An analysis of simulated yield data for pepper shows how genotype × environment interaction in yield can be understood in terms of yield components and their QTLs

Crop Science

Rodrigues, Paulo C.; Heuvelink, Ep; Marcelis, Leo F.M.; Chapman, Scott C.; Eeuwijk, Fred A.

https://doi.org/10.1002/csc2.20476

Received: 2 May 2020 | Accepted: 29 January 2021

DOI: 10.1002/csc2.20476

Crop Science

ORIGINAL RESEARCH ARTICLE

Crop Breeding & Genetics

An analysis of simulated yield data for pepper shows how genotype × environment interaction in yield can be understood in terms of yield components and their QTLs

Paulo C. Rodrigues[1] | **Ep Heuvelink**[2] | **Leo F. M. Marcelis**[2] | **Scott C. Chapman**[3,4] | **Fred A. van Eeuwijk**[5]

[1] Dep. of Statistics, Federal Univ. of Bahia, Av. Adhemar de Barros, s/n-Campus de Ondina 40, Salvador, Bahia 170-110, Brazil

[2] Horticulture and Product Physiology, Wageningen Univ. & Research, PO Box 630, Wageningen 6700 AP, the Netherlands

[3] CSIRO Agriculture and Food, Queensland Bioscience Precinct, 306 Carmody Road, St. Lucia, QLD 4067, Australia

[4] School of Agriculture and Food Sciences, The Univ. of Queensland, Gatton Campus, Warrego Highway, Gatton, QLD 4343, Australia

[5] Biometris, Wageningen Univ. & Research, PO Box 16, Wageningen 6700 AC, the Netherlands

Correspondence
Paulo C. Rodrigues, Dep. of Statistics, Federal Univ. of Bahia, Av. Adhemar de Barros, s/n-Campus de Ondina 40. 170-110, Salvador, Bahia, Brazil.
Email: paulocanas@gmail.com

Assigned to Associate Editor Junping Chen

Abstract

Complex traits like yield are those in which phenotypic variation can be modeled as a linear function of a set of quantitative trait loci (QTLs) with environment dependency. This environment dependency can be observed at a phenotypic level as genotype × environment interaction (GEI) for yield itself and at an underlying genetic level as QTL × environment interaction (QEI). We show how GEI in yield may follow from pleiotropic QTLs for yield components that themselves are not environment dependent. We generated synthetic yield data via a crop growth model and analyzed these data by common statistical models for GEI and QEI. The QTLs for yield were pleiotropic with those for yield components. Such pleiotropy offers a path for improvement of yield under GEI. As a model system we used sweet pepper (*Capsicum annuum* L.) and developed an eco-physiological model for yield with seven genotype-specific inputs or yield components. Synthetic yields were simulated for a back cross population of 500 lines across a factorial combination of four major environmental drivers. The yield components were given a simple QTL basis and produced credible amounts and patterns of GEI for yield. The QEI for yield could be interpreted from the expression of QTLs for yield components and the interaction of these components with the environmental drivers. We see the generation of synthetic yield data via crop growth models followed by an analysis with statistical models for GEI and QEI to quantify the contribution of yield components to GEI as a helpful step in the development of yield prediction models for complex traits across environments that can also serve as a basis for decisions on selection strategies of complex traits.

1 | INTRODUCTION

Genotype × environment interaction (GEI) is the phenomenon observed when the relative performance of genotypes depends on the environment. For example, a genotype that is superior under well-watered conditions

Abbreviations: AMMI, additive main effects and multiplicative interaction; G-P, genotype-to-phenotype; GEI, genotype × environment interaction; GGE, genotypic main effects and genotype × environment interaction; GPAR, photosynthetic active radiation; QEI, quantitative trait locus × environment interaction; QTL, quantitative trait locus; VCOV, variance-covariance.

may yield poorly under dry conditions. A trait that shows strong GEI is difficult to predict, especially when considering predictions for new genotypes in new environments. Genotype × environment interaction is common for complex traits, where phenotypic variation depends on many genes, or quantitative trait loci (QTLs), with relatively small effects, that are also environment dependent. A common example of a complex trait is yield. Understanding of GEI can lead to better predictions of complex traits and is a fundamental requirement for genetic selection and improvement of such traits (Iwanska et al., 2020).

In a statistical genetic context, GEI can be addressed by regressing the GEI part of phenotypic responses on molecular marker variation to identify QTLs that show environment dependency, or QTL × environment interaction (QEI). The QEI can be further modeled in relation to environmental covariables, so that GEI can be predicted from markers linked to QTLs for the complex trait and environmental inputs. A well-known class of genotype-to-phenotype (G-P) models that can be subsumed under this approach are mixed linear and non-linear models with various types of QTL terms. Examples of this approach can be found in Bhakta et al. (2017), Boer et al. (2007), Malosetti et al. (2004, 2013), Millet et al. (2016, 2019), and van Eeuwijk et al. (2005, 2010, 2019).

A physiologically inspired alternative approach to GEI is based on crop growth simulation models. Classical crop models (reviewed by Marcelis et al., 1998; van Ittersum et al., 2003) have been developed to predict crop yield in trials with varying environmental conditions, but these models are usually calibrated for a single genotype. Genetic control of physiological input parameters can be incorporated into crop models with GEI arising as an emerging property. In that case, crop growth models represent a class of G-P models based on prior biological knowledge that has proved to be useful for understanding GEI and QEI (Baldazzi et al., 2016; Bertin et al., 2010; Bustos-Korts et al., 2016; Chapman et al., 2002; Chenu et al., 2009; Cooper et al., 2005; Hammer et al., 2005; Letort et al., 2008; Malosetti et al., 2016; van Eeuwijk et al., 2005, 2010). A particularly strong point of crop growth models in comparison with more statistically oriented G-P models is that they contain explicit representations of development over time and especially this feature may be useful in describing GEI (Bustos-Korts, Boer, et al., 2019; Chenu et al., 2009). A wide spectrum of physiological models with genetic control of genotype specific input parameters is available for better interpretation of GEI and QEI, for traits of varying complexity like yield (Chenu et al., 2008; Tardieu, 2003; Yin et al., 2000, 2004), leaf elongation (Chenu et al., 2008; Reymond et al., 2003, 2004), concentrations of chemical compounds in seed grains (Ishii et al., 2010), flowering time (Zheng et al., 2013), and fruit quality (Quilot et al., 2005).

Most papers that aim at combining crop growth modeling approaches with quantitative genetic approaches give little attention to an integrated statistical analysis and understanding of the patterns of GEI and QEI that occur across environments. In the current paper, a major objective is to investigate GEI and QEI for a complex trait in relation to its known genetic and physiological basis, as generated from a relatively simple crop growth model in which key physiological parameters follow from the assignment of underlying explicit QTL effects. An important question for breeders is whether QTLs for a complex trait will map to genomic locations where QTLs for component traits are known to be present and whether the QTLs for the complex trait will show environment dependency (QEI), even where the component traits are known to possess no environment dependency at all. Some relevant questions include: when we map the complex trait yield, do all QTLs for yield coincide with those of underlying yield components? Is the set of QTLs for yield a subset of the QTLs for yield components, where we know that the yield components are organized according to the structure and dynamics of a physiological model? Are there additional QTLs for yield that are mapped away from QTLs for yield components and, if yes, how do we interpret such QTLs? How much QEI do yield components show? To consider these issues, we simulated yield as a complex trait in pepper (*Capsicum annuum* L.) making use of information from experiments on a recombinant inbred line (RIL) population for pepper (Alimi et al., 2013).

Insight into the complex controls of yield will determine to which extent a dissection of yield in yield components offers a viable breeding strategy for the improvement of yield in the face of strong GEI. In an ideal scenario, a plant breeder could measure or predict a limited set of yield components with little or no QEI and insert those measurements or predictions into a crop growth model together with environmental inputs to predict yield for whichever condition. The nonlinear interactions between the yield components over developmental time feed into emerging properties of the crop growth model that in the end cause GEI in yield. When the yield component can be predicted from a selected set of markers (QTLs) or a full set of markers, like in genomic prediction (de los Campos et al., 2013; Heslot et al., 2015), yield performance for combinations of new genotypes in new environmental conditions could be predicted early on in the breeding cycle from marker profiles together with environmental inputs (Bustos-Korts et al., 2016, 2018; Malosetti et al., 2016).

Currently, various groups develop methods that try to predict yield from marker inputs together with environmental inputs, where the emphasis is on whole genome prediction methods, in contrast with the earlier developed QTL-based methods. These attempts range from statistical approaches using mixed models (Jarquín et al., 2014, 2017) and machine learning and deep learning (A. Montesinos-López et al., 2018; O. A. Montesinos-López et al., 2018) to integrations of crop growth models with mixed model and Bayesian statistical

genetic approaches (Cooper et al., 2016; Messina et al., 2018; Technow et al., 2015). We emphasize that we do not intend here to develop predictive models for GEI along the lines of those authors. This study presents a preliminary step to the building of predictive models and demonstrates how genetically simple yield components without environment sensitivity produce the genetic complexity of realised yield with strong GEI. An improved understanding of yield QTLs via an identification of the corresponding yield components and their QTLs should smoothen the path to better predictive models for yield.

We developed a simple crop growth model containing a small set of component traits. By generating the component traits from a QTL basis, we achieve an integration of crop growth models and statistical genetic models in the spirit of Yin et al. (2000, 2004); for a recent example, see Wallach et al. (2018). We report below on the results of a simulation study in pepper focusing on the questions of (a) whether credible patterns of variation in GEI and QEI for yield could be generated using a simple crop growth model and (b) whether the main effect QTLs, without QEI, used to generate the component traits, the physiological parameters, could be identified in a QTL analysis for yield and whether these yield QTLs showed QEI.

The structure of the paper is as follows. We first describe and justify the structure of our pepper crop growth model, which is a genotype-specific extension of a more general species-specific crop growth model. The complex trait, yield, is produced from a small set of genotype-specific and environment-independent physiological component traits. Values for the component traits were based on prior experiments and literature. Environmental inputs were obtained from actual environmental characterizations in earlier growing seasons. Breeding populations (back-crosses) were simulated in which the variation in the components traits was assigned a genetic basis in terms of one or more underlying QTLs and some residual genetic variation. The simulation framework is thus defined by (a) the structure of the crop growth model and (b) its inputs, the genotype-specific component traits generated from underlying QTLs, and the environmental inputs.

After the description of this framework, we briefly describe published statistical techniques that will be used to analyze the simulated data for the patterns in GEI and QEI. These statistical analyses of the simulated data can be considered as special cases of sensitivity analysis. Finally, we address interpretation of QTL analyses for component traits and the resulting complex trait yield. Pleiotropic QTLs for component traits and complex trait, where the former do not show QEI and the latter does, may allow the identification of beneficial marker profiles for the complex trait.

