Chapter 3
Modelling of Genotype by Environment Interaction and Prediction of Complex Traits across Multiple Environments as a Synthesis of Crop Growth Modelling, Genetics and Statistics

Daniela Bustos-Korts, Marcos Malosetti, Scott Chapman, and Fred van Eeuwijk

Abstract Selection processes in plant breeding depend critically on the quality of phenotype predictions. The phenotype is classically predicted as a function of genotypic and environmental information. Models for phenotype prediction contain a mixture of statistical, genetic and physiological elements. In this chapter, we discuss prediction from linear mixed models (LMMs), with an emphasis on statistics, and prediction from crop growth models (CGMs), with an emphasis on physiology. Three modalities of prediction are distinguished: predictions for new genotypes under known environmental conditions, predictions for known genotypes under new environmental conditions, and predictions for new genotypes under new environmental conditions.

For LMMs, the genotypic input information includes molecular marker variation, while the environmental input can consist of meteorological, soil and management variables. However, integrated types of environmental characterizations obtained from CGMs can also serve as environmental covariable in LMMs. LMMs consist of a fixed part, corresponding to the mean for a particular genotype in a particular environment, and a random part defined by genotypic and environmental

D. Bustos-Korts
Biometris, Wageningen University & Research Centre,
P.O. Box 16, 6700 AC Wageningen, The Netherlands

C.T. de Wit Graduate School for Production Ecology & Resource Conservation (PE&RC), Wageningen, The Netherlands

M. Malosetti • F. van Eeuwijk (✉)
Biometris, Wageningen University & Research Centre,
P.O. Box 16, 6700 AC Wageningen, The Netherlands
e-mail: fred.vaneeuwijk@wur.nl

S. Chapman
CSIRO Plant Industry and Climate Adaptation Flagship, Queensland Bioscience Precinct,
306 Carmody Road, St Lucia, QLD 4067, Australia

X. Yin, P.C. Struik (eds.), *Crop Systems Biology*,
DOI 10.1007/978-3-319-20562-5_3

variances and correlations. For prediction via the fixed part, genotypic and/or environmental covariables are required as in classical regression. For predictions via the random part, correlations need to be estimated between observed and new genotypes, between observed and new environments, or both. These correlations can be based on similarities calculated from genotypic and environmental covariables. A simple type of covariable assigns genotypes to sub-populations and environments to regions. Such groupings can improve phenotype prediction.

For a second type of phenotype prediction, we consider CGMs. CGMs predict a target phenotype as a non-linear function of underlying intermediate phenotypes. The intermediate phenotypes are outcomes of functions defined on genotype dependent CGM parameters and classical environmental descriptors. While the intermediate phenotypes may still show some genotype by environment interaction, the genotype dependent CGM parameters should be consistent across environmental conditions. The CGM parameters are regressed on molecular marker information to allow phenotype prediction from molecular marker information and standard physiologically relevant environmental information.

Both LMMs and CGMs require extensive characterization of genotypes and environments. High-throughput technologies for genotyping and phenotyping provide new opportunities for upscaling phenotype prediction and increasing the response to selection in the breeding process.

3.1 Introduction

The target production area for most arable crops spans a range of environmental conditions. In the absence of diseases and pests (not considered in this review), local environmental conditions are a function of meteorological and soil variables on the one hand and management interventions on the other hand. These conditions will influence the phenotypic response of individual genotypes, and to some extent genotypes will create their 'own' environment, e.g. depending on how they use soil water across the season. The functional form by which environmental inputs are translated into phenotypes is sometimes referred to as the reaction norm (Woltereck 1909; Dobzhansky and Spassky 1963; Sarkar 1999; DeWitt and Scheiner 2004). Reaction norms depend both on environmental inputs and genetic factors. For a given (multi-locus) genotype, the reaction norm is the functional relationship between the phenotype and an environmental gradient, and is often linearised in some way. Modelling of the reaction norms for a set of genotypes is a central objective in many breeding and genetic studies. The prediction of the phenotypic response as a function of genetic and environmental factors is the basis for decisions that involve selection of superior genotypes for a defined environmental range (Hammer et al.2006; Chenu et al. 2011; Sadras et al. 2013).

Several important concepts in breeding and genetics have been defined in relation to the behaviour of reaction norms for a population of genotypes. Firstly, when

the reaction norms are non-constant, genotypes are said to show 'plasticity' (Bradshaw et al. 1965; DeWitt and Scheiner 2004; Sadras and Lawson 2011). Secondly, when the reaction norms for different genotypes are not parallel, this indicates the existence of genotype by environment interaction (GEI) (Finlay and Wilkinson 1963; van Eeuwijk et al. 2005). An extreme form of GEI is cross-over interaction, where the ranking of the genotypes varies with the environmental conditions (Baker 1988; Muir et al. 1992; Crossa et al. 2004). Another important concept in the context of the comparison of reaction norms is adaptation (Wright 1931, 1932; Finlay and Wilkinson 1963; Romagosa and Fox 1993; Cooper and Hammer 1996; Cooper 1999; Romagosa et al. 2013), i.e., some genotypes do better than other ones in a defined set of environmental conditions, the reaction norms of the adapted genotypes are then always above those of the less adapted. Finally, for a given genotype, 'stability' measures quantify the variation around the reaction norm (Lin and Binns 1988; Piepho 1998). So, while plasticity, GEI and adaptation refer to the expected response curve, which may be most simply thought of as the expectation in a linear regression model, stability refers to the variation around this expected response at a defined set of environmental conditions (Slafer and Kernich 1996; DeWitt and Scheiner 2004; van Eeuwijk et al. 2005; van Eeuwijk et al. 2010).

To select genotypes with superior average performance or a given degree of adaptation, predictions need to be made for the phenotype as a function of genotype and environment. These types of predictions occur at various stages in a breeding programme. In the early stages of breeding programmes, seed is limiting and large numbers of new genotypes produced as offspring from crosses between well-chosen parents are evaluated in one or a few trials, normally in small plots. For the earliest stages of a breeding programme, modelling of reaction norms is not possible and selection takes place on the mean performance. At intermediate stages, offspring populations are tested in a limited number of trials at various locations for one or a few years. In those cases when seed is still limiting, it is attractive to use partially replicated designs (Cullis et al. 2006; Smith et al. 2006) so that genotypes can be tested at a larger sample of environmental conditions. Selection can be done on the mean across trials, but there are also possibilities to select for adaptation. At the later stages, when there is sufficient seed for individual genotypes, a limited number of genotypes can be tested in a large number of trials, with again possibilities for selection on wide adaptation to a wide set of environments or narrow adaptation to a limited set of environments (Cooper et al. 2014). Simultaneously, at this stage selection on stability is possible.

When a population of genotypes is evaluated in multiple trials, reaction norms can be fitted to help in describing the observed data efficiently and to allow some form of selection on properties of the reaction norm. To evaluate the predictive quality of reaction norm models, special cross validation (CV) schemes have been proposed. In CV schemes, the data are subdivided in a training set, used to estimate model parameters, and a test set, used to assess prediction accuracy, which is the correlation between predicted and observed values (Meuwissen et al. 2001). For multiple environment data, various CV strategies have been proposed (Crossa et al. 2010, 2014; Burgueño et al. 2012; Heslot et al. 2012, 2013; Zhao et al. 2012). For a

transparent description of CV strategies, it is useful to introduce some notation. When genotypes were tested, evaluated or observed in at least one environment, we indicate this by the letter G. When this was not the case we use nG. For environments the same rule can be defined: E for observed environments, with at least one observed genotype, and nE for environments without observations (new environments). Specific combinations of genotype and environment can have been observed, GE, or not, nGE. Following this terminology, the set [G, E, GE] would indicate a genotype that was observed and an environment that was observed, while also the specific combination of genotype and environment was observed. The combination [G, E, nGE] indicates a genotype and environment that have been observed, but the specific combination of genotype and environment was not observed. This latter situation is typical for unbalanced genotype by environment data.

Figure 3.1 shows four scenarios that are relevant to prediction of phenotypes from genotypes and environments as well as to the calculation of accuracies and CV strategies. Scheme 1 pertains to situations in which both genotypes and environments were observed. Specific combinations of genotypes and environments may

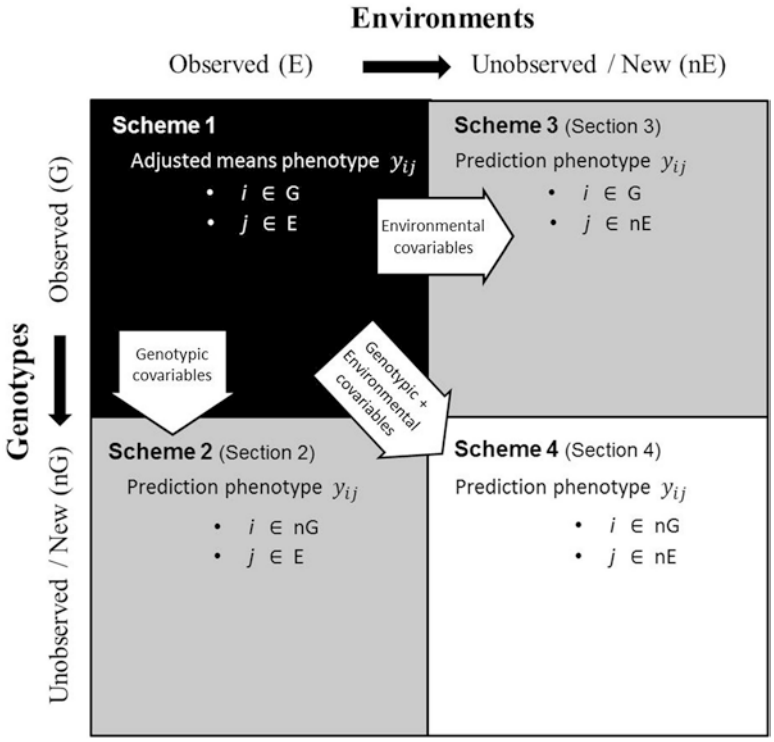

Fig. 3.1 Prediction scenarios, depending on whether genotypes were observed (G) or not observed (nG), and on whether environments were observed (E) or not observed (nE)

be present, [G, E, GE] or absent [G, E, nGE]. Phenotype predictions for Scheme 1 can be made by simple additive models. The Schemes 2, 3 and 4 are more interesting and we will concentrate on those. Potential strategies for assessment of accuracy in genomic prediction are predictions for new genotypes in observed environments [nG, E, nGE] (Scheme 2, Fig. 3.1); predictions for observed genotypes in new environments [G, nE, nGE] (Scheme 3, Fig. 3.1); and predictions for new genotypes in new environments [nG, nE, nGE] (Scheme 4, Fig. 3.1) (Utz et al. 2000; Calus and Veerkamp 2011; Burgueño et al. 2012; Schulz-Streeck et al. 2012; Guo et al. 2013; Crossa et al. 2014). Scheme 4 of CV obviously represents the strictest type of accuracy assessment. (For the notation, whenever nG or nE appears, necessarily nGE needs to appear as well, so for Schemes 2, 3 and 4, we can omit the specification nGE.)

