Incorporating drought and marker associated traits into the DSSAT model

by

Xiaoxing Zhen

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
May 7, 2022

Keywords: peanut, drought tolerance, maize, marker effects, yield, climate change

Approved by

William David Batchelor, Chair, Professor of Biosystems Engineering
Alvaro Sanz-Saez, Assistant Professor of Crop, Soil and Environmental Sciences
Charles Y. Chen, Professor of Crop, Soil, and Environmental Sciences
Jasmeet Singh Lamba, Associate Professor of Biosystems Engineering

Abstract

Most agronomic traits are genetically complex and often show genotype × environment (G × E) interactions which complicate selection and slows down the breeding progress. As model parameters can represent certain genetic characteristics, crop modeling has been commonly used to identify and test desirable traits, to evaluate genetic improvement, and to design the optimum ideotype adapted to future climates. In this work, the DSSAT model was modified to evaluate the potential benefits of drought-tolerant traits in peanut under water-limited conditions, to incorporate genomic information to compute maize model inputs using marker-based information and combined with representative spatial model inputs to evaluate drought tolerant peanut performance under climate change. In Chapter two, the trait of maintaining photosynthesis under water deficit was observed in rainout shelter experiments and incorporated into the DSSAT model as a new drought tolerance cultivar coefficient. This specific trait was shown to be an advantageous trait for peanut varieties, which produced higher simulated rainfed yield with enhanced seasonal evapotranspiration and grain water use efficiency, especially for dry seasons. In Chapter three, we extend this approach to simulate the performance of drought tolerant peanuts for several important peanut production counties in the Southeastern USA. Results showed that a single set of cultivar coefficients and soil parameters could be calibrated to simulate historic peanut growth duration and county-level yields reasonably well. The simulation for future climate change indicated that the rainfed yields will suffer from increasing daytime temperature and an irrigation strategy could potentially offset the heat and drought stress to main higher peanut production in the Southeastern USA. Finally, In Chapter four we developed a methodology to use marker-based prediction of maize model inputs to assess new hybrid performance for plant breeding, and quantitatively assess the effect of genes by explicitly

accounting for G × E interactions. These findings provided a promising insight into the use of crop model in drought-tolerant simulation, marker-based modelling, and regional scale simulation.

Acknowledgments

I would like to express my special thanks to my advisor, Dr. William David Batchelor, for his kindness, guidance and encouragement through my PhD studies in the past four years. Thanks to him, I have a multi-disciplinary research experience and a good committee. I am deeply grateful to my committee members: Dr. Alvaro Sanz-Saez, Dr. Charles Y. Chen, Dr. Jasmeet Singh Lamba for their great support to my research. I also would like to thank Dr. Yin Bao for agreeing to serve as the University Reader of my dissertation.

I would like to thank Dr. Phat Dang, Dr. Di Tian, Dr. Yingjie Xiao, Mr. Bernardo Chaves Cordoba for their great support to my experimental data and statistical analysis. I am also grateful to fellow graduate students, research associates, and visiting scholars from Dr. Batchelor's, Dr. Alvaro's, and Dr. Chen's lab, for their kindness assistance with my lab work and personal life. I am extremely grateful for China Scholarship Council and NIFA-USDA, for their financial support of my study and research in the United States.

Finally, I would like to thank my lovely girlfriend, Weige Huo, for her companionship and encouragement during my PhD study period in Auburn. I would also like to thank my beloved parents, Mr. Qinjian Zhen and Mrs. Liangping Li, for all their love, and support.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

MAS	marker-assisted selection
GS	genomic selection
QTL	quantitative trait locus
GP	genomic prediction
GIS	geographic information system
ET	evapotranspiration
WUE	water use efficiency
SLA	specific leaf area
HI	harvest index
RWC	relative water content
EV	E.V. Smith Research Center of Auburn University at Shorter
HL	Wiregrass Research and Extension Center of Auburn University at Headland
DA	National Peanut Research Laboratory at Dawson
DT	drought tolerant factor
SWFAC	water stress index
PG	daily photosynthesis
R^2	coefficient of determination
RMSE	root mean square error
NRMSE	normalized root mean square error
D value	Willmott's index of agreement
USDA-NASS	US Department of Agriculture-National Agricultural Statistics Service
RCP	representative concentration pathways scenarios

GeoSim	Geospatial Simulation
PEST	Parameter Estimation
PASW	percent available soil water
DUL	drained upper limit
LL	lower limit
RHRF	root hospitality reduction factor
SLPF	soil fertility factor
SRGF	soil root growth factor
MAPE	mean absolute percentage error
E	model efficiency
GWAS	genome-wide association study
CUBIC	Complete-diallel design plus Unbalanced Breeding-like Inter-Cross
HN	Xinxiang, Henan province
HB	Baoding, Hebei province
DTA	days to anthesis
KNPE	kernel number per ear
KWPE	kernel weight per ear
PLSR	partial-least-square regression
P1	degree days from emergence to end of juvenile phase
G2	potential kernel number
G3	potential kernel growth rate
PHINT	phyllochron interval

Chapter 1. Literature review

1.1 Review of literatures

1.1.1 Overview of Peanut and its Genetic Characteristics

Peanut (*Arachis hypogaea L.*) is an annual legume, with geocarpic fruits, and an indeterminate growth habit. It presumably was domesticated in South America ~6000 years ago and then widely distributed in the tropical and subtropical regions of Asia, Africa, and North America which are characterized by high temperature and erratic precipitation (Zhuang et al., 2019). Peanut belongs to the plant family *Fabaceae* and genus *Arachis* (Barkley et al., 2016). The species name, *hypogaea*, means under the earth, which indicates that peanut can flower above the ground but produce their seed below the ground (Holbrook and Stalker, 2003). The cultivated peanut is classified into two subspecies, *subsp hypogaea* and *subsp fastigiata*, based on the presence or absence of floral axes on the main stem (Varshney et al., 2017). Based on location of flowers on the plant, patterns of reproductive nodes on branches, numbers of trichomes and pod morphology, the *subsp hypogaea* is further divided into two botanic varieties, var *hirsute* and var *hypogaea*, while the subsp *fastigiata* is further divided into four botanic varieties, var *aequatoriana*, var *fasitigiata*, var *peruviana*, and var *vulgaris*, based on a range of morphological characteristics (Krapovickas and Gregory 1994).

The *Arachis* genus contains 81 species, mostly diploids (2n=2x=20), while cultivated peanut species are self-pollinating allotetraploid (AABB; 2n=4x=40). This allotetraploid (2n = 4x = 40) originated through the hybridization of two ancient diploid species, *Arachis duranensis* (AA genome, 2n = 2x = 20) and *Arachis ipaensis* (BB genome, 2n = 2x = 20),

followed by a rare spontaneous duplication of chromosomes (Stalker 1997; Barkley et al., 2016). Uniform growth stage descriptions were developed for peanut based on visually observed vegetative (V) and reproductive (R) growth. The V stages contain VE (emergence), V0 (cotyledons are flat and open at or slightly below the soil surface), V1 (one developed node with one fully unfolded tetrafoliolate leaf), and V(N) (one to N developed nodes on the main axis), which is named by the total number of nodes on the main axis. The R stage is determined by the visually observable growth stages including flowering, pegging, fruit development, and seed maturation. The R stages are R1 (beginning bloom), R2 (beginning peg), R3 (beginning pod), R4 (full pod), R5 (beginning seed), R6 (full seed), R7 (beginning maturity), R8 (harvest maturity), and R9 (over mature pod) (Boote 1982; Wang 2019).

In the U.S., cultivated peanut production and marketing has resulted in designation of four market classes which generally correspond to subspecific and varietal groups as follows: Runner (subsp. *hypogaea var. hypogaea*), Virginia (subsp. *hypogaea var. hypogaea*), Spanish (subsp. *fastigiata var. vulgaris*), and Valencia (subsp.*fastigiata var. fastigiata*) (Boipelo 2018). The Runner variety is mostly planted in Georgia, Alabama, Florida, and Mississippi, which accounts for 80% of total U.S. peanut production, while the Virginia variety is primarily planted in South Carolina, North Carolina and Virginia, which accounts for about 15% of total U.S. peanut production (National Peanut Board, 2020; Kumral 2019).

1.1.2 Drought Stress Affecting Peanut Physiology and Production

Drought stress is a major abiotic stress contributing to the reduced agricultural productivity and food safety worldwide (Kambiranda et al., 2011). Drought stress affects physiological processes at the molecular, cellular, and whole-plant levels, including

metabolism, stomatal conductance, photosynthesis, nitrogen fixation, mineral nutrition uptake, water relations, leaf expansion, grain development and yield (Tardieu et al., 2011).

Peanut water relations are affected by several important characteristics, such as relative water content, leaf water potential, stomatal conductance, transpiration rate, and canopy temperature (Reddy et al., 2003). Due to drought stress, leaf water potential, relative water content and rate of transpiration substantially decreased in leaves (Nayyar and Gupta 2006; Siddique et al., 2001; Wang 2019). Leaf photosynthesis is reduced by moisture stress due to reduced stomatal conductance and reductions in leaf area. As water stress increases, transpiration rate is reduced with closing stomata. Consequently, the entry of CO_2 is also reduced (Reddy et al., 2003). Nitrogen fixation by legumes is also reduced by drought stress caused by a reduction in leghemoglobin in nodules, specific nodule activity and number of arid regions. In addition, dry weight of nodules is significantly reduced in moisture stressed plants. Drought stress also delays nodule formation in legumes (Reddi and Reddy, 1995). There is considerable evidence that mineral nutrition (N, P and K) uptake of peanut is reduced by drought stress (Kulkarni et al., 1988).

Drought stress during flowering period directly reduced the rate of flower production (Kambiranda et al., 2011). Peg elongation, which is turgor dependent, is also delayed due to drought stress (Boote and Ketring, 1990). Adequate root zone moisture can keep pegs alive until the pegging zone moisture content is sufficient to allow penetration and initiation of pod development (Skelton and Shear, 1971). Pod and seed development is delayed by dry pegging zone soil, which leads to a large reduction in pod yield, and the reduction percentage also varies among different varieties (Kambiranda et al., 2011). Pod yield can be reduced by 17-

25% due to early season drought and 15-64% due to midseason droughts depending on the cultivar (Wright et al., 1991; Soler et al., 2013). Peanut may experience water stress during reproductive stages, such as pegging and pod development. This results in a drastic reduction in fruit quality and yield (Jogloy et al., 1996; Rucker et al., 1995; Wang, 2019). Drought stress can also reduce biomass, harvest index, water use efficiency, and specific leaf area, while it causes an increase in chlorophyll content and canopy temperature (Jongrungklang et al. 2008; Soler et al., 2013). In addition, Drought stress also influences crop production indirectly by influencing the growth of weeds and intensity of disease and pests (Wang, 2019).

1.1.3 Breeding Efforts for Drought Tolerant Traits

Drought tolerance is an important strategy for peanuts to survive during severe stress conditions (Ravi et al., 2011). Many drought-related traits for plants can reflect drought tolerant ability, including cell protection, avoidance via stomatal closure or reduced leaf area, maintenance of vegetative growth or photosynthesis, high water use efficiency, high nitrogen fixation, large root systems or reduced seed abortion rate under water deficit. These characteristics can be incorporated into peanut to develop cultivars adapted to drought (Parent and Tardieu, 2014; Hammer et al., 2020).

Traditional breeding has been the major avenue for providing modern drought-tolerant peanut varieties to farmers. Conventional breeding is based on phenotypic selection. Through extensive selection based largely on empirical field observations, breeders have been successful in creating high-yielding peanut varieties under water limited conditions (Yin et al., 2003). However, selection for drought tolerance can be time-consuming because drought tolerant traits are controlled by many genes which have been shown to interact with the

environment. Substantial genotype by environment (G × E) interactions impede the breeding process (Ravi et al., 2011). More rapid breeding progress may be achieved by using new genetic and genomic technologies (Varshney et al., 2013).

With advances in high-throughput sequencing platforms and molecular genetic techniques, marker-assisted selection (MAS) has become a common means of molecular breeding. MAS exceeds the conventional selection on handling drought related traits. In addition to administering complex traits, molecular breeding can accurately and efficiently introgress multiple favorable alleles (Varshney et al., 2009). However, MAS is suitable only for traits controlled by a few major genes. Most desired traits are affected by many genes, each with minor effects (Wang et al., 2018). Therefore, the application of MAS in breeding meets with challenges. As an upgraded form of MAS, genomic selection (GS) could be a powerful tool in crop breeding to improve the selection and prediction accuracy for complex quantitative traits (Ali et al., 2020). GS using genome-wide markers simultaneously as predictor variables have been effective solutions to predict performance of the candidates (Xiao et al., 2021). Because selection objectives are no longer limited to the traits controlled by a small number of major genes, GS is a promising quantitative genetic approach to breed drought tolerant peanut varieties (Wang et al., 2018).

1.1.4 Virtual Cultivars Incorporated into Crop Model

Dynamic process-oriented crop models that integrate physical and physiological processes of plant growth and development have been widely used (Jin et al., 2018; Huang et al., 2019). Crop model can be an efficient way to determine optimum traits by determining if a target trait is positive or negative to peanut production under long-term weather conditions

(Tardieu and Tuberosa 2010). To anticipate the long-term climate impact on a breeding program, virtual cultivars incorporating various drought tolerant traits were defined and simulated with process-based crop models to evaluate peanut yield performance under water-limited environments and the adaptation to climate change (Bogard et al., 2020). The desired traits included deeper root depth, maximum fraction of shoot dry matter diverted to root growth under water stress, reduced maximum transpiration rate, early stomata closure, drought tolerant nitrogen fixation, longer seed-filling duration (Sinclair et al., 2010; Singh et al., 2014; Battisti et al., 2017). These drought-tolerance simulations provide guidance on trait selection using hypothetical examples.

1.1.5 Maize and Breeding Efforts for Yield Improvement

Maize (*Zea mays L.*), which provides 42% of the global food calories consumed by humans, has exceeded rice to become the most prominent cereal crop (Luo et al., 2021). Maize is one of the diverse plant species containing tremendous variation. As a result, it is a perfect model crop for genetic and genomic studies (Luo et al., 2020). Breeding efforts have been focused on detecting maize functional genes with an expectation to accelerate genetic improvement. These uncovered functional genes and favorable alleles provided a firm basis for further improving yield through marker-assisted selection or genetic transformation of crops (Gu et al., 2014). Traditional quantitative trait locus (QTL) mapping in crops depends on synthetic population-based linkage analysis or natural population-based linkage disequilibrium analysis. However, it is not very successful due to the large population size, the requirement for high resolution linkage maps, and the chromosomes with low recombination events (Luo et al., 2021). An emerging approach, association mapping, is

becoming popular because it can overcome the low power of linkage analysis, facilitate the detection of marker–trait associations and create fine-mapping of chromosome segments with high recombination rates (Crossa et al., 2017). Association mapping has facilitated the genetic dissection of both simple quantitative traits (e.g., flowering date, kernel number) and complex quantitative traits (e.g., kernel weight, yield) (Luo et al., 2021).

1.1.6 Genomic Prediction and Its Statistical Models

Contrary to QTL and association mapping with limited detected significant loci, phenotype prediction of the crop performance using whole genome markers is promising for breeding advancements. Three statistical prediction models have been developed and tested based on field-measured phenotypic data, including genomic prediction (GP), integrated model, and gene-based modelling. Direct statistical GP models aim to construct a model for predicting phenotype using whole-genome markers jointly. Statistical approaches reported for GP models can be classified into three main categories: parametric methods (e.g., GBLUP, RR-BLUP, LASSO); semiparametric methods (e.g., RKHS); and machine learning methods (e.g., SVM, Random Forest). Genomic prediction (GP) has the potential to accelerate the breeding progress for drought tolerant traits, maintain genetic diversity, and improve complex traits with low heritability (Heslot et al., 2012). The limitation of GP models is the limitation of the explicit environmental inputs.

As crop model parameters can represent certain genetic characteristics, crop modeling has been considered a useful tool to assist breeding (Loomis et al., 1979; Whisler et al., 1986; Boote et al., 1996). Shorter et al. (1991) first proposed collaborative efforts between breeders, physiologists and modellers, using simple biological models as a framework to integrate

physiology with breeding and to evaluate adaptation of genotypes to target environments. In addition, Hammer et al. (1999) proposed that crop models based on physiologically sound mechanisms, can quantify and integrate crop responses to genetic, environmental, and management factors, and thus, have the potential to predict performance of individual genotypes in different managements and environments. Recognizing the potential for incorporating genetic information with process-based crop models, Technow et al. (2015) proposed an integrated statistical approach to combine GP modeling with crop modelling for predicting the grain yield of maize. Following the integrated approach, Onogi et al. (2016) directly linked an eco-physiological model for rice heading date with the GP model, and simultaneously inferred the model input parameters and genome-wide marker effects on the parameters. Oliveira et al. (2021) integrated a dynamic statistical gene-based module into the CSM-CROPGRO-Drybean model to accurately predict the time of first flower appearance. The predictions of these integrated models would be expected to be more accurate and robust than the direct statistical GP model. However, the integrated approach is statistically more challenging (Onogi et al., 2016).

1.1.7 Gene-based Modelling and Its Development

The first attempt to implement the concept of gene-based modelling was published by White and Hoogenboom (1996) in Genegro, a process-oriented model that incorporated the effects of seven genes affecting phenology, growth habit and seed size of common bean (Phaseolus vulgaris L.). They applied linear regression to estimate values of more than 20 model input traits from information about alleles (variants at a gene locus) of seven known genes in the cultivars studied. However, this approach requires extensive data on the genetic

makeup of cultivars, and such data are still not routinely available. Improvements of Genegro was further made included the simulation of the effects of temperature on photoperiod sensitivity regulated by the gene *Tip* and a new function to predict seed weight (Hoogenboom and White, 2003; Hoogenboom et al., 2004).

Photothermal models (e.g., Grimm et al., 1993) that predict time to flowering and flowering duration based on the genetic makeup of *E* loci were developed for soybean (Stewart et al., 2003; Cober et al., 2001; Upadhyay et al., 1994a; Summerfield et al., 1998). Similar studies were conducted in other important agronomic traits (Lark et al., 1995; Orf et al., 1999; Mansur et al., 1993, 1996). To simulate cold acclimation in cereals, Fowler et al. (1999) described a routine whose development was partially guided by information from molecular studies. Further efforts have been published using genetic information to guide the modeling of phenology in soybean (Stewart et al., 2003) and Arabidopsis thaliana (Welch et al., 2003). Messina et al. (2006) also developed a gene-based model for soybean, based on the CROPGRO-Soybean, which contributes toward linking the crop's genetic architecture and whole organism phenotypic expression. The model accurately predicted time to flowering and post-flowering development phases and yield. Based on the CROPGRO-Peanut, Singh et al. (2012) evaluated the genetic traits of peanut for improving productivity and adaptation to climate change in India. The results proposed that the CROPGRO-Peanut model could be used to evaluate the potential benefits of genetic traits to guide breeding of improved peanut varieties.

With recent advancements in high-throughput sequencing, more genome-wide dense molecular markers of various crops have been generated. Gene-based modelling has been

carried out in several crop species with those markers for both simple morphological and complex physiological and quantitative traits, using either measured or optimized parameters, such as, leaf elongation rate and kernel number in maize (Chenu et al., 2008; Amelong et al. 2015); flowering time in wheat (White et al., 2008); canopy cover dynamics and tuber bulking in potato (Khan et al., 2019a; 2019b); tomato fruit sugar concentration (Prudent et al., 2011); photosynthesis and transpiration efficiency in rice leaves (Gu et al., 2012a, 2012b).

1.1.8 U.S. Peanut Production under Climate Change

Within the United States, peanut production is distributed in the Southeastern Coastal Plains which is highly vulnerable to the changing seasonal temperatures and rainfall patterns. This region is characterized by high temperature and erratic precipitation. A significant increase in seasonal temperatures, precipitation anomalies and ambient carbon dioxide level under climate change could be detrimental to peanut production (Vara et al., 2003; Eck et al., 2020). Most of the peanuts in this region are grown under rainfed conditions with sandy or loamy sand soil that have a lower water-holding capacity. Even peanuts grown under irrigation may experience extreme drought and heat because of limited water supply or because irrigation water is applied in amounts and frequencies less than optimal for plant growth. Frequent drought and heat events in the future threaten peanut growth in these areas (Kambiranda et al., 2011).

Grid-based and point-based models, statistical regressions, and greenhouse experiments are four popular analytical methods used to assess the impacts of climate change on crop production (Zhao et al., 2017). With increased availability of spatial and temporal datasets including remotely sensed images, land cover maps, digital soil surveys, gridded weather

datasets, point-based crop growth models are increasingly being used to simulate the impact of climate change, management policies and weather disasters (Manivasagam and Rozenstein, 2020). There is spatial variability in weather, soil, genetics, and management at larger scales that must be represented in the model. Developing representative spatial model inputs would reduce the uncertainty of using the models to study impacts of management practices or climate change.

1.1.9 Regional Simulation with Crop Model and Statistical Learning Model

Increased availability of spatial and temporal datasets has facilitated the implementation and evaluation of regional model simulations (Thorp and Bronson, 2013). The integration of remotely sensed information into the crop models for regional yield estimation can be achieved through data assimilation methods. Many data assimilation algorithms have been developed to improve the accuracy of crop models, such as Kalman Filter (KF), Ensemble Kalman Filter (EnKF), Three-Dimensional Variational Data Assimilation (3DVAR), Four-Dimensional Variational Data Assimilation (4DVAR), Particle Filter (PF) and Hierarchical Bayesian Method (HBM) (Jin et al., 2018). The data assimilation method, which retrieves crop status variables, such as leaf area index (LAI), canopy cover (CC), and Evapotranspiration (ET) from remotely sensed data and uses these variables as inputs to recalibrate and optimize the grain yield simulation ability of the crop growth model (Zhuo et al., 2019). Other approaches are statistical regression-based methods which are commonly used for regional crop yield estimation using satellite data (Wall et al., 2008; Franch et al., 2015). They are based on empirical relationships between historic yields and reflectance-based vegetation indices such as the normalized difference vegetation index (NDVI),

maximum adjusted NDVI (MA-NDVI), enhanced vegetation index (EVI), green chlorophyll vegetation index (GCVI). These methods are typically straightforward to implement without requiring numerous inputs, making them flexible and scalable to new regions, even where field data are scarce (Nguy-Robertson., 2014; Wang et al., 2020).

Crop model integration with geographic information system (GIS) is a potential solution for regional crop model simulations. GIS-based modeling systems create a platform for spatial yield variation analysis that can be used to understand the impacts of spatial variability of weather, soil, and farmer management on yield in large regions (Manivasagam and Rozenstein, 2020). Crop yield estimations have used the GIS-based modeling methods on several scales: field scale (Thorp et al., 2013), regional scale (Resop et al., 2012), national scale (Lv et al., 2017) and global scale (Liu et al., 2007). These methods also show potential for applications in precision agriculture management such as irrigation assessment (McCarthy et al., 2010), drainage water management (Thorp et al., 2008b) and sustainable farm management practices (Rao et al., 2000).

Machine learning techniques, including decision tree, association rule mining, multivariate regression, and artificial neural networks, have been used for crop yield prediction. Machine learning models treat the model output (e.g., yield) as an implicit function of the input variables (e.g., soil, weather, and remotely sensing data), which can be a highly non-linear and complex function (Khaki and Wang, 2019). Deep learning techniques often achieve a better performance compared with traditional machine learning methods (Wang et al., 2020). Convolution Neural Networks (CNN), recurrent neural networks (RNNs), random forest (RF), deep fully connected neural networks (DFNN) and Long Short-Term

Memory (LSTM) networks which are the most popular models among deep learning techniques have been applied to crop yield estimation and forecasting (Khaki et al., 2020).

1.1.10 Challenges in Capturing Spatial Heterogeneity at Regional Scale

Regional-scale simulations cannot consider the level of precisely measured data collected at the field scale. Recent efforts have been focused on developing reasonable methods to capture spatial variability of weather data, crop varieties, soil types, and management practices to minimize errors in yield simulation (Jagtap and Jones, 2002). Model input sampling in either relatively homogeneous administration units (e.g., county) or spatial grid cells (e.g., soil grid) for large regions have been reported with assumptions that enough inputs of the spatially heterogeneous environment could be sampled to reduce aggregation error to an acceptable level (Hansen and Jones, 2000).

At a larger scale, there is spatial heterogeneity in weather, soil, management, genetics, insects, and diseases that must be represented in the model. Climate data are often extracted from local representative weather stations (Xiong et al., 2008; Wang et al., 2020), or gridded weather data from climate simulation or interpolated from weather station networks (Xiong et al., 2007). While planted crop varieties vary spatially, researchers often use a single commonly grown variety to represent many varieties in regional simulations (Jagtap and Jones, 2002). Soil characteristics also vary over the larger scale. For larger scale simulation, soil profile characteristics are often derived from national or global soil survey databases but only one or a few dominant soils are chosen to represent a larger spatial scale. Southworth et al. (2000) divided the midwestern USA into 10 agricultural areas based on climate, soils, land use, and agricultural practices. The predominant soil type in each area was selected to

represent each area. Crop management practices also vary spatially. Researchers often define a representative set of management practices to represent many management practices in the region (Xiong et al., 2008; Jagtap and Jones, 2002). Although higher resolution of model inputs can potentially reduce the aggregation error due to spatial heterogeneity, not all model input data (e.g., management) are available at regional scale and representative model inputs are generally applied uniformly within a subregion (Hansen and Jones, 2000).

1.2 Research objectives

The objectives of this research program are to:

Objective 1: Modify CROPGRO-Peanut model to simulate the drought tolerance of peanut varieties.

1. Estimate crop model genetic coefficients for several drought tolerant peanut varieties grown under both well-water and water-limited conditions.
2. Develop and evaluate a framework to simulate enhanced photosynthesis under drought in the model.
3. Evaluate the impact of this trait on peanut yield for different seasons.

Objective 2: Develop representative spatial model inputs for county-level peanut simulations in the Southeastern USA.

1. Calibrate input parameters for the CROPGRO-Peanut model using historical NASS peanut yields for five counties across the Southeastern USA.
2. Evaluate the calibrated baseline model using data from independent years.
3. Assess the potential effects of future climate change on peanut production and irrigation water use in the Southeastern USA.

Objective 3: Use GWAS and marker-based information to estimate CERES-Maize model inputs.

1. Identify associated genetic markers for essential model input parameters of the CERES-Maize model.

2.Estimate model parameters through conventional model optimization and marker-based statistical prediction.

3. Demonstrate potential application of the marker-based crop modelling as a breeding tool for studying the G × E interactions.

1.3 References

Amelong, A., Gambín, B.L., Severini, A.D. and Borrás, L., 2015. Predicting maize kernel number using QTL information. Field Crops Research, 172, pp.119-131.

Ali, M., Zhang, Y., Rasheed, A., Wang, J. and Zhang, L., 2020. Genomic prediction for grain yield and yield-related traits in chinese winter wheat. International journal of molecular sciences, 21(4), p.1342.

Boote, K. J. 1982. Growth stages of peanut (*Arachis hypogaea L.*). Peanut Science, 9(1), 35-40.

Boote, K.J. & Ketring, D.L. 1990. Peanut. In: Stewart B.A. And Nielson D.R. (Eds), Irrigation Of Agricultural Crops. Asa- Groundnut - A Global Perspective. International Crops Research CSSA-SSSA, Madison.

Boote, K.J., Jones, J.W. and Pickering, N.B., 1996. Potential uses and limitations of crop models. Agronomy journal, 88(5), pp.704-716.

Barkley, N. A., Upadhyaya, H. D., Liao, B., & Holbrook, C. C. 2016. Global resources of genetic diversity in peanut. In Peanuts (pp. 67-109). AOCS Press.

Battisti, R., Sentelhas, P.C., Boote, K.J., Camara, G.M.D.S., Farias, J.R. and Basso, C.J., 2017. Assessment of soybean yield with altered water-related genetic improvement traits under climate change in Southern Brazil. European Journal of Agronomy, 83, pp.1-14.

Boipelo P. N. 2018. Validation of candidate genes associated with leaf spot resistance in cultivated peanut (*Arachis hypogea L.*), Master thesis, Auburn University, Auburn, USA.

Bogard, M., Biddulph, B., Zheng, B., Hayden, M., Kuchel, H., Mullan, D., Allard, V., Gouis, J.L. and Chapman, S.C., 2020. Linking genetic maps and simulation to optimize breeding for wheat flowering time in current and future climates. Crop Science, 60(2), pp.678-699.

Cober, E.R. and Voldeng, H.D., 2001. A new soybean maturity and photoperiod-sensitivity locus linked to E1 and T. Crop Science, 41(3), pp.698-701.

Chenu K, Chapman SC, Hammer GL, McLean G, Salah HBH, Tardieu F. 2008. Short-term responses of leaf growth rate to water deficit scale up to whole-plant and crop levels: an integrated modelling approach in maize. Plant, Cell and Environment 31: 378–391.

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., De Los Campos, G., Burgueño, J., González-Camacho, J.M., Pérez-Elizalde, S., Beyene, Y. and Dreisigacker, S., 2017. Genomic selection in plant breeding: methods, models, and perspectives. Trends in plant science, 22(11), pp.961-975.

Eck, M.A., Murray, A.R., Ward, A.R. and Konrad, C.E., 2020. Influence of growing season temperature and precipitation anomalies on crop yield in the southeastern United States. Agricultural and Forest Meteorology, 291, p.108053.

Fowler, D.B., Limin, A.E. and Ritchie, J.T., 1999. Low-temperature tolerance in cereals: Model and genetic interpretation. Crop Science, 39(3), pp.626-633.

Franch, B., Vermote, E.F., Becker-Reshef, I., Claverie, M., Huang, J., Zhang, J., Justice, C. and Sobrino, J.A., 2015. Improving the timeliness of winter wheat production forecast in

the United States of America, Ukraine and China using MODIS data and NCAR Growing Degree Day information. Remote Sensing of Environment, 161, pp.131-148.

Grimm, S.S., Jones, J.W., Boote, K.J. and Hesketh, J.D., 1993. Parameter estimation for predicting flowering date of soybean cultivars. Crop Science, 33(1), pp.137-144.

Gu J,YinX, StruikPC, Stomph TJ,WangH. 2012a. Using chromosome introgression lines to map quantitative trait loci for photosynthesis parameters in rice (*Oryza sativa L.*) leaves under drought and well watered field conditions. Journal of Experimental Botany 63: 455–469.

Gu J, Yin X, Stomph TJ, Wang H, Struik PC. 2012b. Physiological basis of genetic variation in leaf photosynthesis among rice (*Oryza sativa L.*) introgression lines under drought and well-watered conditions. Journal of Experimental Botany 63: 5137–5153.

Gu, J., Yin, X., Zhang, C., Wang, H. and Struik, P.C., 2014. Linking ecophysiological modelling with quantitative genetics to support marker-assisted crop design for improved yields of rice (*Oryza sativa*) under drought stress. Annals of Botany, 114(3), pp.499-511.

Hammer, G.L., Chapman, S.C., Snell, P., 1999. Crop simulation modelling to improve selection efficiency in plant breeding programs. Proceedings of the ninth Assembly Wheat Breeding Society of Australia, Toowoomba, September 1999, pp. 79–85.

Hansen, J.W. and Jones, J.W., 2000. Scaling-up crop models for climate variability applications. Agricultural Systems, 65(1), pp.43-72.

Hoogenboom, G. and White, J.W., 2003. Improving physiological assumptions of simulation models by using gene-based approaches. Agronomy Journal, 95(1), pp.82-89.

Holbrook, C.C and Stalker, H.T. 2003. Peanut Breeding and Genetic Resources. Plant Breeding Reviews, Volume 22.

Hoogenboom, G., White, J.W. and Messina, C.D., 2004. From genome to crop: integration through simulation modeling. Field Crops Research, 90(1), pp.145-163.

Heslot, N., Yang, H.P., Sorrells, M.E. and Jannink, J.L., 2012. Genomic selection in plant breeding: a comparison of models. Crop Science, 52(1), pp.146-160.

Huang, J., Gómez-Dans, J.L., Huang, H., Ma, H., Wu, Q., Lewis, P.E., Liang, S., Chen, Z., Xue, J.H., Wu, Y. and Zhao, F., 2019. Assimilation of remote sensing into crop growth models: Current status and perspectives. Agricultural and Forest Meteorology, 276, p.107609.

Hammer, G.L., McLean, G., van Oosterom, E., Chapman, S., Zheng, B., Wu, A., Doherty, A. and Jordan, D., 2020. Designing crops for adaptation to the drought and high-temperature risks anticipated in future climates. Crop Science, 60(2), pp.605-621.

Jogloy, S., A. Patanothai, S. Toomsan and T. G. Isleib. 1996. Breeding peanut to fit into Thai cropping systems. Proc. of the Peanut Collaborative Research Support Program-International Research Symposium and Workshop, Two Jima Quality Inn, Arlington, Virginia, USA, 25-31 March, 1996: pp 353-362.

Jagtap, S.S. and Jones, J.W., 2002. Adaptation and evaluation of the CROPGRO-soybean model to predict regional yield and production. Agriculture, Ecosystems & Environment, 93(1-3), pp.73-85.

Jongrungklang, N., Toomsan, B., Vorasoot, N., Jogloy, S., Kesmala, T. and Patanothai, A., 2008. Identification of peanut genotypes with high water use efficiency under drought stress conditions from peanut germplasm of diverse origins. Asian Journal of Plant Sciences.

Jin, X., Kumar, L., Li, Z., Feng, H., Xu, X., Yang, G. and Wang, J., 2018. A review of data assimilation of remote sensing and crop models. European Journal of Agronomy, 92, pp.141-152.

Krapovickas A, Gregory WC. 1994. Taxonomía del género Arachis (*Leguminosae*). Bonplandia, pp.1-186.

H., Ravindra, V., Sojitra, V. K., and Bhatt, D. M. 1988. Growth, nodulation and Nuptake of groundnut (*Arachis hypogaea L.*) as influenced by water deficit stress at different phenophases. Oleagineux (France).

Kambiranda, D. M., Vasanthaiah, H. K., Katam, R., Ananga, A., Basha, S. M., and Naik, K. 2011. Impact of drought stress on peanut (*Arachis hypogaea L.*) productivity and food safety. Plants and Environment. In Tech Publisher.

Khaki, S. and Wang, L., 2019. Crop yield prediction using deep neural networks. Frontiers in plant science, 10, p.621.

Khan, M.S., Struik, P.C., van der Putten, P.E., Jansen, H.J., van Eck, H.J., van Eeuwijk, F.A. and Yin, X., 2019a. A model-based approach to analyse genetic variation in potato using standard cultivars and a segregating population. I. Canopy cover dynamics. Field Crops Research, 242, p.107581.

Khan, M.S., Yin, X., van der Putten, P.E., Jansen, H.J., van Eck, H.J., van Eeuwijk, F.A. and Struik, P.C., 2019b. A model-based approach to analyse genetic variation in potato using standard cultivars and a segregating population. II. Tuber bulking and resource use efficiency. Field Crops Research, 242, p.107582.

Kumral, F. E. 2019. Genome Wide Association Study (GWAS) on Root-Knot Nematode Resistance in Cultivated Peanut, Master thesis, Auburn University, Auburn, USA.

Khaki, S., Wang, L. and Archontoulis, S.V., 2020. A cnn-rnn framework for crop yield prediction. Frontiers in Plant Science, 10, p.1750.

Loomis, R.S., Rabbinge, R. and Ng, E., 1979. Explanatory models in crop physiology. Annual Review of Plant Physiology, 30(1), pp.339-367.

Lark, K.G., Chase, K., Adler, F., Mansur, L.M. and Orf, J.H., 1995. Interactions between quantitative trait loci in soybean in which trait variation at one locus is conditional upon a specific allele at another. Proceedings of the National Academy of Sciences, 92(10), pp.4656-4660.

Liu, J., Williams, J.R., Zehnder, A.J. and Yang, H., 2007. GEPIC–modelling wheat yield and crop water productivity with high resolution on a global scale. Agricultural systems, 94(2), pp.478-493.

Lv, Z., Liu, X., Cao, W. and Zhu, Y., 2017. A model-based estimate of regional wheat yield gaps and water use efficiency in main winter wheat production regions of China. Scientific reports, 7(1), pp.1-15.

Liu, H.J., Wang, X., Xiao, Y., Luo, J., Qiao, F., Yang, W., Zhang, R., Meng, Y., Sun, J., Yan, S. and Peng, Y., 2020. CUBIC: an atlas of genetic architecture promises directed maize improvement. Genome biology, 21(1), pp.1-17.

Luo, J., Wei, C., Liu, H., Cheng, S., Xiao, Y., Wang, X., Yan, J. and Liu, J., 2020. MaizeCUBIC: a comprehensive variation database for a maize synthetic population. Database, 2020.

Luo, Y., Zhang, M., Liu, Y., Liu, J., Li, W., Chen, G., Peng, Y., Jin, M., Wei, W., Jian, L. and Yan, J., 2021. Genetic variation in YIGE1 contributes to ear length and grain yield in maize. New Phytologist.

