Computers and Electronics in Agriculture 235 (2025) 110383

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

In-silico optimization of peanut production in India through envirotyping and ideotyping

Amir Hajjarpoor [a,1,*], Jan Pavlík [b,1], Jan Hora [c], Jakub Konopásek [b], Janila Pusupuleti [d], Vincent Vadez [d,e,f], Afshin Soltani [g], Til Feike [a], Michal Stočes [b], Jan Jarolímek [b], Jana Kholová [b,d,*]

[a] Julius Kühn Institute (JKI) – Federal Research Centre for Cultivated Plants, Institute for Strategies and Technology Assessment, Kleinmachnow, Germany
[b] Department of Information Technologies, Faculty of Economics and Management, Czech University of Life Sciences Prague, Czech Republic
[c] Department of Mathematics, Faculty of Engineering, Czech University of Life Sciences Prague, Czech Republic
[d] International Crops Research Institute for Semi-Arid Tropics (ICRISAT), Patancheru, India
[e] French National Research Institute for Sustainable Development (IRD), UMR DIADE, University of Montpellier, Montpellier, France
[f] Centre d'Etudes Régional pour l'Amélioration de l'Adaptation à la Sécheresse (CERAAS), Thiès, Senegal
[g] Agronomy Group, Gorgan University of Agricultural Sciences and Natural Resources, Gorgan, Iran

ARTICLE INFO

Keywords:
Crop Modeling
High Performance Computing (HPC)
Genotype × Environment × Management (G×E×M) interaction
Index of Goodness (IoG)
Environmental Clustering (EC)

ABSTRACT

Peanut (*Arachis hypogaea* L.) is an important cash crop with significant yield gaps, especially in developing countries. Optimizing peanut production could foster economic growth for a significant number of smallholder farmers across the globe. In this study, we used an *in-silico* cropping system model to simulate and optimize genotype × crop management (G × M) across India that would narrow the existing peanut yield gaps. For that, we simulated diverse G × M combinations across range of environments (E) in India, considering three irrigation regimes typical for managing peanut production systems. Covering whole India in a 0.5°×0.5° resolution, we simulated 60,480 G × M combinations for each grid, summing up to a total of 2.3 billion simulations and 1.02 TB output data. This required well-structured high-performance computing (HPC) approaches, data management, and analytical capacities. For this, we present the concept of a re-usable HPC system with interoperable modules, which can be readily adapted for different simulation setups. We introduced the novel way of analyzing simulation outputs – "Index of Goodness" (IoG) – that aggregates key peanut production characteristics (grain and haulm production) and production risk failure. IoG is a simple way to evaluate the suitability of simulated GxM options from the perspective of end-users, including primary producers and crop improvement programs. The generated output was used to identify the geographic regions (environmental clusters, EC) with high degree of similarities within each of the tested irrigation regimes. For each cluster, we identified a specific suite of GxM to benefit peanut production and prioritize G targets for breeding. In principle, irrigated cropping systems would benefit from high planting densities, long duration and vigorous crop types. With diminishing water availability (particularly in the Thar Desert and SE India), the optimal production included shorter duration crop types which could quickly respond to drought stimuli (i.e. close stomata and conserve soil water upon soil and atmospheric drought exposure). These traits should also be considered in phenotyping strategies to support context-specific breeding.

1. Introduction

Sustainable food production for a growing population is one of the most pressing global challenges (Sustainable Development Goals, SDGs #1, 2). The latest United Nations report (UN, 2024) highlighted that current actions are insufficient to meet the SDGs, requiring significant

* Corresponding authors at: Julius Kühn-Institute (JKI) – Federal Research Centre for Cultivated Plants, Institute for Strategies and Technology Assessment, Kleinmachnow, Germany (A. Hajjarpoor); Department of Information Technologies, Faculty of Economics and Management, Czech University of Life Sciences Prague, Czech Republic (J. Kholová).

E-mail addresses: Amir.Hajjarpoor@julius-kuehn.de (A. Hajjarpoor), kholova@pef.czu.cz (J. Kholová).

[1] Co-first authors.

https://doi.org/10.1016/j.compag.2025.110383
Received 2 August 2024; Received in revised form 31 January 2025; Accepted 3 April 2025
Available online 24 April 2025

investment and acceleration. The socially and environmentally vulnerable agricultural regions, like the Semi-Arid Tropics (SAT), face escalating climate variability, which is likely to further destabilize these systems in the future. Peanut crop is an important oilseed, confectionery, and livestock feed, powering the income of many small-scale farmers in these SAT regions of India and Africa. Peanut cultivation in India makes up to two-thirds of the total rainfed cropland, i.e., more than 5 million hectares (DLD, 2017; Hajjarpoor et al., 2021). The dependence of the rural population well-being on peanut production is evident (Nautiyal and Mejia, 2002) and crucial to the agricultural and socio-economic development of India (e.g., Domhoefer et al., 2022). However, peanut productivity remains much below its potential (on average, including irrigated areas ~1.1 t ha^{-1}; DLD, 2017) with high inter-seasonal variability. To sustain production in SAT Agrosystem, location and climate-optimized solutions are required that would reduce the effect of Genotype by Environment by Management (GxExM) interactions within the target production environment (TPE, Hajjarpoor et al., 2022).

Accelerating analysis, quantification and solutions design to deal with GxExM across spatio-temporal scales requires to transcend traditional approaches that evaluate G (i.e., plant breeding) and M (i.e., agronomy) separately (Cooper et al., 2020; George, 2014; Hammer et al., 2014) or only very limited GxM combinations (e.g. multi-environment trials; METs, Chenu, 2015; Laidig et al., 2024; Ramirez-Villegas et al., 2020; Riedesel et al., 2024; Witcombe et al., 1998). The latter eventually limits the scale of spatio-temporal testing of GxExM options and, consequently, the selection of optimal GxM interventions designed for specific E (Kholová et al., 2020).

Currently, process-based Crop Simulation Models (CSM) play a crucial role in cropping systems design by enabling virtual experimentation (Cooper et al., 2023; Hammer et al., 2014). These models sufficiently represent the main biophysical agrosystem processes and facilitate the design and prioritization of potential solutions to deal with GxExM interactions *in-silico* and at the required scale (Battisti and Sentelhas, 2019; Carcedo et al., 2022; Cooper et al., 2020; Hajjarpoor et al., 2022, 2021; Harrison et al., 2014; Heinemann et al., 2024, 2015; Ramirez-Villegas et al., 2020; Shawon et al., 2024; Sinclair et al., 2020). By simulating potential GxExM, CSMs can handle non-experienced scenarios (Hajjarpoor et al., 2022). This facilitates the prioritization of GxExM options *in-silico* with a higher probability of being superior *in-vivo* to speed up the development of actual solutions (Heinemann et al., 2019).

While the models are used to optimize crop's agronomy with several commercial solutions available (e.g., Yield Prophet,[2] Ypsilon,[3] Cropwin,[4] etc), use of models for development of resilient crop varieties through physiology-led breeding and support climate-smart crop production is just emerging (Chenu et al., 2018, 2011; Welcker et al., 2022; Wu et al., 2019, 2016; Zheng et al., 2018). Outputs from CSMs have also been employed for environmental characterizations, helping breeders refine cultivar testing schemes and identify optimal seed production locations for commercial use (Heinemann et al., 2024). Much of these works, however, focus on major cereal crops while little has been done on legume crops (Hajjarpoor et al., 2018; Kaloki et al., 2019; Justino et al., 2025) and particularly peanut (Hajjarpoor et al., 2021). It is important to note that the practical use of CSMs for high-resolution large-scale simulations poses challenges in data management, computational resources and performance optimization. Some of these advanced computational systems has been described in different studies (Fainges, 2015; Jang et al., 2019; Khabarov et al., 2020; Zheng et al., 2016).

A previous modeling study of Indian peanut systems revealed substantial yield gaps (Hajjarpoor et al., 2021) and identified production regions with higher levels of homogeneity and reduced GxExM effects compared to the conventionally recognized peanut production zones. Therefore, these regions could be the effective geographical targets for specific genetic and agronomic designs that would optimize peanut production. Consequently, here we aim to identify GxM combinations with high potential to enhance/stabilize peanut production across India. In particular, we utilize a grid-based application of the Simple Simulation Model (SSM, Hajjarpoor et al., 2021) to examine the combinations of (i) a range of plant characters ("virtual genotypes" or "ideotypes", G) which could be prioritized for breeding region-specific cultivars and ii) examine the crop management practices (M) which could be combined with these peanut ideotypes to optimize the production in these specific regions.

2. Methodology

2.1. Overview

Fig. 1 visualizes the methodological approaches and workflow. To prioritize crop ideotypes for peanut production environments in India, we created 320 virtual genotypes with contrasting but realistic ranges of crop characters (G: phenology, vigor, and crop responsiveness to soil and atmospheric drought). To optimize the production practices for these virtual genotypes, we analyzed these in the context of agronomy practices feasible in the target production regions in India (M: sowing dates, planting density and irrigation levels across relevant soil types). We conducted a genotype (G) by environment (E) by management (M) factorial system analysis using gridded-based weather data distributed across India. We used a simple simulation model (SSM, Hajjarpoor et al., 2021; Soltani and Sinclair, 2012a), to simulate potential GxM combinations relevant to peanut cropping systems in India (60480 factorial GxM combinations spanning across 40 years within each of 1173 grids resulting in over 2.3 billion simulations;1.02 TB of raw output data). The goodness of the GxExM simulation outputs for the intended end-use was evaluated using Index of Goodness (IoG). The simulations attaining the best IoG were clustered and visualized on the map and the simulation parameters were analyzed within each cluster.

2.2. Data used for model input

The SSM model (2.3) requires crop parameters (G: coefficients defining the crop growth and development functions (2.2.1)), the input to define the environmental context (E; meteorological parameters (2.2.2) and soil (2.2.3)), and crop management practices (M; sowing date, irrigation, fertilization, etc. (2.2.4)):

2.2.1. Crop characteristics (G component)

To capture the biological variability of peanut crop into the crop model, we used combinations of crop model coefficients (G) reflecting observed ranges of phenological development, different rates of above- and below-ground organs development (biomass accumulation, canopy and root growth) and organs functionalities, (stomata closure upon soil and atmospheric drying). These G combinations resulted in virtual crops with variable maturity, vigor, and different capacities to conduct water (details in Table 1). Altogether, we created 320 virtual cultivars spanning the documented ranges of biological variability (Devi et al., 2009; Halilou et al., 2016; Sivasakthi et al., 2018; Vadez et al., 2012).

2.2.2. Meteorological information inputs (E component)

India has a general lack of quality weather information and open-source databases (Hajjarpoor et al., 2021, 2018; Kholová et al., 2013; Ronanki et al., 2022). Therefore, enhancing the spatio-temporal resolution of relevant weather information was necessary for our study. In our previous studies, several sets of gridded weather data (GWD) were tested for their suitability to simulate chickpea (Hajjarpoor et al., 2018)

[2] https://www.yieldprophet.com.au/yp/Home.aspx.

[3] https://ypsilon.services/.

[4] https://www.itk.fr/en/cropwin/.

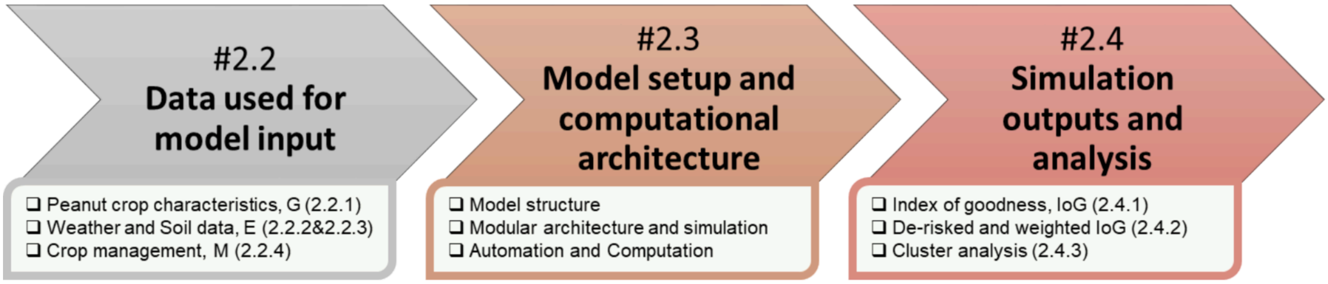

Fig. 1. Process flow used in the methodology of presented work.

and peanut (Hajjarpoor et al., 2021) production systems in India (i.e., IMD (Indian meteorological department), including AgMERRA (both 0.5°x0.5° and 1°x1° resolution, Ruane et al., 2015)), NASAPOWER[5] and MarkSim (Jones et al., 2002; Jones and Thornton, 2000)) by comparing them to available observed weather data (IMD[6]). Using the same approach we found the most relevant data source for our study was AgMERRA 0.5°x0.5°, which, however, contains the weather information only upto in 2010. Therefore, our study combined 31 years of AgMERRA data (1980–2010) with 9 years of NASAPOWER data (2011–2019, the next "best" weather data source) for the modeling analysis (all data are available at Pavlik et al., 2025).

