established a predictor model to estimate leaf area. We collected fully expanded, photosynthetically active, and undamaged leaves from the top, middle, and bottom of a plant not used for phenology. Thirty leaves were sampled per cultivar. The number of lobes, length, and width of the central lobe were measured for each leaf. The leaf was scanned to calculate its area with ImageJ software bundled with Java 8 on 10/29/24 9:51:00 AM. With these data, we established a model (Equation 1) that used the length and width of the central lobe and the number of lobes to estimate the leaf area of cassava (Figure 2).

From the targeted plant (5 plants per plot), three leaves per plant (from top to bottom) were evaluated by the number of lobes, the length, and the width of the central lobe. The leaf area was estimated with the established equation (Equation 1). Then LAI was calculated as the total number of green leaves per plant multiplied by the mean per plant of leaf area and divided by the area occupied of each plant (Equation 2).

Leaf Area = 24 + 0.39 * Lobe length * Lobe width * Number of Lobes Eq. 1

Leaf Area Index = Number of Leaves * Leaf Area / Plant Occupied Area Eq. 2

Yield and yield components

Yield and yield components were collected from 2017 to 2020 in 11 locations (Table 1). At harvest, fresh storage root yield and aboveground biomass (stem + leaves) were weighed separately. A net plot of 20 plants was harvested, excluding all plants on the border of a plot to avoid border effects. The storage root dry weight was calculated as a function of the fresh storage root yield and the dry matter content (Equation 3).

Dry matter content (DMC) in Ibadan was determined from 100 g of fresh roots of each plot grated and oven-dried at 65ºC for 72 hours to constant weight. After the samples were cooled, their dry weights were recorded, and their respective DMC values were determined. For the other locations, the specific gravity method (Wholey & Booth 1979) was used to estimate the DMC per plot. The dry aboveground biomass was calculated as 1/3 of the fresh aboveground biomass one (Streck et al. 2014).

Storage root dry weight = Storage root fresh weight * Dry matter content Eq. 3

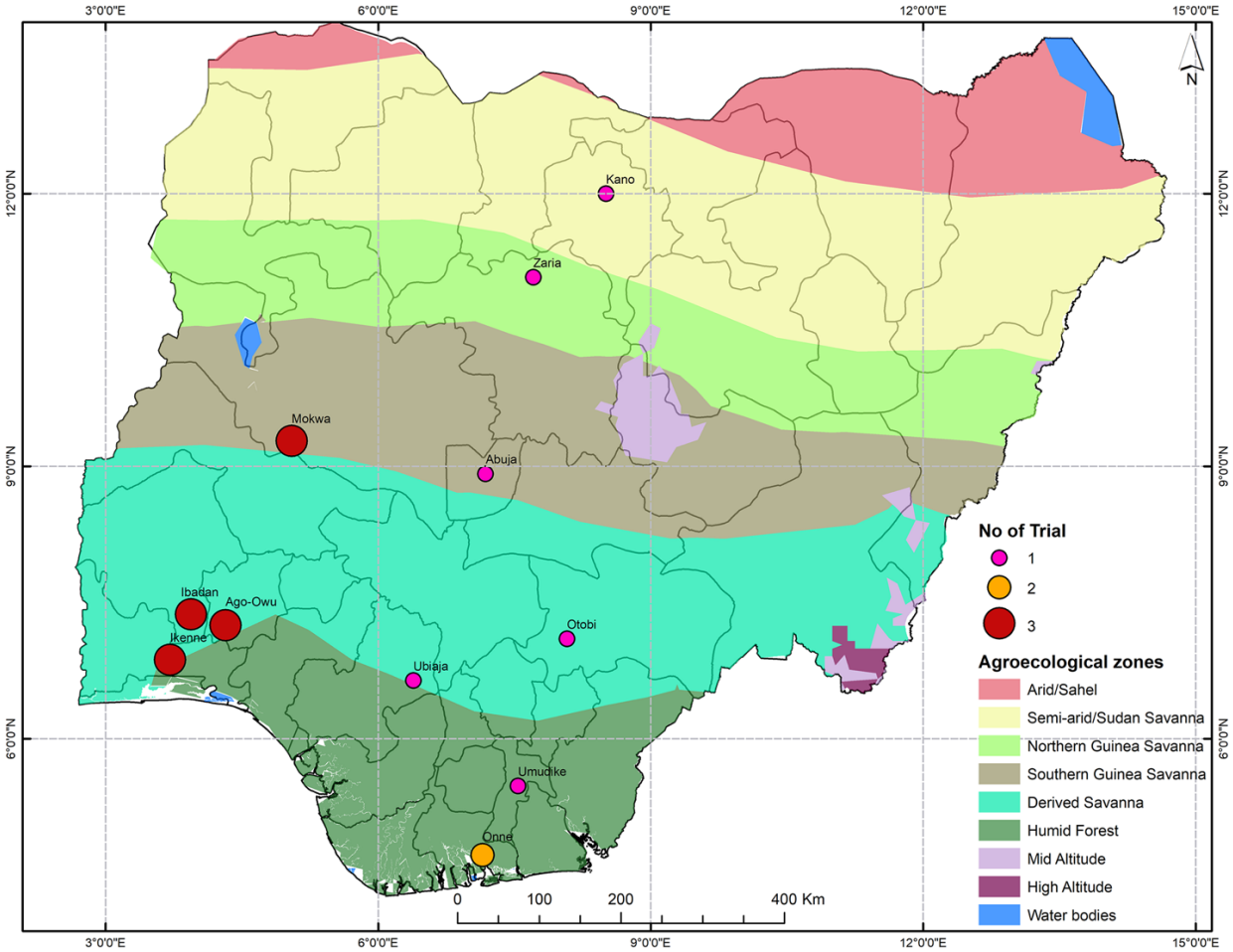


Figure 1. Map of Nigeria showing the locations of multi-environment trials in the IITA cassava breeding program. The agrometeorological zones and the number of tests carried out specified by the colors and the size of the point respectively.

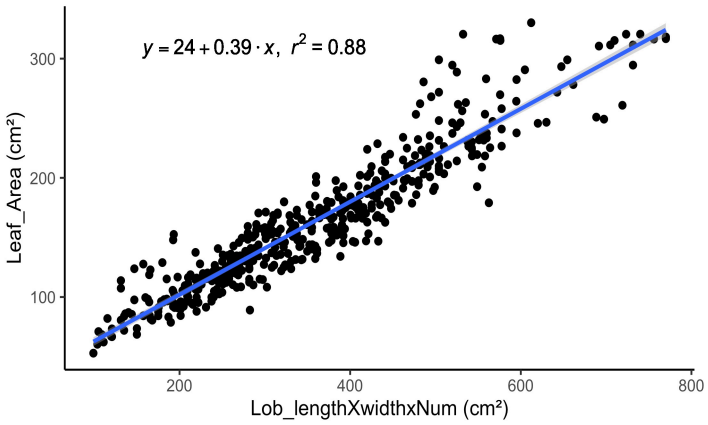


Figure 2. Linear regression between the product of leaf lobe length, leaf lobe width, leaf lobes number width and leaf area for 67 cassava cultivars

Estimating genotype specific parameters

The DSSAT crop model contains genetic parameters at the levels of species, ecotype, and cultivar (Hoogenboom et al. 2019). For model evaluation, coefficients at the level of cultivars, referred as Genotype-Specific Parameters (GSPs), need to be calibrated. We also adjusted two ecotype coefficients directly involved in yield (Moreno-Cadena et al. 2020; Rankine et al. 2021).

The GSPs define differences among cultivars within a crop species (Hoogenboom et al. 2019). The 15 cultivar and 2 ecotype parameters incorporated within CSM-MANIHOT-Cassava that we calibrated for each cultivar are defined in Table 2. The prior distribution of each GSP was provided by Moreno-Cadena et al. (2020).

Two methods were used to estimate GSPs. The first group of GSPs, including the phenological parameters and some growth parameters, were hand-calibrated following the trial and error procedure recommended by Hoogenboom et al. (2019) and used by Phoncharoen et al. (2021b). Simulations were repeated until we could not increase the agreement evaluated by the correlation (r) and the Root Mean Square Error values (RMSE) between the simulated and observed values. That method was adopted because the target variables were measured only in Ibadan 2021. The calibration procedure was as follows:

- Parameters defining plant phenology were calibrated in the respective order of their progress during growth. They included thermal time to first forking (B01ND), second forking (B12ND), third forking (B23ND), and fourth forking (B34ND). The target variables were the time of these events from planting (days after planting).
- The growth parameters measuring the number of branches per fork (BR1FX, BR2FX, BR3FX, and BR4FX) were directly measured. In case of a lack of data, the default value from the model GSPs file was applied.
- Leaf parameters, including the maximum leaf area (LAXS) and leaf life (LLIFA) were calibrated. The target variable was leaf area index (LAI) as computed in Eq. 2.

The second step of the calibration was done by the Generalized Likelihood Uncertainty Estimation (GLUE) method. The calibrated parameters were the specific leaf lamina area (SLAS), leaf petiole fraction (LPEFR), a slope for leaf production (LNSLP), node weight (NODWT) and internode length (NODLT). The GLUE method is one of the first methods to represent prediction uncertainty within the context of Monte Carlo analysis coupled with Bayesian estimation and propagation of uncertainty. The technique uses a likelihood function to measure the closeness-of-fit of modeled and observed data (He et al. 2010). The goal was to minimize the RMSE between simulated and observed values. The target variables were the aboveground biomass dry weight. The GLUE method was used because of its flexibility, ease of implementation as an R program, and its suitability for parallel implementation on distributed computer systems. It has previously been used to estimate GSPs for DSSAT crop models (Hoogenboom et al. 2019) such as wheat (Ibrahim et al. 2016; Li et al. 2018), maize (Mereu et al. 2019), and rice (Buddhaboon et al., 2018) as examples. The main steps of the GLUE program as described by Beven & Binley (1992), He et al. (2010), and updated by He et al. (2021), are summarized in Supplementary Methods.

After these two steps, if the storage root dry weight was systematically underestimated or overestimated across observations of a cultivar, the photosynthetically active radiation conversion factor (PARUE) and PAR extinction coefficient (KCAN) that are ecotype parameters were adjusted manually to remove the bias. Then, the GLUE method was run again as defined above to

estimate the GSPs using new values of PARUE and KCAN.

