

An overview of statistical methods to detect and understand genotype-by-environment interaction and QTL-by-environment interaction

Paulo C. Rodrigues^{1,2}

¹CAST, Faculty of Natural Sciences, University of Tampere, Finland

²Department of Statistics, Federal University of Bahia, Salvador, Brazil
e-mail: paulocanas@gmail.com

SUMMARY

Genotype-by-environment interaction (GEI) is frequently encountered in multi-environment trials, and represents differential responses of genotypes across environments. With the development of molecular markers and mapping techniques, researchers can go one step further and analyse the whole genome to detect specific locations of genes which influence a quantitative trait such as yield. Such a location is called a quantitative trait locus (QTL), and when these QTLs have different expression across environments we talk about QTL-by-environment interaction (QEI), which is the basis of GEI. Good understanding of these interactions enables researchers to select better genotypes across different environmental conditions, and consequently to improve crops in developed and developing countries. In this paper we present an overview of statistical methods and models commonly used to detect and to understand GEI and QEI, ranging from the simple joint regression model to complex eco-physiological genotype-to-phenotype simulation models.

Keywords: Genotype-by-environment interactions; QTL-by-environment interactions; QTL detection; Additive main effects and multiplicative interaction model; Eco-physiological genotype-to-phenotype simulation models.

1. Introduction

One of the main challenges in statistical genetics is to find superior genotypes over a wide range of agro-ecological conditions and also over a number of years. This is also a challenge for farmers, breeders and geneticists, although farmers and breeders often have conflicting interests: breeders want a genotype that can be sold everywhere, and farmers a genotype adapted to their climate and soil

management. To achieve this purpose, multi-environment trials (METs) are conducted, in which a series of genotypes is evaluated over environmental conditions and over time. The data from these trials are usually summarised in a two-way table with genotypes in the rows and environments (locality/year combinations) in the columns, or vice versa. In most of these two-way tables it is possible to find differences between genotypes in their phenotype (e.g. yield) stability along environments; that is, the genotypic and environmental effects are not simply additive, and genotype-by-environment interaction (GEI) is present in the data (van Eeuwijk et al., 2016). GEI is defined by the change of the genetic ranking of genotypes with the environment; for instance, a genotype that is superior in well-watered conditions may yield poorly under dry conditions. The GEI can be expressed either as crossovers, when two different genotypes change in rank order of performance when evaluated in different environments, or as inconsistent responses of some genotypes across environments without changes in rank order. The study and understanding of these interactions is a major challenge, serving to improve complex traits (e.g. yield) across environmental gradients (Annicchiarico, 2002, 2009; Crossa, 1990; Romagosa et al., 2009; van Eeuwijk et al., 2016).

With the development of molecular markers and mapping techniques, researchers can go one step further than GEI and analyse the whole genome to detect specific locations of genes which influence a quantitative trait. Such a location is called a quantitative trait locus (QTL), and when these QTLs have different expression across environments we talk about QTL-by-environment interaction (QEI), which is the basis of GEI (Romagosa et al., 2009; Gauch et al., 2011; Rodrigues et al., 2014). A good understanding of these interactions allows researchers to select better genotypes across different environmental gradients, and consequently to improve crops for developed and, in particular, for developing countries, based on their climate and soil characteristics.

A physiologically inspired alternative approach to GEI is based on crop growth simulation models. Crop growth models represent a class of genotype-to-phenotype models based on prior knowledge (Spitters, 1990; van Ittersum et al.,

2003; Chenu et al., 2009; Rodrigues, 2012) that has proved to be useful for understanding GEI and QEI. These models allow the use of genetic and environmental characteristics to simulate the behaviour of each genotype in each environmental set-up along the growing season (Rodrigues, 2012; Cooper et al., 2009; Bertin et al., 2010; Letort et al., 2008), which is very useful for decision-making in crop science and plant breeding.

2. Genotype-by-environment interactions – the statistical analysis of two-way tables

To better understand GEI and to make predictions for different locations and/or different years, a wide range of statistical methods have been used. They have been applied to the output of extensive experiments and plant breeding programmes conducted under different environmental conditions (or locations) and over several years (van Eeuwijk et al., 2005; Malosetti et al., 2010; Aastveit and Mejza, 1992; Kang and Gauch, 1996).