2 | MATERIALS AND METHODS

2.1 | Description of the genotype-to-phenotype model

The eco-physiological G-P model (Figure 1) is based on the LINTUL crop model (Ezui et al., 2018; van Ittersum et al., 2003), limited to conditions where the only environmental effects are via solar radiation, temperature, and ambient CO_2 concentration [CO_2] (i.e., no stresses due to nutrition, water supply, or biotic effects). This simple model can be articulated in a series of equations with environment inputs and "genotype-specific" parameters (Table 1). The traits crop duration and fruiting period, leaf area, net carbon assimilation rate, and partitioning to fruit are influenced by temperature, whereas daily assimilation is further affected by radiation and [CO_2].

Cumulative dry matter production ($TDM_{i,j}$ for genotype i in environment j; g m^{-2}) is the product of cumulative intercepted light and light use efficiency ($LUE_{i,j;}$ g mol^{-1}):

$$TDM_{i,j} = \sum_{t=t_{0,j}}^{t_{f,j}} \left\{ \left[1 - \exp\left(-K_i \times LAI_{i,j,t} \right) \right] I_{j,t} \right\} \times LUE_{i,j}, \quad (1)$$

where $t_{0,j}$ and $t_{f,j}$ represent the first and last day of the growing season in environment j, K_i is the light extinction coefficient for genotype i, $LAI_{i,j,t}$ represents the leaf area index (m^2 leaf area m^{-2} ground area) for genotype i, in environment j on day t, and $I_{j,t}$ represents the photosynthetic active radiation (PAR; mol m^{-2} d^{-1}) on top of the crop in environment j and on day t.

Light use efficiency (crop growth rate per unit of intercepted PAR) is assumed to increase with CO_2 and with temperature according to a saturating response:

$$LUE_{i,j} = LUE_i^{max} \times \left\{ 1 - \exp\left(c \left[CO_2 \right]_j \right) \right\} \times \left(1 - \exp\left\{ -Z_i \left[I\left(T_j > T_{LUE,j} \right) \left(T_j - T_{LUE,j} \right) \right] \right\} \right) \quad (2)$$

where LUE_i^{max} is the light use efficiency of genotype i, when both CO_2 concentration and temperature are not limiting $LUE_{i,j,}$ c, Z_i, and $T_{LUE,j}$ are scaling constants, $I(T_j > T_{LUE,j})$ an indicator variable, T_j represents the 24-h average temperature, $I(T_j > T_{LUE,j})$ is an indicator variable taking the value 1 when the condition is met, and $[CO_2]_j$ represents the average CO_2 concentration during the day (light period). Further details about the constants used in this model are available in Table 2. In our simulations, CO_2 concentration and temperature are constant over the whole growing season (see below).

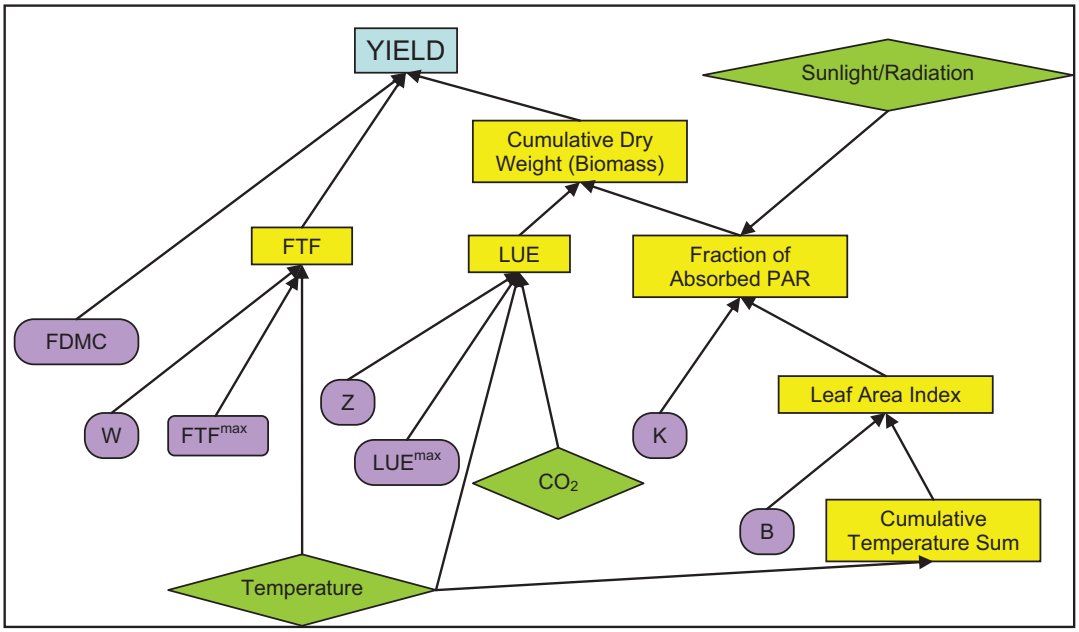

FIGURE 1 Schematic diagram of the crop growth model with seven genotype specific parameters. The diamonds represent input data, rectangles are states, ellipses are parameters and lines represent transfer of matter or information. The seven genotype specific parameters are (a) maximum light use efficiency (LUE^{max}); (b) light extinction coefficient (k); (c) slope for the leaf area increase with temperature sum (B); (d) maximum fraction of dry weight partitioned into the fruits (FTF^{max}, harvest index); (e) slope of the linear reduction in harvest index with temperature above 15 °C (W); (f) fruit dry matter content (FDMC); and (g) slope of the linear reduction in LUE for temperatures below 20 °C (Z). PAR, photosynthetic active radiation

TABLE 1 The seven genotype-specific, environment-independent physiological parameters in the yield model (see caption Figure 1 for full names), parameterized for greenhouse pepper. For each parameter, the mean value and the standard deviation (SD) are given. The last column presents the references for the chosen values (index i refers to genotype i)

Parameter	Mean	SD	References
LUE_i^{max} [a]	0.87	0.174	(Nederhoff, 1994; Heuvelink, 1995)
W_i [b]	0.04	0.011	(Wubs et al., 2009; Wubs et al., 2011)
FTF_i^{max}	0.65	0.04	(Rijsdijk & Houter, 1993; Gelder et al., 2007)
$FDMC_i$	0.0774	0.00508	(Wubs et al., 2009)
Z_i [c]	0.6	0.05	(de Swart et al., 2006)
K_i	0.7	0.04	(Marcelis, Heuvelink & Goudriaan, 1998)
B_i	0.000378	3.78×10^{-5}	(Marcelis et al., 2006)

Note. LUE_i^{max}, light use efficiency of genotype i, when both CO_2 concentration and temperature are not limiting $LUE_{i,j,}$; W_i, time-dependent scaling constant for genotype i; FTF_i^{max}, maximum level of fraction to fruits for genotype i; $FDMC_i$, fruit dry matter content for genotype i; Z_i, time dependent scaling constant for genotype i; K_i, light extinction coefficient for genotype i; B_i, genotype specific slope for the regression of leaf area per shoot (m^2) on temperature sum for genotype i.

[a]Mean value of LUE_i^{max} and c (Table 2) are chosen such that LUE at a CO_2 concentration of 370 μmol mol^{-1} is 0.65 g dry matter (DM) mol^{-1} photosynthetic active radiation (PAR) (Heuvelink, 1995) and the relative increase in LUE when CO_2 concentration rises to 1,000 μmol mol^{-1} agrees with Nederhoff (1994).

[b]Mean value of W_i and T_{FTF} (Table 2) are chosen such that the linear reduction in fraction partitioning to the fruits for temperatures above 15 °C agrees with Wubs et al. (2009; 2011).

[c]Mean value of Z_i and T_{LUE} (Table 2) are chosen such that LUE is not much different between 20 and 25 °C, but is reduced at 15 °C in agreement with De Swart et al. (2006).

Different LUE values for cultivars grown under the same environment have been reported, for example, by Quero et al. (2018), Higashide and Heuvelink (2009) and Li et al. (2014).

The leaf area index ($LAI_{i,j,t}$) is the product of leaf area per shoot and shoot density and is assumed to increase linearly with temperature sum (Marcelis et al., 2006). With a being a genotype-independent intercept and B_i being a genotype-specific slope for the regression of leaf area per shoot (m^2) on temperature sum (°C d), $(T_j - T_{base})(t - t_0)$, the LAI for genotype i in the environment j at day t, $LAI_{i,j,t}$, can be calculated

TABLE 2 Parameterization of the constants in the model for sweet pepper. For each constant, the equation number, the chosen values and reference/section with further explanations are given

Constant	Equation	Value(s)	Reference or section
$t_{0,j}$	(1)	10 Jan. (NL); 10 Sept. (SP)	Section 2.2
$t_{f,j}$	(1)	30 Nov. (NL); 30 Apr. (SP)	Section 2.2
c	(2)	−0.004	(Nederhoff, 1994; Heuvelink, 1995)
$CO_{2,j}$	(2)	370, 1,000 (μmol mol^{-1})	Section 2.2
$T_{LUE,j}$	(2)	13 (°C)	(de Swart et al., 2006)
a	(3)	0.03372	(Marcelis et al., 2006)
T_{base}	(3)	10 (°C)	(Marcelis et al. 2006)
Sd	(3)	7 (m^{-2})	Common practice in the Netherlands
T_j	(3)	15, 20, 25 (°C)	Section 2.2
$RAD_{j,t}$	(4)	Numerical variable	Historical data
F_{PAR}	(4)	0.5	(Goudriaan & Laar, 1994)
Tr_j	(4)	0.75 (NL); 0.60 (SP)	Section 2.2
T_{FTF}	(6)	15 (°C)	(Wubs et al., 2009; Wubs et al., 2011)

Note. NL, the Netherlands; SP, Spain; $t_{0,j}$, first of the growing season in environment j; $t_{f,j}$, last day of the growing season in environment j; c, scaling constant; $CO_{2,j}$, average carbon dioxide concentration during the day for environment j; $T_{LUE,j}$, scaling constant for environment j; a, genotype independent intercept; T_{base}, base temperature; Sd, shoot density; Tj, the 24-h average temperature for environment j; $RAD_{j,t}$, global radiation at day t in environment j; F_{PAR}, fraction of photosynthetically active radiation (PAR) in global radiation; Tr_j greenhouse transmissivity in environment j; T_{FTF}, scaling constant.

as follows:

$$LAI_{i,j,t} = \left\{a + B_i \left[I(T_j > T_{base}) \times \left(T_j - T_{base}\right)\right]\left(t - t_0\right)\right\} \text{ Sd} \quad (3)$$

where T_{base} is the base temperature and $I(T_j > T_{base})$ is an indicator variable, t represents the tth day of the growing season ($t = t_0$ is the day of the first flowering), and Sd is the shoot density.