Table 1: Definition of the CSM- MANIHOT-Cassava genotype parameters in DSSAT

Level	Parameter	Description	Function
Cultivar	B01ND	Duration from planting to first forking (°Cd)	Phenology
	B12ND	Duration from first to second forking (°Cd)	Phenology
	B23ND	Duration from second to third forking (°Cd)	Phenology
	B34ND	Duration from second to third forking (°Cd)	Phenology
	BR1FX	Branch number per fork at fork 1 (no.)	Growth
	BR2FX	Branch number per fork at fork 2 (no.)	Growth
	BR3FX	Branch number per fork at fork 3 (no.)	Growth
	BR4FX	Branch number per fork at fork 4 (no.)	Growth
	LAXS	Maximum leaf area when growing without stress (cm^2)	Growth
	SLAS	Specific leaf lamina area when growing without stress (cm^2 g^{-1})	Growth
	LLIFA	Leaf life from full expansion to start senescence (°Cd)	Growth
	LPEFR	Leaf-petiole fraction (fraction of lamina petiole)	Growth
	LNSLP	Slope for leaf production	Growth
	NODWT	Node weight for the first stem of the shoot before branching at 3400 °Cd (g)	Growth
	NODLT	Mean internode length (cm) for the first stem of the shoot before branching when is lignified (cm)	Growth
Ecotype	PARUE	Photosynthetically active radiation (PAR) conversion factor standard (g dry matter MJ^{-1})	Ecotype
	KCAN	Photosynthetically active radiation extinction parameter (no.)	Ecotype

Statistical analysis

Model performance analysis

The performance of the calibrated CSM-MANIHOT-Cassava model was analyzed using the correlation (r), the Root Mean Square Error (RMSE, Wallach & Goffinet 1987), the normalized root mean square error (nRMSE, Loague & Green 1991) and the index of agreement (d, Willmott 1982). These four statistical parameters were analyzed separately for storage root dry weight. The parameters r (Equation 4), RMSE (Equation 5) and nRMSE (Equation 6) measure the deviation of the simulated from the observed values. The nRMSE is the RMSE value divided by the observed mean. The simulation is considered excellent with nRMSE < 10%, good in the 10-20% range, fair in the 20-30% range, and poor if nRMSE is > 30% (Harb et al. 2016; Rankine et al. 2021). The index of agreement, *d*, (Equation 7) was added because it incorporates both bias and variability. The *d* value varies between 0 and 1. A value of 1 indicates a perfect match, and 0 indicates no agreement at all.

$$\mathrm{r} = \frac{\sum_{i=1}^{M}(S_i - \bar{S})\,(O_i - \bar{O})}{\sqrt{\sum_{i=1}^{M}(S_i - \bar{S})^2 \sum_{i=1}^{M}(O_i - \bar{O})^2}} \qquad \text{Eq. 4}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{M}(S_i - O_i)^2}{M}} \qquad \text{Eq. 5}$$

$$nRMSE = \frac{RMSE\ X\ 100}{\bar{O}} \qquad \text{Eq. 6}$$

$$d = 1 - \frac{\sum_{i=1}^{M}(S_i - O_i)^2}{\sum_{i=1}^{M}(|S_i| + |O_i|)^2} \qquad \text{Eq. 7}$$

M = number of observations

S_i = simulated value

$\bar{S}$ = mean of the simulated values

O_i = observed value

$\bar{O}$ = mean of the observed values

Model sensitivity analysis

The main objective was to identify variables explaining deviations between model predictions and experimental data. The hypothesis was that the differences between the simulated and the observed values could be due either to a bad calibration of the GSPs (leading to a genotype effect on model deviation) or to the fact that some parameters of the environment were not well accounted for in the model (leading to an environment effect on model deviation) or to an interaction between these two causes (leading to a genotype by environment interaction effect on model deviation).

We analyzed deviations using the Additive Main effect and Multiplicative interaction (AMMI) model implemented in the statgenGxE R package (van Rossum et al. 2022). The AMMI model separates additive main effects and multiplicative interaction effects. The model analyses the additive effects with classical analysis of variance (ANOVA) then applies principal component analysis (PCA) to the residual interaction portion of variation (Hongyu et al. 2014; Nowosad et al. 2016; Yan & Rajcan 2002).

The AMMI model is:

$$y_{ij} = \mu + G_i + E_j + \sum_{m=1}^{M} \gamma_{mi}\delta_{mj} + \epsilon_{ij}$$

where:

- M was the number of principal components included,
- y_{ij} was the deviation between observed versus simulated dry storage root yield for genotype i in environment j. That is, y_{ij} = Observed storage root dry weight – simulated storage root dry weight,
- μ was the general mean,
- G_i was the genotypic effect,
- E_j was the environmental effect,
- γ_{mi} was the genotypic score for principal component m and genotype i,
- δ_{mj} was the environment loading for principal component m and environment j,
- ϵ_{ij} was the residual.

For example, if after calibration, observed yields of a genotype across environments are consistently higher than those simulated by the CSM-MANIHOT-Cassava model, G_i will be positive. Likewise, if observed yields across genotypes in an environment are consistently higher than those simulated by CSM-MANIHOT-Cassava, E_j will be positive. Conditions for γ_{mi} and δ_{mj} to be non-zero would require that certain aspects of the genotype interact with certain aspects of the environment to cause a bias in CSM-MANIHOT-Cassava simulations. Explanatory variables considered for these analyses included weather data, soil data, and additional phenotypic data, encompassing diseases such as Cassava Bacterial Blight incidence and severity (CB_), Cassava Green Mite severity (CGM), Cassava Mosaic Disease incidence and severity (CMO), as well as leaf area (LeafArea) and plant morphology, such as Plant forking number (branchP), Stem number (stem_nu), Stem diameter (stem_D), and Plant height (plant_H). The methods used for obtaining additional data were consistent with those described in Rabbi et al. (2014), Rabbi et al. (2017), and Wolfe et al. (2016).

Results

Variation among optimized GSPs

The cultivars presented variation in the optimized parameters controlling phenological events and growth (Table 3). For the phenology, the first forking (B01ND) time presented the highest coefficient of variation (CV = 62%). The second, the third, and the fourth forking time showed substantially lower CVs. The number of branches for the first forking was between 2 and 4, and some cultivars had 4 branches in the second branching. Parameter variation was observed at both the leaf and internode. The leaf production and functioning parameters, such as LAXS and SLAS presented CV values of 19% and 13%, respectively. Similarly, the internode weight (NODWT) and length (NODLT) showed CVs of 32% and 27%, respectively, which were the highest CVs among the GLUE-

362 optimized parameters. Finally, among the parameters affecting photosynthetic activity,
363 PARUE and KCAN had 21% and 15% CV values, respectively

364 **Table 2**. Maximum (Max), minimum (Min), mean and coefficient of variation (CV)
365 values of genotype-specific parameters calibrated for the 67 cassava cultivars in the
366 MANIHOT-Cassava model.

Coefficient	Fitted by	Max	Min	Mean	CV (%)
B01ND (°Cd)	Trial & Error	1200	231	487	62
B12ND (°Cd)	Trial & Error	344	80	247	19
B23ND (°Cd)	Trial & Error	286	101	235	15
B34ND (°Cd)	Trial & Error	250	100	243	11
BR1FX	Trial & Error	3.8	2	2.5	13
BR2FX	Trial & Error	4	2	2.7	17
BR3FX	Trial & Error	3	1	2.1	17
BR4FX	Trial & Error	2	1.4	1.5	4
LAXS (cm^2)	Trial & Error	920	401	754	19
LLIFA (°Cd)	Trial & Error	1800	600	1282	32
SLAS (cm^2 g^{-1})	GLUE	280	149	252	13
LPEFR	GLUE	0.3	0.2	0.3	17
LNSLP	GLUE	1.7	0.7	0.8	19
NODWT (g)	GLUE	6	1.7	3.3	32
NODLT (cm)	GLUE	4	1	2.9	27
PARUE (g dry matter MJ^{-1})	Yield Matching	2.6	1.4	1.8	21
KCAN	Yield Matching	0.8	0.5	0.6	15

367

368 **Model performance**

369 **Performance by cultivar**

370 The analysis included the 67 cultivars evaluated across 16 environments. For each cultivar we
371 calculated the Pearson correlation r of model prediction and observation across the environments,
372 as well as the RMSE and d statistic. The GLUE estimated GSPs improved the accuracy of the dry
373 storage root yield simulated by the model (Figure 3). The correlation coefficient increased from –
374 0.03 to +0.08, the RMSE dropped from 21 t ha^{-1} to 5 t ha^{-1} while d increased from 0.23 to 0.44. All
375 these changes were highly significant using a paired t-test.

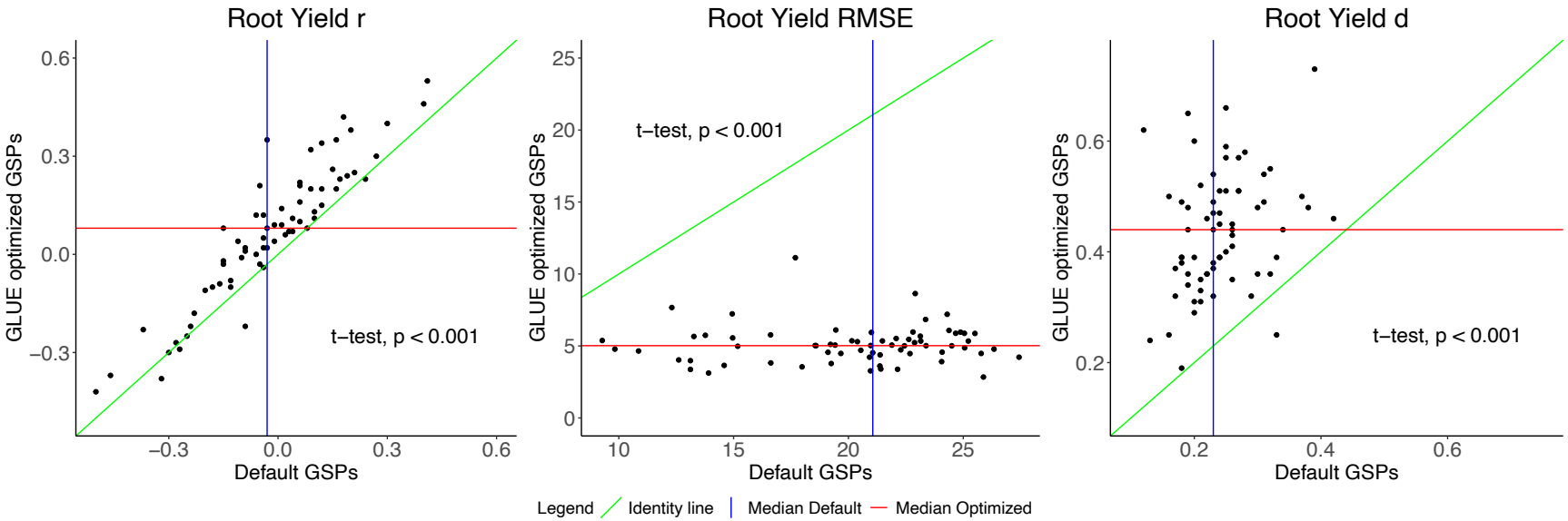


Figure 3. Comparison of the Pearson correlation coefficient (r), the root mean squared error (RMSE), and the index of agreement (d) values for the MANIHOT-Cassava predicted storage root dry weight before and after using General Likelihood Uncertainty Estimation (GLUE) method for Genotype-Specific Parameters (GSPs) estimation. Each point represents one cultivar.

