Constrained AMMI Model: Application to Polish Winter Wheat Post-Registration Data

Jakub Paderewski⋆ and Paulo C. Rodrigues

ABSTRACT

Constrained principal component analysis (C-PCA) describes a two-dimensional data table and assumes a linear dependence of the principal component scores on known additional parameters (i.e., explanatory matrices). In this study, we used C-PCA to generalize the additive main effects and multiplicative interaction (AMMI) model and propose the constrained AMMI model. The constrained AMMI model is interpreted and illustrated when (i) only the environmental principal component parameters have an explanatory data matrix, (ii) only the genotype principal component parameters have an explanatory data matrix, and (iii) both types of parameters have explanatory data matrices. The cross-validation procedure is adapted for model diagnosis. Data for winter wheat (*Triticum aestivum* L.) genotype × location × management × year grain yield, recorded in Poland from multienvironment trials conducted in the post-registration variety testing system, were analyzed and used for model comparison.

J. Paderewski, Dep. of Experimental Design and Bioinformatics, Warsaw Univ. of Life Sciences, Nowoursynowska 159, 02-766 Warsaw, Poland; P.C. Rodrigues, CAST, Faculty of Natural Sciences, Univ. of Tampere, Tampere, Finland, Dep. of Statistics, Federal Univ. of Bahia, Salvador, Brazil. Received 8 June 2017. Accepted 7 May 2018. ⋆Corresponding author (jakub.paderewski@gmail.com). Assigned to Associate Editor Lucía Gutiérrez.

Abbreviations: AMMI, additive main effects and multiplicative interaction; C-AMMI, constrained additive main effects and multiplicative interaction; C-PCA, constrained principal component analysis; C(G)-AMMI, genotypes constrained additive main effects and multiplicative interaction; C(GE)-AMMI, constrained genotype and environment additive main effects and multiplicative interaction; C(E)-AMMI, environments constrained additive main effects and multiplicative interaction; EPC, environment interaction principal component score; G × L × M × Y, genotype × location × management × year; GE, genotype × environment; GPC, genotype interaction principal component score; MET, multienvironment trial; PC, principal component; PCA, principal component analysis; RMSPD, root mean square prediction difference.

Field trials in agronomy are conducted with the aim of evaluating the crop production of different cultivars. Multienvironment trials (METs) (Crossa et al., 1991; Cooper and DeLacy, 1994; Löffler et al., 2005; Samonte et al., 2005) are field trials performed in multiple environments or under different management conditions, which are standard research tools to evaluate the rank of genotypes (Crossa et al., 1991) or evaluate specific adaptations (Romagosa and Fox, 1993). The results are reliable when the data are collected in several sites or seasons (i.e., in multiple environments) because the outcome should be repeatable across the range of environment types. Multienvironment trials can be used to determine what is repeatable across a set of environments, and the reasons for nonrepeatable (interaction) genotype

responses in different environments. They are a tool for describing the yield potentials and yield gaps (Lu and Fan, 2013; Chapagain and Good, 2015).

The genotype × environment (GE) interaction is frequently used in METs and represents the differential responses of genotypes across environmental conditions. The environmental conditions could be related to a control factor (e.g., diverse locations or management levels; Paderewski et al., 2016). The MET data must be analyzed by multidimensional analysis because the genotype responses differ in different environments. The information on interactions between cultivars and environmental conditions, such as management level, cropping systems, and agricultural environments, in addition to the mean yield, should be considered in cultivar recommendations for a given target region (Annicchiarico et al., 2010; Loyce et al., 2012).

One of the models most widely used for such analysis is the additive main effects and multiplicative interaction (AMMI) (Gauch, 2007, 2013; van Eeuwijk et al., 2016) model. However, the standard AMMI model is not appropriate in all cases. For example, the AMMI model needs all GE combinations to be observed (i.e., the two-way data table must be complete, and all the observations in the two-way data table have the same weight in the final model). If a dataset has missing values, some alternatives, such as data imputation (Gauch and Zobel, 1990; Arciniegas Alarcón et al., 2010; Rodrigues et al., 2011; Gauch, 2013; Paderewski, 2013) or model adaptation, to account for different numbers of replications in GE combinations (weighted AMMI model; Rodrigues et al., 2014) can be used. The pattern of missing values also influences the efficiency of the expectation-maximization AMMI procedure (Paderewski and Rodrigues, 2014) and the cultivar recommendations. On the other hand, if some observations differ in importance due to different error variances in the environments or due to data contamination, alternatives such as the weighted AMMI (Rodrigues et al., 2014) or robust AMMI (Rodrigues et al., 2016) models can be used.

One of the outputs of the AMMI analysis is the AMMI biplot, which shows the scores for the genotypes and for the environments, usually plotted in the first two interaction principal components (PCs) space (Gabriel, 1971; Bradu and Gabriel, 1978; Yang et al., 2009). The scores for the genotypes (genotype interaction PC scores [GPCs]) can be described according to the genotypic parameters, such as the genetic marker information, and the scores for the environments (environment interaction PC scores [EPCs]) can be described according to the environmental conditions, such as geographic location, rainfall, radiation, etc. (Paderewski et al., 2016). The description can use linear regression, where EPCs or GPCs are the dependent variables, and the different environmental or genotypic

parameters are the independent variables. However, the $R^2 < 1$, which means that the dependence is not exact. Thus, during the AMMI analysis with the PC scores, there are two sources of inaccuracy: (i) the use of only a few (usually one or two) PCs, and (ii) inaccurate EPC and GPC score description. The first source of inaccuracy can be estimated based on sum of squares for the interaction PCs. The second depends on the R^2 that describes the accuracy of the linear regression. However, the general precision is unknown.

Constrained principal component analysis (C-PCA; Amenta and D'Ambra, 2000) is a generalization of the principal component analysis (PCA) aimed at describing the two-dimensional table of variables (in columns) and individuals (in rows). Constrained PCA assumes a linear dependence of the PC scores on the parameters for the variables or for the individuals (the explanatory datasets). The parameters for the variables (a small number of them) could be used to explain (constrain) the PC scores for the variables, and/or the parameters for the individuals could be used to explain the PC scores for the individuals. Thus, the arrows in the biplot and/or point coordinates will have (a priori) a physical (and known) explanation.

Constrained PCA could be used to describe the genotype yield (or other phenotypic characteristics of interest) in different environments. The GPC and/or EPC would be described by explanatory sets of genotypic parameters and/or environmental parameters (Supplemental Fig. S1; Fig. 2 in Amenta and D'Ambra, 2000). Thus, the GPC scores and/or the EPC scores are functionally dependent on the chosen parameters (explanatory datasets). Hence, this method emphasizes the explanation of the observed genotype responses rather than the model fitting and descriptive ability of PCA.

In this study, we used C-PCA to generalize the AMMI model and to propose the constrained AMMI (C-AMMI) model, where we consider the genotypes as individuals and the environments as variables. The aim is to present an analysis that provides a physical interpretation of the GPC and EPC scores. The C-AMMI model is interpreted and illustrated when: (i) only the EPCs have an explanatory dataset, (ii) only the GPC parameters have an explanatory dataset, and (iii) both GPCs and EPCs have explanatory datasets. The application of the C-AMMI model and comparison with the standard AMMI model were made using the Polish winter wheat (*Triticum aestivum* L.) post-registration data obtained by the Research Centre of Cultivar Testing in Poland

MATERIALS AND METHODS
Materials
Bread wheat cultivars are developed to achieve wide environmental adaptation, but they still differ in environmental adaptations due to the interaction effects on yield associated

with biotic and abiotic resistances (Anderson, 2010). Due to the common crossover GE interactions in yield, the cultivar rankings within environments vary (Gan et al., 2007; Loyce et al., 2008; Mohammadi and Amri, 2013), and specific cultivar recommendations can improve yield (Mandal et al., 2010; Loyce et al., 2012).

The genotype × location × management × year (G × L × M × Y) grain yield data of the winter wheat dataset recorded in Poland from METs conducted in the post-registration variety testing system was considered (Paderewski et al., 2016). Since the G × L × M × Y dataset is highly unbalanced, a balanced subset was considered in this study. The subset without missing values consists of the 24 cultivars tested across the 20 trial locations (Cultivar Testing Stations of the National Centre of Cultivar Testing, COBORU, Poland) in the post-registration trials with two management intensities during the 2006–2007 and 2008–2009 cropping seasons. The locations of the experimental stations are described in a previous paper (Paderewski et al., 2016; for geographical coordinates, see Table 2; for a map of locations, see Fig. 1).