Mansur, L.M., Lark, K.G., Kross, H. and Oliveira, A., 1993. Interval mapping of quantitative trait loci for reproductive, morphological, and seed traits of soybean (*Glycine max L.*). Theoretical and Applied Genetics, 86(8), pp.907-913.

Mansur, L.M., Orf, J.H., Chase, K., Jarvik, T., Cregan, P.B. and Lark, K.G., 1996. Genetic mapping of agronomic traits using recombinant inbred lines of soybean. Crop Science, 36(5), pp.1327-1336.

Messina, C. D., Jones, J. W., Boote, K. J., & Vallejos, C. E. 2006. A gene-based model to simulate soybean development and yield responses to environment. Crop Science, 46(1), 456-466.

McCarthy, A.C., Hancock, N.H. and Raine, S.R., 2010. VARIwise: A general-purpose adaptive control simulation framework for spatially and temporally varied irrigation at sub-field scale. Computers and Electronics in Agriculture, 70(1), pp.117-128.

Manivasagam, V.S. and Rozenstein, O., 2020. Practices for upscaling crop simulation models from field scale to large regions. Computers and Electronics in Agriculture, 175, p.105554.

Nayyar, H., & Gupta, D. 2006. Differential sensitivity of C3 and C4 plants to water deficit stress: association with oxidative stress and antioxidants. Environmental and Experimental Botany, 58(1-3), 106-113.

Nguy-Robertson, A.L., Peng, Y., Gitelson, A.A., Arkebauer, T.J., Pimstein, A., Herrmann, I., Karnieli, A., Rundquist, D.C. and Bonfil, D.J., 2014. Estimating green LAI in four crops: Potential of determining optimal spectral bands for a universal algorithm. Agricultural and forest meteorology, 192, pp.140-148.

National Peanut Board. Peanut Country, USA. 2020. [Online]. Available at https://www.nationalpeanutboard.org/peanut-info/peanut-country-usa.htm

Orf, J.H., Chase, K., Adler, F.R., Mansur, L.M. and Lark, K.G., 1999. Genetics of soybean agronomic traits: II. Interactions between yield quantitative trait loci in soybean. Crop Science, 39(6), pp.1652-1657.

Onogi, A., Watanabe, M., Mochizuki, T., Hayashi, T., Nakagawa, H., Hasegawa, T. and Iwata, H., 2016. Toward integration of genomic selection with crop modelling: the development of an integrated approach to predicting rice heading dates. Theoretical and Applied Genetics, 129(4), pp.805-817.

Oliveira, F.A., Jones, J.W., Pavan, W., Bhakta, M., Vallejos, C.E., Correll, M.J., Boote, K.J., Fernandes, J.M., Hölbig, C.A. and Hoogenboom, G., 2021. Incorporating a dynamic gene-based process module into a crop simulation model. in silico Plants, 3(1), p.diab011.

Prudent, M., Lecomte, A., Bouchet, J.P., Bertin, N., Causse, M. and Génard, M., 2011. Combining ecophysiological modelling and quantitative trait locus analysis to identify key elementary processes underlying tomato fruit sugar concentration. Journal of Experimental Botany, 62(3), pp.907-919.

Parent, B. and Tardieu, F., 2014. Can current crop models be used in the phenotyping era for predicting the genetic variability of yield of plants subjected to drought or high temperature?. Journal of experimental botany, 65(21), pp.6179-6189.

Reddi, G. H. S., and Reddy, T. Y. 1995. Irrigation of principal crops. Efficient use of irrigation water. 2nd edition. Kalyani Publishers. New Delhi, India, 229-259.

Rucker, K.S., Kvien, C.K., Holbrook, C.C. and Hook, J.E., 1995. Identification of peanut genotypes with improved drought avoidance traits. Peanut Science, 22(1), pp.14-18.

Rao, M.N., Waits, D.A. and Neilsen, M.L., 2000. A GIS-based modeling approach for implementation of sustainable farm management practices. Environmental Modelling & Software, 15(8), pp.745-753.

Reddy, T. Y., Reddy, V. R., & Anbumozhi, V. 2003. Physiological responses of groundnut (*Arachis hypogea L.*) to drought stress and its amelioration: a critical review. Plant growth regulation, 41(1), 75-88.

Ravi, K., Vadez, V., Isobe, S., Mir, R.R., Guo, Y., Nigam, S.N., Gowda, M.V.C., Radhakrishnan, T., Bertioli, D.J., Knapp, S.J. and Varshney, R.K., 2011. Identification of several small main-effect QTLs and a large number of epistatic QTLs for drought tolerance related traits in groundnut (*Arachis hypogaea L.*). Theoretical and Applied Genetics, 122(6), pp.1119-1132.

Resop, J.P., Fleisher, D.H., Wang, Q., Timlin, D.J. and Reddy, V.R., 2012. Combining explanatory crop models with geospatial data for regional analyses of crop yield using field-scale modeling units. Computers and electronics in agriculture, 89, pp.51-61.

Skelton, B.J. & Shear, G.M. 1971. Calcium Translocation in the Peanut (*Arachis hypogae L.*). Agronomy Journal, Vol.63, pp. 409–412.

Shorter, R., Lawn, R.J. and Hammer, G.L., 1991. Improving genotypic adaptation in crops–a role for breeders, physiologists and modellers. Experimental Agriculture, 27(2), pp.155-175.

Stalker, H.T. 1997. Peanut (*Arachis Hypogea L*). Field Crops Research. 53(205-217)

Summerfield, R.J., Asumadu, H., Ellis, R.H. and Qi, A., 1998. Characterization of the photoperiodic response of post-flowering development in maturity isolines of soyabean [*Glycine max (L.) Merrill*]'Clark'. Annals of Botany, 82(6), pp.765-771.

Southworth, J., Randolph, J.C., Habeck, M., Doering, O.C., Pfeifer, R.A., Rao, D.G. and Johnston, J.J., 2000. Consequences of future climate change and changing climate variability on maize yields in the midwestern United States. Agriculture, Ecosystems & Environment, 82(1-3), pp.139-158.

Siddique, K. H. M., Regan, K. L., Tennant, D., & Thomson, B. D. 2001. Water use and water use efficiency of cool season grain legumes in low rainfall Mediterranean-type environments. European Journal of Agronomy, 15(4), 267-280.

Stewart, D.W., Cober, E.R. and Bernard, R.L., 2003. Modeling genetic effects on the photothermal response of soybean phenological development. Agronomy Journal, 95(1), pp.65-70.

Sinclair, T.R., Messina, C.D., Beatty, A. and Samples, M., 2010. Assessment across the United States of the benefits of altered soybean drought traits. Agronomy Journal, 102(2), pp.475-482.

Singh, P., Boote, K. J., Kumar, U., Srinivas, K., Nigam, S. N., & Jones, J. W. 2012. Evaluation of Genetic Traits for Improving Productivity and Adaptation of Groundnut to Climate Change in India. Journal of Agronomy and Crop Science, 198(5), 399-413.

Soler, C. M. T., Suleiman, A., Anothai, J., Flitcroft, I., & Hoogenboom, G. 2013. Scheduling irrigation with a dynamic crop growth model and determining the relation between simulated drought stress and yield for peanut. Irrigation science, 31(5), 889-901.

Singh, P., Nedumaran, S., Ntare, B.R., Boote, K.J., Singh, N.P., Srinivas, K. and Bantilan, M.C.S., 2014. Potential benefits of drought and heat tolerance in groundnut for adaptation to climate change in India and West Africa. Mitigation and adaptation strategies for global change, 19(5), pp.509-529.

Thorp, K.R., Jaynes, D.B. and Malone, R.W., 2008b. Simulating the long-term performance of drainage water management across the Midwestern United States. Transactions of the ASABE, 51(3), pp.961-976.

Tardieu, F. and Tuberosa, R., 2010. Dissection and modelling of abiotic stress tolerance in plants. Current opinion in plant biology, 13(2), pp.206-212.

Tardieu, F., Granier, C. and Muller, B., 2011. Water deficit and growth. Co-ordinating processes without an orchestrator? Current opinion in plant biology, 14(3), pp.283-289.

Thorp, K.R. and Bronson, K.F., 2013. A model-independent open-source geospatial tool for managing point-based environmental model simulations at multiple spatial locations. Environmental modelling & software, 50, pp.25-36.

Technow, F., Messina, C.D., Totir, L.R. and Cooper, M., 2015. Integrating crop growth models with whole genome prediction through approximate Bayesian computation. PloS one, 10(6), p.e0130855.

Upadhyay, A.P., R.H. Ellis, R.J. Summerfield, E.H. Roberts, and A. Qi. 1994a. Characterization of photothermal flowering responses in maturity isolines of soyabean [*Glycine max (L.) Merrill*] cv. Clark. Ann. Bot. (London) 74:87–96.

Vara Prasad, P.V., Boote, K.J., Hartwell Allen Jr, L. and Thomas, J.M., 2003. Super-optimal temperatures are detrimental to peanut (*Arachis hypogaea L.*) reproductive processes and yield at both ambient and elevated carbon dioxide. Global Change Biology, 9(12), pp.1775-1787.

Varshney, R. K., Nayak, S. N., May, G. D., & Jackson, S. A. 2009. Next-generation sequencing technologies and their implications for crop genetics and breeding. Trends in biotechnology, 27(9), 522-530.

Varshney, R.K., Mohan, S.M., Gaur, P.M., Gangarao, N.V.P.R., Pandey, M.K., Bohra, A., Sawargaonkar, S.L., Chitikineni, A., Kimurto, P.K., Janila, P. and Saxena, K.B., 2013. Achievements and prospects of genomics-assisted breeding in three legume crops of the semi-arid tropics. Biotechnology advances, 31(8), pp.1120-1134.

Varshney, R. K., Pandey, M. K., & Puppala, N. 2017. The Peanut Genome. Springer International Publishing.

Whisler, F.D., Acock, B., Baker, D.N., Fye, R.E., Hodges, H.F., Lambert, J.R., Lemmon, H.E., McKinion, J.M. and Reddy, V.R., 1986. Crop simulation models in agronomic systems. Advances in agronomy, 40, pp.141-208.

Wright, G.C., Hubick, K.T. and Farquhar, G.D., 1991. Physiological analysis of peanut cultivar response to timing and duration of drought stress. Australian Journal of Agricultural Research, 42(3), pp.453-470.

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

Welch, S.M., Roe, J.L. and Dong, Z., 2003. A genetic neural network model of flowering time control in Arabidopsis thaliana. Agronomy Journal, 95(1), pp.71-81.

Wall, L., Larocque, D. and Léger, P.M., 2008. The early explanatory power of NDVI in crop yield modelling. International Journal of Remote Sensing, 29(8), pp.2211-2225.

White, J.W., Herndl, M., Hunt, L.A., Payne, T.S. and Hoogenboom, G., 2008. Simulation-based analysis of effects of Vrn and Ppd loci on flowering in wheat.

Wang, X., Xu, Y., Hu, Z. and Xu, C., 2018. Genomic selection methods for crop improvement: Current status and prospects. The Crop Journal, 6(4), pp.330-340.

Wang X, 2019. Integration of physiological and molecular approaches for selecting peanut genotypes with superior drought tolerant and nitrogen fixation traits, Master thesis, Auburn University, Auburn, USA.

Wang, X., Huang, J., Feng, Q. and Yin, D., 2020. Winter wheat yield prediction at county level and uncertainty analysis in main wheat-producing regions of China with deep learning approaches. Remote Sensing, 12(11), p.1744.

Xiong, W., Matthews, R., Holman, I., Lin, E. and Xu, Y., 2007. Modelling China's potential maize production at regional scale under climate change. Climatic change, 85(3), pp.433-451.

Xiong, W., Holman, I., Conway, D., Lin, E. and Li, Y., 2008. A crop model cross calibration for use in regional climate impacts studies. Ecological Modelling, 213(3-4), pp.365-380.

Xiao, Y., Jiang, S., Cheng, Q., Wang, X., Yan, J., Zhang, R., Qiao, F., Ma, C., Luo, J., Li, W. and Liu, H., 2021. The genetic mechanism of heterosis utilization in maize improvement. Genome Biology, 22(1), pp.1-29.

Yin, X., Stam, P., Kropff, M. J., & Schapendonk, A. H. 2003. Crop modeling, QTL mapping, and their complementary role in plant breeding. Agronomy Journal, 95(1), 90-98.

Zhao, C., Liu, B., Piao, S., Wang, X., Lobell, D.B., Huang, Y., Huang, M., Yao, Y., Bassu, S., Ciais, P. and Durand, J.L., 2017. Temperature increase reduces global yields of major crops in four independent estimates. Proceedings of the National Academy of Sciences, 114(35), pp.9326-9331.

Zhuang, W., Chen, H., Yang, M., Wang, J., Pandey, M. K., Zhang, C., ... & Garg, V. 2019. The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. Nature genetics, 51(5), 865-876.

Zhuo, W., Huang, J., Li, L., Zhang, X., Ma, H., Gao, X., Huang, H., Xu, B. and Xiao, X., 2019. Assimilating soil moisture retrieved from Sentinel-1 and Sentinel-2 data into WOFOST model to improve winter wheat yield estimation. Remote Sensing, 11(13), p.1618.

Chapter 2. Simulating drought tolerance of peanut (*Arachis hypogaea*) varieties by maintaining photosynthesis under water deficit

2.1 Abstract

Over two-third of global peanuts are grown mainly in seasonally rainfed regions across arid and semi-arid zones where drought is a major yield limiting factor. Breeders are targeting drought adaptive traits by selecting high yielding genotypes under water-limited environments. Recently, several peanut varieties have been developed that exhibit drought tolerant characteristics. Crop models can be used to simulate the impact of these traits for different environments. The overall goal of this study was to develop an approach to simulate drought tolerant traits using the CROPGRO-Peanut model and assess the long-term yield response to these traits. Four peanut varieties and one advanced breeding line variety with varying degrees of drought tolerance response were grown under both field and rainout shelter conditions in 2019 and 2020. The trait of maintaining photosynthesis under water deficit was observed in the rainout shelter experiments and incorporated into the crop model as a new drought tolerance cultivar coefficient. The evaluation results with independent data showed that the modified model simulated peanut growth and yield under water-limited conditions reasonably well. The rainfed yield, seasonal evapotranspiration (ET), and grain water use efficiency (WUE) were simulated for both drought tolerant and baseline peanut varieties at two representative sites using weather data from 1998-2020. The drought tolerant mechanism of maintaining photosynthesis under water deficit was shown to be an advantageous trait for peanut varieties, which produced higher simulated rainfed yield with enhanced seasonal ET and grain WUE, especially for dry seasons. Using sensitivity analysis, the simulated

photosynthesis and yield were sensitive to values of the drought tolerant factor over an expected range of values. Further research is needed on other potential drought tolerant mechanisms, such as maintaining nitrogen fixation, pod harvest index and leaf growth under drought.

Keywords: *Arachis hypogaea L*; Drought tolerance; Photosynthesis; Yield; CROPGRO-Peanut

2.2 Introduction

Drought stress is a major abiotic stress contributing to reduced yields worldwide (Kambiranda et al., 2011). Drought stress affects biochemical and physiological processes at the molecular, cellular, and whole-plant levels, including metabolism, stomatal conductance, photosynthesis, nitrogen fixation, mineral nutrition uptake, water relations, leaf expansion, grain development and yield (Tardieu et al., 2011). Drought avoidance and tolerance are two strategies for plants to survive during severe stress conditions (Ravi et al., 2011). Many physiological traits can be responsible of a crop drought tolerance ability, such as cell protection (Tardieu, 2012), early stomatal closure (Shekoofa et al., 2015) or reduced leaf area (Reddy et al., 2003), maintenance of vegetative growth (Tardieu and Tuberosa 2010), or photosynthesis (Zhang, 2021), high water use (Sinclair, 2011), large root systems (Ye et al., 2018) or reduced seed abortion rate under water deficit (Tardieu, 2012). These cultivar characteristics can be used to develop cultivars adapted to drought (Parent and Tardieu 2014; Hammer et al., 2020).

Peanut (*Arachis hypogea L.*) is an annual legume that has been grown extensively in the

tropical and subtropical regions of Asia, Africa, and North America characterized by high temperature and erratic precipitation (Qin et al., 2012). Drought stress is the major abiotic constraint affecting peanut productivity and quality worldwide (Songsri et al., 2008). Within the United States, peanut production is concentrated in the Southeastern Coastal Plains. Most of the peanuts in this region are grown under rainfed conditions with sandy or loamy sand soil, which is frequently subjected to drought stress of different durations and intensities (Reddy et al., 2003). Peanut grown under irrigation may experience extreme drought and heat because of limited water supply or because irrigation water is applied in amounts and frequencies less than optimal for plant growth (Kambiranda et al., 2011). Breeding peanut varieties adapted to drought through conventional and genomic approaches will provide a long-term solution to overcome drought limitations in peanut production (Chen et al., 2013).

Different selection criteria have been used for breeding drought tolerant peanuts, which depend on physiological surrogates for 1) water use efficiency (WUE) such as specific leaf area (SLA), harvest index (HI), and SPAD chlorophyll meter reading (SCMR), 2) root-related traits such as root length, root volume, nitrogen fixation, 3) photosynthesis traits such as photosynthesis rate, stomatal conductance, relative water content (RWC), and canopy temperature, and 4) yield-related traits such as pod and seed number and above ground biomass (Girdthai et al., 2012; Chen et al., 2013; Zhang, 2021). Because peanut is grown in semi-arid environments with erratic rainfall, multiple drought adaptation mechanisms may exist within the same cultivar (Dang et al., 2013). To minimize yield reduction caused by frequent drought occurrence in the Southeastern Coastal Plains, there is a need to develop better drought adaptive strategies under climate change, including the improvement of

drought response traits through direct or indirect selection (Chen et al., 2013).

Selection for drought tolerance can be time-consuming and labor-intensive because drought-related traits are controlled by many genes which have been shown to interact with the environment. Substantial genotype by environment interactions under long-term climate trends will impede the breeding process (Ravi et al., 2011). Crop simulation models have been commonly used to evaluate genetic improvement and support the breeding of traits adapted to drought under different environments (Boote et al., 2021). These include assisting with multi-environment evaluation of advanced peanut breeding lines, understanding the nature of genotype × environment interactions, identification and evaluation of desirable traits, designing a crop ideotype for a specific environment, and evaluation of breeding strategies for drought tolerance (Boote et al., 2001; Suriharn et al., 2007; Narh et al., 2015). Crop modeling can be an efficient way to determine optimum traits by determining if a given trait is positive or negative to crop production under long-term climate patterns (Tardieu and Tuberosa 2010). In order to anticipate the consequences of climate change in a breeding program, virtual cultivars incorporating various drought tolerant traits have been defined and simulated using crop models to evaluate the peanut adaptation to climate change and the benefits for yield maintenance under water-limited environments (Bogard et al., 2020). The traits included deep root systems, maximum fraction of shoot dry matter diverted to root growth under water stress, reduced maximum transpiration rate, early stomata closure, drought tolerant nitrogen fixation, longer seed-filling duration (Sinclair et al., 2010; Singh et al., 2014; Battisti et al., 2017). These past drought-tolerance simulations provide guidance on trait selection using hypothetical examples, however, there is limited real evidence of the success of these different

physiological mechanisms of drought-tolerance to cultivars. There have been no attempts to quantitatively assess the yield gains of actual drought tolerant traits based on real cultivars using model simulations and experimental data.

The overall goal of this work was to develop a modelling approach to simulate the trait of enhanced photosynthesis under drought using the CROPGRO-Peanut model and combine the modified model with historic climate data to assess the contribution of this trait on peanut yield across long-term seasons. The specific objectives were to 1) estimate crop model genetic coefficients for several drought tolerant peanut varieties grown under both well-water and water-limited conditions; 2) develop and evaluate a framework to simulate enhanced photosynthesis under drought in the model; and 3) evaluate the impact of this trait on peanut yield for different seasons.

2.3 Materials and methods

2.3.1 Experimental Sites and Design

Three experiments were conducted to evaluate peanut drought tolerance performance and collect model input data (Table 2.1). Four peanut varieties (C1: AU-NPL 17, C2: Georgia 06G, C3: AU 16-28, and C4: TUFRunner 297) and one advanced breeding lines (C5: PI 390428) were selected because they showed variations in response to drought intensities in previous test. Peanut varieties (C1-C4) were evaluated in the field experiments (Exp. 1 and Exp. 2) at the E.V. Smith Research Center of Auburn University at Shorter, AL (EV, 32°29′ N, 85°53′ W) and the Wiregrass Research and Extension Center of Auburn University at Headland, AL (HL, 31°22′ N, 85°19′ W) in 2019 and 2020. Each genotype was arranged in a split plot design with

rainfed and irrigation (main split: dryland and full irrigation) with four replications. Each plot had four 12-m rows with a row spacing of 0.91 m and a seeding rate of 20 seeds m^{-2}. Plots were planted in late May and harvested in early October of each year at both sites. Irrigation and agronomic management practices followed best management practices for peanut according to Alabama Peanut Extension Team. The irrigated plots at E.V. Smith, AL, received 165 and 114 mm of irrigation water in 2019 and 2020, respectively. The irrigated plots at Headland, AL, received 104 and 95 mm in 2019 and 2020, respectively.

Peanut varieties (C1-C3, and C5) were evaluated for drought tolerance in the environmental controlled rainout shelter (5.5×12.2 m, Dang et al., 2013) experiment (Exp. 3) in 2019 and 2020 at National Peanut Research Laboratory, Dawson, GA (DA, 31°45′ N, 84°26′ W). In year 2019, five shelters were treated as drought stress experimental plots and one shelter was maintained as a fully watered control plot. Each peanut cultivar was hand planted in a 4 feet single row separated by a 2 feet alley. While in year 2020, three shelters were treated as drought stress experimental plots and one shelter was maintained as a fully watered control plot. Each cultivar hand planted in an 8 feet single row separated by a 2 feet alley (Zhang, 2021). Plots were planted in earlier May and harvest in early October of each year in all rainout shelters. All plots were irrigated before planting to provide uniform germination. Fully watered control plots were fully irrigated throughout the growing season. For drought stress experimental plots, irrigation and rainfall were withheld for 4 weeks in year 2019 and 5 weeks in year 2020 during July and August to create middle season drought, after which they were re-irrigated to recover until harvest (Zhang, 2021). Crop management related to fertilization, weed and pest control was conducted according to University of Georgia best

management practices (Dang et al., 2013).

2.3.2 Weather, Soil and Crop data

The CROPGRO-peanut model, which is distributed with DSSAT v4.7 (Hoogenboom et al., 2019), requires daily weather data including daily maximum/minimum air temperature, solar radiation, and precipitation. Long-term records of weather data from 1998-2020 was obtained from the Auburn University Mesonet (AWIS-AUM, 2021) for EV and HL. Observed weather data for DA in 2019 and 2020 was taken from the University of Georgia Weather Network (UGWN, 2021).

The CROPGRO-Peanut model also required soil profile data including lower limit (LL), drained upper limit (DUL), saturated water holding capacity (SAT), root growth factor (SRGF), bulk density (BD), soil organic carbon (Org. C), clay, silt, total nitrogen (Total N) and pH in different soil layers in the top 200 cm. These soil data for each soil type at each site were taken from the Gridded Soil Survey Geographic database developed by the USDA-National Resources Conservation Service (USDA-NRCS, 2020) and from the SoilGrids dataset with a spatial resolution of 1 km or 250 m (Hengl et al., 2014, 2017).

The crop data collected in 2019 and 2020 for EV and HL include phenology dates (sowing, flowering, first pod and maturity), photosynthesis, stomatal conductance, SPAD, and aboveground biomass (separated into leaves and stems), which was measured once each month for the entire period of the experiment. The pod and seed yield were measured at the harvest maturity stage. The soil water content in the 0- to 45-cm soil profile was recorded regularly during the peanut growing period using PR2 soil moisture probes. The measured crop data in 2019 and 2020 for the DA experiment include phenology data (sowing, maturity),

photosynthesis, stomatal conductance, leaf water content, and specific leaf area, which was collected once a week during the middle season of no irrigation and rainfall period. The aboveground biomass and pod yield were measured at the harvest maturity stage.

2.3.3 Development of Drought Tolerant Peanut Varieties in Model

The C1, C2 and C3 peanut varieties demonstrated drought tolerant traits based on breeding trial data. These varieties were found to be drought tolerant by maintaining photosynthesis under drought conditions in the two-year rainout shelter experiments in DA (Zhang, 2021). In the model, daily photosynthesis (*PG*) is computed using canopy method (Boote et al., 2008) by

$$PG_d = PTSMAX_d * SWFAC,$$
$$SWFAC = TRWUP_d / EP1_d \quad (1)$$

Where *d* is the day of the year, *PTSMAX* is the maximum daily photosynthesis as a function of photosynthetically active radiation, and *SWFAC* is a water deficit index that is calculated from the relationship between potential daily root water uptake over the soil profile (*TRWUP*, cm/d) and actual daily plant transpiration rate (*EP1*, cm/d), which is estimated in the soil water balance module (Ritchie, 1998). When the root water uptake is less than plant transpiration, the water deficit index (*SWFAC*) is less than one and water stress reduces daily photosynthesis. The lower the value of *SWFAC*, the larger the reduction in daily photosynthesis (Figure 2.1).

Under water limited conditions, a drought tolerant (*DT*) factor was developed for drought tolerant peanut varieties to reduce the negative effect of the water stress index (*SWFAC*) on daily photosynthesis:

$$SWFAC = SWFAC * DT \qquad SWFAC < 1 \text{ (water stress occurs)},$$
$$SWFAC = SWFAC \qquad SWFAC \geq 1 \text{ (no water stress)} \tag{2}$$

DT is a drought tolerant factor designed to maintain photosynthesis under water deficit (Figure 2.1). This factor was defined and coded as an additional ecotype coefficient input in the model.

2.3.4 Model Calibration and Evaluation Procedure

The model calibration procedure required three steps. First, the two-year crop data and soil water content measured in irrigated treatments (no water stress) in the EV and HL experiments were used to calibrate the model to obtain the cultivar coefficients of the four peanut varieties (C1-C4). The next step was to use the genetic coefficients calibrated for these peanut varieties (C1-C4) and one advanced breeding line (C5) grown under rainout shelter (middle season drought) in the DA experiment during 2019 and 2020 to calibrate the *DT* value of the modified model to minimize error in simulated and observed pod yield. The two-year rainout shelter experiments did not include the baseline peanut variety (C4). Since C4 was a high yielding variety with less drought tolerance, the range of pod yield for drought sensitive varieties in the experiment was assumed to represent the possible range of pod yield of the C4 peanut variety grown in rainout shelter (Tillman, 2018). The last step was evaluated the optimum *DT* factor and cultivar coefficients of peanut varieties (C1-C4) in rainfed

treatments in the EV and HL experiments to evaluate the modified model performance in simulating peanut biomass, yield and soil water content.

To evaluate the model performance and accuracy for simulating peanut development and growth characteristics, statistical indicators including coefficient of determination (R^2), root mean square error (*RMSE*), normalized root mean square error (*NRMSE*), and Willmott's index of agreement (*D* value, Willmott, 1982) were computed from observed (O_i) and simulated (S_i) variables. Willmott's *D* value is a better indicator of model performance, particularly relative to the 1:1 line, than a correlation coefficient (*r*), and values closer to 1 indicate better prediction while a *D* value of zero indicates no predictability.

2.3.5 Modelling Impacts of Enhanced Photosynthesis under Drought across Seasons

The modified CROPGRO-Peanut model with the optimum *DT* factor was then used to estimate the effects of this drought tolerant mechanism on rainfed yield, seasonal evapotranspiration (ET) and grain water use efficiency (WUE) for both dry and wet seasons at EV and HL in Alabama. The long-term (23 years) seasonal rainfall was classified into two categories: wet and dry seasons for each site. Wet and dry seasons were years with percentage of deviation of averages above and below -10%, respectively (Zhang et al., 2018). The grain WUE was calculated as the ratio of seed yield (kg/ha) to seasonal ET (mm).

$$\text{Grain WUE} = \text{Rainfed yield} / \text{Seasonal ET} \quad (3)$$

2.4 Results

2.4.1 Test Performance of the Newly Modified Model

The cultivar coefficients for the CROPGRO-Peanut model were calibrated to accurately simulate the peanut (C1-C4) growth and yield response to irrigation treatments at EV and HL in 2019 and 2020. The ability of the model to simulate developmental stages was assessed by comparing the simulated durations of three developmental stages with the corresponding observed values (Figure 2.2a and Table 2.2). The model simulated the duration from planting to first flowering and first pod reasonably well for both seasons, with an *RMSE* values of 3.9 and 3.3 days, and *NRMSE* values of 0.12 and 0.07. The model also simulated the maturity dates well for calibration seasons, with an overall *RMSE* value of 5.5 days and a *NRMSE* value of 0.04. The results indicated the model accurately simulated the observed durations of flowering, first pod, and maturity for the four peanut varieties in the EV and HL experiments. There was also good agreement between simulated and observed biomass, pod and seed yield for both seasons at both sites (Figure 2.2b-d and Table 2.2). The *RMSE* and *NRMSE* values for biomass, pod and seed yield were 1551.9 kg/ha and 0.12, 526.5 kg/ha and 0.08, 501.8 kg/ha and 0.10, respectively. The *D*-value, a measure of model predictability, was also high (0.74 for Biomass, 0.90 for pod yield and 0.86 for seed yield).

The genetic coefficients calibrated for the baseline and drought tolerant varieties were used in the rainout shelter experiment (DA, 2019 and 2020) to estimate the *DT* coefficient for the drought tolerant peanut varieties (C1-C3) (Figure 2.3). Varieties C4 and C5 represented the drought sensitive varieties and simulated and observed pod yields were in the range of similar drought sensitive varieties for 2019 and 2020 (Figure 2.3). The optimum value of *DT* = 3 minimized the error between simulated and observed pod yield for the three drought tolerant varieties (C1-C3) for 2019 and 2020 (Figure 2.3, Table 2.3). The *DT* factor for the

two drought susceptible varieties was 1. A *DT* coefficient of 3 gave very good simulated values for the pod yield for the drought tolerant varieties, and simulated well the observed differences in pod yield between the drought susceptible and drought tolerant varieties (Figure 2.3). There was a large difference in simulated and observed pod yield between the drought sensitive (C4, C5) and drought tolerant (C1-C3) varieties for both seasons.

The irrigated treatments in the EV and HL experiments were used to evaluate the model using the calibrated genetic coefficients and calibrated *DT* coefficient. The model simulated days to first flower, first pod and maturity very well as indicated by low values for *RMSE* and *NRMSE*, and the high values of *D* (Figure 2.4e and Table 2.3). This was expected since the genetic coefficients were calibrated for the irrigation treatment in this experiment. Figure 2.4 shows simulated and observed biomass, pod weight and seed weight without (Figure 2.4b, c, d) and with (Figure 2.4f, g, h) the drought tolerant modification. When the *DT* factor was set to 1, the model systematically under simulated biomass, pod and seed weight (Figure 2.4b, c, d). When the *DT* factor was set to the optimum value of 3, the model gave good simulations of biomass, seed and pod weight (Figure 2.4f, g, h). The *RMSE* and *NRMSE* values for biomass, pod and seed yield were 1660 kg/ha and 0.15, 919.6 kg/ha and 0.17, 812.4 kg/ha and 0.21, respectively. The *D*-value, a measure of model predictability, was higher than 0.48 (Table 2.3).

Taking the peanut variety C1 (AU-NPL 17) for example, the simulated values fit well with the observed values for biomass, pod and seed weight at different growth stages in both irrigated (calibration) and rainfed (evaluation) experiments at EV and HL during the 2019 and 2020 (Figure 2.5). In addition, the model accurately simulated the changes in soil water

content in various layers of in the top 45 cm soil profile under both irrigation (calibration) and rainfed (evaluation) conditions for two seasons at EV and HL (Figure 2.A1). These results confirmed that the cultivar coefficients of the peanut varieties and the optimum *DT* value (*DT* = 3) are sufficiently accurate, and the modified CROPGRO-Peanut model successfully simulated growth and yield of drought tolerant peanut varieties in response to water deficit.

2.4.2 Assess Impact of the Enhanced Trait on Peanut Yield across Multiple Seasons

The modified CROPGRO-Peanut model with the *DT* factor (*DT* =3) was used to investigate peanut rainfed yield, seasonal ET and grain WUE responses to different climate patterns (dry and wet seasons) using historical weather data from 1998–2020 at the EV and HL sites (Figure 2.6). At the EV site, eight seasons were classified as dry seasons while 15 seasons were classified as wet seasons. At the HL site, 11 seasons were classified as dry seasons while 12 seasons were classified as wet seasons. The three-drought tolerant peanut varieties (C1-C3) gave higher simulated average seed weight with less yield variation compared to the baseline peanut variety (C4) for both dry and wet seasons at both sites (Figure 2.6a, b). For dry seasons at EV, the average rainfed seed yield for the three-drought tolerant peanut varieties were C1: 3632 kg/ha, C2: 4503 kg/ha, and C3: 4038 kg/ha, which were 48%, 83% and 64% higher than the baseline peanut variety C4 (2460 kg/ha). For the wet seasons at EV, the three-drought tolerant peanut varieties (C1: 3865 kg/ha, C2: 4809 kg/ha, and C3: 4286 kg/ha) also gave 13-25% higher yields compared to the baseline peanut variety C4 (3423 kg/ha). For the HL site, the average rainfed seed yield for varieties C1-C3 (C1: 3586 kg/ha, C2: 4048 kg/ha, and C3: 3645 kg/ha) during the dry seasons were 16-31% higher than the baseline variety (C4: 3102 kg/ha). For wet seasons, the seed yield of the three-drought

tolerant peanut varieties (C1: 4462 kg/ha, C2: 5054 kg/ha, and C3: 4520 kg/ha) was 6-20% higher than the baseline peanut variety (C4: 4227 kg/ha).

In addition to the yield advantage, the seasonal ET and grain WUE of the drought tolerant peanut varieties (C1-C3) were enhanced compared to the baseline peanut variety (C4), especially for dry seasons at both sites (Figure 2.6c-f). For dry seasons at EV, the seasonal increase in ET ranged from 10% to 21% and the increase in grain WUE ranged from 33% to 50% compared to the baseline peanut variety (C4). For the wet seasons at EV, the seasonal ET and grain WUE were higher for drought tolerant peanut varieties (C1-C3), with an increase of 3%-12%, and 9%-26%, respectively.

Compared with baseline peanut variety (C4) under dry seasons at HL, the seasonal ET and grain WUE were 5%-7%, and 14%-25% higher for drought tolerant peanut varieties (C1-C3). The percentage increase under wet seasons were smaller for both the seasonal ET and grain WUE, which were 1%-5%, and 2%-13%, respectively. Overall, these simulation results indicate selection of the trait for maintaining photosynthesis rate under water deficit is desirable across the Southeastern USA and could be a useful trait if incorporated into new germplasms.

2.4.3 Sensitivity Analysis of the Drought Tolerant Factor

The purpose of the drought tolerant (*DT*) factor incorporated into the model was to increase daily gross photosynthesis during drought stress periods. A sensitivity analysis of *DT* was conducted using variety C1 (AU-NPL 17) for the 2019 rainfed experiment at EV. The model was run for different values of *DT* and the simulated daily and cumulative gross

photosynthesis was evaluated. When the *DT* value increased from 1 (non-drought tolerance) to 3 (drought tolerance), the daily gross photosynthesis increased significantly (Figure 2.7a) for the middle growth season period (60-100 days after planting) and the cumulative gross photosynthesis at final harvest increased by 51% (Figure 2.7c). For higher values of *DT* (Range: 3-9), the gross photosynthesis did not change much for the middle season period but increased consistently during the late growth season (Figure 2.7b). The increase of the cumulative gross photosynthesis at final harvest was also small, which were only 9% (*DT* = 5), 3% (*DT* = 7) and 2% (*DT* = 9), respectively. A *DT* value of 3 was optimum for the drought tolerant peanut varieties grown under field conditions. The similar results were also indicated by Figure 2.4. The *DT* value did affect the daily photosynthesis and subsequent growth the peanut varieties (Figure 2.4a). With *DT* =1, the biomass, pod and seed yield of the three-drought tolerant peanut varieties (C1-C3) were underestimated (Figure 2.4b, c, d), while with *DT* = 3, the prediction of biomass, pod and seed yield were optimized (Figure 2.4f, g, h).

The change in pod yield response to a range of *DT* values for the three-drought tolerant peanut varieties (C1-C3) grown under rainout shelter was also simulated using the modified CROPRO-Peanut model (Figure 2.8a, b). For both years (2019 and 2020) and three peanut varieties (C1-C3), the percent change in pod yield was sensitive to the changes in *DT* values (Range: 3-9). When *DT* values decreased from 3 to 1, the model simulated a 20% decrease in pod yield for both seasons. However, the percent change in pod yield plateaued when *DT* values were larger than 9, which indicate the modified model was not sensitive to *DT* values that were too large.