To produce phenotype predictions for new genotypes (nG) from observed genotypes (G), it is essential to use statistical models that allow us to connect the new genotypes to the observed genotypes. The connections between nG and G can be achieved by the inclusion of explicit genotypic covariables in the statistical model, and/or by borrowing information via the correlation structure among genotypes, defined by their genetic similarities. Analogously, for predicting new environments, there needs to be a connection between nE and E via explicit environmental covariables and/or the correlation structure among environments. The latter correlation structure is an expression of environmental similarity as estimated from environmental characterizations.

In this chapter, we introduce linear mixed models (LMMs) as our default class of statistical prediction models. LMMs can be described as consisting of two parts: (1) a fixed part, corresponding to the mean; and (2) a random part defined by variances and covariances. Predictions in LMMs can be obtained via the fixed and the random part, although the statistical mechanism for prediction in those two cases is different. As an illustration, we provide an LMM for the phenotype of genotype i in environment j: $\underline{y}_{ij} = \mu_j + x_i\alpha_j + \beta_i z_j + \underline{GE}_{ij} + \underline{e}_{ij}$ (van Eeuwijk et al. 2010). The fixed part of this model is given by the expectation, or mean, for genotype i in environment j: $\mu_{ij} = \mu_j + x_i\alpha_j + \beta_i z_j$. Here μ_j is a fixed intercept (mean) for environment j, x_i is a genotypic covariable, for example a molecular marker, α_j is an environment specific slope corresponding to x_i. When x_i is a molecular marker, α_j is an environment specific quantitative trait locus (QTL) effect (Malosetti et al. 2004; Boer et al. 2007). For the environments, z_j is an environmental covariable, for example, a drought stress index, and β_i is a corresponding genotype specific slope, for example a genotype-specific sensitivity to drought stress.

For prediction via the fixed part, we use genotypic and/or environmental covariables as in classical regression (van Eeuwijk et al. 1996). Besides values for the covariable, x_i and z_j, prediction requires that we have estimates for the slopes, α_j and β_i. These can be obtained by fitting a model for the mean to training data, where we need to select suitable genotypic and/or environmental covariables. For prediction, we combine the estimated slopes in the training set with the values for genotypic and/or environmental covariables in the test set.

The random part of the model is determined by the terms $\underline{GE}_{ij}$ and $\underline{e}_{ij}$, the first term representing the (residual) genotypic effect of genotype *i* in environment *j*, the second term containing experimental (block) and measurement errors. (Random terms in model formulations are underlined.) The random terms are assumed to have a Gaussian distribution, with expectation zero and proper variance-covariance structures. The important random term for prediction purposes is $\underline{GE}_{ij}$. For this term, the correlations among genotypes on the one hand and the correlations among environments on the other hand determine the predictive properties of the LMM. Thus, for predictions via the random part of the LMM, correlations need to be estimated between observed and new genotypes (Scheme 2), observed and new environments (Scheme 3), or both (Scheme 4). Correlations among genotypes can be estimated from genotypic covariables, including molecular markers, and pedigree data, or a combination of genotypic covariables and pedigree. Correlations among environments follow from environmental covariables. Although important, we will largely ignore the error term $\underline{e}_{ij}$ in the remainder of this chapter. See Smith et al. (2001a) and Smith et al. (2005) for discussion on models for $\underline{e}_{ij}$.

The realization of the predictive potential of LMMs depends on the selection of genotypic covariables and environmental covariables, for the fixed part as well as for the random part. Physiological knowledge on genotypes and environments can help in the choice of covariables for inclusion in LMMs. For example, knowledge on the structure and use of crop growth models (CGMs) can help in the dissection of complex traits (Chapman et al. 2002b; Edmeades et al. 2004; Reynolds et al. 2009a), thereby suggesting genotypic and environmental covariables for inclusion in predictive LMMs. A CGM can suggest writing a complex target trait as a function of a set of simpler component traits and a set of environmental input variables (Yin et al. 2003, 2004; Chenu et al. 2008; Hammer et al. 2010). These component traits are traditionally related to physiological parameters in CGMs (see Chaps. 4, 5, 6, 7, 8, and 9 of this book). The CGM produces GEI as an emerging property of the interaction between the physiological parameters and the environmental information (Chapman et al. 2002a, 2008; Hammer et al. 2002, 2006, 2010). Interpreting the CGM as a function that transforms physiological parameters and environmental inputs into a complex trait, we can understand that when the CGM can be approximated by a linear function, the component traits may be entered as genotypic covariables and the environmental inputs as environmental covariables in an LMM for the complex trait.

In Sect. 3.2, we will discuss how statistical LMM models can be used to predict phenotypic responses for new genotypes in observed environments (Scheme 2; [nG, E, nGE]), observed genotypes in unobserved (new) environments (Scheme 3; [G, nE, nGE]), or new genotypes in new environments (Scheme 4; [nG, nE, nGE]). In Sect. 3.3, we will discuss the use of CGMs to predict the performance of genotypes for environments in which they were not tested. Section 3.4 will discuss the contribution of high throughput genotyping and phenotyping to models for phenotype prediction. Strategies to group genotypes and environments will also be discussed in this Section. We finish with some concluding remarks in Sect. 3.5.

3.2 Statistical Models to Predict Phenotypic Performance

Section 3.2.1 presents statistical models for predicting the phenotype of genotypes that were so far not tested in the environments for which we want to predict, although we do have information about these environments from phenotypic evaluations for other genotypes [nG, E, nGE], Scheme 2 in Fig. 3.1. The connection between observed genotypes (G) and not observed genotypes (nG) will come from explicit genotypic covariables and/or the genetic correlations among genotypes. Section 3.2.2 describes statistical models for predicting phenotypes in environments that were not used to test genotypes, although we do have phenotypic information about these genotypes in other environments [G, nE, nGE], Scheme 3 in Fig. 3.1. The connection between observed environments (E) and unobserved environment (nE), will result from the inclusion of explicit environmental covariables and/or the correlations among environments calculated on the basis of environmental characterizations. Section 3.2.3 discusses the most challenging prediction scenarios; predicting the phenotype of genotypes that were not tested so far, for environments that neither were tested [nG, nE, nGE], Scheme 4 in Fig. 3.1. Here, both explicit genotypic and environmental covariables are required for prediction.

3.2.1 Statistical Models to Predict Performance of Unobserved Genotypes in Observed Environments [nG, E, nGE]

Quantitative traits are determined by many loci, with allelic effects varying in magnitude. Specific genomic regions significantly associated with phenotypic variation may be identified as quantitative trait loci, or QTLs (see Chap. 1 of this book by Baldazzi et al.). Besides QTLs, or instead thereof, many other loci with small additive effects (polygenic effects) can contribute to phenotypic variation. None of these loci with small effects might by itself have an important phenotypic effect, but these loci together can still make a sizeable contribution to the phenotype. Model 3.1, includes loci with relatively large quantitative effects (QTLs) together with loci that have small effects.

$$\underline{y}_{ij}^{t} = \mu_j + \sum_{q=1}^{Q} x_{iq}\alpha_{jq} + \underline{G}_{ij} + \underline{e}_{ij} \tag{3.1}$$

In the multi-environment Model 3.1, y_{ij}^{t} represents the target trait, t, (for example, yield) of genotype i in environment j, μ_j is a fixed intercept term for each environment, x_{iq} is a genotypic covariable that represents DNA information of genotype i at QTL position q, and α_{jq} is the additive effect of the fixed QTL q in environment j. $\underline{G}_{ij}$ represents the residual genetic effect (polygenic effects) for genotype i in environment j. The matrix with elements $\underline{G}_{ij}$, $\left\{\underline{G}_{ij}\right\}$, has a multivariate normal distribution with zero mean, 0, and, as we will see later, a highly structured variance-covariance

matrix $\boldsymbol{\Sigma}$; $\left\{\underline{G}_{ij}\right\} \sim MVN\left(0, \Sigma\right)$. (For notational simplicity, we will omit the dimensions of the various matrices.) $\boldsymbol{\Sigma}$ defines the genetic variances and covariance for any two pairs of observations, $\underline{y}_{ij}^{t}$ and $\underline{y}_{i'j'}^{t}$, and depends on the genetic and environmental similarities of the two genotypes, *i* and *i'*, and the two environments *j* and *j'*. The term $\underline{e}_{ij}$ stands for a non-genetic residual, $\left\{\underline{e}_{ij}\right\} \sim MVN\left(0, \boldsymbol{R}\right)$, with $\boldsymbol{R}$ often allowing for specific residual variances per environment.

A simplification of Model 3.1 omits the genetic residual, $\underline{G}_{ij}$, and is appropriate when QTLs account for all of the genetic variation:

$$\underline{y}_{ij}^{t} = \mu + \sum_{q=1}^{Q} x_{iq}\alpha_{jq} + \underline{e}_{ij} \tag{3.2}$$

When Model 3.2 fits the data well, the performance of the unobserved genotype *i* in environment *j* can be predicted as;

$$\hat{\underline{y}}_{ij}^{t} = \hat{\underline{\mu}}_{j} + \sum_{q=1}^{Q} x_{iq}\hat{\underline{\alpha}}_{jq}$$

Compared with single-environment QTL models, multi-environment QTL models like Model 3.1 or Model 3.2 are more powerful in picking up QTLs and generally explain a larger amount of the genetic variance (Piepho 2000; Piepho and Möhring 2005; Mathews et al. 2008; Alimi et al. 2013). It has been shown that jointly considering multivariate phenotypes (i.e., the phenotype in multiple environments) allows a substantially greater separation between genotype classes than when considering univariate phenotypes (i.e., phenotype in a single environment) (Stephens 2013).