In a chronological analysis of research on GEI and QEI, Rodrigues (2012) found that the number of publications on GEI together with QTL analyses had remained almost constant between 2005 and 2012. Research on regression-based techniques continues to increase within GEI analysis, and since 1990 has been the most commonly used statistical tool. The particular case of factorial regression models represents less than 10% of total research on regression for GEI. Research articles which use graphical techniques such as biplots (Gabriel, 1971) or genotype main effect plus genotype-by-environment interaction (GGE) biplots (Yan and Kang, 2002) have increased sharply in number, especially since 2004. There is also a clear increase in research on singular value decomposition techniques such as principal component analysis (PCA) and additive main effects and multiplicative interaction (AMMI) models (Gauch, 1992). The steep increase in the use of biplots is, in part, explained by the increase in the use of PCA and AMMI models, which also use these graphical representations in their outputs.

2.1. Statistical models based on regression and singular value decomposition

The simplest model to describe phenotypic observations along environments is the additive model without interaction terms. In this case, the expected phenotypic response for genotype i , $i = 1, \dots, I$, in environment j , $j = 1, \dots, J$, equals the grand mean plus the genotype and environment main effects (both expressed as deviations from the grand mean), that is

$$y_{i,j} = \mu + G_i + E_j + \varepsilon_{i,j}. \quad (1)$$

The additive model is the basis of all models with interaction, but it is only applicable when there is no GEI in the two-way table with genotypes in the rows and environments in the columns, that is, when the phenotypic response across environments is a set of parallel lines. If there is interaction between genotypes and environments, model (1) can be written to account for GEI, that is

$$y_{i,j} = \mu + G_i + E_j + (G.E)_{i,j} + \varepsilon_{i,j}, \quad (2)$$

where $(G.E)_{i,j}$ represents the GEI term for genotype i and environment j . The full interaction model (2) has as many parameters to be estimated as genotype-by-environment combinations, which is associated with less precise tests and represents a less parsimonious model. An alternative extension of the additive model (1) was first proposed by Finlay and Wilkinson (1963), where the phenotypic responses across environments are regressed on the phenotypic mean over environments (a measure of productivity or biological quality in the absence of other environmental characterisations). The GEI is expressed by the I slopes β_i , and the model can be written as

$$y_{i,j} = \mu + G_i + E_j + \beta_i E_j + \varepsilon_{i,j}. \quad (3)$$

Another regression-based model was presented by Gusmão (1985), where the (physical) block information is used to correct for spatial effects. In this way the phenotypic responses per block are regressed across environments, resulting in $i \times b$ regressions, where b is the number of blocks. Other studies and

generalisations related to Gusmão's approach were presented by Pereira et al. (2007, 2012b).

A further alternative to the full interaction model is the additive main effects and multiplicative interaction (AMMI) model (Gollob, 1968; Mandel, 1969; Bradu and Gabriel, 1978; Gauch, 1988; Gauch, 1992; Paderewski et al., 2011; Hongyu et al., 2014; Rodrigues et al., 2016), which is more flexible than the Finlay and Wilkinson regression, because it can partition the interaction into $N = \min(I - 1, J - 1)$ terms. It combines analysis of variance (ANOVA) and principal component analysis (PCA), with ANOVA performed first, and then PCA (i.e. singular value decomposition) applied to the resultant GEI matrix (Gauch, 1992). The model can be written as

$$\begin{aligned} y_{i,j} &= \mu + G_i + E_j + \sum_{n=1}^N \lambda_n \gamma_{i,n} \delta_{j,n} + \varepsilon_{i,j} \\ &= \mu + G_i + E_j + \sum_{n=1}^N a_{i,n} b_{j,n} + \varepsilon_{i,j}, \end{aligned} \quad (4)$$

where λ_n is the singular value for interaction principal component (IPC) n , $\gamma_{i,n}$ is the left singular vector for genotype i in component n , $\delta_{j,n}$ is the right singular vector for environment j in component n , $\varepsilon_{i,j}$ is the residual for genotype i in environment j , and N is the number of retained components. Rodrigues et al. (2011) and Pereira et al. (2012a) made comparisons in terms of "robustness" with different incidence rates of missing values, and in terms of dominant/winner genotypes.