2.2.3. Soil information (E component)

Gridded soil data at a spatial resolution of 250 m are available for India.[7] However, these data only include generic soil profiles and do not adequately represent peanut production conditions. Given the scale of our study (entire Indian subcontinent) and size of the grids (0.5°x0.5°), where each grid can contain different soil types, we decided to simulate all possible combinations of soil types and effective rooting depths typical for peanut cultivation. In this way, the simulated results of all other GxM combinations can be further contextualized for local soil conditions based on local experts' and farmers' knowledge. To reflect the soil heterogeneity of peanut cultivation systems in India, we referred to literature and reconstructed the parameters of all relevant soil types for peanut production systems in India, using sources such as the Handbook of Agriculture (Trivedi, 2011), ISRIC, NBSS-LUP, and expert knowledge from NBPGR and ICRISAT (details in Hajjarpoor et al, 2021). We used combinations ofsoil textures and effective rooting depths, resulting in six different soil types (details in Table 2). We simulated all soil types in each grid to gain a comprehensive view of the soil's effect.

2.2.4. Crop management information (M component)

In India, peanut is produced in all three main cropping seasons; rainy ("Kharif", sowing ~ May-September, ~82 % of total production), post-rainy ("Rabi", sowing ~ October-January, ~10 % of total production) and summer (sowing ~ January-February, ~8% of total production). Therefore, in the current study, we focused on the major season – Kharif.

Since sowing window in Kharif season can vary from May to September across India (Hajjarpoor et al., 2021; Trivedi, 2011), a range of sowing windows was set accordingly from mid of May to end of August with two-week intervals. The criteria for sowing was inspired by the regular management practices and was set to be initiated when a moisture content would likely support successful germination, i.e., when a minimum of 20 mm water in the soil profile would have accumulated since the beginning of sowing window. As the dry season precedes peanut sowing, it was also assumed that there was little or no transpirable water in the soil profile early in the season. This is a valid assumption since around 80 % of annual rainfall is received over the Indian subcontinent during Kharif (June to September; Parthasarathy et al., 1994). If these conditions were not met, sowing was forced at the last day of two-week intervals with 20 mm irrigation to establish the crop, at minimum.

The recommended sowing density for peanut cultivation in India is 33 plants m^{-2} (ICAR-DGR, 2016). Based on that, three sowing densities were chosen for simulations as standard with 33 plants m^{-2}, lower and higher densities, respectively, with 22 and 44 plants m^{-2} as alternatives.

Being capable of fixing atmospheric nitrogen through rhizobial symbiosis, peanut requires only a small amount of basal N application for establishment prior to the formation of nodules and this is well-captured by the SSM model (Soltani and Sinclair, 2011). This requirement was accounted for in simulations by reflecting the recommended basal fertilizer dose of 20 kg N ha^{-1} (Rachaputi et al., 2021; Trivedi, 2011) as the initial soil nitrogen content.

Despite we study the Kharif (rainy) season, its well documented the rains themselves are not sufficient to support the crop in some parts of India (e.g. North-West and Southern semi-arid regions) and farmers do provide additional irrigation (DLD, 2017; Hajjarpoor et al., 2021; Raju et al., 2018). To investigate the irrigation management practices, we simulated rainfed, partial-irrigated, and fully irrigated systems (details in Table 3). We used the three modes in the SSM modules which allow to simulate rainfed crop, partial irrigation (40 mm water at the beginning of flowering and another 40 mm water at the beginning of seed set which is broadly representative of farmer practice across the regions) and full irrigation module (using the fraction of transpirable soil water (FTSW) threshold of 0.5 to trigger irrigation as used by Vadez et al. (2017) and Hajjarpoor et al. (2021).

2.3. Model setup, computational architecture and preparation for large scale simulation

SSM model, originally written in VisualBasic (VBA) programming language (which is the embedded version in Excel known as "Macros", Soltani and Sinclair, 2012a), was developed to capture the mechanistic nature of the key plant processes, i.e., the concepts of resource capture, resource use efficiency, and mass partitioning to the grain. SSM simulates phenological development, leaf development and senescence, dry matter production and partitioning, plant nitrogen balance, yield formation and soil water balance. Responses of crop processes to environmental factors of solar radiation, CO_2, photoperiod, temperature, nitrogen and water availability, and genotype differences are included in the model (For more details on relational diagrams refer to Figs. 9.3, 12.2 and 14.2 of Soltani and Sinclair, 2012a). The model arbitrates the outputs of simulated interactions of the crop (G) with environmental (E) and crop management (M) in a daily time step, thus reproducing the system dynamics in the daily steps.

The model was rewritten in C# programming language to automate, speed-up, and effectively process large number of GxExM factorial simulations. This allowed optimization of the model performance by allowing automated batch processing as well as multithreaded parallel processing. In developing the C# processing program, a lot of focus was

[5] https://power.larc.nasa.gov/.
[6] https://dsp.imdpune.gov.in/.
[7] https://www.isric.org.

Table 1
Overview of simulated genetic (G) factors and their levels in the SSM model set-up. These crop characteristics combinations resulted in 320 virtual crops with variable maturity, vigor, and different capacities to conduct water.

Crop characteristics	Corresponding SSM coefficient	Range of coefficients tested	Relevance/references
Crop duration [biological days, bd]			
Emergence to flowering	bdEMRR1	15.66 17.4 19.14 20.88	The coefficient specifies the number of biological days with optimal photoperiod, temperature, and soil water required for the crop to complete a specific phenological phase (Soltani and Sinclair, 2012a). These range of parameters reflects the duration of three main peanut product profiles for India at ICRISAT; short, medium and long duration (internal documents, expert knowledge).
First pod to first seed	bdR3R5	9.45 10.5 11.55 12.6	
Duration of seed filling	bdR5R7	55.8 62 68.2 74.4	
FTSW and VPD response			
The fraction of transpirable soil water (FTSW) that triggers stomata closure	WSSG	0.2 0.35 0.45 0.55	WSSG coefficient specifies the level of soil dryness when plant starts closing the stomata (measured as FTSW threshold). Higher WSSG, means that stomata starts closing in wetter soil. The reported range of FTSW in peanut is between 0.22 and 0.71 (Devi et al., 2009).
The vapor pressure deficit (VPD) threshold that triggers stomata closure	vpd_resp	No response 1.5 kPa 2.2 kPa 2.9 kPa 3.6 kPa	Plant increases transpiration linearly with increasing VPD (vpd_resp – no response). Plant stops increasing transpiration at specific VPD level. Plants limiting their transpiration produce less biomass with higher transpiration efficiency (Sinclair et al., 2010; Soltani and Sinclair, 2012b).
Vigor related			
Max Root depth	MEED	1000 mm 1200 mm	Plant allometry is, upto certain extent, regulated from same genetic regions (Sivasakthi et al., 2018; Vadez et al., 2012). Vigorous cultivar is expected to grow and expand their organs faster (root and shoot). Accordingly, vigorous crop type linked together the roots that expanded faster and reached deeper , with leaves that emerged and expanded faster.
Root extension rate	GRTDP	35 mm/day 38.5 mm/day	
Phyllochron	phyl	56 °C/leaf 50.4 °C/leaf	
Leaf area development rate	PLAPOW	2.75 3.025	(e.g. Halilou et al., 2016)

Table 2
Soil characterizes used in the model set-up and corresponding SSM parameters. Soil sAturation Limit (SAT, $m^3 m^{-3}$), soil Drained Upper Limit (DUL, $m^3 m^{-3}$), volumetric EXTRactable water content (EXTR, $m^3 m^{-3}$), and Soil effective rooting DEPth (SOLDEP, mm).

SOLTEX	SAT	DUL	EXTR	SOLDEP	Relevance/references
Alfisol	0.40	0.25	0.09	600 900 1200	Based on Handbook of Agriculture (Trivedi, 2011), ISRIC, NBSS-LUP, and expert knowledge from NBPGR and ICRISAT (details in Hajjarpoor et al, 2021).
Vertisol	0.45	0.38	0.13	600 900 1200	

put into areas of optimization and reusability. Therefore, the tool follows a modular architecture with a high degree of independence of individual modules (Fig. 2).

The modular architecture allowed for adjusting or improving the simulation processing individually and ensured high performance of the developed software tool. The software was run on a high-power processing PC for simulations. The computer had two AMD EPYC 7281 16-core processors, with 128 GB of RAM and a 1 TB SSD hard disk drive for storage. This processing capacity made it possible to execute the simulations in parallel by utilizing multithreading. The overall processing took approximately two days and generated 1.02 TB of output data. Considering the total number of 1173 grid locations and 60,480 simulations on each grid, this averages to a speed of approximately 400 simulations per second.

2.4. Raw simulations output and its analysis

Each of the 60,480 simulated GxExM combinations consisted of raw simulation output in 41 rows of data (i.e. 41 simulation seasons) within the single csv file (one file per grid). This data contained the summary of the simulated production output (yield, biomass etc.) as well as specification of each particular GxExM simulation setup, which enabled further output sorting, filtering and analysis.

2.4.1. Evaluation of simulations based on the anticipated crop value and production risks for end-users; Index of Goodness "IoG"

To evaluate simulated GxExM combinations, we designed a custom index ("Index of Goodness", further as IoG) which represented the overall performance of each combination based on the crop value as per our understanding of farmer's requirements; In India, peanut is primarily used for grain production (Rachaputi et al., 2021; Rathnakumar et al., 2013; Trivedi, 2011), however, non-negligible crop value relates to haulm production which is used as the livestock fodder (Blümmel et al., 2012; Rachaputi et al., 2021; Rathnakumar et al., 2013). Importantly, majority of peanut producers in India are small-holder farmers (Nautiyal and Mejia, 2002; Rathnakumar et al., 2013), whose livelihood might be at considerable risk in the case of peanut crop failure. These ground-realities were reflected in the single IoG number.

Table 3
Overview of simulated agronomy management (M) factors and their modified levels in the model set-up.

Modified factor	Corresponding SSM coefficient	Range of M variation tested	Relevance/references
Sowing window (within Kharif season)	Pdoy	16th May –31th May 1st June –15th June 16th June-30th June 1st July-15th July 16th July-31st July 1st Aug-15th Aug 16th Aug-31st Aug	Details on Kharif season in India in Hajjarpoor et al., 2021; Trivedi, 2011
Planting density [plant m^{-2}]	PDEN	22 (Low) 33 (Optimum) 44 (High)	Details on recommended planting density in the Handbook of Agriculture (Trivedi, 2011) and Annual Report 2015–16 (ICAR-DGR, 2016).
Irrigation level	IRGW	Rainfed: No irrigation Partial irrigation Full irrigation	Partial irrigation rule in SSM: 40 mm water at the beginning of flowering and beginning of seed growth Full irrigation rule in SSM: irrigate before soil moisture limits the crop growth (i.e. irrigate when FTSW < 0.5)

The simulations were separated into groups (Table 4). That represented major environmental drivers of production and unlikely to vary within the production system (e.g., access to irrigation or soil properties are "fixed" for the farmer, details in Table 4).

The process of IoG calculation was conducted in the following steps (Fig. 3):

(i). To account for the frequency of crop failure, instead of calculating average yield/biomass we opted to select the value corresponding to the 10^{th} percentile. In our respective dataset, this meant using the 5th smallest value from the 40 years for each GxExM simulation setup.

We choose the 10^{th} percentile over the average because 1) unlike the average, the 10^{th} percentile captures the absolute level of the simulation output, as well as the variance of the results (i.e., a dataset with higher variance results in the 10^{th} percentile being lower than average). 2) the 10^{th} percentile is unaffected by any abnormally high simulated values in "good years" (top outliers). From the point of view of small-holder farmers who would likely prefer a stable output, this choice would essentially represent a 90 % guarantee of those minimum production outputs compared to average, where crop production failures can be "hidden" (i.e. balanced by high yields in non-failure years).

(ii). We normalized the results from i) by converting from absolute to relative values within a particular simulation group (see Table 4). For this, we calculated average for the 10^{th} percentile and its standard deviation for each group

Average: $\mu = \frac{1}{n}\ \Sigma x$;

where n is the number of simulations in the group and x is the 10^{th} percentile result.