The PAR incident on the crop on day t in environment j, $I_{j,t}$, is the product of (a) global radiation at day t in environment j, $RAD_{j,t}$, (b) a fraction of PAR in global radiation (F_{PAR}), and (c) greenhouse transmissivity in environment j (Tr_j):

$$I_{j,t} = RAD_{j,t} \times F_{PAR} \times Tr_j. \quad (4)$$

Fresh yield is calculated from cumulative dry matter production by multiplying the latter with a partitioning index ($FTF_{i,j}$) and dividing by fruit dry matter content ($FDMC_i$):

$$Yield_{i,j} = TDM_{i,j} \times FTF_{i,j} \times \frac{1}{FDMC_i} \quad (5)$$

where partitioning index $FTF_{i,j}$ decreases linearly with temperature from a genotype-specific maximum level, FTF_i^{Max}, that is not limited by high temperatures (since high temperature stimulates abortion of flowers and fruit, hence a reduced partitioning to the fruits; Wubs et al., 2009):

$$FTF_{i,j} = FTF_i^{Max}\left\{1 - W_i\left[I\left(T_j > T_{FTF}\right) \times \left(T_j - T_{FTF}\right)\right]\right\}, \quad (6)$$

in which W_i and T_{FTF} are scaling constants, whereas $I(T_j > T_{FTF})$ is an indicator variable.

The model was programmed in the R language (R Core Team, 2019).

2.2 | Parameterization of the model

A breeding population of greenhouse pepper genotypes (described below) was simulated by assigning values to the seven genotype-specific, and environment-independent, physiological component traits of the G-P model above. For each trait, we assumed a Gaussian distribution with mean values based on a priori knowledge, as specified in Table 1. There were no physiological reasons to believe that the component traits would be moderately or strongly dependent, and therefore they were assumed to be independent.

2.3 | Environments

The genetic and environmental configurations for our data simulations were inspired by real experiments on yield in pepper as described in Alimi et al. (2013). Thirty six environments were defined, a three-by-two-by-two-by-three full factorial combination of four environmental factors: (a) three levels of daily radiation (low, average, and high radiation) based on historical annual weather data (1994, 2000, and 2008 for Spain; 1998, 2003, and 2007 for the Netherlands), (b) two countries (Spain and The Netherlands), (c) two levels of CO_2 concentration (370 μmol mol^{-1} for the open

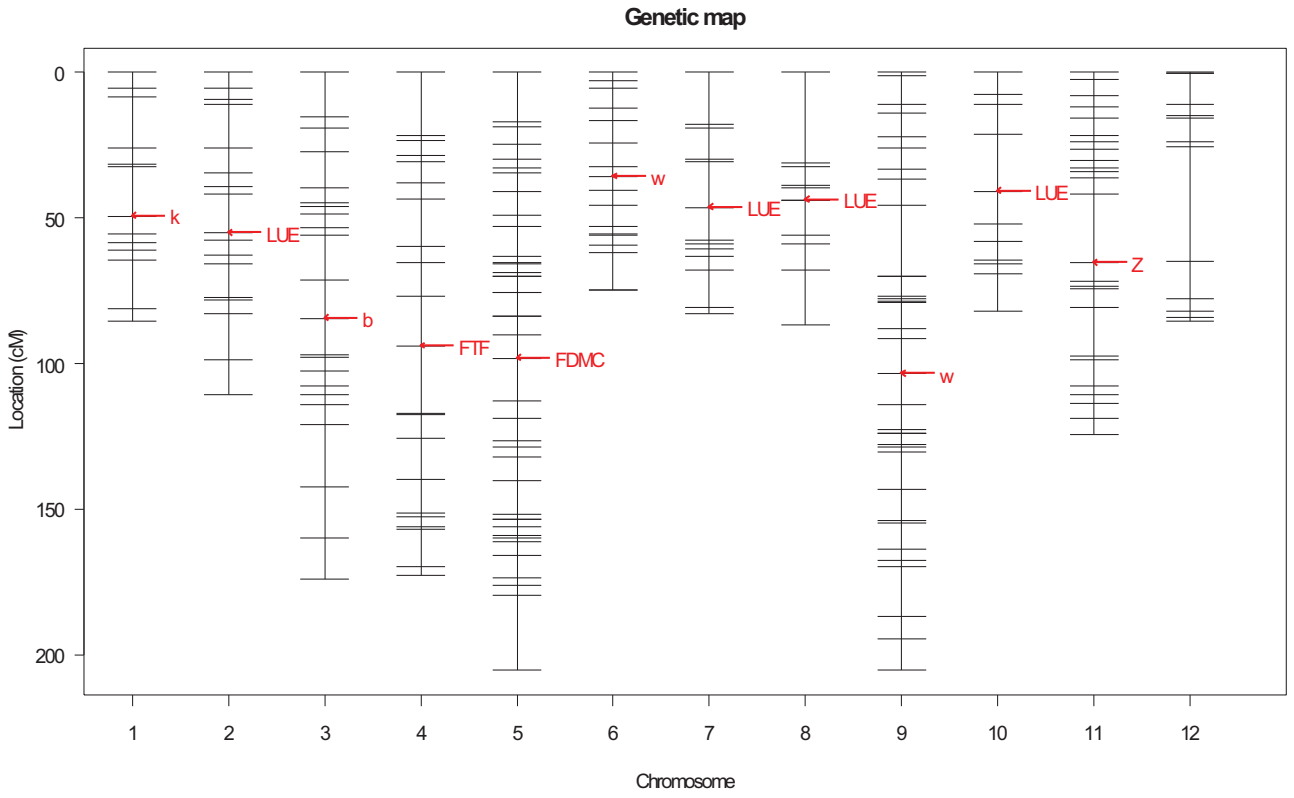

FIGURE 2 Genetic map for pepper, based on the lengths of chromosome and number of markers per chromosome in Barchi et al. (2007; 2009). The marker positions were taken randomly. The arrows with the name of the seven genotype-specific crop model parameters indicate the place where the quantitative trait loci (QTLs) were placed. See Figure 1 for definitions of parameters

environment, and 1,000 μmol mol^{-1} for the closed greenhouse with CO_2 enrichment), and (d) three levels of daily average temperature (15, 20, and 25 ˚C). The growing season in Spain was considered to start on 10 September and end on 30 April (232 d) and in the Netherlands this was from 10 January to 30 November (324 d). The greenhouse transmissivity (Tr) was considered to be 0.75 for the Netherlands (high tech glasshouses) and 0.60 for Spain ("paral" plastic greenhouses).

2.4 | Simulation of the population

To study the generation of GEI for a complex trait by simulating yields for a set of genotypes belonging to a segregating breeding population, we chose a backcross with the phenotypes predicted by the LINTUL-based model described above for each genotype in each environment. Hence, the simulation model combines genotype-specific physiological parameters (no GEI), with environment-specific inputs, to generate yield phenotypes that should show GEI. Yield was simulated for a population of 500 backcross lines (each line is a new independent draw of seven genotypic parameter values), for each of the 36 environments, resulting in a two-way 500 genotypes × 36 environments data table. A population size of 500 lines can be expected to produce clear test profiles for the QTLs.

Chromosome lengths and numbers of markers were based on the pepper population described by Barchi et al. (2007). A single map was generated with marker positions drawn from a continuous uniform distribution defined over the full length of the corresponding chromosomes, and ensuring markers were present at both ends of each chromosome (Figure 2). Marker positions and alleles were generated by the function sim.map in package qtl (Broman & Sen, 2009) of R software, and for convenience, QTLs for the seven genotype-specific physiological input parameters were assigned one at a time to 11 out of 12 chromosomes, and they were positioned at the marker closest to the middle of the chromosome (Table 3).

Yield given by Equation 5 depends on seven physiological parameters, each of which was considered to depend on a given number of QTLs (Figure 2, Table 3). Since LUE^{max} and W were found to have stronger impact on the final phenotypic data [higher proportion of variance explained in sensitivity analysis, see Table 4 below, and higher $-\log_{10}(P)$ values in a preliminary QTL analysis], we decided to make these parameters dependent on more than one QTL. We now express the physiological parameters of Equation 5 in terms

TABLE 3 Genetic architecture in the studied model. For each genotype specific parameter, the number of quantitative trait loci (QTLs) responsible for its genetic variation, the location of the QTLs and their heritabilities are presented. All the 11 QTLs were placed next to the closest marker to the middle of the given chromosome

Parameter	No. of QTLs	Location of the QTLs (chromosome)	h^2
LUE^{max}	4	2, 7, 8, 10	.12 each
W	2	6, 9	.16 each
FTF^{max}	1	4	.64
FDMC	1	5	.64
Z	1	11	.64
K	1	1	.95
B	1	3	.95

Note. LUE^{max}, maximum light use efficiency; *W*, time dependent scaling parameter; FTF^{max}, maximum level of fraction to fruits; FDMC, fruit dry matter content; *Z*, time dependent scaling parameter; *K*, light extinction coefficient; *B*, genotype specific slope for the regression of leaf area per shoot (m^2) on temperature sum.

TABLE 4 Yield analyzed in terms of genotype-specific physiological parameters and environmental characterizations (3 temperature levels, 3 radiation levels, 2 countries, and 2 CO_2 concentrations) using a factorial regression model. The percentages of explained variation are given for genotypic and environmental main effects as well as genotype × environment interaction. Results for 10 simulation runs are presented as minimum, maximum, and average across those runs. Bold fonts represent the maximum percentages of explained variance for the main effects and for the interaction

	Main effects			Interaction		
Parameter	Min.	Max.	Avg.	Min.	Max.	Avg.
	%					
Physiological parameters						
LUE^{max}	**64.85**	**74.81**	**69.82**	13.71	14.29	13.93
W	11.69	13.23	12.68	**24.59**	**27.08**	**25.88**
FTF^{max}	6.09	6.87	6.43	0.81	1.58	1.32
FDMC	7.15	8.41	7.66	1.14	2.03	1.59
Z	0.29	0.51	0.37	1.16	1.51	1.32
B	0.66	0.90	0.74	0.27	0.51	0.39
K	0.21	0.46	0.30	0.09	0.24	0.17
Environmental Variable						
Country	**41.04**	**41.98**	**41.49**	8.78	9.57	9.16
Temperature	19.50	20.19	19.85	**28.49**	**30.37**	**29.41**
CO_2	16.36	17.12	16.71	3.64	4.32	3.92
Radiation	9.10	9.63	9.37	1.90	2.32	2.11

Note: All other combinations of genotypic parameters and environmental variables represent at most 1.1% of the main effects or interactions. LUE^{max}, maximum light use efficiency; *W*, time dependent scaling parameter; FTF^{max}, maximum level of fraction to fruits; FDMC, fruit dry matter content; *Z*, time dependent scaling parameter; *K*, light extinction coefficient; *B*, genotype specific slope for the regression of leaf area per shoot (m^2) on temperature sum.

of QTL effects. For example, when the parameter LUE^{max} for genotype *i* would depend on a single QTL the model is as follows:

$$LUE_i^{max} = g\_LUE_i^{max} = x_i\alpha + g^*\_LUE_i^{max} \quad (7)$$

where we first assume the phenotypic differences for the physiological parameters to be equal to the genetic differences, or the heritability for the physiological parameters is 1. Next, we partition the genetic differences in a QTL part and a (polygenic) genetic residual, with x_i a function of the QTL genotype, α the QTL allele substitution effect for the crop growth parameter, assumed to be constant across all environments, and $g^*\_LUE_i^{max}$ the residual of the genetic effect for LUE^{max}. An expansion of the genetic and residual approach as shown in Equation 7 can be inserted in Equation 5 for each of the physiological parameters. Before inserting the component traits as defined in Equation 7, the mean and variance of the component traits were scaled to comply with the specifications given in Table 1.

