Australian Journal of Crop Science AJCS
AJCS 8(4):640-645 (2014) ISSN:1835-2707

The usefulness of EM-AMMI to study the influence of missing data pattern and application to Polish post-registration winter wheat data

JakubPaderewski[1] and Paulo Canas Rodrigues[2,3]

[1]Department of Experimental Design and Bioinformatics, Warsaw Univ. of Life Sciences, Nowoursynowska 159, 02-766 Warsaw, Poland
[2]CMA–Center for Mathematics and Applications, Faculty of Sciences and Technology, Nova University of Lisbon, 2829-516 Caparica, Portugal
[3]Universidade Europeia, Laureate International Universities, Lisbon, Portugal

***Corresponding author: jakub.paderewski@omega.sggw.waw.pl**

Abstract

The study of genotype-by-environment interaction (GEI) is of key importance in plant sciences because an understanding of this allows a great improvement in complex phenotypic traits. Genotypes and environments constitute a two-way factorial design. The phenotypic data for these studies, usually arranged in two-way data tables with genotypes and environments (location-year combinations). In plant breeding programs some genotypes are often discarded and others included from year to year, which results in the presence of missing values in these data sets. Several options are available for dealing with missing values in two-way data tables. One of the most widely used alternatives is the imputation of the missing cells using an expectation-maximization (EM) algorithm together with the additive main effects and multiplicative interaction (AMMI) model. In this paper we present a simulation study to investigate the influence of the pattern of missing values on the efficiency of the expectation-maximization AMMI (EM-AMMI) algorithm. Four scenarios are considered: one with cells missing completely at random; and three patterns with cells not missing at random (block-diagonal pattern, diagonal pattern and block-diagonal pattern with checks). The results are compared in terms of precision to estimate the missing cells and genotype selection.

Keywords: Additive, main effects, multiplicative interaction model, expectation maximization, genotype-by-environment interaction, missing data, plant breeding programs, winter wheat.
Abbreviations: AMMI, additive main effects and multiplicative interaction; ANOVA, analysis of variance; EM, expectation-maximization; GEI_genotype-by-environment interaction; GGE_genotype main effect plus genotype-by-environment interaction; GLY_genotype x location x year; MCAR_missing completely at random; MET_multi-environment trials; NMAR_not missing at random; PCA_principal component analysis; RMSPD_root mean squares predictive difference; SHMM_shifted multiplicative model; SVD_singular value decomposition.

Introduction

Most agricultural research is done to improve quality and maximize the complex trait yield. Multi-environment trials (MET) provide the base for evaluating genetic improvements for yield and are essential to give recommendations of genotypes that could have wide or narrow adaptation (Gauch, 1992). The data from these trials are usually collected in two-way tables genotypes –by– environments (the environments may be the combinations of locations and years). The understanding the genotype-by-environment interaction (GEI) is one of the purposes for plant breeders and agronomists. Data from METs often have the presence of GEI, especially crossover interaction, where two different genotypes change in rank order of performance when evaluated in different environments. This phenomenon complicates the selection of superior genotypes because of difficulties in predicting the complex phenotypic trait yield for new locations and/or new years (Gauch, 1992; Yan and Kang, 2002). These crossover effects are also the reason to conduct trials in many locations and, sometimes, over several years. Since genotypes react differently to different environmental conditions, there is a need to select the best testing sites to identify superior and stable genotypes. Besides the standard regression based techniques (Finlay and Wilkinson, 1963; Pereira et al. 2012) and the mixed linear methodology (Piepho, 1997; Galwey, 2006), the most widely used methods to analyze data from METs are the additive main effects and multiplicative interaction (AMMI) model (Gauch, 1988, 1992, Paderewski et al. 2011) and the genotype main effect plus genotype-by-environment interaction (GGE) model (Yan and Kang, 2002). These methods (AMMI and GGE) -based on singular value decomposition (SVD) of matrices -break the interaction down into several components, allowing a separation between signal and noise. An important issue to be considered when analyzing data from METs is how to deal with missing values. In addition to the case of missing observations caused by natural factors (e.g. diseases, pests, animals, etc.), which are usually missing in a particular place and rarely missing in all replications, i.e. missing cells in the final table of means, the experiments often have some genotypes which are discarded and/or included in the trials (i.e. the missing cells have a clear pattern). Many statistical methods require complete data sets, e.g. SVD-based techniques such as principal component analysis (PCA),