2.5 Discussions

Drought is the major yield-reducing factor for peanut in the Southeastern USA. Drought is caused by unpredictable and intermittent periods of water deficit which commonly occurs in the peanut growing region of the U.S. Because most peanuts are grown under rainfed conditions, drought tolerance is the priority trait for breeding high yielding peanut varieties over the long term (Soler et al., 2013; Wada et al., 2013). Using crop models as breeding tools in assisting the multi-environment evaluation of breeding lines is promising for peanut breeders as it will improve breeding efficiency and help estimate how drought tolerant traits may perform in different environments (Narh et al., 2015). One physiological mechanism that drought tolerant plants employ is maintaining a higher photosynthesis and water use under drought stress than drought susceptible varieties (Polania et al., 2016; Sanz Saez et al., 2019). In this work, a method was developed to simulate peanuts that maintain a high level of photosynthesis under drought stress. A new drought tolerance ecotype coefficient (*DT*) was incorporated into the CROPGRO-Peanut model to simulate this physiological response.

2.5.1 Peanut Multiple Drought Tolerance Strategies

Tolerance to abiotic stresses is an ambiguous concept, even after distinguishing different strategies (e.g., escape, avoidance and tolerance). The crop's sensitivity to drought stress is related with the timing and intensity of the stress (Boote at al., 2021). Early-season water deficit stress may not significantly reduce yield and quality, but rather, can sometimes lead to increased yields (Devi et al., 2019). Water stress during reproductive stages, such as pegging and pod development, results in a drastic reduction in yield. The magnitude of reduction can be attributed to peanut varieties and agronomic practices (Dang et al., 2013). Drought tolerance in peanut may be attributed to multiple adaptative traits, such as deeper root depth in

the soil profile or altered rooting distribution, partitioning of assimilates to root growth under water stress, partial stomatal closure, improved harvest index, or shortening the grain filling phase under water deficit (Tardieu and Tuberosa, 2010). There are many complex interactions that arise out of drought effects on multiple singular processes (Tardieu et al., 2011).

A given trait itself confers positive or negative effects on crop production depending on the environmental scenario. The long-term effects of drought tolerance traits on peanut are variable depending upon the rainfall pattern and the soil properties at the target sites, as well as the approach adopted in the model for simulated responses to soil water deficit (Tardieu, 2012). These various plant traits have been incorporated into model to create virtual cultivars for adaptation to climate change (Boote, 2004). It was concluded that enhancing yield potential traits, including maximum leaf photosynthesis rate (LFMAX), fraction of daily growth partitioned to pod (XFRT) and seed-filling duration (SFDUR), each by 10% could increase pod yield by 9%-14% across different sites under both current and future climates (Singh et al., 2013). Opportunities exist to investigate the consequences of the specific trait-maintained photosynthesis under water deficit through crop modelling that integrates genetic, climate, and soil process interactions. Such process-based modelling methods allow a quick estimation of the effect of the modified trait across different climate, soil, and management combinations (Zhao et al., 2019).

2.5.2 Effects of the Maintaining Photosynthesis on Crop Production

The target trait of the maintaining photosynthesis under water deficit was observed in rainout shelter experiments and chosen to incorporate into the CROPGRO-Peanut model

because that it was the primary drought mechanism in our peanut varieties and was explicitly related to dry matter accumulation and yield (Figure 2.3). Photosynthesis is more resilient to soil water deficit than processes linked to leaf expansion and development (Muller et al., 2011). Photosynthesis is reduced by soil moisture stress due to a partial reduction in stomatal conductance and reductions in leaf area. As moisture stress increases, stomata start closing as a mechanism to reduce transpiration. As a result, the entry of CO_2 is also reduced (Reddy et al., 2003). Biomass accumulation is subsequently reduced because of decreased stomatal opening and reduced carbon assimilation by leaves (Figure 2.1). Effects of soil water deficit are imposed on the photosynthetic rate that is also the driving factor for pod addition rate, seed growth rate and final seed number. Genotypes with enhanced photosynthesis thanks to a more effective use of water (Blum, 2009; Polania et al., 2016) under water deficit would have greater aboveground biomass during reproductive stages which significantly increases grain number as compared with baseline genotypes (Zhao et al., 2019). However, the elevated transpiration associated with maximum biomass during dry periods could also lead to total water depletion and plant death before the end of the crop cycle if common mid-season drought suffered in the Southeastern USA develops in a terminal drought (Tardieu and Tuberosa, 2010).

Whether maintaining photosynthesis under water deficit leads to increased or decreased yield depends on local climate conditions and stored soil water. There is a lack of systematic quantification of the enhanced photosynthesis across a range of climate conditions at multiple sites (Chenu et al., 2018). In this study, we simulated peanut behavior related to maintaining photosynthesis under multiple climatic scenarios to test the overall effects on yield. The long-

term simulation of the trait for maintaining photosynthesis rate under water deficit confirmed that consistent yield gains could be obtained in dry seasons and no risk of yield loss in wet seasons, which was attributed to interactions among cultivar differences (e.g., C1, C2, C3), weather conditions (e.g., dry and wet), and soil types (e.g., loam and sandy loam) (Figure 2.6a, b). Grain yield under water-limited conditions was strongly associated with the extent of seasonal ET and crop water use, where a higher seasonal ET and grain WUE was observed for drought tolerant peanut varieties, especially for dry seasons at the EV location (Figure 2.6c-f).

2.5.3 Model Sensitivity to the Newly Designed Drought Tolerant Factor

The simulation of responses to soil water deficit varies across current crop models due to differences in model structures and parameter values (Asseng et al., 2013). The CROPGRO-Peanut model with the daily potential evapotranspiration options has no VPD (vapor pressure deficit) effect on photosynthesis or on stomatal function. Effects of water deficit on photosynthesis are simulated through reduced rates of accumulation of dry matter (Boote et al., 2008). The *SWFAC* signals are used to regulate peanut growth processes and photosynthesis using growth phase-specific modifiers (Figure 2.1). When *SWFAC* is less than 1.0, rooting depth extension is more rapid, leaf senescence is accelerated, crop life cycle may be accelerated or delayed, and nitrogen mobilization is more rapid during seed fill. The species file contains corresponding coefficients for each given crop life cycle phase to designate accelerated or delayed development under water deficit (Boote et al., 2001). However, these regulation coefficients are species traits that are constant for all peanut varieties within a species. This work suggests that these regulation coefficients for drought responses should be cultivar specific for peanut varieties. The *DT* factor defined in this study

to maintain photosynthesis under water deficit is an example of the need for cultivar specific drought tolerant traits. The *DT* value of 3 that was optimized for drought tolerant peanut varieties in this study was able to accurately simulate drought tolerant peanut growth and yield under water deficit (Table 2.3). Furthermore, more drought tolerant ecotype coefficients may need to be defined for other desirable traits, for example, high nitrogen fixation under drought, higher pod harvest index, and leaf growth maintenance.

A sensitivity analysis was conducted to better understanding of how *DT* factors in crop models affect peanut photosynthesis and harvest yield (Parent and Tardieu 2014). The modified model gave different responses to *DT* factors over a range of values (Range: 1-9) (Figure 2.7 and Figure 2.8). Peanut grown in the southeastern USA usually experience unpredictable and intermittent periods of water deficit during reproductive stages at middle seasons (Hamidou et al., 2012). Simulated changes in *DT* factors from 1 to 3 mainly affected daily gross photosynthesis during middle seasons, and drastically influenced the cumulative gross photosynthesis and final pod yield at the same time (Figure 2.7c). As *DT* factors were increasing from 3 to 9, the increase of gross photosynthesis occurred later in the season when pegging and pod development were beginning to slow down. Therefore, the change of pod yield caused by change in photosynthesis becomes reaches a plateau, where the peanut photosynthesis and yield were not sensitive to higher values of *DT*. The approach used in the modified model to maintain photosynthesis under drought gave good simulations of plant growth and yield for drought tolerant varieties in this study.

2.6 Conclusions

Simulating drought tolerance of peanut varieties can enhance the process of breeding new germplasms adapted to drought. In this study, the peanut trait of maintaining photosynthesis under water deficit was observed in rainout shelter trials and incorporated into CROPGRO-Peanut model. The optimum *DT* factor was determined under rainout shelter conditions and gave good simulations of drought tolerant peanut growth, yield and soil water content for independent experiments. The long-term simulation of drought tolerant and baseline peanut varieties showed that the trait of maintaining photosynthesis under water deficit had a consistent yield advantage in dry seasons and no risk of yield loss in wet seasons, which was associated with higher seasonal ET and grain WUE. Further sensitivity analysis based on the modified model confirmed the ability of the model to accurately simulate the change of photosynthesis and yield in respond to changes in *DT* factors. The drought simulation approach develop in this study represents a first attempt to simulate a single drought tolerant mechanism for peanut using experimental data. This method can be expanded to consider other drought tolerant traits as data become available.

2.7 References

Asseng, S., Ewert, F., Rosenzweig, C., Jones, J.W., Hatfield, J.L., Ruane, A.C., Boote, K.J., Thorburn, P.J., Rötter, R.P., Cammarano, D. and Brisson, N., 2013. Uncertainty in simulating wheat yields under climate change. Nature climate change, 3(9), pp.827-832.

Agricultural Weather Information Service (AWIS)- Auburn University Mesonet (AUM), 2021. http://awis.aumesonet.com/ (accessed 2 August 2021).

Boote, K.J., Kropff, M.J. and Bindraban, P.S., 2001. Physiology and modelling of traits in crop plants: implications for genetic improvement. Agricultural Systems, 70(2-3), pp.395-420.

Boote, K.J., Sau, F., Hoogenboom, G. and Jones, J.W., 2008. Experience with water balance, evapotranspiration, and predictions of water stress effects in the CROPGRO model. Response of crops to limited water: Understanding and modeling water stress effects on plant growth processes, 1, pp.59-103.

Blum, A., 2009. Effective use of water (EUW) and not water-use efficiency (WUE) is the target of crop yield improvement under drought stress. Field crops research, 112(2-3), pp.119-123.Bogard, M., Ravel, C., Paux, E., Bordes, J., Balfourier, F., Chapman, S.C., Le Gouis, J. and Allard, V., 2014. Predictions of heading date in bread wheat (*Triticum aestivum L.*) using QTL-based parameters of an ecophysiological model. Journal of experimental botany, 65(20), pp.5849-5865.

Bogard, M., Biddulph, B., Zheng, B., Hayden, M., Kuchel, H., Mullan, D., Allard, V., Gouis, J.L. and Chapman, S.C., 2020. Linking genetic maps and simulation to optimize breeding for wheat flowering time in current and future climates. Crop Science, 60(2), pp.678-699.

Battisti, R., Sentelhas, P.C., Boote, K.J., Camara, G.M.D.S., Farias, J.R. and Basso, C.J., 2017. Assessment of soybean yield with altered water-related genetic improvement traits under climate change in Southern Brazil. European Journal of Agronomy, 83, pp.1-14.

Boote, K.J., Jones, J.W. and Hoogenboom, G., 2021. Incorporating realistic trait physiology into crop growth models to support genetic improvement. in silico Plants, 3(1), p.diab002.

Chen, C.Y., Nuti, R.C., Rowland, D.L., Faircloth, W.H., Lamb, M.C. and Harvey, E., 2013. Heritability and genetic relationships for drought-related traits in peanut. Crop Science, 53(4), pp.1392-1402.

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. and Hammer, G.L., 2018. Integrating modelling and phenotyping approaches to identify and screen complex traits: transpiration efficiency in cereals. Journal of experimental botany, 69(13), pp.3181-3194.

Dang, P.M., Chen, C.Y. and Holbrook, C.C., 2013. Evaluation of five peanut (*Arachis hypogaea*) genotypes to identify drought responsive mechanisms utilising candidate-gene approach. Functional Plant Biology, 40(12), pp.1323-1333.

Devi, M.J., Sinclair, T.R., Vadez, V., Shekoofa, A. and Puppala, N., 2019. Strategies to Enhance Drought Tolerance in Peanut and Molecular Markers for Crop Improvement. Genomics Assisted Breeding of Crops for Abiotic Stress Tolerance, Vol. II, pp.131-143.

Godfray, H. C. J. et al. 2010. Food security: the challenge of feeding 9 billion people. Science 327, 812–818.

Girdthai, T., Jogloy, S., Vorasoot, N., Akkasaeng, C., Wongkaew, S., Patanothai, A. and Holbrook, C.C., 2012. Inheritance of the physiological traits for drought resistance under

terminal drought conditions and genotypic correlations with agronomic traits in peanut. SABRAO J Breed Genet, 44, pp.240-262.

Grassini, P., van Bussel, L.G., Van Wart, J., Wolf, J., Claessens, L., Yang, H., Boogaard, H., de Groot, H., van Ittersum, M.K. and Cassman, K.G., 2015. How good is good enough? Data requirements for reliable crop yield simulations and yield-gap analysis. Field Crops Research, 177, pp.49-63.

Hengl, T., de Jesus, J.M., MacMillan, R.A., Batjes, N.H., Heuvelink, G.B., Ribeiro, E., Samuel-Rosa, A., Kempen, B., Leenaars, J.G., Walsh, M.G., 2014. SoilGrids1km-global soil information based on automated mapping. PLoS One. 9, e105992.

Hengl, T., Mendes de Jesus, J., Heuvelink, G.B., Ruiperez Gonzalez, M., Kilibarda, M., Blagotic, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G., Ribeiro, E., Wheeler, I., Mantel, S., Kempen, B., 2017. SoilGrids250m: Global gridded soil information based on machine learning. PLoS One. 12, e0169748.

Hoogenboom, G., C.H. Porter, V. Shelia, K.J. Boote, U. Singh, J.W. White, L.A. Hunt, R. Ogoshi, J.I. Lizaso, J. Koo, S. Asseng, A. Singels, L.P. Moreno, and J.W. Jones. 2019. Decision Support System for Agrotechnology Transfer (DSSAT) Version 4.7.5 (https://DSSAT.net). DSSAT Foundation, Gainesville, Florida, USA.

Hammer, G.L., McLean, G., van Oosterom, E., Chapman, S., Zheng, B., Wu, A., Doherty, A. and Jordan, D., 2020. Designing crops for adaptation to the drought and high-temperature risks anticipated in future climates. Crop Science, 60(2), pp.605-621.

Jongrungklang, N., Toomsan, B., Vorasoot, N., Jogloy, S., Kesmala, T. and Patanothai, A., 2008. Identification of peanut genotypes with high water use efficiency under drought stress conditions from peanut germplasm of diverse origins. Asian Journal of Plant Sciences.

Kambiranda, D.M., Vasanthaiah, H.K., Katam, R., Ananga, A., Basha, S.M. and Naik, K., 2011. Impact of drought stress on peanut (*Arachis hypogaea L.*) productivity and food safety. Plants and environment, pp.249-272.

Liu, F., Jensen, C.R. and Andersen, M.N., 2004. Drought stress effect on carbohydrate concentration in soybean leaves and pods during early reproductive development: its implication in altering pod set. Field crops research, 86(1), pp.1-13.

Mothilal, A., & Ezhil, A. 2010. Combining ability analysis for yield and its components in groundnut (*Arachis hypogaea L.*). Electronic Journal of Plant Breeding, 1(2), 162-166.

Muller, B., Pantin, F., Génard, M., Turc, O., Freixes, S., Piques, M. and Gibon, Y., 2011. Water deficits uncouple growth from photosynthesis, increase C content, and modify the relationships between C and growth in sink organs. Journal of experimental botany, 62(6), pp.1715-1729.

Narh, S., Boote, K.J., Naab, J.B., Jones, J.W., Tillman, B.L., Abudulai, M., Sankara, P., M'Bi Bertin, Z., Burow, M.D., Brandenburg, R.L. and Jordan, D.L., 2015. Genetic Improvement of Peanut Cultivars for West Africa Evaluated with the CSM-CROPGRO-Peanut Model. Agronomy Journal, 107(6), pp.2213-2229.

Parent, B. and Tardieu, F., 2014. Can current crop models be used in the phenotyping era for predicting the genetic variability of yield of plants subjected to drought or high temperature?. Journal of experimental botany, 65(21), pp.6179-6189.

Polania, J.A., Poschenrieder, C., Beebe, S. and Rao, I.M., 2016. Effective use of water and increased dry matter partitioned to grain contribute to yield of common bean improved for drought resistance. Frontiers in plant science, 7, p.660.

Qin, H., Feng, S., Chen, C., Guo, Y., Knapp, S., Culbreath, A., He, G., Wang, M.L., Zhang, X., Holbrook, C.C. and Ozias-Akins, P., 2012. An integrated genetic linkage map of cultivated peanut (*Arachis hypogaea L.*) constructed from two RIL populations. Theoretical and Applied Genetics, 124(4), pp.653-664.

Ritchie, J.T., 1985. A user-orientated model of the soil water balance in wheat. In Wheat growth and modelling (pp. 293-305). Springer, Boston, MA.

Ritchie, J.T., 1998. Soil water balance and plant water stress. In Understanding options for agricultural production (pp. 41-54). Springer, Dordrecht.

Reddy, T. Y., Reddy, V. R., & Anbumozhi, V. 2003. Physiological responses of groundnut (*Arachis hypogea L.*) to drought stress and its amelioration: a critical review. Plant growth regulation, 41(1), 75-88.

Ravi, K., Vadez, V., Isobe, S., Mir, R.R., Guo, Y., Nigam, S.N., Gowda, M.V.C., Radhakrishnan, T., Bertioli, D.J., Knapp, S.J. and Varshney, R.K., 2011. Identification of several small main-effect QTLs and a large number of epistatic QTLs for drought

tolerance related traits in groundnut (*Arachis hypogaea L.*). Theoretical and Applied Genetics, 122(6), pp.1119-1132.

Suriharn, B., Patanothai, A., Pannangpetch, K., Jogloy, S. and Hoogenboom, G., 2007. Determination of cultivar coefficients of peanut lines for breeding applications of the CSM-CROPGRO-Peanut model. Crop science, 47(2), pp.607-619.

Songsri, P., Jogloy, S., Kesmala, T., Vorasoot, N., Akkasaeng, C., Patanothai, A., & Holbrook, C. C. 2008. Heritability of drought resistance traits and correlation of drought resistance and agronomic traits in peanut. Crop Science, 48(6), 2245-2253.

Sinclair, T.R., Messina, C.D., Beatty, A. and Samples, M., 2010. Assessment across the United States of the benefits of altered soybean drought traits. Agronomy Journal, 102(2), pp.475-482.

Sinclair, T.R., 2011. Challenges in breeding for yield increase for drought. Trends in plant science, 16(6), pp.289-293.

Soler, C. M. T., Suleiman, A., Anothai, J., Flitcroft, I., & Hoogenboom, G. 2013. Scheduling irrigation with a dynamic crop growth model and determining the relation between simulated drought stress and yield for peanut. Irrigation science, 31(5), 889-901.

Singh, P., Nedumaran, S., Boote, K.J., Gaur, P.M., Srinivas, K. and Bantilan, M.C.S., 2014. Climate change impacts and potential benefits of drought and heat tolerance in chickpea in South Asia and East Africa. European Journal of Agronomy, 52, pp.123-137.

Singh, P., Nedumaran, S., Ntare, B.R., Boote, K.J., Singh, N.P., Srinivas, K. and Bantilan, M.C.S., 2014. Potential benefits of drought and heat tolerance in groundnut for

adaptation to climate change in India and West Africa. Mitigation and adaptation strategies for global change, 19(5), pp.509-529.

Shekoofa, A., Rosas‐Anderson, P., Sinclair, T.R., Balota, M. and Isleib, T.G., 2015. Measurement of limited‐transpiration trait under high vapor pressure deficit for peanut in chambers and in field. Agronomy Journal, 107(3), pp.1019-1024.

Sanz‐Saez, A., Maw, M.J., Polania, J.A., Rao, I.M., Beebe, S.E. and Fritschi, F.B., 2019. Using carbon isotope discrimination to assess genotypic differences in drought resistance of parental lines of common bean. Crop Science, 59(5), pp.2153-2166.

Tardieu, F., 2003. Virtual plants: modelling as a tool for the genomics of tolerance to water deficit. Trends in plant science, 8(1), pp.9-14.

Tardieu, F. and Tuberosa, R., 2010. Dissection and modelling of abiotic stress tolerance in plants. Current opinion in plant biology, 13(2), pp.206-212.

Tilman, D., Balzer, C., Hill, J. and Befort, B.L., 2011. Global food demand and the sustainable intensification of agriculture. Proceedings of the national academy of sciences, 108(50), pp.20260-20264.

Tardieu, F., Granier, C. and Muller, B., 2011. Water deficit and growth. Co-ordinating processes without an orchestrator? Current opinion in plant biology, 14(3), pp.283-289.

Tardieu, F., 2012. Any trait or trait-related allele can confer drought tolerance: just design the right drought scenario. Journal of experimental botany, 63(1), pp.25-31.

Tillman, B.L., 2018. Registration of ‘TUFRunner ‘297’’peanut. Journal of Plant Registrations, 12(1), pp.31-34.

University of Georgia Weather Network (UGWN), 2021. http://weather.uga.edu/ (accessed 2 August 2021).

US Department of Agriculture (USDA) - National Resources Conservation Service (NRCS), 2020. http://www.nrcs.usda.gov/wps/portal/nrcs/site/soils/home/ (accessed 2 July 2021).

Wada, Y., Wisser, D., Eisner, S., Flörke, M., Gerten, D., Haddeland, I., Hanasaki, N., Masaki, Y., Portmann, F.T., Stacke, T. and Tessler, Z., 2013. Multimodel projections and uncertainties of irrigation water demand under climate change. Geophysical research letters, 40(17), pp.4626-4632.

Ye, H., Roorkiwal, M., Valliyodan, B., Zhou, L., Chen, P., Varshney, R.K. and Nguyen, H.T., 2018. Genetic diversity of root system architecture in response to drought stress in grain legumes. Journal of Experimental Botany, 69(13), pp.3267-3277.

Zhang, D., Li, R., Batchelor, W.D., Ju, H. and Li, Y., 2018. Evaluation of limited irrigation strategies to improve water use efficiency and wheat yield in the North China Plain. PloS one, 13(1), p.e0189989.

Zhao, Z., Rebetzke, G.J., Zheng, B., Chapman, S.C. and Wang, E., 2019. Modelling impact of early vigour on wheat yield in dryland regions. Journal of experimental botany, 70(9), pp.2535-2548.

Zhang, Q., 2021. Uncovering different physiological mechanisms of peanut drought tolerance under mid-season drought in automated Rain-out Shelters. Master dissertation. Auburn University.

Table 2.1 Peanut varieties, soil types, water treatments, seasonal weather and years of data used for CROPGRO-Peanut model calibration and evaluation.

Experiment numbers	Locations	Peanut varieties	Soil types	Water treatments	T_{max} (◦C)	T_{min} (◦C)	P (mm)	Years
Exp. 1	EV Smith research center at Shorter, Alabama (EV)	C1: AU-NPL 17 C2: Georgia 06G C3: AU16-28 C4: TUFRunner 297	Loam	Dryland and Irrigation	32.0	19.9	306	2019/2020
Exp. 2	Wiregrass research and extension center at Headland, Alabama (HL)		Sandy loam	Dryland and Irrigation	32.1	21.3	495	2019/2020
Exp. 3	National Peanut Research Laboratory at Dawson, Georgia (DA)	C1: AU-NPL 17 C2: Georgia 06G C3: AU16-28 C5: PI 390428	Sandy loam	Rainout shelter	32.3	20.9	575	2019/2020

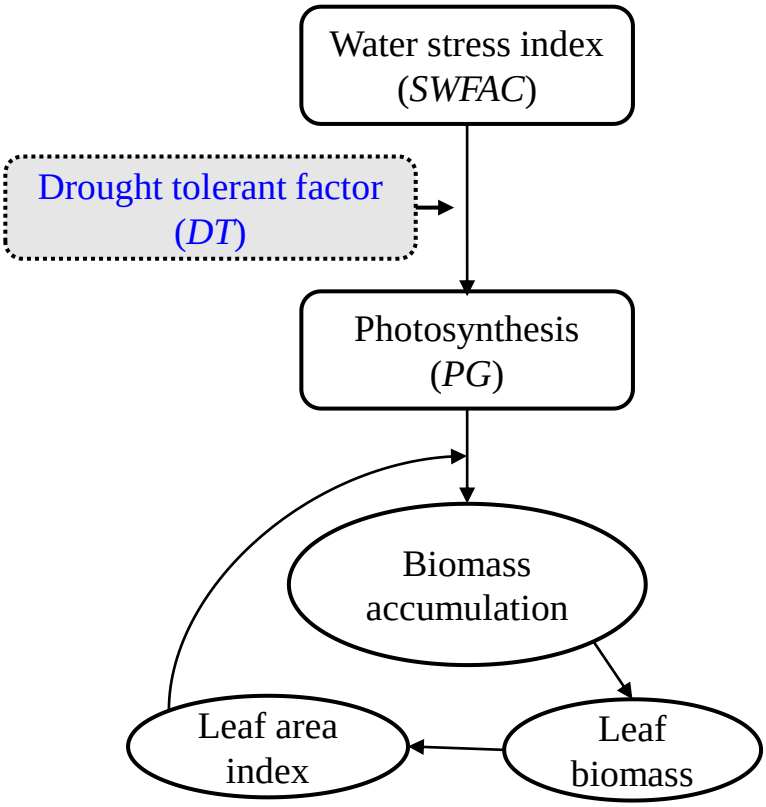


Figure 2.1 Schematic representation indicated the daily gross photosynthesis (*PG*) modified in CROPGRO-Peanut model for a peanut variety with drought tolerance.

Table 2.2 CROPGRO-Peanut model calibration for four peanut varieties (C1-C4) listed in Table 2.1 grown under irrigated conditions in EV and HL during 2019 and 2020.

Items	**Calibration (Irrigated field in EV and HL)**			
	R^2	***RMSE***	***NRMSE***	***D-value***
Days from sowing to first flowering (d)	0.31	3.9	0.12	0.58
Days from sowing to first pod (d)	0.44	3.3	0.07	0.69
Days from sowing to maturity (d)	0.43	5.5	0.04	0.68
Biomass (kg/ha)	0.42	1551.9	0.12	0.74
Pod yield (kg/ha)	0.70	526.5	0.08	0.90
Seed yield (kg/ha)	0.60	501.8	0.10	0.86

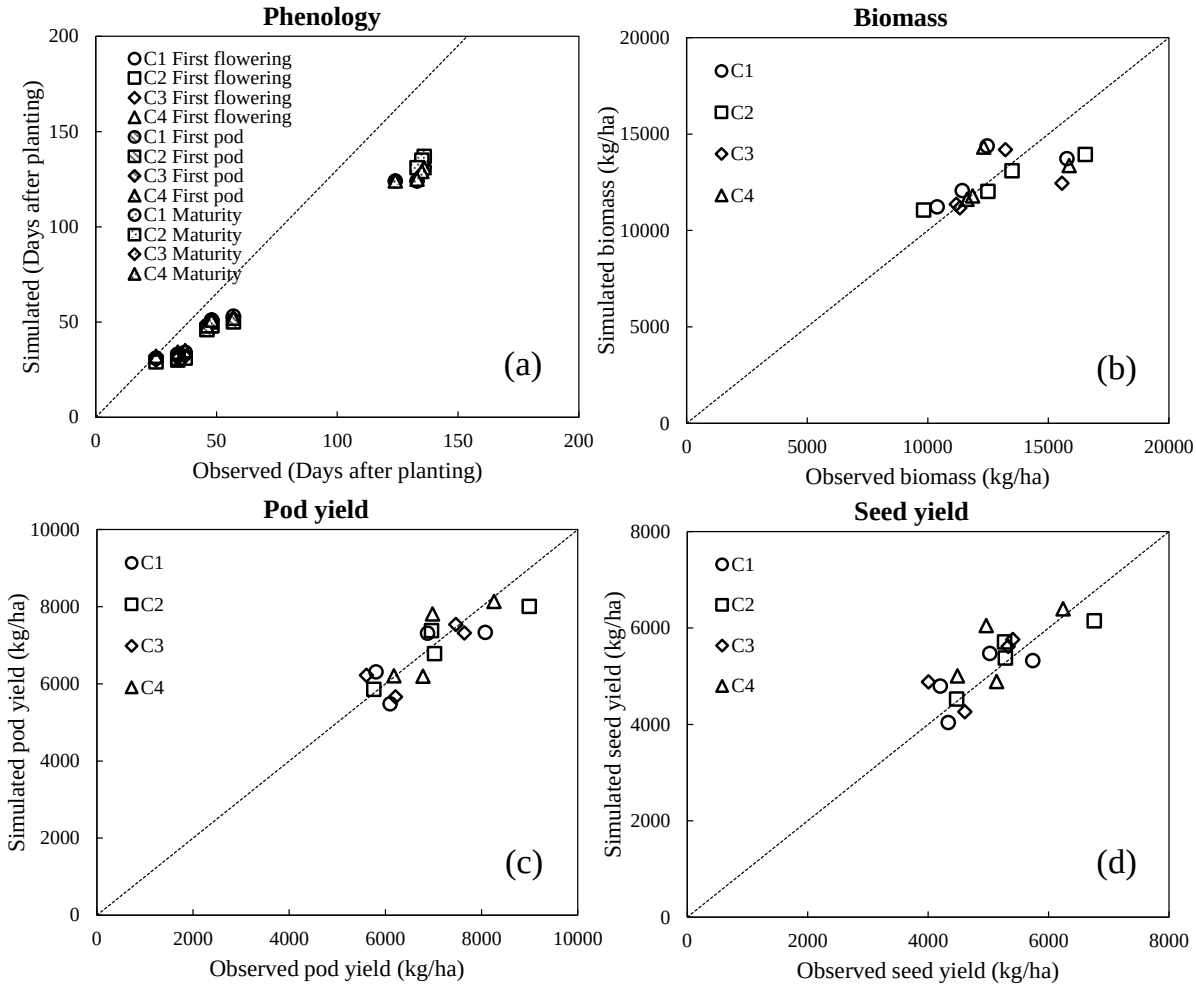


Figure 2.2 Comparison of simulated and observed values for model calibration (a-d) of four peanut varieties (C1-C4, listed in Table 2.1) grown under irrigated conditions at EV and HL during 2019 and 2020.

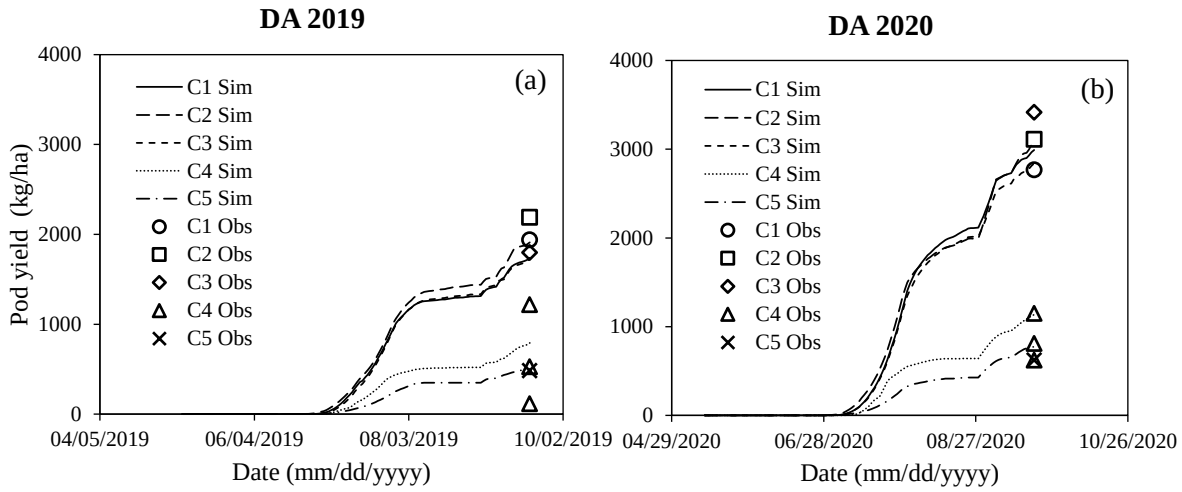


Figure 2.3 Simulated (Sim) and observed (Obs) pod yields for model evaluation (a-b) of five peanut varieties (C1-C5, listed in Table 2.1) with drought tolerant modification (*DT* = 3) grown under rainout shelter in DA during 2019 and 2020.

Table 2.3 CROPGRO-Peanut model evaluation for five peanut varieties (C1-C5, listed in Table 2.1) with drought tolerant modification (*DT* = 3) grown under rainfed conditions in EV and HL and rainout shelter in DA during 2019 and 2020.

Items	**Evaluation (Dryland in EV and HL)**				**Evaluation (Rainout shelter in DA)**			
	R^2	***RMSE***	***NRMSE***	***D-value***	***R^2***	***RMSE***	***NRMSE***	***D-value***
Days from sowing to first flowering (d)	0.34	3.8	0.12	0.60	-	-	-	-
Days from sowing to first pod (d)	0.30	3.7	0.07	0.67	-	-	-	-
Days from sowing to maturity (d)	0.53	6.3	0.05	0.67	0.02	4.7	0.04	0.50
Biomass (kg/ha)	0.15	1660.5	0.15	0.52	0.55	1330.9	0.28	0.83
Pod yield (kg/ha)	0.04	919.6	0.17	0.52	0.95	265.2	0.15	0.98
Seed yield (kg/ha)	0.04	812.4	0.21	0.48	-	-	-	-

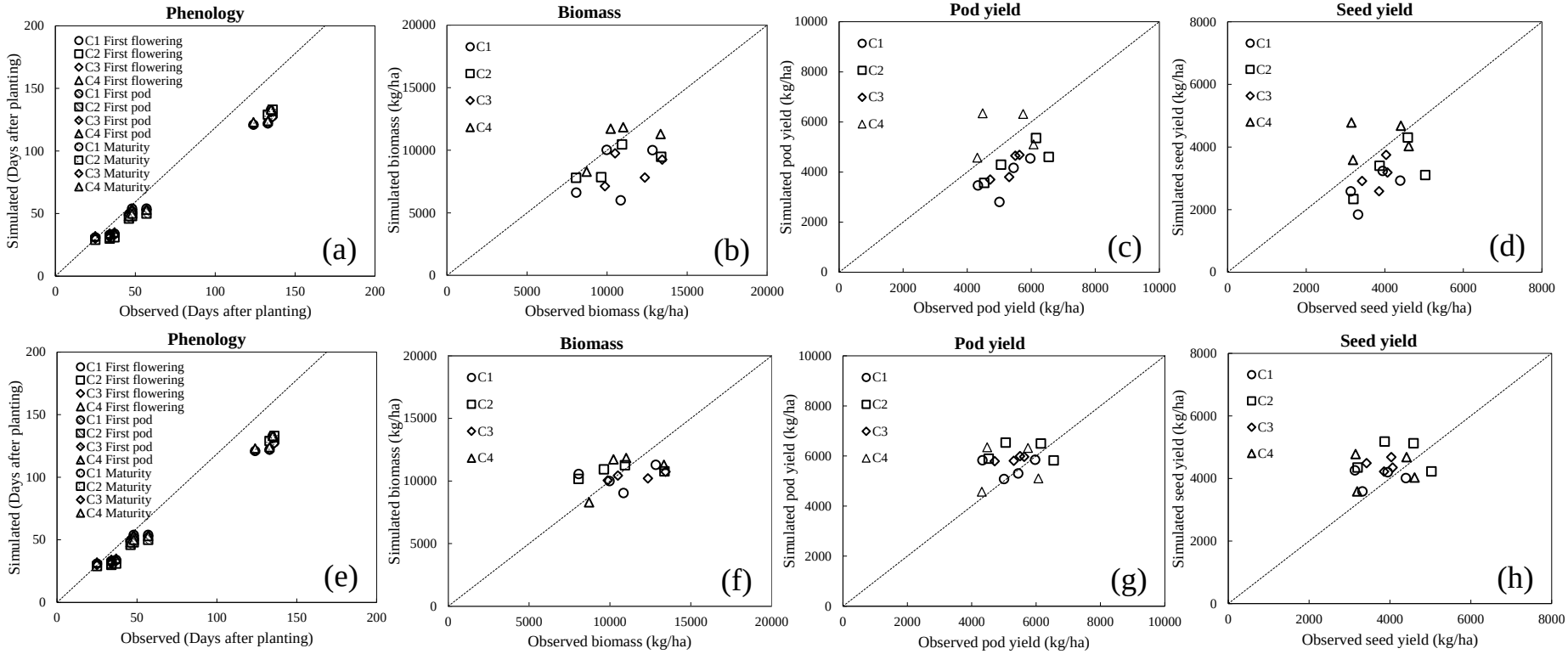


Figure 2.4 Comparison of simulated and observed values for model evaluation of four peanut varieties (C1-C4, listed in Table 2.1) without (*DT* = 1) (a-d) and with (e-h) the drought tolerant modification (*DT* = 3) grown under rainfed conditions in EV and HL during the 2019 and 2020.