Boxplots for one random realization of the simulated yields across environments (500 × 36 table) can be found in Supplemental Figure S1. These simulated data were subjected to an extensive study of GEI and QEI. For each environment, the realized average yield was calculated and subsequently a normally distributed error was added to the yields such that the coefficient of variation became 10%. This in an attempt to add a realistic nongenetic error variance.

2.5 | Sensitivity analyses

When dealing with a simulation model comprising multiple parameters, sensitivity analysis is typically used to ascertain the absolute and relative importance of the individual parameters. Here, the sensitivity of yield to the parameters was accomplished by applying multiple statistical methods for investigating two-way tables of genotype × environment means: factorial regression (van Eeuwijk et al., 1996), additive main effects and multiplicative interaction (AMMI) analysis (Gauch, 1988; Gollob, 1968; Mandel, 1969), and principal component analysis, or GGE (genotypic main effects and GEI) analysis (Yan & Kang, 2002). An overview of these techniques is presented in Malosetti et al. (2013), Rodrigues (2018), van Eeuwijk (1995), and van Eeuwijk et al. (2016), and their application in this paper is briefly presented below.

2.6 | Factorial regression

Factorial regression can best be understood as the imposition of contrasts on the levels of the row and column factor in a two-way table (Denis, 1988; van Eeuwijk et al., 1996,2016).

We can use contrasts in the direction of the genotypes to partition the original variation between genotypes in a part due to a contrast and a residual. This is valid for both the genotype main effect and the GEI. Our intention is to use yield components to define contrasts on the genotypes. As the yield components were generated to be uncorrelated, the interpretation of the decomposition of genotype main effect and GEI is relatively straightforward. For the environments, we can focus on the parts of the environmental main effect and the GEI that can be attributed to the initial four environment condition generating factors: country, temperature, CO_2, and radiation.

2.7 | Bilinear models: AMMI and GGE

As a follow up on the above sensitivity analyses by factorial regression, where we used explicitly defined covariates or contrasts, we also studied the series of simulated yields for the 500 backcross lines in 36 environments with explorative linear-bilinear techniques (Gauch et al., 2008, 2011; Paderewski et al., 2011; Paderewski & Rodrigues, 2014; van Eeuwijk et al., 2016). These techniques combine additive and multiplicative terms. Well-known representatives of this class of models are (a) the model underlying principal components analysis (PCA) of the genotype × environment table, also called GGE biplot model (see Yan and Kang, 2003), and (b) the AMMI model, which is a combination of ANOVA for the genotypic and environmental main effects and PCA for the residuals from additivity (Gauch, 1988, 1992; Gollob, 1968; Mandel, 1969). Useful generalizations of the AMMI model have also been proposed by Rodrigues et al. (2014, 2016), Assis et al. (2018), and Paderewski and Rodrigues (2018). The GGE model is

$$y_{i,j} = \mu + E_j + \sum_{n=1}^{N} b_{i,n} z_{j,n} + \varepsilon_{i,j} \quad (8a)$$

whereas the AMMI model can be written as

$$y_{i,j} = \mu + G_i + E_j + \sum_{n=1}^{N} b_{i,n} z_{j,n} + \varepsilon_{i,j} \quad (8b)$$

where $y_{i,j}$ is the yield of genotype i in environment j, μ is the grand mean, G_i are the genotype mean deviations (genotype means minus the grand mean), E_j are the environment mean deviations, $b_{i,n}$ and $z_{i,n}$ are the genotypic and environmental parameters (scores) for the nth multiplicative interaction term (i.e., the genotype and environment principal component scores and loadings for PCA axis n), N is the number of interaction principal component (IPC) axes retained, and $\varepsilon_{i,j}$ is a residual. In the GGE model, the genetic main effects and GEI are modeled simultaneously as a sum of multiplicative terms, whereas in the AMMI model only the GEI is modelled multiplicatively.

2.8 | QTL analysis

For the QTL analysis we used the mixed model QTL framework described in several papers (Boer et al., 2007; Malosetti et al., 2004, 2013) as implemented in GenStat (Boer et al., 2015; Payne et al., 2011). The major point of interest was whether QEI for yield could be detected and whether we could interpret this QEI in terms of QTLs for the underlying physiological parameters. The QTL model that we applied used explicit marker-derived information to describe the GEI in terms of QTLs in their dependence on the environments (i.e., the QEI). The inclusion of this marker information, genetic predictors, allows testing whether the phenotypic trait (e.g., yield) is affected by the DNA variation at a particular genome position, and whether this effect depends on the environment. The genotype × environment two-way table of means is first described by an ANOVA model with a GEI term, $(GE)_{i,j}$, whereafter the sum of genotypic main effect and GEI is modeled in terms of QTLs with environment-specific effects in a mixed linear model definition following Boer et al. (2007):

$$\begin{aligned} y_{i,j} &= (\mu + E_j) + [G_i + (GE)_{i,j}] \\ &= (\mu_j) + \left(\sum_{p=1}^{P} x_{p,i} \alpha_{p,j} + \varepsilon_{i,j} \right) \end{aligned} \quad (9)$$

where μ_j is the intercept for each environment, $x_{p,i}$ is derived from marker genotype information for genotype i, $\alpha_{p,j}$ the QTL allele substitution effect for environment j, P is total number of QTLs underlying $y_{i,j}$ (e.g., yield), and $\varepsilon_{i,j}$ follows a multivariate normal distribution with zero mean vector and a given variance-covariance (VCOV) matrix. The choice of the best VCOV structure was done following the procedure described in Malosetti et al. (2004) and Boer et al. (2007).

3 | RESULTS

3.1 | Factorial regression analysis

Table 4 shows the results of various types of factorial regression on the simulated genotype × environment tables of means. For the genotype main effect in yield, we see that the variation in LUE^{max} was dominant (i.e., the most explanatory covariable), whereas W, FTF^{max}, and FDMC contributed to a lesser extent to differences in genotypic means. For the GEI, W seems the most important genotypic variable, followed by LUE.

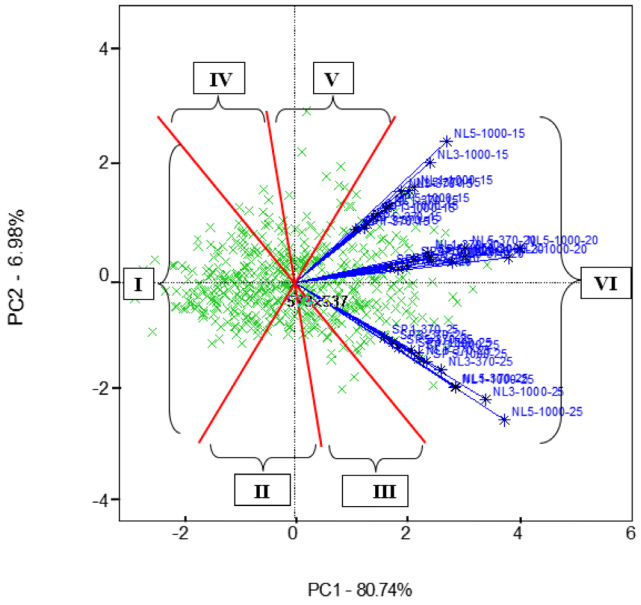

FIGURE 3 GGE (genotypic main effects and genotype × environment interaction) biplot for one random realization of yield (from samples of genotype specific parameters) of the two-way table with 500 genotypes and 36 environments. The abscissa shows the Principal Component 1 (PC1) scores, and the ordinate shows the PC2 scores. The 36 environments are marked by their code names (e.g., NL1-370-15 represents a Dutch environment with the lowest yearly average radiation in the considered historical period [NL1], $CO_2 = 370$ µmol mol^{-1}, and daily average temperature of 15 ˚C). The first and second axes explain a total of 87.72%. The different sections of the plot define various type of crossover interactions. See text for more details

For the environmental main effects (i.e., the differences between the environments as averaged across genotypes), country was the most important factor, probably because of the differences in length of the season and the differences in level and dynamics of radiation between Spain and the Netherlands. Temperature and CO_2 were about half as important as country, whereas radiation differences (low, average, and high radiation levels within country) were again about half the effect of temperature and CO_2, and about a quarter of that for country. Interactions between environmental factors were not found to add substantially to the average differences between environments. For environmental variables that describe GEI, it is mainly temperature that had influence, whereas country (i.e., season length and radiation record) also had influence, but three times less than temperature. Other factors or interactions could be ignored.

3.2 | GGE and AMMI analysis

Figure 3 shows the GGE biplot for the simulated virtual phenotype (i.e., yield), from one run of 500 genotypes grown in 36 environments (Supplemental Figure S1). The variation due to environments follows principally from temperature differences, in correspondence with the results of the factorial regressions. The longer vectors for the Netherlands and for high [CO_2] environments indicate that proportionally more variance is explained in the biplot for these environments, compared with the Spanish and the low-[CO_2] environments. The longer duration of the Netherlands environments would allow a greater expression of yield differences between genotypes. Similarly, high-[CO_2] environments, which generated greater yield differences, also have longer vectors. The near-orthogonal relationship between the 15 and 25 ˚C environments indicates that genotype performance was almost uncorrelated between these two types of environments. Zones of cross over interactions between temperature regimes occurred in the Sectors II, III, IV and V. Sector VI shows genotypes that were above average in yield everywhere, Sector I shows genotypes that were below average everywhere. Sector II shows genotypes that were below average in 20 ˚C, but above average at 25 ˚C, Sector III shows genotypes that were below average at 15 ˚C, but above average at 20 ˚C. In a similar way, the Sectors IV and V can be interpreted in terms of cross over interactions.

The interpretation of GGE and AMMI biplots is very similar. Figure 3 shows that cross over interactions can be generated for the complex trait (yield) from a set of component traits without GEI. The axes of the AMMI2 biplot (Supplemental Figure S2) can be interpreted as a function of temperature and radiation for IPC1 (interaction PC1) and as a function of radiation, CO_2 and country for IPC2 (interaction PC2). The environments appear as three diagonal bands in the plot, with from left to right diagonals for 25, 20, and 15 ˚C. Supplemental Figure S2 thus endorses the results from the factorial regression analysis and the GGE biplot analysis (Table 4, Figure 3), both emphasizing the dominant role of temperature. Besides temperature, also country plays an important role, as shown by the factorial regression as well, with Spanish environments being located in the upper right corner of the plot and Dutch environments in the left and lower parts. Carbon dioxide pushes environments to the lower left of the plot within the diagonal groups defined by the temperatures, less than what the factor country does, but more than what radiation does.