Standard deviation: $\sigma = \sqrt{\frac{\Sigma(x-\mu)^2}{n}}$;

Consequently, two z-score values to represent the rank of each simulation within its group for grain and haulm yield were calculated:

Z-score $= \frac{x-\mu}{\sigma}$

This z-score value therefore represents how many standard deviations each simulation is above or below the average (withing the same group as per Table 4).

(iii). Finally, to express the IoG as a single number, the z-scores for grain and haulm yields were weighted by a factor of 70 % for yield / 30 % for biomass and combined to reflect the crop value anticipated by farmers:

$\text{IoG} = 0.7 \times \text{z-score}_{\text{yield}} + 0.3 \times \text{z-score}_{\text{haulm biomass}}$

The resulting IoG tended to follow normal distribution, and the values were within the range typical for z-scores (approx. -3 to $+3$, which would capture 99.7 % of the data).

2.4.2. De-risked and weighted IoG

IoG is conceptualized to guide the decisions of two types of end-users:

- <u>Primary producers (farmers, farmer associations and advisory services)</u> to optimize the crop production practices based on their context: i) location ii) soil conditions and iii) capacity to provide irrigation
- <u>Crop improvement programs (breeding, phenotyping)</u> to breed context-specific peanut using the simulated ideotypes as a "blueprint"

Nevertheless, both of these end-user groups might be facing several risks by trying to adopt the optimal G and M based on the IoG calculated above. After the discussions with these two groups of stakeholders, we have accounted for enhanced IoG calculations further:

- <u>De-risked IoG for primary producers</u> Crop producers might face logistical / resource-related difficulties in adhering to the recommendations. In some instances, farmers cannot adhere to sowing within the optimal sowing window (due to previous crop logistics or socio-ethnic context). In other cases, optimal crop density cannot be achieved (e.g. because of seed viability issues). For these anticipated reasons we calculated "de-risked IoG". De-risked IoG extends IoG calculated in 2.4.1 and incorporates a weightage factor of similar simulations to de-risk the uncertainty that farmers might deviate from the recommended optimal practices;

For sowing date, the "de-risked" IoG considers the simulations with directly preceding or following sowing window as its neighbors and incorporates them with a total weightage of 30 %. This means that any particular simulation would contribute 70 % of its own IoG and 30 % of IoG calculated for the preceding or following sowing window [15 % / 70 % / 15 %]. For simulations with only one neighbor, the weightage will be 30 % / 70 %. Unlike the IoG, which assumes that producers will follow the recommendations perfectly, the "de-risked" IoG in-builds 30 % uncertainty that farmer misses the optimal sowing window and de-prioritizes simulations with "bad neighboring simulations".

In the case of planting density, there is uncertainty that farmers might not be able to procure sufficient seeding material to achieve the desired plating density or that the seed is of low quality. Here, we considered that the optimal simulation itself contributes 60 % to the

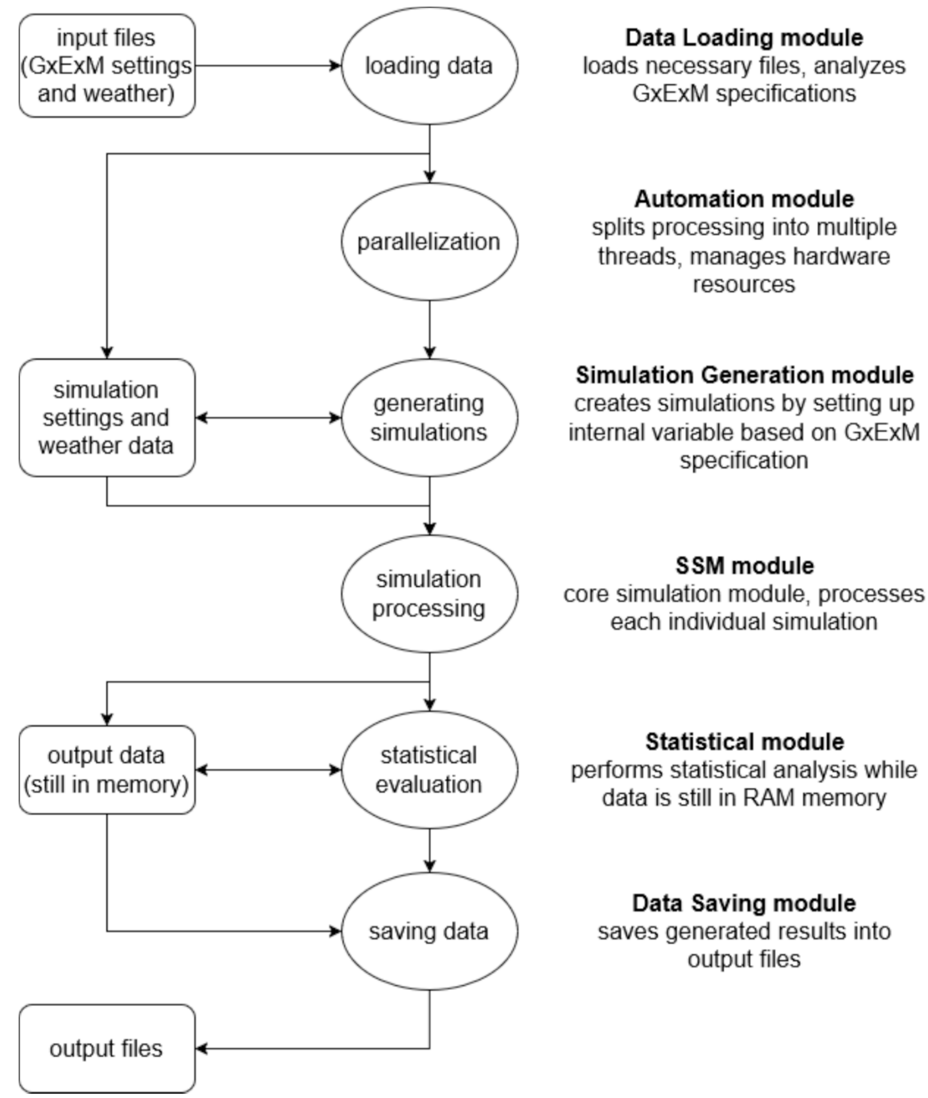

Fig. 2. Modular architecture of the SSM processing software tool, simplified to show the main task performed by each module, the sequence in which the modules are used and the flow of main data sets between modules.

Table 4
Environmental (E, Table 2) and management (M, Table 3) parameters determining fixed "simulation groups", i.e., the parameters that are of major influence on production and/or cannot be altered within the production system. Index of Goodness (IoG) was calculated separately within these 18 ExM groups.

E/M parameter defining the simulation group	Simulation group description (# of groups)
Irrigation	Irrigated, partially irrigated, rainfed (3)
Soil depth	Shallow, medium, deep (3)
Soil texture	Alfisol, vertisol (2)

"derisked IoG" with 40 % chance that the density will be lower than recommended. With three possible optimal planting density levels (44, 33 and 22), we accounted for that the planting density 22 would contributes 100 % its own IoG (as it has no neighbor), planting density 33 would contribute 60 % its own IoG + 40 % IoG of neighbor simulation with density 22, planting density 44 would contribute 60 % its own IoG + 30 % IoG of simulation with density 33 + 10 % of IoG of simulation with density 22.

In principle, the "de-risked" IoG accounts for the IoG value of the neighboring simulations to prioritize GxExM combinations that do not exhibit a sharp decline in production due to slight shifts in the simulation multi-dimensional space. In practice, this should account for the situations where we anticipated the stakeholders would face difficulties to adhere to recommendations while still realizing the "next best" options with higher probability.

- Weighted IoG for crop improvement programs accounts for the uncertainty of crop phenotyping accuracy – i.e., breeders may not be able to select varieties with the optimal G parameter (e.g., phenotyping tools are not available). For these anticipated reasons, we calculated a weighted version of IoG for four simulated G parameters (crop duration, vigor, crop responsiveness to soil and atmospheric drought, Table 1).

That means four additional values were calculated in each of these dimensions, and weighted (Table 5). By comparing these weighted versions to the base IoG, we could compare the "sensitivity" of IoG to changes in each of the G parameter dimensions. In this way, the magnitude of the difference between the original IoG and weighted IoG determines which of the genetic parameters is the most important to breed for to enhance production in a given ExM context (see 2.4.3).

2.4.3. Cluster analysis and visualization

The main idea of the clustering was to identify the geographic

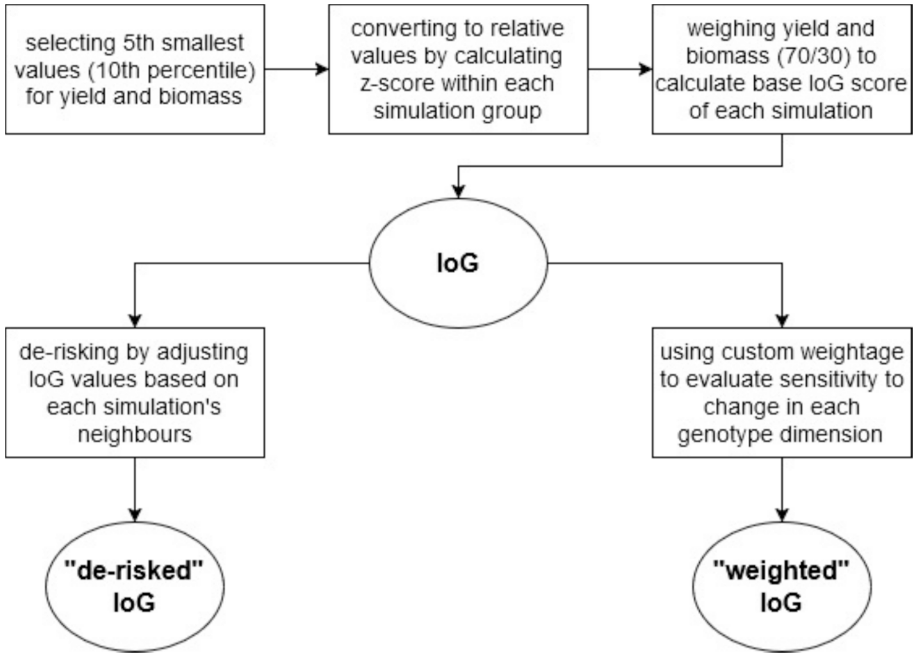

Fig. 3. Process of calculation of IoG, “de-risked” IoG (end-user target, primary producers) and “weighted” IoG (end-user target, crop improvement programs).

Table 5
Genotype parameters (G), the number of their simulated levels (from Table 1) and the weightage factor used for calculating the IoG sensitivity to a single G dimension (“weighted IoG”).

G	# of simulated levels	Weightage of each level for IoG
Crop duration	4	25 %
Vigor	2	50 %
VPD responsiveness	5	20 %
Responsiveness to soil drought	4	25 %

regions (E) with similar production drivers/limitations, for the “fixed” situations with maximum effect on production – i.e., irrigation (Table 4). Within clusters we could expect less GxExM effects on production and more homogeneous crop responses to the system intervention (G × M). For this, we assessed the simulations within each of the grids in the following steps (note: grids with a yield less than 100 kg ha^{-1} at 90th percentile were excluded. This was mostly the case of high-altitude locations in Himalayan regions):

- We selected the simulations within different irrigated conditions (average across six groups of soils, see Table 4) that attained maximum de-risked IoG value. For all 28 input-output parameters of these simulations, a correlation matrix was first generated. Then we utilized the findCorrelation function from the caret package (Classification And REgression Training) in R to select the unique and the most influential parameters driving the variability in the given dataset (Kuhn, 2008). This narrowed down the selection to 13–14 variables within different irrigation regimes.
- Consequently, we ran principle component analysis (PCA) only with these selected parameters (incl. latitude and longitude). The loadings of the first number of PCs (~3–4), which explained >85 % of the dataset, were analyzed by Partitioning Around Medoids (PAM) clustering methods (Kaufman and Rousseeuw, 1990).
- For each of three different irrigated systems, the range of clusters was visualized on the map (ggplot2 R package) and the results were summarized for the most sensible number of clusters (Nbclust R package). The significance of the differences in G, E and M parameters between these clusters was evaluated by ANOVA and mean comparison tests.