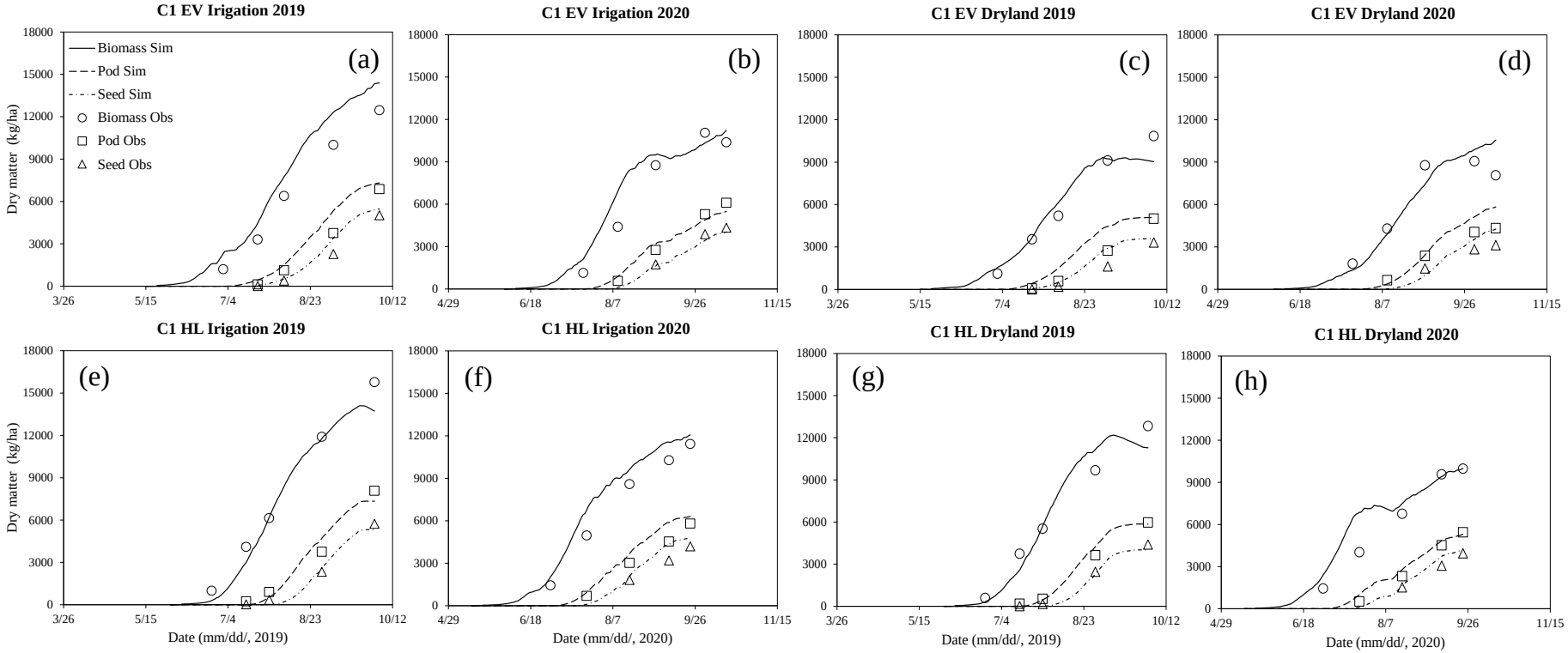


Figure 2.5 Simulated (Sim) and observed (Obs) values for biomass, pod, and seed weight of a drought-tolerant peanut variety (C1, listed in Table 2.1) grown in the irrigated (calibration) and rainfed (evaluation) experiments at EV and HL during 2019 and 2020.

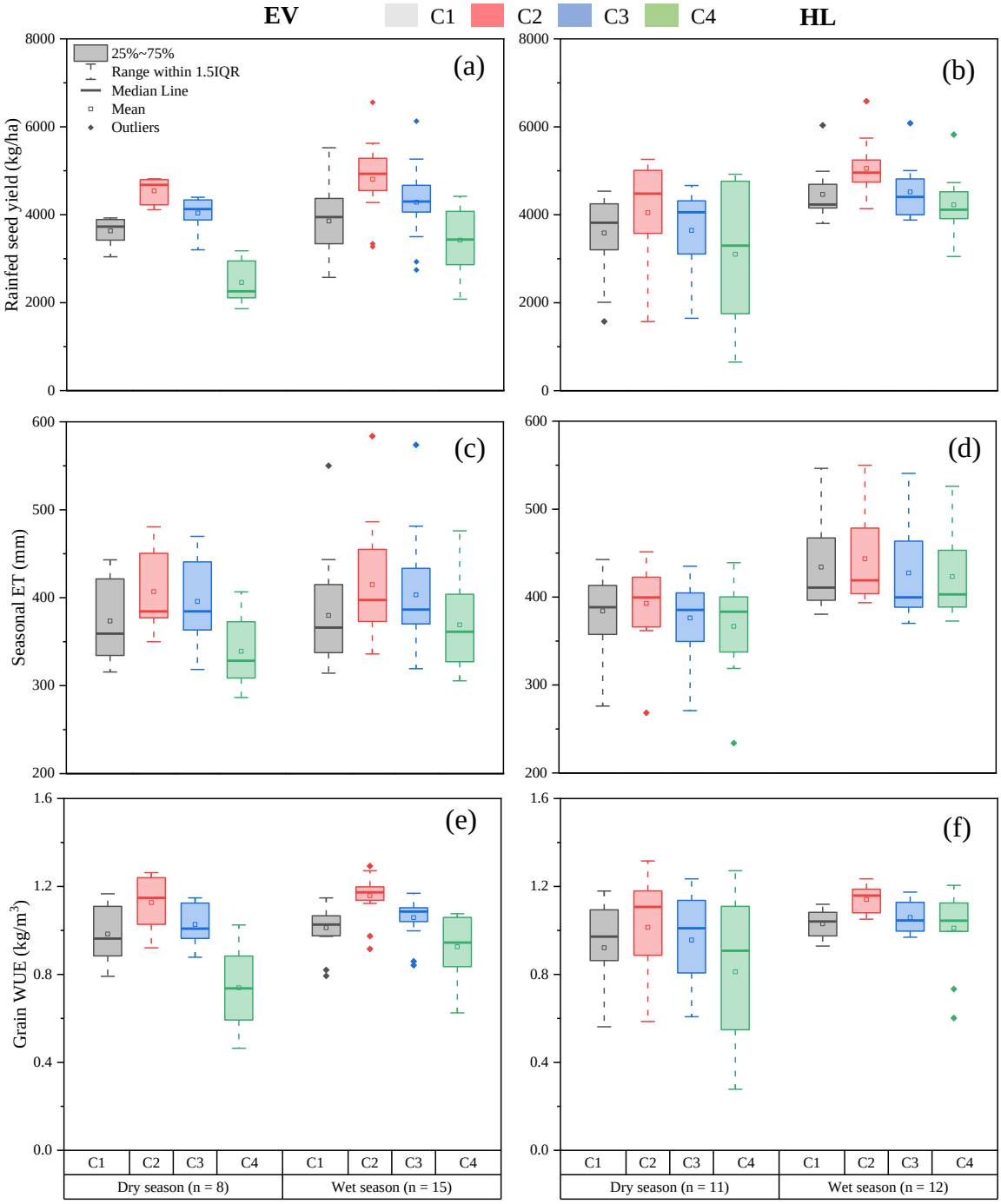


Figure 2.6 Simulated rainfed peanut grain weight (a-b), seasonal evapotranspiration (ET) (c-d) and grain water use efficiency (WUE) (e-f) for dry and wet seasons over the period 1998-2020 at EV (a, c, e) and HL (b, d, f) sites. The drought tolerant coefficient was *DT*=3.

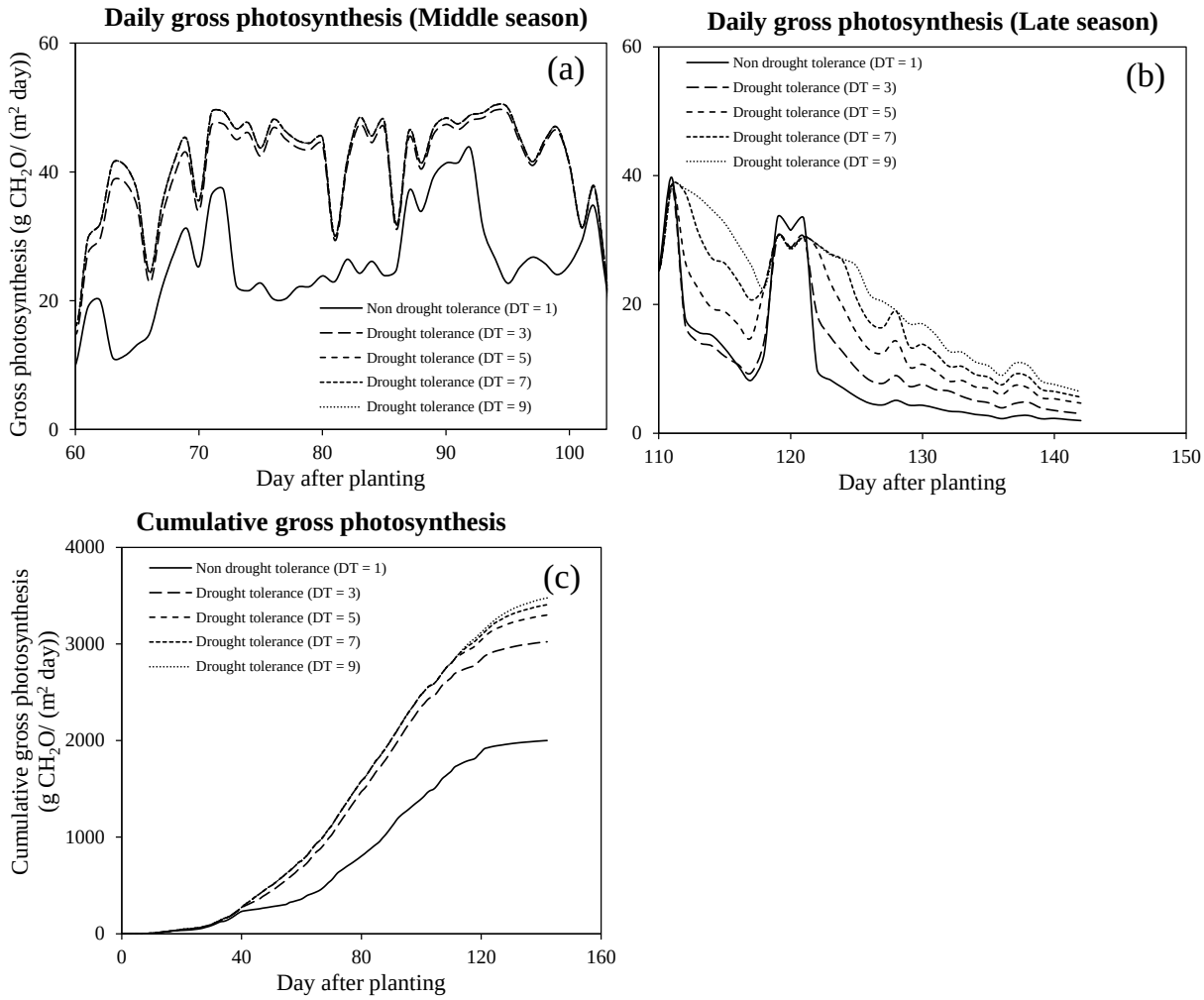


Figure 2.7 Sensitivity of simulated gross photosynthesis to changes in the *DT* coefficient for a drought-tolerant peanut variety (C1, listed in Table 2.1) for the 2019 EV rainfed experiment. Figure a and b show the daily gross photosynthesis during the middle and late season period while figure c shows the cumulative gross photosynthesis throughout the entire growing season.

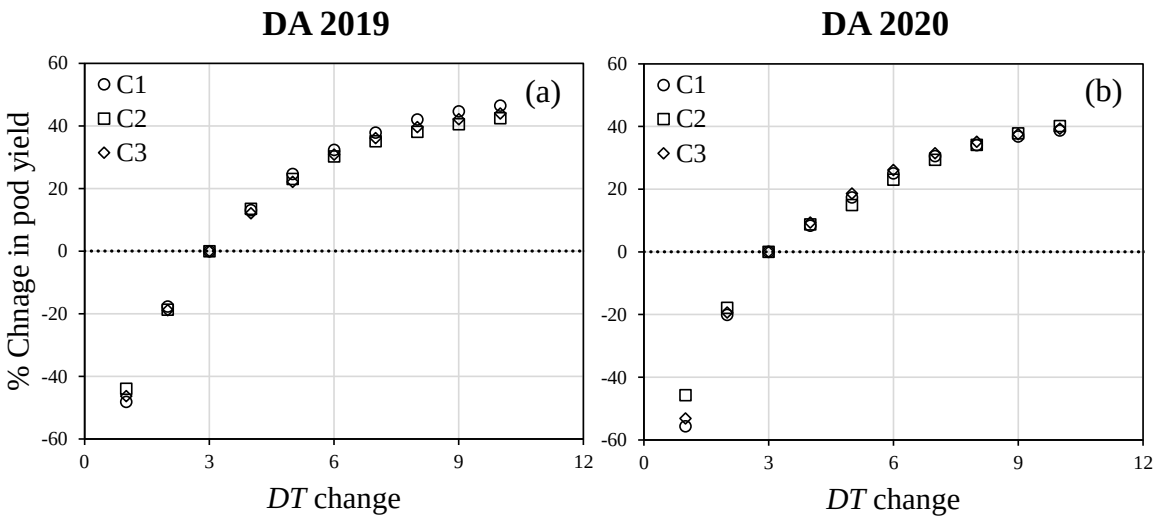
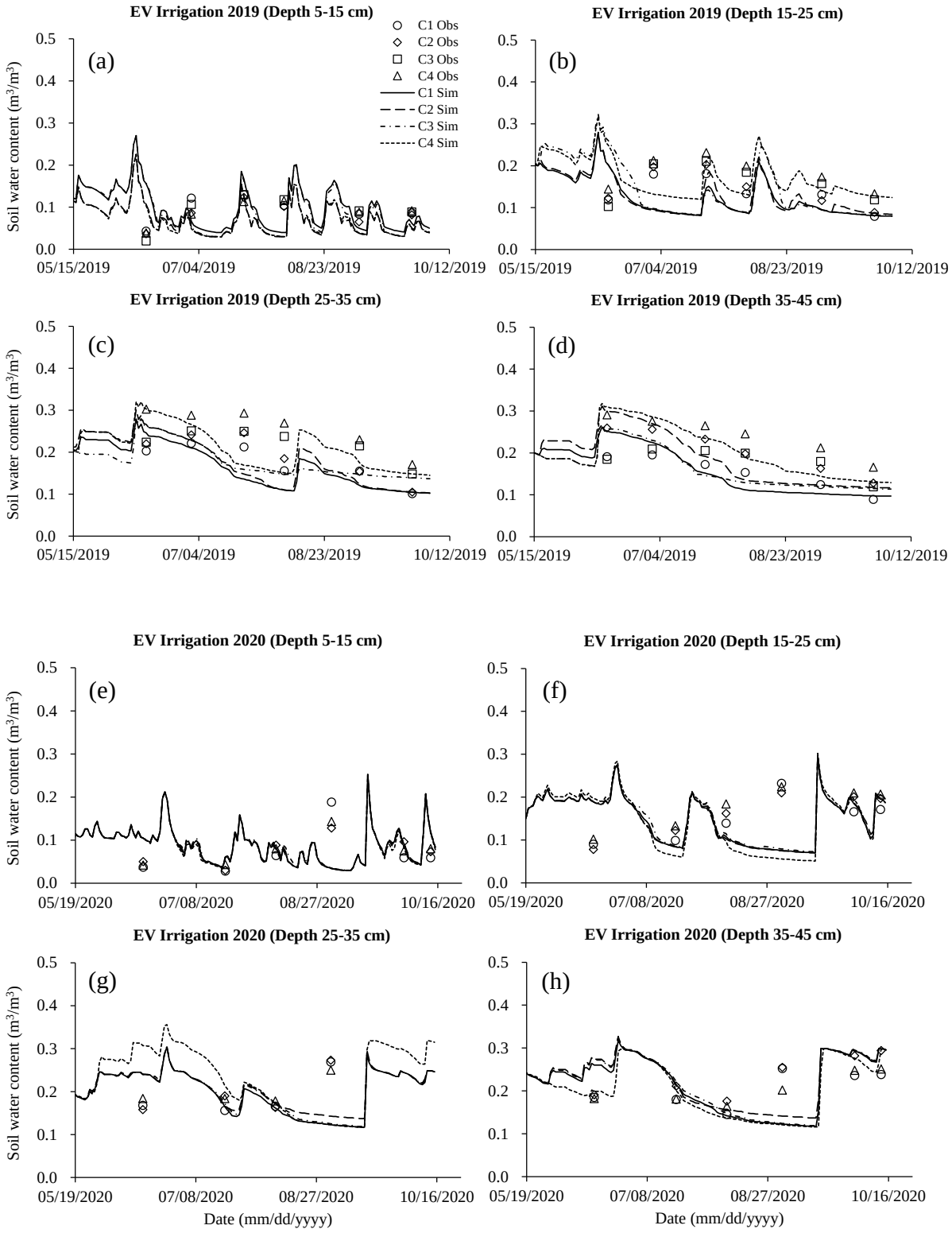
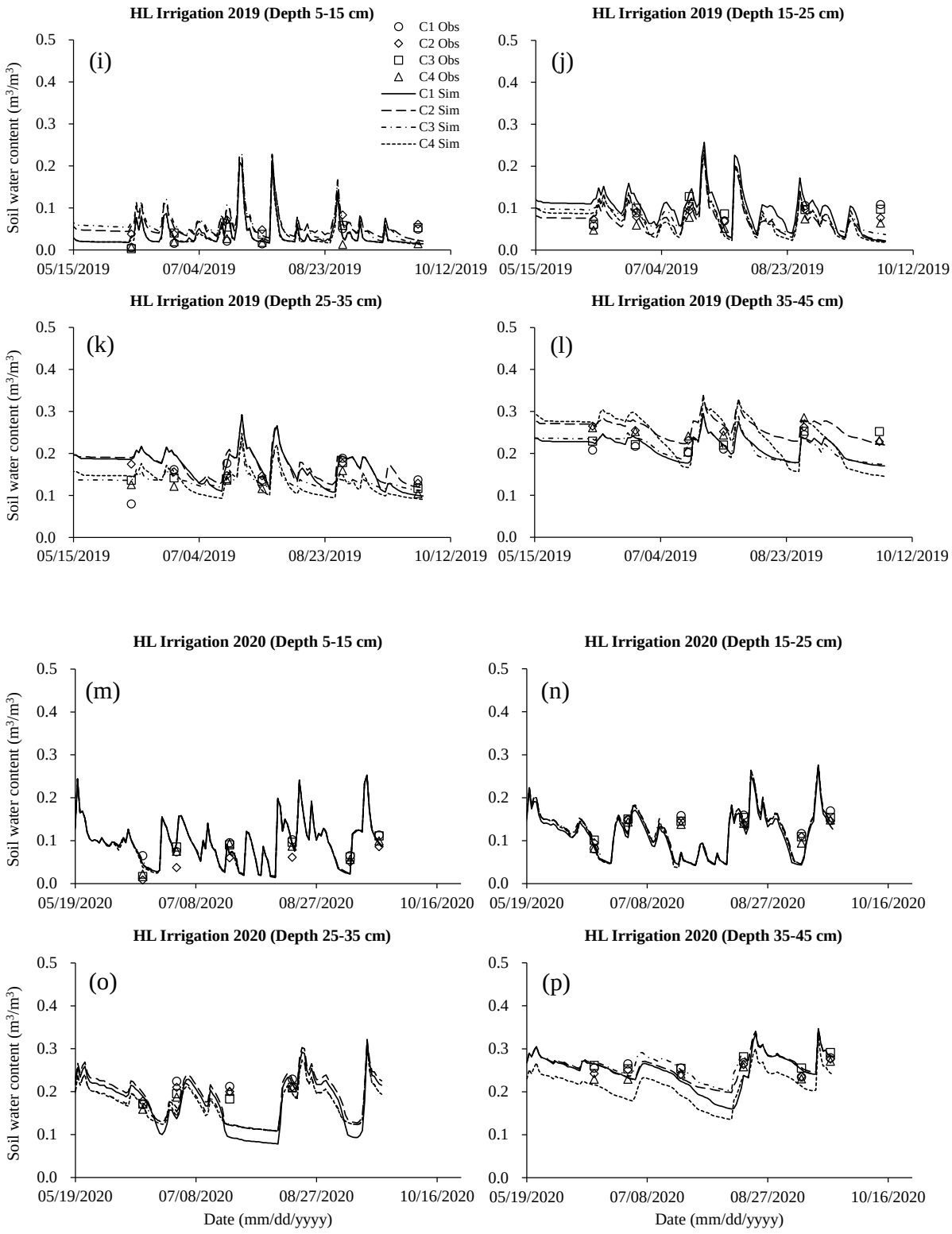
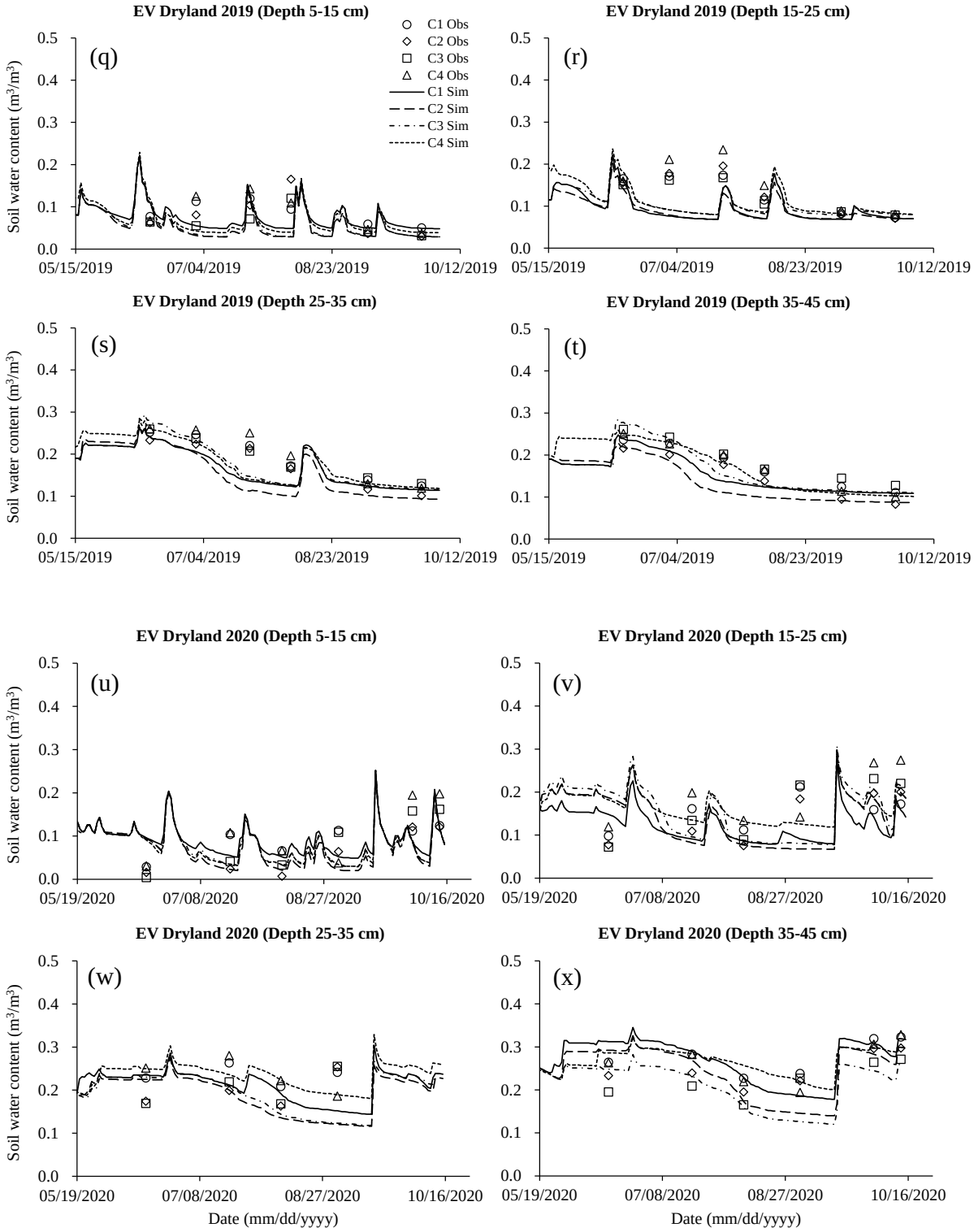


Figure 2.8 Sensitivity analyses indicated by change in pod yield response to *DT* change for three drought tolerant peanut varieties (C1-C3, listed in Table 2.1) grown under rainout shelter in DA during the 2019 (a) and 2020 (b).


EV Irrigation 2019 (Depth 5-15 cm)
EV Irrigation 2019 (Depth 15-25 cm)
EV Irrigation 2019 (Depth 25-35 cm)
EV Irrigation 2019 (Depth 35-45 cm)
EV Irrigation 2020 (Depth 5-15 cm)
EV Irrigation 2020 (Depth 15-25 cm)
EV Irrigation 2020 (Depth 25-35 cm)
EV Irrigation 2020 (Depth 35-45 cm)
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
C1 Obs
C2 Obs
C3 Obs
C4 Obs
C1 Sim
C2 Sim
C3 Sim
C4 Sim
Soil water content (m3/m3)
0.0
0.1
0.2
0.3
0.4
0.5
05/15/2019
07/04/2019
08/23/2019
10/12/2019
05/19/2020
07/08/2020
08/27/2020
10/16/2020
Date (mm/dd/yyyy)


HL Irrigation 2019 (Depth 5-15 cm)
HL Irrigation 2019 (Depth 15-25 cm)
HL Irrigation 2019 (Depth 25-35 cm)
HL Irrigation 2019 (Depth 35-45 cm)
HL Irrigation 2020 (Depth 5-15 cm)
HL Irrigation 2020 (Depth 15-25 cm)
HL Irrigation 2020 (Depth 25-35 cm)
HL Irrigation 2020 (Depth 35-45 cm)
(i)
(j)
(k)
(l)
(m)
(n)
(o)
(p)
C1 Obs
C2 Obs
C3 Obs
C4 Obs
C1 Sim
C2 Sim
C3 Sim
C4 Sim
Soil water content (m3/m3)
Date (mm/dd/yyyy)
05/15/2019
07/04/2019
08/23/2019
10/12/2019
05/19/2020
07/08/2020
08/27/2020
10/16/2020


EV Dryland 2019 (Depth 5-15 cm)
EV Dryland 2019 (Depth 15-25 cm)
EV Dryland 2019 (Depth 25-35 cm)
EV Dryland 2019 (Depth 35-45 cm)
EV Dryland 2020 (Depth 5-15 cm)
EV Dryland 2020 (Depth 15-25 cm)
EV Dryland 2020 (Depth 25-35 cm)
EV Dryland 2020 (Depth 35-45 cm)
(q)
(r)
(s)
(t)
(u)
(v)
(w)
(x)
C1 Obs
C2 Obs
C3 Obs
C4 Obs
C1 Sim
C2 Sim
C3 Sim
C4 Sim
Soil water content (m3/m3)
0.0
0.1
0.2
0.3
0.4
0.5
05/15/2019
07/04/2019
08/23/2019
10/12/2019
05/19/2020
07/08/2020
08/27/2020
10/16/2020
Date (mm/dd/yyyy)

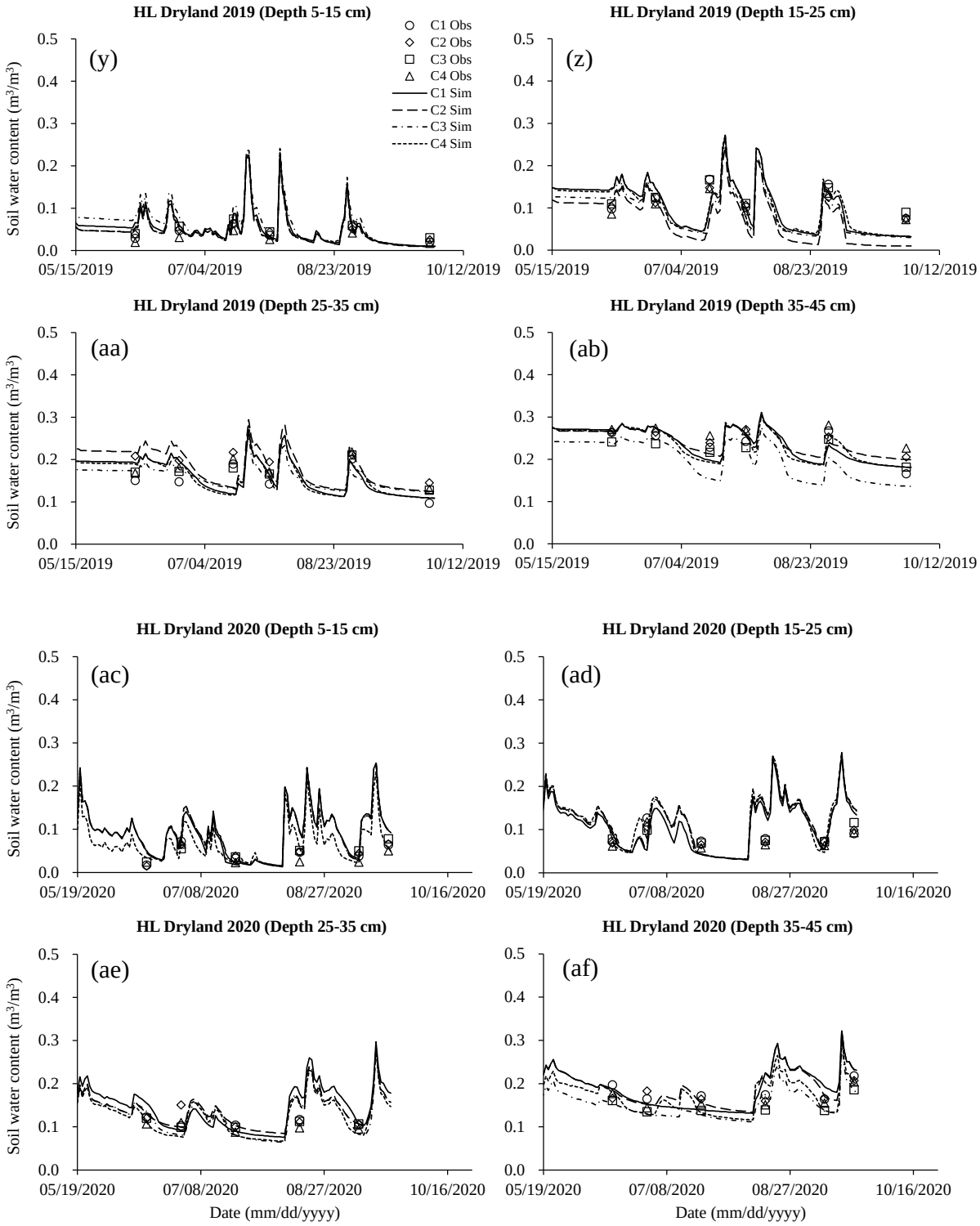


Figure 2.A1. Simulated (Sim) and observed (Obs) soil water content in the 0- to 45-cm soil profile during the growing season of four peanut varieties (C1-C4, listed in Table 2.1) under irrigation (a-p) and dryland (q-af) at EV and HL during 2019 and 2020.

Chapter 3. County-level calibration strategy to evaluate peanut (*Arachis hypogaea*) irrigation water use under different climate change scenarios

3.1 Abstract

Regional-scale simulation of crop yield is challenging due to the spatial variability of soil properties, crop varieties, management practices and weather conditions. Point-based crop models are commonly used for spatial simulation with increased availability of high-resolution spatial datasets. However, it is still difficult to calibrate crop models well due to the spatial variability of model inputs. The focus of this work was to determine if a single set of cultivar and soil parameters could be calibrated to simulated county-level peanut yield, and to evaluate the effects of irrigation to mitigate potential climate change impacts on peanut yield. Model input data for fourteen seasons and five major peanut producing counties was assembled and used for model calibration. Three seasons of data were withheld and used for an independent evaluation. Overall, peanut growth duration and county-level yields were simulated well with a set of optimum cultivar and soil parameters for each county. The model calibration showed that simulated maturity dates and yields were in good agreement with the observed county level values reported by NASS, giving an overall R^2 of 0.71 and 0.73, and *RMSE* values of 6 days and 333 kg/ha, respectively. The model also gave good simulations of maturity dates and yield for the three evaluation seasons, with an overall R^2 of 0.83 and 0.76 and *RMSE* values of 5 days and 429 kg/ha, respectively. The results from future climate simulations indicated that the rainfed yields will suffer from increasing daytime temperature and an irrigation strategy could potentially offset the heat and drought stress to maintain

higher peanut production in the Southeastern USA. This study provides a calibration and evaluation strategy that aggregates spatial heterogeneity of model inputs. This approach can enhance the accuracy of simulating the impact of climate change on crop production by providing a method to calibrate and evaluate the underlying model at larger spatial scales.

Keywords: CROPGRO-Peanut; Cultivar coefficients; Soil parameters; Yield; Climate change

3.2 Introduction

The global demand for agricultural crops is expected to double by 2050 due to increases in population, rising global middle class, and use of food crops for bioenergy production (Godfray et al., 2010; Tilman et al., 2011). Furthermore, climate change is expected to have adverse impacts on crop production around the world (Lobell et al., 2011, Rosenzweig et al., 2014). Simulating crop production is critical to understanding the impact of climate change and plan for mitigation strategies on yield and food security. Crop models are often used to simulate the impact of climate change, management policies and weather disasters at the county scale (Johnston 2013).

Dynamic process-oriented crop models that integrate physical and physiological processes of plant growth and development have been widely used (Jin et al., 2018; Huang et al., 2019). Most prevailing crop models are point-based and simulate growth processes over a homogeneous area. Physical and physiological assumptions in crop models are based on uniform field growth situations and commonly tested at the plot or field scale (Thorp et al., 2008). The typical input data requirements include weather, crop management practices,

genetic information and soil water holding characteristics, and it is assumed that these data are uniform over the simulation area (Batchelor et al., 2020). The model calibration process usually involves trial-and-error approaches and optimization of input parameters to minimize error (Ma et al., 2020). The accuracy of the trial-and-error approach usually depends on the user's experience to refine cultivar coefficients and/or soil parameter values until a good match between simulated and observed values are obtained. Automatic optimization methods have been developed and are increasingly being used to estimate model parameters (Thorp et al., 2013). Compared to the conventional trial-and-error approaches, automatic optimization methods are more objective and efficient and can increase model accuracy (He et al., 2010a).

With increased availability of spatial and temporal datasets including remotely sensed images, land cover maps, digital soil surveys, gridded weather datasets, and county scale crop yield reports, crop growth models are increasingly being used to simulate crop growth and yield at a larger scale (Manivasagam and Rozenstein, 2020). Point-based models are commonly used in spatial simulations for yield gap and yield potential analysis (Liu et al., 2021), irrigation water management (Kothari et al., 2019), leaf growth estimation (Zhen et al., 2018), fertilizer management (Yang et al., 2020), and climate impact on yield (Adhikari et al., 2016). There are clear benefits in using crop models for analyzing regional production because policy formulation and decision-making are rarely implemented at the plot or field level and policy makers need information at larger spatial scales. This is especially true for the adaptation to the future climate change on agriculture systems (Shin et al., 2020; Tofa et al., 2021).

Regional-scale simulations cannot consider the level of detailed data collected at the field or farm scale. Efforts have been focused on development of reasonable methods to capture spatial heterogeneity of climate data, crop varieties, soil properties, and management practices to minimize errors in yield simulation (Jagtap and Jones, 2002; Grassini et al., 2015; Manivasagam and Rozenstein, 2020). At a larger spatial scale, there is variability in weather, soil, genetics, and management that must be represented in the model. Climate data are often taken from local representative meteorological stations, or gridded weather data from climate simulation or interpolated from weather station networks (Adhikari et al., 2016). While crop varieties and soil types vary spatially, researchers often use several commonly grown varieties and dominant soil types to represent many varieties and soil types in regional simulations (Tofa et al., 2021). Relative to many farmers management practices, researchers often define a representative set of management practices to represent management within the region (Wang et al., 2020).

Peanut (*Arachis hypogea L.*) is an annual legume crop that has been grown extensively in the tropical and subtropical regions of Asia, Africa, and North America which are characterized by high temperature and erratic precipitation (Qin et al., 2012). Within the United States, peanut production is concentrated in the Southeastern Coastal Plains which is highly vulnerable to the changing seasonal temperatures and rainfall patterns. A significant increase in seasonal temperatures, ambient carbon dioxide level and precipitation anomalies could be detrimental to peanut growth in the future (Vara et al., 2003; Eck et al., 2020). Most of the peanuts in this region are grown under rainfed conditions with sandy or loamy sand soil

that have a lower water-holding capacity. Even peanut grown under irrigation may experience extreme drought and heat because of limited water supply or because irrigation water is applied in amounts and frequencies less than optimal for plant growth (Kambiranda et al., 2011). Therefore, it is important to understand how peanut yields and irrigation requirements may change under climate change scenarios to allow farmers and policymakers to determine how to maximize profits and reduce losses (Jin et al., 2018).