A remarkable feature of the AMMI model is its ability to create mega-environments, that is, groups of environments with a similar response to the variable of interest (e.g. yield) (Gauch and Zobel, 1997) and to produce biplot graphs that are very useful to delineate mega-environments. A useful role of mega-environments is the possibility of more accurate prediction for new years and/or new environments with similar environmental conditions as in a given mega-environment.

The AMMI model, in its standard form, implicitly assumes that all entries of the two-way data table have equal weights and that no outliers are present in the data (Gauch, 1992; Paderewski and Rodrigues, 2014). Several data imputation

techniques have been proposed to deal with the missing data, such as those of Arciniegas-Alarcón (2010, 2014) and Gauch and Zobel (1990). Rodrigues et al. (2014) proposed a weighted AMMI algorithm where different weights are assigned to columns and/or rows and/or particular cells of the two-way data table, and Rodrigues et al. (2016) proposed a robust version of the AMMI model that accounts for data contamination. Moreover, Josse et al. (2014) and references therein discuss the use of the Bayesian AMMI model to study genotype-by-environment data.

A similar alternative to the GEI model is the GGE model (Yan and Kang, 2002, 2003; Yang et al., 2009) which applies PCA (i.e. singular value decomposition) to the two-way table without environmental main effects, that is:

$$y_{i,j} = \mu + E_j + \sum_{n=1}^N a_{i,n}b_{j,n} + \varepsilon_{i,j}, \quad (5)$$

where $a_{i,n}$ and $b_{j,n}$ are genotypic and environmental parameters (scores) for the bilinear term n . Both the AMMI and GGE models are more useful when using graphical representations such as biplots (Gabriel, 1971).

Gauch et al. (2008), and references therein, present detailed comparisons between the AMMI and GGE models.

2.2. Factorial regression – including environmental and genotypic information in the model

When specific environmental (and/or genotypic) information is available (e.g. rainfall, radiation, temperature, marker information), the linear-bilinear model recommended for use is the biadditive factorial regression model, also termed reduced rank factorial regression (Denis, 1988; van Eeuwijk et al., 1996; van Eeuwijk, 1995) because it allows this extra information to be included in the model. Considering the simple case in which the interaction is due to two environmental variables Z_{1j} and Z_{2j} , the model can be written as

$$y_{i,j} = \mu + G_i + E_j + \beta_{1,i}Z_{1,j} + \beta_{2,i}Z_{2,j} + \varepsilon_{i,j}, \quad (6)$$

where $\beta_{1,i}$ and $\beta_{2,i}$ are the genotypic sensitivities to the two respective environmental variables. This model is an extension of the Finlay–Wilkinson

regression (1.2) in which the interaction is written based on several real environmental variables. This allows a physiological interpretation of the GEI in terms of real environmental information. The generalisation to the case when H environmental covariates $Z_{1,j}, \dots, Z_{H,j}$ are available is straightforward:

$$y_{i,j} = \mu + G_i + E_j + \sum_{h=1}^H \beta_{h,i} Z_{h,j} + \varepsilon_{i,j}. \quad (7)$$

A similar expression can be obtained when, instead of only H environmental covariates, we also have information about K genotypic covariates (e.g. physiological parameters or marker information). The generalisation to H environmental covariates $Z_{1,j}, \dots, Z_{H,j}$, and K genotypic covariates $X_{1,i}, \dots, X_{K,i}$, can be written as:

$$y_{i,j} = \mu + G_i + E_j + \sum_{h=1}^H \beta_{h,i} Z_{h,j} + \sum_{k=1}^K X_{k,i} \tau_{k,j} + \sum_{k=1}^K \sum_{h=1}^H \varphi_{k,h} X_{k,i} Z_{h,j} + \varepsilon_{i,j}. \quad (8)$$

These coefficients are not genotype- or environment-dependent. The coefficients $\beta_{h,i}$ are genotypic sensitivities to the environmental covariables $Z_{h,j}$, and the $\tau_{k,j}$ denote environmental weighting constants with respect to the genotypic covariable $X_{k,i}$ (Baril et al., 1995). The parameters $\varphi_{k,h}$ represent coefficients with respect to cross-products of the genotypic covariables $X_{k,i}$ and environmental covariables $Z_{h,j}$. Further generalisations are possible depending on the research objectives (van Eeuwijk et al., 1996; Romagosa et al., 2009).