3. Results

3.1. *High performance computing and simulation outputs analyses*

3.1.1. *HPC and computational architecture*

For our work, it was required to generate, process and analyze 60,480 factorial GxM combinations of simulations (>2.3 billion simulations; 1.02 TB of raw output data). Therefore, while developing the C# tool for simulation and processing program, optimization and reusability elements were focused. For that, we developed a modular architecture that allowed for easy adjusting or improving individual elements of the simulations’ processing independently. In our case, this modular system became particularly advantageous, for iterating the analyses and flexible hypothesis testing using the generated simulation outputs for which the statistical module was frequently altered and changed. The input and simulation generation module used object abstraction, which allowed processing of different GxExM setups without change in the actual source code. The multithreading module managed input data sources so that these could be shared between simulations without unnecessary duplication for each processing thread, thereby increasing processing efficiency. Altogether, this system achieved a computation speed between 400 and 500 simulations per second (~40 h to generate all required simulations). This fast turnaround allowed for more experimentation and fine-tuning of the methodology.

3.1.2. *Inputs setup for effective simulation management*

Simulation for each of the grid locations required the weather files and the settings file specifying the combination of GxExM parameters. The factorial GxExM simulations needed to take into account that reflecting one specific system condition (e.g., soil or genotype) requires a combination of several specific variables (e.g., alfisol was achieved by a combination of 3 soil variables of SSM: SAT, DUL and EXTR; Table 2). These sub-variables were linked together and generated one factor in the combination. The second specificity is that certain parameters determined the simulation group (Table 4). Since it was required to evaluate outputs belonging to particular simulation group (Table 4, section

2.4.1), it was necessary for the simulations to be calculated in a specific order to make sure simulations from the same group formed a continuous block of memory. That way the statistical processing did not require any sorting or reorganizing of the data.

3.1.3. Simulation output data analysis

The outputs for each simulation were stored in RAM until all simulations for the grid were concluded. Then the statistical module performed further analyses (averages, medians, variances for selected output columns within each simulation group), including the calculation of the base IoG for each simulation. This was done so that the utilization of RAM as a processing resource could be predicted (it scales linearly with the number of simulations) and, accordingly, the processing could be optimized and split into the appropriate number of parallel threads to utilize memory in an efficient manner. The ability for statistical evaluation to occur immediately after finishing the simulation reduced the

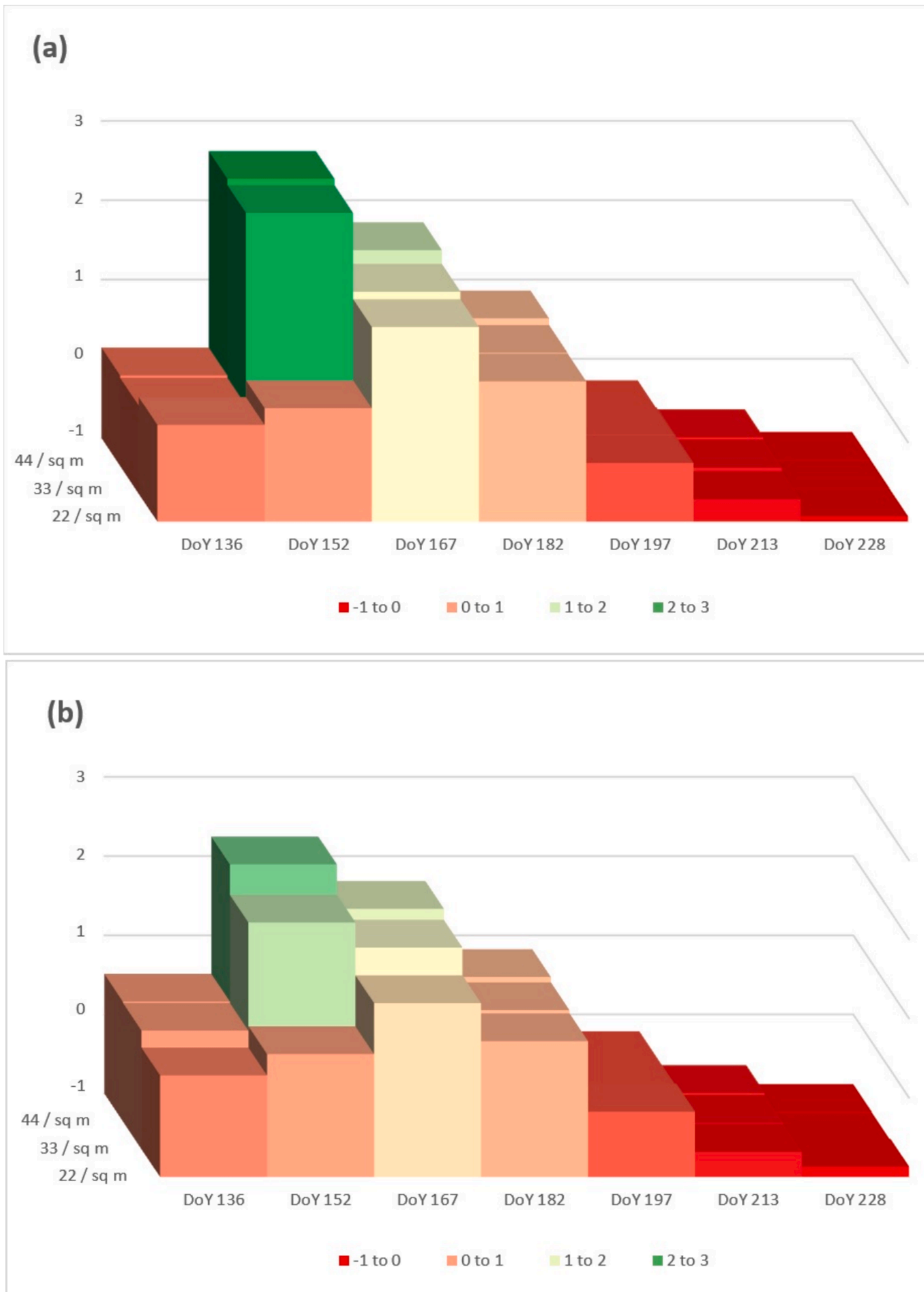

Fig. 4. Visualization of Index of Goodness (IoG, a) and de-risked IoG (b) example data. On the charts is IoG/de-risked IoG (y-axis) based on the combination of sowing day (x-axis) and planting density (z-axis). The weightage used during the "de-risking" process considers neighboring simulations, so any high IoG values (Dark green bars in chart a) with low-value neighbors are reduced (Pale green bars in chart b). This adjustment is only one example of 21,114 separate values (18 groups in each of 1173 grid locations) helping primary producers in choosing options with lower risk when when comparing all adjustment values. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

amount of read / write operations as well as time needed for simulation's evaluation.

3.2. Evaluation metrics for simulations

Second part of the simulation output evaluation was done separately (not as part of the simulation software) by a custom-made Python script in order to calculate the "de-risked" and "weighted" version of IoG. This multidimensional re-calculation eliminated situations where a single simulation would have high IoG but would be at an extreme value within its simulation neighborhood. Fig. 4 shows an example of the effect of "de-risking" the IoG in a 3D bar chart. Because the weightage used during the "de-risking" process takes into account neighboring simulations, any high IoG values with low-value neighbors are being reduced. Therefore, the "de-risking" process prioritizes such GxExM combinations, where a slight deviation from the optimal parameter value does not cause a sharp decline in production. This "de-risked IoG" was then used to compare the importance of particular GxExM for farmers (for Section 3.3.1).

Weightage of IoG ("weighted IoG") was introduced to quantify the importance of the G factors for peanut production (for section 3.3.2). The magnitude of the difference between IoG and the weighted IoG is proportionate to the production loss due to the deviation from identified optimum G level. Therefore, the maximum difference between IoG and weighted IoG signified that the particular genetic factor is more important for the optimized productions in target ExM context. Consequently the G factors with comparatively larger drop when weighted were considered more important for peanut crop "ideotype" as the diversion from that particular level of genetic factor would affect the production comparatively more. Weighted IoG can then be seen as a way to quantify and compare the consequences of deviating from this ideotype in particular context (Section 3.3.2, Table 7) and thus used to prioritize selection for peanut genetic features specific for ExM breeding programs.

3.3. Environmental clusters and their characters

Irrigated system with the highest yield potential (~5 t ha^{-1}) resulted in the most relatively homogeneous peanut production (Fig. 5). The production decreased and became increasingly heterogeneous in partially irrigated and, moreover, in the rainfed systems (Fig. 5). The optimal number of clusters increased alongside environmental heterogeneity with 4 clusters identified under irrigated conditions, and 5 and 7 clusters in partially irrigated and rainfed conditions, respectively (Fig. 5).

The transition of the systems from full to partial irrigation resulted in the disaggregation of the Northern (sub-Himalayan) environmental unit (cluster #2 transformed into #2 and #5, Fig. 5). The transition from

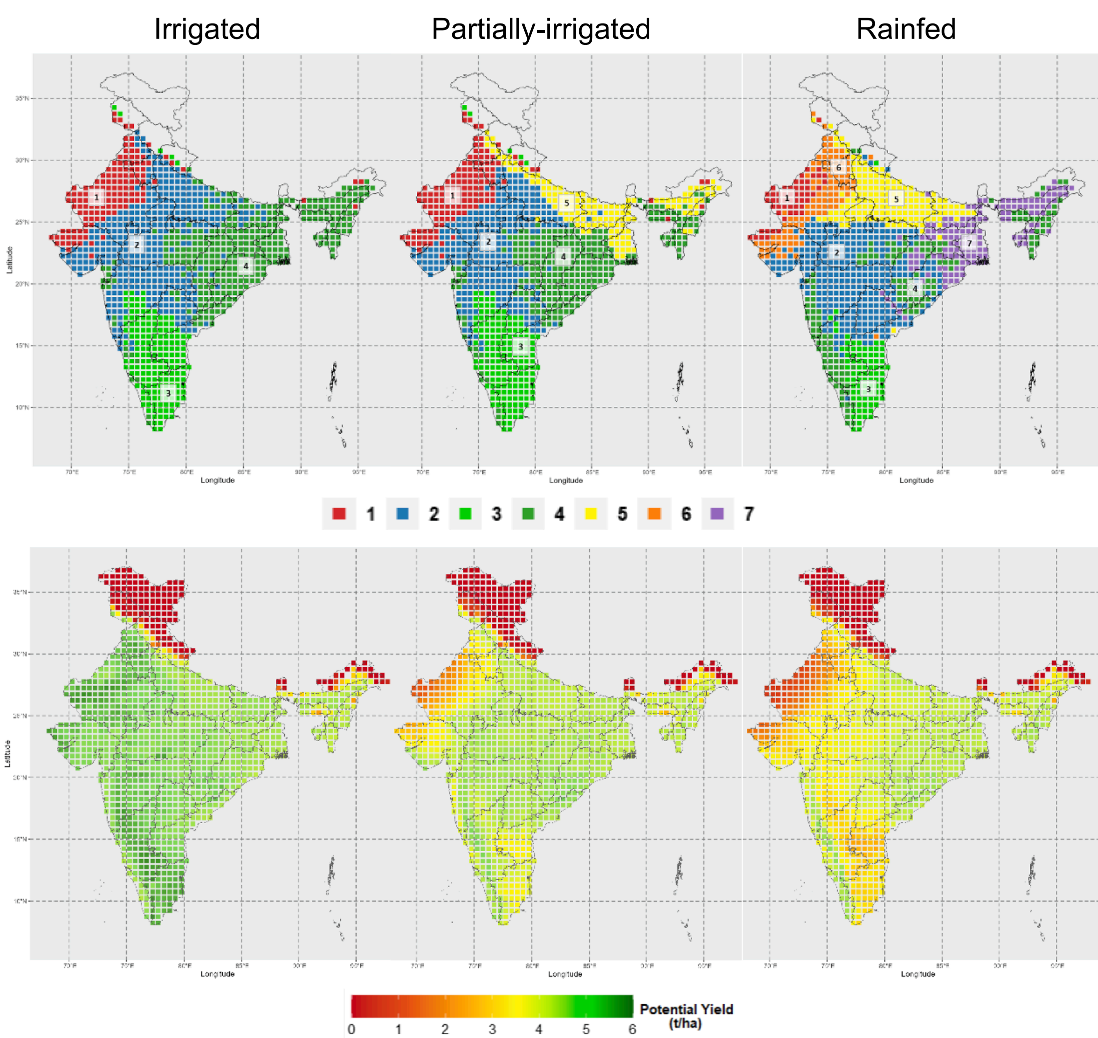

Fig. 5. Environmental clustering (above) and the associated yield potential (below, t ha^{-1}) for different irrigation scenarios. Grids yielding less than 0.1 t ha^{-1} were excluded from clustering to prevent misleading results with unreal conditions of growing peanut (mostly the case of high altitude locations in Himalayan regions). Heterogeneity is increased by water limitation resulting in more clusters from irrigated to partially irrigated and rainfed systems.

partial irrigation to rainfed systems further separated the North-West regions (Rajasthan and N Gujarat) into two distinct clusters (cluster #1 transformed into #1 and #6, Fig. 5) and separated the"dispersed" cluster #4 influenced by the ocean proximity into two distinct cluster (cluster #1 transformed into #4 and #7, Fig. 5). Table 6 provides the overview of the identified environmental clusters, peanut production (P) and the underlying optimal G, and M characters for prevailing E (aligned with the states in India).