shifted multiplicative model (SHMM), GGE and AMMI. To analyze incomplete data tables, where some combinations of genotypes and environments were not observed, researchers have three main options: (i) drop the genotypes with missing cells; or (ii) drop environments with missing cells; or (iii) impute the missing cells by the appropriate procedure, before the analysis. In this paper we will be interested in the latter option. When a data set from an MET has missing values, researchers often assumed that the missing values are located randomly (Arciniegas-Alarcón et al., 2010; Bergamo et al., 2008; Pereira et al. 2012; Rodrigues et al. 2012; Yan 2013). In these cases, imputation techniques are often used to estimate the data before analysis or try to infer the results without imputation. However, many times there is a clear pattern on the shape of the missing values. In fact, cases when the values are not missing at random (NMAR; Little and Rubin, 2002) and have a specific pattern are more common than cases where the values are missing completely at random (MCAR). To the best of our knowledge, the influence of a pattern NMAR on the estimation of the missing values in METs has not yet been studied. Gauch and Zobel (1990) proposed an expectation-maximization (EM)-AMMI algorithm to estimate the missing cells in two-way data tables. The EM-AMMI algorithm was also described by Gauch (1992), and was implemented in the open source software MATHODEL 3.0 (Gauch 2007). The EM-AMMI algorithm imputes the missing cells according to both the main effects and the interaction effects, based on a parsimonious AMMI model. It works as follows: (i) the initial values for the missing cells are calculated as the grand mean plus the main effects of rows (genotypes) and columns (environments); (ii) the parameters of the parsimonious AMMI model are computed; (iii) the adjusted means of the AMMI model are re-calculated using the new AMMI parameters; (iv) the values of the missing cells are replaced by the new estimates, accordingly to the parsimonious AMMI model; (v) the steps (*ii*) to (*iv*) are repeated until convergence. The aim of this paper is to study the influence of the pattern of missing values on the efficiency of the EM-AMMI algorithm. This efficiency will be compared for missing cells in four scenarios: NMAR and MCAR (block-diagonal pattern, diagonal pattern, and block-diagonal pattern with checks). This study will be conducted in terms of precision to estimate the missing cells and in terms of genotype selection.

Results

Leave-one-out cross-validation procedure

A complete two-way data table with Genotype–by–combination of Locations and Years is considered. The ANOVA table for a three-factor mixed model G×L×Y is presented in Table S2 and the AMMI analysis is presented in Table S1. The EM-AMMI analysis was conducted according to the leave-one-out cross-validation procedure and the RMSPD computed for every number of principal components (from 0 to 3). This resulted in RMSPD of 5.935, 5.473, 6.116 and 5.660, respectively for the four possible numbers of principal components, at the convergence criterion which consisted in checking if the maximum change in the predicted cell was less than 0.001. Therefore, the EM-AMMI with one principal component—EM-AMMI1—is the best option.

The convergence

The four patterns of missing cells (one MCAR and three NMAR, Fig. 1) were randomly generated with a different proportion of missing cells and those data sets were imputed by EM-AMMI with different number of principal components (10000 times for each combination of pattern, proportion of missing values and number of principal components). The EM-AMMI estimation of missing cells for different combinations was only considered when this iterative procedure converges, being the RMSPD only computed in these cases. The particular case of EM-AMMI0, unlike the higher members of this model family, does not involve any iterative calculations, so this 100% convergence is automatic. The limit for the number of iterations was 1000 and the convergence criterion was that the maximum change in predicted cells was less than 0.05. The cases that converged were: 100% for EM-AMMI0; more than 98.5% for EM-AMMI1, considering all combinations of the proportion of missing cells and shape of missing cells; more than 82% for EM-AMMI2; and more than 77.4% for EM-AMMI3. In this manner, a three-way classification table was obtained: the pattern of missing cells-by-proportion of missing cells-by-number of principal components, and analyzed according to a three-factor analysis of variance (ANOVA) model (Table 1).