2.3. Linear mixed model – taking into account the variance structure of the data

A more elaborate approach to understanding GEI is the mixed model framework (Galwey, 2006; Verbeke and Molenberghs, 2009). This methodology combines the modelling of the mean and the variance, and provides a powerful tool to analyse GEI. The main advantage of these models is the ability to model the heterogeneity of variance across environments and correlations between environments. Unlike the models presented above, where all terms except the residual are fixed, the mixed linear model (Searle, 1971) provides a framework where the fixed effects can be combined with several random terms (Lourenço

et al., 2017). Residual maximum likelihood (REML) (Patterson and Thompson, 1971; Searle et al., 1992) is used to estimate variances and random parameters. A mixed model for a two-way table indexed by genotypes and environments is

$$y_{i,j} = \mu + G_i + E_j + (G.E)_{i,j} + \varepsilon_{i,j} \quad (9)$$

where the model parameters are defined as before. Typically, E_j is fixed and G_i , $(G.E)_{i,j}$ and $\varepsilon_{i,j}$ are random, following a normal distribution with zero mean and a variance specific to the term (Boer et al., 2007; Malosetti et al., 2004). Piepho (1997) proposed an analysis of genotype-by-environment data using mixed models with multiplicative terms which can be seen as a mixed effects model version of the fixed effects AMMI model.

3. QTL-by-environment interaction

QTL scans are a quick and relatively straightforward way to find a rough localisation of genetic factors that influence a complex trait. These QTL scans are usually conducted in multiple environments in order to increase generality and reliability (Broman and Sen, 2009; Alimi et al., 2012), but often inconsistent QTL detections are found. Inconsistent results raise questions about the validity and utility of these QTLs, especially those that are only marginally significant or infrequent. Inconsistent QTLs can emerge from a mixture of two causes: (i) from false positives and false negatives due to inadequate population size or imperfect statistical models and noisy phenotypic data; and (ii) from actual QTL-by-environment interaction (QEI) due to a given allele increasing a phenotypic trait in only some environments while having no detectable effect or even a significant negative effect in other environments (Gauch et al., 2011).

When dealing with QEI instead of GEI, the previously described fixed and mixed models can be easily adapted. For example, the QTL model with interaction can be written as

$$y_{i,j} = \mu + QTL_i + G_i^* + E_j + (QTL.E)_{i,j} + (G.E)_{i,j}^* + \varepsilon_{i,j}, \quad (10)$$

where QTL_i is the QTL main effect, $(QTL.E)_{i,j}$ is the QEI, G_i^* is the genotypic residual, and $(G.E)_{i,j}^*$ is the residual from the interaction. More details on these models and how to include genetic information such as marker information can be found in van Eeuwijk et al. (2005) and Romagosa et al. (2009).

A major point of interest is whether QEI can be detected for the phenotypic trait of interest and whether this QEI can be interpreted in terms of underlying QTLs for physiological parameters or molecular markers. The QTL model of interest uses explicit marker-derived information to describe the GEI in terms of QTLs as dependent on the environments (i.e. the QEI). The inclusion of this marker information (genetic predictors) allows one to test whether the phenotypic trait (e.g. yield) is affected by the DNA at a particular genome position, and whether this effect depends on the environment. Romagosa et al. (1996) proposed the use of the AMMI model for QTL mapping, and Gauch et al. (2011) followed up on that approach, applying a parsimonious AMMI model to the phenotypic data in order to gain accuracy, and then using the AMMI predicted values to obtain the QTL scans, so as to order the environments by AMMI scores and enable the analysis of patterns with ecological or biological interpretation. A three-stage approach that is able to account for differences in error variance across environments, proposed by Rodrigues et al. (2014), uses a weighted AMMI algorithm, instead of the standard AMMI model, to obtain the predicted values that are used to obtain the QTL scans.

If, instead of a fixed effects model, a mixed effects framework is considered, the model can be defined as (Boer et al., 2007):

$$\begin{aligned} y_{i,j} &= [\mu + E_j] + [G_i + (G.E)_{i,j}] \\ &= [\mu_j] + [\sum_{k=1}^K x_{k,i} \alpha_{k,j} + \theta_{i,j}] \end{aligned} \quad (11)$$

where μ_j is the intercept for each environment, $x_{k,i}$ is derived from marker genotype information for genotype i , $\alpha_{k,j}$ is the QTL allele substitution effect for environment j , K is the total number of QTLs underlying $y_{i,j}$ (e.g. yield), and $\theta_{i,j}$ follows a multivariate normal distribution with zero mean vector and a given variance-covariance (VCOV) matrix. The choice of the best VCOV structure can

be made by following the procedure described in Malosetti et al. (2004) and Boer et al. (2007).