Another simplification of Model 3.1 occurs when we assume that there are no large discrete genetic effects in the form of QTLs that drive phenotypic differences, but that genetic effects are exclusively of a polygenic nature. A prediction model that generalizes the single environment genomic best linear unbiased prediction (G-BLUP) approach of (Meuwissen et al. 2001) to multi-environment prediction can be defined as:

$$\underline{y}_{ij}^{t} = \mu_{j} + \underline{G}_{ij} + \underline{e}_{ij} \tag{3.3}$$

In Model 3.3, the distribution of the polygenic effects $\underline{G}_{ij}$ is $\left\{\underline{G}_{ij}\right\} \sim MVN\left(0, \Sigma\right)$. Since $\boldsymbol{\Sigma}$ is a function of the genetic and environment similarities, the larger the similarity of unobserved genotypes with observed genotypes, and the larger the similarity of observed environments with unobserved environments, the more information is available for phenotype prediction, and the higher is the prediction accuracy (Crossa et al. 2006; Albrecht et al. 2014). Analogous to the classical partitioning of genetic and environmental effects, the covariance matrix $\boldsymbol{\Sigma}$ can be partitioned into a 'genotypic' variance-covariance matrix ($\boldsymbol{\Sigma}^{G}$), and an 'environmental' variance-covariance matrix ($\boldsymbol{\Sigma}^{E}$), such that $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{\boldsymbol{G}} \otimes \boldsymbol{\Sigma}^{E}$, i.e., the Kronecker product of the

genotypic variance-covariance matrix and the environmental variance-covariance matrix (West et al. 2006; Smith et al. 2005). It is important to realize that although $\boldsymbol{\Sigma}^E$ is called an 'environmental' variance-covariance matrix, $\boldsymbol{\Sigma}^E$ reflects genetic correlations among environments, and so plays a role in forming predictions in the multi-environment context. Examples of commonly used models for these two covariance matrices are given below.

$\boldsymbol{\Sigma}^G$ can be modelled as $\boldsymbol{\Sigma}^G = \boldsymbol{A}$, where $\boldsymbol{A}$ corresponds to the expected additive relationship matrix calculated from the coefficients of coancestry estimated from the pedigree, or to the realized additive relationship matrix estimated from molecular markers (Piepho et al. 2008). If the one step prediction with statistical models uses pedigree information, $\underline{G}_{ij}$ is commonly called "breeding value" (Falconer and Mackay 1996; Piepho et al. 2008). On the other hand, if the prediction uses molecular marker information, it is called "genomic estimated breeding value" (Burgueño et al. 2012; Piepho 2009).

In the multi-environment context, genotypic variances tend to change across environments with consequent changes in genotypic correlations for pairs of these environments. A flexible variance-covariance structure across environments $\boldsymbol{\Sigma}^E$, is required to achieve higher prediction accuracies. One flexible and parsimonious model for variances and covariances/correlations across environments is the factor analytic model (FA) (Table 3.1) (Smith et al. 2001a, 2005; Mathews et al. 2008).

The decision about when it is convenient to use Models 3.1, 3.2, or 3.3 depends on the genetic architecture of the target trait. If the trait is regulated by a few QTLs

Table 3.1 Variance-covariance models for the environmental covariance (Σ^E), ordered by increasing number of parameters. For simplicity, these examples assume three environments (m=3)

Name	Number of parameters	Structure
Identity	1	$\begin{bmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix}$
Compound symmetry	2	$\begin{bmatrix} \sigma^2+\varphi & \varphi & \varphi \\ \varphi & \sigma^2+\varphi & \varphi \\ \varphi & \varphi & \sigma^2+\varphi \end{bmatrix}$
Factor analytic, order 1	2 m	$\begin{bmatrix} \lambda_1^2+\psi_1 & \lambda_1\lambda_2 & \lambda_1\lambda_3 \\ \lambda_2\lambda_1 & \lambda_2^2+\psi_2 & \lambda_2\lambda_3 \\ \lambda_3\lambda_1 & \lambda_3\lambda_2 & \lambda_3^2+\psi_3 \end{bmatrix}$
Unstructured	m(m+1)/2	$\begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{bmatrix}$

with large effects, a QTL model (Model 3.2) might provide the largest prediction accuracy. On the other hand, traits like grain yield, which are regulated by many genes with small effects might not show any significant QTL that can be included in Model 3.2. In this case, Model 3.3, whose predictions we will call GE-BLUPs because they can account for GEI, should integrate the large number of small additive effects into a multi-environment prediction model. For the intermediate case when traits have a few QTLs with large effects, and many other loci with very small additive effects, Model 3.1 is adequate. Bernardo (2014) suggested that it is convenient to consider QTLs (or genes) as fixed effects when they account for more than 10 % of the genetic variance. The simulations made by Bernardo (2014) show that the most adequate model depends on the genetic architecture of the trait, i.e., on the number of QTLs and the magnitudes of the QTL effects.

3.2.2 *Statistical Models to Predict Performance of Observed Genotypes in New Environments*

After genotypes have been phenotyped in some environments, it can be useful to predict their performance in other environments that were not used for evaluation. New environments could, for example, include future trials at known locations, which implies that none of the genotypes were observed in that environment yet [G, nE, nGE]. Thus, the correlation between observed environments and the predicted environments cannot be estimated from phenotypic data, or direct observations on the complex trait. In this case, we may use environmental covariables, like meteorological, soil or management covariables, as predictors in models for the mean or define correlations between environments in models for the variance-covariance structure.

Models for the mean that can be used to predict phenotypes in unobserved environments usually correspond to factorial regression models that incorporate environmental covariables. These models explicitly estimate the sensitivity of the QTL to environmental covariables (Model 3.4) (Campbell et al. 2004; Boer et al. 2007; Laperche et al. 2007; Malosetti et al. 2013; Romagosa et al. 2013). Hence, model parameters can have biological interpretation.

$$\underline{y}_{ij} = \mu_j + \sum_{q=1}^{Q} x_{iq}\left(\gamma_q + \delta_q z_j\right) + \underline{G}_{ij} + \underline{e}_{ij} \tag{3.4}$$

In Model 3.4, the additive effects (α_{jq}) of the fixed QTL q in environment j of Model 3.1 are replaced by a regression formulation, $\left(\gamma_q + \delta_q z_j\right)$, in which the effect of QTL q is a function of the environmental covariable z_j, and so changes over environments. When the covariable z_j is centered, the intercept term, γ_q, corresponds to the effect of the QTL in the average environment, while the slope δ_q corresponds to the sensitivity of the QTL q to the environmental covariable z_j. Although Model 3.4 does not explicitly restrict the environmental covariables to a particular range, it should

be considered that crops respond differently to covariables in the environmental extremes (e.g., too cold or too warm). So, the sensitivity of the genotype to the environmental covariables cannot be assumed constant outside the range of environments in which δ_q was estimated. A second issue that needs to be taken into account is that models like Model 3.4 do not make explicit in which phenological stage the environmental covariable is considered. Since the sensitivity of a crop to the environment varies throughout the development, environmental covariables included in the prediction model need to coincide with the developmental timing used to estimate the sensitivity.

For example, Boer et al. (2007) analysed grain yield and grain moisture for F5 maize testcross progenies evaluated across 12 environments in the U.S. corn belt. Since QTLs did not have a constant effect across environments (QTL by environment interaction), QTL effects were modelled conditional on longitude and year, both consequences of temperature differences during critical stages of the development. This factorial regression model allows prediction of yield and moisture at any location provided that temperatures during specific developmental stages are contained within those of the observed environments.

A second example is shown by Malosetti et al. (2004), who identified QTLs conferring differential sensitivity of grain yield to temperature during heading in a double haploid barley population. In a model like Model 3.4, the average daily temperature range during heading was the most important environmental covariable explaining differential QTL expression, i.e., the QTL allele from the parental line Steptoe conferred an extra grain yield of 0.112 t ha^{-1} for each extra degree Celsius during heading. Hence, yield could be predicted for unobserved environments if the average temperature for such environments was available. In that sense, Model 3.4 is closer to CGM than Model 3.1 because Model 3.4 explicitly represents environments on a continuous scale.

The second way to use environmental information for prediction is using environmental covariables to estimate similarities (covariances) among environments, analogous to the way molecular markers are used to characterize similarity among genotypes. If environmental covariables are considered as an indicator of environmental similarity, they can be used to estimate the environmental variance-covariance matrix in Model 3.3. Hence, $\boldsymbol{\Sigma}^E = \boldsymbol{\Omega}$, where $\boldsymbol{\Omega}$ is the variance-covariance matrix that accounts for the similarity in environmental conditions. The larger the covariance between observed and unobserved environments, the more information can be shared to make the predictions. The genotypic covariance $\boldsymbol{\Sigma}^G$ can be modelled as explained in Sect. 3.2.1 by imposing an additive relationship matrix to define $\boldsymbol{\Sigma}^G = \boldsymbol{A}$, where $\boldsymbol{A}$ can be estimated from the pedigree and/or from molecular markers.

Using multiple climatic variables to model the environmental covariance, as proposed by Jarquín et al. (2013) shows promise as a tool to predict genotypic performance in unobserved environments. However, many environmental covariables are correlated and not all need to be included in the model. Mechanistic CGMs such as APSIM have shown to be a good integrative tool to select subsets of variables that characterize environmental similarity (Chapman2008).

3.2.3 Statistical Models to Predict Performance of Unobserved Genotypes in New Environments

Section 3.2.1 presented models that used genotypic covariables to predict the phenotype on unobserved genotypes. Section 3.2.2 described how environmental covariables can be used in factorial regression models for prediction, and how to estimate the environmental covariance of a random term, necessary for prediction along the random part of an LMM. This Sect. 3.2.3 will combine both situations, aiming to predict the phenotype of genotypes that have not been tested yet for environments that have not been used for evaluation.

When predicting unobserved genotypes in new environments, both genotypic and environmental covariables are needed. In factorial regression-type of models, prediction of unobserved genotypes is possible, provided that the additive effects of each QTL allele can be estimated from the tested genotypes. The phenotypes of unobserved genotypes can also be predicted in new environments, provided that the sensitivity of the QTL effects along an environmental gradient (e.g., temperature), can be estimated from observed environments. In the example of Malosetti et al. (2004) presented in Sect. 3.2.2, phenotype prediction is possible for any environment provided the temperature remains within the range used to estimate the QTL sensitivity to temperature.