3.3.1. Characters of peanut system clusters within different irrigation management scenarios

Our modeling analysis indicated that irrigated peanut systems (Table 6a, Table S1a) would benefit from intensified production methodologies. These would include production management practices that maximize planting density (~44 plants m^{-2}) combined with long-duration genotypes (>73bd) with high vigor. The identified clusters mostly differed in climatic variables, resulting in different lengths of growing seasons (139–164 days) and amount of intercepted radiation (Clusters #4 and #2 had comparatively low intercepted radiation).

Our model representation of partially irrigated systems could be still considered as high-production environments (Clusters #2, 3, 4, 5) with the exception of cluster #1 where a decline in production due to irrigation deficit was apparent (from yield 5.50, haulm 3.62 [t ha^{-1}] in irrigated system to yield 3.71, haulm 2.86 [t ha^{-1}]; compare Table 6a, b and Table S1a, b). The clustering analysis discriminated the sub-Himalayan region (cluster #5), which had early sowing dates (159 DOY) and higher in-season rainfall (1132 mm/season). The optimal suite of intensified M and G for clusters #2, 3, and 4 was similar to irrigated systems (high planting densities, crop duration, and vigor) except that the optimal crop type (G) included early crop responsiveness to drying atmospheric and soil. These G factors were rather similar to cluster #1 (and all clusters in rainfed systems, below) with the occurrence of droughts.

Some of the clusters identified in rainfed systems retained the characteristics of well-endowed environments (clusters # 2, 4, 5, 7) and so followed the suite of optimal M and G (high planting densities, longer crop duration and vigorous crop types). Nevertheless, even with the optimized GxM, the production decline from irrigated and partially irrigated conditions within clusters #1, 6 was very apparent (>1t ha^{-1} of yield loss; compare Table 6 a and b with c). Cluster #6 was more vulnerable to early and, to some extent, late droughts, whereas cluster #1 experienced frequent early and severe terminal droughts and the effective growing seasons were shorter compared to other clusters (96–125 days). In all clusters, the drought adaptive G factors gained importance compared to irrigated and partially irrigated systems. In clusters #1, 2, 3, 5, and 6 a crop ideotype would include both early responsiveness to soil and atmospheric drought, while in clusters #4 and 7, the crop ideotype would include mostly atmospheric drought responsiveness.

3.3.2. Cluster-specific crop ideotypes

Within the differently irrigated systems and individual clusters (3.3.1), we analyzed the importance of each tested G parameter for the production quantified as the proportional change in IoG and weighted IoG (Table 7). This aimed to assess the importance of the deviation from the "ideal" G value (i.e., crop ideotype) measured as the proportional drop in the value of weighted IoG. The analysis highlighted the comparative importance of crop duration (phenology) and vigor over crop responsiveness to soil and atmospheric drying in irrigated systems. In partially irrigated systems, the crop responsiveness to atmospheric drying was the most important of all tested G parameters. The crop responsiveness to atmospheric drying remained the most important G factor in rainfed systems, along with the crop responsiveness to soil drying which became the second most important parameter across most of the clusters.

4. Discussion

4.1. Utilization of the HPC system architecture

Simulation software such as SSM or APSIM was primarily built for biologists to test hypotheses on small-scale datasets (Hammer et al., 2010; Soltani and Sinclair, 2012a), so the single PC suffices the computations. Such scales are sufficient when the simulation software is used to evaluate and extrapolate field experimentation that typically involves thousands of different options at most (e.g. Chenu et al., 2018; Collins and Chenu, 2021; Diancoumba et al., 2024; Hajjarpoor et al., 2021, 2018; Holzworth et al., 2014; Kholova et al., 2014; Sinclair et al., 2010; Vadez et al., 2017). For many simulation scenarios, a single PC can process such a number of simulations in a matter of minutes (hours at most), and some software have provision to organize factorial simulation runs of that order of magnitude. The fact that the computation time and utilization of hardware resources was not considered an issue in such type of applications can be corroborated by the fact that even the publications of the global leaders in model development rarely mention the IT aspects of the calculations (e.g. SSM, DSSAT, APSIM, ECOMERISTEM development teams). However, computational demands of simulation models can vary based on their complexity (Boote et al., 2021; Tardieu et al., 2020). For instance, models that operate on hourly time steps (e.g. to simulate photosynthesis, Wu et al., 2019) or functional-structural plant models (FSPMs) require considerably higher computational resources. Large-scale applications of CSM are computation intensive and we could trace only a few literature sources where data processing efficiency was systematically examined (Fainges, 2015; Zheng et al., 2016). However, it is clear that the efficiency, throughput and optimization of computation systems remain a key challenge for applications of CSMs in agriculture (Holzworth et al., 2018; Jang et al., 2019; Khabarov et al., 2020; Montañana et al., 2020).

For a large-scale computation presented hereby (2.3 billion simulations, 1.02 TB of raw output data), the optimization and efficient use of hardware resources was an issue. At the presented scale, a one-order magnitude increase in computational efficiency would turn years of simulation processing into weeks and is a new type of interdisciplinary research that requires the cooperation of biologists and IT experts in "Big Data" (which is traditionally difficult). Therefore, one of the novelties of our work is the architecture for big data management and processing. Particularly emphasizing the C# tool using modular simulation and processing programs focusing on flexible optimization and reusability. Such modular architecture allowed effective simulation output analysis in our work and could be used in the future, for example, to swap out the SSM core simulation module with a different crop growth model while utilizing the surrounding modules for data processing. The architecture ease further development – e.g. to suit other end-users requirements considering a real-time simulation.

4.2. Index of Goodness (IoG) emerges as a key innovation, providing a more intuitive and relevant framework for assessing simulation guided by end-users needs

Another hurdle inherently linked to "big data'" relates to their evaluation and interpretation. In our case, the data analytics was inspired by the anticipated end-users (primary peanut producers and peanut improvement programs) and co-designed with their participation and feedback. Primary producers (e.g. farmers, farmer associations and advisory services) need to optimize the crop production practices specific to their production context while crop improvement programs (breeding, phenotyping) need to identify context-specific peanut ideotypes which can be further realized by "physiology-assisted breeding" or "trait-specific breeding" (Welcker et al., 2022). In the literature, very few analytical approaches are available to inspire the analysis for such specific user needs however, maximizing average yield is a common practice in selecting optimal scenarios within crop adaptation strategies

Table 6
Description of main clusters for three irrigation scenarios (fully irrigated (a), partially irrigated (b), rainfed (c)). Alongside, the optimal crop characteristics (G) by management (M) options among different clusters with their environment characters (E), and production potential (P). The statistical analysis of GxExM factors within clusters is in Suppl. Material Table S1a, b, c.

Table 6a / #Cluster*	G**	E	M	P (t ha-1)
Irrigated system				
#1 visualized in red Thar desert region: W-Rajasthan, N-Gujarat, Haryana, Punjab	Crop duration~73bd, vigorous crop type**	Season~141 days, Rainfall~342 mm/season, SRdn~2533 MJ m^{-2}/season, ~17.9 MJ m^{-2}/day, minT/maxT ~ 21.9/33.8°C	Sowing ~ 5th July (186 DOY), density ~ 44 plant m^{-2}	Yield~5.50; Haulm ~3.62
#2 visualized in blue E-Rajasthan, S-Gujarat, Madhya Pradesh, Uttar Pradesh, W-Bihar, Maharashtra, Telangana, parts of Karnataka	Crop duration~74bd, vigorous crop type**	Season~151 days, Rainfall ~441 mm/season, SRdn ~2494 MJ m^{-2}/season, ~16.5 m^{-2}/day, minT/maxT ~19.5/30.9°C	Sowing ~ 4th August (216 DOY), density ~ 44 plant m^{-2}	Yield~5.45; Haulm ~ 3.67
#3 visualized in bright green Andhra Pradesh, parts of Karnataka, Tamil Nadu, S-Telangana	Crop duration ~74bd, vigorous crop type closing stomata at FTSW~0.40	Season length ~139 days, Rainfall ~506 mm/season, SRdn~2399 m^{-2}/season, 17.2 MJ m^{-2}/day, minT/maxT~21.8/31.0°C	Sowing ~ 15th July (196 DOY), density ~ 44 plant m^{-2}	Yield~5.60; Haulm~3.80
#4 visualized in dark green Chhattisgarh, Jakharand, West Bengal, Odisha, East-states of India, Ocean-proximal parts of Kerala, Goa and Maharashtra	Crop duration ~74bd, vigorous crop type closing stomata at FTSW~0.39	Season length~164 days Rainfall~ 506 mm/season, SRdn ~2585MJ m^{-2}/season, 15.8 MJ m^{-2}/day, minT/maxT ~18.8/28.9°C	Sowing~ 2nd August (214 DOY), density ~ 44 plant m^{-2}	Yield~5.27; Haulm~3.33
Table 6b / #Cluster*	**G****	**E**	**M**	**P (t ha^{-1})**
Partially Irrigated system				
#1 visualized in red Thar desert region: W-Rajasthan, N-Gujarat, Haryana, Punjab	Crop duration ~71bd, vigorous crop type closing stomata at FTSW~0.47	Season length ~117 days, Rainfall ~422 mm/season, SRdn ~2322 MJ m^{-2}/season, 19.8 MJ m^{-2}/day, minT/maxT ~25.7/35.9°C	Sowing ~ 5th June (156 DOY), density ~ 42 plant m^{-2}	Yield~3.71; Haulm~2.86
#2 visualized in blue E-Rajasthan, S-Gujarat, Madhya Pradesh, Uttar Pradesh, Maharashtra, Telangana, parts of Karnataka	Crop duration ~74bd, vigorous crop type	Season length~127 days, Rainfall ~756 mm/season, SRdn ~2121 MJ m^{-2}/season, 16.7 MJ m^{-2}/day, minT/maxT ~23.8/32.4°C	Sowing ~ 19th June (170 DOY), density ~42 plant m^{-2}	Yield~5.08; Haulm~2.97
#3 visualized in bright green Andhra Pradesh, Kerala, Karnataka, Tamil Nadu, S-Telangana	Crop duration ~74bd, vigorous crop type	Season length ~133 days, Rainfall ~597 mm/season, SRdn ~2248 MJ m^{-2}/season, 16.8 MJ m^{-2}/day, minT/maxT ~21.9/30.6°C	Sowing ~ 17th July (198 DOY), density ~ 42 plant m^{-2}	Yield~4.80; Haulm~3.64
#4 visualized in dark green Dispersed: Chhattisgarh, Jakharand, Odisha, E-Madhya Pradesh, East-states of India (except of Arunachal Pradesh), Kerala, Goa, Ocean-proximal parts of Karnataka and Maharashtra	Crop duration ~74bd, vigorous crop type	Season length~139 days, Rainfall ~1095mm/season, SRdn ~2143 MJ m^{-2}/season, 15.4 MJ m^{-2}/day, minT/maxT ~22.2/30.0°C	Sowing ~ 2nd July (183 DOY), density ~ 42 plant m^{-2}	Yield~5.13; Haulm~2.83
#5 visualized in dark yellow Sub-Himalayan regions of Uttar Pradesh, Uttarakhand, Bihar	Crop duration (74bd), vigorous crop type closing stomata at FTSW~0.39	Season length ~130 days, Rainfall ~1132mm/season, SRdn ~2117 MJ m^{-2}/season, 16.7 MJ m^{-2}/day, minT/maxT ~24.4/32.4°C	Sowing ~ 8th June (159 DOY), density ~ 42 plant m^{-2}	Yield~5.00; Haulm~3.17

(continued on next page)

Table 6 (*continued*)