Comparison of the model precision

As can be concluded from Table 1, all main effects, two-way interactions and three-way interactions are significant ($p < 0.001$). Overall, all these factors interact and influence the RMSPD resulting from the EM-AMMI imputation procedure. The mean RMSPD for every combination: pattern of missing cells–by–proportion of missing cells–by–number of principal components, was calculated with great precision because of the number of repetitions (greater than 7739) and the small mean square of error (Table 1). Fig. 2 shows the behavior of the RMSPD with the increasing proportion of missing cells for the four patterns and the four possible numbers of principal components. For the EM-AMMI0 it seems that no clear interaction between the proportion of missing cells and the pattern of missing cells (Fig. 2a) is present, because the values for the RMSPD are almost overlapped for the different patterns of missing cells. However, from the ANOVA table (Table S3 of the supplementary material), we can conclude that there is a significant interaction between these two factors ($p = 0.026$). The same interaction is present for the other possible numbers of principal components (with $p < 0.001$). For the best model—the EM-AMMI, the MCAR pattern generated the smallest RMSPD values. The differences to the other patterns were more significant when increasing the proportion of missing values (Fig. 2b). When analyzing the two models which overfit noise (EM-AMMI2, Fig. 2c; and EM-AMMI3, Fig. 2d) by considering more principal components than the optimal number, a completely different pattern is visible in the plots: when the proportion of missing cells is smaller than 20%, the diagonal pattern (NMAR2) showed a lower RMSPD; when the proportion of missing cells is greater than 20%, the MCAR showed a lower RMSPD (Fig. 2). The patterns block-diagonal (NMAR1) and block-diagonal with checks (NMAR3) showed similar

Table 1. ANOVA table presenting the influence of three factors (pattern of missing cells –by– proportion of missing cells –by– number of principal components) on the RMSPD value.

Source	SS (type III)	df	MS	F-value	p-value
MPtype[a]	155024	3	51674.8	4531.5	< 0.001
PC[b]	9556431	3	3185477.0	279340.9	< 0.001
Mprop[c]	260072	10	26007.2	2280.6	< 0.001
MPtype[a]×PC[b]	123341	9	13704.6	1201.8	< 0.001
MPtype[a]×Mprop[c]	88129	30	2937.6	257.6	< 0.001
PC[b]×Mprop[c]	119986	30	3999.5	350.7	< 0.001
MPtype[a]×PC[b]×Mprop[c]	90867	90	1009.6	88.5	< 0.001
Error	18744763	1643766	11.4		

[a)]Type of missing cells pattern; [b)]Number of principal components used in EM-AMMI; [c)]Proportion of missing cells in the data set. MPtype, pattern of missing cells; PC, Principal component; Mprop, proportion of missing cells.

behavior and the presence of checks does not seem to make much difference when carrying out imputation of missing values using the EM-AMMI algorithm. When analyzing the plots for the EM-AMMI2 and EM-AMMI3 models, unlikely behavior is visible for the patterns of missing values MCAR, NMAR1 and NMAR3 (Fig. 2c,d) for proportions of missing cells between 5% and 20%, where a clear peak for the RMSPD is observed. This is likely to be caused by an increase in the number of principal components, which represents an overestimating of the noise in the data. When increasing the proportion of missing cells, the influence of the overfit of noise caused by "too many" principal components decreases and, consequently, between 5 and 20% of missing cells the RMSPD decreases and starts to increase again, as expected, above 20% of missing cells. This overfit of noise does not affect the RMSPD for the diagonal pattern of missing cells (NMAR2).