In models that entirely rely on the use of the variance-covariance structures imposed on genotypes and environments, prediction of unobserved genotypes in new environments is possible via the reconstruction of the full covariance matrix $\boldsymbol{\Sigma}$ from its components, $\boldsymbol{\Sigma}^G$ and $\boldsymbol{\Sigma}^E$. For the genotypic part, this runs via explicit pedigree information or information from genotypic covariables (molecular markers), while for the environmental part correlations between environments can be estimated from environmental characterization (temperature, precipitation, soil characteristics, etc.). Note that while in Sect. 3.2.1, $\boldsymbol{\Sigma}^G$ was calculated from genotypic covariables, and $\boldsymbol{\Sigma}^E$ was estimated from the phenotypic data on the target trait, here both $\boldsymbol{\Sigma}^G$ and $\boldsymbol{\Sigma}^E$ are estimated from explicit covariables.

3.3 Crop Growth Models to Predict Genotypic Performance

The algorithms in a CGM predict the target trait (e.g., grain yield) as a non-linear combination of underlying intermediate phenotypes (also commonly called “components”, e.g., biomass), which are calculated indirectly from a set of inputs to the CGM that typically comprise environment (soil, weather, and nutrients) data and CGM parameters derived from prior experimentation. GEI in the target trait is then a consequence of the interactions between the intermediate phenotypes (Chapman

et al. 2003; Tardieu 2003; Tardieu et al. 2005; Chenu et al. 2009; Makumburage et al. 2013).

Considering the CGM in reverse, we can state that the value of the target trait is able to be 'dissected' into these intermediate phenotypes (See Chap. 7 of this book by Hammer et al.). Although these intermediate phenotypes are likely to show less GEI than the target trait, they still correspond to an integration of genotypic responses to environmental conditions (e.g., they may show GEI). Ideally, a complete dissection of the target trait would comprise of a set of CGM input parameters that depend only on the genotype (for example, a genotypic sensitivity of development rate to the air temperature), and environmental covariables (Model 3.4), i.e., CGM parameters that do not show GEI (Slafer 2003; Yin et al. 2003; Bertin et al. 2010; Alam et al. 2014). The target trait for genotype *i* in environment *j* can be written as a function of CGM parameters and environmental inputs as follows:

$$\underline{y}_{ij}^{t} = \int f(\underline{\boldsymbol{y}}_{i}^{P}; \boldsymbol{z}_{j})\, dt + \underline{e}_{ij} \tag{3.5}$$

In Model 3.5, $\underline{y}_{ij}^{t}$ represents the target trait for genotype *i* in environment *j*, which is modelled as a function of multiple CGM parameters, $\underline{\boldsymbol{y}}_{i}^{P}$ (with *P* for parameter in the superscript), and multiple environmental inputs, $\boldsymbol{z}_{j}$, integrated over time (Fig. 3.2). The function *f* (;) embodies the algorithms that transform CGM parameters into intermediate phenotypes as well as the interactions between the intermediate phenotypes that lead to the target trait.

A commonly-studied CGM is APSIM, which currently has modules for several crops, e.g., wheat, canola, sorghum (Keating et al. 2003; Holzworth et al. 2014). In the case of APSIM-Wheat, growth (biomass accumulation) and development (phenological events, the functionality of plant structures or appearance of new structures) are calculated on a daily basis (Wang et al. 2002). The final phenotype (e.g., grain yield) is calculated as a function of a series of intermediate phenotypes. Examples of intermediate phenotypes are biomass, grain number and radiation interception on any given day or accumulated to a given day (Fig. 3.2). Intermediate phenotypes depend on CGM parameters that are genetically determined, and which modulate the phenotypic response to environmental covariables. Examples of CGM parameters are vernalization requirement and sensitivity to photoperiod, which are regulated by the *VRN* and the *PPD* alleles (Zheng et al. 2013).

CGM parameters, $\underline{\boldsymbol{y}}_{i}^{P}$, for phenotyped genotypes can be directly observed, estimated or calculated from the phenotypic measurements. However, given that CGM parameters depend on the genotype, they can also be predicted from genotypic covariables, i.e., molecular marker information. When we can identify the genetic basis of physiological parameters in terms of underlying QTLs, or, equivalently, when we can predict the physiological parameters from marker information, we can effectively predict the target trait from marker information and environmental inputs provided the intermediate traits and their interactions have been correctly identified

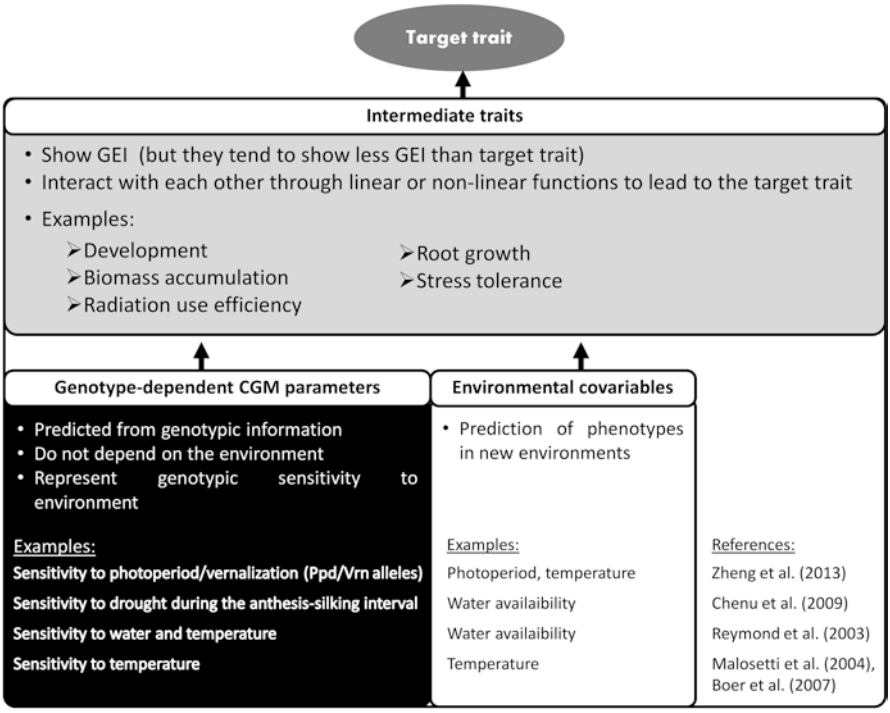

Fig. 3.2 Representation of the information flow in a CGM. The *black box* corresponds to CGM parameters that are dependent on the genotype, the *white box* represents environmental covariables and the *grey box* represents the intermediate phenotypes. Examples of different crops/traits/models are given for each category

and implemented in the CGM. Hence, predicted CGM parameters enable to predict the phenotype of genotypes that have not been observed yet. The prediction for individual CGM parameters ($\underline{y}_i^P$) would look like Model 3.6:

$$\underline{y}_i^P = \mu + \sum_{q=1}^{Q} x_{iq}\alpha_q + \underline{G}_i + \underline{e}_i \tag{3.6}$$

Like Model 3.1, Model 3.6 can be modified to include (i) only the QTLs, in a QTL model (Model 3.7) or (ii) only the polygenic effects ($\underline{G}_i$) , in a genomic prediction model with the random effects $\underline{G}_i$ being structured by a genetic relationship matrix (Model 3.8).

$$\underline{y}_i^P = \mu + \sum_{q=1}^{Q} x_{iq}\alpha_q + \underline{e}_i \tag{3.7}$$

$$\underline{y}_i^P = \mu + \underline{G}_i + \underline{e}_i \tag{3.8}$$

If more than one CGM parameter is to be predicted from molecular markers and/or pedigree information, Models 3.6, 3.7, and 3.8 could also be expanded to a multi-trait prediction model that takes into account possible correlations among the CGM parameters, in a model that is similar to the multi-environment Model 3.1. Modelling traits simultaneously allows to gain power for QTL detection and to detect QTLs with pleiotropic effects (Alimi et al. 2013; Stephens 2013).

Predictions for multiple CGM parameters, $\hat{\underline{y}}_i^P$, can be used as input in Model 3.5 to calculate intermediate phenotypes, and produce the prediction for the target trait, $\hat{\underline{\mu}}_{ij}^t$, in Model 3.9.

$$\hat{\underline{\mu}}_{ij}^t = \int f(\hat{\underline{y}}_i^P; \boldsymbol{z}_j)\, dt \tag{3.9}$$

In Model 3.9, the prediction accuracy of the target trait depends on the accuracy of the prediction of each of the components, and on the ability of the functions that transform CGM parameters into intermediate phenotypes to correctly describe the processes leading to the target trait.

CGMs with known/predicted genotypic parameters are a potentially useful tool to understand which traits can be advantageous in a given environment, and also to identify management practices that contribute to improved crop productivity (Yin et al. 2004; Hammer et al. 2006; Reynolds et al. 2009b; Harrison et al. 2014). In the context of adaptation to climate change, Zheng et al. (2012) modelled how phenology of current wheat varieties would influence their adaptation to future environments, which are expected to show different CO_2 and precipitation levels. In their second paper, Zheng et al. (2013) demonstrated that the flowering time of spring wheat genotypes can be modelled using the composition of their *VRN1* and *Ppd-D1* alleles together with responses derived from a single experiment with four environments: +/− treatments for vernalisation and extended photoperiod. Allelic combinations of loci *Vrn-A1*, *Vrn-B1*, *Vrn-D1*, and *Ppd-D1* were used to predict APSIM-wheat parameters of a population of genotypes. From a single experiment (replicated in 2 years), they validated the model with more than 250 wheat genotypes across the entire Australian wheat belt, and were able to simulate flowering time for any weather records in the wheat belt. These conclusions can be useful to guide breeders in the process of determining which alleles should be considered in the selection process.

Bogard et al. (2014) extended this approach further to model the drivers of flowering time in winter wheat as functions of major genes as well as SNPs derived from association mapping, i.e., allowing prediction of unknown genotypes (but with known genes and SNPs) in new environments. In both Zheng et al. (2013) and

Bogard et al. (2014), the predictions for heading date using the gene-based predicted parameters corresponded well with the observed dates to heading. Attributes that contributed to a successful phenotype prediction were (i) a well-defined CGM for heading date (Slafer and Rawson 1994), (ii) a well-defined set of environmental covariables with corresponding CGM parameters, and (iii) a genetic basis of the CGM parameters (Snape et al. 2001).

The same approach has also shown to be successful for other more complex and less heritable traits such as grain yield under drought. For example, Chenu et al. (2009) used APSIM to model the impact of QTLs controlling the intermediate traits leaf and silk elongation on maize grain yield. The intercept and slope of these intermediate traits in response to meristem temperature, evaporative demand and soil water deficit were genotype-dependent (Reymond et al. 2003, 2004).