Table 6c / #Cluster*	G**	E	M	P (t ha^{-1})
Rainfed system				
#1 visualized in red Hyper-arid regions of Thar desert: parts of W-Rajasthan, N-Gujarat	Crop duration~66.6bd, closing stomata at soil FTSW=0.55, VPD ~1.5 kPa	Season length ~96 days, Rainfall~210 mm/season, SRdn ~1967 MJ m^{-2}/season, 20.5 MJ m^{-2}/day, minT/maxT ~26.8/37.2°C), frequent early and terminal drought stress.	Sowing ~ 13th June (164 DOY); density ~38.7 plant m^{-2}	Yield~1.79; Haulm 1.83
#2 visualized in blue S-E-Gujarat, Madhya Pradesh, Telangana, Chhattisgarh, Maharashtra, Telangana, N-E Andhra Pradesh	Crop duration~74.1bd, vigorous crop closing stomata at VPD ~1.7 kPa	Season length ~96 days, Rainfall ~210 mm/season, SRdn~1967 MJ m^{-2}/season, 20.5 MJ m^{-2}/day, minT/maxT~26.8/37.2°C,	Sowing ~ 19th June (171 DOY), density ~ 43.7 plant m^{-2}	Yield~4.83; Haulm~2.87
#3 visualized in bright green Tamil Nadu, C-S Andhra Pradesh, S-E-Karnataka	Crop duration ~73.2bd, vigorous crop type closing stomata at FTSW~0.49, VPD ~1.5 kPa	Season length~96 days, Rainfall~210 mm/season, SRdn~1967 MJ m^{-2}/season, 20.5 MJ m^{-2}/day, minT/maxT ~26.8/37.2°C,	Sowing ~ 9th August (221 DOY), density ~ 43.7 plant m^{-2}	Yield~3.97; Haulm~3.05
#4 visualized in dark green Dispersed: parts of ocean influenced regions of Jakharand, Odisha, E-Madhya Pradesh, East-states of India (particularly Manipur, Mizoram, Nagaland), Ocean-proximal parts of Kerala, Goa and Maharashtra, E-Meghalaya	Crop duration ~(73.1bd), vigorous crop type responding to ~ VPD ~2.5 kPa	Season length~96 days, Rainfall~210 mm/season, SRdn~1967 MJ m^{-2}/season, 20.5 MJ m^{-2}/day, minT/maxT ~26.8/37.2°C,	Sowing ~ 25th June (176 DOY), density ~ 43.9 plant m^{-2}	Yield~5.05; Haulm~3.04
#5 visualized in dark yellow Uttar Pradesh, Uttarakhand, Bihar	Crop duration ~73.9bd, vigorous crop type closing stomata at VPD ~1.5 kPa	Season length ~96 days, Rainfall ~210 mm/season, SRdn~1967 MJ m^{-2}/season, 20.5 MJ m^{-2}/day, minT/maxT~26.8/37.2°C,	Sowing ~ 14th June (165 DOY), density ~ 43.8 plant m^{-2}	Yield~4.83; Haulm 3.14
#6 visualized in orange Arid regions of Thar desert: Punjab, Haryana, E-Rajasthan, E and N-Gujarat	Crop duration ~ 72bd, vigorous crop type closing stomata ~ FTSW~0.49, VPD~1.5 kPa	Season length ~96 days, Rainfall ~210 mm/season, SRdn~1967 MJ m^{-2}/season, 20.5 MJ m^{-2}/day, minT/maxT (26.8/37.2°C), occurrence of early drought stress.	Sowing ~ 8th June (159 DOY), density ~ 41 plant m^{-2}	Yield~3,43; Haulm~2.60
#7 visualized in purple Eastern ocean-influenced regions of West Bengal, Assam, Jharkhand, S-E-Bihar, Arunachal Pradesh, Meghalaya, Tripura	Crop duration~73.8bd, vigorous crop type closing stomata at VPD ~2.4 kPa	Season length~96 days, Rainfall~210 mm/season, SRdn~1967 MJ m^{-2}/season, 20.5 MJ m^{-2}/day, minT/maxT~26.8/37.2°C	Sowing ~ 9th June (160 DOY), density ~ 44 plant m^{-2}	Yield~5.07; Haulm~3.01

*Visualized colors are referring to Fig. 5 on environmental clustering.

**In a vigorous cultivar, root and shoot developing traits are improved, including maximum root depth and extension rate, phyllochron, and leaf area development rate.

(Zheng et al., 2018). To advance this approach, we developed an "Index of Goodness" (IoG) which is a single number that encompasses the key production parameters important for the primary producers (peanut grain for consumption and sale, haulm to feed animals, probability of crop failure). IoG was then used to evaluate and compare the simulations to find out the optimal suite of production practices specific to farmer conditions ("derisked IoG") and to identify and compare the key features of ideotypes to prioritize genetic factors for effective context-specific

Table 7
The importance of genetic factors over each other for the clusters' contexts in different irrigated peanut production systems. Each number represents the proportional importance of each tested G factor: phenology, vigor, responsiveness to soil drying (Res_Soil), and atmospheric drying (Res_Atm); [all in %]. This table uses a color-coded heatmap to visualize the data. The higher the value, the greener the cell representing the importance of the G factor for the production success (measured as % change in IoG) of that optimal ideotype in the respective cluster and system.

System	Cluster	Phenology [%]	Vigor [%]	Res_Soil [%]	Res_Atm [%]
Irrigated	#1	29	38	4	7
	#2	28	36	4	8
	#3	34	36	4	8
	#4	29	37	4	6
Partially irrigated	#1	19	13	19	64
	#2	15	12	14	60
	#3	20	15	20	54
	#4	19	17	20	54
	#5	20	17	18	50
Rainfed	#1	24	17	32	56
	#2	27	18	22	39
	#3	27	19	32	33
	#4	28	22	25	33
	#5	25	17	23	43
	#6	26	22	29	40
	#7	26	22	29	40

crop improvement ("weighted IoG"). Moreover, the IoG can be readily re-calculated to test different hypotheses from the available data (available at Pavlik et al., 2025).

4.3. Environmental clusters for targeted production optimization and ideotype breeding

De-risked IoG was used to find the best combination of GxM parameters, which were consequently used to identify the geographical units (E) that have higher levels of homogeneity and are expected to respond to changes in G and M factors similarly (environmental clusters). The intention was not only to re-investigate the zoning systems (as in Hajjarpoor et al. 2021) but also to provide quantitative decision-making support to optimize the suite of peanut production practices (G and M) within the identified environmental clusters. The result could minimize peanut productivity gaps caused by a lack of crop and management adaptation and prevent resource waste on varieties that might underperform in particular regions (Heinemann et al., 2019). The detailed summary of the identified clusters within each tested irrigation regime is in Table 6 and Table S1. In general, the simulated peanut production reflected well the expected system heterogeneity (e.g. Hajjarpoor et al., 2021; Trivedi, 2011; Witcombe et al., 1998) and was, as expected, on the higher end of crop production potential; The average current grain yield was 1.1 t ha^{-1} across India (IndiaStat, last five years), average simulated grain yield was 2.6–4.76 t ha^{-1} in irrigated and rainfed conditions across India (Hajjarpoor et al 2021), and potential optimized grain yield in this study was 4.1, 4.7 and 5.5 t ha^{-1} for rainfed, partially irrigated and irrigated systems, respectively.

The common M and G components to optimize production in irrigated systems included high planting densities (~44 plants m^{-2}) with long duration (>73bd) and vigorous crop types. The production was also found to be the most sensitive to crop phenology- and vigor-related G parameters, signifying that the ideotype bred for irrigated systems needs to primarily focus on these traits. Based on environmental clustering, most irrigated clusters, particularly clusters #4 and #2, were radiation-limited. In these conditions, it might be pertinent to further investigate use of higher radiation-use efficient cultivars (*in-vivo* or *in-silico*, using models simulating detailed photosynthetic processes (e.g., Abshenas et al., 2022; Wu et al., 2019). With declining water availability in partially irrigated and rainfed systems, the optimized production started to decline (similarly reported by Rathore et al., 2021). The environmental clusters of rainfed regime most affected by drought stress were clusters #1, 6 (arid regions of Thar Desert; experiencing early and late season water deficit). Despite the regions belonging to cluster #3 are reportedly affected by droughts (e.g. Hajjarpoor et. al. 2021), we did not detect the drought stress when the GxM was optimized. This might signify the droughts could be, at least up to a certain extent, avoided with relevant suit of practices; In these clusters # 1, 3, 6, the production benefited from conservative agronomic practices and crop capacity to regulate water transfer pathways (e.g. early stomata closure upon soil (FTSW > 0.4) and atmospheric drought (VPD ~ 1.5 kPa)). Both of these physiological processes allowed for conservative soil water use and benefiting the grain filling period under terminal type of drought stress in *in-vivo* experiments (Kholová et al., 2010a, 2010b; Sinclair et al., 2005) and is well reflected in simulations (e.g. Kholová et al., 2013; Sinclair et al., 2010; Soltani and Sinclair, 2012b). Hot and dry conditions during reproduction phase increase the risk of preharvest aflatoxin contamination (Rachaputi et al., 2021), although the GxM optimization of this study may reduce this risk in a prone area of cluster #3. Worth to note is that, currently, cluster #3 (SE-India) encompasses the major peanut producing area (details in Hajjarpoor et al., 2021), thus an important target area for breeders and agronomists. With depleting ground-water irrigation availability, development of drought-adapted ideotypes for these specific geographies might be a priority target. This agrees with our analysis that showed the production in these water-limited systems (1, 6, and, up to a certain extent, clusters 3 and 7) was also the most sensitive to deviation from the genetic parameters that support crop adaptation to drought (Table 7). These water-saving physiological functions should be considered a priority in breeding “climate-ready” crops for such a context.

We would like to mention the analysis of the geographies not currently cultivated with peanut that might be of some interest to experts. Interestingly, with GxM optimization, the model predicted favorable production environments in the Eastern Plains of India (clusters #4 and #7), where peanut production diminished decades ago. The decline in peanut production in Eastern India can be attributed to the Green Revolution’s emphasis on high-yielding varieties of wheat and rice, leading to reduced groundnut cultivation (Talawar, 2004). Additionally, we also simulated peanut crop in the regions with high rainfall and humidity and high incidences of pest and diseases where peanut is not grown – the Northern sub-Himalayan region (Janila et al., 2016; Rathnakumar et al., 2013). Pest and diseases are not captured by the current version of the model (see section 4.4.) but provide a hint of production potential provided the pest and disease resistant crop would be developed. These estimates might provide the basis to investigate the economic benefits related to integrating this cash crop into these systems.

4.4. Limitations of the study and directions for further development

There are several limitations that need to be considered while interpreting and, particularly, adopting the presented research;

1) Crop simulation modeling tools are only an imperfect reflection of reality biased by the gaps in our current knowledge. In our case, we need to consider that the modeling tool used in this study (SSM, Sinclair and Soltani, 2012a) does not include the algorithms to reproduce crop responses to extreme temperatures and salinity stresses nor responses to biotic stresses (also see Hajjarpoor et al 2021).
2) There is always a “dilemma of scales” involved when choosing the level of the model inputs appropriate to capturing the trends for target geographies (e.g. Kholová et al., 2020; Tardieu et al., 2020). For our study, we chose the sensible balance between the available data, data quality and resolution, and end-user needs (e.g., breeders/farm advisory services). Upon further end-user requests and data availability, the resolution of simulations might be increased for the specific geographies. Some of the analyses might also expand, for example, with seasonal forecasts (e.g., IMD forecast services).
3) This study identified Optimal GxM combinations based on past weather records. Ultimately, conducting similar analyses for future climatic scenarios (e.g., using GCM scenarios, Ruane et al., 2015) is important to understand the shifts in optimal GxM in changing climates.
4) For seamless use of the generated outputs by different type of end users, we plan to develop a common platform for generated data browsing and visualization (co-designed with a key pool of end users). This might, eventually, involve the actual simulation runs so the end-users could test their own specific hypotheses beyond the hereby simulated options.
5) We generated a similar simulation series not only for Kharif but also for rabi (post-rainy) and summer seasons. These should also be analyzed and will be important to advise farmers and guide breeders’ interventions beyond the Kharif season.

5. Conclusion

The presented peanut production system analysis was intended as a step toward context-specific system optimization and environment-specific breeding. We used an *in-silico* crop growth model (SSM) and generated large-scale simulations to test the suitability of factorial combination of crop type and crop management practices (G and M; 2.3 billion simulations, 1.02 TB output data) for peanut production systems with different irrigation management across India (resolution 0.5°x0.5°). This necessitated the use of advanced HPC and data management/analytical methods. For this, we developed a re-usable HPC system with interoperable modules that can be readily adapted for different simulation setups. Up to our knowledge, this is the first introduction of the “Index of Goodness” (IoG, grain and haulm production and production failure risk) – a simple index to evaluate the simulated GxM options from the perspective anticipated by end-users (primary producers and crop improvement programs). From the generated output, we identified environmental clusters (ECs) for systems with different irrigation management. Within each EC we identified a specific suite of GxM to optimize production and prioritized G targets for breeding. Our analysis highlighted the importance of crop duration and vigor traits in development of ideotypes for irrigated systems, while prioritizing ideotypes that show and early response to soil and atmospheric drought for water-limited systems. These, for the first time, provide quantitative, context specific GxM options for primary producers and phenotyping targets to support context-specific peanut breeding. Efforts are underway to develop the means to better support the end-users needs by *in-silico* system analyses.