Additional information about genotypes and environments has been collected for this yield trial (COBORU, 2008, 2009, 2010). The explanatory variables for the G × L × M × Y grain yield data for cultivars are frost resistance, height, lodging, time to earing, wax maturity, and susceptibility to diseases.

For environments (the location × management × year combinations), the explanatory variables are

- Year: there are two dummy variables, Year 2007 (the variable was equal to 1 for Year 2007 and equal to 0 otherwise) and Year 2009 (equal to 1 for Year 2009 and equal to 0 otherwise). These variables describe the yield compared with that in Year 2008, so they express the impact of years.
- Geographical coordinates of locations, which express the impact of locations.
- Soil classification (according to the Polish bonitation system for arable land) and soil reaction in locations used for growing cultivars in the years, expressed as the impact of location × year combinations (because of changes in the exact position of the trials in the test stations, the soil in the location could change across years).
- Rainfall (accumulated rainfall specified for the four periods: winter, April and May, June, and July and August) per location and per year, also expressed as the impact of location × year combinations.
- Type of management (standard = 0, or high = 1). The standard crop management only included an N fertilization rate appropriate for the general nutrient status of the field. The high crop management included N fertilization at a 40 kg ha^{-1} higher rate than the appropriate application, fungicide use, application of growth regulator (trinexapac–ethyl), and foliar compound fertilization.

Constrained Principal Component Analysis

The PCA method describes the values in a two-dimensional table—for example, the description of the individuals (in the rows of the two-way data table) by the different variables (in columns of the two-way data table). The PCA algorithm computes the scores for the individuals and the scores for the variables to maximize the model fitting to the data for a given number of PCs. Constrained PCA requires that the PC scores for individuals and/or the PC scores for variables are linear combinations of external parameters (i.e., the external parameters explain the PCs). In this study, the genotypes are the individuals and the environments are the variables, so that the two-way data table describes the response of the genotypes (in the rows of the two-way data table) to the different environmental conditions (in the columns of the two-way data table). Thus, the GPC and EPC scores could be described by the genotypic and/or environmental variables included in the external explanatory datasets.

Let **Y** be the matrix, with the yield for the genotypes (in the rows) observed in each environment (in the columns). When two explanatory matrices are available, one with genotype characterizations (the **A** matrix, with genotypes in the rows and genotypic parameters in the columns, which will be the restrictions for the PCs) and another with environments characterizations (the **B** matrix, with environments in the rows and environmental parameters in the columns), the yield matrix can be decomposed into four matrices (Amenta and D'Ambra, 2000) using orthogonal projections of the studied matrix on the explanatory matrices space or perpendicular to the explanatory matrices space.

Let $\mathbf{P_M}$ be the matrix of orthogonal projections onto the subspace generated by any matrix **M**, that is, $\mathbf{P_M} = \mathbf{M}(\mathbf{M}^\mathrm{T}\mathbf{M})^{-1}\mathbf{M}^\mathrm{T}$, where superscript T is the transposition operator. The projection on the perpendicular subspace could be calculated as the difference between the identity matrix (**Id**) and the matrix of projections on the source matrix (without a perpendicular matrix calculation), that is, $\mathbf{P_M}^\perp = \mathbf{Id} - \mathbf{P_M}$. The equations for the decomposition are as follows:

$$\mathbf{Y_{A,B}} = \mathbf{P_A}\mathbf{Y}\mathbf{P_B}; \mathbf{Y_A}^\perp{}_{,\mathbf{B}} = \mathbf{P_A}^\perp\mathbf{Y}\mathbf{P_B}; \mathbf{Y_{A,B}}^\perp$$
$$= \mathbf{P_A}\mathbf{Y}\mathbf{P_B}^\perp; \mathbf{Y_A}^\perp{}_{\mathbf{B}}^\perp = \mathbf{P_A}^\perp\mathbf{Y}\mathbf{P_B}^\perp \qquad [1]$$

$$\mathbf{Y_{A,B}} + \mathbf{Y_A}^\perp{}_{,\mathbf{B}} + \mathbf{Y_{A,B}}^\perp + \mathbf{Y_A}^\perp{}_{,\mathbf{B}}^\perp = \mathbf{Y} \qquad [2]$$

where $\mathbf{Y_{A,B}}$ is the part of variability in matrix **Y** that is explained by both (i) genotypic explanatory data (**A**) and (ii) environmental explanatory data (**B**). This is the most important matrix. The $\mathbf{Y_{A,B}}$ matrix is the orthogonal projection of matrix **Y** on the subspace generated by genotype parameters in matrix **A** (orthogonal projection of the column vectors of matrix **Y** on the subspace of explanatory parameters in matrix **A**) and the orthogonal projection of such row vectors on the subspace generated by environmental parameters in matrix **B**. $\mathbf{Y_A}^\perp{}_{,\mathbf{B}}$ is the part of the variability of **Y** that is explained by the environmental explanatory data (**B**), but not by the genotypic explanatory data (**A**). This component is the orthogonal projection of **Y** on the subspace generated by matrix **B** and on the subspace generated by the perpendicular matrix for **A**. $\mathbf{Y_{A,B}}^\perp$ is the part of the variability of **Y** that is explained by the genotypic explanatory data (**A**), but not by environmental explanatory data (**B**). $\mathbf{Y_A}^\perp{}_{,\mathbf{B}}^\perp$ is the part of the variability of **Y** that is not explained by the genotypic and environmental explanatory datasets **A** and **B**.

AMMI Model

The AMMI model (Gauch, 1992, 2007, 2013; Paderewski et al., 2011, 2016) describes the mean response from a two-factor experimental design organized in a two-way data table. The first factor denotes genotypes, whereas the second factor denotes environments (locations in the simplest case), and can be written as

$$\sum_{t=1}^{T} u_{i,t}\lambda_t v_{t,j} = \overline{Y}_{i,j} + m - g_i - e_j$$

where $u_{i,t}$ is the GPC parameter of the tth interaction PC on the ith genotype, $v_{t,j}$ is the EPC parameter of the tth PCs on the jth environment, λ_t is the singular value of singular value decomposition, $\overline{Y}_{i,j}$ is the mean yield of ith genotype in jth environment, m is the overall mean, g_i is the main effect of the genotypes, and e_j is the environmental main effect. The square root of λ_t jointly with the appropriate PC parameters (GPC and EPC for symmetric scaling; Yan and Kang, 2003) can be used to obtain the PC scores (Gauch, 1992). The AMMI analysis consists of two steps: (i) subtracting the overall mean and main effects (double centering), and (ii) singular value decomposition of the GE interaction matrix obtained in the first step. For graphical illustration and data interpretation, biplots can be made (Gabriel, 1971).

Constrained AMMI Models

C(GE)-AMMI Model

Here, we propose the generalization of the C-PCA to be used in the context of the AMMI model, (i.e., the C-AMMI model). This analysis consists of two steps. First, the data matrix with the phenotypic data (e.g., yield) is double centered (i.e., the genotypic and environmental main effects are removed from the data with an ANOVA model, as in the standard AMMI model). Then, C-PCA is applied to the resulting data table (i.e., to the residuals of the ANOVA model in the first step), replacing the standard singular value decomposition used in the standard AMMI model.

Let **Y** be the matrix with phenotypic data (e.g., yield) containing the genotypes in rows and the environments in columns, and **Z** be the resulting double-centered matrix (i.e., the GE interaction matrix in the standard AMMI model). When two explanatory matrices are used—**A**, with rows describing the GPC scores − characterization (constraints) for the GPC scores (constraints), and **B**, with rows describing the EPC scores − characterization (constraints) for the EPC—the GE interaction matrix **Z** can be decomposed into four matrices as follows:

$$\mathbf{Z} = \mathbf{Z}_{\mathbf{A,B}} + \mathbf{Z}_{\mathbf{A}^\perp,\mathbf{B}} + \mathbf{Z}_{\mathbf{A,B}^\perp} + \mathbf{Z}_{\mathbf{A}^\perp,\mathbf{B}^\perp} \quad [3]$$

where $\mathbf{Z}_{\mathbf{A,B}}$ is the part of variability in the GE interaction matrix **Z** that is explained by both (i) genotypic explanatory data and (ii) environmental explanatory data. The other three components in Eq. [3] have less importance as, according to Eq. [2], they reflect the variability contained in the **Z** matrix associated with the data fitting ability of the constrained genotype and environment AMMI [C(GE)-AMMI] model if there was no explanatory matrix **A** for genotypes and/or no explanatory matrix **B** for the environments.