Unfortunately, the identification of CGM parameters is sometimes less straightforward for complex traits like grain yield. Yin et al. (2000) showed an example in barley with a successful estimation of QTL effects for the CGM parameters, but with a poor prediction of grain yield. The correlation between the observed CGM parameters, i.e. phenotype of CGM parameters, and the QTL predictions of the same parameters was high. However, the correlation between yield predictions of the CGM, whether phenotype based or QTL-prediction based, and observed yield was not high. The cause of the poor predictions did not reside in the fact that the CGM parameters were replaced by predicted parameters from the QTL model, but in the fact that the CGM was unable to predict yield from its component traits. Similar work has been recently reported by Gu et al. (2014) on grain yield of rice crop, using a new CGM, which gave more promising results. However, efforts to improve CGM for predicting complex traits like grain yield are still strongly needed.

The example from Yin et al. (2000) shows that although the integrated statistical and CGM modelling allows for a larger flexibility, it might result in more complex and fragile models, because the approach can break down at the level of the estimation of the CGM parameters and at the level of the integration of these CGM parameters to calculate the intermediate phenotypes. However, even if the CGMs are not fully able to predict the target trait, it is valuable to develop models of intermediate traits as well as for yield per se. Breeders can still be interested to recombine lines with high levels of proven intermediate traits with the expectation that these should on average result in better yield when further crossing and selection is done, i.e., because the selection on intermediate traits should already have improved part of the physiological adaptation pathway (Cooper et al. 2014). If breeders select mainly on yield per se, then it may be less likely that selected genotypes will also have high radiation use efficiency (RUE) or transpiration efficiency (TE) or traits for which genetic variation was not expressed in the given selection environment.

The examples of Zheng et al. (2013), Bogard et al. (2014), and Chenu et al. (2009) show that CGMs are a tool to integrate complex information from the genotypic, organ, and crop level (see also examples reviewed in Chap. 9 of this book by Yin et al.). Dissection of a target trait into component traits at different levels of biological organisation allows phenotype prediction for the target trait in the face of genotype by environment and QTL by environment interactions for that same trait.

Hence, the combined approach of statistical QTL modelling and CGM is an alternative to model complex GEI interactions (Yin et al. 2004).

3.4 Further Considerations

3.4.1 Classification of Environments

Sections 3.2.2 and 3.2.3 presented models to predict the performance of genotypes in new environments ([G, nE, nGE] or [nG, nE, nGE]). However, if there are repeatable patterns that allow to classify environments, these patterns might help to reduce the complexity of $\boldsymbol{\Sigma}^E$ and thereby improve the accuracy of prediction.

One example of repeatable patterns that often justifies to group environments is the presence of regions. Here, we understand 'regions' (or mega-environments) as a group of locations where genotypes perform consistently across years (Bull et al. 1992; Gauch and Zobel 1997; Basford and Cooper1998; Yan et al. 2000). Environments inside the same region are expected to be more homogeneous in terms of genotypic ranking, i.e., less GEI inside the regions (e.g. Atlin et al. 2000; Burgueño et al. 2008). In dryland production areas, other groupings may relate to characteristics of the soil (shallow/deep, low/high water holding capacity) and the management of the crop (sowing date, row spacing arrangement, etc.). De la Vega and Chapman (2010) showed how multiple component traits related to yield for a complex set of mega-environments in Argentina.

If locations can be grouped into regions, it is generally convenient to breed for specific adaptation to those regions, instead of broad adaptation across regions (Atlin et al. 2000, 2011). In this case, predictions can be produced for the whole of a region, or for new environments within a region. Precision of yield estimates might still benefit from the information of neighbouring regions by means of the covariance structure in a mixed model (Piepho and Möhring 2005; Kleinknecht et al. 2013).

When phenotypes are not available for all the locations of interest, environmental covariables can also be used to classify environments, and reduce the complexity of $\boldsymbol{\Sigma}^E$. Classifying environments into regions on the basis of environmental similarity, potentially allows to (i) predict new environments (as discussed in Sects. 3.2.2 and 3.2.3), and also (ii) define the target population of environments, where a particular genotype is to be grown (Chapman et al. 2000a; Hammer et al. 2002; Chenu et al. 2011). CGMs are a powerful tool to identify relevant environmental factors (Chapman 2008; Messina et al. 2011), and the periods when the crop is most sensitive to those factors (Chenu et al. 2013). For example, considering drought seasonal patterns could give a better indication of the environment types, instead of the total rainfall per year (Chapman et al. 2000a, b).

A further application of explicit environmental characterization is to weight environments based on their expected relevance for future years (Podlich et al.

1999). This means that environmental conditions that are more likely to occur receive more weight when doing the predictions, compared to less likely environmental conditions.

3.4.2 Population Structure

Sections 3.2 and 3.3 discussed different models to predict phenotypes of unobserved genotypes using molecular marker information. In those sections, $\boldsymbol{\Sigma}^G$ had the structure of the genomic relationship matrix, without explicitly specifying sub-populations. However, genetic relatedness between training and test sets largely influences prediction accuracy (Windhausen et al. 2012; Riedelsheimer et al. 2013). Hence, when there is strong population structure, it is necessary to define whether prediction will be done among or within populations. When predictions are limited to specific sub-populations, accuracy is commonly larger than when predicting across sub-populations, or when correcting for population structure (Daetwyler et al. 2012; Guo et al. 2014).

Methods to consider population structure in the model for genomic prediction can be based on the incorporation of the eigenvectors of the genotype by molecular marker data matrix (Patterson et al. 2006; Janss et al. 2012). Another option is to consider population structure in the design of the cross-validation scheme, for example by a stratified cross-validation design conditional on known population structure to ensure that each sub-population is equally represented in the training and validation sets (Albrecht et al. 2014; Guo et al. 2014).

3.4.3 Next Generation Sequencing

With the recent development of next generation sequencing technologies, genotyping costs have been largely reduced, allowing improving the genotypic characterization of important crops as barley, wheat and potato (Poland et al. 2012b; Uitdewilligen et al. 2013). In sequence-based genotyping approaches, marker discovery and genotyping are completed at the same time, allowing for faster genotyping processes (Poland and Rife 2012). The shorter time needed is thanks to the combination of restriction enzymes, sequencing, imaging, and genome alignment and assembly methods (Metzker 2010; Elshire et al. 2011).

These technologies permit the genotyping of larger populations of plants with higher marker density and increased mapping resolution (Varshney et al. 2014). Larger marker density increases the chances of including causal loci that otherwise would not have been considered in models for phenotype prediction (Spindel et al. 2013). More loci in the model means increased genomic prediction accuracy (Poland et al. 2012a). However, models for phenotype prediction have diminishing

returns on additional markers once the point of "marker saturation" has been reached, which depends on the genetic diversity of the population (Jannink et al. 2010; Heffner et al. 2011; Poland et al. 2012a).

Other questions regarding larger numbers of markers that remain not fully answered are: (i) how imputation of missing genotype data or haplotype inferences may affect prediction accuracies when genotyping by sequencing is used (Crossa et al. 2013), (ii) how to reduce the computational time needed because of the large number of markers (Verbyla and Cullis 2012), and (iii) how to improve model diagnostics, distinguishing between loci with large effects, and loci with smaller effects (Bernardo 2014).

3.4.4 High-Throughput Phenotyping to Input to Models for Phenotype Prediction

Mixed models and CGM discussed in Sects. 3.2 and 3.3 are promising tools for phenotype prediction. However, these models require the phenotyping of multiple genotypes, traits and environments. With the reduction of genotyping costs, evaluating the populations phenotypically has become the limiting factor (Cobb et al. 2013).

High-throughput phenotyping platforms can either measure the target trait directly, or measure one or more traits that are correlated with the target trait. The use of CGMs allows estimation of hard-to-measure traits such as seasonal water use, given inputs of leaf area over time and canopy thermal characteristics, for example. Correlated traits measured by high-throughput phenotyping platforms can be used as inputs in models like Model 3.1. To do so, traits must: (i) have high genetic correlation with the target trait in the target environment, (ii) be less affected by environment (have a larger heritability) than the target trait, and (iii) provide an easy and reliable measurement, which is less expensive than the target trait itself (Bänziger 2000; Araus et al. 2008; Prasanna et al. 2013). When measuring correlated traits, high-throughput phenotyping platforms could be particularly useful for obtaining detailed non-destructive measurements of plant characteristics that collectively provide reliable estimates of trait phenotypes (Cabrera-Bosquet et al. 2012; Prasanna et al. 2013; Cooper et al. 2014).

High-throughput phenotyping platforms are commonly used under two scenarios: (i) precise phenotyping under controlled environments that aim at representing different levels of environmental quality, and (ii) phenotyping in environments that correspond to a sample of environments in the field. The main advantage of controlled environments is that screening protocols can be more easily standardized, ensuring that plants are exposed to fairly reliable levels of stress. Hence, controlled environments offer the stability to search for attractive phenotypes or genotypes in a specific type of environment, e.g. drought stress (Cobb et al. 2013; Passioura 2012).

Growth under controlled conditions is usually different from that under field conditions. Hence, high-throughput phenotyping platforms in controlled environments might not lead to the identification of important yield-determining processes and promising genotypes in the field (Passioura 2012). This limits the application for phenotyping to specific stages of the crop (e.g., early vigour), or to traits that are correlated with the target trait (e.g., carbon isotope discrimination as an indicator of water use efficiency (Passioura 2012; Prasanna et al. 2013).

Popular high-throughput phenotyping techniques are those based on spectral technologies or remote sensing, such as near infrared spectroscopy (NIRS), or image analysis. These techniques are a powerful tool that can provide information about multiple traits from only one or few images, and can be applied in controlled conditions as well as in field trials.

One example of how phenotypes obtained by image analysis can be included in phenotype prediction is shown by van der Heijden et al. (2012). Here, QTLs for leaf area were identified from the 3D representation of the plant canopy reconstructed from stereo images. The QTLs for leaf area from the image analysis agreed with the QTLs detected when using manually measured leaf areas, showing the potential of stereo images to characterize phenotypically breeding populations.

Image analysis introduces potentially larger measurement errors than conventional measurement techniques. For that reason, image information should be first carefully selected with the aid of statistical and physiological knowledge, in an automatized and standardized fashion, before incorporating it in the genetic analysis (Eberius and Lima-Guerra 2009; Hartmann et al. 2011). Hence, models accounting separately for the measurement error and for the experimental (plot) error should be considered (Smith et al. 2001b).