Table 5 gives a summary ANOVA table for simulated yield analysed with the AMMI2 model. The ranges for the proportions of variance explained by genotypes, environments, and GEI were [0.30; 0.36], [0.55; 0.64], and [0.12; 0.14], respectively. In the simulated phenotypic data, the GEI was responsible for 29.0% (mean value for the 10 runs, with values between 28.1 and 29.7%) of the genotype related sum of squares (SS; i.e., GGE). The first two IPCs were responsible for 16.0% of the GGE SS (Table 5).

TABLE 5 ANOVA for yield from the additive main effects and multiplicative interaction (AMMI) model with two interaction principal components. In the column for the sum of squares (SS), the mean values of 10 independent runs of our simulation model are reported, and in parens the range (minimum and maximum). For the mean squares (MS), only the values associated with the mean SS are reported. The grand (yield) mean is 20.014 kg m^{-2}

Source	df	SS (in 10^3)	MS
Total	17,999	1,274.9 (1133.5; 1331.8)	70.83
Genotypes	499	400.5 (377.2; 427.5)	802.69
Environments	35	727.6 (719.9; 737.0)	20,788.28
GEI	17,465	163.1 (159.2; 170.4)	9.34
IPC1	533	52.5 (49.5; 57.3)	98.45
IPC2	531	37.9 (35.6; 41.2)	71.42
Residual	16,401	72.7 (71.7; 74.1)	4.43

Note. GEI, genotype × environment interaction; IPC, interaction principal component.

TABLE 6 Quantitative trait locus (QTL) effects and (standard errors) on simulated yield for 10 QTLs as detected in a genome wide QTL scan (chr. 2–11) for several subsets of environments. The influence of the QTL increases with the absolute value of the QTL effects. The last column has the mean standard error (SE) for each environmental group

Parameter Variable[a]	LUE	*B*	FTF	FDMC	*W*	LUE	LUE	*W*	LUE	*Z*	
Chromosome	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**SE**
SP	4.46	0.87	3.85	−3.70	−1.72	3.49	3.91	−1.58	4.29	1.68	0.86
NL	6.84	1.05	6.41	−5.82	−2.10	5.70	5.90	−1.78	6.68	2.68	0.71
CO_2 = 370 µmol mol^{-1}	5.10	0.80	4.65	−4.14	−1.40	4.12	4.44	−1.17	5.07	1.97	0.85
CO_2 = 1,000 µmol mol^{-1}	6.20	1.12	5.62	−5.39	−2.42	5.07	5.36	-2.19	5.89	2.38	0.73
T = 15 °C	4.56	1.45	4.62	−3.88	0.25	4.21	3.91	1.04	4.66	3.47	0.62
T = 20 °C	6.77	1.11	6.29	−5.86	−1.06	5.67	6.05	−0.86	6.69	1.97	0.84
T = 25 °C	5.62	0.32	4.49	−4.55	−4.92	3.90	4.75	−5.22	5.10	1.09	0.89
SP, temp. = 15 °C	3.58	1.23	3.46	−3.01	0.20	3.18	3.10	0.83	3.74	2.78	0.68
SP, temp. = 20 °C	5.28	0.98	4.58	−4.39	−1.37	4.08	4.69	−1.33	4.98	1.44	0.91
SP, T temp. = 25 °C	4.52	0.39	3.52	−3.71	−3.98	3.20	3.94	−4.24	4.15	0.81	0.97
NL, temp. = 15 °C	5.55	1.66	5.77	−4.75	0.31	5.23	4.71	1.25	5.58	4.16	0.55
NL, T temp. = 20 °C	8.26	1.24	8.00	−7.32	−0.75	7.26	7.42	−0.39	8.39	2.51	0.76
NL, temp. = 25 °C	6.71	0.24	5.47	−5.39	−5.87	4.61	5.56	−6.21	6.06	1.37	0.80

[a]NL, the Netherlands; SP, Spain.

3.3 | QTL analyses

We have chosen one run (one seed) out of the 10 runs for the model in Equation 5 for illustration of a QTL analysis, other runs produced comparable results.

A preliminary analysis of the VCOV structure was carried out in order to model the genetic variances and correlations across environments. Following the procedure described by Boer et al. (2007) and Malosetti et al. (2004), both Akaike information criterion (AIC) and Schwarz information criterion (SIC, also known as Bayesian information criterion) indicated the best model to be the factor analytic with two multiplicative terms (FA2).

The genetic architecture used in this study comprised 11 QTLs for yield components (Table 3, Figure 2).

The results from the QTL analysis on simulated yield, using composite interval mapping (Zeng, 1994), and the factor analytic model with two multiplicative terms as VCOV structure, are presented in Table 6 and Figure 4. As shown there and in Supplemental Table S1, 10 out of the 11 simulated QTLs for yield components were found back as yield QTLs in many of the environments, often showing QEI, variations across environments for the QTL effect size.

We can observe QEI related to country for the yield QTL associated with LUE^{max} (chr. 2, 7, 8, and 10), FTF^{max} (chr. 4), and FDMC (chr. 5), when changing from Spanish to Dutch

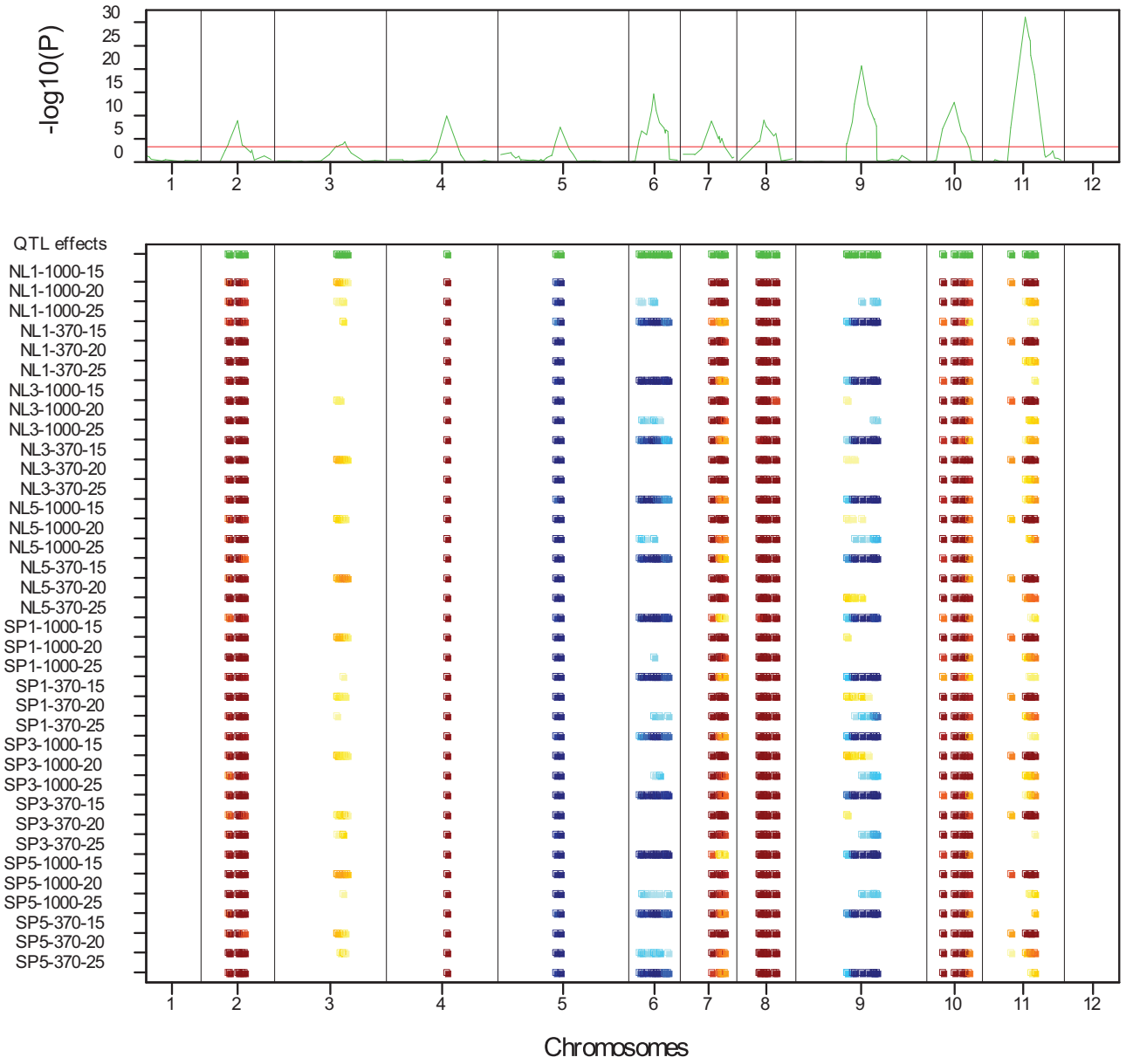

FIGURE 4 Genome quantitative trait locus (QTL) scan for the yield data. The top panel presents the profile expressed as $-\log_{10}(P)$ values for a test on the presence of a QTL effect in any environment. The red horizontal line is the 5% genome wide significance threshold. The bottom panel depicts the environment specific QTL effects with environment labels (left-hand side). The green line in the first row of the bottom panel summarizes the profile of top panel. For the subsequent lines corresponding to QTL tests for individual environments, blue represents an increasing effect of the allele of the reference parent, whereas red shows a decreasing effect. Higher intensity points to stronger effects, whereas lighter indicates weaker effects

locations. This finding is in agreement with the factorial regressions in Table 4 and the AMMI biplot (Supplemental Figure S2), where the IPC2 is a function of the country.

For *W* (chr. 6 and 9) and *Z* (chr. 11), we also clearly observed QEI in yield related to daily average temperature. The yield QTL following from the QTL for *B* on chr. 3 had a consistent effect across environments. No yield QTL was found corresponding to *K* (chr. 1).

Table 6 compares the yield QTL substitution effects for each of the QTLs in relation to country, CO_2 level, temperature level, and the combination of country and temperature. The effects in Table 6 can be interpreted as averages across the 36 environment-specific QTL effects obtained from the initial multi-environment QTL analysis. The differences in QTL effects between Spain and The Netherlands are not that large, and comparable with the differences induced by temperature. Carbon dioxide differences cause even smaller QTL effect differences. Largest QTL effect differences occur as a consequence of the combined effect of country and temperature, especially at 20 ˚C.

When considering the subsets of environments categorized by temperature in Table 6, it is clear that the yield QTL effects induced by the component QTLs for LUE^{max}, FTF^{max}, and FDMC present a curvilinear trend across temperature levels, whereas the remaining QTLs have a linear trend.

We conclude that the patterns observed in the QEI for yield, follow from the expression of QTLs for components, and can be understood in terms of the nature of the underlying yield components and their interaction with specific environmental factors. The QTL analyses in Table 6 are in good agreement with the factorial regressions in Table 4 and the GGE and AMMI biplots in Figure 3 and Supplemental Figure S2.