Discussion

Standard statistical methods are usually not able to analyze datasets with missing data. Since the AMMI model is the result of a combination of two of the most widely used statistical procedures—ANOVA and PCA—, it is no exception. Moreover, when some genotypes are discarded in plant breeding programs and others included from year to year, the amount of missing data (with a strong pattern) in the two-way table can be huge. In this paper we have approached one of the most widely used techniques to deal with missing data in genotype-by-environment trials, the imputation procedure based on the EM-AMMI algorithm (Gauch and Zobel, 1990, Gauch, 1992). When using the EM-AMMI algorithm, the imputation considers both the additive component (i.e. main effects) and the multiplicative component (i.e. interaction). Gauch and Zobel (1990), when proposing the EM-AMMI, stated that 'No problems with numerical instability or local minims have been noted' and 'Nevertheless, further theoretical and empirical study of stability would be desirable'. Another approach, proposed for application in GGE biplot analysis, was developed by Yan (2013), where only the multiplicative component is of interest because a change in other model parameters 'will not affect the relative values and the rank of the genotypes in the environment'. Yan (2013) concluded that the validity of the predicted values seems to be dependent on the size of the two-way table and on the proportion of missing cells.

The simulation study presented in this paper considered four patterns of missing cells: one MCAR—which is quite unlikely in plant breeding programs, but more likely when the missing cells are due to natural causes such as diseases, pests or animals—and three NMAR cases: block-diagonal pattern, diagonal pattern and block-diagonal pattern with checks. The AMMI with one principal component being the more parsimonious model, the results shown that fewer multiplicative terms (underfit signal) result in a very similar RMSPD within different patterns of missing cells; and more multiplicative terms (overfit noise) result in similar shapes of MCAR, NMAR block-diagonal patter and NMAR block-diagonal pattern with checks, along with a percentage of missing cells with higher RMSPD between 5% and 20% of missing cells. The clear message is that, in general, there is a great penalty to be paid for fitting an EM-AMMI model that is too complex, i.e. with too many interaction principal components (Fig. 2). The particular case described in Fig. 1b shows a particularly common situation, a gradually changing

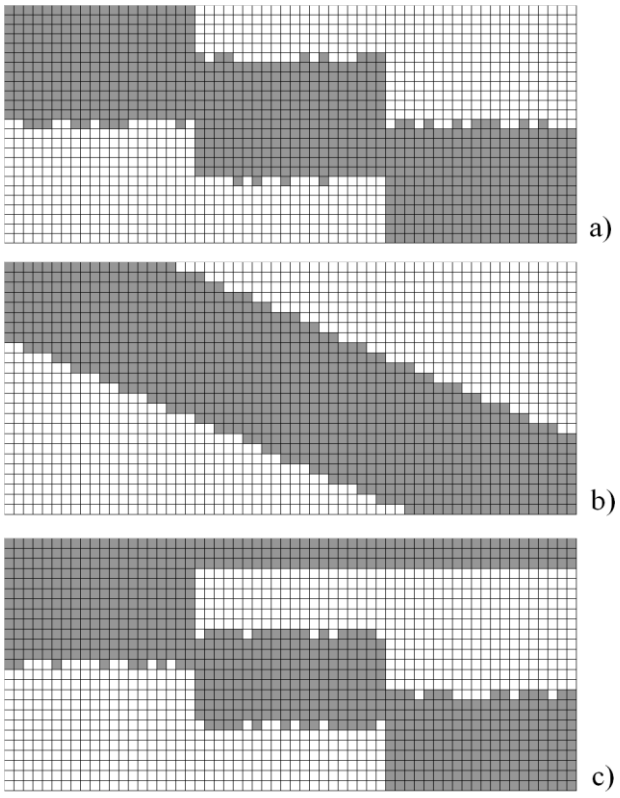

Fig 1. The three different NMAR patterns of missing cells (white squares) considered in this study.
Fig 1a) block-diagonal pattern
Fig 1b) diagonal pattern
Fig 1c) block-diagonal pattern with checks.