C(G)-AMMI and C(E)-AMMI Models

Particular models are also considered for cases when only environmental variables or only genotypic characteristics are available. That is, the C-AMMI models without constraining matrix **B** [C(G)-AMMI, with only parameters that will explain the GPC scores] or without constraining matrix **A** [C(E)-AMMI, with only parameters that will explain the EPC scores].

The results can be seen in the biplot. The interpretation is similar to the AMMI analysis, that is, the neighboring genotypes showed similar interaction patterns to the environmental conditions. The environments with similar EPC scores generated similar interaction effects with the genotypes. In addition, the genotypes with coordinates consistent with some of the environments (GPCs and EPCs are in the same direction), showed positive interactions with those environments. Therefore, the GPC and EPC have explanations (by linear relationships from the explanatory variables).

The tests used in the AMMI analysis for the significance of PCs cannot be used for C-AMMI because the matrix $\mathbf{Z}_{\mathbf{A,B}}$ in Eq. [3] is very specific (i.e., only the first few PCs have nonzero singular values, Table 1). If P_G is the number of explanatory parameters of genotypes, and P_E is the number of environmental explanatory parameters, then the number of PCs of the $\mathbf{Z}_{\mathbf{A,B}}$ matrix is less than or equal to the minimum of P_G and P_E. Thus, the df is different from the AMMI model. Moreover, we assume that the experimental error should be more concentrated in the three other matrices (especially in $\mathbf{Z}_{\mathbf{A}^\perp,\mathbf{B}^\perp}$) in the decomposition (because of the physical explanation), and the $\mathbf{Z}_{\mathbf{A,B}}$ matrix should have a low noise level.

The exhaustive leave-one-out cross-validation procedure (Geisser, 1993, Paderewski and Rodrigues, 2014; Hadasch et al., 2017) was adapted and generalized to evaluate the prediction ability of AMMI and C-AMMI. Because the algorithm divides the original sample into a training set and a validation set in all possible ways, the model ratings given by the cross-validation could be treated as the total population sampling, and the validation should be treated as the final mark. Since the AMMI model does not account for missing values, an adaption was considered as follows: when one cell from the complete GE matrix is moved to the validation dataset, that cell in the GE matrix (that is currently missing) is replaced by the mean yield of all entries in the two-way data table. Thus, the dataset in hand is the complete GE matrix and, at the same time, one value is not used for the modeling. The leave-one-out cross-validation procedure was conducted, and the root mean square prediction differences (RMSPDs) (Gauch and Zobel, 1988, 1990; Paderewski, 2013; Paderewski and Rodrigues, 2014; Hadasch et al., 2017) were calculated, meaning the difference between the original cell value, validation set, and model estimation value was calculated for each cell separately, one by one, and the square root of mean squares is the RMSPD value. A higher RMSPD represents a lower ability for model prediction.

The RMSPD statistic was also used for the explanatory variable selections. The RMSPD statistic was calculated to select the C(GE)-AMMI model, considering a backward selection where each explanatory variable was singularly removed. The RMSPDs of the impoverished models were compared with the one from the full model. Thus, the significance of the environmental and genotypic variables was tested.

Mega-Environments

The locations with similar environmental conditions ordered genotypes (e.g., based on yield) in a similar way. Such location groups are called mega-environments (Gauch and Zobel, 1997; Löffler et al., 2005; Chenu et al., 2011). The procedure for constructing mega-environments can be based on a single winner (the same winning genotype for each location), on a group of top genotypes (similar composition of the top-yielding group of genotypes for each location), or on a similar order of all genotypes. The dataset under consideration in this paper is a four-way classification, and the winner for the single location depends on the year and management level, so the mega-environments were created based on the EPC scores. Aiming to compare the results with previous studies (based on AMMI1 analysis; Paderewski et al., 2016), the locations were clustered according to the averaged EPC1 scores by the cluster analysis using Ward's method.

Simulation Study

A simulation study was considered, where the data were obtained from the observed data, keeping the number of genotypes and environments, and the genotype and environmental explanatory sets of variables. The first two PCs of matrix $\mathbf{Z}_{A,B}$ coming from the observed yield analysis are considered to be the true GE interaction effects in our simulation study. Therefore, the C(GE)-AMMI model with two PCs is appropriate to fit the simulated data. We assume that the remaining interaction (variability described by the remaining PCs of $\mathbf{Z}_{A,B}$ and by the matrices $\mathbf{Z}_{A^\perp,B}$, $\mathbf{Z}_{A,B}^\perp$, and $\mathbf{Z}_{A^\perp,B}^\perp$) is associated with noise (the SD for the error term is 0.54). The simulated datasets were created as the sum of the true GE interaction effects (i.e., the signal. the same for all simulated sets) and the random noise (different for each simulation run), which is assumed to follow a normal distribution with a mean of zero and a SD of 0.54. For each of the models [C(GE)-AMMI with zero, one, two, or three PCs, C(G)-AMMI 2, C(E)-AMMI 2, and AMMI 2], the simulation was repeated 50 times, and the RMSPD was computed to diagnose the model prediction ability. The aim of the simulation study was to check if the proposed approach of the cross-validation method is efficient in model diagnosis.

RESULTS AND DISCUSSION
Simulation Study

The simulation study shows that choosing the model with the smallest of RMSPD statistic is the right choice; thus, the RMSPD statistic is a good indicator during the model diagnosis. The RMSPD for C(GE)-AMMI models with zero to three PCs were 0.5906, 0.5553, 0.5520, and 0.5548, respectively, and the lower value was obtained for two PCs. The RMSPD values for the C(G)-AMMI, C(E)-AMMI and AMMI models were equal to 0.5564, 0.5699, and 0.5789, respectively. The SDs of the means were 0.0013, 0.0012, 0.0009, 0.0011, 0.0011, 0.0011, and 0.0018. The best model is C(GE)-AMMI 2, which uses both the additional explanatory datasets. Thus, the RMSPD found the model that was used to create the simulated dataset. Similar levels of RMSPDs for METs and RMSPDs for simulated sets were the consequence of reflecting the trial character.

The AMMI Model

The AMMI analysis was applied to the phenotypic data for location × management × year combinations (treated as environmental conditions). Some of the AMMI1 results from the same data were described in Paderewski et al. (2016) in a different context. The first two GPC scores and EPC loadings are presented in Fig. 1a and 1b, respectively. The reason that the biplot was decomposed into two plots is the large number of environmental conditions (120 location × management × year combinations). The genotypes, Alcazar, Rapsodia, Boomer, Anthus, Kris, and Batuta (with negative GPC1), had positive interaction effects of the yield with the environmental conditions that occurred in 2008 (which also had a negative EPC1), but negative interactions with those in 2009 (environmental conditions which positive EPC1). The rest of the genotypes (especially Mewa, Legenda, Sukces, Smuga, Turnia, and Zyta had positive or close to zero interaction effects with 2009 but were negative with 2008. The management levels had close to zero interaction effects, which means that the management intensity did not influence the genotype ranking. The environmental scores seem to be separated on the upper left and lower right sides of Fig. 1b. The location Nowa Wieś Ujska for 2007 (for both management levels) showed very different scores than other locations, so it might be seen as an outlier.

The C(GE)-AMMI Model

Given the explanatory variables available for genotypes and environments, the GE interaction matrix was decomposed into the four matrices in Eq. [3]. The sum of squares for those matrices was 144 for $\mathbf{Z}_{A,B}$ and 117, 256, and 439 (Table 1) for the other three matrices in Eq. [3]. Thus, the environmental and genotypic explanatory variables account for 12% of the variability in the phenotypic data.

Aiming at properly choosing the number of PCs, the RMSPD was calculated for the C(GE)-AMMI model with different numbers of PCs to access the prediction ability of the model. Considering the number of PCs in the AMMI model from zero (model without interaction) to three, the RMSPD obtained values were 0.611, 0.5821, 0.5801, and 0.5851, respectively (i.e., a model with two PCs should be considered).