4 | DISCUSSION

4.1 | The importance of studying and understanding the GEI and QEI in simulation studies

Genotype-to-phenotype crop growth models have been widely used to study and understand the behavior of plant growth and development along the growing season. These studies focus mostly on the analysis of GEI and QTLs and, sometimes, on the analysis of QEI (see Section 1 for examples). **I**n this paper, however, we conduct an **e**xtensive statistical analysis of GEI and QEI to understand the relation between the physiological parameters and the yield phenotype. Another approach in the same spirit was presented by Letort et al. (2008), but only one environment was considered by these authors. In a large simulation study of sorghum "genotypes" grown in 600 dryland environments, Hammer et al. (2006) demonstrated how the magnitude of the yield QTL effects was related to the relative importance of each of four genotype-specific parameters controlling crop maturity, transpiration efficiency, stay-green, and utilization of carbohydrate reserves.

As in many engineering areas of research, simulation studies and synthetic data are powerful tools for complementing real world questions—in this case, questions related to breeding programs. Their use opens the possibility of controlling all the input parameters to better "model" the reality (Chapman et al., 2002; Hammer et al., 2006). The current simulations show how to obtain insight in the factors determining a complex trait like yield by using additional genotypic and environmental (covariate) information in the analysis of multiple environment data for the complex trait. A QEI analysis for yield using information on yield components and environmental characterizations allows the partial unravelling of the genotype-to-phenotype function and the genetic architecture involved. Bustos-Korts, Boer, et al. (2019) and Bustos-Korts, Malosetti, et al. (2019) simulated wheat (*Triticum aestivum* L.) data using the Agricultural Production Systems sIMulator (APSIM) to study, among other things, the possibilities to increase prediction accuracies in multi-trait genomic prediction. This work is reminiscent of the work presented in this paper on pepper.

4.2 | Simulated GEI and QEI generated with a crop growth model and analyzed by statistical models

The integration of statistical genetics and crop growth modeling for reliable and robust prediction of phenotypic traits, on the basis of genotypic-specific and stable physiological parameters and environmental characterizations, is the object of extensive research in plant sciences (see Section 1). A challenge in the application of these models is to obtain realistic estimates for the parameters for a large set of genotypes.

In this study, we considered a parsimonious crop growth model with a small number of parameters and determined to which extent complex GEI in yield could be generated and whether such GEI could be modeled and interpreted in terms of underlying yield components by a set of statistical models developed to analyze patterns in GEI and QEI. Despite its simplicity, the crop growth model simulated GEI and QEI, including crossovers (Figure 3), for yield. While focusing on GEI in yield (cf., QEI), Chapman (2008) demonstrated for sorghum [*Sorghum bicolor* (L.) Moench] simulations with four traits in 30 environments (including drought), that a biplot approach could be used to interpret GEI resulting from relative changes in biomass at different times during a season. Chapman (2008) also demonstrated that the biplot for GEI of simulated final yield was far more complex than that for final biomass, which also showed the statistical complexities that are generated by the (physiological) processes transforming biomass into yield.

It needs to be admitted here that GEI and QEI with crossovers could only be obtained by "penalization" of yield components in relation to environmental factors (e.g., higher fruit abortion at higher temperatures, drought influence, pests, etc.); as is the experience of breeders, stress impacts are responsible for much of the observed GEI and QEI.

4.3 | Benefitting from pleiotropy of QTLs for yield and yield components

Yield QTLs were detected at the exact same place where the underlying yield component QTLs were allocated in the simulation. These QTLs were detected in final yield data. Breeders search for yield QTLs that are stable across environments (i.e., QTLs that are well described by QTL main effects). The yield QTLs induced by the QTLs for LUE^{max} were clearly the most important ones in our simulations, followed by FDMC and FTF^{max} (Table 6, Supplemental Table S1, and lower panel of

Figure 4). In real conditions, a multi-trait multi-environment QTL analysis for yield and yield components will identify yield QTLs that are pleiotropic with LUE^{max}, FDMC, and FTF^{max}. Yield QTLs corresponding to *W* showed QEI with higher $-\log_{10}(P)$ values at higher temperatures, whereas the yield QTL corresponding to *Z* had lower $-\log_{10}(P)$ values at higher temperatures (Figure 4, Table 6). The yield QTLs corresponding to the components *K* and *B* were not consistently detected and apparently less crucial for yield. This is apparent in Table 6 and Supplemental Table S1 and in the lower panel of Figure 4. At the environmental side, we observed the lack of importance of CO_2 for GEI in final yield.

At the level of overall GEI as analyzed by factorial regression and AMMI, the main environmental driver for GEI was temperature and, to a lesser degree, country (total radiation over growing season), whereas on the genotypic side, variation in the yield components *W* and LUE^{max} was identified as inducing GEI. When we look at the nature of the QTLs that were identified (i.e., the yield components to which the QTLs corresponded), it comes as no surprise that the QTL for *W* was strongly expressed and easily identified as a yield QTL. Similar, the QTLs for LUE^{max} were expected to be picked up by QTL analysis for yield. It is less evident why the QTLs for FDMC, FTF^{max}, and *Z* are also strongly expressed alongside the QTLs for *W* and LUE^{max}, whereas the QTLs for *K* and *B* have no or a minor effect on yield. What breeders can learn from this is that yield improvement in tomato (*Solanum lycopersicum* L.) for Dutch and Spanish growing conditions will not benefit from attention for *K* and *B*.

For the QEI, the QTLs for LUE^{max}, FTF^{max}, and FDMC had the strongest expression at the intermediate temperature of 20 ˚C, whereas *W* was strongest at 15 ˚C and *Z* at 25 ˚C. The pattern of temperature dependence for all these major yield QTLs was only slightly modified by Country (Table 6, Figure 4, Supplemental Table S1). From the structure of the crop growth model, it is not straightforward to predict which yield components will cause QEI and GEI and how QTL expression will vary with environmental conditions. A priori, it seems hard to define an environmental dependence signature for a yield QTL that would allow the identification of the underlying yield component QTL. The only way to create insight in environmental dependence patterns for QEI as occurring in yield QTLs because of yield components interacting with the environment is by running crop growth simulations for many forms of genetic diversity combined with specified environmental diversity. Such simulations may lead to the formulation of environmental expression keys for yield components in QTL analyses for yield. Closely related to the development of such QEI environmental expression signatures is the work on environmental characterization by crop growth simulation, where the correlation structure of yield and yield components is used to define the class of environmental (stress) conditions that is pertinent to a set of yield trials (Bustos-Korts, Boer, et al., 2019; Bustos-Korts, Malosetti, 2019; van Eeuwijk et al., 2019). In the latter work, the importance of the inclusion in the analysis of the dynamics of the yield components during the growing season is emphasized. In the current study, these dynamics were not considered.

For breeders, the identification of underlying physiological parameters and their QTLs as being responsible for GEI and QEI in yield can assist in developing more efficient strategies for the improvement of yield. For example, if a crop growth model with a limited number of environment-independent physiological parameters can predict yield under GEI, it suggests that it could be sensible to develop phenotyping techniques to cheaply assess or approximate those parameters, whereas genomic prediction techniques may help to produce predictive models for the same parameters as functions of marker profiles. The integrated use of statistical genetic modeling and crop growth modeling in combination with modern phenotyping, genotyping, and envirotyping techniques to arrive at a new generation of predictive models for yield was foreseen by Cooper et al. (2014). Recent papers showing the possibilities of such a hybrid approach to predictive modeling are Messina et al. (2018), Millet et al. (2016, 2019), and van Eeuwijk et al. (2019). The current paper provides an illustration of how synthetic yield data generated by a crop growth model and analyzed by statistical models can help to develop a framework for quantifying the importance of yield components and their QTLs for improvements of yield under GEI.

ACKNOWLEDGMENTS

P.C. Rodrigues acknowledges financial support from the National Council for Scientific and Technological Development (CNPq) through the project UNIVERSAL-MCTI/CNPq 421580/2018-6, and the CNPq grant “bolsa de produtividade PQ-2” 305852/2019-1. Fred van Eeuwijk worked on this paper within the context of the EPPN2020 project. The authors would like to thank Marcos Malosetti and Martin Boer for the fruitful discussions on QEI procedures in Genstat.

AUTHOR CONTRIBUTIONS

Paulo C. Rodrigues: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Software, Visualization, Writing-original draft, Writing-review & editing; **Ep Heuvelink**: Conceptualization, Data curation, Formal analysis, Methodology, Supervision, Validation, Visualization, Writing-original draft, Writing-review & editing; **Leo F. M. Marcelis**: Conceptualization, Data curation, Writing-original draft, Writing-review & editing; **Scott C. Chapman**: Conceptualization, Writing-original draft, Writing-review & editing; **Fred A. Van Eeuwijk**: Conceptualization, Data curation, Formal analysis, Funding acquisition, Methodology, Supervision, Validation,

Visualization, Writing-original draft, Writing-review & editing.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Paulo C. Rodrigues https://orcid.org/0000-0002-1248-9910

REFERENCES

Alimi, N. A., Bink, M. C. A. M., Dieleman, J. A., Magán, J. J., Wubs, A. M., Palloix, A., & van Eeuwijk, F. A. (2013). Multi-trait and multi-environment QTL analyses of yield and a set of physiological traits in pepper. *Theoretical and Applied Genetics*, *126*, 2597–2625. https://doi.org/10.1007/s00122-013-2160-3

Assis, T. O. G., Dias, C. T. S., & Rodrigues, P. C. (2018). A weighted AMMI algorithm for nonreplicated data. *Pesquisa Agropecuária Brasileira*, *53*, 557–565. https://doi.org/10.1590/s0100-204x2018000500004

Baldazzi, V., Bertin, N., Genard, M., Gautier, H. Desnoues, E., & Quilot-Turion, B. (2016). Challenges in integrating genetic control in plant and crop models. In X. Yin, & P. Struik (Eds.), *Crop systems biology*. Springer. https://doi.org/10.1007/978-3-319-20562-5_1

Barchi, L., Bonnet, J., Boudet, C., Signoret, P., Nagy, I., Lanteri, S., Palloix, A., & Lefebvre, V. (2007). A high-resolution, intraspecific linkage map of pepper (*Capsicum annuum* L.) and selection of reduced recombinant inbred line subsets for fast mapping. *Genome*, *50*, 51–60. https://doi.org/10.1139/g06-140

Barchi, L., Lefebvre, V., Sage-Palloix, A. M., Lanteri, S., & Palloix, A. (2009). QTL analysis of plant development and fruit traits in pepper and performance of selective phenotyping. *Theoretical and Applied Genetics*, *118*, 1157–1171. https://doi.org/10.1007/s00122-009-0970-0

Bertin, N., Martre, P., Genard, M., Quilot, B., & Salon, C. (2010). Under what circumstances can process-based simulation models link genotype to phenotype for complex traits? Case-study of fruit and grain quality traits. *Journal of Experimental Botany*, *61*, 955–967. https://doi.org/10.1093/jxb/erp377

Bhakta, M. S., Gezan, S. A., Michelangeli, J. A. C., Carvalho, M., Zhang, L., Jones, J. W., Boote, K. J., Correll, M. J., Beaver, J., Osorno, J. M., Colbert, R., Rao, I., Beebe, S., Gonzalez, A., Ricaurte, J., & Vallejos, C. E. (2017). A predictive model for time-to-flowering in the common bean based on QTL and environmental variables. *G3: Genes, Genomes, Genetics*, *7*, 3901–3912. https://doi.org/10.1534/g3.117.300229

Boer, M. P., Cave, V., Jansen, J., Malosetti, M., Mathews, K., Murray, D., van Eeuwijk, F. A., & Welham, S. (2015). *A guide to QTL analysis in Genstat*. VSN International.