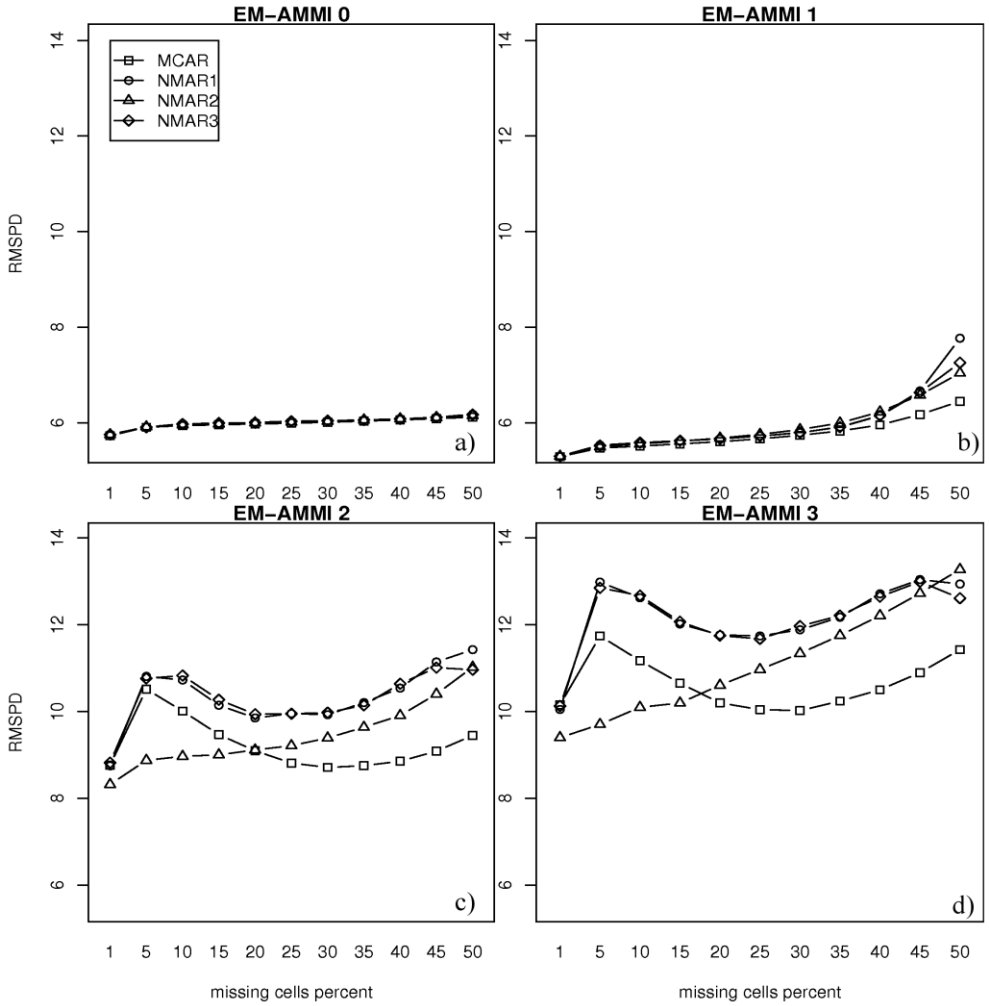

Fig 2. Behavior of the RMSPD for the combination pattern of missing cells–by–proportion of missing cells–by–number of principal components.
Fig 2a) EM-AMMI0
Fig 2b) EM-AMMI1
Fig 2c) EM-AMMI2
Fig 2d) EM-AMMI3.

roster of genotypes and environments over time. Although this is a tidy diagram, in practice plant breeding databases may have the genotypes and environments listed in an ecologically random order, i.e. have a pattern obtained from Fig. 1b by making several permutations of rows and columns. In those cases, the rows and columns can be arranged accordingly to reciprocal averaging scores, which automatically optimizes the placement of presences along the matrix diagonal (Gauch et al. 1977). When the database includes several blocks or subsets of the entire data matrix with few missing values, then TWINSPAN can be used to arrange the data (Hill et al. 1975, Hill 1979).