Previous research using crop models to simulate production at the county or regional scale rely on making an un-calibrated baseline model run and comparing results to an alternative management practice or climate scenario (Salazar et al., 2012; Leng et al., 2016). Relative yield differences are used to characterize the impact of the alternative management scenarios or climate change (Shin et al., 2020). There have been no attempts to assess the underlying accuracy in simulating historical peanut yields for multiple years at the county-level, even though the US Department of Agriculture-National Agricultural Statistics Service (USDA-NASS, 2020) reports county-level yields on an annual basis (Huang et al., 2021). Developing representative spatial model inputs that can be evaluated against historical county-level yields would reduce the uncertainty of using the models to study impacts of management practices or climate change at the county scale. The objectives of this work were to: (1) calibrate input parameters for the CROPGRO-Peanut model using historical NASS peanut yields for five counties across the Southeastern USA; (2) evaluate the calibrated baseline model using data from independent years; and (3) assess the potential effects of future climate change on peanut production and irrigation water use in the Southeastern USA.

3.3 Materials and methods

3.3.1 Description of the Study Area

The study area consisted of the top peanut producing counties by planted area in each of the five states in the Southeastern USA peanut belt. These states produce approximately 80% of US peanuts (USDA-NASS, 2020). The counties in this study included Santa Rosa (Florida), Houston (Alabama), Worth (Georgia), Orangeburg (South Carolina) and Martin (North Carolina) (Figure 3.1; Table 3.1). These counties ranged from 75.5°W to 88.5°W longitude and from 25.0°N to 36.5°N latitude. Table 3.2 shows a summary of weather conditions for each county during the growing seasons included in this study.

3.3.2 Crop Model Input Data

3.3.2.1 County Level Weather and Soil Data

The CROPGRO-Peanut model, which is included with DSSAT v4.7 (Hoogenboom et al., 2019), requires daily weather data input. A single representative point of weather was selected to represent average daily weather conditions within each county. Observed historical weather data were obtained from the cooperative observer program of the National Weather Service (NWS-COP, 2020) from 2006-2019 for each county. The nearest weather station to the centroid of each county was determined as the representative weather station. The projected daily weather data for the mid-century (2050s) and end-century (2080s) under representative concentration pathways scenarios (RCP 4.5 and 8.5) were generated by MarkSimGCM (http://gismap.ciat.cgiar.org/MarkSimGCM/) using spatial downscaling methods (Jones and

Thornton, 2013). The ensemble of two general circulation models (GCMs) including GFDL-ESM2G and GFDL-ESM2M developed by National Oceanic and Atmospheric Administration (NOAA) were used in this analysis (Dunne et al., 2012). Fourteen replicate years of each target future year (2050s and 2080s) were extracted to compare with the 14-year baseline period (2006-2019). Summary details of the climate scenarios are shown in Table 3.2.

The CROPGRO-Peanut model also required soil profile data in different soil layers in the top 200 cm (Table 3.A1). Our hypothesis was that a single soil profile can be calibrated to represent the seasonal water stress patterns affecting peanut yield at the county scale. These soil data were extracted from the Gridded Soil Survey Geographic database developed by the USDA-National Resources Conservation Service (USDA-NRCS, 2020) and from the SoilGrids dataset with a spatial resolution of 1 km or 250 m (Hengl et al., 2014, 2017). Properties for the most dominant soil type for peanuts within each county were used to represent an initial estimate of soil properties in each county. These soil properties were later refined through model calibration. The detailed soil profile characteristics for each representative soil type in each county are shown in Table 3.A1.

3.3.2.2 Peanut Phenology, Management and Yield Data

Historical county-level peanut yields for model calibration and evaluation were taken from the US Department of Agriculture-National Agricultural Statistics Service (USDA-NASS, 2020). Yields were reported at 10% moisture content and adjusted to 0% moisture content for model comparison. A state-level record of peanut phenology and management practices were

taken from the US Department of Agriculture-Risk Management Agency (USDA-RMA, 2020) and USDA-NASS for each county over the study period (Table 3.1).

The Georgia 06G (Runner) variety was chosen to represent the genetics planted in these top production counties from 2006-2019, as it was the most widely planted peanut variety in the peanut belt during that period (Branch, 2007). The peanut management information including sowing and maturity dates and sowing density are shown in Table 3.1. Peanuts in these counties are predominately grown under rainfed conditions, thus yields were simulated without irrigation (Brown and Pervez, 2014; USDA-NASS, 2017). The NASS reported yield data was used directly for model simulation comparison without detrending because the historical peanut yields did not show an increasing trend during our study period. We assumed that changes in agricultural technologies (e.g., narrow rows, pest management) did not critically affect yields during our study period and the interannual yield variability was mainly due to other factors (Wang et al., 2020).

3.3.3 Optimization Software and Package

Thorp and Bronson (2013) developed a geospatial toolbox called Geospatial Simulation (GeoSim) that is distributed as a plug-in for the open-source Quantum GIS environment. GeoSim was designed as a model independent optimizer that can estimate spatial model inputs to minimize an objective function. In our study, each county was mapped in QGIS 2.18.28 (QGIS.org, 2021) and a database was developed containing spatial soil parameters to be optimized. The CROPGRO-Peanut model was modified to accept optimum parameters from the GeoSim optimizer for simulating yields. The GeoSim software was then used to

optimize several soil parameters to minimize errors between simulated and observed yields for calibration years (Table 3.3).

The DSSAT-PEST package was used to calibrate cultivar coefficients to minimize errors between simulated and observed county-level yields. The main program file of the DSSAT-PEST package was a program in the R programming language which was executed using R 4.0.3 (R Core Team, 2021). Ma et al. (2020) reported that DSSAT-PEST gave accurate estimations of cultivar coefficients for different crops with good optimization efficiency. In this study, the cultivar coefficients were adjusted using the DSSAT-PEST package to minimize errors between simulated and observed peanut maturity dates and county-level yields reported by NASS.

3.3.4 Model Calibration and Evaluation Procedure

Yield data from about fourteen seasons were analyzed for each county and characterized as low (i.e., bottom 25% of yields), average (i.e., 25-75% yield) and high (i.e., top 25% of yield) yielding seasons. One season was randomly selected from each category for the model evaluation years and the remainder of the years were used for model calibration (Table 3.1). The calibration strategy was to sequentially adjust soil, cultivar and then soil coefficients a second time within their range of uncertainty to minimize errors between simulated and observed peanut maturity dates and yields for each county.

The CROPGRO-Peanut model was modified to allow the GeoSim optimizer to change three key soil parameters, including percent available soil water (*PASW*), root hospitality

reduction factor (*RHRF*) and the soil fertility factor that affects daily photosynthesis (*SLPF*). The optimization parameter *PASW* was used to adjust the lower limit (*LL*) in each soil layer to change available water holding capacity using the following equations (1)- (3).

The available soil water (*ASW*) for each soil layer was defined as:

$$ASW_i = DUL_i - LL_i, \quad (1)$$

where i was the number of soil layers, *DUL* was drained upper limit (field capacity) for layer i and *LL* was the lower limit for layer i.

The updated available soil water (*uASW*) determined by the optimizer was calculated by the optimization parameter *PASW*:

$$uASW_i = ASW_i + ASW_i \times PASW / 100, \quad (2)$$

Finally, the updated *LL* for each soil layer was adjusted by the following:

$$LL_i = DUL_i - uASW_i, \quad (3)$$

The range of *PASW* was set from -40 to 40% in this study. When the value of *PASW* is positive, the *ASW* is increased by reducing the default *LL* in each layer of the soil profile. Larger values of *PASW* give higher available soil water content in each soil layer.

The root hospitality reduction factor (*RHRF*) was used to adjust the soil root growth factor (*SRGF*) in each soil layer at soil layer depths (*DS*) below 60 cm. The *SRGF* factor was defined as 1.0 in the top 60 cm soil layer (Table 3.A1). In this study, the value of *RHRF* ranged from -0.1 indicating less or no root growth to 0 indicating more root growth below 60 cm. The crop model was modified to use the *RHRF* optimization parameter to adjust the soil *SRGF* factor in each soil layer by:

$$SRGF_i = 1,\ DS_i \leq 60 \text{ cm}$$
$$SRGF_i = RHRF \times (DS_i - 60) + 1,\ DS_i > 60 \text{ cm} \quad (4)$$

where DS_i was the depth of soil in layer i.

The soil fertility effect on photosynthesis (*SLPF*) is used by the model to represent limitations in crop growth attributed to factors other than water and nitrogen (i.e., pH, micronutrients). In the CROPGRO-Peanut model, daily photosynthesis is computed as a function of genetics, daily temperature, water and nitrogen stress, and CO_2. Daily photosynthesis can be modified by *SLPF* by:

$$PG_d = PGDAY / 44.0 \times 30.0 \times SLPF, \quad (5)$$

where d is the day of the year, *PGDAY* is the maximum daily photosynthesis as a function of photosynthetically active radiation, and *SLPF* is a reduction factor due to often unknown soil properties. An *SLPF* value of 1 indicates no reduction in daily canopy photosynthesis, while a value <1 reduces daily photosynthesis each day during the season.

The model calibration procedure included three steps. First, the model was calibrated using the GeoSim optimizer with initial cultivar coefficients for each county to obtain a first estimate of soil properties to minimize error in county-level yields. In Step two, the DSSAT-PEST optimizer was used to estimate cultivar coefficients using the first estimate of soil parameters from GeoSim (from step one). The final step was to use the calibrated cultivar coefficients to re-estimate soil parameters using the GeoSim optimizer and the optimum cultivar coefficients from DSSAT-PEST. The lower and upper bounds, and definition of the peanut cultivar coefficients and soil parameters are shown in Table 3.3.

3.3.5 Model Evaluation Metrics

To evaluate the model performance and accuracy for simulating maturity dates and yields, statistical indicators including coefficient of determination (R^2), root mean square error (*RMSE*), normalized root mean square error (*NRMSE*), mean absolute percentage error (*MAPE*) and model efficiency (*E*) were computed from observed (O_i) and simulated (S_i) values.

$$R^2 = 1 - \sum (O_i - S_i)^2 / \sum (O_i - O_a)^2, \quad (6)$$

$$RMSE = [(\sum (O_i - S_i)^2) / n]^{0.5}, \quad (7)$$

$$NRMSE = RMSE / O_a \times 100, \quad (8)$$

$$MAPE = 1 / n \times \sum | (O_i - S_i) / O_i |, \quad (9)$$

$$E = 1 - \sum (O_i - S_i)^2 / \sum (O_i - O_a)^2, \quad (10)$$

Where O_i is the i-th observation value, S_i is the i-th simulation value, O_a is the average value of a series observations and n is the sample number.

3.3.6 Assessing Climate Change Impacts on Peanut Production

The evaluated CROPGRO-Peanut model was then used to estimate the impact of climate change on peanut yields and irrigation water demand. The effects of climate change on both rainfed and irrigated yields were simulated for the 14-year baseline (2006-2019) weather data, while the future climate scenarios for year 2050s were simulated at 487 ppm CO_2 (RCP 4.5) and 541 ppm CO_2 (RCP 8.5) and year 2080s were simulated at 531 ppm CO_2 (RCP 4.5) and 758 ppm CO_2 (RCP 8.5) based on the Intergovernmental Panel on Climate Change (IPCC)

projections (Meinshausen et al., 2013). Other variables such as soil, variety and management practices were held constant. Irrigation management was simulated with the auto-irrigation using a management depth of 60 cm of soil and irrigation efficiency of 90%, assuming that full irrigation to meet peanut water demand was the dominant practice. Automatic irrigation was assumed to start when 60% of available water was depleted in the root zone, which was a common irrigation strategy (USDA-NASS, 2017). Analyses of variance (ANOVA) and Tukey's test were conducted using the GLM procedure of SAS (SAS Institute, 2013) to identify the climate change scenarios that resulted in statistically different simulation outputs. A Pearson correlation analysis was performed by SAS (SAS Institute, 2013) to determine the relationship between simulated peanut yields and growing season mean maximum temperature (T_{max}), minimum temperature (T_{min}), total precipitation (P), evapotranspiration (ET), irrigation amount.

3.4 Results

3.4.1 Optimum County-Level Cultivar Coefficients and Soil Parameters

Parameter calibration is a common practice to overcome uncertainties in model inputs, parameters, model structure and data errors (Bilionis et al., 2015; Huang et al., 2016). A three-step parameter estimation technique as described above was implemented to minimize error between simulated and observed peanut maturity dates and county-level yields. This resulted in a single set of soil properties and cultivar coefficients that represented the yield response of many soils and different varieties and management practices in each county (Table 3.4).

After calibration, the representative cultivar coefficient *EM-FL* ranged from 16.9 to 27.5 with an average value of 20.4 photothermal days (Table 3.4). The values of *FL-SH* and *FL-SD* did not vary much in different counties, with average values of 8.7 ± 2.1 and 16.7 ± 0.9 photothermal days, respectively. The *SD-PM* coefficient ranged from 60.4 photothermal days in Santa Rosa County to 78.0 photothermal days in Houston County (Table 3.4). The calibrated phenological parameters *EM-FL*, *FL-SH*, *FL-SD*, and *SD-PM* gave a good fit between simulated and observed peanut maturity dates for each county (Figure 3.2a, b).

Santa Rosa and Houston Counties had lower values for the *LFMAX* coefficient (1.49 and 1.31 mg CO_2/ (m^2 s), respectively) compared to other counties (1.50 mg CO_2/ (m^2 s)), indicating the composite of peanut varieties planted in these two counties had lower maximum leaf photosynthesis rate under optimum conditions. The coefficients *SLAR* and *SIZELF* are related to leaf area index (LAI) in the model. The calibration gave average values of 258.9 ± 27.5 cm^2/g for *SLAR* and 19.8 ± 2.1 cm^2/g for *SIZELF*. The simulated harvest yields for each county were calibrated by adjusting the representative cultivar coefficients *XFRT*, *WTPSD*, *SFDUR*, and *PODUR*. All the counties had a value of 1.00 for *XFRT*. In the crop model, weight per seed and seed number are used to determine total seed weight. The calibrated value of *WTPSD* was lower in Houston and Worth Counties (0.360 and 0.395 g, respectively) while the optimum *SFDUR* and *PODUR* were relatively higher compared to other counties. This indicated that the optimizer favored the creation of smaller seeds with a longer filling duration to represent the different varieties planted in these counties.

The initial soil properties for each county were taken from the dominant soil type in each county. Three soil parameters were adjusted for this soil type to represent the county-level yield response resulting from many different soil types in the county (Table 3.4). The optimizer reduced the available soil water (negative values of *PASW*) from the initial soil properties for Santa Rosa County, which had higher average seasonal rainfall (655 mm) compared to the other counties (Table 3.2). However, the optimizer added water holding capacity (positive values for *PASW*) to the other counties which had lower average seasonal rainfall, which indicated that drought stress was too severe using the default available water holding capacity in soil profile (Table 3.4). The second soil input that was adjusted was the soil root growth factor (*SRGF*) for each soil layer (Table 3.A1). This input factor governs the root growth (i.e., root length volume) that is allocated to each soil layer. The slope of tapering was designed as an optimization parameter *RHRF* in the model, where a value of 0 gave an *SRGF* factor of 1.0 below 60 cm and a value close to -0.1 cut the roots off at 60 cm depth (Table 3.3). Intermediate values create root structures with different maximum root depths and distributions, which impact root water uptake and water stress. The optimizer selected values of *RHRF* for the slope of *SRGF* close to 0 (range from -0.001 to -0.012), which gave *SRGF* average values of 0.8 ± 0.2 in the soil layer 60-100 cm and 0.5 ± 0.4 in the soil layer 100-200 cm, respectively (Table 3.4, Table 3.A1). This had the effect of simulating a deeper root system and reducing water stress by allowing roots to access deep soil moisture later in the growing season.

The third soil parameter considered for calibration was *SLPF*, which can mimic the effects of poor management, reduction in plant stands, as well as the impact of spatial and systematic foliar diseases such as late leafspot (*Cercosporidium personatum*) which is common in peanuts. This parameter effectively reduces daily simulated photosynthetic rate and has a large effect on crop growth and yield. The optimizer was used to estimate values of *SLPF* in conjunction with the slope of *SRGF* and *PASW* to minimize error in annual peanut yields. The optimum values of *SLPF* ranged from 0.6 to 0.8 with an average of 0.7 (Table 3.4).

3.4.2 Model Calibration and Evaluation

The ability of the model to simulate peanut growth duration was assessed by comparing the simulated and observed values of maturity dates. The model simulated the maturity dates reasonably well for calibration years, with an overall R^2 value of 0.71, *RMSE* value of 6 days, *NRMSE* value of 0.04 and *MAPE* of 3% (Figure 3.2a). The calibrated model was evaluated for three separate years in each county that were randomly selected to represent a low, average, and high yielding seasons. Evaluation of the model indicated that the simulated peanut maturity dates agreed well with observed values, which was indicated by a R^2 value of 0.83 and values of *RMSE*, *NRMSE,* and *MAPE* of 5 days, 0.04 and 4%, respectively (Figure 3.2b).

Table 3.5 and Figure 3.2c and d show the simulated and observed yields for fourteen seasons and five counties after soil and cultivar coefficient calibration. Overall, the model gave a very good simulation of yields across the peanut belt. The calibrated model accurately

simulated the inter-annual variability of yields with an overall R^2 of 0.73. The overall *RMSE*, *NRMSE* and *MAPE* were 333 kg/ha, 0.09 and 8%, respectively, which is very good for large-scale simulation (Figure 3.2c). The overall model efficiency (*E*) was 0.73, which indicated that the model was a much better predictor of seasonal yields than the observed mean yields (Table 3.5). The final evaluation results confirmed a good yield simulation for independent years resulting from the model calibration method. The model simulation accuracy for yield was high with an overall R^2 of 0.76, *RMSE* of 429 kg/ha, *NRMSE* of 0.12 and *MAPE* of 10% (Figure 3.2d). The *E* was 0.71 further indicating the predictive skill of the model was high (Table 3.5).

The R^2 index for the calibration and evaluation years of individual counties was larger than 0.57 except for Santa Rosa and Worth Counties (Figure 3.3a, e, f). The indicators *RMSE* (range: 169 kg/ha to 440 kg/ha), *NRMSE* (range: 0.06 to 0.12) and *MAPE* (range: 5% to 11%) were reasonable for the calibration years for all five counties. For the evaluation years, the *RMSE* (range: 210-624 kg/ha), *NRMSE* (range: 0.06-0.18) and *MAPE* (range: 5%-16%) confirmed that the model estimated the peanut growth and yields reasonably well in individual counties. All of the *E* values were positive for both calibration and evaluation years, which further indicated a good model efficiency (Table 3.5). In this study, we found that the model underestimated high-level yields and overestimated low-level yields in some counties (e.g., Figure 3.3e, d). However, there were many years and counties where the simulation agreed very well with the observations, giving both higher R^2 and lower *RMSE*, *NRMSE* and *MAPE* (e.g., Figure 3.3c, h, j).

3.4.3 Model Applications for Climate Change Impacts

A case study was conducted to demonstrate the potential use of the county level calibration and evaluation procedure to study the impact of climate change on peanut production and potential irrigation water requirement at county scale (Figure 3.4 and Table 3.6). The calibrated CROPGRO-Peanut model was used to estimate the peanut rainfed and irrigated yields, seasonal ET, and irrigation amount for the baseline (2006-2019) and future periods (2050s and 2080s) under RCP 4.5 and 8.5 climate scenarios. Fourteen replicates of RCP 4.5 and 8.5 weather data were compared to 14 seasons of historical weather data (baseline, 2006-2019) in order to evaluate the statistical significance of changes in yield.

For future climate scenarios using current varieties and management practices, simulated rainfed yields decreased from 1.4%-37.4% under RCP 4.5 and 8.5 scenarios relative to the baseline period in four counties (Table 3.6), which was caused by the significant negative impacts of increased future temperature (Figure 3.4a, b). In contrast, simulated rainfed yields increased (range: 28.2%-42.4%) for Martin County, NC (Table 3.6), which would benefit from increasing temperature (Figure 3.4a, b). However, these rainfed yield changes were not statistically significant for all the RCPs and climate scenarios which was probably due to mix-effects of the projected increasing temperature, rainfall, and CO_2 levels. Relative to the baseline, there was no significant difference in rainfed ET under different climate scenarios for Houston, Worth, and Orangeburg Counties. A significant increase ($p< 0.05$, range: 6.4%-17.2%) in rainfed ET was observed for future climate scenarios in both Santa Rosa and Martin Counties (Table 3.6).

The response of peanut yield to climate change scenarios under irrigation conditions was also simulated. A significant increase ($p< 0.05$, range: 8.5%-61.8%) of irrigated yield was observed in Santa Rosa, Houston and Martin Counties (Table 3.6). For Worth and Orangeburg Counties, the projected irrigated yield (Range: -8.5%-2.7%) varied around the baseline yield, with the highest decline occurring in the year 2080s under RCP 8.5 due to the significant negative effects of higher daytime temperature (Figure 3.4d, e). The irrigated ET for all counties showed a consistent significant increase ($p< 0.05$, range: 7.0%-26.8%) under both RCP 4.5 and 8.5 in future periods (2050s and 2080s), which had significant positive correlation with irrigated yields (Figure 3.4h). The irrigation water requirement for Santa Rosa, Houston and Worth Counties was projected to increase significantly ($p< 0.05$, range: 33.6%-220.5%) under future periods compared with the baseline periods in order to maximize yields. However, there was no significant difference observed in Orangeburg and Martin Counties related to irrigation amount (Table 3.6). For drier counties (e.g., Worth and Orangeburg Counties) with relatively lower seasonal rainfall and increasing temperature, a larger amount of irrigation water (Range: 161-330 mm) was required in the baseline and future periods in order to maximize yields, especially under the RCP 8.5 scenarios. While, for wet counties with consistently higher seasonal rainfall (e.g., Santa Rosa and Martin Counties), a less irrigation amount (Range: 60-167 mm) during growth period can achieve higher peanut yields in the future (Table 3.6).

3.5 Discussion

The focus of this research was to determine if one set of soil properties and cultivar coefficients could be calibrated to represent county scale peanut production. One challenge is that much uncertainty exists at the county scale in the distribution of crop varieties, soil properties, weather conditions, crop management practices and diseases within a county. The calibration strategy developed in our work provided excellent simulations of peanut growth duration and county-level yields, which could potentially be used for climate change studies (Jin et al., 2018; Huang et al., 2019; Wang et al., 2020).

3.5.1 Model Performance and Sources of Uncertainties

Unlike other crops, genetic diversity in commercial peanut varieties is limited. There were a limited number of peanut varieties available for planting, with Georgia 06G dominating the market in the Southeastern USA since 2006 (Brown et al., 2005; Branch and Culbreath, 2018). Therefore, the cultivar coefficients of Georgia 06G (Runner) was selected as the initial model inputs for our study period and was extracted from previous research (Prostko et al., 2012; Re et al., 2020). The approach in this work assumed that we could develop a set of representative cultivar coefficients that represented the inter-annual yield response of the mix of cultivars planted in each county. Adjustments made by the optimizer reflected the uncertainty in varieties planted within each county. Thus, the calibrated representative cultivar coefficients represented a composite of all varieties planted, harvested, and reported in the NASS datasets (Suriharn et al., 2011; Putto et al., 2013; Narh et al., 2015).

Counties are comprised of many different soil types with different soil properties. Peanuts are usually planted in sandy or loamy sand soil with limited water holding capacity in

the Southeastern USA (Devi et al., 2013; Mylavarapu et al., 2014). Therefore, soil water availability was a dominant factor controlling peanut growth and inter-annual variation of peanut yield in this region (Huang et al., 2021). Water stress in the CROPGRO-Peanut model is a function of available soil water determined by the difference between the drained upper limit (*DUL*), lower limit (*LL*), thickness of a soil layer and the root length volume of roots in each soil layer. Water stress occurs when the evapotranspiration demand exceeds the root water uptake. To account for different soil types within the county, the soil parameter *PASW* was used to adjust the lower limit (*LL*) uniformly across soil layers to increase or decrease water holding capacity and ultimately, the pattern of water stress (Singh et al., 2012). In the model, root length density in each soil layer governs how much water can be taken up from that soil layer to satisfy transpiration demand. An optimization parameter, *RHRF*, was incorporated into the model to adjust the soil root growth factor (*SRGF*) in each soil layer below 60 cm in a linear fashion. Several studies have reported the need to modify the *SRGF* distributions to calibrate soil water content and enhance drought tolerance of crop cultivars (Singh et al., 2014; Battisti et al., 2017). Our results show that the values for *SRGF* below 60 cm had to be increased in deeper soil layers to increase rooting depth and distribution (Table 3.4). This was required to simulate inter-annual drought patterns that minimized error between simulated and observed yield. Some peanut cultivars grown under sandy soils and drought conditions have been reported to develop roots more quickly to reach moisture deeper in the soil profile (Songsri et al., 2008).

The crop model has a soil fertility coefficient that is used to reduce daily photosynthesis resulting from poor fertility, or sub-optimum pH. The soil fertility factor (*SLPF*) is a 0 to 1 factor that affects the rate of daily photosynthesis and biomass production and accounts for the site-specific soil nutrient effects other than nitrogen (Narh et al., 2015). In this work, the optimizer found values for *SLPF* for each county that minimized error between simulated and observed yields. *SLPF* values ranged from 0.6 for Houston County to 0.8 for Santa Rosa and Martin Counties. Soils in the Southeastern USA were generally formed with abundant rainfall under temperate forests. Many soils in this region are acidic and low in fertility (Mylavarapu et al., 2014). The values of *SLPF* in this work were consistent with *SLPF* values reported before, where an *SLPF* value was 0.5 for low fertility soils and 1.0 for highly fertile soils (Jones et al., 1989; Narh et al., 2015).

The input parameters for management practices such as sowing and harvesting dates, population, and irrigation and fertilizer dates and amount are not available at the county level. Farmers follow different crop management practices depending on their soil, climate, and socio-economic conditions. The assumption of uniform crop management across a county can introduce errors for some years and counties (e.g., Figure 3.3e, d) (Manivasagam and Nagarajan, 2018). In addition to this, tomato spotted wilt, early and late leaf spot, stem and root rot, and white mold are commonly endemic diseases for peanuts planted in the Southeastern USA, which was not simulated by crop model (Branch and Brenneman, 2009; Branch and Fletchler, 2017; Standish et al., 2019). It is likely that the calibration of soil parameters *SLPF* mostly accounted for these unknown factors in each county.

3.5.2 Applications of Model for Climate Impacts Assessment

The case study demonstrated the potential use of this method of county-level model calibration to be used to evaluate future peanut production and potential irrigation demand under different climate change scenarios. If future production systems in the Southeastern USA remain non-irrigated with current varieties and management, most of the counties will experience a slight decline in yield, which is primarily due to future daytime temperatures frequently exceeding the optimum temperatures for peanuts (Figure 3.4a, b) (Vara et al., 2003; Eck et al., 2020; Shin et al., 2020). However, rainfed yield in Martin County would benefit from the increasing temperature because it is located in a relatively cooler climate zone that could potentially benefit from increased temperatures and rainfall in the future (Table 3.2, Figure 3.4a-c). Using the automatic irrigation strategy in the crop model, we determined the potential yield under optimum irrigation for the different climate change scenarios. Irrigation could potentially offset the negative impacts of heat and drought stress under climate change, which would significantly increase the yield under future climate scenarios in most of the counties (Figure 3.4i) (Lobell and Bonfils, 2008; Adhikari et al., 2016). To accomplish this, irrigation amount would have to increase in the future to cope with the significant increase in seasonal temperature and ET (Kothari et al., 2019). However, the actual irrigation demand varied among the five counties in the future scenarios, which can be attributed to spatial variability of seasonal rainfall and soil conditions across the Southeastern USA (Figure 3.4, Table 3.A1). Overall, supplemental irrigation during drought stress is critical to produce high

yield and top-quality peanuts under future climate change scenarios in the Southeastern Coastal Plain in the USA.

3.6 Conclusions

Simulating crop yield at the county scale is important for yield forecasting and studying the impact of climate change on production. Previous studies have shown that it is difficult to calibrate crop models for historical county-level yields over multiple seasons due to uncertainty or unavailability of many model inputs. The goal of this work was to determine if a single representative variety and soil type could be calibrated to simulate the long term inter-annual yield response of peanuts, and to evaluate the irrigation water demand under different climate change scenarios. Five major peanut producing counties across the Southeastern Coastal Plain in the USA peanut belt were selected to test this concept. Eleven cultivar coefficients and three soil parameters were calibrated using about eleven seasons of historical NASS data. The calibrated cultivar coefficients and soil parameters were evaluated using three additional seasons of data. Overall, the model simulated the peanut growth duration and county-level yields reasonably well for both calibrated and evaluated seasons, with high values of R^2 and low values of model errors (e.g., *RMSE*, *NRMSE, MAPE*). A case study was conducted using the calibrated model to evaluate the benefit of increasing irrigation water use for maintaining higher peanut production by reducing the negative impacts of climate change across the Southeastern USA. It was found that irrigation could mitigate the impact of climate change under different climate scenarios. The model calibration strategy developed in this study could be potentially used for other applications such as yield forecasting.

3.7 References

Adhikari, P., Ale, S., Bordovsky, J.P., Thorp, K.R., Modala, N.R., Rajan, N., Barnes, E.M., 2016. Simulating future climate change impacts on seed cotton yield in the Texas High Plains using the CSM-CROPGRO-Cotton model. Agric. Water. Manag. 164, 317-330.

Batchelor, W., Suresh, L., Zhen, X., Beyene, Y., Wilson, M., Kruseman, G., Prasanna, B., 2020. Simulation of maize lethal necrosis (MLN) damage using the CERES-maize model. Agronomy. 10, 710.

Battisti, R., Sentelhas, P.C., Boote, K.J., Camara, G.M.d.S., Farias, J.R., Basso, C.J., 2017. Assessment of soybean yield with altered water-related genetic improvement traits under climate change in Southern Brazil. Eur. J. Agron. 83, 1-14.

Bilionis, I., Drewniak, B., Constantinescu, E., 2015. Crop physiology calibration in the CLM. Geosci. Model. Dev. 8, 1071-1083.

Branch, W., 2007. Registration of 'Georgia‐06G' peanut. J. Plant. Regist. 1, 120-120.

Branch, W., Brenneman, T., 2009. Field evaluation for the combination of white mould and tomato spotted wilt disease resistance among peanut genotypes. Crop. Prot. 28, 595-598.

Branch, W., Culbreath, A., 2018. Transgressive segregation and long-term consistency for high TSWV field resistance in the 'Georgia-06G' peanut cultivar. Plant. Health. Prog. 19, 201-206.

Branch, W., Fletcher, S., 2017. Combination of Disease Resistance, Drought Tolerance, and Dollar Value among Runner and Virginia-Type Peanut Cultivars in Georgia. Peanut. Sci. 44, 42-46.

Brown, J.F., Pervez, M.S., 2014. Merging remote sensing data and national agricultural statistics to model change in irrigated agriculture. Agric. Syst. 127, 28-40.

Brown, S.L., Culbreath, A.K., Todd, J.W., Gorbet, D.W., Baldwin, J.A., Beasley, J.P., Jr., 2005. Development of a Method of Risk Assessment to Facilitate Integrated Management of Spotted Wilt of Peanut. Plant. Dis. 89, 348-356.

Dunne, J.P., John, J.G., Adcroft, A.J., Griffies, S.M., Hallberg, R.W., Shevliakova, E., Stouffer, R.J., Cooke, W., Dunne, K.A., Harrison, M.J., 2012. GFDL's ESM2 global coupled climate-carbon earth system models. Part I: Physical formulation and baseline simulation characteristics. J. Clim. 25, 6646-6665.

Devi, J.M., Rowland, D.L., Payton, P., Faircloth, W., Sinclair, T.R., 2013. Nitrogen fixation tolerance to soil water deficit among commercial cultivars and breeding lines of peanut. Field. Crops. Res. 149, 127-132.

Eck, M.A., Murray, A.R., Ward, A.R., Konrad, C.E., 2020. Influence of growing season temperature and precipitation anomalies on crop yield in the southeastern United States. Agric. For. Meteorol. 291, 108053.

Godfray, H.C., Beddington, J.R., Crute, I.R., Haddad, L., Lawrence, D., Muir, J.F., Pretty, J., Robinson, S., Thomas, S.M., Toulmin, C., 2010. Food security: the challenge of feeding 9 billion people. Science. 327, 812-818.

Grassini, P., van Bussel, L.G., Van Wart, J., Wolf, J., Claessens, L., Yang, H., Boogaard, H., de Groot, H., van Ittersum, M.K., Cassman, K.G., 2015. How good is good enough? Data

requirements for reliable crop yield simulations and yield-gap analysis. Field. Crops. Res. 177, 49-63.

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. Agric. Syst. 103, 256-264.

Hengl, T., de Jesus, J.M., MacMillan, R.A., Batjes, N.H., Heuvelink, G.B., Ribeiro, E., Samuel-Rosa, A., Kempen, B., Leenaars, J.G., Walsh, M.G., 2014. SoilGrids1km-global soil information based on automated mapping. PLoS One. 9, e105992.

Hengl, T., Mendes de Jesus, J., Heuvelink, G.B., Ruiperez Gonzalez, M., Kilibarda, M., Blagotic, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G., Ribeiro, E., Wheeler, I., Mantel, S., Kempen, B., 2017. SoilGrids250m: Global gridded soil information based on machine learning. PLoS One. 12, e0169748.

Hoogenboom, G., C.H. Porter, V. Shelia, K.J. Boote, U. Singh, J.W. White, L.A. Hunt, R. Ogoshi, J.I. Lizaso, J. Koo, S. Asseng, A. Singels, L.P. Moreno, and J.W. Jones. 2019. Decision Support System for Agrotechnology Transfer (DSSAT) Version 4.7.5 (https://DSSAT.net). DSSAT Foundation, Gainesville, Florida, USA.

Huang, J., Gómez-Dans, J.L., Huang, H., Ma, H., Wu, Q., Lewis, P.E., Liang, S., Chen, Z., Xue, J.-H., Wu, Y., 2019. Assimilation of remote sensing into crop growth models: Current status and perspectives. Agric. For. Meteorol. 276, 107609.

Huang, M., Ray, J., Hou, Z., Ren, H., Liu, Y., Swiler, L., 2016. On the applicability of surrogate‐based Markov chain Monte Carlo‐Bayesian inversion to the Community Land Model: Case studies at flux tower sites. J. Geophys. Res. Atmos. 121, 7548-7563.

Huang, J., Hartemink, A.E., Kucharik, C.J., 2021. Soil-dependent responses of US crop yields to climate variability and depth to groundwater. Agric. Syst. 190, 103085.

Jagtap, S.S., Jones, J.W., 2002. Adaptation and evaluation of the CROPGRO-soybean model to predict regional yield and production. Agric. Ecosyst. Environ. 93, 73-85.

Jin, X., Kumar, L., Li, Z., Feng, H., Xu, X., Yang, G., Wang, J., 2018. A review of data assimilation of remote sensing and crop models. Eur. J. Agron. 92, 141-152.

Johnston, R.Z., 2013. Using the CERES-Maize model to create a geographically explicit grid based estimate of corn yield under climate change scenarios. Dissertation. University of Arkansas.

Jones, J., Boote, K., Hoogenboom, G., Jagtap, S., Wilkerson, G., 1989. SOYGRO V5. 42, Soybean crop growth simulation model. User's guide. Florida. Agri. Exp. Station J. 8304, 83.

Jones, J.W., Hoogenboom, G., Porter, C.H., Boote, K.J., Batchelor, W.D., Hunt, L., Wilkens, P.W., Singh, U., Gijsman, A.J., Ritchie, J.T., 2003. The DSSAT cropping system model. Eur. J. Agron. 18, 235-265.

Jones, P.G., Thornton, P.K., 2013. Generating downscaled weather data from a suite of climate models for agricultural modelling applications. Agric. Syst. 114, 1-5.

Kambiranda, D.M., Vasanthaiah, H.K., Katam, R., Ananga, A., Basha, S.M., Naik, K., 2011. Impact of drought stress on peanut (*Arachis hypogaea L.*) productivity and food safety. Plants. Environ. 249-272.

Kothari, K., Ale, S., Attia, A., Rajan, N., Xue, Q., Munster, C.L., 2019a. Potential climate change adaptation strategies for winter wheat production in the Texas High Plains. Agric. Water. Manag. 225, 105764.

Kothari, K., Ale, S., Bordovsky, J.P., Thorp, K.R., Porter, D.O., Munster, C.L., 2019b. Simulation of efficient irrigation management strategies for grain sorghum production over different climate variability classes. Agric. Syst. 170, 49-62.

Leng, G., Zhang, X., Huang, M., Yang, Q., Rafique, R., Asrar, G.R., Ruby Leung, L., 2016. Simulating county‐level crop yields in the Conterminous United States using the Community Land Model: The effects of optimizing irrigation and fertilization. J. Adv. Model. Earth Syst. 8, 1912-1931.

Lobell, D.B., Bonfils, C., 2008. The effect of irrigation on regional temperatures: A spatial and temporal analysis of trends in California, 1934–2002. J. Clim. 21, 2063-2071.

Lobell, D.B., Schlenker, W., Costa-Roberts, J., 2011. Climate trends and global crop production since 1980. Science. 333, 616-620.