382 **Performance by environment**

383 The r, RMSE, and d statistics could also be calculated across cultivars within environments. In this
384 per-environment analysis, we also observed reasonable agreement between the simulated and the
385 observed storage root dry weight. A total of 69% of the environments had an nRMSE ≤ 20%
386 (considered excellent to good). Some environments such as Ikenne17 and Mokwa18 were very
387 well simulated with nRMSE values of 9% and 10%, RMSE of 2.2 and 2.7 tha^{-1} and d values of
388 0.74 and 0.63 respectively (Table 4). Equally, substantial deviation of the model was observed in
389 some environments such as Ikenne18 and Onne18 (nRMSE ≥ 30%). Given this observation,
390 statistical analyses of simulation results were performed to determine the causes of the model
391 deviations.

392 **Table 3**. Agreement between simulated (Sim) and observed (Obs) mean storage root dry
393 weight of the cultivars across environments for MANIHOT-Cassava calibration

Location	Environment	Obs	Sim	r	RMSE ($t\ ha^{-1}$)	nRMSE (%)	d	Total cultivars
Abuja	Abuja19	8.9	9.4	0.28	3.0	13	0.58	67
Ago-Owo	Ago17	11.3	9.7	0.28	2.7	14	0.51	36
	Ago18	9.3	11.4	-0.1	4.5	24	0.23	36
	Ago19	14.4	9.9	0.48	5.3	20	0.47	36
Ibadan	Ibadan17	14.2	11.3	0.22	4.7	13	0.51	33
	Ibadan18	16	13.4	0.48	4.3	11	0.62	57
	Ibadan19	8.6	8.4	0.06	3.2	14	0.41	66
Ikenne	Ikenne17	11.4	11.7	0.56	2.2	9	0.74	67
	Ikenne18	5.4	14.3	0.36	9.4	57	0.25	67
	Ikenne19	13.5	11.6	0.6	3.8	11	0.69	67
Mokwa	Mokwa17	12.1	3	0.35	9.7	26	0.38	63
	Mokwa18	6.4	5.3	0.43	2.7	10	0.63	67
	Mokwa19	6	6.4	0.25	2.4	17	0.47	36
Onne	Onne18	6.1	10.9	0.38	5.8	33	0.36	67
Ubiaja	Ubiaja19	5	10.1	0.44	5.7	32	0.38	67
Umudike	Umudike18	11.4	8.8	0.51	3.7	13	0.61	67

394

395 **Genotype and environment effects on model prediction error**

396 The objective of the AMMI analysis was to identify the possible causes that could be ascribed to
397 cultivars or environments of poor model performance in predicting root yield. The analysis was
398 applied to the difference between observed and simulated storage root dry weight (observed –
399 simulated = "O-S"). The AMMI model evaluated 66 genotypes across 16 environments. One
400 genotype out of the 67 was not included because it had more than 10% missing data. The ANOVA
401 analysis revealed a significant effect only for the environment explaining model error (Table 5).
402 The effect of the genotype and environment interaction was also highly significant ($p < 0.001$).
403 The first three interaction principal components (PCs) explained 50% of the GEI variation.

404 The biplot of model prediction errors indicated that the environment effects were strongest at
405 Ago18 and Mokwa17 whose effects were opposite on the component of GEI explained by PC1
406 (Figure 5).

Table 4. Analysis of variance for the model storage root dry weight (t ha^{-1}) deviation (observed – simulated) using the AMMI model. With degrees of freedom (df), mean square (Mean Sq), the F value and P value and the proportion of sums of squares (% Total SS).

Source	df	Mean Sq	F value	P value	% Total SS	Cumulative Proportion (%)
Environment	15	1153.7	164	<0.001***	70	
Genotype	66	6.8	1.5	0.55	2	
Interactions	990	7			28	
PC1	80	19.7	4.3	<0.001***		23
PC2	78	14.8	3.2	<0.001***		39
PC3	76	10.1	2.2	<0.001***		50
Residuals	756	4.5				

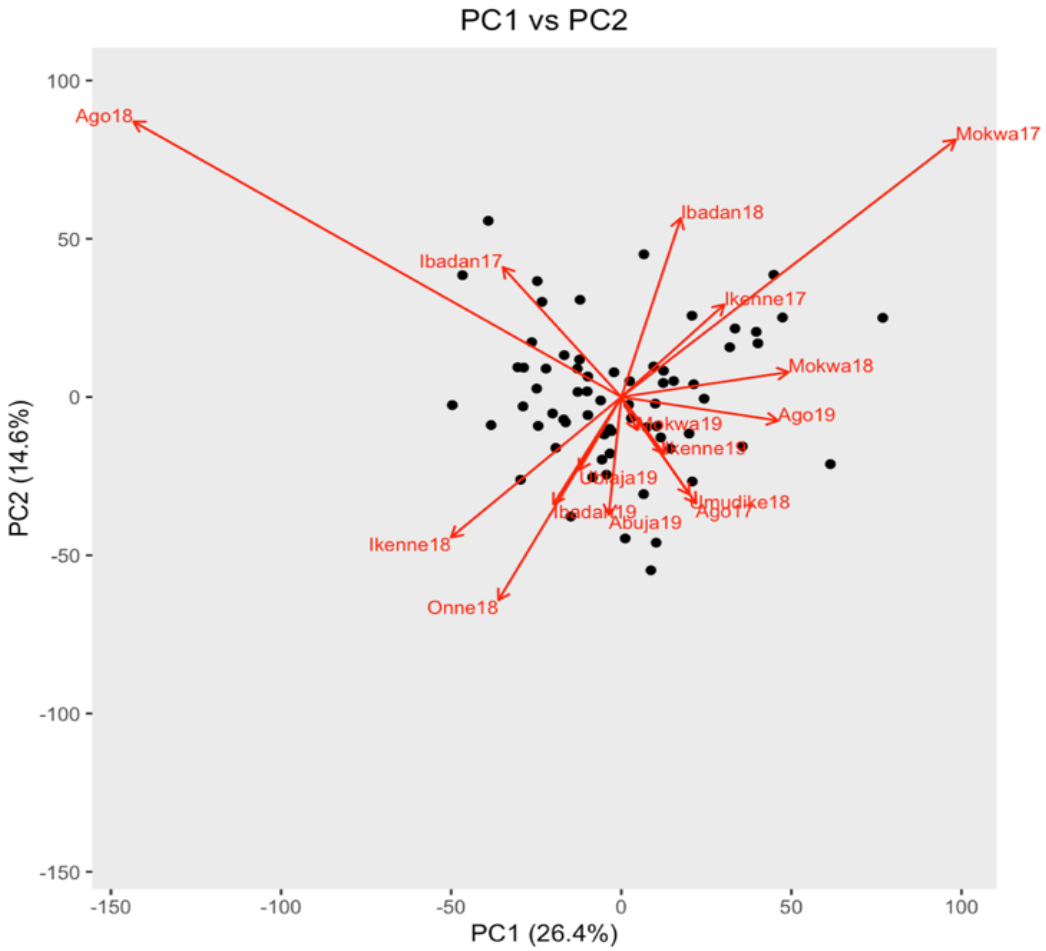


Figure 4. Biplot of genotypes and environment interaction with PC1 and PC2. Genotypes are black points, and environments are red arrows.

Effects of environment characteristics on model prediction error

To identify characteristics of environments that affected model accuracy, we correlated weather and soil profile data with the AMMI analysis output including the mean value of model deviations (i.e., "O-S") by environment (envMean) and the environment loading values on the three PCs.

Many input weather parameters affected the model performance (Figure 6). Indeed, the solar radiation, maximum temperature, wind, and daily photosynthetic radiation were positively correlated with the deviations while rainfall and relative humidity were negatively correlated (p-values < 0.01). Root yields in dry (low rain) environments marked by high temperatures and high solar radiation were underestimated by the model, while yields in wet (high rain) environments with low temperatures and low solar radiation were overestimated (Figure 7). These weather parameters affected estimation more in the first six months of plant growth: parameters at 6 months after planting (parameter abbreviation with "_6") had the strongest correlations.

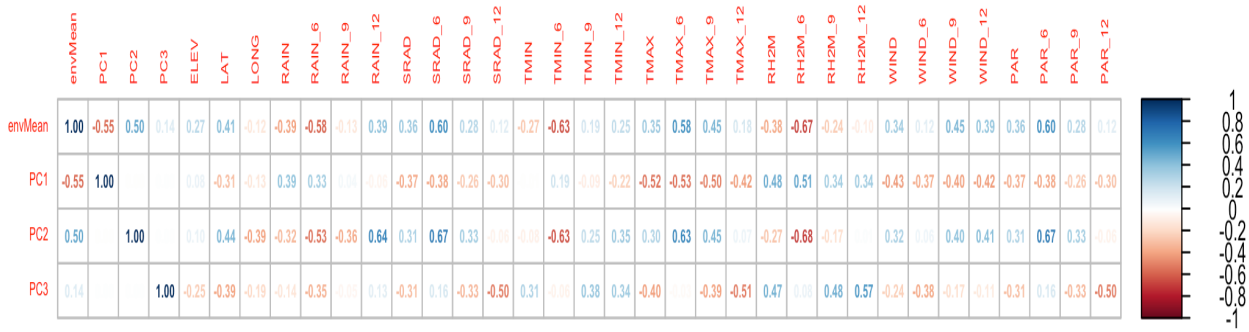


Figure 5. Correlation between the mean value of the model deviation ("Observed – Simulated") by environment (envMean), the environment loading values on PC1, PC2 and PC3, and weather parameters The suffix number ("_x") defines the months after planting, e.g., TMAX_6 is the maximum temperature at 6 MAP. Abbreviations are: ELEV: location altitude (m), LAT: location latitude, LONG: location longitude, RAIN: rainfall from planning to harvest; SRAD: solar radiation, TMIN: minimum daily temperature, TMAX: daily maximum temperature, RH2M: relative humidity, WIND: average wind speed at 2m, PAR: Photosynthetically active radiation.