3.5 Concluding Remarks

This chapter discussed several approaches that aim at predicting the phenotype in a multi-environment context. These approaches ranged from pure statistical models and pure CGMs, to a combination of both types of models. Special attention was given to different prediction scenarios; unobserved genotypes, new environments, and the combination of both. How prediction accuracy can profit from the large availability of environmental and genotypic information was also discussed, aiming at integrating physiological and statistical knowledge. Phenotypic and genomic data start to become abundant. The challenge for better phenotype prediction and more effective selection lies in more sophisticated procedures for selection of genotypic and environmental covariables in models for phenotype prediction, separating the signal from the noise.

Acknowledgements Daniela Bustos-Korts thanks Becas Chile (CONICYT) for the financial support in form of a PhD scholarship. Marcos Malosetti and Fred van Eeuwijk worked on this chapter as part of a project financed by the Generation Challenge Program – Integrated Breeding Platform (https://www.integratedbreeding.net/).

References

Alam MM, Mace ES, van Oosterom EJ, Cruickshank A, Hunt CH, Hammer GL, Jordan DR (2014) QTL analysis in multiple sorghum populations facilitates the dissection of the genetic and physiological control of tillering. TAG Theor App Genet (Theor angew Genet) 127(10):2253–2266. doi:10.1007/s00122-014-2377-9

Albrecht T, Auinger H-J, Wimmer V, Ogutu J, Knaak C, Ouzunova M, Piepho H-P, Schön C-C (2014) Genome-based prediction of maize hybrid performance across genetic groups, testers, locations, and years. TAG Theor App Genet (Theor angew Genet) 127:1375–1386. doi:10.1007/s00122-014-2305-z

Alimi NA, Bink MCAM, Dieleman JA, Magán JJ, Wubs AM, Palloix A, van Eeuwijk FA (2013) Multi-trait and multi-environment QTL analyses of yield and a set of physiological traits in pepper. Theor Appl Genet 126(10):2597–2625. doi:10.1007/s00122-013-2160-3

Araus JL, Slafer GA, Royo C, Serret MD (2008) Breeding for yield potential and stress adaptation in cereals. Crit Rev Plant Sci 27(6):377–412. doi:10.1080/07352680802467736

Atlin GN, Baker RJ, McRae KB, Lu X (2000) Selection response in subdivided target regions. Crop Sci 40(1):7–13. doi:10.2135/cropsci2000.4017

Atlin GN, Kleinknecht K, Singh GP, Piepho HP (2011) Managing genotype x environment interaction in plant breeding programs: a selection theory approach. J Indian Soc Agric Stat 65(2):237–247

Baker RJ (1988) Tests for crossover genotype-environmental interactions. Can J Plant Sci 68(2):405–410. doi:10.4141/cjps88-051

Bänziger M (2000) Breeding for drought and nitrogen stress tolerance in maize: from theory to practice. Cimmyt, Mexico

Basford KE, Cooper M (1998) Genotype x environment interactions and some considerations of their implications for wheat breeding in Australia. Aust J Agric Res 49(2):153–174. doi:10.1071/A97035

Bernardo R (2014) Genomewide selection when major genes are known. Crop Sci 54(1):68–75. doi:10.2135/cropsci2013.05.0315

Bertin N, Martre P, Génard 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. J Exp Bot 61(4):955–967. doi:10.1093/jxb/erp377

Boer MP, Wright D, Feng L, Podlich DW, Luo L, Cooper M, van Eeuwijk FA (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(3):1801–1813. doi:10.1534/genetics.107.071068

Bogard M, Ravel C, Paux E, Bordes J, Balfourier F, Chapman SC, Le Gouis J, Allard V (2014) Predictions of heading date in bread wheat (Triticum aestivum L.) using QTL-based parameters of an ecophysiological model. J Exp Bot 65(20):5849–5865. doi:10.1093/jxb/eru328

Bradshaw AD, Caspari EW, Thoday JM (1965) Evolutionary significance of phenotypic plasticity in plants. Adv Genet 13:115–155

Bull JK, Cooper M, DeLacy IH, Basford KE, Woodruff DR (1992) Utility of repeated checks for hierarchical classification of data from plant breeding trials. Field Crops Res 30((1–2)):79–95. doi:10.1016/0378-4290(92)90058-H

Burgueño J, Crossa J, Cornelius PL, Yang R-C (2008) Using factor analytic models for joining environments and genotypes without crossover genotype×environment interaction. Crop Sci 48(4):1291–1305. doi:10.2135/cropsci2007.11.0632

Burgueño J, De los Campos G, Weigel K, Crossa J (2012) Genomic prediction of breeding values when modeling genotype×environment interaction using pedigree and dense molecular markers. Crop Sci 52(2):707–719. doi:10.2135/cropsci2011.06.0299

Cabrera-Bosquet L, Crossa J, von Zitzewitz J, Serret MD, Araus JL (2012) High-throughput phenotyping and genomic selection: the frontiers of crop breeding converge. J Integr Plant Biol 54(5):312–320. doi:10.1111/j.1744-7909.2012.01116.x

Calus M, Veerkamp R (2011) Accuracy of multi-trait genomic selection using different methods. Genet Sel Evol 43(1):26

Campbell BT, Baenziger PS, Eskridge KM, Budak H, Streck NA, Weiss A, Gill KS, Erayman M (2004) Using environmental covariates to explain genotype × environment and QTL × environment interactions for agronomic traits on chromosome 3A of wheat. Crop Sci 44(2):620–627. doi:10.2135/cropsci2004.6200

Chapman S (2008) Use of crop models to understand genotype by environment interactions for drought in real-world and simulated plant breeding trials. Euphytica 161(1–2):195–208. doi:10.1007/s10681-007-9623-z

Chapman SC, Cooper M, Hammer GL, Butler DG (2000a) Genotype by environment interactions affecting grain sorghum. II. Frequencies of different seasonal patterns of drought stress are related to location effects on hybrid yields. Aust J Agric Res 51(2):209–222. doi:10.1071/AR99021

Chapman SC, Hammer GL, Butler DG, Cooper M (2000b) Genotype by environment interactions affecting grain sorghum. III. Temporal sequences and spatial patterns in the target population of environments. Aust J Agric Res 51(2):223–234. doi:10.1071/AR99022

Chapman SC, Cooper M, Hammer GL (2002a) Using crop simulation to generate genotype by environment interaction effects for sorghum in water-limited environments. Aust J Agric Res 53(4):379–389. doi:10.1071/AR01070

Chapman SC, Hammer GL, Podlich DW, Cooper M (2002b) Linking bio-physical and genetic models to integrate physiology, molecular biology and plant breeding. In: Kang M (ed) Quantitative genetics, genomics, and plant breeding. CAB International, Wallingford, pp 167–187

Chapman S, Cooper M, Podlich D, Hammer G (2003) Evaluating plant breeding strategies by simulating gene action and dryland environment effects. Agron J 95(1):99–113. doi:10.2134/agronj2003.9900

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 Environ 31(3):378–391. doi:10.1111/j.1365-3040.2007.01772.x

Chenu K, Chapman SC, Tardieu F, McLean G, Welcker C, Hammer GL (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(4):1507–1523. doi:10.1534/genetics.109.105429

Chenu K, Cooper M, Hammer GL, Mathews KL, Dreccer MF, Chapman SC (2011) Environment characterization as an aid to wheat improvement: interpreting genotype–environment interactions by modelling water-deficit patterns in North-Eastern Australia. J Exp Bot 62(6):1743–1755. doi:10.1093/jxb/erq459

Chenu K, Deihimfard R, Chapman SC (2013) Large-scale characterization of drought pattern: a continent-wide modelling approach applied to the Australian wheatbelt – spatial and temporal trends. New Phytol 198(3):801–820. doi:10.1111/nph.12192

Cobb JN, Declerck G, Greenberg A, Clark R, McCouch S (2013) Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype-phenotype relationships and its relevance to crop improvement. Theor Appl Genet 126(4):867–887. doi:10.1007/s00122-013-2066-0

Cooper M (1999) Concepts and strategies for plant adaptation research in rainfed lowland rice. Field Crops Res 64(1–2):13–34. doi:10.1016/S0378-4290(99)00048-9

Cooper M, Hammer GL (1996) Plant adaptation and crop improvement. CAB International, Wallingford

Cooper M, Messina CD, Podlich D, Totir LR, Baumgarten A, Hausmann NJ, Wright D, Graham G (2014) Predicting the future of plant breeding: complementing empirical evaluation with genetic prediction. Crop Pasture Sci 65(4):311–336

Crossa J, Yang R-C, Cornelius P (2004) Studying crossover genotype × environment interaction using linear-bilinear models and mixed models. JABES 9(3):362–380. doi:10.1198/108571104x4423

Crossa J, Burgueño J, Cornelius PL, McLaren G, Trethowan R, Krishnamachari A (2006) Modeling genotype × environment interaction using additive genetic covariances of relatives for predicting breeding values of wheat genotypes. Crop Sci 46(4):1722–1733. doi:10.2135/cropsci2005.11-0427

Crossa J, de los Campos G, Perez P, Gianola D, Burgueño J, Araus JL, Makumbi D, Singh RP, Dreisigacker S, Yan JB, Arief V, Banziger M, Braun HJ (2010) Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. Genetics 186(2):713-U406. doi:10.1534/genetics.110.118521

Crossa J, Beyene Y, Kassa S, Pérez P, Hickey JM, Chen C, de los Campos G, Burgueño J, Windhausen VS, Buckler E, Jannink J-L, Lopez Cruz MA, Babu R (2013) Genomic prediction in maize breeding populations with genotyping-by-sequencing. G3 Genes Genomes Genet 3(11):1903–1926. doi:10.1534/g3.113.008227

Crossa J, Perez P, Hickey J, Burgueno J, Ornella L, Ceron-Rojas J, Zhang X, Dreisigacker S, Babu R, Li Y, Bonnett D, Mathews K (2014) Genomic prediction in CIMMYT maize and wheat breeding programs. Heredity 112(1):48–60. doi:10.1038/hdy.2013.16

Cullis BR, Smith AB, Coombes NE (2006) On the design of early generation variety trials with correlated data. J Agric Biol Environ Stat 11(4):381–393. doi:10.1198/108571106x154443