4. Eco-physiological genotype-to-phenotype models

All of the models described above are intended to analyse data after it has been collected in multi-environment trials. This procedure of collecting data is expensive and time-consuming, and has limitations regarding the number of genotypes, traits and environmental conditions considered. Simulation tools such as genotype-to-phenotype models have proved to be useful for gaining better understanding of GEI and QEI (van Eeuwijk et al., 2010).

A physiologically inspired alternative approach for collecting field data is based on physiological crop growth simulation models. Crop growth models represent a class of genotype-to-phenotype (G-P) models with a prior biological structure (Spitters, 1990; van Ittersum et al., 2003) that can be used to help understand GEI and QEI (Tardieu, 2003; van Eeuwijk et al., 2005; Letort et al., 2008; Chenu et al., 2009; Cooper et al., 2009; Bertin et al., 2010; van Eeuwijk et al., 2010). These models allow the simulation of the trait of interest (e.g. yield) along the growing season (i.e. every day) and require as input: (i) genotypic information on the crop at hand, i.e. the genetic map with the position of the markers in the chromosomes and marker information; (ii) information about the physiological parameters of the model for each of the considered genotypes; and (iii) environmental characteristics of the study locations (i.e. weather, soil, etc.).

The integration of statistical genetics and crop growth modelling for reliable and robust prediction of phenotypic traits, on the basis of genotype-specific and stable physiological parameters and environmental characteristics, is the subject of extensive research in the plant sciences (Tardieu, 2003; Chenu et al., 2008; Malosetti et al., 2010). These G-P models allow the inclusion of genetic information such as previously found QTLs for the trait and/or QTLs for the physiological parameters, which will result in a more parsimonious and meaningful model. A particular strength of crop growth models in comparison

with the more statistical G-P models is that they contain explicit representations of development over time, which may be useful in describing GEI (Chenu et al., 2009). Otherwise, “time” would be an extra dimension on the phenotypic observations and harder to collect. Recently, a wide spectrum of physiological models have been proposed for better interpretation of GEI and QEI, focusing on traits of varying complexity such as yield (Yin et al., 2000; Tardieu, 2003; Yin et al., 2004; Chenu et al., 2008), leaf elongation (Reymond et al., 2003; Reymond et al., 2004; Chenu et al., 2008), chemical concentration in seed grains (Ishii et al., 2010) and fruit quality (Quilot et al., 2005). Rodrigues (2012) developed a seven-component eco-physiological model for yield in pepper that simulates yield and yield components, where the yield components were given a simple QTL basis, and showed how credible patterns of GEI and QEI for yield can be simulated from genotype-specific yield components with a simple QTL basis.

5. Discussion

In this paper we have described some of the standard techniques used to analyse and to structure genotype-by-environment interaction (GEI), provided a general overview of fixed and mixed effects models commonly used to detect and to understand quantitative trait locus (QTL)-by-environment interaction (QEI), and given a brief introduction and review of eco-physiological genotype-to-phenotype models. Despite the wide range of existing references and techniques to explore and better understand GEI and QEI, not all of them are available to all breeders and researchers. In some cases, the statistical methods are too complex to be computationally implemented and applied by non-statisticians; in other cases, although these complex techniques are already well implemented in statistical packages, the software is commercial and too expensive for budgets in developing countries, where statistical improvements are slow to arrive. However, free software (e.g. QTL Cartographer, MATMODEL, R/agricolae and R/qtl) is available for most of the basic methods discussed in this paper.

Acknowledgements

P.C. Rodrigues acknowledges financial support from the National Council for Scientific and Technological Development (CNPq) through project UNIVERSAL-MCTI/CNPq 448775/2014-0.