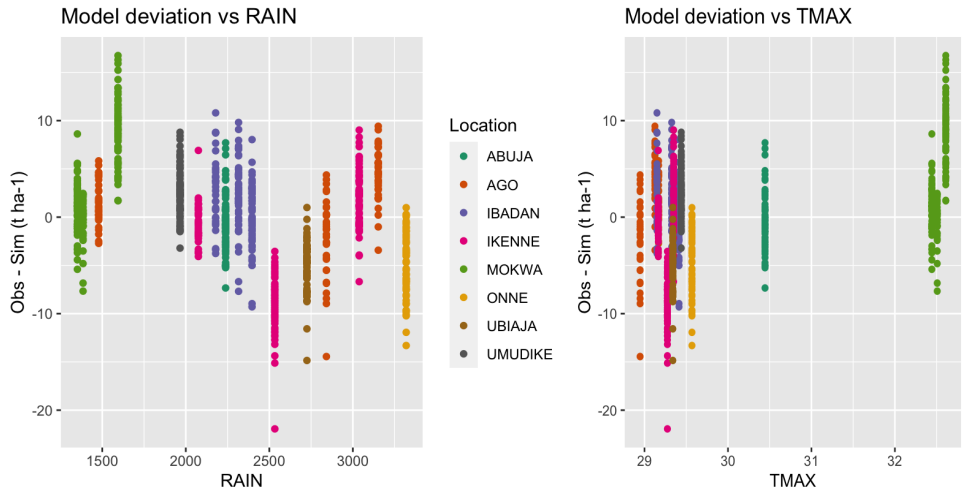


Figure 6. Scatterplot of the mean model deviation ("Observed – Simulated") against the annual total rain by environment (RAIN) and the annual mean maximum temperature by environment (TMAX)

Subsoil characteristics at 15 and 100 cm were strongly correlated, and we only present results at 15cm. Only the soil organic carbon (SLOC), significatively affected MANIHOT-Cassava performance showing a negative correlation with the model deviations (Figure 8).

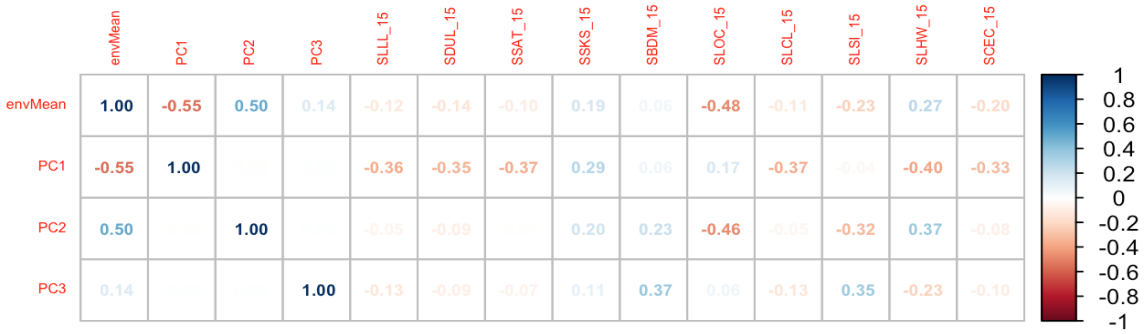


Figure 7. Correlation between the mean value of the model deviation ("Observed – Simulated") by environment (envMean), the environment loading values on PC1, PC2 and PC3, and soil profile parameters. Soil parameters abbreviations are SLLL: Lower limit of plant extractable soil water (cm^3), SDUL: Drained upper limit (cm^3), SSAT: Upper limit, saturated (cm^3), SSKS: Saturated hydraulic conductivity (cm/h), SBDM: Bulk density (g/cm^3), SLOC: Organic carbon (%), SLCL: Clay (%), SLSI: Silt (%), SLHW: pH in water, SCEC: Cation exchange capacity (cmol/kg).

Discussion

Variation in cultivar GSPs

Crop Growth Models are usually calibrated in non-stress environments. These environments enable the expression of the cultivar's inherent growth and potential without confounding them with stress factors. Our objective here, however, was to calibrate with breeding trial data to avoid requiring the breeding program to allocate resources to model calibration experiments. We assume that the lack of ideal environments can be compensated for by having data from many environments.

Across the 15 GSPs, we used three calibration approaches: Trial & Error, GLUE, and Yield Matching (Table 3). For the Trial & Error calibration we only used data from the Ibadan 2021 trial for which we had estimates of leaf area and of branch phenology and number of forks. These parameters were fixed in part because we found that the GLUE approach could not handle more than the five GSPs that were calibrated with it. Because GLUE samples from the prior distribution of parameters it is not super-efficient at narrowing in on values with maximum posterior density (He et al. 2010). Recent efforts have been made to parallelize GLUE (Berton Ferreira 2024) to make it more computationally feasible. Finally, we set PARUE and KCAN to come close to matching the average yield of the cultivars. This approach was chosen to reduce the impact of genotype main effects on model calibration and instead capture sources of genotype by environment interaction. The approach is not unlike post hoc adjustments for harvest index to match harvested yield to model biomass simulation (as might have been done, for example, in Wellens et al. 2022). One consequence of that choice is that model deviations were barely affected by genotype (Table 5) so that we were not able to explore effectively genotype characteristics that might affect model performance. Adjusting to match genotype main effect, however, arithmetically could not affect the correlation for a genotype between simulation and observation across environments. Thus, the model improvement in that correlation from a median across cultivars of -0.03 for the default parameters to 0.08 was due to calibration of the other parameters. While the optimized correlation is not high, the improvement was statistically significant (Figure 3).

Across calibration approaches, we found evidence of variability between the IITA cultivars. Concerning plant phenology, the CV among cultivars was 62% for thermal time to first branching. This parameter establishes the difference between early- and late-branching genotypes (Moreno-Cadena et al. 2020). Similar results were obtained on the elite cassava cultivars from the National Crops Resource Research Institute breeding program of Uganda (Ibrahim et al. 2020) using standard statistical analysis. For Thai genotypes, Phoncharoen et al. (2021b) obtained a very high variation 97%, unlike the Jamaica genotypes where Rankine et al. (2021) found a CV of only 10%. Branching is a significant event in the cassava's life cycle. Phenology not only determines the rate of production of leaves and the architecture of the plant but also accounts for the plant's ability to flower, because an inflorescence is always coupled with branching (Halsey et al. 2008; Jennings & Iglesias 2001). Thus, the level and timing of branching influence flower availability and fertility (Ibrahim et al. 2020). The B12ND, B23ND and B34ND parameters and their corresponding number of branches (BRnFX) could be interpreted because not all the cultivars showed a fork when the data was collected between 5 and 6 months after planting. For those cultivars, the model default values were used (Moreno-Cadena et al. 2020).

There was also variation among the cultivars for the leaf parameters. These results indicate that the germplasm of IITA was characterized by high variability of leaf production, growth, as well as leaf functioning as shown in the PARUE and KCAN parameters. The importance of getting leaf parameters well calibrated is that the MANIHOT-Cassava model uses a surplus of assimilate approach to determine when to allocate to storage roots. We were only able to measure LAI in one environment. More data on cassava LAI across studies might improve model performance. At least two different data collection times during the growing season are needed to calibrate the genetic parameters (Banterng et al. 2009; Phoncharoen al. 2021b; Rankine et al. 2021; Suriharn et al. 2007).

The individual node weight (NODWT) showed a high variation across the target germplasm followed by node length (NODLT) and leaf production slope (LNSLP). According to Moreno-Cadena et al. (2020), NODWT and LNSLP are both essential GSPs affecting plant total biomass with the latter being particularly important in warm environments. The NODWT parameter is important because it affects aboveground photosynthate demand which must be satisfied before the remaining assimilates are allocated to the storage roots. A high value for NODWT can lead to a decrease in storage root dry weight (Moreno-Cadena et al. 2021; Phoncharoen et al. 2021b).

Comparison to other model calibrations

Model calibration across 67 cultivars with final fine calibration using the GLUE method improved the prediction of the storage root dry weight. Strong model improvements for the RMSE and d statistics were no doubt partially due to the yield matching approach we used to calibrate PARUE and KCAN. The improvement in r that we observed was due to other GSP adjustments, however. Similar methodology was successfully used to calibrate the CSM-CERES-Wheat and CSM-CERES-Maize models for winter wheat and maize grain yield simulation, achieving RMSE of 0.27 and 1.57 t ha^{-1}, respectively, and d values of 0.98 and 0.85, respectively (Ibrahim et al. 2016; Mereu et al. 2019). However, RMSE values of 2.19 and 3.72 t ha^{-1} were obtained for durum wheat and common wheat grain yield simulation, respectively, which are similar to the results we obtained in some environments (Mereu et al. 2019). The cassava storage root yield for six locations in Nigeria was simulated with an RMSE of 4.9 t dry matter ha^{-1} using the LINTUL-Cassava model (Adiele et al. 2021). The nRMSE values obtained by environment were similar to those reported by Phoncharoen et al. in two diferrent studies (2021a; 2021b). Finally, Wellens et al. (2022) achieved similar normalized RMSE in the range of 11% to 33%.