Daetwyler HD, Kemper KE, van der Werf JHJ, Hayes BJ (2012) Components of the accuracy of genomic prediction in a multi-breed sheep population. J Anim Sci 90(10):3375–3384. doi:10.2527/jas.2011-4557

De la Vega AJ, Chapman SC (2010) Mega-environment differences affecting genetic progress for yield and relative value of component traits. Crop Sci 50(2):574–583. doi:10.2135/cropsci2009.04.0209

DeWitt TJ, Scheiner SM (2004) Phenotypic plasticity: functional and conceptual approaches. Oxford University Press, Oxford

Dobzhansky T, Spassky B (1963) Genetics of natural populations. XXXIV. Adaptive norm, genetic load and genetic elite in drosophila pseudoobscura. Genetics 48(11):1467–1485

Eberius M, Lima-Guerra J (2009) High-throughput plant phenotyping – data acquisition, transformation, and analysis. In: Edwards D, Stajich J, Hansen D (eds) Bioinformatics. pp 259–278. doi:10.1007/978-0-387-92738-1_13

Edmeades GO, McMaster GS, White JW, Campos H (2004) Genomics and the physiologist: bridging the gap between genes and crop response. Funct Plant Biol 90(1):5–18. doi:10.1016/j.fcr.2004.07.002

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6(5), e19379. doi:10.1371/journal.pone.0019379

Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics. 4th edn. Harlow, GB: Longman

Finlay K, Wilkinson G (1963) The analysis of adaptation in a plant-breeding programme. Aust J Agric Res 14(6):742–754. doi:10.1071/AR9630742

Gauch HG, Zobel RW (1997) Identifying mega-environments and targeting genotypes. Crop Sci 37(2):311–326. doi:10.2135/cropsci1997.0011183X003700020002x

Gu J, Yin X, Zhang C, Wang H, Struik PC (2014) Linking ecophysiological modelling with quantitative genetics to support marker-assisted crop design for improved yields of rice (Oryza sativa) under drought stress. Ann Bot 114(3):499–511. doi:10.1093/aob/mcu127

Guo Z, Tucker DM, Wang D, Basten CJ, Ersoz E, Briggs WH, Lu J, Li M, Gay G (2013) Accuracy of across-environment genome-wide prediction in maize nested association mapping populations. G3 Genes Genomes Genet 3(2):263–272. doi:10.1534/g3.112.005066

Guo Z, Tucker D, Basten C, Gandhi H, Ersoz E, Guo B, Xu Z, Wang D, Gay G (2014) The impact of population structure on genomic prediction in stratified populations. Theor Appl Genet 127(3):749–762. doi:10.1007/s00122-013-2255-x

Hammer GL, Kropff MJ, Sinclair TR, Porter JR (2002) Future contributions of crop modelling – from heuristics and supporting decision making to understanding genetic regulation and aiding crop improvement. Eur J Agron 18(1–2):15–31. doi:10.1016/S1161-0301(02)00093-X

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 Plant Sci 11(12):587–593. doi:10.1016/j.tplants.2006.10.006

Hammer GL, van Oosterom E, McLean G, Chapman SC, Broad I, Harland P, Muchow RC (2010) Adapting APSIM to model the physiology and genetics of complex adaptive traits in field crops. J Exp Bot 61(8):2185–2202. doi:10.1093/jxb/erq095

Harrison MT, Tardieu F, Dong Z, Messina CD, Hammer GL (2014) Characterizing drought stress and trait influence on maize yield under current and future conditions. Glob Chang Biol 20(3):867–878. doi:10.1111/gcb.12381

Hartmann A, Czauderna T, Hoffmann R, Stein N, Schreiber F (2011) HTPheno: an image analysis pipeline for high-throughput plant phenotyping. BMC Bioinf 12(1):148

Heffner EL, Jannink J-L, Sorrells ME (2011) Genomic selection accuracy using multifamily prediction models in a wheat breeding program. Plant Genome 4(1):65–75. doi:10.3835/plantgenome2010.12.0029

Heslot N, Yang H-P, Sorrells ME, Jannink J-L (2012) Genomic selection in plant breeding: a comparison of models. Crop Sci 52(1):146–160. doi:10.2135/cropsci2011.06.0297

Heslot N, Jannink J-L, Sorrells ME (2013) Using genomic prediction to characterize environments and optimize prediction accuracy in applied breeding data. Crop Sci 53(3):921–933. doi:10.2135/cropsci2012.07.0420

Holzworth DP, Huth NI, deVoil PG, Zurcher EJ, Herrmann NI, McLean G, Chenu K, van Oosterom EJ, Snow V, Murphy C, Moore AD, Brown H, Whish JPM, Verrall S, Fainges J, Bell LW, Peake AS, Poulton PL, Hochman Z, Thorburn PJ, Gaydon DS, Dalgliesh NP, Rodriguez D, Cox H, Chapman S, Doherty A, Teixeira E, Sharp J, Cichota R, Vogeler I, Li FY, Wang E, Hammer GL, Robertson MJ, Dimes JP, Whitbread AM, Hunt J, van Rees H, McClelland T, Carberry PS, Hargreaves JNG, MacLeod N, McDonald C, Harsdorf J, Wedgwood S, Keating BA (2014) APSIM – evolution towards a new generation of agricultural systems simulation. Environ Model Software. doi:10.1016/j.envsoft.2014.07.009

Jannink J-L, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. Brief Funct Genomics 9(2):166–177. doi:10.1093/bfgp/elq001

Janss L, De los Campos G, Sheehan N, Sorensen D (2012) Inferences from genomic models in stratified populations. Genetics 192(2):693–704. doi:10.1534/genetics.112.141143

Jarquín D, Crossa J, Lacaze X, Cheyron P, Daucourt J, Lorgeou J, Piraux F, Guerreiro L, Pérez P, Calus M, Burgueño J, Campos G (2013) A reaction norm model for genomic selection using high-dimensional genomic and environmental data. Theor Appl Genet 3:1–13. doi:10.1007/s00122-013-2243-1

Keating BA, Carberry PS, Hammer GL, Probert ME, Robertson MJ, Holzworth D, Huth NI, Hargreaves JNG, Meinke H, Hochman Z, McLean G, Verburg K, Snow V, Dimes JP, Silburn M, Wang E, Brown S, Bristow KL, Asseng S, Chapman S, McCown RL, Freebairn DM, Smith CJ (2003) An overview of APSIM, a model designed for farming systems simulation. Eur J Agron 18(3–4):267–288. doi:10.1016/S1161-0301(02)00108-9

Kleinknecht K, Möhring J, Singh KP, Zaidi PH, Atlin GN, Piepho HP (2013) Comparison of the performance of best linear unbiased estimation and best linear unbiased prediction of genotype effects from zoned Indian maize data. Crop Sci 53(4):1384–1391. doi:10.2135/cropsci2013.02.0073

Laperche A, Brancourt-Hulmel M, Heumez E, Gardet O, Hanocq E, Devienne-Barret F, Le Gouis J (2007) Using genotype × nitrogen interaction variables to evaluate the QTL involved in wheat tolerance to nitrogen constraints. Theor Appl Genet 115(3):399–415. doi:10.1007/s00122-007-0575-4

Lin CS, Binns MR (1988) A method of analyzing cultivar x location x year experiments: a new stability parameter. Theor Appl Genet 76(3):425–430. doi:10.1007/bf00265344

Makumburage GB, Richbourg HL, LaTorre KD, Capps A, Chen C, Stapleton AE (2013) Genotype to phenotype maps: multiple input abiotic signals combine to produce growth effects via attenuating signaling interactions in maize. G3 Genes Genomes Genet 3(12):2195–2204. doi:10.1534/g3.113.008573

Malosetti M, Voltas J, Romagosa I, Ullrich SE, van Eeuwijk FA (2004) Mixed models including environmental covariables for studying QTL by environment interaction. Euphytica 137(1):139–145. doi:10.1023/B:EUPH.0000040511.46388.ef

Malosetti M, Ribaut J-M, van Eeuwijk FA (2013) The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis. Front Physiol 4:1–17. doi:10.3389/fphys.2013.00044

Mathews K, Malosetti M, Chapman S, McIntyre L, Reynolds M, Shorter R, van Eeuwijk F (2008) Multi-environment QTL mixed models for drought stress adaptation in wheat. Theor Appl Genet 117(7):1077–1091. doi:10.1007/s00122-008-0846-8

Messina CD, Podlich D, Dong Z, Samples M, Cooper M (2011) Yield–trait performance landscapes: from theory to application in breeding maize for drought tolerance. J Exp Bot 62(3):855–868. doi:10.1093/jxb/erq329

Metzker ML (2010) Sequencing technologies – the next generation. Nat Rev Genet 11(1):31–46

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157(4):1819–1829

Muir W, Nyquist WE, Xu S (1992) Alternative partitioning of the genotype-by-environment interaction. Theor Appl Genet 84(1–2):193–200. doi:10.1007/bf00224000

Passioura JB (2012) Phenotyping for drought tolerance in grain crops: when is it useful to breeders? Funct Plant Biol 39(11):851–859. doi:10.1071/FP12079

Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. PLoS Genet 2(12), e190. doi:10.1371/journal.pgen.0020190

Piepho HP (1998) Methods for comparing the yield stability of cropping systems. J Agron Crop Sci 180(4):193–213. doi:10.1111/j.1439-037X.1998.tb00526.x

Piepho H-P (2000) A mixed-model approach to mapping quantitative trait loci in barley on the basis of multiple environment data. Genetics 156(4):2043–2050

Piepho HP (2009) Ridge regression and extensions for genomewide selection in maize. Crop Sci 49(4):1165–1176

Piepho HP, Möhring J (2005) Best linear unbiased prediction of cultivar effects for subdivided target regions. Crop Sci 45(3):1151–1159. doi:10.2135/cropsci2004.0398

Piepho H, Möhring J, Melchinger A, Büchse A (2008) BLUP for phenotypic selection in plant breeding and variety testing. Euphytica 161(1):209–228. doi:10.1007/s10681-007-9449-8

Podlich DW, Cooper M, Basford KE, Geiger HH (1999) Computer simulation of a selection strategy to accommodate genotype environment interactions in a wheat recurrent selection programme. Plant Breed 118(1):17–28. doi:10.1046/j.1439-0523.1999.118001017.x

Poland JA, Rife TW (2012) Genotyping-by-sequencing for plant breeding and genetics. Plant Genome 5(3):92–102. doi:10.3835/plantgenome2012.05.0005