The first two PCs explain 81.1% of $\mathbf{Z}_{A,B}$ variability (69.5 and 11.6% for PC1 and PC2, respectively). At the same time, the first two PCs explain 12% of the total sum of squares of the $\mathbf{Z}$ matrix (Table 1). The environmental variables explained less of the variation in the phenotypic data than the genotypic characterizations, but most of the variability (45%) was the result of experimental random error or was not accounted for by genotypic and environmental parameters.
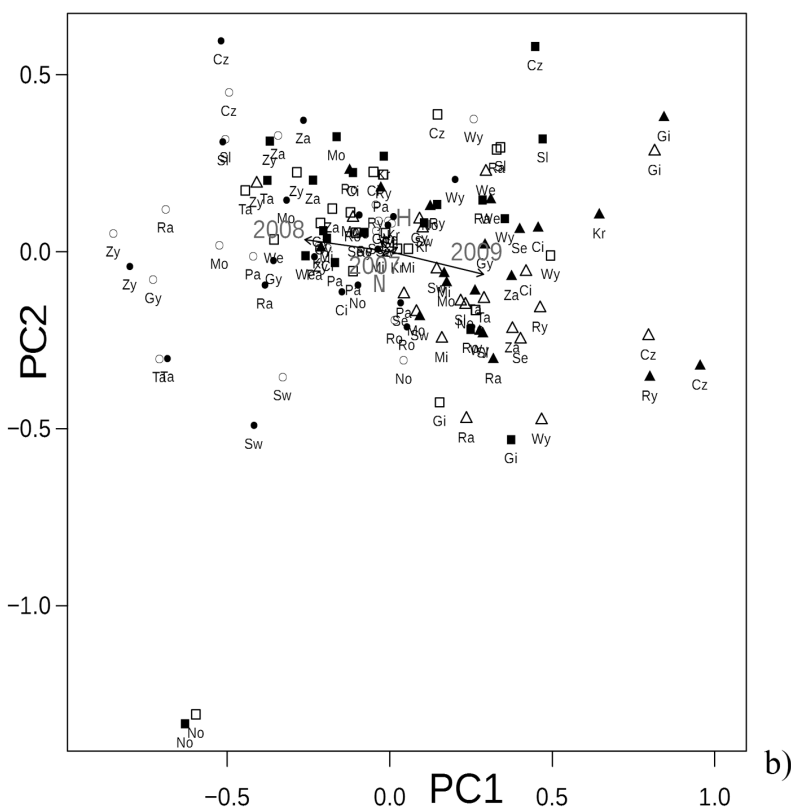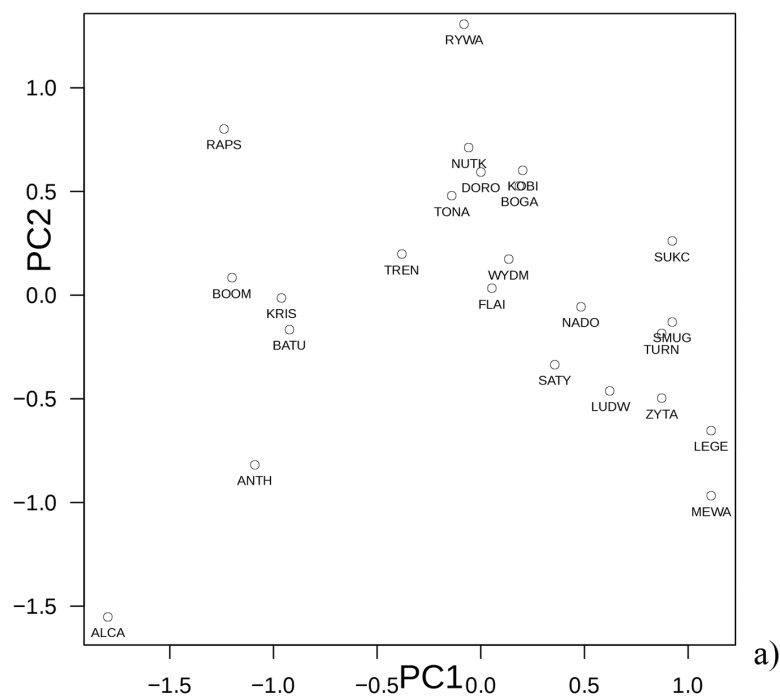
a)

b)

Fig. 1. The (a) genotype and (b) environment principal component (PC) scores according to additive main effects and multiplicative interaction analysis of winter wheat multienvironment trial data. Each location (Subplot b) consists of the six environmental conditions: squares for 2007, circles for 2008, and triangles for 2009, with open markers for standard management and filled markers for intensive management. Arrows indicate the averaged parameters for the years (2007, 2008, and 2009) and management intensity (normal or high). The cultivars: ALCA, Alcazar; ANTY, Anthus; BATU, Batuta; BOGA, Bogatka; BOOM, Boomer; DORO, Dorota; FLAI, Flair; KOBI, Kobiera; KRIS, Kris; LEGE, Legenda; LUDW, Ludwig; MEWA, Mewa; NADO, Nadobna; NUTK, Nutka; RAPS, Rapsodia; RYWA, Rywalka; SATY, Satyna; SMUG, Smuga; SUKC, Sukces; TONA, Tonacja; TREN, Trend; TURN, Turnia; WYDM, Wydma; ZYTA, Zyta. The trial locations: Ci, Cicibór; Cz, Czesławice; Gi, Głębokie; Gy, Głubczyce; Kr, Krościna Mała; Mo, Marianowo; Mi, Masłowice; No, Nowa Wieś Ujska; Pa, Pawłowice; Ra, Radostowo; Ro, Rawinowo; Ry, Rychliki; Se, Seroczyn; Sl, Słupia; Sw, Świebodzin; Ta, Tarnów; We, Węgrzce; Wy, Wyczechy; Za, Zadąbrowie; Zy, Zybiszów.

Comparisons between Models

When considering the models, C(G)-AMMI, C(E)-AMMI and AMMI, with two PCs, the RMSPD values obtained were 0.576, 0.574 and 0.562, respectively (Supplemental Table S1). When the C(GE)-AMMI was compared with the other models, the conclusion was that the C(GE)-AMMI 2 model (and, consequently, the C(GE)-AMMI 2 biplot) was less effective in predicting the phenotypic data. The best option, with two PCs,

was the AMMI model. This might be because the small collection of additional parameters (explanatory datasets for genotypes and environments) does not provide a good representation of the genotypic and environmental characteristics, and should be improved in the future.

During the AMMI biplot interpretation, i.e., when the estimated pattern of genotype response to the environmental condition is explained by some genotype traits or environment parameters, an additional error that is

Table 1. Sum of squares for the constrained additive main effects and multiplicative interaction (C-AMMI) principal components (PCs) with or without genotypes and environmental explanatory variables.

Component†	C(GE)-AMMI	C(G)-AMMI	C(E)-AMMI	AMMI
$Z_{A,B}$ total	144.0	419.7	261.1	975.5
PC1	100.2	204.0	134.3	252.4
PC2	16.7	59.6	29.5	86.9
PC3	9.7	56.9	23.0	77.8
PC4	9.1	44.2	21.0	68.0
PC5	5.5	32.4	19.2	61.7
PC6	2.8	22.5	11.6	54.8
PC7–PC11	–	–	22.6	195.9
PC12–PC23	–	–	–	177.9
$Z_{A^\perp,B}$ total	117.1	555.9	–	–
PC1	40.7	94.4	–	–
PC2	20.6	70.6	–	–
PC3	16.1	53.0	–	–
PC4	15.7	48.3	–	–
PC5	6.9	40.7	–	–
PC6	5.9	39.8	–	–
PC7–PC11	11.3	129.4	–	–
PC12–PC17	–	79.7	–	–
$Z_{A,B^\perp}$ total	275.6	–	714.4	–
PC1	106.4	–	130.6	–
PC2	48.8	–	73.6	–
PC3	40.6	–	63.6	–
PC4	37.9	–	53.2	–
PC5	24.3	–	52.1	–
PC6	17.6	–	44.2	–
PC7–PC23	–	–	297.0	–
$Z_{A^\perp,B^\perp}$ total	438.8	–	–	–
PC1	66.7	–	–	–
PC2	62.0	–	–	–
PC3	41.8	–	–	–
PC4	38.6	–	–	–
PC5	36.6	–	–	–
PC6	28.8	–	–	–
PC7–PC17	164.4	–	–	–

† The interaction components according to Eq. [3].