Finally, our model fit statistics could have been affected by the fact that we did not have explicitly measured local soil and weather data. Lack of accurate local soil profile data and specific topsoil chemical analysis data could lead to the model prediction errors. To assess this possibility further, we calibrated the model under both nitrogen-limited and non-nitrogen-limited conditions; phosphorus not being an important limiting factor in the CSM-MANIHOT-Cassava model (Moreno-Cadena et al. 2021). Model accuracy was better under non-nitrogen-limited conditions. The weather data we used came from NASA sources; a local weather station would have been preferred. In future research we will also seek to use CHIRPS rainfall data (Funk et al. 2015).

Model variables to investigate further

To our knowledge no cassava study has separately calibrated a model for as many cultivars as we have here. We believe that the extensive set of cultivars we calibrated lends robustness to our

conclusions on model performance. The analysis of model prediction errors showed that root yield prediction error was linked to the input weather parameters SRAD, TMAX, wind, rain, and RH2M (Figure 7). The fact that genotypes clustered together when plotted by environments also reflects the importance of environment on model prediction error (Figure 5). The correlation between weather parameters and model prediction error could be used to classify the environments and identify physiological relationships to be improved in the model.

The underestimated environments were characterized by low rain resulting in low relative humidity. The model integrates rainfall as plant extractable soil water by subtracting surface runoff, drainage, and soil evaporation (Hoogenboom et al. 2019). As root water uptake decreases, stomatal conductance decreases through stomatal closure, which reduces carbon assimilation and transpiration as well as nutrient absorption (Moreno-Cadena et al. 2020). In addition, prior studies with the model showed that under warm temperatures water availability could be a limiting factor for cassava root yield (Moreno-Cadena et al. 2021; Phoncharoen et al. 2021a; Rankine et al. 2021). Cassava stomata are sensitive to vapor pressure deficit (VPD) and reduce gas exchange more than other crops, which can limit photosynthesis while also saving water (Oguntunde & Alatise 2007; Vongcharoen et al. 2018). The model currently incorporates a drought stress factor that integrates a VPD factor for photosynthesis. Our calibration study suggests that this stress factor might need to be adjusted. The improved cultivars from the IITA breeding program are selected for high performance including in water stress environments in the north of Nigeria. These cultivars could therefore be more tolerant than assumed by the model. A drought condition assumed by the model could still allow normal root growth for these improved genotypes. At the same time, increased rainfall led to higher predicted storage root dry weight but did not affect the observed values. The model requires defining thresholds for water and nitrogen stress, leaf and root senescence, and root water uptake (Moreno-Cadena 2021). These are currently species coefficients that were not modified by our study. Modifying these settings could improve model performance in more extreme wet or dry environments.

Finally, the IITA improved cultivars might not follow the photosynthetic assimilate allocation strategy of CSM-MANIHOT-Cassava model which feeds the storage root growth with the remaining assimilate after having satisfied the aboveground sink. The spill over model is elegant and has scientific support (Cock & Connor 2021) but is not a consensus (e.g., Gray 2000). Several studies showed that improved cassava cultivars preferentially allocate photosynthetic assimilates to storage roots after their initiation (Lahai et al. 1999; Turyagyenda et al. 2013). The partitioning of the assimilate and the dynamics of the source-sink carbon allocation during the life cycle of the plant are important phenomena to understand to better simulate cassava's growth. In a study contrasting two cassava varieties, one genotype favored stem growth while the other favored the storage roots once they were initiated (Chiewchankaset et al. 2022). Finally, a given cultivar need not be limited to a single allocation strategy but that, at a stage of its growth and according to its genotype and environmental factors, the cultivar adopts different mixtures of strategies.

Data Availability

The datasets and scripts used in the analyses presented here can be found on GitHub: https://github.com/Opamelas83/RDSSATMICH.

Conflict of Interest

The authors declare no competing interests.

Author Contributions

PO, J-LJ conceived and designed the study; PO, IR, SK Implemented field trials and methodology, PO, J-LJ and PM-C curated the data, performed data analyses, and wrote the manuscript, PO, J-LJ, PM-C and GH edited the manuscript; J-LJ and SK Provided overall coordination and leadership.

Funding

This work was supported by the Bill & Melinda Gates Foundation through the "Next Generation Cassava Breeding project" (https://www.nextgencassava.org; grant no. INV-007637) managed by Cornell University.

Acknowledgments

The authors thank the Bill & Melinda Gates Foundation for their financial support. Thanks to the staff of the International Institute of Tropical Agriculture (IITA) of the cassava breeding program that assisted for implementing the field experiments and collecting the data.

References

Adebayo, W. G. (2023). Cassava production in africa: A panel analysis of the drivers and trends. Heliyon, 9(9), e19939. https://doi.org/10.1016/j.heliyon.2023.e19939

Adiele, J. G., Schut, A. G. T., Van Den Beuken, R. P. M., Ezui, K. S., Pypers, P., Ano, A. O., Egesi, C. N., & Giller, K. E. (2021). A recalibrated and tested LINTUL-Cassava simulation model provides insight into the high yield potential of cassava under rainfed conditions. European Journal of Agronomy, 124, 126242. https://doi.org/10.1016/j.eja.2021.126242

Baguma, J. K., Mukasa, S. B., Nuwamanya, E., Alicai, T., Omongo, C., Hyde, P. T., Setter, T. L., Ochwo-Ssemakula, M., Esuma, W., Kanaabi, M., Iragaba, P., Baguma, Y., & Kawuki, R. S. (2023). Flowering and fruit-set in cassava under extended red-light photoperiod supplemented with plant-growth regulators and pruning. BMC Plant Biology, 23(1), 335. https://doi.org/10.1186/s12870-023-04349-x

Bakare, M. A., Kayondo, S. I., Aghogho, C. I., Wolfe, M. D., Parkes, E. Y., Kulakow, P., Egesi, C., Rabbi, I. Y., & Jannink, J.-L. (2022). Exploring genotype by environment interaction on cassava yield and yield related traits using classical statistical methods. PLOS ONE, 17(7), e0268189. https://doi.org/10.1371/journal.pone.0268189

Banterng, P., Hoogenboom, G., Patanothai, A., Singh, P., Wani, S. P., Pathak, P., Tongpoonpol, S., Atichart, S., Srihaban, P., Buranaviriyakul, S., Jintrawet, A., & Nguyen, T. C. (2009). Application of the Cropping System Model (CSM)-CROPGRO-Soybean for Determining Optimum Management Strategies for Soybean in Tropical Environments: Optimum Management

Strategies for Soybean Cropping. Journal of Agronomy and Crop Science, 196(3), 231–242. https://doi.org/10.1111/j.1439-037X.2009.00408.x

Basak, J. K., Ali, Muhammad A., Islam, Md. Nazrul I., & Alam, Jobair B. (2009). Assessment of the effect of climate change on boro rice production in Bangladesh using DSSAT model. 38 (2), 9.

Berton Ferreira, T., V. Shelia, C. Porter, P. Moreno Cadena, M. Salmeron Cortasa, et al. 2024. Enhancing crop model parameter estimation across computing environments: Utilizing the GLUE method and parallel computing for determining genetic coefficients. Computers and Electronics in Agriculture 227: 109513.

Beven, K., & Binley, A. (1992). The future of distributed models: Model calibration and uncertainty prediction. Hydrological Processes, 6(3), 279–298. https://doi.org/10.1002/hyp.3360060305

Boote, K. J. (2019). Advances in Crop Modeling for a Sustainable Agriculture. Burleigh Dodds Science Publishing, Cambridge, United Kingdom (Boote, K.J. (Ed.)). Burleigh Dodds Science Publishing. https://shop.bdspublishing.com/store/bds/detail/workgroup/3-190-82506.

Buddhaboon, C., A. Jintrawet, and G. Hoogenboom. 2018. Methodology to estimate rice genetic coefficients for the CSM-CERES-Rice model using GENCALC and GLUE genetic coefficient estimators. *The Journal of Agricultural Science (Cambridge)* 156(4):482-492.

Chai, Y., Pardey, P. G., & Silverstein, K. A. T. (2022). Scientific selection: A century of increasing crop varietal diversity in US wheat. Proceedings of the National Academy of Sciences, 119(51), e2210773119. https://doi.org/10.1073/pnas.2210773119

Chiaka, J. C., Zhen, L., Yunfeng, H., Xiao, Y., Muhirwa, F., & Lang, T. (2022). Smallholder Farmers Contribution to Food Production in Nigeria. Frontiers in Nutrition, 9, 916678. https://doi.org/10.3389/fnut.2022.916678

Chiewchankaset, P., Thaiprasit, J., Kalapanulak, S., Wojciechowski, T., Boonjing, P., & Saithong, T. (2022). Effective Metabolic Carbon Utilization and Shoot-to-Root Partitioning Modulate Distinctive Yield in High Yielding Cassava Variety. Frontiers in Plant Science, 13, 832304. https://doi.org/10.3389/fpls.2022.832304

Cock, J. H., Franklin, D., Sandoval, G., & Juri, P. (1979). The Ideal Cassava Plant for Maximum Yield 1. Crop Science, 19(2), 271–279. https://doi.org/10.2135/cropsci1979.0011183X001900020025x

Cock, J.H., and D.J. Connor. 2021. Cassava. In: Sadras, V.O. and Calderini, D.F., editors, Crop Physiology Case Histories for Major Crops. Elsevier. p. 588–633

Covarrubias-Pazaran, G., Gebeyehu, Z., Gemenet, D., Werner, C., Labroo, M., Sirak, S., Coaldrake, P., Rabbi, I., Kayondo, S. I., Parkes, E., Kanju, E., Mbanjo, E. G. N., Agbona, A., Kulakow, P., Quinn, M., & Debaene, J. (2022). Breeding Schemes: What Are They, How to

Formalize Them, and How to Improve Them? Frontiers in Plant Science, 12, 791859. https://doi.org/10.3389/fpls.2021.791859

de Andrade, L. R. B., Sousa, M. B. e, Wolfe, M., Jannink, J.-L., de Resende, M. D. V., Azevedo, C. F., & de Oliveira, E. J. (2022). Increasing cassava root yield: Additive-dominant genetic models for selection of parents and clones. Frontiers in Plant Science, 13, 1071156. https://doi.org/10.3389/fpls.2022.1071156

Gray, V. 2000. A comparison of two approaches for modelling cassava (*Manihot esculenta* Crantz.) crop growth. Ann. Bot. 85(1): 77–90.