Boer, M. P., Wright, D., Feng, L. Z., Podlich, D. W., Luo, L., Cooper, M., & van Eeuwijk, F. A. (2007). A mixed-model quantitative trait loci (QTL) analysis for multiple-environment trial data using environmental covariables for QTL-by-environment interactions, with an example in maize. *Genetics*, *177*, 1801–1813. https://doi.org/10.1534/genetics.107.071068

Broman, K. W., & Sen, S. (2009). *A guide to QTL mapping with R/qtl*. Springer.

Bustos-Korts, D., Boer, M. P., Malosetti, M., Chapman, S., Chenu, K., Zheng, B., & van Eeuwijk, F. A. (2019). Combining crop growth modelling and statistical genetic modelling to evaluate phenotyping strategies. *Frontiers in Plant Science*, *10*, 1491. https://doi.org/10.3389/fpls.2019.01491

Bustos-Korts, D., Malosetti, M., Chapman, S., & van Eeuwijk, F. A. (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 X. Yin, & P. Struik (Eds.), *Crop systems biology* (pp. 55–82). Springer. https://doi.org/10.1007/978-3-319-20562-5_3

Bustos-Korts, D., Malosetti, M., Chenu, K., Chapman, S., Boer, M. P., Zheng, B., & van Eeuwijk, F. A. (2019). From QTLs to adaptation landscapes: Using genotype-to-phenotype models to characterize G × E over time. *Frontiers in Plant Science*, *10*, 1540. https://doi.org/10.3389/fpls.2019.01540

Bustos-Korts, D., Romagosa, I., Borràs-Gelonch, G., Casas, A. M., Slafer, G. G. A., & van Eeuwijk, F. A. (2018). Genotype by environment interaction and adaptation. In *Encyclopedia of sustainability science and technology*. Springer. https://doi.org/10.1007/978-1-4939-2493-6_199-3

Chapman, S. C. (2008). Use of crop models to understand genotype by environment interactions for drought in real-world and simulated plant breeding trials. *Euphytica*, *161*, 195–208. https://doi.org/10.1007/s10681-007-9623-z

Chapman, S. C., Cooper, M., & Hammer, G. L. (2002). Using crop simulation to generate genotype by environment interaction effects for sorghum in water-limited environments. *Australian Journal of Agricultural Research*, *53*, 379–389. https://doi.org/10.1071/AR01070

Chenu, K., Chapman, S. C., Hammer, G. L., Mclean, G., Salah, H. B. H., & 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. https://doi.org/10.1111/j.1365-3040.2007.01772.x

Chenu, K., Chapman, S. C., Tardieu, F., McLean, G., Welcker, C., & Hammer, G. L. (2009). Simulating the yield impacts of organ-level quantitative trait loci associated with drought response in maize: A "gene-to-phenotype" modeling approach. *Genetics*, *183*, 1507–1523. https://doi.org/10.1534/genetics.109.105429

Cooper, M., Messina, C. D., Podlich, D., Totir, L. R., Baumgarten, A., Hausmann, N. J., Wright, D., & Graham, G. (2014). Predicting the future of plant breeding: Complementing empirical evaluation with genetic prediction. *Crop and Pasture Science*, *65*, 311–336. https://doi.org/10.1071/CP14007

Cooper, M., Podlich, D. W., & Smith, O. S. (2005). Gene-to-phenotype models and complex traits. *Australian Journal of Agricultural Research*, *56*, 895–918. https://doi.org/10.1071/AR05154

Cooper, M., Technow, F., Messina, C., Gho, C., & Totir, L. R. (2016). Use of crop growth models with whole-genome prediction: Application to a maize multienvironment trial. *Crop Science*, *56*, 2141–2156. https://doi.org/10.2135/cropsci2015.08.0512

Cooper, M., van Eeuwijk, F. A., Hammer, G. L., Podlich, D. W., & Messina, C. (2009). Modeling QTL for complex traits: Detection and context for plant breeding. *Current Opinion in Plant Biology*, *12*, 231–240. https://doi.org/10.1016/j.pbi.2009.01.006

de los Campos, G., Hickey, J. M., Pong-Wong, R., Daetwyler, H. D., & Calus, M. P. L. (2013). Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics*, *193*, 327–345. https://doi.org/10.1534/genetics.112.143313

de Swart, E. A. M., Marcelis, L. F. M., & Voorrips, R. E. (2006). Variation in relative growth rate and growth traits in wild and cultivated *Capsicum* accessions grown under different temperatures. *Journal of Horticultural Science & Biotechnology*, *81*, 1029–1037.

Denis, J. B. (1988). Two way analysis using covarites. *Statistics*, *19*, 123–132. https://doi.org/10.1080/02331888808802080

Ezui, K. S., Leffelaar, P. A., Franke, A. C., Mando, A., & Giller, K. E. (2018). Simulating drought impact and mitigation in cassava using the LINTUL model. *Field Crops Research*, *219*, 256–272. https://doi.org/10.1016/j.fcr.2018.01.033

Gauch, H. G. (1988). Model selection and validation for yield trials with interaction. *Biometrics*, *44*, 705–715. https://doi.org/10.2307/2531585

Gauch, H. G. (1992). *Statistical analysis of regional yield trials: AMMI analysis of factorial designs*. Elsevier.

Gauch, H. G., Piepho, H.-P., & Annicchiarico, P. (2008). Statistical analysis of yield trials by AMMI and GGE: Further considerations. *Crop Science*, *48*, 866–889. https://doi.org/10.2135/cropsci2007.09.0513

Gauch, H. G., Rodrigues, P. C., Munkvold, J. D., Heffner, E. L., & Sorrells, M. (2011). Two new strategies for detecting and understanding QTL by environment interactions. *Crop Science*, *51*, 96–113. https://doi.org/10.2135/cropsci2010.04.0206

Gelder, A. D., Raaphorst, M., Hoon, M.d., & Breugem, F. (2007). *Paprikateelt in de gesloten kas : Resultaten bij Themato in 2006*. Wageningen UR.

Gollob, H. F. (1968). A statistical model which combines features of factor analysis and analysis of variance techniques. *Psychometrika*, *33*, 73–115. https://doi.org/10.1007/BF02289676

Goudriaan, J., & Laar, H. H. V. (1994). *Modelling potential crop growth processes : Textbook with exercises*. Kluwer.

Hammer, G., Cooper, M., Tardieu, F., Welch, S., Walsh, B., van Eeuwijk, F., Chapman, S., & Podlich, D. (2006). Models for navigating biological complexity in breeding improved crop plants. *Trends in Plant Science*, *11*, 587–593. https://doi.org/10.1016/j.tplants.2006.10.006

Hammer, G. L., Chapman, S., van Oosterom, E., & Podlich, D. W. (2005). Trait physiology and crop modelling as a framework to link phenotypic complexity to underlying genetic systems. *Australian Journal of Agricultural Research*, *56*, 947–960. https://doi.org/10.1071/AR05157

Heslot, N., Jannink, J.-L., & Sorrells, M. E. (2015). Perspectives for genomic selection applications and research in plants. *Crop Science*, *55*, 1–12. https://doi.org/10.2135/cropsci2014.03.0249

Heuvelink, E. (1995). Growth, development and yield of a tomato crop: Periodic destructive measurements in a greenhouse. *Scientia Horticulturae*, *61*, 77–99. https://doi.org/10.1016/0304-4238(94)00729-Y

Higashide, T., & Heuvelink, E. (2009). Physiological and morphological changes over the past 50 years in yield components in tomato. *Journal of the American Society for Horticultural Science*, *134*, 460–465. https://doi.org/10.21273/JASHS.134.4.460

Ishii, T., Hayashi, T., & Yonezawa, K. (2010). Categorization of quantitative trait loci by their functional roles: QTL analysis for chemical concentration in seed grains. *Crop Science*, *50*, 784–793. https://doi.org/10.2135/cropsci2009.01.0015

Iwanska, M., Paderewski, J., Stepien, M., & Rodrigues, P. C. (2020). Adaptation of winter wheat cultivars to different environments: A case study in Poland. *Agronomy*, *10*, 632. https://doi.org/10.3390/agronomy10050632

Jarquín, D., Crossa, J., Lacaze, X., Cheyron, P. D., Daucourt, J., Lorgeou, J., Piraux, F., Guerreiro, L., Pérez, P., Calus, M., Burgueño, J., & de los Campos, G. (2014). A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theoretical and Applied Genetics*, *127*, 595–607. https://doi.org/10.1007/s00122-013-2243-1

Jarquín, D., Silva, C. L., Gaynor, R. C., Poland, J., Fritz, A., Howard, R., Battenfield, S., & Crossa, J. (2017). Increasing genomic-enabled prediction accuracy by modeling genotype × environment interactions in Kansas wheat. *The Plant Genome*, *10*(2). https://doi.org/10.3835/plantgenome2016.12.0130

Letort, V., Mahe, P., Cournede, P. H., De Reffye, P., & Courtois, B. (2008). Quantitative genetics and functional-structural plant growth models: Simulation of quantitative trait loci detection for model parameters and application to potential yield optimization. *Annals of Botany*, *101*, 1243–1254. https://doi.org/10.1093/aob/mcm197

Li, T., Heuvelink, E., van Noort, F., Kromdijk, J., & Marcelis, L. F. M. (2014). Responses of two Anthurium cultivars to high daily integrals of diffuse light. *Scientia Horticulturae*, *179*, 306–313. https://doi.org/10.1016/j.scienta.2014.09.039

Malosetti, M., Bustos-Korts, D., Boer, M. P., & van Eeuwijk, F. A. (2016). Predicting responses in multiple environments: Issues in relation to genotype x environment interactions. *Crop Science*, *56*, 2210–2222. https://doi.org/10.2135/cropsci2015.05.0311

Malosetti, M., Ribaut, J.-M., & van Eeuwijk, F. A. (2013). The statistical analysis of multi-environment data: Modeling genotype-by-environment interaction and its genetic basis. *Frontiers in Physiology*, *4*, 44. https://doi.org/10.3389/fphys.2013.00044

Malosetti, M., Voltas, J., Romagosa, I., Ullrich, S. E., & van Eeuwijk, F. A. (2004). Mixed models including environmental covariables for studying QTL by environment interaction. *Euphytica*, *137*, 139–145. https://doi.org/10.1023/B:EUPH.0000040511.46388.ef

Mandel, J. (1969). Partitioning of interaction in analysis of variance. *Journal of Research of the National Bureau of Standards—B. Mathematical Sciences*, *73*, 309–328.