Liu, Z., Ying, H., Chen, M., Bai, J., Xue, Y., Yin, Y., Batchelor, W.D., Yang, Y., Bai, Z., Du, M., 2021. Optimization of China's maize and soy production can ensure feed sufficiency at lower nitrogen and carbon footprints. Nature Food, 1-8.

Ma, H., Malone, R.W., Jiang, T., Yao, N., Chen, S., Song, L., Feng, H., Yu, Q., He, J., 2020. Estimating crop genetic parameters for DSSAT with modified PEST software. Eur. J. Agron. 115, 126017.

Ma, L., Hoogenboom, G., Saseendran, S., Bartling, P., Ahuja, L.R., Green, T.R., 2009. Effects of estimating soil hydraulic properties and root growth factor on soil water balance and crop production. Agron. J. 101, 572-583.

Manivasagam, V., Nagarajan, R., 2018. Rainfall and crop modeling-based water stress assessment for rainfed maize cultivation in peninsular India. Theor. Appl. Climatol. 132, 529-542.

Manivasagam, V., Rozenstein, O., 2020. Practices for upscaling crop simulation models from field scale to large regions. Comput. Electron. Agric. 175, 105554.

Meinshausen, M., Smith, S.J., Calvin, K., Daniel, J.S., Kainuma, M.L., Lamarque, J.-F., Matsumoto, K., Montzka, S.A., Raper, S.C., Riahi, K., 2011. The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. Clim. Chang. 109, 213-241.

Mylavarapu, R., Mitchell, C., Savoy, H., 2014. Soils of the Southeastern US. Soil test methods from the southeastern United States, 2. Soil test methods from the southeastern United States. Southern Cooperative Series Bulletin 419.

Narh, S., Boote, K.J., Naab, J.B., Jones, J., Tillman, B.L., Abudulai, M., Sankara, P., M'Bi Bertin, Z., Burow, M.D., Brandenburg, R.L., 2015. Genetic Improvement of Peanut Cultivars for West Africa Evaluated with the CSM-CROPGRO-Peanut Model. Agron. J. 107, 2213-2229.

National weather service- Cooperative observer program (NWS-COP), 2020. http://www.weather.gov/ (accessed 2 July 2021).

Prostko, E.P., Kemerait, R.C., Webster, T.M., 2012. Georgia-06G, Florida-07, and Tifguard Peanut Cultivar Response to Chlorimuron. Weed. Technol. 26, 429-431.

Putto, C., Patanothai, A., Jogloy, S., Boote, K., Hoogenboom, G., 2013. Determination of plant traits that affect genotype× location (G× L) interaction in peanut using the CSM-CROPGRO-Peanut model. Int. J. Plant. Prod. 7, 537-568.

QGIS.org, 2021. QGIS Geographic Information System. QGIS Association. https://www.qgis.org/en/site/ (accessed 2 July 2021).

Qin, H., Feng, S., Chen, C., Guo, Y., Knapp, S., Culbreath, A., He, G., Wang, M.L., Zhang, X., Holbrook, C.C., Ozias-Akins, P., Guo, B., 2012. An integrated genetic linkage map of cultivated peanut (*Arachis hypogaea L.*) constructed from two RIL populations. Theor. Appl. Genet. 124, 653-664.

Re, M.I.Z., Rath, S., Dukes, M.D., Graham, W., 2020. Water and Nitrogen Budget Dynamics for a Maize-Peanut Rotation in Florida. Trans. ASABE. 63, 2003-2020.

Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A.C., Muller, C., Arneth, A., Boote, K.J., Folberth, C., Glotter, M., Khabarov, N., Neumann, K., Piontek, F., Pugh, T.A., Schmid, E., Stehfest, E., Yang, H., Jones, J.W., 2014. Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. Proc. Natl. Acad. Sci. U.S.A. 111, 3268-3273.

R Core Team, 2021. R: A language and environment for statistical computing. http://www.R-project.org/ (accessed 2 July 2021).

Salazar, M., Hook, J., y Garcia, A.G., Paz, J., Chaves, B., Hoogenboom, G., 2012. Estimating irrigation water use for maize in the Southeastern USA: A modeling approach. Agric. Water. Manag. 107, 104-111.

SAS Institute Inc. 2013. SAS® 9.4 Statements: Reference. Cary, NC: SAS Institute Inc.

Shin, D., Cocke, S., Baigorria, G.A., Romero, C.C., Kim, B.-M., Kim, K.-Y., 2020. Future Crop Yield Projections Using a Multi-model Set of Regional Climate Models and a Plausible Adaptation Practice in the Southeast United States. Atmosphere 11, 1300.

Singh, P., Boote, K., Kumar, U., Srinivas, K., Nigam, S., Jones, J., 2012. Evaluation of Genetic Traits for Improving Productivity and Adaptation of Groundnut to Climate Change in India. J. Agron. Crop. Sci. 198, 399-413.

Singh, P., Nedumaran, S., Ntare, B., Boote, K.J., Singh, N.P., Srinivas, K., Bantilan, M., 2014. Potential benefits of drought and heat tolerance in groundnut for adaptation to climate change in India and West Africa. Mitig. Adapt. Strateg. Glob. Chang. 19, 509-529.

Songsri, P., Jogloy, S., Kesmala, T., Vorasoot, N., Akkasaeng, C., Patanothai, A., Holbrook, C., 2008. Heritability of drought resistance traits and correlation of drought resistance and agronomic traits in peanut. Crop. Sci. 48, 2245-2253.

Standish, J.R., Culbreath, A.K., Branch, W.D., Brenneman, T.B., 2019. Disease and Yield Response of a Stem-rot-resistant and -Susceptible Peanut Cultivar under Varying Fungicide Inputs. Plant. Dis. 103, 2781-2785.

Suriharn, B., Patanothai, A., Boote, K., Hoogenboom, G., 2011. Designing a peanut ideotype for a target environment using the CSM‐CROPGRO‐Peanut model. Crop. Sci. 51, 1887-1902.

Thorp, K.R., Bronson, K.F., 2013. A model-independent open-source geospatial tool for managing point-based environmental model simulations at multiple spatial locations. Environ. Model. Softw. 50, 25-36.

Thorp, K.R., DeJonge, K.C., Kaleita, A.L., Batchelor, W.D., Paz, J.O., 2008. Methodology for the use of DSSAT models for precision agriculture decision support. Comput. Electron. Agric. 64, 276-285.

Tilman, D., Balzer, C., Hill, J., Befort, B.L., 2011. Global food demand and the sustainable intensification of agriculture. Proc. Natl. Acad. Sci. U.S.A. 108, 20260-20264.

Tofa, A.I., Kamara, A.Y., Babaji, B.A., Akinseye, F.M., Bebeley, J.F., 2021. Assessing the use of a drought-tolerant variety as adaptation strategy for maize production under climate change in the savannas of Nigeria. Sci. Rep. 11, 1-16.

US Department of Agriculture (USDA)-Farm Service Agency (FSA), 2020. https://www.fsa.usda.gov (accessed 2 July 2021).

US Department of Agriculture (USDA)-National Agricultural Statistics Service (NASS), 2017. https://www.nass.usda.gov/AgCensus/index.php (accessed 2 July 2021).

US Department of Agriculture (USDA) - National Agricultural Statistics Service (NASS), 2020. https://quickstats.nass.usda.gov/ (accessed 2 July 2021).

US Department of Agriculture (USDA) - National Resources Conservation Service (NRCS), 2020. http://www.nrcs.usda.gov/wps/portal/nrcs/site/soils/home/ (accessed 2 July 2021).

US Department of Agriculture (USDA) - Risk Management Agency (RMA), 2020. https://www.rma.usda.gov/ (accessed 2 July 2021).

Vara Prasad, P., Boote, K.J., Hartwell Allen Jr, L., Thomas, J.M., 2003. Super‐optimal temperatures are detrimental to peanut (*Arachis hypogaea L.*) reproductive processes and yield at both ambient and elevated carbon dioxide. Glob. Chang. Biol. 9, 1775-1787.

Wang, X., Huang, J., Feng, Q., Yin, D., 2020. Winter wheat yield prediction at county level and uncertainty analysis in main wheat-producing regions of China with deep learning approaches. Remote. Sens. 12, 1744.

Yan, W., Jiang, W., Han, X., Hua, W., Yang, J., Luo, P., 2020. Simulating and predicting crop yield and soil fertility under climate change with fertilizer management in Northeast China based on the decision support system for agrotechnology transfer model. Sustainability. 12, 2194.

Zhen, X., Shao, H., Zhang, W., Huo, W., Batchelor, W.D., Hou, P., Wang, E., Mi, G., Miao, Y., Li, H., 2018. Testing a bell-shaped function for estimation of fully expanded leaf area in modern maize under potential production conditions. Crop. J. 6, 527-537.

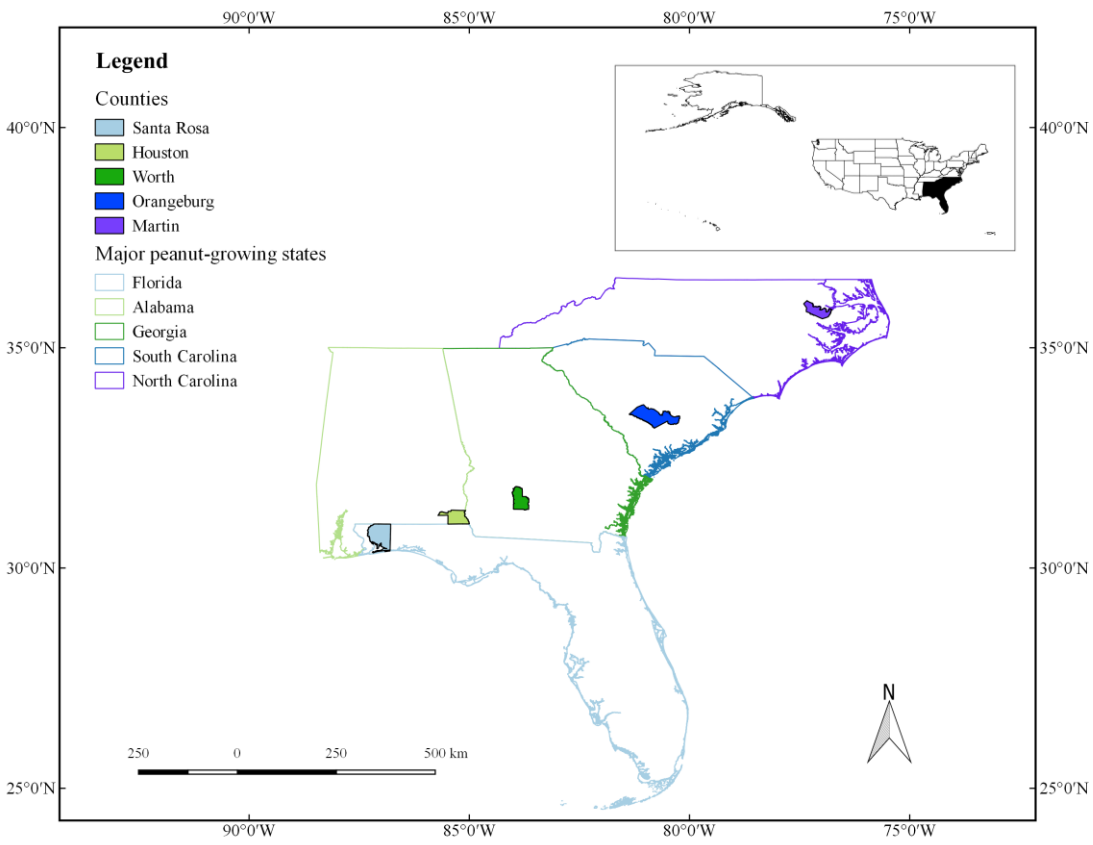


Figure 3.1 Study sites for the main peanut producing counties in each state in the Southeastern Coastal Plains, USA.

Table 3.1 Study sites, major peanut varieties, management practices and data resources for model simulation.

County/State	Major variety	Sowing date	Maturity date	Density (pl^f/m^2)	Years used for calibration	Years used for evaluation	Daily weather	Soil properties	Peanut phenology	Observed county yield
Santa Rosa, FL[a]	Georgia 06G	4/25-5/30	9/10-11/30	13	2007-2009, 2011-2015, 2017-2018	2006, 2010, 2016	NWS-COP[g]	gSSURGO[h]; SoilGrids for DSSAT	USDA[i]-NASS[j]; USDA-RMA[k]	USDA-NASS
Houston, AL[b]	Georgia 06G	4/25-5/25	9/22-10/22	13	2006-2009, 2011, 2013, 2015-2017	2010, 2012, 2014				
Worth, GA[c]	Georgia 06G	5/6-5/31	9/25-10/31	13	2006-2007, 2009-2012, 2014, 2016-2019	2008, 2013, 2015				
Orangeburg, SC[d]	Georgia 06G	4/25-5/15	10/2-11/1	13	2006-2009, 2011-2013, 2016-2019	2010, 2014, 2015				
Martin, NC[e]	Georgia 06G	5/10-5/30	10/10-10/30	13	2006-2009, 2012-2016, 2018-2019	2010, 2011, 2017				

a. Florida; b. Alabama; c. Georgia; d. South Carolina; e. North Carolina; f. Plants number; g. National Weather Service-Cooperative Observer Program; h. Gridded Soil Survey Geographic database; i. United States Department of Agriculture; j. National Agricultural Statistics Service; k. Risk Management Agency.

Table 3.2 Baseline and projected mean maximum temperature (T_{max}), minimum temperature (T_{min}), total precipitation (P) and atmospheric CO_2 concentration (CO_2) during the peanut growing season (from sowing to maturity) for top peanut producing counties of each state across the Southeastern USA.

County/State	Climate scenario	N[a]	T_{max} (◦C)	T_{min} (◦C)	P (mm)	CO_2 (ppm)
Santa Rosa, FL	Baseline (2006-2019)	14	32.1	20.8	655	380
	RCP 4.5 (2050s)	14	33.9	22.4	758	487
	RCP 8.5 (2050s)	14	34.4	23.0	691	541
	RCP 4.5 (2080s)	14	34.4	22.8	739	531
	RCP 8.5 (2080s)	14	35.8	24.3	685	758
Houston, AL	Baseline (2006-2019)	14	31.8	21.2	543	380
	RCP 4.5 (2050s)	14	34.0	21.2	660	487
	RCP 8.5 (2050s)	14	34.5	21.8	632	541
	RCP 4.5 (2080s)	14	34.4	21.6	685	531
	RCP 8.5 (2080s)	14	36.0	23.1	599	758
Worth, GA	Baseline (2006-2019)	14	32.7	21.4	431	380
	RCP 4.5 (2050s)	14	34.4	22.0	473	487
	RCP 8.5 (2050s)	14	34.9	22.6	480	541
	RCP 4.5 (2080s)	14	34.9	22.4	502	531
	RCP 8.5 (2080s)	14	36.4	23.9	465	758
Orangeburg, SC	Baseline (2006-2019)	14	32.4	19.7	495	380
	RCP 4.5 (2050s)	14	33.7	21.1	620	487
	RCP 8.5 (2050s)	14	34.2	21.7	612	541
	RCP 4.5 (2080s)	14	34.1	21.5	619	531
	RCP 8.5 (2080s)	14	35.7	23.0	610	758
Martin, NC	Baseline (2006-2019)	14	30.1	19.6	603	380
	RCP 4.5 (2050s)	14	31.4	19.6	730	487
	RCP 8.5 (2050s)	14	31.8	20.1	669	541
	RCP 4.5 (2080s)	14	31.7	20.0	696	531
	RCP 8.5 (2080s)	14	33.3	21.4	656	758

a. Sample numbers.

Table 3.3 List of cultivar coefficients and soil parameters in the CROPGRO-Peanut model used for model calibration.

Variable	Units	Lower bound	Upper bound	Definition of variable	References
Cultivar coefficients					
EM-FL	Photothermal days	15	28	Time between plant emergence and flower appearance	
FL-SH	photothermal days	5	11	Time between first flower and first pod	
FL-SD	photothermal days	16	25	Time between first flower and first seed	
SD-PM	photothermal days	50	85	Time between first seed and physiological maturity	
LFMAX	mg CO_2/ (m^2 s)	1	1.5	Maximum leaf photosynthesis rate at 30 ◦C, 350 vpm CO_2, and high light	
SLAR	cm^2/g	230	290	Specific leaf area of cultivar under standard growth conditions	Suriharn et al., 2011; Ma et al., 2020
SIZELF	cm^2	16	21	Maximum size of full leaf (three leaflets)	
XFRT	-	0.57	1.00	Maximum fraction of daily growth that is partitioned to seed + shell	
WTPSD	g	0.36	1.20	Maximum weight per seed	
SFDUR	photothermal days	23	46	Seed filling duration for pod cohort at standard growth conditions	
PODUR	photothermal days	14	34	Time required for cultivar to reach final pod load under optimal conditions	
Soil parameters					
PASW	-	-40	40	Percent available soil water	
RHRF	-	-0.1	0	Root hospitality reduction factor	Jones et al., 2003; Ma et al., 2009
SLPF	-	0	1.0	Soil fertility factor	

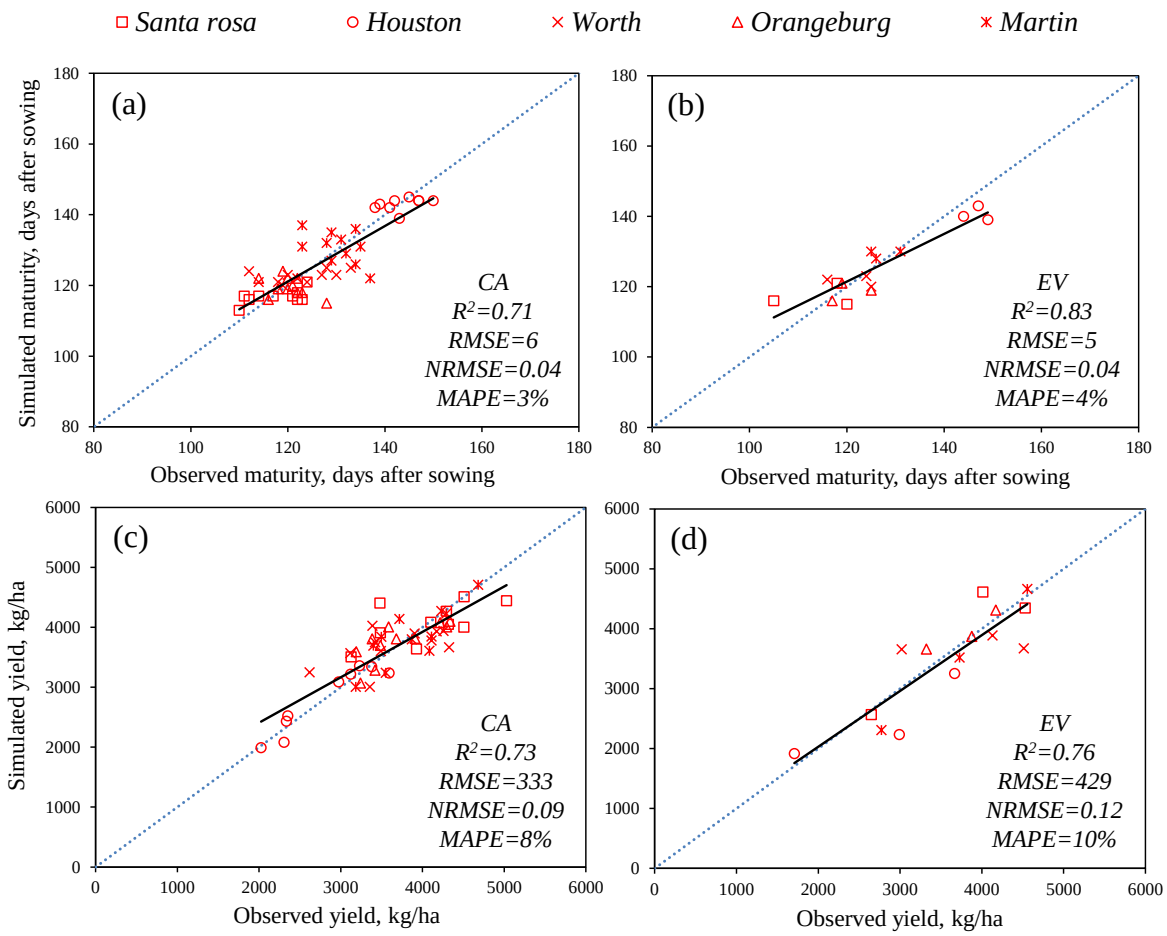
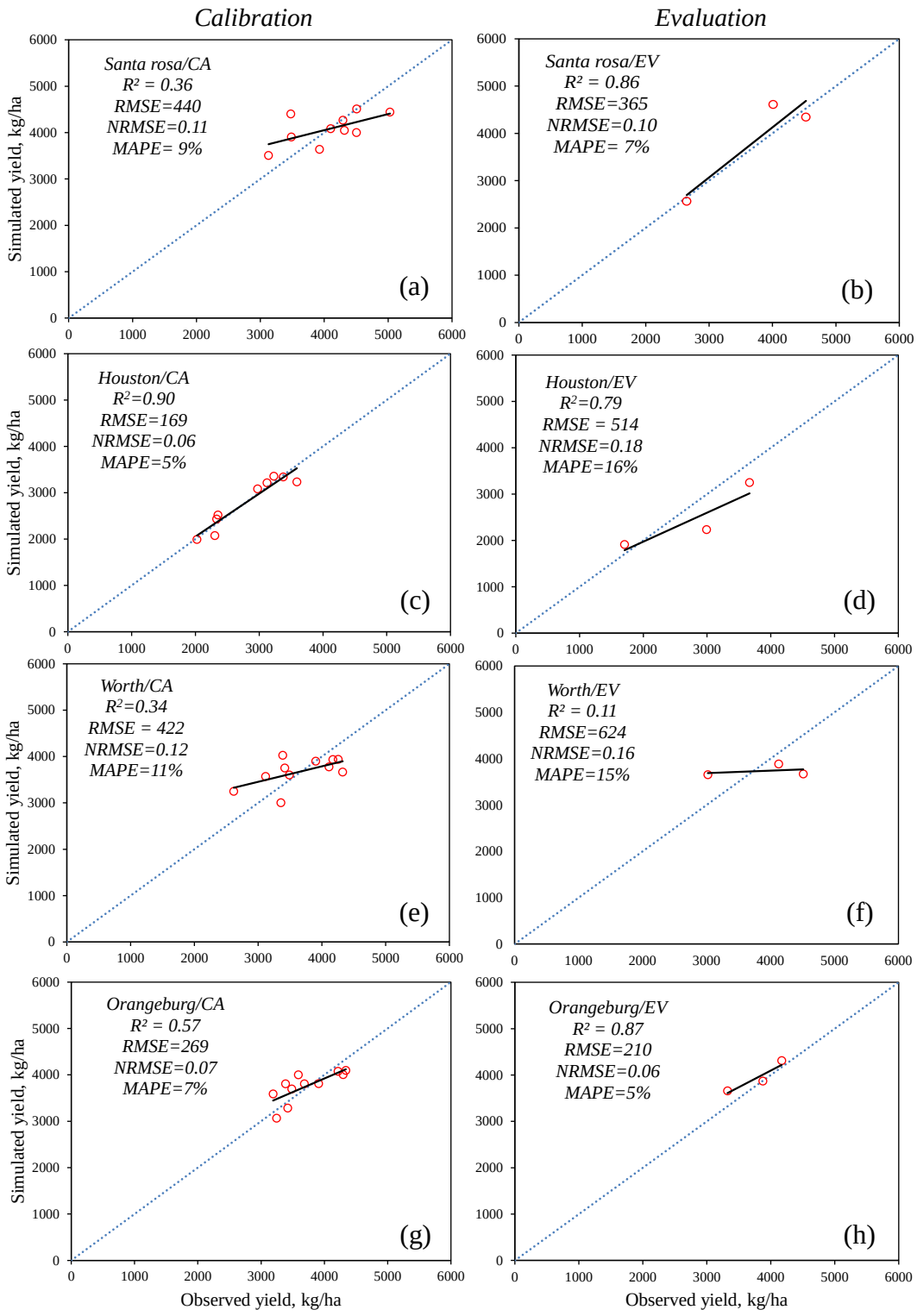


Figure 3.2 Simulated and observed maturity dates, harvest yields for the top peanut producing counties in each state across the Southeastern USA. Figures a and c show the final calibration (CA) results while figures b and d show the final evaluation (EV) results for all the counties. All yields were converted to 0% moisture. The dashed line and solid line represent the 1:1 line and regression trend line, respectively.

Table 3.4 Representative cultivar coefficients and soil parameters calibrated for each county.

County/State	Cultivar coefficients											Soil parameters		
	EM-FL **(Pd[a])**	***FL-SH*** **(Pd)**	***FL-SD*** **(Pd)**	***SD-PM*** **(Pd)**	***LFMAX*** **(mg CO_2/ (m^2 s))**	***SLAR*** **(cm^2/g)**	***SIZELF*** **(cm^2)**	***XFRT***	***WTPSD*** **(g)**	***SFDUR*** **(Pd)**	***PODUR*** **(Pd)**	***PASW***	***RHRF***	***SLPF***
Santa Rosa, FL	21.3	11.0	17.6	60.4	1.49	290	20.6	1.00	0.807	36.6	16.9	-18.5	-0.001	0.8
Houston, AL	27.5	6.8	16.0	78.0	1.31	268	20.8	1.00	0.360	45.9	14.0	24.8	-0.012	0.6
Worth, GA	18.1	7.0	18.0	69.2	1.50	230	20.7	1.00	0.395	32.7	25.3	39.8	-0.005	0.7
Orangeburg, SC	16.9	11.0	16.0	64.9	1.50	276	21.0	1.00	1.200	36.8	27.9	36.8	-0.004	0.7
Martin, NC	18.3	7.7	16.2	73.5	1.50	230	16.0	1.00	1.200	32.1	22.2	0.50	-0.001	0.8

a. Photothermal days


Calibration
Evaluation
Santa rosa/CA
R² = 0.36
RMSE=440
NRMSE=0.11
MAPE= 9%
(a)
Santa rosa/EV
R² = 0.86
RMSE=365
NRMSE=0.10
MAPE= 7%
(b)
Houston/CA
R²=0.90
RMSE=169
NRMSE=0.06
MAPE=5%
(c)
Houston/EV
R²=0.79
RMSE = 514
NRMSE=0.18
MAPE=16%
(d)
Worth/CA
R²=0.34
RMSE = 422
NRMSE=0.12
MAPE=11%
(e)
Worth/EV
R² = 0.11
RMSE=624
NRMSE=0.16
MAPE=15%
(f)
Orangeburg/CA
R² = 0.57
RMSE=269
NRMSE=0.07
MAPE=7%
(g)
Orangeburg/EV
R² = 0.87
RMSE=210
NRMSE=0.06
MAPE=5%
(h)
Simulated yield, kg/ha
Observed yield, kg/ha
0
1000
2000
3000
4000
5000
6000

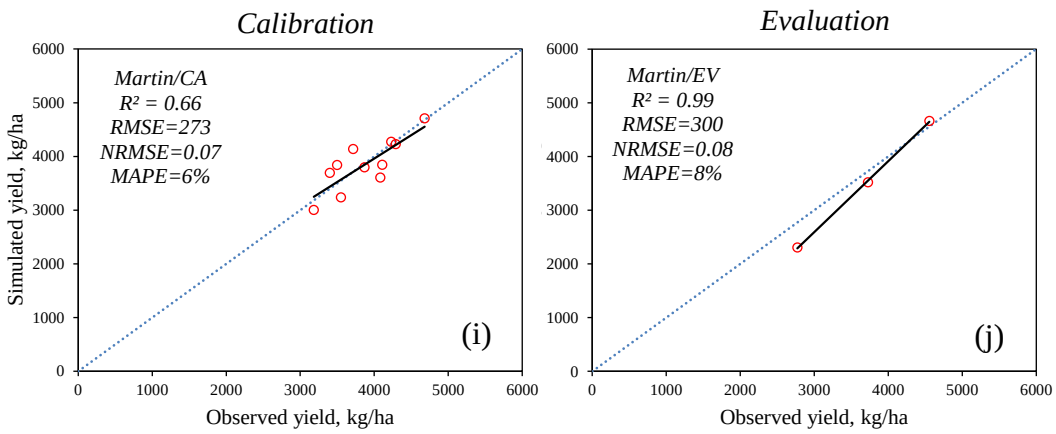


Figure 3.3 Calibration and evaluation results for top peanut producing counties in each state across the Southeastern USA. Figures a, c, e, g, and i show the final calibration (CA) results while figures b, d, f, h, and j show the final evaluation (EV) results. All yields were converted to 0% moisture. The dashed line and solid line represent the 1:1 line and regression trend line, respectively.

Table 3.5 Results of the CROPGRO-Peanut model calibration and evaluation for simulating NASS reported county-level yields in top peanut producing counties of each state across the Southeastern USA.

County/State	**N[c]**	***R²***	***RMSE* (kg/ha)**	***NRMSE***	***MAPE***	***E***
CA[a]						
Santa Rosa, FL	10	0.36	440	0.11	9%	0.36
Houston, AL	9	0.90	169	0.06	5%	0.90
Worth, GA	11	0.34	422	0.12	11%	0.34
Orangeburg, SC	11	0.57	269	0.07	7%	0.55
Martin, NC	11	0.66	273	0.07	6%	0.59
Overall	52	0.73	333	0.09	8%	0.73
EV[b]						
Santa Rosa, FL	3	0.86	365	0.10	7%	0.79
Houston, AL	3	0.79	514	0.18	16%	0.60
Worth, GA	3	0.11	624	0.16	15%	0.03
Orangeburg, SC	3	0.87	210	0.06	5%	0.64
Martin, NC	3	0.99	300	0.08	8%	0.83
Overall	15	0.76	429	0.12	10%	0.71

a. Calibration; b. Evaluation; c. Sample numbers.

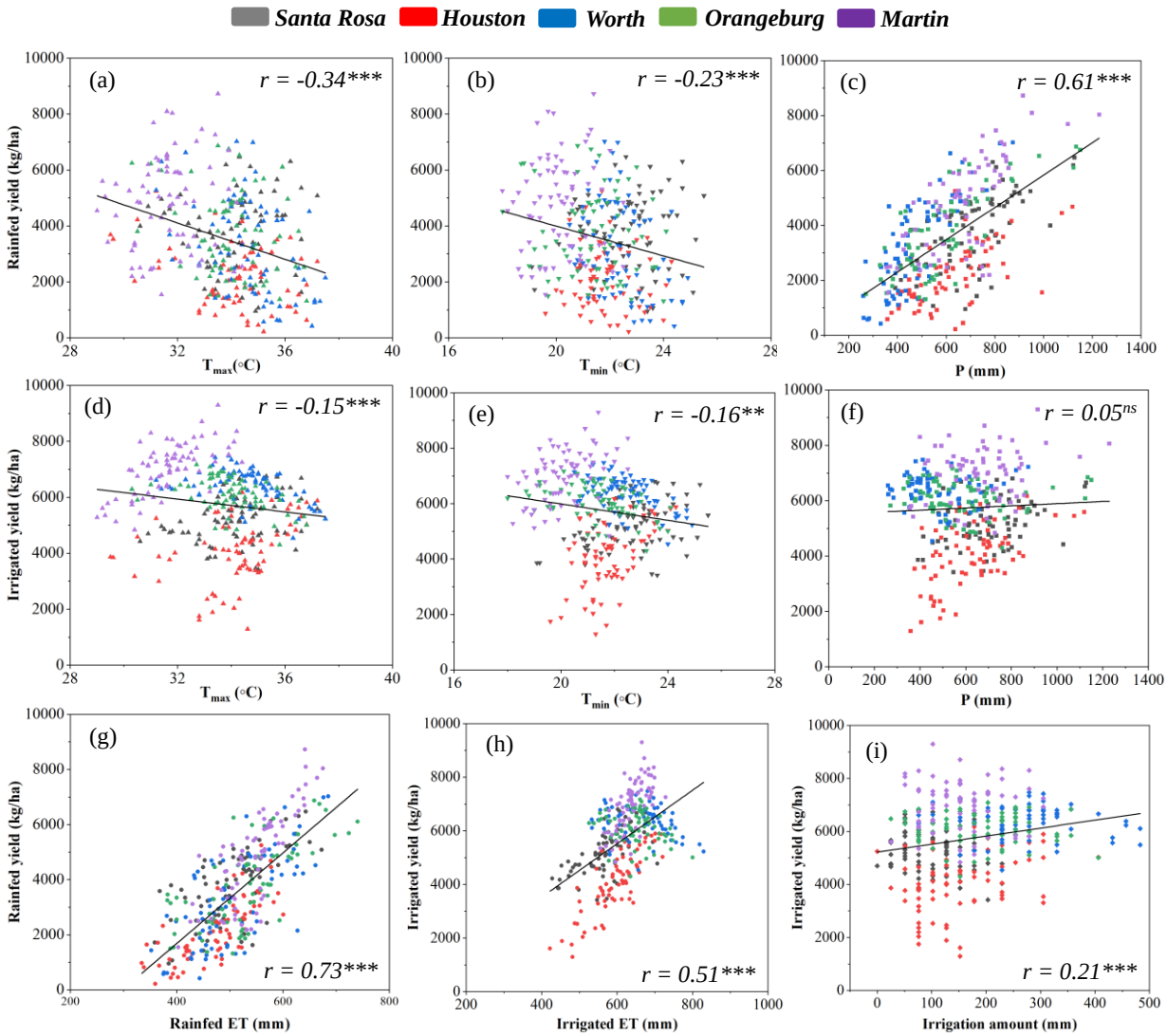


Figure 3.4 Correlation between growing season mean maximum temperature (T_{max}), minimum temperature (T_{min}), total precipitation (P), evapotranspiration (ET), irrigation amount and simulated peanut yields in the baseline and future periods across the Southeastern USA. Figures a, b, c, and g show the correlation for rainfed yield while figures d, e, f, h, and i show the correlation for irrigated yield. The solid line represents the regression trend line. ^{ns} Not significant. * Significant at 0.05 level. ** Significant at 0.01 level. *** Significant at 0.001 level.

Table 3.6 Effects of climate scenarios on simulated rainfed and irrigated yields, seasonal evapotranspiration (ET), and irrigation amount in top peanut producing counties of each state across the Southeastern USA. Mean values with the same letter are not significantly different at $p< 0.05$ based on Tukey's test.

County/State	Climate scenario	N[a]	Rainfed yield (kg/ha)	Rainfed ET (mm)	Irrigated yield (kg/ha)	Irrigated ET (mm)	Irrigation amount (mm)
Santa Rosa, FL	Baseline (2006-2019)	14	3926a	465b	4553b	494c	60b
	RCP 4.5 (2050s)	14	3789a	515ab	5219ab	577ab	103ab
	RCP 8.5 (2050s)	14	3552a	495ab	5219ab	568b	123a
	RCP 4.5 (2080s)	14	3870a	514ab	5281ab	576ab	109ab
	RCP 8.5 (2080s)	14	3920a	545a	5763a	621a	125a
Houston, AL	Baseline (2006-2019)	14	2571a	484a	3261b	533c	82c
	RCP 4.5 (2050s)	14	1994a	469a	3537b	585bc	109c
	RCP 8.5 (2050s)	14	2057a	470a	4029b	600ab	122c
	RCP 4.5 (2080s)	14	2279a	477a	4089b	600ab	194b
	RCP 8.5 (2080s)	14	1904a	476a	5276a	650a	261a
Worth, GA	Baseline (2006-2019)	14	3714a	474a	6442ab	573c	189b
	RCP 4.5 (2050s)	14	3479a	511a	6619a	662b	272a
	RCP 8.5 (2050s)	14	3002a	506a	6390ab	675ab	287a
	RCP 4.5 (2080s)	14	3555a	521a	6511a	670b	258ab
	RCP 8.5 (2080s)	14	2323a	519a	5897b	726a	330a
Orangeburg, SC	Baseline (2006-2019)	14	3983a	484a	6145a	582c	161a
	RCP 4.5 (2050s)	14	3898a	539a	6256a	645b	178a
	RCP 8.5 (2050s)	14	3703a	540a	6057a	662ab	194a
	RCP 4.5 (2080s)	14	3917a	547a	6181a	658ab	183a
	RCP 8.5 (2080s)	14	3210a	548a	5714a	696a	229a
Martin, NC	Baseline (2006-2019)	14	3765a	506b	5627c	601b	163a
	RCP 4.5 (2050s)	14	5359a	567a	7039b	644a	133a
	RCP 8.5 (2050s)	14	4909a	550ab	7265b	643a	158a
	RCP 4.5 (2080s)	14	5249a	559ab	7343ab	649a	147a
	RCP 8.5 (2080s)	14	4824a	555ab	7857a	661a	167a

a. Sample numbers.

Table 3.A1 Soil profile characteristics for top peanut producing counties in each state across the Southeastern USA.