REFERENCES

- Aastveit A.H., Mejza S. (1992): A selected bibliography on statistical methods for the analysis of genotype x environment interaction. *Biuletyn Oceny Odmian*, 24-25: 83-97.
- Alimi N.A., Bink M.C.A.M., Dieleman J.A., Nicolai M., Wubs M., Heuvelink E., Magan J.J., Voorrips R.E., Jansen J., Rodrigues P.C., Vercauteren A., Vuylsteke M., Song Y., Glasbey C., Barocsi A., Lefebvre V., Palloix A., van Eeuwijk F.A. (2012): Genetic and QTL analyses of yield and a set of physiological traits in pepper. *Euphytica* 190: 181–201.
- Arciniegas-Alarcón S., García-Peña M., Krzanowski W.J., Dias C.T.S. (2014): An alternative methodology for imputing missing data in trials with genotype-by-environment interaction: some new aspects. *Biometrical Letters* 51: 75-88.
- Arciniegas-Alarcón S., Peña M.G., Dias C.T.S., Krzanowski W.J. (2010): An alternative methodology for imputing missing data in trials with genotype-by-environment interaction. *Biometrical Letters* 47: 1-14.
- Annicchiarico P. (2009): Coping with and exploiting genotype-by-environment interactions. *In*: Ceccarelli, S., E.P., G. & Weltzien, E. (eds.) *Plant breeding and farmer participation*. Rome: FAO.
- Annicchiarico P. (2002): *Genotype x Environment Interactions – Challenges and Opportunities for Plant Breeding and Cultivar Recommendations*. FAO Plant Production and Protection Papers [Online].
- Baril C.P., Denis J.B., Wustman R., van Eeuwijk F.A. (1995): Analyzing Genotype by Environment Interaction in Dutch Potato Variety Trials Using Factorial Regression. *Euphytica* 82: 149-155.
- 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.
- 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.
- Bradu D., Gabriel K.R. (1978): Biplot as a Diagnostic Tool for Models of 2-Way Tables. *Technometrics* 20: 47-68.
- Broman K.W., Sen S. (2009): *A Guide to QTL Mapping with R/qtl*, New York, Springer-Verlag.

- 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.
- 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.
- 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.
- Crossa J. (1990): Statistical analyses of multilocation trials. *Advances in Agronomy* 44: 55-85.
- Denis J. B. (1988): Two-way analysis using covariables. *Statistics* 19: 123-132.
- Finlay K.W., Wilkinson G.N. (1963): Analysis of Adaptation in a Plant-Breeding Programme. *Australian Journal of Agricultural Research* 14: 742-754.
- Gabriel K.R. (1971): Biplot Graphic Display of Matrices with Application to Principal Component Analysis. *Biometrika* 58: 453-467.
- Galwey N. (2006): *Introduction to mixed modelling: beyond regression and analysis of variance*, Chichester, England; Hoboken, NJ, Wiley.
- Gauch H.G. (1992): *Statistical analysis of regional yield trials: AMMI analysis of factorial designs*, Amsterdam, Elsevier.
- Gauch H.G. (1988): Model Selection and Validation for Yield Trials with Interaction. *Biometrics* 44: 705-715.
- Gauch H.G., Rodrigues P.C., Munkvold J.D., Heffner E.L., Sorrells M. (2011): Two New Strategies for Detecting and Understanding QTL x Environment Interactions. *Crop Science* 51: 96-113.
- 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.
- Gauch H.G., Zobel R.W. (1997): Identifying mega-environments and targeting genotypes. *Crop Science* 37: 311-326.
- Gauch H.G., Zobel R.W. (1990): Imputing Missing Yield Trial Data. *Theoretical and Applied Genetics* 79: 753-761.
- Gollob H.F. (1968): A Statistical Model Which Combines Features of Factor Analysis and Analysis of Variance Techniques. *Psychometrika* 33: 73-115.
- Gusmão L. (1985): An Adequate Design for Regression-Analysis of Yield Trials. *Theoretical and Applied Genetics* 71: 314-319.
- Hongyu K., García-Peña M., Araújo L.B., Dias C.T.S. (2014): Statistical analysis of yield trials by AMMI analysis of genotype × environment interaction. *Biometrical Letters* 51: 89-102.
- 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.
- Josse J., van Eeuwijk F., Piepho H.-P., Denis J.B. (2014): Another look at Bayesian analysis of AMMI models for genotype-environment data. *Journal of Agricultural, Biological, and Environmental Statistics* 19: 240-257.