Materials and methods

Plant materials

The data set contains the post-registration trials made in Poland by the Research Center of Cultivar Testing (COBORU) in the growing seasons 2006/2007, 2007/2008 and 2008/2009. The trials were carried out at two levels of crop management intensity: standard (the standard fertilization suited to the soil conditions of a given station); and intensive (not used in this paper). The grain density ranged from 400 to 550 grains per m^2, depending on the cultivar and soil at a location. The trials were carried out in two factorial split-block designs (management levels were arranged in sub-blocks and within each sub-block cultivars were randomly allocated) with two replications. The size of sub-sub-plots was 11m by 1.5m and the harvesting area was 10m by 1.5m. The main types of soil in Poland are podzol and brown podzolic soils. The average annual precipitation for the whole country is 600mm. The data set includes the genotype × location × year (GLY) classification of post-registration trials of winter wheat. The genotypes changed from year to year, with the "best" kept for more than one year. The well-tested genotypes or genotypes with a lower yield are removed from the trials. Some genotypes are added in the following year and some are removed. Every year, the number of genotypes is similar but the GLY classification contains a large proportion of missing cells. Since, with time, the number of the same genotypes being tested decreases, the observed values of GEI combinations are close to the 'diagonal' (Fig. 1). To achieve the aim of this paper we used a complete data set so that we can dispose of the observed values in cells marked as missing and then use the EM-

AMMI algorithm. A complete data set (subset of the post-registration trials) contains 25 genotypes growing in 20 locations from 2007 to 2009. The combination of locations × years were treated as environments and had a two-way table of 25 genotypes and 60 environments.

Simulation of missing data

The influence of a given pattern in missing data was evaluated for four cases. The patterns considered, which occur more frequently in METs, were the MCAR pattern and three NMAR patterns: block-diagonal pattern, diagonal pattern and block-diagonal pattern with checks. These four patterns were randomly generated based on a complete data set with a different proportion of missing cells (with eleven levels: 1%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45% and 50%) and the missing cells imputed by EM-AMMI with a different number of principal components (from 0 to 3). Every combination of pattern shape by proportion of missing cells and by number of PCs used for imputation in EM-AMMI was repeated 10000 times. To obtain a two-way table with a MCAR pattern of missing values, after choosing the proportion of missing values (α), the missing cells were obtained by randomly deleting α% of the values. For that, a matrix with the same size of the original data, without replications, was created with random values and the positions of α% smallest values were deleted in the original data table. The necessary criterion of unambiguous computation of the EM-AMMI algorithm with three principal components, in order to have four observations in every row and every column (Gauch and Zobel, 1990), was fulfilled. Moreover, the pattern of observed values must connect the genotypes (and the environments) in each series of experiments. As the simulated data sets are not checked to confirm whether the genotypes and environments are connected we have decided to use the minimal number of observed values twice as the minimum suggested by Gauch and Zobel (1990). Therefore, data sets generated with an MCAR pattern that contain a genotype or an environment with less than eight observed values were discarded. When assuming that the missing values were NMAR, the patterns were created by analogy to a standard post-registration complete data set, where the locations are kept unchanged over the period of several years (Fig. 1). Those patterns often exist in multi-year trials, when some genotypes change from year to year. Three different cases are discussed:

Missing values are placed with a clear "diagonal" pattern by blocks corresponding to the years (Fig. 1a). The number of genotypes is denoted by *G* and the number of years is denoted by *Y*. The simulation for the location of the missing values was conducted as follows: (i) randomly order the genotypes; (ii) define α, the proportion of missing cells; (iii) assume that, for every combination of year and location, *int(αG)*—the integer part of *αG*—genotypes are observed and they are the closest to (i.e. centered in) genotype position number $(\alpha G+1)/2+G(1-\alpha)(y-1)/(Y-1)$, y=1,…,Y. Then, the genotypes between indexes $G(1-\alpha)(y-1)/(Y-1)+1$ and $G(1-\alpha)(y-1)/(Y-1)+\alpha G$, y=1,…,Y, are considered as observed. The remaining genotypes are missing. If the limits of the interval of indexes are not integers, the two external genotypes are marked as missing randomly, so that the proportion of missing cells is fulfilled.