Soil depth (cm)	LL (mm/mm)	DUL (mm/mm)	SAT (mm/mm)	SRGF	BD (g/cm^3)	Org. C (%)	Clay (%)	Silt (%)	Total N (%)	pH
Santa Rosa, FL[a]										
0-5	0.15	0.24	0.39	1.0	1.46	2.18	20.6	33.8	0.12	5.1
5-15	0.16	0.26	0.40	1.0	1.48	1.85	22.5	32.9	0.09	5.2
15-30	0.17	0.27	0.40	1.0	1.51	1.41	24.9	31.7	0.07	5.3
30-60	0.19	0.29	0.41	1.0	1.56	0.90	27.4	30.4	0.06	5.4
60-100	0.19	0.28	0.41	1.0	1.62	0.52	27.2	29.8	0.05	5.5
100-200	0.17	0.27	0.40	0.9	1.67	0.30	25.5	29.5	0.05	5.7
Houston, AL[b]										
0-5	0.12	0.27	0.40	1.0	1.51	1.68	24.3	34.8	0.12	5.3
5-15	0.13	0.29	0.41	1.0	1.53	1.42	26.3	33.9	0.09	5.3
15-30	0.14	0.30	0.41	1.0	1.56	1.09	29.0	32.7	0.07	5.4
30-60	0.16	0.32	0.42	1.0	1.61	0.70	31.4	31.3	0.06	5.5
60-100	0.16	0.32	0.42	0.5	1.67	0.41	31.4	30.6	0.05	5.7
100-200	0.15	0.30	0.41	0.0	1.72	0.23	29.6	30.3	0.05	5.9
Worth, GA[c]										
0-5	0.07	0.23	0.39	1.0	1.60	1.84	19.2	33.1	0.12	5.0
5-15	0.08	0.24	0.39	1.0	1.62	1.56	21.0	32.3	0.09	5.1
15-30	0.09	0.26	0.40	1.0	1.65	1.19	23.5	31.1	0.07	5.2
30-60	0.11	0.27	0.40	1.0	1.70	0.76	25.6	29.9	0.06	5.3
60-100	0.11	0.27	0.40	0.8	1.76	0.44	25.7	29.2	0.05	5.5
100-200	0.10	0.26	0.40	0.3	1.81	0.25	24.1	28.9	0.05	5.6
Orangeburg, SC[d]										
0-5	0.07	0.24	0.39	1.0	1.52	2.18	19.3	36.1	0.12	5.2
5-15	0.08	0.25	0.40	1.0	1.54	1.84	21.1	35.1	0.09	5.3
15-30	0.10	0.27	0.40	1.0	1.57	1.41	23.7	33.9	0.07	5.4
30-60	0.11	0.28	0.41	1.0	1.62	0.90	26.0	32.6	0.06	5.5
60-100	0.11	0.28	0.40	0.8	1.68	0.52	25.9	32.0	0.05	5.6
100-200	0.10	0.27	0.40	0.5	1.73	0.30	24.3	31.6	0.05	5.8
Martin, NC[e]										
0-5	0.11	0.23	0.39	1.0	1.54	2.72	17.7	34.5	0.12	5.1
5-15	0.12	0.24	0.39	1.0	1.56	2.30	19.5	33.7	0.09	5.2
15-30	0.13	0.25	0.40	1.0	1.59	1.75	22.0	32.5	0.07	5.3
30-60	0.15	0.27	0.40	1.0	1.64	1.12	24.3	31.3	0.06	5.4
60-100	0.15	0.26	0.40	1.0	1.70	0.65	24.2	30.6	0.05	5.5
100-200	0.14	0.25	0.40	0.9	1.75	0.38	22.5	30.3	0.05	5.7

a. Florida; b. Alabama; c. Georgia; d. South Carolina; e. North Carolina.

Chapter 4. Combining genomics and crop modelling to simulate maize (*Zea mays L.*) yield and its component traits

4.1 Abstract

Simulation of the adaptive capacity of existing and new germplasms using process-based crop model and genetic information can efficiently assist in determining the potential of well-adapted genotypes for target environments. To achieve the integration of genomic prediction with crop modelling, we developed an integrated marker-based modelling approach by detecting associated genetic markers for essential model input parameters and incorporating the genetic effects of these associated markers into CERES-Maize model. The performance of the marker-based modelling in simulating yield and its component traits was tested using four observed sub-datasets including: (i) observed genotypes grown in observed environments; (ii) observed genotypes phenotyped in new environments; (iii) new genotypes in characterized environments; and (iv) new genotypes in new environments. One outcome of this study is to improve marker-based modelling and the prediction of genotype performance for plant breeding. The marker-based model in this study reasonably simulated the anthesis date, kernel number, kernel weight and yield for all four sub-datasets. Another outcome is to assist in quantitatively assessing the effect of genes by explicitly accounting for genotype by environment interactions. The prediction performance of marker-based modelling was either affected by new genotypes or new environments depending on the traits being simulated. Several limiting factors for marker-based modelling need to be considered, including phenotyping traits and environments, marker effects, statistical methods, and model input parameters. Breeding programs could further exploit marker-based modelling to predict adaptation in diverse environmental and management conditions for new genotypes before they are globally distributed for multilocation yield testing.

Keywords: Genomic prediction; Marker-based modelling; Genotype–environment interactions; Yield; *Zea mays L*

4.2 Introduction

Maize (*Zea mays L*.) is one of the most important global staple crops, and is planted over most of China owing to its adaptation to a wide range of temperatures and precipitation regimes (Xiong et al., 2007). Maize is also one of the most diverse crop species and used as a model plant in genetic studies (Liu et al., 2020). Breeding efforts have focused on detecting maize functional genes with an expectation to accelerate genetic improvement. These discovered functional genes and favorable alleles provide a firm basis for further yield improvement through marker-assisted selection or genetic transformation of crops (Gu et al., 2014). To select the desired traits for ideotypes, breeders cross a panel of breeding materials with the favorable alleles for multiple generations at multiple sites, which can be time-consuming and labor-intensive (Xiao et al., 2021). Most agronomic traits are genetically complex and substantial genotype by environment (G × E) interactions under long-term climate impedes the breeding process (Ravi et al., 2011).

Accurate phenotype prediction using genomic data is important for plant breeding and management because it facilitates the design and selection of new genotypes (G). Genomic prediction (GP) models using genome-wide markers simultaneously as predictor variables have been effective solutions to predict hybrid performance for plant breeding (Xiao et al., 2021). GP has the potential to accelerate the breeding process, reduce the cost of inbred line and hybrid development, maintain genetic diversity, and improve complex traits with low heritability (Heslot et al., 2012). Direct statistical GP models aim to construct a statistical model for predicting phenotype using whole-genome DNA markers jointly. Statistical methods reported for GP models can be classified into three main categories: parametric methods (e.g., GBLUP, RR-

BLUP, LASSO); semiparametric methods (e.g., RKHS); and machine learning methods (e.g., SVM, Random Forest). Both additive effects and nonadditive (dominance and epistasis) effects have been incorporated in those GP models (Crossa et al., 2017). Although GP models aim to predict phenotypes based on whole-genome markers, current progress for predicting phenotypes of plants using genetic information alone has reached a bottleneck. It is quite difficult to predict phenotypes of complex physiological and quantitative traits in diverse environments (E) because of the G × E interaction (Hu et al., 2019). A GP model has no explicit environmental inputs so it cannot be used to predict the performance of a new genotype in a new environment (Onogi et al., 2016).

Another approach for phenotype prediction is a systematic modelling approach known as process-based crop modelling. Crop simulation models have been commonly used to evaluate genetic improvement and support the breeding of target traits (Boote et al., 2021). These include assisting with multi-environment evaluation of advanced peanut breeding lines, assessing the adaptation of a new genotypes to a region, understanding the G × E interactions, identification and evaluation of desirable traits, and designing an ideotype for a target environment (Boote et al., 2001; Suriharn et al., 2007; Tardieu and Tuberosa 2010; Narh et al., 2015). Recognizing the potential for incorporating genetic information with process-based crop models, Technow et al. (2015) proposed an integrated statistical approach to combine GP models with crop modelling for predicting the grain yield of maize. Following the integrated approach, Onogi et al. (2016) directly linked an eco-physiological model for rice heading date with a GP model and simultaneously inferred the model input parameters and genome-wide marker effects on the parameters. Oliveira et al. (2021) integrated a dynamic statistical gene-based module into the CSM-CROPGRO-Drybean model to accurately predict the time of first flower appearance. The

predictions of these integrated models would be expected to be more accurate and robust than the direct statistical GP model. However, the integrated approach is statistically more challenging (Onogi et al., 2016).

A straightforward way (two-step approach) to use GP with crop modelling is to predict the model input parameters using GP methods. Based on estimated marker effects, phenotypes of untested breeding lines in untested environments can be simulated via the prediction of model parameter values. This 'marker-based modelling' approach can dissect complex traits into physiologically relevant component traits, integrate effects of significant QTLs on the component traits over time and space at the whole-plant level and predict a complex trait of various allele combinations under different climatic scenarios (Gu et al., 2014). This 'marker-based modelling' approach has been carried out in several crop species, for both simple morphological and complex physiological and quantitative traits, using either measured or optimized parameters, such as leaf elongation rate and kernel number in maize (Chenu et al., 2008; Amelong et al. 2015); flowering time in wheat (White et al., 2008); canopy cover dynamics and tuber bulking in potato (Khan et al., 2019a; 2019b); tomato fruit sugar concentration (Prudent et al., 2010); and photosynthesis and transpiration efficiency in rice leaves (Gu et al., 2012a, 2012b).

A major challenge for marker-based modelling is to accurately predict phenotypic differences for various genotypes grown in diverse environments, which can include (i) observed genotypes phenotyped in new environments; (ii) new genotypes in characterized environments; and (iii) new genotypes in new environments. Most previous studies linking crop modelling with genetic information were conducted on bi-parental mapping populations with one environment representing only a small part of the available genetic diversity and environmental variability. To

improve grain yield further, a deeper understanding of the physiologically relevant component traits contributing to grain yield, and the interaction between genotypes with the environments is important. The goal of this work were to 1) identify associated genetic markers for essential model input parameters of the CERES-Maize model; 2) estimate model parameters through conventional model optimization and marker-based statistical prediction; 3) demonstrate potential applications of the marker-based crop modelling as a breeding tool for studying the G × E interactions.

4.3 Materials and methods

4.3.1 Plant Material and Field Experiments

Plant material consisted of 282 maize lines developed from the Complete-diallel design plus Unbalanced Breeding-like Inter-Cross (CUBIC) population, as previously reported by Liu *et al.* (2020) and Luo *et al.* (2020). These maize lines were derived from the crossing of 1428 maternal lines with the Zheng58 and Jing724 testers, which have diverse genetic backgrounds covering six heterotic groups: Reid, Lancaster, waxy, tropic, P-population, and X-population germplasm (Xiao et al., 2021). Field trials were conducted in the year 2014 and 2015 at five sites distributed in the major maize producing areas in Northern China to collect field-measured phenotypic data including anthesis date, kernel number, kernel weight, and grain yield, and to detect significant markers associated with these traits (Liu et al., 2020). The five sites had various environmental growing conditions with different soil types, water treatments, and weather conditions (Table 4.1).

To derive crop model input parameters (Table 4.2), two of the five sites located in Xinxiang, Henan province (HN, 35.2° N, 113.8° E) and Baoding, Hebei province (HB, 38.7° N, 115.8° E) were selected in this study for CERES-Maize model calibration and evaluation. The

maize lines were planted with a completely randomized design. About 17 individual plants were planted in a row for each inbred line or F1 hybrid, as described by Xiao *et al.* (2021). Plots were planted in early June and harvested in middle October of each year at both sites. Irrigation and agronomic management practices followed best management practices for maize at local agricultural experimental stations. Plots were kept free of water stress, nutrient stress, weeds, insects, and diseases.

4.3.2 Weather, Soil and Phenotypic Data

The CERES-Maize model, which is distributed with DSSAT v4.7 (Hoogenboom et al., 2019), requires daily weather data including daily maximum/minimum air temperature, solar radiation, and precipitation. The required weather data for each site in year 2014 and 2015 were obtained from China meteorological Administration (http://data.cma.cn). The model also required soil profile data including lower limit (LL), drained upper limit (DUL), saturated water holding capacity (SAT), root growth factor (SRGF), bulk density (BD), soil organic carbon (Org. C), clay, silt, total nitrogen (Total N) and pH in different soil layers in the top 200 cm. These soil data for each soil type at each site were taken from the China Soil Scientific Database (http://www.soil.csdb.cn/) and from the global SoilGrids dataset with a spatial resolution of 1 km or 250 m (Hengl et al., 2014, 2017).

The phenotypic data collected in 2014 and 2015 include three representative agronomic traits: days to anthesis (DTA), measured as the interval from sowing to the day of pollen shed for half of the individuals; kernel number per ear (KNPE) and kernel weight per ear (KWPE), measured as the average kernel number and weight of five fully formed ears in the middle of each row to avoid edge effects within the plots. The kernel number, kernel weight and grain yield were measured at the harvest maturity stage.

4.3.3 Procedure for Estimating Crop Model Input Parameters Using Markers

To achieve the objectives of the marker-based crop modelling, we followed a pair-wise methodology (Figure 4.1). The individual steps are described in the following sections.

4.3.3.1 Conventional Model Parameters, Model Calibration and Evaluation

A set of conventional genotype-specific model parameters (Table 4.2) for each maize line was determined based on field-measured data collected in the two-year experiment at HN and HB. These model input parameters, covering various phenological characteristics, reproductive growth and leaf appearance of the maize, are: P1 (Degree days from emergence to end of juvenile phase), P2 (Photoperiod sensitivity coefficient), P5 (Degree days from silking to physiological maturity), G2 (Potential kernel number), G3 (Potential kernel growth rate), and PHINT (Phyllochron interval). The field-measured data in the year 2014 at two sites were used to calibrate the CERES-Maize model using the conventional trial-and-error method by finding the genotype-specific parameters that minimize error between simulated and observed phenotype traits. The calibrated model was then evaluated using independent datasets from the year 2015 at two sites. Statistical indicators including coefficient of determination (R^2), root mean square error (*RMSE*), and normalized root mean square error (*NRMSE*) were computed from observed (O_i) and simulated (S_i) variables to evaluate model performance. The comparisons between simulated and observed values of anthesis date, kernel number, kernel weight and grain yield for conventional crop model simulations can be found in Table 4.A1.

4.3.3.2 Identification of Significant Markers Associated with Model Parameters

Whole-genome resequencing and GWAS analysis of the CUBIC population including the 282 maize lines used in this study has been previously reported by Liu *et al.* (2020) and Xiao *et*

al. (2021). Two methods of GWAS analysis (sGWAS and hGWAS) were conducted and integrated with QTL mapping using both high-quality marker and phenotyping of the 23 agronomic traits. Based on results of multiple omics studies for the CUBIC population, many QTL regions were narrowed to a few candidates or to a single causal gene (Liu et al., 2020). These above genotypic data analysis procedures detected 839 genetic markers associated with anthesis date (P1 and PHINT), 322 genetic markers associated with kernel number (G2), 399 genetic markers associated with kernel weight (G3), which were further used for marker-based crop modelling in this study.

4.3.3.3 Statistical Models to Predict Marker-based Crop Model Parameters

First, all of the associated markers related with anthesis date (839 markers), kernel number (322 markers), and kernel weight (399 markers) were filtered for collinearity. The filtering procedure consisted of identifying colinear markers ($r \geq 0.9$) among associated genetic markers. After the filtering procedure, 137 (anthesis date), 100 (kernel number) and 146 (kernel weight) remaining genetic markers were used to train a partial-least-square regression (PLSR) model to compute marker-based values of four essential model parameters (P1, G2, G3, and PHINT) for each maize line (Table 4.3). The PLSR model was particularly adapted to the case of high-dimensional data to avoid the multicollinearity problem. It was also adapted to the case that the number of markers is much higher than the number of phenotypic observations (Abdi, 2003; Kadam et al., 2019; Bogard et al., 2020). The performance of the PLSR model was evaluated by prediction accuracy (i.e., the Pearson correlation coefficient between conventional and marker-based model parameters), coefficient of determination (R^2), and normalized root mean square error (*NRMSE*). The estimated marker-based crop model input parameters were then used in the

CERES-Maize model to simulate the marker-based values of anthesis date, kernel number, kernel weight and grain yield.

When training the PLSR statistical model for marker-based model parameters, the overall 283 genotypes were randomly divided into a training (235 observed genotypes; 83% of the maize population) and a testing (47 new genotypes; 17% of the maize population) set, which were considered as observed genotype (observed G) and new genotype (new G), respectively. By using the hierarchical sampling approach, the random separation of the maize population had a minimal effect on the population structure as the testing set and training set genotypes have similar structure (Heslot et al., 2012). The approximate 4:1 sample ratio considered the balance between prediction precision and phenotyping cost, which is commonly used in the seed industry to perform G-to-P assisted breeding (Xiao et al., 2021). The two experimental sites (HN and HB) in two years (2014/2015) were also divided as training environment (HN-2014/2015, Observed E) and testing environment (New E, HB-2014/2015). The training and testing division strategy separated the whole phenotyping datasets of 283 genotypes at two sites in HN and HB into four independent sub-datasets: observed G and E (235 observed genotypes, HN-2014/2015); observed G and new E (235 observed genotypes, HB-2014/2015); new G and observed E (47 new genotypes, HN-2014/2015); new G and E (47 new genotypes, HB-2014/2015). The performance of the marker-based model simulation was then evaluated based on the four sub-datasets.

4.3.4 Comparison of Strategies for Marker-based Crop Modelling

The number of essential model parameters that were linked to associated genetic markers to predict yield and its component traits ranged from two (P1 + PHINT) to four (P1 + PHINT + G2 + G3) in this study. Three different strategies were tested to estimate the relative importance of the essential model parameters to marker-based crop modelling. The first strategy was that only

model input parameters P1 and PHINT were predicted using associated genetic markers and the other model input parameters came from the conventional calibrated parameter values. The second strategy was that model input parameters P1, PHINT and G2 were predicted using associated genetic markers and the others were taken from the conventional calibration process. The third strategy was model input parameters P1, PHINT, G2 and G3 were predicted using associated genetic markers and the others were taken from the conventional calibration approach. The performance of the marker-based crop modelling in simulating yield and its component traits for the above four independent sub-datasets, with three different sets of marker-based model input parameters (P1 + PHINT, P1 + PHINT + G2, and P1 + PHINT + G2 + G3), were compared to determine the best strategy for marker-based crop modelling.

4.4 Results

4.4.1 Marker-based Estimation of Essential Model Parameters

The most immediately practical approach for employing genetic information is to estimate genotype-specific model parameters as a function of the genetic markers present at known loci. Based on the additive effects estimated by the partial-least-square regression (PLSR) analysis and allele information at each detected locus, marker-based values for each of the four essential model parameters (P1, G2, G3 and PHINT) were calculated for both training (n = 235) and testing (n = 47) sets of maize populations of 283 genotypes (Table 4.3).

In the training data sets (n = 235), there was a high correlation between conventional and marker-based values of the four model input parameters. The marker-based prediction accuracy ranged from 0.69 to 0.85 and the marker-based model parameters accounted for 48% - 72% of variation in conventional model parameters. The marker-based prediction error indicated by *NRMSE* ranged from 0.11 - 0.22. However, for the independent testing data sets (n = 47), the

marker-based prediction accuracy decreased especially for model parameters G3, which gave a value of 0.17. The marker-based model parameters only accounted for 0.03% - 21% of variation in conventional model parameters and the marker-based prediction error indicated by *NRMSE* ranged from 0.19 to 0.47. Overall, the model parameter G2 was best predicted and the G3 was worst predicted by associated genetic markers due to the quality of detected significant QTLs (Table 4.3).

4.4.2 Performance of Marker-based Modelling for Yield and Its Relevant Component Traits

The marker-based model capabilities to predict the yield and its relevant component traits (anthesis date, kernel number, kernel weight) were tested using four independent sub-datasets (observed G and E; observed G and new E; new G and observed E; new G and E) of 283 genotypes that were collected during 2014 and 2015 at HN and HB.

The marker-based model was good at simulating simple morphological traits. There was good agreement between observed and simulated values for time to flowering for all four sub-datasets (Figure 4.2). However, there was a consistent tendency for increased values of *RMSE* and *NRMSE* from 3.9 to 6.8 days and from 0.07 to 0.11 when testing four sub-datasets following the order of observed G and E; observed G and new E; new G and observed E; new G and E. This decrease in predictive ability was apparently caused by adding the new genotypes (G) and new environment (E) into the datasets. The marker-based model also simulated kernel number very well in two sub-datasets of observed and new G at observed E, with lower *RMSE* (469 and 586 number/m^2, respectively) and *NRMSE* (0.15 and 0.18, respectively) (Figure 4.3a, c). When changing to the other two sub-datasets of observed and new G at new E, both *RMSE* (721 and 798 number/m^2, respectively) and *NRMSE* (0.21 and 0.23, respectively) increased for the prediction error of marker-based values of kernel number (Figure 4.3b, d). Thus, the prediction

accuracy of marker-based kernel number was influenced more by new environments (E) than by new genotypes (G).

It was challenging to accurately simulated complex physiological and quantitative traits from genetic markers. The prediction accuracy of marker-based kernel weight was influenced more by new genotypes (G) than by new environments (E). The marker-based model accounted for 20% and 18% of variation in measured kernel weight using the two sub-datasets of observed G at both observed and new E (Fig. 4.4a, b). In contrast, it only accounted for 14% and 10% for new G grown at both observed and new E (Fig. 4.4c, d). The prediction error indicated by *RMSE* and *NRMSE* also shows the same tendency. The range of *RMSE* and *NRMSE* values increased from 56 – 61 to 61 – 74 mg/grain and from 0.19 – 0.22 to 0.21 – 0.27 with sub-datasets changing from observed genotypes (Fig. 4.4a, b) to new genotypes ((Fig. 4.4c, d). The marker-based crop modelling for grain yield also gave good simulations of observed genotypes (G) compared to new genotypes (G). The R^2 of the two sub-datasets of observed G with observed and new E were 0.26 and 0.14, respectively (Fig. 4.5a, b). However, the R^2 decreased to 0.10 and 0.11, respectively, for the observed G with observed and new E (Fig. 4.5b, c). The prediction performance of yield was sensitive to both new G and new E. The best marker-based modelling performance with the lowest values of *RMSE* (1845 kg/ha) and *NRMSE* (0.21) was achieved using the sub-datasets of observed G and E (Fig. 4.5a), compared to the *RMSE* (Range: 2172 – 2476 kg/ha) and *NRMSE* (Range: 0.22 – 0.29) of other three sub-datasets: observed G and new E; new G and observed E; new G and E (Fig. 4.5b, c, d).

4.4.3 Comparison of Different Strategies for Marker-based Modelling

Three different sets of essential model parameters were computed using associated genetic markers to determine the best strategy for marker-based crop modelling (Table 4.4). For all four

independent sub-datasets (observed G and E; observed G and new E; new G and observed E; new G and E), the marker-based simulation of anthesis date was not affected by the three different strategies for marker-based modelling since time to flowering was only related to the model parameters P1 and PHINT. Compared to the strategy that only two model input parameters (P1 + PHINT) were predicted using associated genetic markers, the three model parameters strategy with newly added marker-based model parameter G2 slightly reduced the error of the marker-based prediction of kernel number for all four sub-datasets with less values of *RMSE* or *NRMSE*. There was no further change of the model accuracy when adding the fourth marker-based model parameter G3.

As more marker-based parameters were added to the model, the accuracy of predicted kernel weight decreased, especially for the four model parameters strategy with the marker-based model parameter values for G3. The R^2 for all four independent sub-datasets consistently decreased and the model error indicated by *RMSE* and *NRMSE* generally increased as more model inputs were estimated by the marker-based technique. The performance of maker-based prediction of grain yield was improved for the two sub-datasets of observed G and new G at new E when more marker-based parameters were added to the model, indicated by lower values of *RMSE* and *NRMSE*. In contrast, the values of *RMSE* and *NRMSE* increased with more marker-based model parameters for the other two sub-datasets of observed G and new G at observed E. The proportion (R^2) of the explained variance of observed grain yield for marker-based modelling slightly decreased for the two sub-datasets of observed G at both observed and new E. However, a decrease in R^2 was found in the sub-datasets of new G at both observed and new E with increasing use of the marker-based model parameters.

4.5 Discussions

Conventional breeding for ideotypes for target environments is quite difficult because of the G×E interaction and the nature of the genetic complex for economic traits (Wang et al., 2018). Process-based crop models together with genome wide prediction are a powerful tool for marker selection and designing ideotypes for different crops and cultivation environments (Gu et al., 2014; Yin et al., 2018). In this paper, we proposed an integrated marker-based modelling approach by integrating essential model input parameters of the CERES-Maize model with their associated genetic markers. Critical findings are discussed below in detail.

4.5.1. Challenges in Predicting Model Parameters Using Limited Genetic Markers and Statistical Regression

In this study, the CUBIC population we used was created by crossing 1428 previously reported inbred lines, which was more genetically diversity than the previous publications with bi-parental mapping populations (Liu et al., 2021). Only detected QTL-associated markers through GWAS were used in the PLSR model to predict the input parameter values of crop models (Table 4.3). Although GWAS analyses has been successfully used to identify thousands of trait-associated genes for plant agronomic traits in the past decade, their ability to predict phenotypes remain limited because detected significant loci can only explain a small proportion of trait variation (Hu et al., 2019). It is also difficult to find the same QTLs across multiple environments or in different genetic backgrounds (Crossa et al., 2017). With limited numbers of associated genetic markers detected, the prediction accuracy of four essential model parameters (P1, G2, G3 and PHINT) based on PLSR (Table 4.3) was not good in independent testing data sets (R^2: 0.03 to 0.21). Since detected QTLs usually capture only a part of the genetic variance and QTL effect sizes tend to be overestimated, prediction of model parameters using the whole-genome markers is suggested for the further research (Onogi et al., 2016).

There are difficulties inherent in accurately linking the associated genetic markers with model parameters through appropriate statistical regression methods, such as the PLSR model we used (Table 4.3). The prediction of four essential model parameters (P1, G2, G3 and PHINT) was reasonable in the training data sets (R^2: 0.48 to 0.72) based on PLSR. However, for a molecular marker effect to be incorporated into a crop growth model, most statistical regression models (e.g., multiple linear regression) reported consider only simple additive effects (Gu et al., 2014; Kadam et al., 2019). It is suitable for simpler traits controlled by a small number of major genes. However, most economic traits are complex and affected by multiple genes, with both additive effects and nonadditive (dominance and epistasis) effects (Wang et al., 2018). An ideal statistical regression model used to link genetic markers with model parameters should consider incorporating these nonadditive marker effects in the gene-based crop modelling (Figure 4.6). Scientists have successfully used statistical methods, such as, simple linear regression and mixed linear regression models, for gene-based prediction of simpler traits (e.g., flower date, leaf area) (White et al., 2008; Zheng et al., 2013; Khan et al., 2019; Bogard et al., 2014; 2020). Enhancing the marker-based modelling capability for predicting complex quantitative traits (e.g., yield) may require machine and deep learning algorithms that can represent both additive and non-additive genetic effects (Figure 4.6).

4.5.2. Benefits of Marker-based Modelling in Dissecting Complex Traits for Studying G × E Interactions

Physiological dissection of target traits with marker-based modelling provides an avenue by which process-based crop models could be used to integrate molecular genetic technologies and crop improvement (Amelong et al, 2015). A complex trait like grain yield is an emergent property of interactions between simpler traits and environment × management, which can be

dissected into genetic controls related to simpler related component traits and then easily assess genetic variation for each component trait. An ideal marker-based modelling approach should be extensible to physiological traits that more directly affect grain yield (White et al., 2008). This was achieved in our study by dissecting maize yield into three yield physiologically relevant component traits (anthesis date, kernel number, and kernel weight) and four essential model parameters (P1, G2, G3, and PHINT) of CERES-Maize models (Table 4.2). These component traits and related model input parameters were a priori considered to be important for yield variation among the maize lines (Cooper et al., 2016).

We found that the prediction performance of marker-based modelling was largely affected by the types of traits predicted (Figure 4.6). The marker-based prediction performance of complex traits (kernel weight and yield) (Figure 4.4 and Figure 4.5) was generally reduced compared to simpler traits (anthesis date and kernel number) in this study (Figure 4.2 and Figure 4.3). For simple morphological traits, such as anthesis date and kernel number, marker-assistant selection comprises of selecting individuals with QTL-associated markers that have major effects; markers not significantly associated with a trait are not used. A higher prediction performance was observed for anthesis date and kernel number in four independent sub-datasets with both observed and new genotypes under both observed and new environments (Figure 4.2 and Figure 4.3). Unlike previous studies that only simulated limited traits that were phenotyped with limited mapping populations and environments, the results in our research were more powerful in accounting for G × E interactions for multiple traits (Chenu et al., 2008; White et al., 2008; Zheng et al., 2013; Amelong et al. 2015). For complex quantitative traits such as kernel weight and yield, improvement depends on the selection of biological interactions between multiple genes (Figure 4.5). Complex crop traits are also affected by age-dependent expression

of genes and gene effects. These underlying genetic complexities of complex traits can cause the prediction accuracy of marker-based kernel weight and yield (Figure 4.4 and Figure 4.5) to be more sensitive to new genotypes (G) than new environment (E).

For our study, the phenotyping environments at different sites in Northern China have spatial heterogeneity in soil types, weather condition, and management practices, which are part of the sources of possible prediction bias for marker-based modelling (Table 4.1 and Figure 4.6). Phenotyping traits observed in these experimental fields are often highly variable from site to site (e.g., HN and HB) and from growing season to growing season (e.g., 2014 and 2015). That can lead to prediction accuracy of marker-based kernel number (Figure 4.3) being more sensitive to new environments (E) rather than new genotypes (G). As a result, phenotyping environments of experimental sites are often grouped according to similarities in growing environment and germplasm types grown. In addition, most economic traits of interest to plant breeders are emergent outcomes of many interacting gene effects and physiological processes rather than being simpler traits directly driven by singular genetics. An emergent outcome from a specific genetic trait may benefit crop yield in one environment but have little or negative effect in another environment, which shows complex G × E interactions. This is reflected by the prediction performance of marker-based yield (Figure 4.5) being highly sensitive to both new environments (E) and new genotypes (G). Those situations imply that crop growth models in the future may need improvement in the simulation of detailed biological processes and dissection of complex traits into physiologically relevant component traits to better predict those emergent outcomes (Boote et al., 2021).

4.5.3. Contribution of Model Input Parameters to Marker-based Modelling

It is recognized that the model input parameters can represent certain genetic characteristics. With genetic-specific parameters, crop modeling has been considered a powerful tool to help breeders understand genetic by environment by management (G x E x M) interactions of crop production (Boote et al., 2021). In fact, those genotype-specific model parameters (P1, G2, G3, and PHINT) are artificial constructs that reproduce different growth cycles, photoperiod sensitivity, productivities and kernel number/growth rate traits without considering molecular genetics information (Table 4.2 and Table 4.4). Simple models are typically designed with less model parameters. In contrast, sophisticated crop models are designed with more model parameters. We found that using more model input parameters for marker-based modelling did not always improve marker-based modelling. When more marker-based model parameters (e.g., G2, G3) were added to the crop model, the prediction of kernel number improved but kernel weight was worse for all four sub-datasets with both observed and new genotypes under both observed and new growth environments (Table 4.4).

Although model parameters can provide higher prediction accuracy in crop models when they are independently estimated for each cultivar, these parameters may not accurately represent the genetic architecture of the associated crop phenotype or process (Hwang et al. 2017). The crop models do not use information on variation in genes among the genotypes. Instead, parameters of crop models are estimated primarily using phenotypic data obtained from field trials, and their genetic basis is largely unknown (Kromdijk et al., 2014; Oliveira et al., 2021). By connecting genetic markers with model parameters (P1, G2, G3, and PHINT), marker-based modelling showed promise in accounting for the genotype-phenotype relationships and G × E interactions. When more marker-based model parameters (e.g., G2, G3) were added to the crop models, we found that the prediction of yield in this study was improved for the two sub-datasets

of observed G and new G at new E but was worse for the other two sub-datasets of observed G and new G at observed E (Table 4.4). Increasing the number of parameters may have led to the complexity of the genetic effects by increasing the number of genes involved in the determinism of each parameter (Figure 4.6).

In addition to natural random errors from field-measured data, another potential source of error is the two-step approach we adopted that first fits genetic-specific parameters of crop models, and then re-estimate the new parameters associated with these QTL/genes. The systematic errors of crop models and PLSR model can be accumulated in the two-step approach (Zheng et al., 2013). Multiple sets of possible parameter values for traditional crop models can produce very similar responses during model optimization procedures (Boote et al., 2021). The specific designed crop models for gene-based modelling, such as GECROS, with input parameters that are directly related to the phenotyping traits breeders score for selection, can likely further facilitate the use of crop modelling integrated with genomic prediction in support of breeding (Gu et al., 2014).

4.6 Conclusions

Simulating genetic effects of associated markers on maize yield and its components traits can enhance the process of breeding new germplasms adapted to new environments. Compared with direct statistical genomic prediction, process-based crop models with explicit environmental inputs have excellent potential for predicting the behavior of a genotype in an environment that differs from that in which QTLs were detected. Four essential model input parameters of CERES-Maize model were integrated with associated genetic markers to design a marker-based crop modelling approach. The results showed that the marker-based model can accurately predict yield and its component traits (anthesis date, kernel number and kernel weight) with diverse

genotypes and environments by combining phenotypic and genotypic data. The prediction performance of complex quantitative traits for marker-based modelling was generally reduced compared to simpler morphological traits. To further improve the prediction accuracy of marker-based modelling, the special design of model parameters, the incorporation of non-additive genetic effects, the usage of statistical learning methods, and the consideration of phenotyping traits or environments are important. The marker-based modelling developed in this study can be a useful breeding tool in assisting geneticists and plant breeders for studying the genotype by environment interactions and selection of breeding lines to improve crop yield.

4.7 References

Abdi H. 2003. Partial least squares (PLS) regression. In: Lewis-Beck MS, Bryman A, Liao TF, eds. The SAGE encyclopedia of social sciences research methods. Thousand Oaks, CA: Sage Publishers, 792–795.

Amelong, A., Gambín, B.L., Severini, A.D. and Borrás, L., 2015. Predicting maize kernel number using QTL information. Field Crops Research, 172, pp.119-131.

Acharya S, Correll M, Jones JW, Boote KJ, Alderman PD, Hu Z, Vallejos CE. 2017. Reliability of genotype-specific parameter estimation for crop models: insights from a Markov chain Monte-Carlo estimation approach. Transactions of the ASABE 60:1699–1712.

Boote, K.J., Kropff, M.J. and Bindraban, P.S., 2001. Physiology and modelling of traits in crop plants: implications for genetic improvement. Agricultural Systems, 70(2-3), pp.395-420.

Bustos-Korts D, Malosetti M, Chapman S, van Eeuwijk F. 2016. Modelling of genotype by environment interaction and prediction of complex traits across multiple environments as a synthesis of crop growth modelling, genetics and statistics. In: Yin X, Struik PC, eds. Crop

systems biology: narrowing the gaps between crop modelling and genetics. Cham: Springer International Publishing, 55–82.

Bogard, M., Ravel, C., Paux, E., Bordes, J., Balfourier, F., Chapman, S.C., Le Gouis, J. and Allard, V., 2014. Predictions of heading date in bread wheat (*Triticum aestivum L.*) using QTL-based parameters of an ecophysiological model. Journal of experimental botany, 65(20), pp.5849-5865.

Bogard, M., Biddulph, B., Zheng, B., Hayden, M., Kuchel, H., Mullan, D., Allard, V., Gouis, J.L. and Chapman, S.C., 2020. Linking genetic maps and simulation to optimize breeding for wheat flowering time in current and future climates. Crop Science, 60(2), pp.678-699.

Boote, K.J., Jones, J.W. and Hoogenboom, G., 2021. Incorporating realistic trait physiology into crop growth models to support genetic improvement. in silico Plants, 3(1), p.diab002.

Chenu K, Chapman SC, Hammer GL, McLean G, Salah HBH, Tardieu F. 2008. Short-term responses of leaf growth rate to water deficit scale up to whole-plant and crop levels: an integrated modelling approach in maize. Plant, Cell and Environment 31: 378–391.

Cooper, M., Technow, F., Messina, C., Gho, C. and Totir, L.R., 2016. Use of crop growth models with whole‐genome prediction: application to a maize multi environment trial. Crop Science, 56(5), pp.2141-2156.

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., De Los Campos, G., Burgueño, J., González-Camacho, J.M., Pérez-Elizalde, S., Beyene, Y. and Dreisigacker, S., 2017. Genomic selection in plant breeding: methods, models, and perspectives. Trends in plant science, 22(11), pp.961-975.

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. and Hammer, G.L., 2018. Integrating modelling and phenotyping approaches to identify and screen complex traits: transpiration efficiency in cereals. Journal of experimental botany, 69(13), pp.3181-3194.

Gu J,YinX, StruikPC, Stomph TJ,WangH. 2012a.Using chromosome introgression lines to map quantitative trait loci for photosynthesis parameters in rice (*Oryza sativa L.*) leaves under drought and well watered field conditions. Journal of Experimental Botany 63: 455–469.