Halsey, M. E., Olsen, K. M., Taylor, N. J., & Chavarriaga-Aguirre, P. (2008). Reproductive Biology of Cassava (Manihot esculenta Crantz) and Isolation of Experimental Field Trials. Crop Science, 48(1), 49–58. https://doi.org/10.2135/cropsci2007.05.0279

Han, E., A.V.M. Ines, and J. Koo. 2019. Development of a 10-km resolution global soil profile dataset for crop modeling applications. Environ. Model. Softw. 119: 70–83.

Harb, O. M., Abd El Hay, G. H., Hager, M. A., & Abou El-Enin, M. M. (2016). Calibration and Validation of DSSAT V.4.6.1, CERES and CROPGROModels for Simulating No-Tillage in Central Delta, Egypt. Agrotechnology, 05(02). https://doi.org/10.4172/2168-9881.1000143

He, J., Jones, J. W., Graham, W. D., & Dukes, M. D. (2010). Influence of likelihood function choice for estimating crop model parameters using the generalized likelihood uncertainty estimation method. Agricultural Systems, 103(5), 256–264. https://doi.org/10.1016/j.agsy.2010.01.006

He, J., Porter, C. H., Wilkens, P. W., & Jones, J. W. (2021). Guidelines for Installing and Running GLUE Program (p. 8) [Program updated]. https://dssat.net/

Hongyu, K., García-Peña, M., Araújo, L. B. de, & Santos Dias, C. T. dos. (2014). Statistical analysis of yield trials by AMMI analysis of genotype × environment interaction. Biometrical Letters, 51(2), 89–102. https://doi.org/10.2478/bile-2014-0007

Hoogenboom, G., J.W. Jones, P.C.S. Traore, and K.J. Boote. 2012. Experiments and data for model evaluation and application. *In*: p. 9-18. [J. Kihara, D. Fatondji, J.W. Jones, G. Hoogenboom, R. Tabo, and A. Bationo, editors] Improving Soil Fertility Recommendations in Africa using the Decision Support Systems for Agrotechnology Transfers (DSSAT). Springer, Dordrecht, the Netherlands.

Hoogenboom, G., Porter, C. H., Boote, K. J., Shelia, V., Wilkens, P. W., Singh, U., White, J. W., Asseng, S., Lizaso, J. I., Moreno-Cadena, L. P., Pavan, W., Ogoshi, R., Hunt, L. A., Tsuji, G. Y., & Jones, J. W. (2019). The DSSAT crop modeling ecosystem. In K. Boote, Advances in Crop Modeling for a Sustainable Agriculture. (Boote, K.J. (Ed.), pp. 173–216). Burleigh Dodds Science Publishing. https://doi.org/10.19103/AS.2019.0061.10

Hoogenboom, G., Porter, C. H., Shelia, V., Boote, K. J., Singh, U., White, J. W., Pavan, W., Oliveira, F. A. A., Moreno-Cadena, L. P., Lizaso, J. I., Asseng, S., Pequeno, D. N. L., Kimbal, B.

A., Thorp, K. R., Jones, J. W., Jones, M. R., Cuadra, S. V., Vianna, M. S., Villalobos, F. J., … Hunt, A. L. (2021). Decision Support System for Agrotechnology Transfer (DSSAT) (4.8) [Computer software]. DSSAT Foundation. https://dssat.net/

Ibrahim, J., Baguma, Y., Abincha, W., Gibson, P., & Edema, R. (2020). Flowering problems and their possible solution in cassava breeding. Journal of Scientific Agriculture, 83–89. https://doi.org/10.25081/jsa.2020.v4.6220

Ibrahim, O. M., Gaafar, A. A., Wali, A. M., Tawfik, M. M., & El-Nahas, M. M. (2016). Estimating Cultivar Coefficients of a Spring Wheat Using GenCalc and GLUE in DSSAT. Journal of Agronomy, 15(3), 130–135. https://doi.org/10.3923/ja.2016.130.135

Jennings, D. L., & Iglesias, C. (2001). Breeding for crop improvement. In R. J. Hillocks & J. M. Thresh (Eds.), Cassava: Biology, production and utilization (1st ed., pp. 149–166). CABI Publishing. https://doi.org/10.1079/9780851995243.0149

Joaqui Barandica, O., Pérez, J. C., Lenis, J. I., Calle, F., Morante, N., Pino, L., Hershey, C. H., & Ceballos, H. (2016). Cassava Breeding II: Phenotypic Correlations through the Different Stages of Selection. Frontiers in Plant Science, 7. https://doi.org/10.3389/fpls.2016.01649

Lahai, M. T., George, J. B., & Ekanayake, I. J. (1999). Cassava (Manihot esculenta Crantz) Growth Indices, Root Yield and its Components in Upland and Inland Valley Ecologies of Sierra Leone. Journal of Agronomy and Crop Science, 182(4), 239–248. https://doi.org/10.1046/j.1439-037x.1999.00299.x

Letort, V., Mahe, P., Cournede, P.-H., de Reffye, P., & Courtois, B. (2007). Quantitative Genetics and Functional-Structural Plant Growth Models: Simulation of Quantitative Trait Loci Detection for Model Parameters and Application to Potential Yield Optimization. Annals of Botany, 101(8), 1243–1254. https://doi.org/10.1093/aob/mcm197

Li, Z., He, J., Xu, X., Jin, X., Huang, W., Clark, B., Yang, G., & Li, Z. (2018). Estimating genetic parameters of DSSAT-CERES model with the GLUE method for winter wheat (Triticum aestivum L.) production. Computers and Electronics in Agriculture, 154, 213–221. https://doi.org/10.1016/j.compag.2018.09.009

Loague, K., & Green, R. E. (1991). Statistical and graphical methods for evaluating solute transport models: Overview and application. Journal of Contaminant Hydrology, 7(1–2), 51–73. https://doi.org/10.1016/0169-7722(91)90038-3

Mereu, V., Gallo, A., & Spano, D. (2019). Optimizing Genetic Parameters of CSM-CERES Wheat and CSM-CERES Maize for Durum Wheat, Common Wheat, and Maize in Italy. Agronomy, 9(10), 665. https://doi.org/10.3390/agronomy9100665

Messina, C. D., Technow, F., Tang, T., Totir, R., Gho, C., & Cooper, M. (2017). Leveraging biological insight and environmental variation to improve phenotypic prediction: Integrating crop growth models (CGM) with whole genome prediction (WGP) [Preprint]. Genetics. https://doi.org/10.1101/100057

Moreno-Cadena, L. P., Hoogenboom, G., Cock, J. H., Ramirez-Villegas, J., Pypers, P., Kreye, C., Tariku, M., Ezui, K. S., Becerra Lopez-Lavalle, L. A., & Asseng, S. (2021). Modeling growth, development and yield of cassava: A review. Field Crops Research, 267, 108140. https://doi.org/10.1016/j.fcr.2021.108140

Moreno-Cadena, L. P., Hoogenboom, G., Fisher, M. J., Ramirez-Villegas, J., Prager, S. D., Becerra Lopez-Lavalle, L. A., Pypers, P., Mejia de Tafur, M. S., Wallach, D., Muñoz-Carpena, R., & Asseng, S. (2020). Importance of genetic parameters and uncertainty of MANIHOT, a new mechanistic cassava simulation model. European Journal of Agronomy, 115, 126031. https://doi.org/10.1016/j.eja.2020.126031

nextgencassava.org. (2022, March 6). https://www.nextgencassava.org/who-we-are/nextgen-fellows/

Nowosad, K., Liersch, A., Popławska, W., & Bocianowski, J. (2016). Genotype by environment interaction for seed yield in rapeseed (Brassica napus L.) using additive main effects and multiplicative interaction model. Euphytica 208(1), 187–194. https://doi.org/10.1007/s10681-015-1620-z

Oguntunde, P. G., & Alatise, M. O. (2007). Environmental regulation and modelling of cassava canopy conductance under drying root-zone soil water. Meteorological Applications, 14(3), 245–252. https://doi.org/10.1002/met.24

Oliveira, L. A., da Silva, C. P., da Silva, A. Q., Mendes, C. T. E., Nuvunga, J. J., Nunes, J. A. R., Parrella, R. A. da C., Baleste, M., & Filho, J. S. de S. B. (2022). Bayesian GGE model for heteroscedastic multienvironmental trials. Crop Science, 62(3), 982–996. https://doi.org/10.1002/csc2.20696

Phoncharoen, P., Banterng, P., Vorasoot, N., Jogloy, S., Theerakulpisut, P., & Hoogenboom, G. (2021). Identifying Suitable Genotypes for Different Cassava Production Environments—A Modeling Approach. Agronomy, 11(7), 1372. https://doi.org/10.3390/agronomy11071372

Phoncharoen, P., Poramate, B., Moreno-Cadena, L. P., Vorasoot, N., Jogloy, S., Theerakulpisut, P., & Hoogenboom, G. (2021). Performance of the CSM–MANIHOT–Cassava model for simulating planting date response of cassava genotypes. Field Crops Research, 264, 108073. https://doi.org/10.1016/j.fcr.2021.108073

Putto, W., A. Patanothai, S. Jogloy, and G. Hoogenboom. 2008. Determination of mega-environments for peanut breeding using the CSM-CROPGRO-Peanut model. *Crop Science* 48(3):973-982.

R Core Team. (2023). R: A language and environment for statistical computing (Vienna, Austria; 4.3.0) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/.