For example, in a three-year data set, if α=0.6 and G=25, in the first year the first *int*(αG) genotypes (from 1 to 15) are marked as observed, in the second year the middle *int*(αG) genotypes (from 6 to 20) are observed and in the final year, the last *int*(αG) genotypes (from 11 to 25) are marked as observed. Other cells are missing.

Missing values are placed with a clear pattern far from the "diagonal" without taking into account the years (Fig. 1b). The number of environments (combinations location-year) is denoted as E. The simulation for the location of the missing values was done as follows: (i) randomly order the genotypes and the environments; (ii) define α, the proportion of missing cells; (iii) for each of environments calculate the difference between indexes of the cells (according to the new order of genotypes) and the position on the diagonal (for the j^{th} environment the diagonal is placed at genotypes' index equal to $1+(i-1)(G-1)/(E-1)$); (iv) the part of cells (genotypes by environments combinations) that have the difference closer to zero were marked as observed and others as missed. This pattern can be seen as what is observed in post-registration trials when the number of locations is small and number of years big.

Missing values are placed with a "diagonal" pattern by blocks corresponding to the years (as in the first case) but with some genotypes observed in all environments (Fig. 1c). There are also some rows without missing cells. Those patterns occur in trials with check cultivars. According to pre-registration trials run in Poland, only three genotypes appear in all years (this is usually used number of check cultivars). At the beginning of the simulations these three genotypes were randomly chosen to be observed in all years, and the pattern of missing values for other genotypes was obtained as described in the first case. The simulation was done as follows: (i) randomly order the genotypes; (ii) randomly choose the three genotypes which will appear in all environments; (iii) simulate a missing cells pattern for other genotypes, according to the first case described above.

The precision of the EM-AMMI estimation of the missing values was evaluated by comparing the estimated missing values with the values in the complete (original) data set. This is done by using root mean squares predictive difference (RMSPD, Gauch and Zobel, 1988, 1990, Dias and Krzanowski, 2003).

The analysis of the entire dataset

In real field experiments, researchers have the data set with missing cells and must decide on the procedure for analyzing that data. If the alternative selected is to impute missing values, this can be done by cross-validation (using some observations to estimate the model parameters and the rest for validation of that model, Stone 1974) and the model with minimum average RMSPD to the observed values (Gauch and Zobel 1988, 1990, Eastment and Krzanowski 1982, Dias and Krzanowski 2003) should be chosen. The EM-AMMI algorithm was run, taking into consideration all cells except the one being predicted and this was repeated for all cells (leave-one-out cross-validation procedure). The RMSPD for every number of principal components (from 0 to 3) was computed to evaluate the optimal number of interaction principal components.

Conclusion

When modeling two-way data tables, conducting a model diagnosis in order to choose the most parsimonious AMMI model (i.e. to choose the "right" number of multiplicative terms) is of key importance for AMMI analysis and to make the proper agricultural recommendations. In this paper three factors with a direct impact on model diagnosis were studied: the proportion of missing values, the number of principal

components in the AMMI model and the pattern of the missing cells. All three factors have shown significant interaction with each other with regard to the RMSPD between the observed and imputed values by the EM-AMMI procedure. The EM-AMMI algorithm could be used for NMAR patterns but a minor loss of estimation accuracy is to be reckoned with. A big penalty was observed when fitting an EM-AMMI model that is too complex, i.e. that has too many multiplicative terms.