Marcelis, L. F. M., Elings, A., Dieleman, J. A., Brajeul, E., Bakker, M. J., & Heuvelink, E. (2006). Modelling dry matter production and partitioning in sweet pepper. *Acta Horticulturae*, *718*, 121–128. https://doi.org/10.17660/ActaHortic.2006.718.13

Marcelis, L. F. M., Heuvelink, E., & Goudriaan, J. (1998). Modelling biomass production and yield of horticultural crops: A review. *Scientia Horticulturae*, *74*, 83–111. https://doi.org/10.1016/S0304-4238(98)00083-1

Messina, C. D., Technow, F., Tang, T., Totir, R., Gho, C., & Cooper, M. (2018). Leveraging biological insight and environmental variation to improve phenotypic prediction: Integrating crop growth models (CGM) with whole genome prediction (WGP). *European Journal of Agronomy*, *100*, 151–162. https://doi.org/10.1016/j.eja.2018.01.007

Millet, E. J., Kruijer, W., Coupel-Ledru, A., Prado, S. A., Cabrera-Bosquet, L., Lacube, S., Charcosset, A., Welcker, C., van Eeuwijk, F. A., & Tardieu, F. (2019). Genomic prediction of maize yield across European environmental conditions. *Nature Genetics*, *51*, 952–956. https://doi.org/10.1038/s41588-019-0414-y

Millet, E. J., Welcker, C., Kruijer, W., Negro, S., Coupel-Ledru, A., Nicolas, S. D., Laborde, J., Bauland, C., Praud, S., Ranc, N., Presterl, T., Tuberosa, R., Bedo, Z., Draye, X., Usadel, B., Charcosset, A., Van Eeuwijk, F., & Tardieu, F. (2016). Genome-wide analysis of yield in Europe: Allelic effects vary with drought and heat scenarios. *Plant Physiology*, *172*, 749–764. https://doi.org/10.1104/pp.16.00621

Montesinos-López, A., Montesinos-López, O. A., Gianola, D., Crossa, J., & Hernández-Suárez, C. M. (2018). Multi-environment genomic

prediction of plant traits using deep learners with dense architecture. *G3: Genes, Genomes, Genetics*, *8*, 3813–3828. https://doi.org/10.1534/g3.118.200740

Montesinos-López, O. A., Montesinos-López, A., Crossa, J., Gianola, D., Hernández-Suárez, C. M., & Martín-Vallejo, J. (2018). Multi-trait, multi-environment deep learning modeling for genomic-enabled prediction of plant traits. *G3: Genes, Genomes, Genetics*, *8*, 3829–3840. https://doi.org/10.1534/g3.117.300309

Nederhoff, E. M. (1994). *Effects of CO_2 concentrations on photosynthesis, transpiration, and production of greenhouse fruit vegetable crops*. Wageningen University.

Paderewski, J., Gauch, H. G., Madry, W., Drzazga, T., & Rodrigues, P. C. (2011). Yield response of winter wheat to agro-ecological conditions using additive main effects and multiplicative interaction and cluster analysis. *Crop Science*, *51*, 969–980. https://doi.org/10.2135/cropsci2010.05.0278

Paderewski, J., & Rodrigues, P. C. (2014). The usefulness of EM-AMMI to study the influence of missing data pattern and application to Polish post-registration winter wheat data. *Australian Journal of Crop Science*, *8*, 640–645.

Paderewski, J., & Rodrigues, P. C. (2018). Constrained AMMI model: Application to Polish winter-wheat post-registration data. *Crop Science*, *58*, 1458–1469. https://doi.org/10.2135/cropsci2017.06.0347

Payne, R. W., Murray, D. A., Harding, S. A., Baird, D. B., & Soutar, D. M. (2011). *An introduction to GenStat for windows* (14th ed.). VSN International.

Quero, G., Bonnecarrère, V., Fernández, S., Silva, P., Simondi, S., & Borsani, O. (2018). Light-use efficiency and energy partitioning in rice is cultivar dependent. *Photosynthesis Research*, *140*, 51–63. https://doi.org/10.1007/s11120-018-0605-x

Quilot, B., Genard, M., Lescourret, F., & Kervella, J. (2005). Simulating genotypic variation of fruit quality in an advanced peach x *Prunus davidiana* cross. *Journal of Experimental Botany*, *56*, 3071–3081. https://doi.org/10.1093/jxb/eri304

R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Reymond, M., Muller, B., Leonardi, A., Charcosset, A., & Tardieu, F. (2003). Combining quantitative trait loci analysis and an ecophysiological model to analyze the genetic variability of the responses of maize leaf growth to temperature and water deficit. *Plant Physiology*, *131*, 664–675. https://doi.org/10.1104/pp.013839

Reymond, M., Muller, B., & Tardieu, F. (2004). Dealing with the genotype × environment interaction via a modelling approach: A comparison of QTLs of maize leaf length or width with QTLs of model parameters. *Journal of Experimental Botany*, *55*, 2461–2472. https://doi.org/10.1093/jxb/erh200

Rijsdijk, A. A., & Houter, G. (1993). Validation of a model for energy consumption, CO_2 consumption and crop production (epc-model). *Acta Agriculturae*, *328*, 125–131.

Rodrigues, P. C. (2018). Statistical methods to detect and understand genotype-by-environment interaction and QTL-by-environment interaction. *Biometrical Letters*, *55*, 123–138. https://doi.org/10.2478/bile-2018-0009

Rodrigues, P. C., Monteiro, A., & Lourenço, V. M. (2016). A robust additive main effects and multiplicative interaction model for the analysis of genotype-by-environment data. *Bioinformatics*, *32*, 58–66.

Rodrigues, P. C., Malosetti, M., Gauch, H. G., & van Eeuwijk, F. A. (2014). A weighted AMMI algorithm to study genotype-by-environment interaction and QTL-by-environment interaction. *Crop Science*, *54*, 1555–1570. https://doi.org/10.2135/cropsci2013.07.0462

Tardieu, F. (2003). Virtual plants: Modelling as a tool for the genomics of tolerance to water deficit. *Trends in Plant Science*, *8*, 9–14. https://doi.org/10.1016/S1360-1385(02)00008-0

Technow, F., Messina, C. D., Totir, L. R., & Cooper, M. (2015). Integrating crop growth models with whole genome prediction through approximate Bayesian computation. *PLOS ONE*, *10*, e0130855. https://doi.org/10.1371/journal.pone.0130855

van Eeuwijk, F. A. (1995). Linear and bilinear models for the analysis of multi-environment trials .1. An inventory of models. *Euphytica*, *84*, 1–7. https://doi.org/10.1007/BF01677551

van Eeuwijk, F. A., Bink, M. C. A. M., Chenu, K., & Chapman, S. C. (2010). Detection and use of QTL for complex traits in multiple environments. *Current Opinion in Plant Biology*, *13*, 193–205. https://doi.org/10.1016/j.pbi.2010.01.001

van Eeuwijk, F. A., Bustos-Korts, D. V., & Malosetti, M. (2016). What should students in plant breeding know about the statistical aspects of genotype × environment interactions? *Crop Science*, *56*, 2119–2140. https://doi.org/10.2135/cropsci2015.06.0375

van Eeuwijk, F. A., Bustos-Korts, D., Millet, E. J., Boer, M. P., Kruijer, W., Thompson, A., Malosetti, M., Iwata, H., Quiroz, R., Kuppe, C., & Muller, O. (2019). Modelling strategies for assessing and increasing the effectiveness of new phenotyping techniques in plant breeding. *Plant Science*, *282*, 23–39. https://doi.org/10.1016/j.plantsci.2018.06.018

van Eeuwijk, F. A., Denis, J. B., & Kang, M. S. (1996). Incorporating additional information on genotypes and environments in models for two-way genotype by environment tables. In M. S. Kang & H. G. Gauch (Eds.), *Genotype by environment interaction: New perspectives* (pp. 15–49). CRC Press.

van Eeuwijk, F. A., Malosetti, M., Yin, X. Y., Struik, P. C., & Stam, P. (2005). Statistical models for genotype by environment data: From conventional ANOVA models to eco-physiological QTL models. *Australian Journal of Agricultural Research*, *56*, 883–894. https://doi.org/10.1071/AR05153

van Ittersum, M. K., Leffelaar, P. A., van Keulen, H., Kropff, M. J., Bastiaans, L., & Goudriaan, J. (2003). On approaches and applications of the Wageningen crop models. *European Journal of Agronomy*, *18*, 201–234. https://doi.org/10.1016/S1161-0301(02)00106-5

Wallach, D., Hwang, C., Correll, M. J., Jones, J. W., Boote, K., Hoogenboom, G., Gezan, S., Bhakta, M., & Vallejos, C. E. (2018). A dynamic model with QTL covariables for predicting flowering time of common bean (*Phaseolus vulgaris*) genotypes. *European Journal of Agronomy*, *101*, 200–209. https://doi.org/10.1016/j.eja.2018.10.003

Wubs, A. M., Heuvelink, E., Hemerik, L., & Marcelis, L. F. M. (2011). Stochastic dynamic simulation of fruit abortion: A case study of sweet pepper. *Acta Horticulturae*, *893*, 765–772. http://dx.doi.org/10.17660/ActaHortic.2011.893.82

Wubs, A. M., Heuvelink, E., & Marcelis, L. F. M. (2009). Abortion of reproductive organs in sweet pepper (*Capsicum annuum* L.): A review. *Journal of Horticultural Science & Biotechnology*, *84*, 467–475.

Yan, W., & Kang, M. S. (2002). *GGE biplot analysis: A graphical tool for breeders, geneticists, and agronomists*. CRC Press.

Yan, W., & Kang, M. S. (2003). *GGE biplot analysis : A graphical tool for breeders, geneticists, and agronomists*. CRC Press.

Yin, X. Y., Chasalow, S. D., Dourleijn, C. J., Stam, P., & Kropff, M. J. (2000). Coupling estimated effects of QTLs for physiological traits to a crop growth model: Predicting yield variation among recombinant inbred lines in barley. *Heredity*, *85*, 539–549. https://doi.org/10.1046/j.1365-2540.2000.00790.x

Yin, X. Y., Struik, P. C., & Kropff, M. J. (2004). Role of crop physiology in predicting gene-to-phenotype relationships. *Trends in Plant Science*, *9*, 426–432. https://doi.org/10.1016/j.tplants.2004.07.007

Zeng, Z. B. (1994). Precision mapping of quantitative trait loci. *Genetics*, *136*, 1457–1468. https://doi.org/10.1093/genetics/136.4.1457

Zheng, B., Biddulph, B., Li, D., Kuchel, H., & 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*, 3747–3761. https://doi.org/10.1093/jxb/ert209

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Rodrigues PC, Heuvelink E, Marcelis LFM, Chapman SC, van Eeuwijk FA. An analysis of simulated yield data for pepper shows how genotype × environment interaction in yield can be understood in terms of yield components and their QTLs. *Crop Science*. 2021;1-17. https://doi.org/10.1002/csc2.20476