CRediT authorship contribution statement

Amir Hajjarpoor: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis. **Jan Pavlík:** Writing – review & editing, Writing – original draft, Visualization, Software, Formal analysis, Data curation. **Jan Hora:** Software, Methodology. **Jakub Konopásek:** Writing – review & editing, Resources. **Janila Pusupuleti:** Investigation, Writing – review & editing. **Vincent Vadez:** Writing – review & editing, Validation. **Afshin Soltani:** Writing – review & editing, Validation, Software. **Til Feike:** Writing – review & editing, Validation. **Michal Stočes:** Writing – review & editing, Resources. **Jan Jarolímek:** Writing – review & editing, Resources. **Jana Kholová:** Writing – review & editing, Writing – original draft, Supervision, Conceptualization.

Funding sources

The results and knowledge included herein have been obtained owing to support from the following institutional grant. Internal grant agency of the Faculty of Economics and Management,Czech University of Life Sciences Prague, grant no. IGA 2022B0006 – “Life Sciences 4.0. Plus”. Part of the time the first author spent on this paper was dedicated to the Make Our Planet Great Again (MOPGA) ICARUS project (Improve Crops in Arid Regions and Future Climates) as well as SORGHUM project; the methods developed here will be beneficial for these projects. The project "SORGHUM" is funded by the Federal Ministry of Food and Agriculture as part of the German Climate Protection Programme 2022.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

We thank the national experts from the ICAR network in India for their valuable contributions. Additionally, we deeply appreciate the Crop Physiology Laboratory and Groundnut Breeding teams at ICRISAT for their consultancy and data sharing, which were instrumental to our research. The first author would especially like to thank the CERES Team at IRD, France, for providing support and hospitality.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compag.2025.110383.

Data availability

We uploaded the source data to the Zenodo research repository with the DOI 10.5281/zenodo.14699852 to be reused. The pipeline module coding, for the time being, will be enabled upon request.

References

Abshenas, M., Kamkar, B., Soltani, A., Kazemi, H., 2022. Predicting the effects of climate change on physiological parameters determining wheat yield in 2050 (case study: Golestan Province, Iran). Environ. Monit. Assess. 194. https://doi.org/10.1007/s10661-022-10428-7.

Battisti, R., Sentelhas, P.C., 2019. Characterizing Brazilian soybean-growing regions by water deficit patterns. Field Crop. Res. 240, 95–105. https://doi.org/10.1016/j.fcr.2019.06.007.

Blümmel, M., Ratnakumar, P., Vadez, V., 2012. Opportunities for exploiting variations in haulm fodder traits of intermittent drought tolerant lines in a reference collection of groundnut (Arachis hypogaea L.). Field Crop. Res. 126, 200–206. https://doi.org/10.1016/j.fcr.2011.10.004.

Boote, K.J., Jones, J.W., Hoogenboom, G., 2021. Incorporating realistic trait physiology into crop growth models to support genetic improvement. in silico Plants 3. Doi: 10.1093/insilicoplants/diab002.

Carcedo, A.J.P., Mayor, L., Demarco, P., Morris, G.P., Lingenfelser, J., Messina, C.D., Ciampitti, I.A., Heinemann, A.B., Kantar, M.B., 2022. Environment Characterization in Sorghum (Sorghum bicolor L .) by Modeling Water-Deficit and Heat Patterns in the Great Plains Region , United States 13, 1–13. Doi: 10.3389/fpls.2022.768610.

Chenu, K., 2015. Characterizing the crop environment - nature, significance and applications. In: Crop Physiology: Applications for Genetic Improvement and Agronomy, Second Edition. Academic Press, pp. 321–348. https://doi.org/10.1016/B978-0-12-417104-6.00013-3.

Chenu, K., Cooper, M., Hammer, G.L., Mathews, K.L., Dreccer, M.F., Chapman, S.C., 2011. Environment characterization as an aid to wheat improvement: interpreting genotype-environment interactions by modelling water-deficit patterns in North-Eastern Australia. J. Exp. Bot. 62, 1743–1755. https://doi.org/10.1093/jxb/erq459.

Chenu, K., Van Oosterom, E.J., Mclean, G., Deifel, K.S., Fletcher, A., Geetika, G., Tirfessa, A., Mace, E.S., Jordan, D.R., Sulman, R., Hammer, G.L., 2018. Integrating modelling and phenotyping approaches to identify and screen complex traits: transpiration efficiency in cereals. J. Exp. Bot. 69, 3181–3194. https://doi.org/10.1093/jxb/ery059.

Collins, B., Chenu, K., 2021. Improving productivity of Australian wheat by adapting sowing date and genotype phenology to future climate. Clim. Risk Manag. 32, 100300. https://doi.org/10.1016/j.crm.2021.100300.

Cooper, M., Messina, C.D., Tang, T., Gho, C., Powell, O., Podlich, M., Dean, W., Technow, F., Hammer, G.L., 2023. Predicting Genotype × Environment × Management (G × E × M) Interactions for the Design of Crop Improvement Strategies: Integrating Breeder, Agronomist, and Farmer Perspectives. Plant Breed. Revi.

Cooper, M., Tang, T., Gho, C., Hart, T., Hammer, G., Messina, C., 2020. Integrating genetic gain and gap analysis to predict improvements in crop productivity. Crop Sci. 1–23. https://doi.org/10.1002/csc2.20109.

Devi, M.J., Sinclair, T.R., Vadez, V., Krishnamurthy, L., 2009. Peanut genotypic variation in transpiration efficiency and decreased transpiration during progressive soil drying. Field Crop. Res. 114, 280–285. https://doi.org/10.1016/j.fcr.2009.08.012.

Diancoumba, M., Kholová, J., Adam, M., Famanta, M., Clerget, B., Traore, P.C.S., Weltzien, E., Vacksmann, M., Mclean, G., Hammer, G.L., 2024. APSIM - based modeling approach to understand sorghum production environments in Mali. Doi: 10.1007/s13593-023-00909-5.

DLD, 2017. District Level Database [WWW Document]. URL http://data.icrisat.org/dld/index.html (accessed 5.15.22).

Domhoefer, M., Chakraborty, D., Hufnagel, E., Claußen, J., Wörlein, N., Voorhaar, M., Anbazhagan, K., Choudhary, S., Pasupuleti, J., Baddam, R., Kholova, J., Gerth, S., 2022. X-ray driven peanut trait estimation: computer vision aided agri-system transformation. Plant Methods 18. https://doi.org/10.1186/s13007-022-00909-8.

Fainges, J.L., 2015. Using APSIM , C # and R to Create and Analyse Large Datasets, in: MODSIM2015, 21st International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand. Queensland, Australia, pp. 333–339. Doi: 10.36334/MODSIM.2015.B1.Fainges.

George, T., 2014. Why crop yields in developing countries have not kept pace with advances in agronomy. Glob. Food Sec. 3, 49–58. https://doi.org/10.1016/j.gfs.2013.10.002.

Hajjarpoor, A., Kholová, J., Pasupuleti, J., Soltani, A., Burridge, J., Degala, S.B., Gattu, S., Murali, T.V., Garin, V., Radhakrishnan, T., Vadez, V., 2021. Environmental characterization and yield gap analysis to tackle genotype-by-environment-by-management interactions and map region-specific agronomic and breeding targets in groundnut. Field Crop Res. 267, 108160. https://doi.org/10.1016/j.fcr.2021.108160.

Hajjarpoor, A., Nelson, W.C.D., Vadez, V., 2022. How process-based modeling can help plant breeding deal with G x E x M interactions. Field Crop Res. 283, 108554. https://doi.org/10.1016/j.fcr.2022.108554.

Hajjarpoor, A., Vadez, V., Soltani, A., Gaur, P., Whitbread, A., Suresh Babu, D., Gumma, M.K., Diancoumba, M., Kholová, J., 2018. Characterization of the main chickpea cropping systems in India using a yield gap analysis approach. Field Crop Res. 223, 93–104. https://doi.org/10.1016/j.fcr.2018.03.023.

Halilou, O., Hissene, H.M., Clavijo Michelangeli, J.A., Hamidou, F., Sinclair, T.R., Soltani, A., Mahamane, S., Vadez, V., 2016. Determination of coefficient defining leaf area development in different genotypes, plant types and planting densities in peanut (Arachis hypogeae L.). Field Crop Res. 199, 42–51. https://doi.org/10.1016/j.fcr.2016.09.013.

Hammer, G.L., McLean, G., Chapman, S., Zheng, B., Doherty, A., Harrison, M.T., Van Oosterom, E., Jordan, D., 2014. Crop design for specific adaptation in variable dryland production environments. Crop. Pasture Sci. 65, 614–626. https://doi.org/10.1071/CP14088.

Hammer, G.L., van Oosterom, E., McLean, G., Chapman, S.C., Broad, I., Harland, P., Muchow, R.C., 2010. Adapting APSIM to model the physiology and genetics of complex adaptive traits in field crops. J. Exp. Bot. 61, 2185–2202. https://doi.org/10.1093/jxb/erq095.

Harrison, M.T., Tardieu, F., Dong, Z., Messina, C.D., Hammer, G.L., 2014. Characterizing drought stress and trait influence on maize yield under current and future conditions. Glob. Chang. Biol. 20, 867–878. https://doi.org/10.1111/gcb.12381.

Heinemann, A.B., Barrios-Perez, C., Ramirez-Villegas, J., Arango-Londoño, D., Bonilla-Findji, O., Medeiros, J.C., Jarvis, A., 2015. Variation and impact of drought-stress patterns across upland rice target population of environments in Brazil. J. Exp. Bot. 66, 3625–3638. https://doi.org/10.1093/jxb/erv126.

Heinemann, A.B., Costa-Neto, G., da Matta, D.H., Fernandes, I.K., Stone, L.F., 2024. Harnessing crop models and machine learning for a spatial-temporal characterization of irrigated rice breeding environments in Brazil. Field Crop Res. 315, 109452. https://doi.org/10.1016/j.fcr.2024.109452.

Heinemann, A.B., Ramirez-Villegas, J., Rebolledo, M.C., Costa Neto, G.M.F., Castro, A.P., 2019. Upland rice breeding led to increased drought sensitivity in Brazil. Field Crop Res. 231, 57–67. https://doi.org/10.1016/j.fcr.2018.11.009.

Holzworth, D., Huth, N.I., Fainges, J., Brown, H., Zurcher, E., Cichota, R., Verrall, S., Herrmann, N.I., Zheng, B., Snow, V., 2018. APSIM Next Generation: Overcoming challenges in modernising a farming systems model. Environ. Model. Softw. 103, 43–51. https://doi.org/10.1016/j.envsoft.2018.02.002.

Holzworth, D.P., Huth, N.I., DeVoil, P.G., Zurcher, E.J., Herrmann, N.I., McLean, G., Chenu, K., van Oosterom, E.J., Snow, V., Murphy, C., Moore, A.D., Brown, H., Whish, J.P.M., Verrall, S., Fainges, J., Bell, L.W., Peake, A.S., Poulton, P.L., Hochman, Z., Thorburn, P.J., Gaydon, D.S., Dalgliesh, N.P., Rodriguez, D., Cox, H., Chapman, S., Doherty, A., Teixeira, E., Sharp, J., Cichota, R., Vogeler, I., Li, F.Y., Wang, E., Hammer, G.L., Robertson, M.J., Dimes, J.P., Whitbread, A.M., Hunt, J., van Rees, H., McClelland, T., Carberry, P.S., Hargreaves, J.N.G., MacLeod, N., McDonald, C., Harsdorf, J., Wedgwood, S., Keating, B.A., 2014. APSIM – Evolution towards a new generation of agricultural systems simulation. Environ. Model. Softw. 62, 327–350. https://doi.org/10.1016/j.envsoft.2014.07.009.

ICAR-DGR, 2016. Annual Report 2015-16, ICAR-Directorate of Groundnut Research. Junagadh, 362 001, Gujarat, India.

Jang, W.S., Lee, Y., Neff, J.C., Im, Y., Ha, S., Doro, L., 2019. Development of an EPIC parallel computing framework to facilitate regional/global gridded crop modeling with multiple scenarios: a case study of the United States. Comput. Electron. Agric. 158, 189–200. https://doi.org/10.1016/j.compag.2019.02.004.