Acknowledgments

Paulo C. Rodrigues acknowledges financial support from Portuguese National Science Foundation (Fundação para a Ciência e Tecnologia), through the projects PTDC/AGR-PRO/2335/2012 and PEst-OE/MAT/UI0297/2011 (CMA).

References

Arciniegas-Alarcón SA, Peña MG, Dias CTS, Krzanowski, WJ (2010) An alternative methodology for imputing missing data in trials with genotype-by-environment interaction. Biom Letters, 47: 1-14.

Bergamo GC, Dias CTDS, Krzanowski WJ (2008) Distribution-free multiple imputation in an interaction matrix though singular value decomposition. Sci Agric, 65: 422-427.

Cornelius PL (1993) Statistical tests and retention of terms in the additive main effects and multiplicative interaction model for cultivar trials. Crop Sci, 33:1186-1193

Dias C, Krzanowski WJ (2003) Model selection and cross validation in additive main effect and multiplicative interaction models. Crop Sci, 43: 865-873.

Eastment HT, Krzanowski WJ (1982) Cross-validatory choice of the number of components from a principal component analysis. Technometrics, 24: 73-77.

Finlay KW, Wilkinson GN (1963) Analysis of adaptation in a plant-breeding programme. Austr J Agric Res, 14: 742-754.

Galwey N (2006) Introduction to mixed modelling: beyond regression and analysis of variance. Chichester, England ; Hoboken, NJ, Wiley.

Gauch HG (1988) Model selection and validation for yield trials with interaction. Biometrics, 44: 705-715.

Gauch HG (1992) Statistical analysis of regional yield trials: AMMI analysis of factorial designs. Amsterdam, Elsevier.

Gauch HG (2007) MATMODEL version 3.0: open source software for AMMI and related analyses. Crop and Soil Sci, Cornell University, Ithaca, New York.

Gauch HG, Zobel RW (1988) Predictive and postdictive success of statistical analyses of yield trials. Theor Appl Genet, 76:1-10.

Gauch HG, Whittaker R, Wentworth T (1977) A comparative study of reciprocal averaging and other ordination techniques. J Ecol: 157-174.

Gauch HG, Zobel RW (1990) Imputing missing yield trial data. Theor Appl Genet, 79: 753-761.

Hill MO, Bunce R, Shaw M (1975) Indicator species analysis, a divisive polythetic method of classification, and its application to a survey of native pinewoods in Scotland. J Ecol: 597-613.

Hill MO (1979) TWINSPAN - a FORTRAN program for arranging multivariate data in an ordered two-way table by classification of the individuals and attributes Section of Ecology and Systematics, Cornell University: New York. 90 pp.

Little RJA, Rubin DB (2002) Statistical analysis with missing data (2nd ed.). New York: Wiley.

Paderewski J, Gauch H, Mądry W, Drzazga T, Rodrigues P (2011) Yield response of winter wheat to agro-ecological conditions using additive main effects and multiplicative interaction and cluster analysis. Crop Sci, 51(3): 969-980.

Pereira D, Rodrigues PC, Mejza S, Mexia JT (2012) A comparison between joint regression analysis and the AMMI model: a case study with barley. Stat Comp and Simul, 82: 193–207.

Piepho HP (1997) Analyzing genotype-environment data by mixed models with multiplicative terms. Biometrics, 53: 761-766.

Rodrigues PC, Pereira D, Mexia JT (2011) A comparison between JRA and AMMI: the robustness with increasing amounts of missing data. Sci Agric, 68: 679–686.

Stone M (1974) Cross-validatory choice and assessment of statistical predictions, J. R. Stat. Soc. B 36 (1) 111–147.

Yan W (2013) Biplot analysis of incomplete two-way data. Crop Sci, 53:48-57.

Yan W, Kang MS (2002) GGE biplot analysis: A graphical tool for breeders, geneticists, and agronomists, Boca Raton, Florida, CRC Press.