‡ The C-AMMI models according to the occurrence of explanatory matrices—for genotypes and environments [C(GE)], for only genotypes [C(G)], and for only environments [C(E)].

not contained in the RMSPD statistic appears, which is not the case for the C(GE)–AMMI model and the biplot. Therefore, the obtained results did not discredit the C(GE)–AMMI model, and this model seems to be a good choice. Moreover, for the C(GE)–AMMI and C(E)–AMMI models, the EPC scores have an explicit explanation (based on the environmental explanatory variables), while in the AMMI and C(G)–AMMI models, it needs to be found, and represents an extra and subjective source of possible inaccuracy not accounted for when computing the RMSPD. The C(GE)–AMMI also includes the genotype explanatory variables (potential reasons for different adaptability), which allows for the characterization of similar adaptability patterns of types of genotypes. Unfortunately, the biplot interpretation is subjective, and cannot be used to perform any statistical test.

The results of the RMSPD statistic are only binding for comparisons within one model with different numbers of components or for the comparison of different models if the yield estimation (numerical values) would serve another statistical analysis. On the other hand, when the analysis is to be used for making a simplistic conclusion (e.g., from the biplot), it is worthwhile to improve this next stage (by adding the explanation for PCs) at the expense of slightly lower RMSPD values for the model. This is why the small advantage of the AMMI model measured by the RMSPD statistic is not crucial.

Tests for the Explanatory Variables

The RMSPD statistic was used for the explanatory variable selection (both, environmental and genotypic variables). The RMSPD was calculated for the C(GE)–AMMI 2 model, considering a backward selection where each explanatory variable was removed. The RMSPD for the subsets of genotype explanatory variables without: frost resistance, height, lodging, time to earing, wax maturity and susceptibility to diseases were 0.5802, 0.5900, 0.5803, 0.5807, 0.5803 and 0.5836, respectively. Since all sub-models had a higher RMSPD than the full model, all genotypic explanatory variables were kept in the model. Additionally, the environmental explanatory variables were tested in a similar way with RMSPD values (for subsets of explanatory variables that consist of all variables except one) equal to 0.5816, 0.5841, 0.5807, 0.5849, 0.5804, 0.5802, 0.5803, 0.5813 0.5811, 0.5803 and 0.5813, respectively (according to the order of the variable in the materials and methods section). Since all sub-models had a higher RMSPD than the full model, all environmental explanatory variables were kept in the model.

Interpretation of the Genotypes

The interpretation of C(GE)–AMMI is very similar to that of the AMMI analysis. The AMMI2 biplot for the C(GE)–AMMI analysis (Fig. 2a and 2b) showed the interaction pattern, and after considering the main effects of genotypes, it allows the adaptability to be specified. The main difference compared with the standard AMMI2 biplots is that, after the C(GE)–AMMI analysis, the PCs have a physical interpretation.

In the C(GE)–AMMI analysis, the first two GPCs (GPC1 and GPC2) have a physical interpretation because the PC scores are a linear combination of the genotype and environmental parameters:

$$GPC1 = 19.476 - 0.036FR + 0.069H - 0.219L \\ + 0.236TE - 0.296WM - 0.373SC \quad [4]$$

$$GPC2 = -1.105 - 0.062FR - 0.004H - 0.218L \\ + 0.32TE - 0.133WM - 2.543SC \quad [5]$$
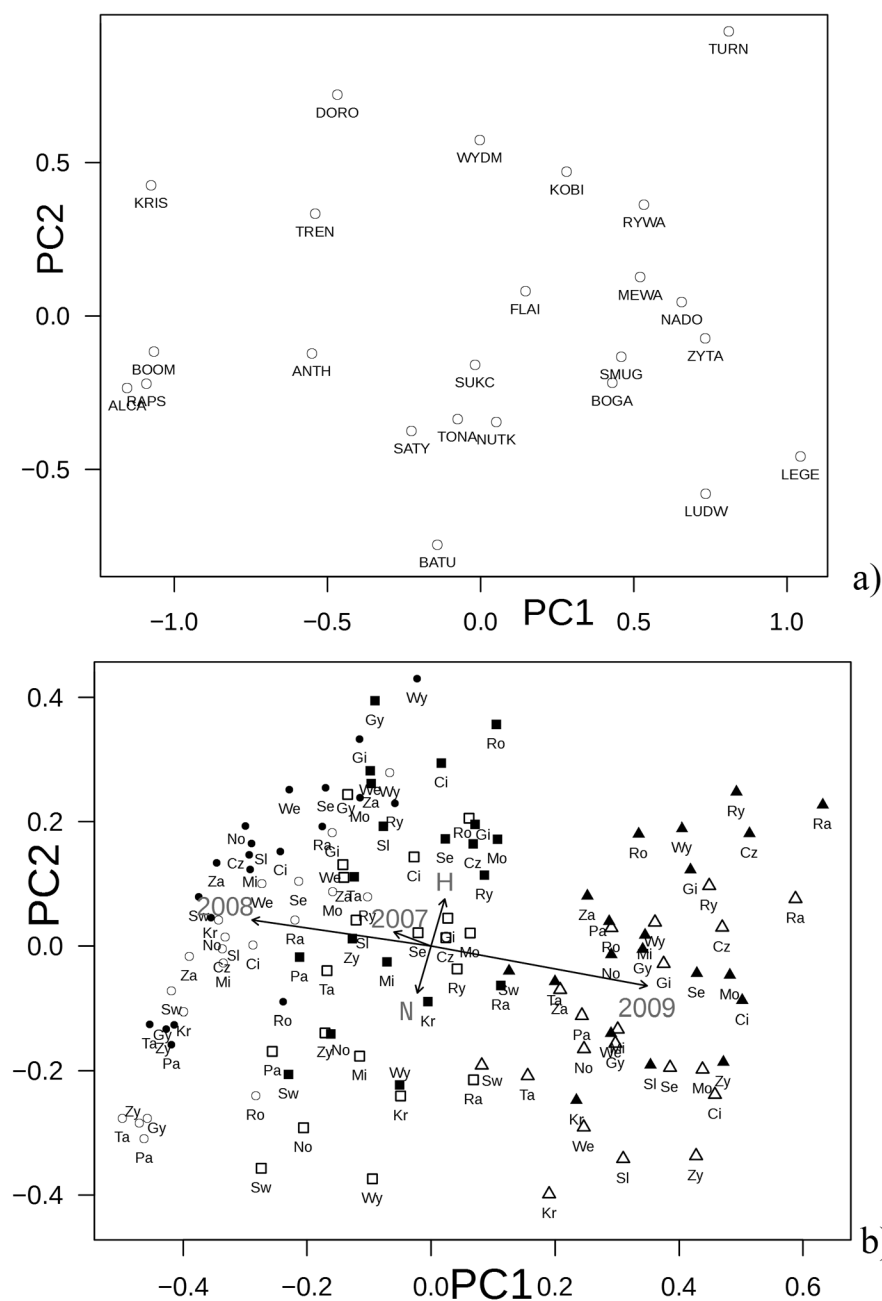
Fig. 2. The (a) genotype and (b) environment principal component (PC) scores according to the constrained genotype and environment additive main effects and multiplicative interaction [C(GE)-AMMI] analysis of winter wheat multienvironment trial data. Each location (Subplot b) consists of the six environmental conditions: squares for 2007, circles for 2008, and triangles for 2009, with open markers for standard management and filled markers for intensive management. Arrows indicate the averaged parameters for the years (2007, 2008, and 2009) and management intensity (normal or high). The cultivars: ALCA, Alcazar; ANTY, Anthus; BATU, Batuta; BOGA, Bogatka; BOOM, Boomer; DORO, Dorota; FLAI, Flair; KOBI, Kobiera; KRIS, Kris; LEGE, Legenda; LUDW, Ludwig; MEWA, Mewa; NADO, Nadobna; NUTK, Nutka; RAPS, Rapsodia; RYWA, Rywalka; SATY, Satyna; SMUG, Smuga; SUKC, Sukces; TONA, Tonacja; TREN, Trend; TURN, Turnia; WYDM, Wydma; ZYTA, Zyta. The trial locations: Ci, Cicibór; Cz, Czesławice; Gi, Głębokie; Gy, Głubczyce; Kr, Krościna Mała; Mo, Marianowo; Mi, Masłowice; No, Nowa Wieś Ujska; Pa, Pawłowice; Ra, Radostowo; Ro, Rawinowo; Ry, Rychliki; Se, Seroczyn; Sl, Słupia; Sw, Świebodzin; Ta, Tarnów; We, Węgrzce; Wy, Wyczechy; Za, Zadąbrowie; Zy, Zybiszów.

where FR is frost resistance, *H* is height, *L* is lodging, TE is time to earing, WM is wax maturity, and SC is susceptibility to disease.