Janila, P., Variath, M.T., Pandey, M.K., Desmae, H., Motagi, B.N., Okori, P., Manohar, S. S., Rathnakumar, A.L., Radhakrishnan, T., Liao, B., Varshney, R.K., 2016. Genomic Tools in Groundnut Breeding Program: Status and Perspectives. Front. Plant Sci. 7. https://doi.org/10.3389/fpls.2016.00289.

Jones, P.G., Thornton, P.K., 2000. MarkSim: software to generate daily weather data for Latin America and Africa. Agron. J. 92, 445–453.

Jones, P.G., Thornton, P.K., Diaz, W., Wilkens, P.W., Jones, A.L., 2002. MarkSim: A computer tool that generates simulated weather data for crop modeling and risk assessment. Centro Internacional de Agricultura Tropical, Cali, Colombia.

Kaloki, P., Luo, Q., Trethowan, R., Tan, D.K.Y., 2019. Can the development of drought tolerant ideotype sustain Australian chickpea yield? Int. J. Biometeorol. 63, 393–403. https://doi.org/10.1007/s00484-019-01672-7.

Kaufman, L., Rousseeuw, P.J., 1990. Partitioning Around Medoids (Program PAM). In: Finding Groups in Data: an Introduction to Cluster Analysis. John Wiley & Sons Inc, pp. 68–125.

Khabarov, N., Smirnov, A., Balkovič, J., Skalský, R., Folberth, C., Van Der Velde, M., Obersteiner, M., 2020. Heterogeneous compute clusters and massive environmental simulations based on the EPIC model. Modelling 1, 215–224. https://doi.org/10.3390/modelling1020013.

Kholová, J., Adam, M., Diancoumba, M., Marrou, H., Hammer, G., Hajjarpoor, A., Chenu, K., 2020. Sorghum; General crop modelling tools guiding principles and use of crop

models in support of crop improvement programs in developing countries., in: Tonapi, V.A., Talwar, H.S., Kumar, A., Bhat, B.V., Reddy, C.R., Dalton, T.J. (Eds.), Sorghum in the 21st Century: Food – Fodder – Feed – Fuel for a Rapidly Changing World. Springer Singapore, pp. 189–207. Doi: 10.1007/978-981-15-8249-3.
Kholová, J., Hash, C.T., Kakkera, A., Koová, M., Vadez, V., 2010a. Constitutive water-conserving mechanisms are correlated with the terminal drought tolerance of pearl millet [Pennisetum glaucum (L.) R. Br.]. J. Exp. Bot. 61, 369–377. https://doi.org/10.1093/jxb/erp314.
Kholová, J., Hash, C.T., Kumar, P.L., Yadav, R.S., Koová, M., Vadez, V., 2010b. Terminal drought-tolerant pearl millet [Pennisetum glaucum (L.) R. Br.] have high leaf ABA and limit transpiration at high vapour pressure deficit. J. Exp. Bot. 61, 1431–1440. https://doi.org/10.1093/jxb/erq013.
Kholová, J., McLean, G., Vadez, V., Craufurd, P., Hammer, G.L., 2013. Drought stress characterization of post-rainy season (rabi) sorghum in India. Field Crop Res. 141, 38–46. https://doi.org/10.1016/j.fcr.2012.10.020.
Kholova, J., Murugesan, T., Kaliamoorthy, S., Malayee, S., Baddam, R., Hammer, G.L., McLean, G., Deshpande, S., Hash, C.T., Craufurd, P.Q., Vadez, V., 2014. Modelling the effect of plant water use traits on yield and stay-green expression in sorghum. Funct. Plant Biol. 41, 1019. https://doi.org/10.1071/fp13355.
Kuhn, M., 2008. Building Predictive Models in R Using the caret Package. J. Stat. Softw. 28, 1–26. https://doi.org/10.18637/jss.v028.i05.
Laidig, F., Feike, T., Lichthardt, C., Schierholt, A., Piepho, H.P., 2024. Breeding progress of nitrogen use efficiency of cereal crops, winter oilseed rape and peas in long-term variety trials. Theor. Appl. Genet. 137, 1–21. https://doi.org/10.1007/s00122-023-04521-9.
Montañana, J.M., Marangio, P., Hervás, A., 2020. Agris on-line Papers in Economics and Informatics Open Source Framework for Enabling HPC and Cloud Geoprocessing Services XII, 61–76. Doi: 10.7160/aol.2020.120405.Introduction.
Nautiyal, P.C., Mejia, D., 2002. Groundnut: post-harvest operations.
Parthasarathy, B., Munot, A.A., Kothawale, D.R., 1994. All-India monthly and seasonal rainfall series: 1871?1993. Theor. Appl. Climatol. 49, 217–224. https://doi.org/10.1007/BF00867461.
Pavlik, J., Hajjarpoor, A., Kholova, J., 2025. India - Peanut production simulations - SSM. Doi: 10.5281/zenodo.14699852.
Rachaputi, R., Chauhan, Y.S., Wright, G.C., 2021. Chapter 11 - Peanut, in: Crop Physiology Case Histories for Major Crops. Elsevier Inc., pp. 360–382. Doi: 10.1016/B978-0-12-819194-1.00011-6.
Raju, B.M.K., Kumar, R.N., Rao, C.A.R., Samuel, J., Rao, K. V, Rao, A.V.M.S., Osman, M., Swapna, N., 2018. Unreaped Yield Potentials in Major Rainfed Crops and Scope for Bridging Yield Gaps-A Decision Support System. ICAR-Central Research Institute for Dryland Agriculture, Santoshnagar, Hyderabad-500059.
Ramirez-Villegas, J., Molero Milan, A., Alexandrov, N., Asseng, S., Challinor, A.J., Crossa, J., 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., Vadez, V., White, J.W., Reynolds, M., 2020. CGIAR modeling approaches for resource-constrained scenarios: I. Accelerating crop breeding for a changing climate. Crop Sci. 60, 547–567. https://doi.org/10.1002/csc2.20048.
Rathnakumar, A.L., Singh, R., Parmar, D.L., Misra, J.B., 2013. Groundnut a crop profile and compendium of notified varieties of India.
Rathore, V.S., Nathawat, N.S., Bhardwaj, S., Yadav, B.M., Kumar, M., Santra, P., Kumar, P., Reager, M.L., Yadava, N.D., Yadav, O.P., 2021. Optimization of deficit irrigation and nitrogen fertilizer management for peanut production in an arid region. Sci. Rep. 11. https://doi.org/10.1038/s41598-021-82968-w.
Riedesel, L., Möller, M., Piepho, H.P., Rentel, D., Lichthardt, C., Golla, B., Kautz, T., Feike, T., 2024. Site conditions determine heat and drought induced yield losses in wheat and rye in Germany. Environ. Res. Lett. 19. https://doi.org/10.1088/1748-9326/ad24d0.
Ronanki, S., Pavlík, J., Masner, J., Jarolímek, J., Stočes, M., Subhash, D., Talwar, H.S., Tonapi, V.A., Srikanth, M., Baddam, R., Kholová, J., 2022. An APSIM-powered framework for post-rainy sorghum-system design in India. Field Crop Res. 277, 108422. https://doi.org/10.1016/j.fcr.2021.108422.
Ruane, A.C., Goldberg, R., Chryssanthacopoulos, J., 2015. Climate forcing datasets for agricultural modeling: Merged products for gap-filling and historical climate series estimation. Agric. For. Meteorol. 200, 233–248. https://doi.org/10.1016/j.agrformet.2014.09.016.
Shawon, A.R., Memic, E., Kottmann, L., Uptmoor, R., Hackauf, B., Feike, T., 2024. Comprehensive evaluation of the DSSAT-CSM-CERES-Wheat for simulating winter rye against multi-environment data in Germany. Agron. J. 1844–1868. https://doi.org/10.1002/agj2.21590.
Sinclair, T.R., Hammer, G.L., van Oosterom, E.J., 2005. Potential yield and water-use efficiency benefits in sorghum from limited maximum transpiration rate. Funct. Plant Biol. 32, 945–952.
Sinclair, T.R., Messina, C.D., Beatty, A., Samples, M., 2010. Assessment across the United States of the Benefits of Altered Soybean Drought Traits. Agron. J. 102, 475–482. https://doi.org/10.2134/agronj2009.0195.
Sinclair, T.R., Soltani, A., Marrou, H., Ghanem, M., Vadez, V., 2020. Geospatial assessment for crop physiological and management improvements with examples using the simple simulation model. Crop Sci. https://doi.org/10.1002/csc2.20106.
Sivasakthi, K., Thudi, M., Tharanya, M., Kale, S.M., Kholová, J., Halime, M.H., Jaganathan, D., Baddam, R., Thirunalasundari, T., Gaur, P.M., Varshney, R.K., Vadez, V., 2018. Plant vigour QTLs co-map with an earlier reported QTL hotspot for drought tolerance while water saving QTLs map in other regions of the chickpea genome. BMC Plant Biol. 18. https://doi.org/10.1186/s12870-018-1245-1.
Soltani, A., Sinclair, T.R., 2012a. Modeling physiology of crop development, growth and yield. CABI, Wallingford. Doi: 10.1079/9781845939700.0000.
Soltani, A., Sinclair, T.R., 2012b. Identifying plant traits to increase chickpea yield in water-limited environments. Field Crop Res. 133, 186–196. https://doi.org/10.1016/j.fcr.2012.04.006.
Soltani, A., Sinclair, T.R., 2011. A simple model for chickpea development, growth and yield. Field Crop Res. 124, 252–260. https://doi.org/10.1016/j.fcr.2011.06.021.
Talawar, S., 2004. Peanut in India: History, production and utilization. Peanuts in local and Global Food Systems Series Report.
Tardieu, F., Granato, I.S.C., Van Oosterom, E.J., Parent, B., Hammer, G.L., 2020. Are crop and detailed physiological models equally 'mechanistic' for predicting the genetic variability of whole-plant behaviour? The nexus between mechanisms and adaptive strategies. in silico Plants 2. Doi: 10.1093/insilicoplants/diaa011.
Trivedi, T.P., 2011. Handbook of Agriculture. Directorate of Knowledge Management in Agriculture, Indian Council of Agricultural Research, 6th Revise. ed. New Delhi, India.
UN, 2024. The sustainable development goals report 2024. United Nations.
Vadez, V., Halilou, O., Hissene, H.M., Sibiry-Traore, P., Sinclair, T.R., Soltani, A., 2017. Mapping Water Stress Incidence and Intensity, Optimal Plant Populations, and Cultivar Duration for African Groundnut Productivity Enhancement. Front. Plant Sci. 8, 432. https://doi.org/10.3389/fpls.2017.00432.
Vadez, V., Soltani, A., Sinclair, T.R., 2012. Modelling possible benefits of root related traits to enhance terminal drought adaptation of chickpea. Field Crop Res. 137, 108–115. https://doi.org/10.1016/j.fcr.2012.07.022.
Welcker, C., Spencer, N.A., Turc, O., Granato, I., Chapuis, R., Madur, D., Beauchene, K., Gouesnard, B., Draye, X., Palaffre, C., Lorgeou, J., Melkior, S., Guillaume, C., Presterl, T., Murigneux, A., Wisser, R.J., Millet, E.J., van Eeuwijk, F., Charcosset, A., Tardieu, F., 2022. Physiological adaptive traits are a potential allele reservoir for maize genetic progress under challenging conditions. Nat. Commun. 13, 1–13. https://doi.org/10.1038/s41467-022-30872-w.
Witcombe, J.R., Virk, D.S., Farrington, J., 1998. Seeds of choice: Making the most of new varieties for small farmers.
Wu, A., Hammer, G.L., Doherty, A., Von Caemmerer, S., Farquhar, G.D., 2019. Quantifying impacts of enhancing photosynthesis on crop yield. Nat. Plants 5, 380–388. https://doi.org/10.1038/s41477-019-0398-8.
Wu, A., Song, Y., Van Oosterom, E.J., Hammer, G.L., 2016. Connecting biochemical photosynthesis models with crop models to support crop improvement. Front. Plant Sci. 7, 1–16. https://doi.org/10.3389/fpls.2016.01518.
Zheng, B., Chapman, S., Chenu, K., 2018. The value of tactical adaptation to El Niño-Southern Oscillation for East Australian wheat. Climate 6, 1–15. https://doi.org/10.3390/cli6030077.
Zheng, B., Holland, E., Chapman, S.C., 2016. A standardized workflow to utilise a grid-computing system through advanced message queuing protocols. Environ. Model. Softw. 84, 304–310. https://doi.org/10.1016/j.envsoft.2016.07.012.