Gu J, Yin X, Stomph TJ, Wang H, Struik PC. 2012b. Physiological basis of genetic variation in leaf photosynthesis among rice (*Oryza sativa L.*) introgression lines under drought and well-watered conditions. Journal of Experimental Botany 63: 5137–5153.

Gu, J., Yin, X., Zhang, C., Wang, H. and Struik, P.C., 2014. Linking ecophysiological modelling with quantitative genetics to support marker-assisted crop design for improved yields of rice (*Oryza sativa*) under drought stress. Annals of Botany, 114(3), pp.499-511.

Hammer, G.L., van Oosterom, E., McLean, G., Chapman, S.C., Broad, I., Harland, P. and Muchow, R.C., 2010. Adapting APSIM to model the physiology and genetics of complex adaptive traits in field crops. Journal of experimental botany, 61(8), pp.2185-2202.

Heslot, N., Yang, H.P., Sorrells, M.E. and Jannink, J.L., 2012. Genomic selection in plant breeding: a comparison of models. Crop science, 52(1), pp.146-160.

Hengl, T., de Jesus, J.M., MacMillan, R.A., Batjes, N.H., Heuvelink, G.B., Ribeiro, E., Samuel-Rosa, A., Kempen, B., Leenaars, J.G., Walsh, M.G., 2014. SoilGrids1km-global soil information based on automated mapping. PLoS One. 9, e105992.

Hengl, T., Mendes de Jesus, J., Heuvelink, G.B., Ruiperez Gonzalez, M., Kilibarda, M., Blagotic, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G., Ribeiro, E., Wheeler, I., Mantel, S., Kempen, B., 2017. SoilGrids250m: Global gridded soil information based on machine learning. PLoS One. 12, e0169748.

Hwang C, Correll MJ, Gezan SA, Zhang L, Bhakta MS, Vallejos CE, Boote KJ, Clavijo-Michelangeli JA, Jones JW. 2017. Next generation crop models: a modular approach to model early vegetative and reproductive development of the common bean (*Phaseolus vulgaris L*). Agricultural Systems 155:225–239.

Hu, X., Xie, W., Wu, C. and Xu, S., 2019. A directed learning strategy integrating multiple omic data improves genomic prediction. Plant biotechnology journal, 17(10), pp.2011-2020.

Hoogenboom, G., C.H. Porter, V. Shelia, K.J. Boote, U. Singh, J.W. White, L.A. Hunt, R. Ogoshi, J.I. Lizaso, J. Koo, S. Asseng, A. Singels, L.P. Moreno, and J.W. Jones. 2019. Decision Support System for Agrotechnology Transfer (DSSAT) Version 4.7.5 (https://DSSAT.net). DSSAT Foundation, Gainesville, Florida, USA.

Kromdijk J, Bertin N, Heuvelink E, Molenaar J, de Visser PH, Marcelis LF, Struik PC. 2014. Crop management impacts the efficiency of quantitative trait loci (QTL) detection and use: case study of fruit load×QTL interactions. Journal of Experimental Botany 65, 11–22

Kadam, N.N., Jagadish, S.K., Struik, P.C., van der Linden, C.G. and Yin, X., 2019. Incorporating genome-wide association into eco-physiological simulation to identify markers for improving rice yields. Journal of experimental botany, 70(9), pp.2575-2586.

Khan, M.S., Struik, P.C., van der Putten, P.E., Jansen, H.J., van Eck, H.J., van Eeuwijk, F.A. and Yin, X., 2019. A model-based approach to analyse genetic variation in potato using standard cultivars and a segregating population. I. Canopy cover dynamics. Field Crops Research, 242, p.107581.

Khan, M.S., Yin, X., van der Putten, P.E., Jansen, H.J., van Eck, H.J., van Eeuwijk, F.A. and Struik, P.C., 2019. A model-based approach to analyse genetic variation in potato using standard cultivars and a segregating population. II. Tuber bulking and resource use efficiency. Field Crops Research, 242, p.107582.

Luo, J., Wei, C., Liu, H., Cheng, S., Xiao, Y., Wang, X., Yan, J. and Liu, J., 2020. MaizeCUBIC: a comprehensive variation database for a maize synthetic population. Database, 2020.

Liu, H.J., Wang, X., Xiao, Y., Luo, J., Qiao, F., Yang, W., Zhang, R., Meng, Y., Sun, J., Yan, S. and Peng, Y., 2020. CUBIC: an atlas of genetic architecture promises directed maize improvement. Genome biology, 21(1), pp.1-17.

Narh, S., Boote, K.J., Naab, J.B., Jones, J.W., Tillman, B.L., Abudulai, M., Sankara, P., M'Bi Bertin, Z., Burow, M.D., Brandenburg, R.L. and Jordan, D.L., 2015. Genetic Improvement of Peanut Cultivars for West Africa Evaluated with the CSM‐CROPGRO‐Peanut Model. Agronomy Journal, 107(6), pp.2213-2229.

Onogi, A., Watanabe, M., Mochizuki, T., Hayashi, T., Nakagawa, H., Hasegawa, T. and Iwata, H., 2016. Toward integration of genomic selection with crop modelling: the development of an integrated approach to predicting rice heading dates. Theoretical and Applied Genetics, 129(4), pp.805-817.

Oliveira, F.A., Jones, J.W., Pavan, W., Bhakta, M., Vallejos, C.E., Correll, M.J., Boote, K.J., Fernandes, J.M., Hölbig, C.A. and Hoogenboom, G., 2021. Incorporating a dynamic gene-based process module into a crop simulation model. in silico Plants, 3(1), p.diab011.

Ravi, K., Vadez, V., Isobe, S., Mir, R.R., Guo, Y., Nigam, S.N., Gowda, M.V.C., Radhakrishnan, T., Bertioli, D.J., Knapp, S.J. and Varshney, R.K., 2011. Identification of several small main-effect QTLs and a large number of epistatic QTLs for drought tolerance related traits in groundnut (*Arachis hypogaea L.*). Theoretical and Applied Genetics, 122(6), pp.1119-1132.

Suriharn, B., Patanothai, A., Pannangpetch, K., Jogloy, S. and Hoogenboom, G., 2007. Determination of cultivar coefficients of peanut lines for breeding applications of the CSM‐CROPGRO‐Peanut model. Crop science, 47(2), pp.607-619.

Tardieu, F. and Tuberosa, R., 2010. Dissection and modelling of abiotic stress tolerance in plants. Current opinion in plant biology, 13(2), pp.206-212.

Technow F, Messina CD, Totir LR, Cooper M (2015) Integrating crop growth models with whole genome prediction through approximate Bayesian computation. PLoS ONE 10:e0130855.

White, J.W. and Hoogenboom, G., 2003. Gene‐based approaches to crop simulation: Past experiences and future opportunities. Agronomy Journal, 95(1), pp.52-64.

White, J.W., Herndl, M., Hunt, L.A., Payne, T.S. and Hoogenboom, G., 2008. Simulation-based analysis of effects of Vrn and Ppd loci on flowering in wheat.

Wang, X., Xu, Y., Hu, Z. and Xu, C., 2018. Genomic selection methods for crop improvement: Current status and prospects. The Crop Journal, 6(4), pp.330-340.

Xiong, W., Matthews, R., Holman, I., Lin, E. and Xu, Y., 2007. Modelling China's potential maize production at regional scale under climate change. Climatic change, 85(3), pp.433-451.

Xiao, Y., Jiang, S., Cheng, Q., Wang, X., Yan, J., Zhang, R., Qiao, F., Ma, C., Luo, J., Li, W. and Liu, H., 2021. The genetic mechanism of heterosis utilization in maize improvement. Genome Biology, 22(1), pp.1-29.

Yin, X., Stam, P., Kropff, M.J. and Schapendonk, A.H., 2003. Crop modeling, QTL mapping, and their complementary role in plant breeding. Agronomy Journal, 95(1), pp.90-98.

Yin, X., van der Linden, C.G. and Struik, P.C., 2018. Bringing genetics and biochemistry to crop modelling, and vice versa. European Journal of Agronomy, 100, pp.132-140.

Zheng, B., Biddulph, B., Li, D., Kuchel, H. and Chapman, S., 2013. Quantification of the effects of VRN1 and Ppd-D1 to predict spring wheat (*Triticum aestivum*) heading time across diverse environments. Journal of experimental botany, 64(12), pp.3747-3761.

Table 4.1 Maize populations, soil types, water treatments, and seasonal weather conditions for the five phenotyping sites distributed in the major maize producing areas in Northern China.

Phenotyping sites	Latitude, longitude	Maize populations	Soil type	Water treatments	T_{max} (◦C)	T_{min} (◦C)	P (mm)	Years	References
Xinxiang, Henan province (HN)	35.2° N, 113.8° E	CUBIC[a]	Tidal	Irrigation	29.6	20.3	389	2014/2015	Liu et al., 2020; Luo et al., 2020, Xiao et al., 2021
Baoding, Hebei province (HB)	38.7° N, 115.8° E		Brown	Irrigation	29.7	18.9	348		
Changping, Beijing city (BJ)	40.2° N, 116.2° E		Brown	Irrigation	30.9	20.4	339		
Shengyang, Liaoning province (LN)	41.8° N, 123.4° E		Loam	Dryland	28.3	16.3	309		
Yushu, Jilin province (JL)	44.8° N, 126.6° E		Black	Dryland	25.8	15.3	391		

a. Complete-diallel design plus Unbalanced Breeding-like Inter-Cross.

Table 4.2 Details of genotype-specific parameters of the CERES-Maize model that classified into three categories.

Parameters	Description	Unit	Related traits
Phenological characteristics			
P1	Degree days (base 8 °C) from emergence to end of juvenile phase	°C days	Anthesis
P2	Photoperiod sensitivity coefficient (0-1.0)	-	Tassel initiation; Total Leaf number
P5	Degree days (base 8 °C) from silking to physiological maturity	°C days	Maturity
Reproductive growth			
G2	Potential kernel number	kernels/ear	Kernel number
G3	Potential kernel growth rate	mg/(kernel day)	Kernel weight
Leaf appearance			
PHINT	Degree days required for a leaf tip to emerge (phyllochron interval)	°C days per tip	Total leaf number; Anthesis date

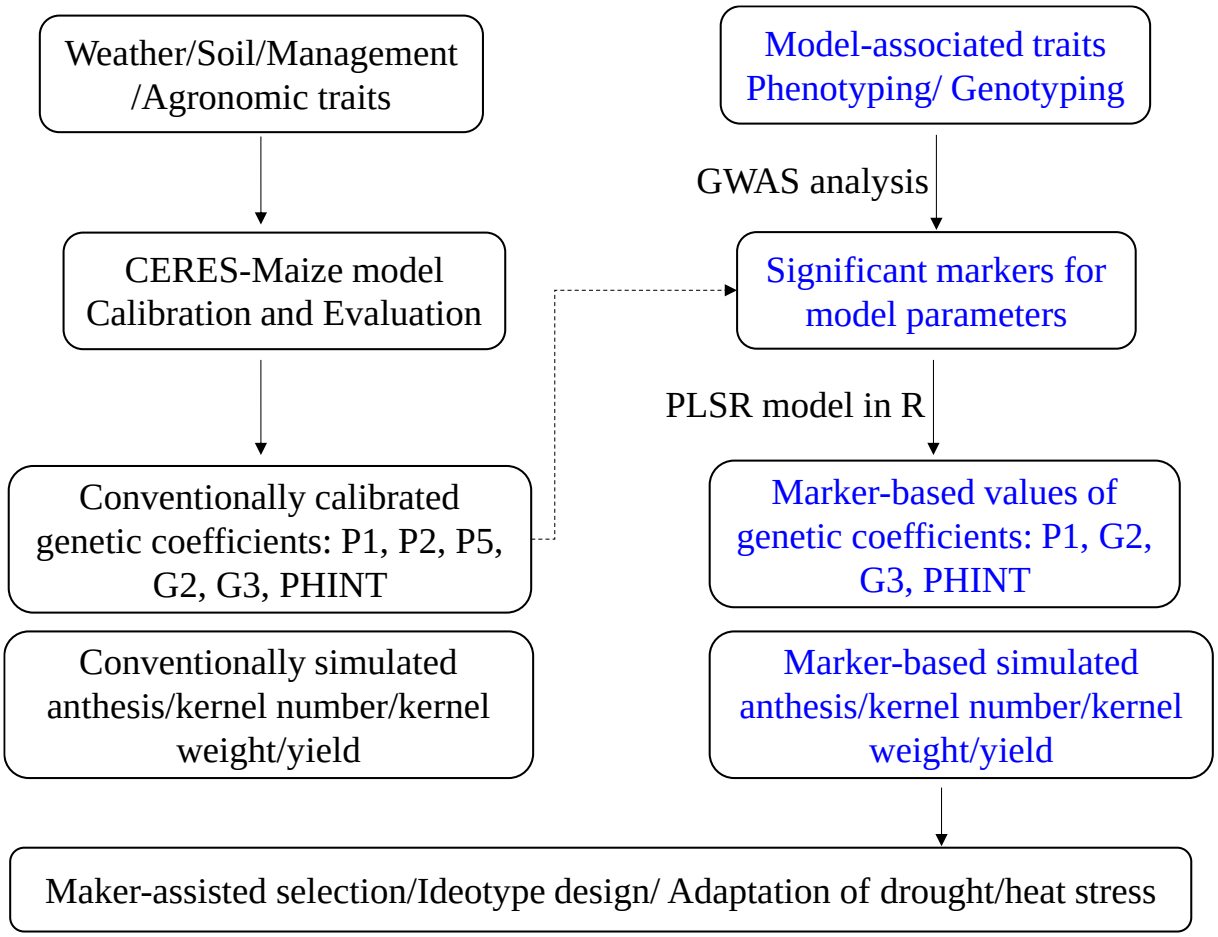


Figure 4.1 Diagram of the methodology used in this study, which combines a partial-least-square regression (PLSR) model and a crop model, CERES-Maize, into the marker-based crop modelling.

Table 4.3 Total number of filtered significant genetic markers used in partial-least-square regression (PLSR) for prediction of four essential CERES-Maize model input parameters of the maize training (n = 235) and testing (n = 47) genotypes in the 2014 and 2015 experiments at Henan (HN) and Hebei (HB) provinces.

Parameters	**Significant markers**	**Training (n = 235)**			**Testing (n = 47)**		
		Prediction accuracy	***R^2***	***NRMSE***	***Prediction accuracy***	***R^2***	***NRMSE***
P1	137	0.69	0.48	0.12	0.25	0.06	0.19
G2	100	0.85	0.72	0.11	0.46	0.21	0.19
G3	146	0.78	0.60	0.22	0.17	0.03	0.47
PHINT	137	0.71	0.51	0.16	0.27	0.07	0.29

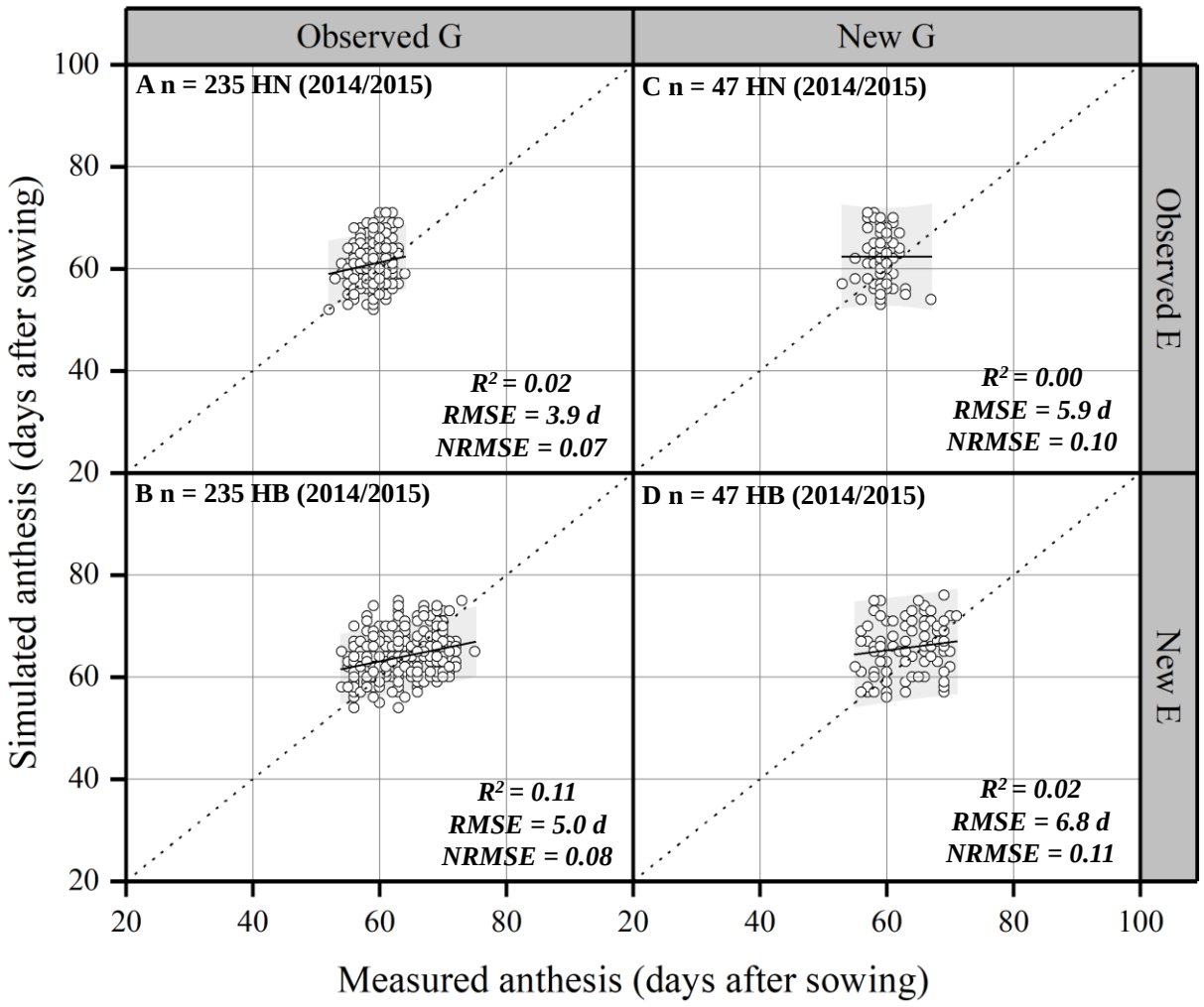


Figure 4.2 Relationship between marker-based simulated and measured anthesis dates in Observed (A, B) and new (C, D) genotypes (G) of maize under observed (A, C) and new (B, D) environments (E) in the 2014 and 2015 experiments at Henan (HN) and Hebei (HB) provinces.

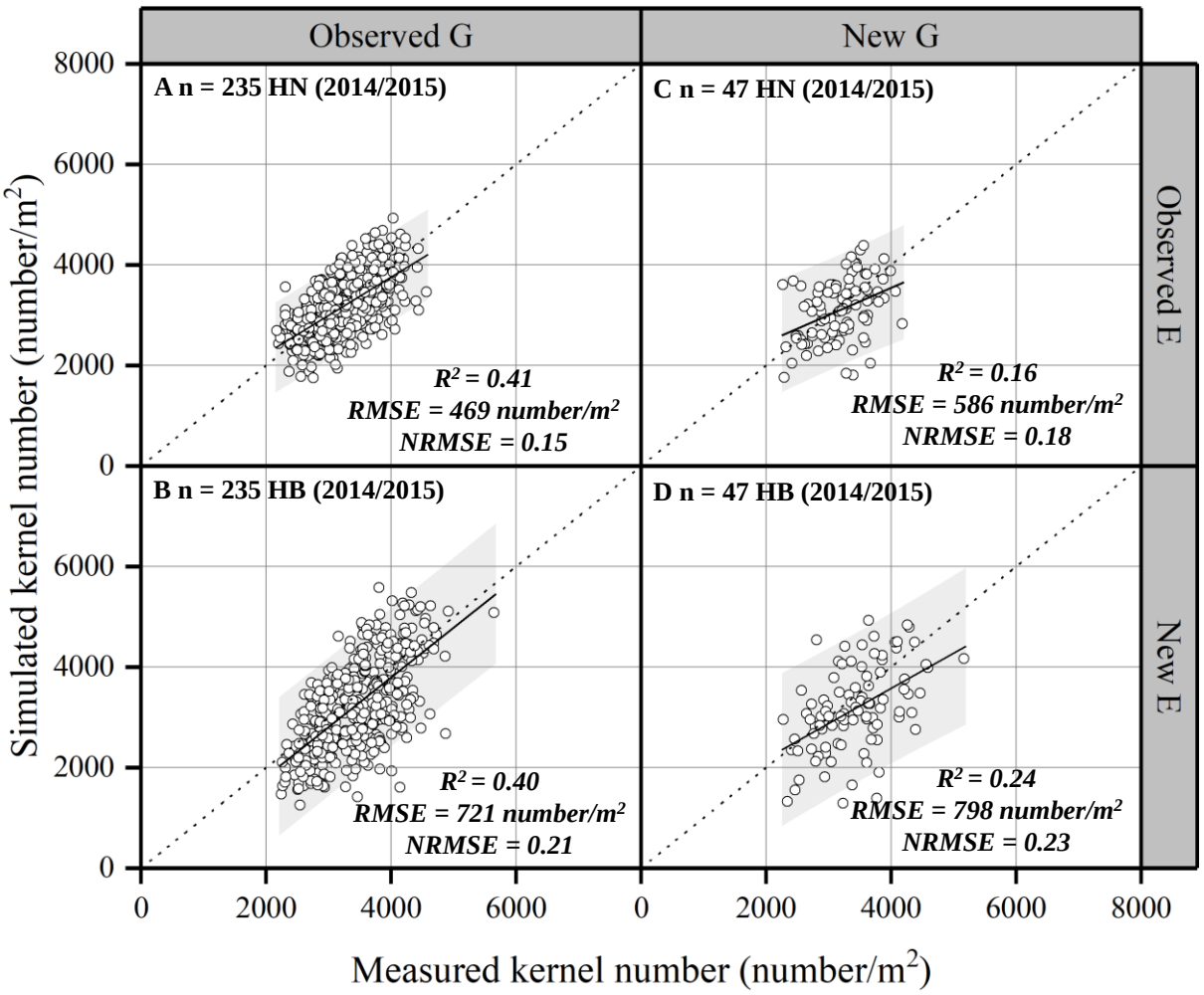


Figure 4.3 Relationship between marker-based simulated and measured kernel number in Observed (A, B) and new (C, D) genotypes (G) of maize under observed (A, C) and new (B, D) environments (E) in the 2014 and 2015 experiments at Henan (HN) and Hebei (HB) provinces.

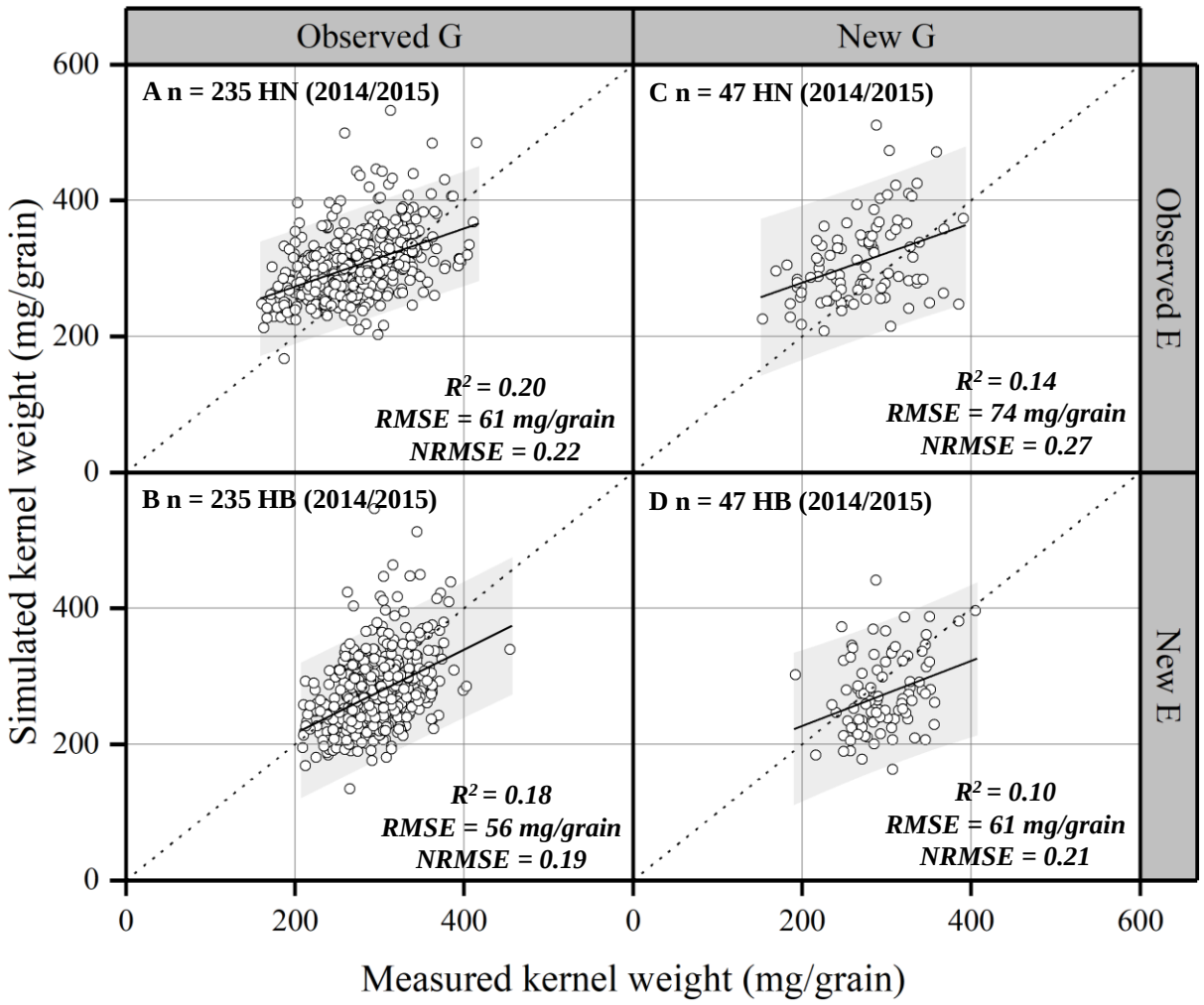


Figure 4.4 Relationship between marker-based simulated and measured kernel weight in Observed (A, B) and new (C, D) genotypes (G) of maize under observed (A, C) and new (B, D) environments (E) in the 2014 and 2015 experiments at Henan (HN) and Hebei (HB) provinces.

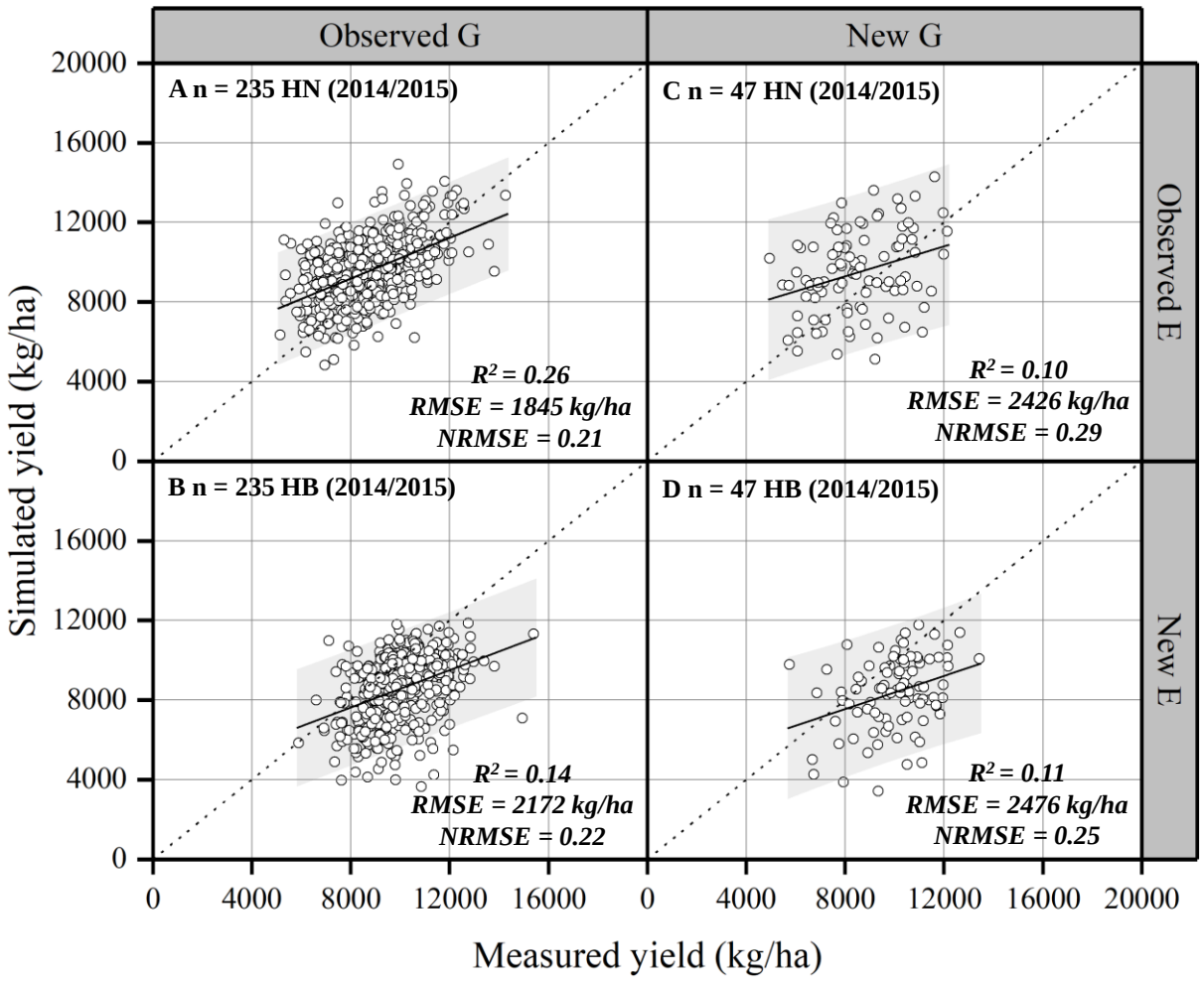


Figure 4.5 Relationship between marker-based simulated and measured yield in Observed (A, B) and new (C, D) genotypes (G) of maize under observed (A, C) and new (B, D) environments (E) in the 2014 and 2015 experiments at Henan (HN) and Hebei (HB) provinces.

Table 4.4 Comparison of three different strategies for marker-based crop modelling based on four independent sub-datasets. Two model input parameters (P1 + PHINT), three model input parameters (P1 + PHINT + G2), and four model input parameters (P1 + PHINT + G2 + G3) were predicted using associated genetic markers. The other model input parameters came from the conventional calibrated parameter values.

Strategy	**Observed G and E**			**Observed G and new E**			**New G and observed E**			**New G and E**		
	R^2	***RMSE***	***NRMSE***	***R^2***	***RMSE***	***NRMSE***	***R^2***	***RMSE***	***NRMSE***	***R^2***	***RMSE***	***NRMSE***
Anthesis dates (days after sowing)												
P1 + PHINT	0.02	3.9	0.07	0.11	5.0	0.08	0.00	5.9	0.10	0.02	6.8	0.11
P1 + PHINT + G2	0.02	3.9	0.07	0.11	5.0	0.08	0.00	5.9	0.10	0.02	6.8	0.11
P1 + PHINT + G2 + G3	0.02	3.9	0.07	0.11	5.0	0.08	0.00	5.9	0.10	0.02	6.8	0.11
Kernel number (number/m^2)												
P1 + PHINT	0.45	478	0.15	0.43	728	0.21	0.21	618	0.19	0.24	869	0.25
P1 + PHINT + G2	0.41	469	0.15	0.40	721	0.21	0.16	586	0.18	0.24	798	0.23
P1 + PHINT + G2 + G3	0.41	469	0.15	0.40	721	0.21	0.16	586	0.18	0.24	798	0.23
Kernel weight (mg/grain)												
P1 + PHINT	0.32	49	0.18	0.28	52	0.18	0.27	50	0.19	0.28	58	0.20
P1 + PHINT + G2	0.29	52	0.19	0.24	53	0.18	0.24	55	0.20	0.27	62	0.21
P1 + PHINT + G2 + G3	0.20	61	0.22	0.18	56	0.19	0.14	74	0.27	0.10	61	0.21
Yield (kg/ha)												
P1 + PHINT	0.30	1630	0.19	0.17	2305	0.23	0.27	1800	0.21	0.19	3130	0.31
P1 + PHINT + G2	0.30	1678	0.19	0.16	2230	0.22	0.17	1834	0.22	0.16	2759	0.28
P1 + PHINT + G2 + G3	0.26	1845	0.21	0.14	2172	0.22	0.10	2426	0.29	0.11	2476	0.23

Figure 4.6 Prediction bias of marker-based modelling caused by marker effects, phenotyping traits, phenotyping environments, model parameters, and statistical methods.

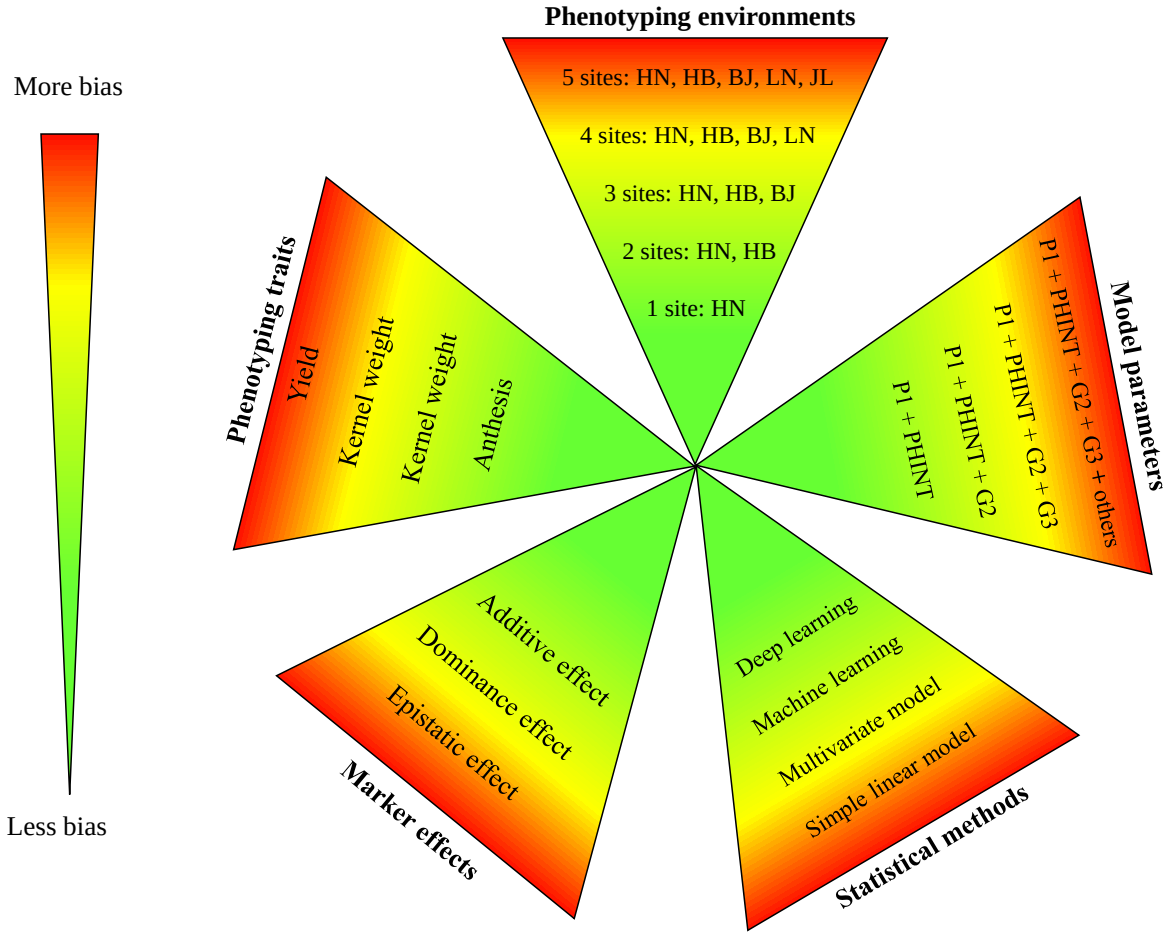


Table 4.A1 The comparisons between simulated and observed values of anthesis date, kernel number, kernel weight and grain yield for conventional crop model simulation in calibration year 2014 and evaluation year 2015.

Traits	Unit	Calibration (Year = 2014)			Evaluation (Year = 2015)		
		R^2	*RMSE*	*NRMSE*	R^2	*RMSE*	*NRMSE*
Anthesis date	days after sowing	0.21	3.7	0.06	0.64	2.9	0.04
Kernel number	number/m^2	0.66	392	0.11	0.40	548	0.17
Kernel weight	mg/grain	0.01	57	0.21	0.49	30	0.10
Grain yield	kg/ha	0.17	1647	0.18	0.30	1733	0.18