Rabbi, I. Y., Hamblin, M. T., Kumar, P. L., Gedil, M. A., Ikpan, A. S., Jannink, J.-L., & Kulakow, P. A. (2014). High-resolution mapping of resistance to cassava mosaic geminiviruses in cassava using genotyping-by-sequencing and its implications for breeding. Virus Research, 186, 87–96. https://doi.org/10.1016/j.virusres.2013.12.028

Rabbi, I. Y., Kayondo, S. I., Bauchet, G., Yusuf, M., Aghogho, C. I., Ogunpaimo, K., Uwugiaren, R., Smith, I. A., Peteti, P., Agbona, A., Parkes, E., Lydia, E., Wolfe, M., Jannink, J.-L., Egesi, C., & Kulakow, P. (2022). Genome-wide association analysis reveals new insights into the genetic architecture of defensive, agro-morphological and quality-related traits in cassava. Plant Molecular Biology, 109(3), 195–213. https://doi.org/10.1007/s11103-020-01038-3

Rabbi, I. Y., Udoh, L. I., Wolfe, M., Parkes, E. Y., Gedil, M. A., Dixon, A., Ramu, P., Jannink, J., & Kulakow, P. (2017). Genome-Wide Association Mapping of Correlated Traits in Cassava: Dry Matter and Total Carotenoid Content. The Plant Genome, 10(3), plantgenome2016.09.0094. https://doi.org/10.3835/plantgenome2016.09.0094

Ramirez-Villegas, J., Molero Milan, A., Alexandrov, N., Asseng, S., Challinor, A. J., Crossa, J., Van Eeuwijk, F., Ghanem, M. E., Grenier, C., Heinemann, A. B., Wang, J., Juliana, P., Kehel, Z., Kholova, J., Koo, J., Pequeno, D., Quiroz, R., Rebolledo, M. C., Sukumaran, S., … Reynolds, M. (2020). CGIAR modeling approaches for resource-constrained scenarios: I. Accelerating crop breeding for a changing climate. Crop Science, 60(2), 547–567. https://doi.org/10.1002/csc2.20048

Rankine, D., Cohen, J., Murray, F., Moreno-Cadena, L. P., Hoogenboom, G., Campbell, J., Taylor, M., & Stephenson, T. (2021). Evaluation of DSSAT-MANIHOT-Cassava model to determine potential irrigation benefits for cassava in Jamaica. Agronomy Journal, 113(6), 5317–5334. https://doi.org/10.1002/agj2.20876

Shackelford, G. E., Haddaway, N. R., Usieta, H. O., Pypers, P., Petrovan, S. O., & Sutherland, W. J. (2018). Cassava farming practices and their agricultural and environmental impacts: A systematic map protocol. Environmental Evidence, 7(1), 30. https://doi.org/10.1186/s13750-018-0142-2

Streck, N. A., Pinheiro, D. G., Junior Zanon, A., Gabriel, L. F., Rocha, T. S. M., Souza, A. T. de, & Silva, M. R. da. (2014). Efeito do espaçamento de plantio no crescimento, desenvolvimento e produtividade da mandioca em ambiente subtropical. Bragantia, 73(4), 407–415. https://doi.org/10.1590/1678-4499.0159

Suriharn, B., Patanothai, A., Pannangpetch, K., Jogloy, S., & Hoogenboom, G. (2007). Determination of Cultivar Coefficients of Peanut Lines for Breeding Applications of the CSM-CROPGRO-Peanut Model. Crop Science, 47(2), 607–619. https://doi.org/10.2135/cropsci2006.01.0050

Tovihoudji, P. G., Akponikpè, P. B. I., Agbossou, E. K., & Bielders, C. L. (2019). Using the DSSAT Model to Support Decision Making Regarding Fertilizer Microdosing for Maize Production in the Sub-humid Region of Benin. Frontiers in Environmental Science, 7, 13. https://doi.org/10.3389/fenvs.2019.00013

Tsuji, G. Y., Hoogenboom, G., & Thornton, P. K. (Eds.). (1998). Understanding Options for Agricultural Production (Vol. 7). Springer Netherlands. https://doi.org/10.1007/978-94-017-3624-4

Turyagyenda, L. F., Kizito, E. B., Ferguson, M., Baguma, Y., Agaba, M., Harvey, J. J. W., & Osiru, D. S. O. (2013). Physiological and molecular characterization of drought responses and identification of candidate tolerance genes in cassava. AoB Plants, 5(0), plt007–plt007. https://doi.org/10.1093/aobpla/plt007

Tyagi, S., Singh, N., Sonkar, G., & Mall, R. K. (2019). Sensitivity of evapotranspiration to climate change using DSSAT model in sub humid climate region of Eastern Uttar Pradesh. Modeling Earth Systems and Environment, 5(1), 1–11. https://doi.org/10.1007/s40808-018-0513-2

van Rossum, B.-J., van Eeuwijk, F., Boer, M., Malosetti, M., Bustos-Korts, D., Millet, E. J., Paulo, J., Verouden, M., Kruijer, W., & Wehrens, R. (2022). Package 'statgenGxE' CRAN. R Package Version 4.0.1. ttps://CRAN.R-project.org/package=statgenGxE

Vongcharoen, K., Santanoo, S., Banterng, P., Jogloy, S., Vorasoot, N., & Theerakulpisut, P. (2018). Seasonal variation in photosynthesis performance of cassava at two different growth stages under irrigated and rain-fed conditions in a tropical savanna climate. Photosynthetica, 56(4), 1398–1413. https://doi.org/10.1007/s11099-018-0849-x

Wallach, D., & Goffinet, B. (1987). Mean Squared Error of Prediction in Models for Studying Ecological and Agronomic Systems. Biometrics, 43(3), 561. https://doi.org/10.2307/2531995

Wallach, D., Martre, P., Liu, B., Asseng, S., Ewert, F., Thorburn, P. J., Ittersum, M., Aggarwal, P. K., Ahmed, M., Basso, B., Biernath, C., Cammarano, D., Challinor, A. J., De Sanctis, G., Dumont, B., Eyshi Rezaei, E., Fereres, E., Fitzgerald, G. J., Gao, Y., … Zhang, Z. (2018). Multimodel ensembles improve predictions of crop–environment–management interactions. Global Change Biology, 24(11), 5072–5083. https://doi.org/10.1111/gcb.14411

Wellens, J., D. Raes, E. Fereres, J. Diels, C. Coppye, et al. 2022. Calibration and validation of the FAO AquaCrop water productivity model for cassava (Manihot esculenta Crantz). Agric. Water Manag. 263(107491): 107491.

White, J. W., and G. Hoogenboom. 1996. Simulating effects of genes for physiological traits in a process-oriented crop model. *Agronomy Journal* 88(3):416-422.

Wholey, D. W., & Booth, R. H. (1979). A comparison of simple methods for estimating starch content of cassava roots. Journal of the Science of Food and Agriculture, 30(2), 158–164. https://doi.org/10.1002/jsfa.2740300210

Willmott, C. J. (1982). Some Comments on the Evaluation of Model Performance. Bulletin of the American Meteorological Society, 63(11), 1309–1313. https://doi.org/10.1175/1520-0477(1982)063<1309:SCOTEO>2.0.CO;2

Wolfe, M. D., Chan, A. W., Kulakow, P., Rabbi, I., & Jannink, J.-L. (2021). Genomic mating in outbred species: Predicting cross usefulness with additive and total genetic covariance matrices. Genetics, 219(3), iyab122. https://doi.org/10.1093/genetics/iyab122

Wolfe, M. D., Rabbi, I. Y., Egesi, C., Hamblin, M., Kawuki, R., Kulakow, P., Lozano, R., Carpio, D. P. D., Ramu, P., & Jannink, J. (2016). Genome-Wide Association and Prediction Reveals

Genetic Architecture of Cassava Mosaic Disease Resistance and Prospects for Rapid Genetic Improvement. The Plant Genome, 9(2), plantgenome2015.11.0118. https://doi.org/10.3835/plantgenome2015.11.0118

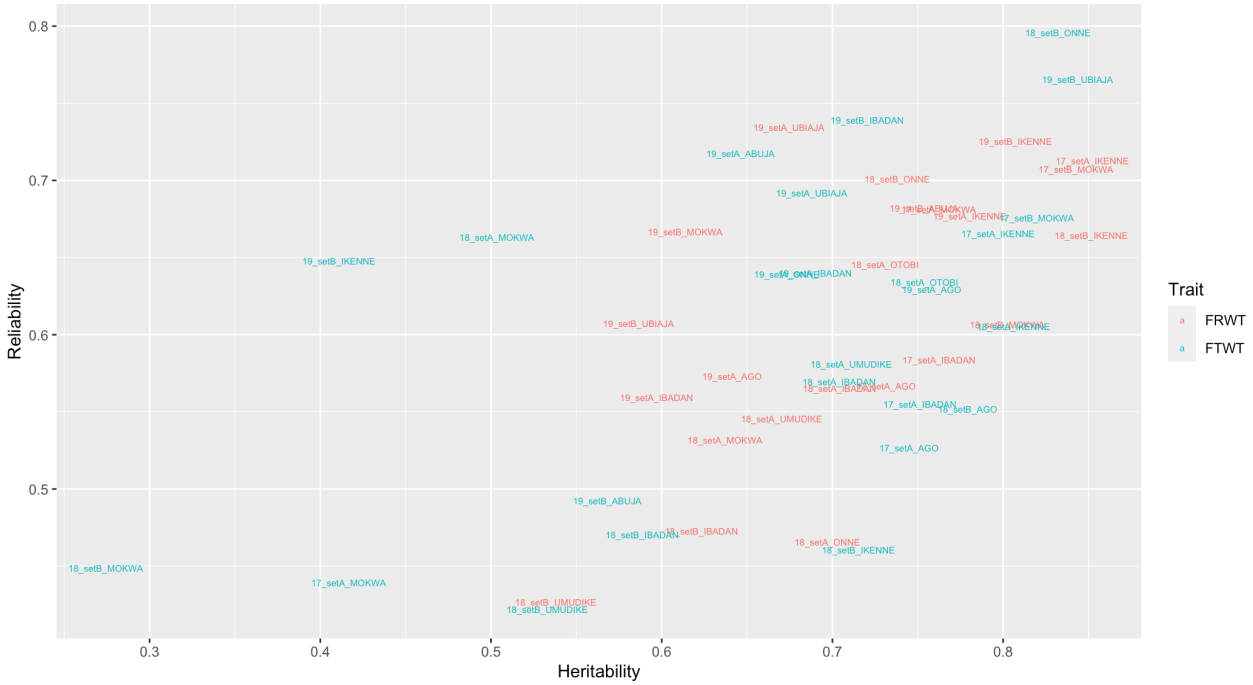
Yan, W., & Rajcan, I. (2002). Biplot Analysis of Test Sites and Trait Relations of Soybean in Ontario. Crop Science, 42(1), 11–20. https://doi.org/10.2135/cropsci2002.1100

Yu, Y., Cheng, Q., Wang, F., Zhu, Y., Shang, X., Jones, A., He, H., & Song, Y. (2023). Crop/Plant Modeling Supports Plant Breeding: I. Optimization of Environmental Factors in Accelerating Crop Growth and Development for Speed Breeding. Plant Phenomics, 5, 0099. https://doi.org/10.34133/plantphenomics.0099

Evaluating a Cassava Crop Growth Model by Optimizing Genotypic-Specific Parameters Using Multi-environment Trial Breeding Data

Pamelas M. Okoma, Siraj S. Kayondo, Ismail Y. Rabbi, Patricia L. Moreno-Cadena, Gerrit Hoogenboom and Jean-Luc Jannink

Supplementary Data



Supplementary Figure 1: Plot of the reliability and heritability of the single trials used for the study.