Poland J, Endelman J, Dawson J, Rutkoski J, Wu S, Manes Y, Dreisigacker S, Crossa J, Sánchez-Villeda H, Sorrells M, Jannink J-L (2012a) Genomic selection in wheat breeding using genotyping-by-sequencing. Plant Genome 5(3):103–113. doi:10.3835/plantgenome2012.06.0006

Poland JA, Brown PJ, Sorrells ME, Jannink J-L (2012b) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS One 7(2), e32253. doi:10.1371/journal.pone.0032253

Prasanna B, Araus J, Crossa J, Cairns J, Palacios N, Das B, Magorokosho C (2013) High-throughput and precision phenotyping for cereal breeding programs. In: Gupta PK, Varshney RK (eds) Cereal genomics II. pp 341–374. doi:10.1007/978-94-007-6401-9_13

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 Physiol 131(2):664–675. doi: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. J Exp Bot 55(407):2461–2472. doi:10.1093/jxb/erh200

Reynolds M, Foulkes MJ, Slafer GA, Berry P, Parry MAJ, Snape JW, Angus WJ (2009a) Raising yield potential in wheat. J Exp Bot 60(7):1899–1918. doi:10.1093/jxb/erp016

Reynolds M, Manes Y, Izanloo A, Langridge P (2009b) Phenotyping approaches for physiological breeding and gene discovery in wheat. Ann App Biol 155(3):309–320. doi:10.1111/j.1744-7348.2009.00351.x

Riedelsheimer C, Endelman JB, Stange M, Sorrells ME, Jannink J-L, Melchinger AE (2013) Genomic predictability of interconnected biparental maize populations. Genetics 194(2):493–503. doi:10.1534/genetics.113.150227

Romagosa I, Fox P (1993) Genotype × environment interaction and adaptation. In: Hayward MD, Bosemark NO, Romagosa I, Cerezo M (eds) Plant breeding: principles and prospects. Springer, Dordrecht, pp 373–390

Romagosa I, Borràs-Gelonch G, Slafer G, Eeuwijk F (2013) Genotype by environment interaction and adaptation. In: Savin R, Costa-Pierce B, Misztal I, Whitelaw CB, Christou P (eds) Sustainable food production. Springer, New York, pp 846–870. doi:10.1007/978-1-4614-5797-8_199

Sadras VO, Lawson C (2011) Genetic gain in yield and associated changes in phenotype, trait plasticity and competitive ability of South Australian wheat varieties released between 1958 and 2007. Crop Pasture Sci 62(7):533–549. doi:10.1071/CP11060

Sadras VO, Rebetzke GJ, Edmeades GO (2013) The phenotype and the components of phenotypic variance of crop traits. Field Crop Res 154:255–259. doi:10.1016/j.fcr.2013.10.001

Sarkar S (1999) From the reaktionsnorm to the adaptive norm: the norm of reaction, 1909–1960. Biol Philos 14(2):235–252. doi:10.1023/a:1006690502648

Schulz-Streeck T, Ogutu JO, Karaman Z, Knaak C, Piepho HP (2012) Genomic selection using multiple populations. Crop Sci 52(6):2453–2461. doi:10.2135/cropsci2012.03.0160

Slafer GA (2003) Genetic basis of yield as viewed from a crop physiologist's perspective. Ann Appl Biol 142(2):117–128. doi:10.1111/j.1744-7348.2003.tb00237.x

Slafer G, Kernich G (1996) Have changes in yield (1900–1992) been accompanied by a decreased yield stability in Australian cereal production? Aust J Agric Res 47(3):323–334. doi:10.1071/AR9960323

Slafer G, Rawson H (1994) Sensitivity of wheat phasic development to major environmental factors: a re-examination of some assumptions made by physiologists and modellers. Funct Plant Biol 21(4):393–426. doi:10.1071/PP9940393

Smith A, Cullis B, Thompson R (2001a) Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. Biometrics 57(4):1138–1147. doi:10.1111/j.0006-341X.2001.01138.x

Smith AB, Cullis BR, Appels R, Campbell AW, Cornish GB, Martin D, Allen HM (2001b) The statistical analysis of quality traits in plant improvement programs with application to the mapping of milling yield in wheat. Aust J Agric Res 52(12):1207–1219. doi:10.1071/AR01058

Smith AB, Cullis BR, Thomson R (2005) The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. J Agric Sci 143(06):449–462. doi:10.1017/S0021859605005587

Smith AB, Lim P, Cullis BR (2006) The design and analysis of multi-phase plant breeding experiments. J Agric Sci 144(05):393–409. doi:10.1017/S0021859606006319

Snape JW, Butterworth K, Whitechurch E, Worland AJ (2001) Waiting for fine times: genetics of flowering time in wheat. Euphytica 119(1–2):185–190. doi:10.1023/a:1017594422176

Spindel J, Wright M, Chen C, Cobb J, Gage J, Harrington S, Lorieux M, Ahmadi N, McCouch S (2013) Bridging the genotyping gap: using genotyping by sequencing (GBS) to add high-density SNP markers and new value to traditional bi-parental mapping and breeding populations. Theor Appl Genet 126(11):2699–2716. doi:10.1007/s00122-013-2166-x

Stephens M (2013) A unified framework for association analysis with multiple related phenotypes. PLoS One 8(7), e65245. doi:10.1371/journal.pone.0065245

Tardieu F (2003) Virtual plants: modelling as a tool for the genomics of tolerance to water deficit. Trends Plant Sci 8(1):9–14. doi:10.1016/S1360-1385(02)00008-0

Tardieu F, Reymond M, Muller B, Granier C, Simonneau T, Sadok W, Welcker C (2005) Linking physiological and genetic analyses of the control of leaf growth under changing environmental conditions. Aust J Agric Res 56(9):937–946. doi:10.1071/AR05156

Uitdewilligen JGAML, Wolters A-MA, D'hoop BB, Borm TJA, Visser RGF, van Eck HJ (2013) A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato. PLoS One 8(5), e62355. doi:10.1371/journal.pone.0062355

Utz HF, Melchinger AE, Schön CC (2000) Bias and sampling error of the estimated proportion of genotypic variance explained by quantitative trait loci determined from experimental data in maize using cross validation and validation with independent samples. Genetics 154(4):1839–1849

van der Heijden G, Song Y, Horgan G, Polder G, Dieleman A, Bink M, Palloix A, van Eeuwijk F, Glasbey C (2012) SPICY: towards automated phenotyping of large pepper plants in the greenhouse. Funct Plant Biol 39(11):870–877. doi:10.1071/FP12019

van Eeuwijk F, Denis J, Kang M (1996) Incorporating additional information on genotypes and environments in models for two-way genotype by environment tables. In: Kang MS, Gauch HG Jr (eds) Genotype-by-environment interaction. Taylor & Francis Group, Boca Raton, Florida, pp 15–50

van Eeuwijk FA, Malosetti M, Yin X, Struik PC, Stam P (2005) Statistical models for genotype by environment data: from conventional ANOVA models to eco-physiological QTL models. Aust J Agric Res 56(9):883–894. doi:10.1071/AR05153

van Eeuwijk FA, Bink MCAM, Chenu K, Chapman SC (2010) Detection and use of QTL for complex traits in multiple environments. Curr Opin Plant Biol 13(2):193–205. doi:10.1016/j.pbi.2010.01.001

Varshney RK, Terauchi R, McCouch SR (2014) Harvesting the promising fruits of genomics: applying genome sequencing technologies to crop breeding. PLoS Biol 12(6), e1001883. doi:10.1371/journal.pbio.1001883

Verbyla AP, Cullis BR (2012) Multivariate whole genome average interval mapping: QTL analysis for multiple traits and/or environments. Theor Appl Genet 125(5):933–953

Wang E, Robertson MJ, Hammer GL, Carberry PS, Holzworth D, Meinke H, Chapman SC, Hargreaves JNG, Huth NI, McLean G (2002) Development of a generic crop model template in the cropping system model APSIM. Eur J Agron 18(1–2):121–140. doi:10.1016/S1161-0301(02)00100-4

West BT, Welch KB, Galecki AT (2006) Linear mixed models: a practical guide using statistical software. Boca Raton, Florida

Windhausen VS, Atlin GN, Hickey JM, Crossa J, Jannink J-L, Sorrells ME, Raman B, Cairns JE, Tarekegne A, Semagn K, Beyene Y, Grudloyma P, Technow F, Riedelsheimer C, Melchinger AE (2012) Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. G3 Genes Genomes Genet 2(11):1427–1436. doi:10.1534/g3.112.003699

Woltereck R (1909) Weitere experimentelle Untersuchungen über Artveränderung, speziel über das Wesen quantitativer Artunterschiede bei Daphnien. Verhandlungen der deutschen zoologischen Gesellschaft 19:110–173

Wright S (1931) Evolution in mendelian populations. Genetics 16(2):97

Wright S (1932) The roles of mutation, inbreeding, crossbreeding and selection in evolution. In: Proceedings of the 6th international congress of genetics, Brooklin, NY, pp 356–366

Yan W, Hunt LA, Sheng Q, Szlavnics Z (2000) Cultivar evaluation and mega-environment investigation based on the GGE biplot. Crop Sci 40(3):597–605. doi:10.2135/cropsci2000.403597x

Yin X, Chasalow SD, Dourleijn CJ, Stam P, Kropff MJ (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(6):539–549. doi:10.1046/j.1365-2540.2000.00790.x

Yin X, Stam P, Kropff MJ, Schapendonk AHCM (2003) Crop modeling QTL mapping, and their complementary role in plant breeding. Agron J 95(1):90–98. doi:10.2134/agronj2003.9000

Yin X, Struik PC, Kropff MJ (2004) Role of crop physiology in predicting gene-to-phenotype relationships. Trends Plant Sci 9(9):426–432. doi:10.1016/j.tplants.2004.07.007

Zhao Y, Gowda M, Longin F, Würschum T, Ranc N, Reif J (2012) Impact of selective genotyping in the training population on accuracy and bias of genomic selection. Theor Appl Genet 125(4):707–713. doi:10.1007/s00122-012-1862-2

Zheng B, Chenu K, Fernanda Dreccer M, Chapman SC (2012) Breeding for the future: what are the potential impacts of future frost and heat events on sowing and flowering time requirements for Australian bread wheat (Triticum aestivum) varieties? Glob Chang Biol 18(9):2899–2914. doi:10.1111/j.1365-2486.2012.02724.x

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. J Exp Bot 64(12):3747–3761. doi:10.1093/jxb/ert209