Some of genotypes showed very similar responses (Fig. 2a) in that they had similar trait values in the genotype explanatory matrix, and the distance between those values determined the similar yield estimations (e.g., Boomer, Rapsodia, and Alcazar). Some other genotypes showed different responses (e.g., the trio Boomer, Rapsodia, and Alcazar with Legenda or Turnia).

More generally, the genotypes that had lower GPC1 and GPC2 were more resistant to frost, were shorter, had less lodging, and had earlier earing but reached wax maturity later and had higher resistance to disease (Eq. [4 and 5]).

Interpretation of the Environments

The AMMI2 biplot for C(GE)–AMMI analysis was done without genotypes to increase the readability (Fig. 2b). There were 120 environmental conditions (location × management × year combinations), which are denoted as squares in the plots. Instead of those 24 genotype points, Fig. 2b depicts the arrows of the average EPC scores (EPC1 and EPC2) for each year, and the averaged EPCs scores for the management level. The main factor that determines the yield is the year. The management has only a small impact on the interaction pattern.

The first two EPCs (EPC1 and EPC2) can be written as follows:

$$EPC1 = -4.82 + 0.025E + 0.078N - 0.07Y7 - 0.26Y8$$
$$+ 0.33Y9 - 0.057Sl + 0.086pH - 0.0003Rw \quad [6]$$
$$- 0.0016R1 + 0.0005R2 + 0.0006R3 + 0.04Mn$$

$$EPC2 = -3.63 + 0.031E + 0.058N - 0.05Y7 - 0.05Y8$$
$$+ 0.10Y9 - 0.065Sl + 0.052pH + 0.0013Rw \quad [7]$$
$$- 0.0032R1 - 0.0034R2 + 0.0007R3 + 0.15Mn$$

where E and N in stand for geographical coordinates in degrees east and north, respectively; Y7, Y8, and Y9 stand for the years 2007, 2008, and 2009, respectively; Sl is the soil classification; pH indicates the soil reaction; and Rw, R1, R2, and R3 are the rainfall accumulated in four seasons (winter, April and May, June and July, respectively).

According to the EPC description in Eq. [6 and 7], the environments located in the northeast part of Poland (because of the positive coefficient of the latitude and longitude [N and E]) with weaker (negative coefficient of Sl in Eq. [6]) and more acidic soil, lower rainfall during winter and in April and May, higher rainfall during June, July, and August, and only standard management resulted in higher EPC1 values. Most of those traits seem to be characterized by poor locations. Moreover, higher EPC1 values were found in the years 2007 or 2009 (i.e., those years that had lower yield than 2008 by 0.13 t ha^{-1}). The influence on EPC2 was similar (Eq. [7]), except for the winter rainfall and rainfall during June, which had an opposite influence. This means that higher values of EPC1 and EPC2 could be interpreted as weaker environmental conditions.

The averaged EPC scores can be found in the trellis plot (Supplemental Fig. S2), aiming at a clearer (than on the standard biplot, Fig. 2b) presentation and interpretation of the effect of years on genotype yields in individual locations. Each subplot contains the data connected with a single location. It contains (0,0) points, averaged EPC1 and EPC2 scores for the appropriate location, and averaged EPCs for each location × year combination for that location. Since there were no explanatory variables describing any of the management × location, management × year, or management × location × year combinations (interaction of management with other environmental factors), the influence of management was the same for each location and year. The lack of this variable is justified by a very weak interaction with management for this dataset (Paderewski et al., 2016). That is, the sum of squares for the genotype × management and genotype × management × other factors effects were relatively low compared with the genotype × location or × year effects.

Each location had both positive and negative EPC1 values depending on the year. However, for example, for location Świebodzin, the EPC1 was positive and close to zero in 2009, and strictly negative in 2008 and 2007. On the other hand, the location Rychliki had a close to zero negative EPC1 value in 2008, and positive values in 2007

and 2009. The influence of the factor "year" is changeable in the locations, but the 2009 always had the highest and positive EPC1 values, whereas 2008 had the lowest and negative EPC1 values (except for the Wyczechy location).

Mega-Environments

The genotypes with negative GPC1 coordinates (Fig. 2a) had higher yields with better environmental conditions. Therefore, it seems that the genotypes are more resistant to frost, are lower, have less lodging, have earlier earring, reach wax maturity later and with higher resistance to disease, and are better adapted to locations with greater yield potential. Based on the EPC1 values from the C(GE)-AMMI analysis, the plot of the EPC1 scores (proposed in Paderewski et al., 2016) is presented in Fig. 3. This plot can be used for interpretation when the environments are the combinations of locations and other factors. The years had high impact on EPC1 scores, and they changed the genotype rankings in this area. In 2008, and 2007, the highest yield usually comes from
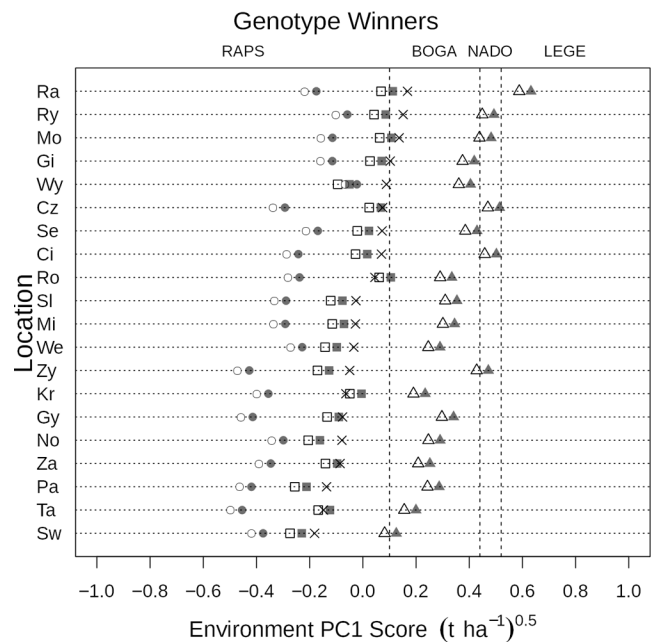

Fig. 3. The first environmental principal component (PC) scores plot with the best (winning) genotype choice according to the constrained genotype and environment additive main effects and multiplicative interaction [C(GE)-AMMI] analysis for the winter-wheat multienvironment trial data. Each location consists of the six environmental conditions: squares for 2007, circles for 2008, and triangles for 2009, with open markers for standard management and filled markers for intensive management. The average environment PC score for each location was marked by x. The dashed vertical lines separate the environmental conditions in which the same genotype was the winner. The trial locations: Ci, Cicibór; Cz, Czesławice; Gi, Głębokie; Gy, Głubczyce; Kr, Krościna Mała; Mo, Marianowo; Mi, Masłowice; No, Nowa Wieś Ujska; Pa, Pawłowice; Ra, Radostowo; Ro, Rawinowo; Ry, Rychliki; Se, Seroczyn; Sl, Słupia; Sw, Świebodzin; Ta, Tarnów; We, Węgrzce; Wy, Wyczechy; Za, Zadąbrowie; Zy, Zybiszów.

Rapsodia, whereas in 2009, it was from Bogatka, in four cases from Nadobna, and once from Legenda. In locations with better growing conditions (they were defined by the EPC1 equation , Eq. [6], and these were the environments that were contained in the bottom part of Fig. 3, i.e., from Świebodzin [Sw] to Słupia [Sl]), the Rapsodia genotype had the greatest advantage in 2007 and 2008, whereas in the worse locations (the upper part of the graph), the Rapsodia yielded higher values only in 2008.