FRWT= Fresh storage root weight; FTWT= Fresh above ground biomass weight

Supplementary Table 1: Physical and chemical soil properties in 8 locations

ID / Location	Depth (cm)	SLLL	SDUL	SSAT	SRGF	SSKS	SBDM	SLOC	SLCL	SLSI	SLNI	SLHW	SCEC
NG04365455 / UMUDIKE	5	0.164	0.278	0.403	1	0.51	1.29	2.1	27.22	26.27	0.12	5.21	17.6
	15	0.176	0.292	0.407	0.85	0.41	1.31	1.78	29.25	25.35	0.09	5.28	15.4
	30	0.192	0.309	0.414	0.7	0.3	1.34	1.35	31.99	24.06	0.07	5.38	15
	60	0.207	0.326	0.421	0.5	0.22	1.39	0.86	34.56	22.92	0.06	5.49	15.6
	100	0.206	0.324	0.419	0.38	0.23	1.45	0.51	34.44	22.26	0.05	5.63	15.7
	200	0.196	0.311	0.414	0.05	0.28	1.5	0.28	32.65	22.12	0.05	5.81	15.7
NG04408648 / ONNE	5	0.166	0.27	0.399	1	0.54	1.24	2.18	27.59	19.8	0.12	4.41	10.7
	15	0.177	0.283	0.403	0.85	0.43	1.26	1.85	29.54	19.12	0.09	4.48	9.5
	30	0.194	0.302	0.409	0.7	0.3	1.29	1.4	32.33	18.03	0.07	4.57	9.2
	60	0.209	0.319	0.415	0.5	0.22	1.34	0.9	34.83	16.91	0.06	4.69	9.6
	100	0.208	0.318	0.414	0.38	0.22	1.4	0.52	34.68	16.41	0.05	4.82	9.6
	200	0.198	0.305	0.408	0.05	0.28	1.46	0.29	33.03	16.13	0.05	5	9.6
NG04309279 / UBIAJA	5	0.154	0.264	0.399	1	0.63	1.23	2.79	25.52	25.17	0.12	5.82	25.4
	15	0.165	0.277	0.402	0.85	0.51	1.25	2.37	27.49	24.26	0.09	5.9	22.3
	30	0.181	0.295	0.408	0.7	0.37	1.28	1.8	30.23	23.1	0.07	5.99	21.6
	60	0.196	0.311	0.414	0.5	0.28	1.33	1.16	32.71	21.97	0.06	6.1	22.6
	100	0.196	0.31	0.413	0.38	0.28	1.39	0.68	32.75	21.32	0.05	6.24	22.7
	200	0.186	0.297	0.408	0.05	0.35	1.45	0.37	30.96	20.88	0.05	6.42	22.7
NG04340916 / IKENNE	5	0.195	0.312	0.415	1	0.28	1.4	2.47	32.57	22.94	0.12	5.1	15.6
	15	0.208	0.326	0.42	0.85	0.22	1.42	2.09	34.67	22.25	0.09	5.17	13.6
	30	0.225	0.344	0.428	0.7	0.15	1.45	1.59	37.55	21.25	0.07	5.26	13.2
	60	0.239	0.359	0.436	0.5	0.11	1.5	1.02	39.97	20.14	0.06	5.38	13.8
	100	0.239	0.359	0.435	0.38	0.11	1.56	0.59	40.08	19.45	0.05	5.52	13.8
	200	0.228	0.346	0.428	0.05	0.14	1.62	0.33	38.08	18.86	0.05	5.7	13.8
NG04283331 / AGO	5	0.136	0.237	0.391	1	0.94	1.31	2.43	22.72	22.25	0.12	4.9	9.9
	15	0.148	0.25	0.394	0.85	0.75	1.33	2.07	24.69	21.45	0.09	4.97	8.6
	30	0.164	0.269	0.398	0.7	0.55	1.36	1.57	27.4	20.29	0.07	5.06	8.4
	60	0.179	0.285	0.403	0.5	0.42	1.41	1.01	29.78	19.06	0.06	5.18	8.8
	100	0.178	0.283	0.402	0.38	0.42	1.47	0.59	29.67	18.58	0.05	5.32	8.8
	200	0.167	0.268	0.397	0.05	0.54	1.52	0.33	27.79	18.32	0.05	5.5	8.8
NG04261714 / IBADAN	5	0.106	0.187	0.384	1	1.88	1.35	1.23	17.82	16.66	0.12	5.34	9.1
	15	0.116	0.2	0.385	0.85	1.55	1.37	1.04	19.56	16.1	0.09	5.41	8
	30	0.131	0.218	0.387	0.7	1.17	1.4	0.8	22.02	15.51	0.07	5.51	7.9
	60	0.146	0.235	0.389	0.5	0.89	1.45	0.51	24.37	14.59	0.06	5.62	8
	100	0.144	0.232	0.388	0	0.93	1.51	0.3	24.1	14.07	0	5.75	8.1
	200	0.134	0.218	0.386	0	1.16	1.57	0.19	22.46	13.61	0	5.94	8.1
NG04171020 / MOKWA	5	0.198	0.312	0.413	1	0.27	1.3	1.64	33.07	20.59	0.12	5.82	17.3
	15	0.21	0.325	0.419	0.85	0.21	1.32	1.39	35.11	19.91	0.09	5.89	15.1
	30	0.225	0.342	0.426	0.7	0.15	1.35	1.06	37.65	19	0.07	5.99	14.7
	60	0.239	0.357	0.433	0.5	0.11	1.41	0.68	39.98	17.99	0.06	6.1	15.4
	100	0.238	0.356	0.432	0.38	0.11	1.47	0.39	39.91	17.27	0.05	6.24	15.5
	200	0.229	0.344	0.426	0.05	0.14	1.52	0.21	38.32	16.87	0.05	6.41	15.5
NG04184010 / ABUJA	5	0.137	0.23	0.389	1	1.01	1.21	2.19	22.94	17.15	0.12	5.55	14.1
	15	0.149	0.243	0.391	0.85	0.81	1.22	1.86	24.85	16.45	0.09	5.61	12.3
	30	0.165	0.262	0.395	0.7	0.59	1.26	1.42	27.49	15.64	0.07	5.7	12
	60	0.179	0.279	0.4	0.5	0.43	1.31	0.91	29.95	14.8	0.06	5.82	12.4
	100	0.178	0.277	0.399	0	0.44	1.37	0.52	29.78	14.23	0	5.96	12.5
	200	0.169	0.264	0.395	0	0.55	1.42	0.31	28.18	13.73	0	6.14	12.5

SLLL: Lower limit of plant extractable soil water (cm^3)
SDUL: Drained upper limit (cm^3)
SSAT: Upper limit, saturated (cm^3)
SRGF: Soil Root growth factor
SSKS: Saturated hydraulic conductivity (cm/h)
SBDM: Bulk density (g/cm^3)
SLOC: Organic carbon (%)
SLCL: Clay (%)
SLSI: Silt (%)
SLNI: Total nitrogen (%)
SLHW: pH in water
SCEC: Cation exchange capacity (cmol/kg)

Supplementary Methods

The Generalized Likelihood Uncertainty Estimation (GLUE) program as described by Beven & Binley (1992), He et al. (2010), and updated by He et al. (2021).

a. Develop prior parameter distributions

Random sampling, based on the prior distribution of parameters, is generally implemented in GLUE. The ranges of the parameters were defined and stored in a file formatted for the GLUE program under DSSAT 4.8.0.

b. Generate random parameter sets from the prior parameter distributions

The parameters sets were generated using a Monte Carlo sampling of distributions implemented in a MATLAB program. From the point of view of Monte Carlo sampling in the GLUE method, more parameter sets lead to more stable results (Li et al., 2018). Thus, a total of 10,000 (the minimum recommended is 3,000) random parameter sets were generated from the prior distribution.

c. Run the MANIHOT-Cassava model

The MANIHOT-Cassava model was run with each parameter set generated above by introducing the set into the standard genetic input file (cultivars file). For each run, the program identified all experiments in the DSSAT cassava data files involving the accession selected and these were used in the coefficient estimation process. The crop growth simulation outputs of dry aboveground biomass for each parameter set were tabulated for use in the GLUE likelihood calculations.

d. Calculate the likelihood values and probability

The observed data on dry aboveground biomass were used along with the corresponding simulated outputs to compute the likelihood value. The Likelihood Function is the product of these individual likelihood values. The Gaussian likelihood function used in this study was recommended by He *et al.* (2010):

$$L[\theta_i|O] = \Pi_{j=1}^{M} \frac{1}{\sqrt{2\pi\sigma_o^2}} \exp\left[\frac{(O_j - P_j(\theta_i))^2}{2\sigma_o^2}\right] \qquad \text{Eq. 4}$$

$L[\theta_i|O]$ = likelihood value of parameter set i, given the set of observations
θ_i = ith parameter set (I = 1, 2, 3, ...N). N = 10,000 in our case.
O_j = jth observation
$P_j(\theta_i)$ = model output for the jth environment under parameter set θ_i
σ_o^2 = variance of model error
M = number of observations.
The posterior probability $p(\theta_i)$, of each parameter set θ_i is computed with the Bayesian equation as follows:

$$p(\theta_i) = \frac{L(\theta i|O)}{\sum_{i=1}^{N} L(\theta i|O)} \qquad \text{Eq. 5}$$

e. Construct posterior distribution and statistics

The pair of parameter sets and probabilities [θ_i, $p(\theta_i)$] were used to compute the mean, and the variance of each parameter θ_i:

$$u_{post}(\theta) = \sum_{i=1}^{N} p(\theta_i) \,.\, \theta_i \qquad \text{Eq. 6}$$

$$\sigma_{post}^2(\theta) = \sum_{i=1}^{N} p(\theta_i) \,.\, \left(\theta_i - u_{post}(\theta)\right)^2 \qquad \text{Eq. 7}$$

$u_{post}(\theta_i)$ = mean of the posterior distribution
σ_{post}^2 = variance of the posterior distribution