- Kang M.S., Gauch H.G. (1996): Genotype -by- Environment Interaction, Boca Raton, CRC Press.
- 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.
- Lourenço V., Rodrigues P.C., Pires A.M., Piepho H.-P. (2017): A robust DF-REML framework for variance components estimation in genetic studies. *Bioinformatics* 33: 3584-3594.
- Malosetti M., Ribaut J.M., van Eeuwijk F.A. (2010): The analysis of multi-environment data: modeling genotype by environment and QTL by environment interaction. In: Monneveux P. & Ribaut J.M. (eds.) *Drought phenotyping in crops: from theory to practice*.
- 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.
- Mandel J. (1969): Partitioning of Interaction in Analysis of Variance. *Journal of Research of the National Bureau of Standards Section B-Mathematical Sciences* B 73: 309-&.
- 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., Gauch H.G., Madry W., Drzazga T., Rodrigues P.C. (2011): Yield Response of Winter Wheat to Spatial Conditions Using AMMI and Cluster Analysis. *Crop Science* 51: 969-980.
- Patterson H.D., Thompson R. (1971): Recovery of Inter-Block Information When Block Sizes Are Unequal. *Biometrika* 58: 545-&.
- Pereira D., Rodrigues P.C., Mejza S., Mexia J.T. (2012a): A comparison between Joint Regression Analysis and the AMMI model: a case study with barley. *Journal of Statistical Computation and Simulation* 82: 193-207.
- Pereira D., Rodrigues P.C., Mejza I., Mejza S., Mexia J.T. (2012b): Analyzing genotypes by environment interaction by curvilinear regression. *Scientia Agricola* 69: 357-363.
- Pereira D.G., Mexia J.T., Rodrigues P.C. (2007): Robustness of Joint Regression Analysis. *Biometrical Letters*,44: 105-128.
- Piepho H.P. (1997): Analyzing genotype-environment data by mixed models with multiplicative terms. *Biometrics* 53: 761-766.
- 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.
- Reymond M., Muller B., Tardieu F. (2004): Dealing with the genotypexenvironment 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.
- 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.

- 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.
- Rodrigues P.C. (2012): New strategies to detect and understand genotype-by-environment interactions and QTL-by-environment interactions. PhD Thesis, Faculty of Sciences and Technology, Nova University of Lisbon, Lisboa, Portugal.
- Rodrigues P.C., Pereira D.G., Mexia J.T. (2011): A comparison between JRA and AMMI: the robustness with increasing amounts of missing data. *Scientia Agricola* 68: 679-686.
- Romagosa I., van Eeuwijk F.A., Thomas W.T.B. (2009): Statistical analyses of genotype by environment data. In: Carena M.J. (ed.) *Cereals*. Springer, New York: 291-331.
- Romagosa I., Ullrich S.E., Han F., Hayes P.M. (1996): Use of the additive main effects and multiplicative interaction model in QTL mapping for adaptation in barley. *Theoretical and Applied Genetics* 93: 30-37.
- Searle S.R. (1971): *Linear models*, New York, Wiley.
- Searle S.R., Casella G., McCulloch C.E. (1992): *Variance components*, New York, Wiley.
- Spitters C.J.T. (1990): Crop growth models: their usefulness and limitations. *Acta Horticulture* 267: 349-368.
- Tardieu F. (2003): Virtual plants: modelling as a tool for the genomics of tolerance to water deficit. *Trends in Plant Science* 8: 9-14.
- van Eeuwijk F.A., Bustos-Kortsa D.V., Malosetti M. (2016): What Should Students in Plant Breeding Know About the Statistical Aspects of Genotype \times Environment Interactions? *Crop Science* 56: 2119-2140.
- 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.
- 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.
- 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: Kang M.S. & Gauch H.G. (eds.) *Genotype by Environment Interaction: New Perspectives*. Boca Raton: CRC Press.
- 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.
- 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.
- Verbeke G., Molenberghs G. (2009): *Linear mixed models for longitudinal data*, New York, Springer.
- Yan W., Kang M.S. (2003): *GGE biplot analysis: a graphical tool for breeders, geneticists, and agronomists*, Boca Raton, Fla., CRC Press.

- Yan W., Kang M.S. (2002): GGE biplot analysis: A graphical tool for breeders, geneticists, and agronomists, Boca Raton, Florida, CRC Press.
- Yang R.C., Crossa J., Cornelius P.L., Burgueno J. (2009): Biplot Analysis of Genotype x Environment Interaction: Proceed with Caution. *Crop Science* 49: 1564-1576.
- 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.
- 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.