The locations were clustered according to their averaged EPC1 scores by cluster analysis using Ward's method (Supplemental Fig. S3). Two mega-environments (Gauch and Zobel, 1997; Paderewski et al., 2016) were identified: M–E1, consisting of 11 locations with lower averaged EPC1 values (from Świebodzin [Sw] to Słupia [Sl], Fig. 3, or from Świebodzin [Sw] to Zybiszów [Zy], Supplemental Fig. S3), and M–E2, with nine locations (Rawinowo [Ro] to Radostowo [Ra] in Fig. 3, or from Radostowo [Ra] to Czesławice [Cz] in Supplemental Fig. S3). The EPCs scores are slightly different from those calculated according the standard AMMI model (Fig. 1b and 2b, or Fig. 3 vs. Paderewski et al., 2016, Fig. 3). The Alcazar, Rapsodia, and Boomer genotypes had the lowest GPC1 values in both the AMMI and C(GE)-AMMI models, and Legenda had the highest GPC1 value (Fig. 1a, 2a, and 4). The genotype Bogatka had a small positive GPC1 in both cases.

The rankings in both mega-environments were similar (Supplemental Fig. S4ab) in both the AMMI and C(GE)-AMMI analyses. The C(GE)-AMMI analysis (based on studied genotypic and environmental parameters)

emphatically recommended the Rapsodia genotype in M-E1, and Rapsodia and Bogatka in M-E2 (Fig. 3).

The M-E2 mega-environment consists of nine locations: (i) five locations that slightly preferred the Rapsodia genotype (i.e., Rawinowo, Cicibór, Seroczyn, Czesławice, and Wyczechy), and (ii) four locations where Bogatka was slightly better than Rapsodia (i.e., Głębokie, Marianowo, Rychliki, and Radostowo). The M-E2 mega-environment (Czesławice, Wyczechy, and Głębokie), according to the standard AMMI model (Paderewski et al., 2016), where Legenda was recommended, was placed in the middle of the current M-E2; in two of these locations, Rapsodia was slightly better than Legenda according to C(GE)-AMMI. The locations in which Legenda was slightly better according C(GE)-AMMI were more diversified in the AMMI analysis— Marianowo favored Rapsodia, Radostowo favored Rapsodia in 1 yr and Bogatka in 2 yr, and Rychliki favored Legenda in 2009, Rapsodia in 2008, and Rapsodia or Bogatka in 2007 (depending on the management intensity).

The C(G)-AMMI and C(E)-AMMI Models

The C(G)-AMMI and C(E)-AMMI were calculated. The additional explanatory sets could describe both genotypes and environments, or only one factor. If there is only a set describing genotypes, the C(G)-AMMI analysis should be performed. Thus, the third and fourth matrices in Eq. [1] are absent (matrices with zeros). The variability of the third matrix in C(GE)-AMMI is included in the first matrix in C(G)-AMMI, and the variability of the fourth matrix is included in the second matrix (Table 1).
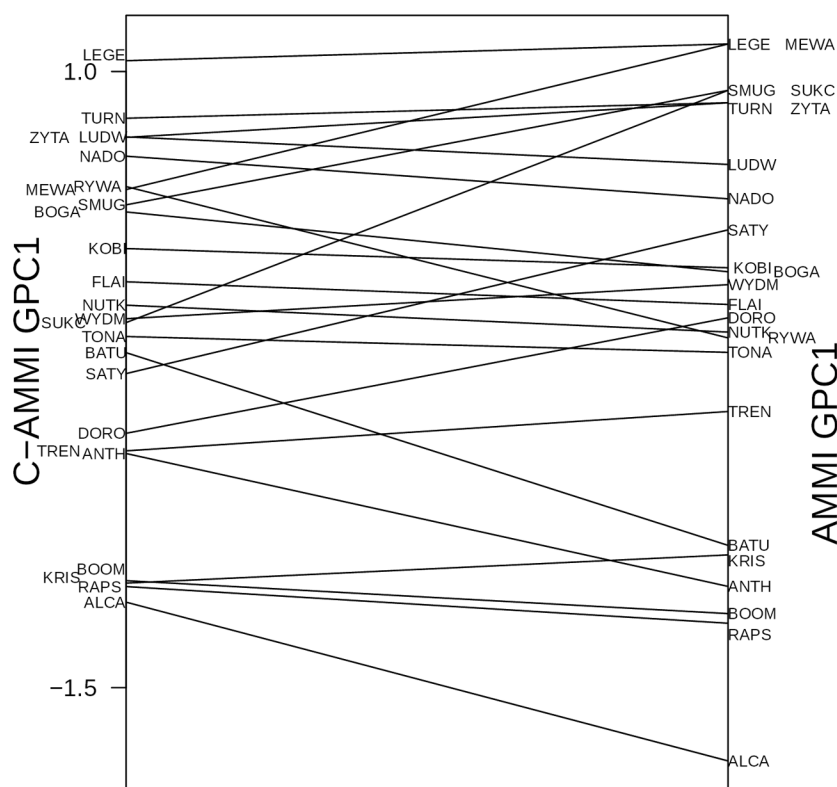

Fig. 4. The comparison of the first genotype principal component (GPC) scores for the constrained genotype and environment additive main effects and multiplicative interaction [C(GE)-AMMI] and additive main effects and multiplicative interaction (AMMI) analyses.

By analogy, the C(E)-AMMI contains only extra information regarding environments, and no constrains for GPC scores are included. The EPC1 scores plots for the C(G)-AMMI and C(E)-AMMI models are presented in Supplemental Fig. S5ab. In all the analyses, Rapsodia was the best genotype for most locations. The winners were also Bogatka and Legenda. The C(G)-AMMI analysis also indicated Nadobna as the best (winning) genotype for some environmental combinations.

The interaction component scores plots of the AMMI2 analysis (Fig. 1) can be compared with C(GE)-AMMI (Fig. 2). The first PC retained 26% and the second 9% of the total sum of squares (Table 1). The biplot for the first two PCs (Fig. 1) more precisely described the interaction effects than did the C(GE)-AMMI biplot (12% of the sum of squares for both components jointly, Fig. 2, Table 1), C(E)-AMMI (17%, Table 1), and C(G)-AMMI (27%, Table 1), but the PC scores have no substantive explanation.

Advantages of the C-AMMI Model

The understanding of the GE interaction seems to be more interesting and useful than precisely describing the yield response of genotypes to the studied environmental conditions, as this understanding allows the yield to be predicted in other environmental conditions. However, to achieve a better and biological understanding of the interaction, additional environmental variables and/or genotypic characterizations are needed.

In this paper, we generalized the AMMI model using some features from the constrained PCA and proposed the C-AMMI model, where environmental and genotypic variables are used to help understand the GE interaction. Caution should be exercised when choosing environmental or genotypic variables. If too many variables are used, the C-AMMI analysis tends to be similar to the AMMI analysis, with the GPC and/or EPC scores described by linear regression. On the other hand, a lack of appropriate variables makes it impossible to establish a real interaction pattern. Since we decided not to use any parameters that described the interaction of genotype × management by other environmental factors (location and/or year) in environmental covariates because of the very weak interaction with management (Paderewski et al., 2016), such an interaction was not visible in the C(GE)-AMMI and C(E)-AMMI results (e.g., EPC1 scores in Fig. 3 and Supplemental Fig S5b). The high management type shifted the EPC1 scores to the left with constant distance, whereas for the AMMI and C(G)-AMMI analyses, the pattern is more complex (Paderewski et al., 2016, or Supplemental Fig S5a, respectively).

The C(GE)-AMMI model results in different megaenvironments and different rankings for the genotypes when compared with the AMMI model. In our application, the C(GE)-AMMI model explained slightly less variability (i.e., sum of squares) of the data, but there is a great gain in terms of interpretation: the genotypic scores and the environmental loadings have a physical and biological meaning because they depend on the genotypic and environmental covariates, respectively. The RMSPD statistic allows the choice of the number of PCs, and it is usable for the explanatory variable diagnosis. The prediction ability depends on the number of explanatory variables for the genotypes, the number of environmental explanatory variables, and whether they are properly chosen. It needs further investigation.

There are two main advantages of the proposed C-AMMI model: (i) the results have a similar interpretation as those of the AMMI model but provide a physical interpretation of the GPC and EPC scores, and (ii) it can be a good alternative to more complex experiments, including both qualitative and quantitative explanatory variables. The developed C-AMMI analysis could be applied to different METs for the evaluation of different genotypes and environmental conditions.

Conflict of Interest

The authors declare that there is no conflict of interest.

Supplemental Material Available

The supplemental material contains the following tables and figures: a table of RMSPD values for different C-AMMI models, a table of correlations between yields estimations for the same environmental conditions for different models, a figure that presents the relations between the datasets for the C(GE)-AMMI analysis, the EPC1 and EPC2 scores of the C(GE)-AMMI analysis for each location separately, the dendrogram of the cluster analysis for grouping the locations based on their averaged EPC1 scores, the comparison of the genotype main effects for C(GE)-AMMI 1 and AMMI1, the EPC1 scores for locations, and the best (winning) genotypes according the C(G)-AMMI and C(E)-AMMI analyses.

Acknowledgments

References

Amenta, P., and L. D'Ambra. 2000. Constrained principal component analysis with external information. Statistica Applicata 12:7–29.

Anderson, W.K. 2010. Closing the gap between actual and potential yield of rainfed wheat. The impacts of environment, management and cultivar. Field Crops Res. 116:14–22. doi:10.1016/j.fcr.2009.11.016

Annicchiarico, P., E. Chiapparino, and M. Perenzin. 2010. Response of common wheat varieties to organic and conventional production systems across Italian locations and implications for selection. Field Crops Res. 116:230–238. doi:10.1016/j.fcr.2009.12.012

Arciniegas Alarcón, S., M. Garcia-Peña, C.T.S. Dias, and W.J. Krza-nowski. 2010. An alternative methodology for imputing miss-ing data in trials with genotype-by-environment interaction. Biometr. Lett. 47:75–88.

Bradu, D., and R.K. Gabriel. 1978. The biplot as a diagnostic tool for models of two-way tables. Technometrics 20:47–68. doi:10.1080/00401706.1978.10489617

Chapagain, T., and A. Good. 2015. Yield and production gaps in rain-fed wheat, barley, and canola in Alberta. Front. Plant Sci. 6:990. doi:10.3389/fpls.2015.00990

Chenu, K., M. Cooper, G.L. Hammer, K.L. Mathews, M.F. Dreccer, and S.C. Chapman. 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. 62:1743–1755. doi:10.1093/jxb/erq459

COBORU. 2008. Wyniki porejestrowych doświadczeń odmi-anowych. Zboża ozime 2007 PDO. Vol. 55. Centralny Ośrodek Badania Odmian Roślin Uprawnych, Słupia Wielka, Poland.

COBORU. 2009. Wyniki porejestrowych doświadczeń odmi-anowych. Zboża ozime 2008 PDO. Vol. 64. Centralny Ośrodek Badania Odmian Roślin Uprawnych, Słupia Wielka, Poland.

COBORU. 2010. Wyniki porejestrowych doświadczeń odmi-anowych. Zboża ozime 2009 PDO. Vol. 71. Centralny Ośrodek Badania Odmian Roślin Uprawnych, Słupia Wielka, Poland.

Cooper, M., and I.H. DeLacy. 1994. Relationships among analytical methods used to study genotypic variation and genotype-by-envi-ronment interaction in plant breeding multi-environment experi-ments. Theor. Appl. Genet. 88:561–572. doi:10.1007/BF01240919

Crossa, J., P.N. Fox, W.H. Pfeiffer, S. Rajaram, and H.G. Gauch. 1991. AMMI adjustment for statistical analysis of an international wheat yield trial. Theor. Appl. Genet. 81:27–37. doi:10.1007/BF00226108

Gabriel, K.R. 1971. The biplot graphic display of matrices with appli-cation to principal components analysis. Biometrika 58:453–467. doi:10.1093/biomet/58.3.453

Gan, Y., S. Malhi, F. Brandt, F. Katepa-Mupondwa, and H. Kutcher. 2007. *Brassica juncea* canola in the Northern Great Plains: Responses to diverse environment and nitrogen fertilization. Agron. J. 99:1208–1218. doi:10.2134/agronj2006.0296

Gauch, H.G. 1992. Statistical analysis of regional yield trials: AMMI analysis of factorial designs. Elsevier Sci., New York.

Gauch, H.G. 2007. MATMODEL version 3.0: Open source software for AMMI and related analyses. Cornell Univ., Ithaca, NY. http://css.cals.cornell.edu/people/hugh-gauch (accessed 18 Dec. 2014).

Gauch, H.G. 2013. A simple protocol for AMMI analysis of yield tri-als. Crop Sci. 53:1860–1869. doi:10.2135/cropsci2013.04.0241

Gauch, H.G., and R.W. Zobel. 1988. Predictive and postdictive suc-cess of statistical analyses of yield trials. Theor. Appl. Genet. 76:1–10. doi:10.1007/BF00288824

Gauch, H.G., and R.W. Zobel. 1990. Imputing missing yield trial data. Theor. Appl. Genet. 79:753–761. doi:10.1007/BF00224240

Gauch, H.G., and R.W. Zobel. 1997. Identifying mega-environments and targeting genotypes. Crop Sci. 37:311–326. doi:10.2135/cropsci1997.0011183X003700020002x

Geisser, S. 1993. Predictive inference: An introduction. Chapman and Hall, London. doi:10.1007/978-1-4899-4467-2

Hadasch, S., J. Forkman, and H.P. Piepho. 2017. Cross-validation in AMMI and GGE models: A comparison of methods. Crop Sci. 57:264–274. doi:10.2135/cropsci2016.07.0613

Löffler, C.M., J. Wei, T. Fast, J. Gogerty, S. Langton, M. Bergman, et al. 2005. Classification of maize environments using crop simu-lation and geographic information systems. Crop Sci. 45:1708–1716. doi:10.2135/cropsci2004.0370

Loyce, C., J.M. Meynard, C. Bouchard, B. Rolland, P. Lonnet, P. Bataillon, et al. 2008. Interaction between cultivar and crop

management effects on winter wheat diseases, lodging, and yield. Crop Prot. 27:1131–1142. doi:10.1016/j.cropro.2008.02.001

Loyce, C., J.M. Meynard, C. Bouchard, B. Rolland, P. Lonnet, P. Bataillon, et al. 2012. Growing winter wheat cultivars under dif-ferent management intensities in France: A multicriteria assess-ment based on economic, energetic and environmental indicators. Field Crops Res. 125:167–178. doi:10.1016/j.fcr.2011.08.007

Lu, C., and L. Fan. 2013. Winter wheat yield potentials and yield gaps in the North China Plain. Field Crops Res. 143:98–105. doi:10.1016/j.fcr.2012.09.015

Mandal, N.P., P.K. Sinha, M. Variar, V.D. Shukla, P. Perraju, A. Mehta, et al. 2010. Implications of genotype × input interac-tions in breeding superior genotypes for favorable and unfavor-able rainfed upland environment. Field Crops Res. 118:135–144. doi:10.1016/j.fcr.2010.05.004

Mohammadi, R., and A. Amri. 2013. Genotype × environment interaction and genetic improvement for yield and yield stabil-ity of rainfed durum wheat in Iran. Euphytica 192:227–249. doi:10.1007/s10681-012-0839-1

Paderewski, J. 2013. An R function for imputation of missing cells in two-way data sets by EM-AMMI algorithm. Commun. Biom-etry Crop Sci. 8:60–69.

Paderewski, J., H.G. Gauch, W. Mądry, T. Drzazga, and P.C. Rodrigues. 2011. Yield response of winter wheat to agro-eco-logical conditions using additive main effects and multiplica-tive interaction and cluster analysis. Crop Sci. 51:969–980. doi:10.2135/cropsci2010.05.0278

Paderewski, J., and P.C. Rodrigues. 2014. The usefulness of EM-AMMI to study the influence of missing data pattern and appli-cation to Polish post-registration winter wheat data. Aust. J. Crop Sci. 8:640–645.

Paderewski, J., H.G. Gauch, W. Mądry, and E. Gacek. 2016. AMMI analysis of four-way genotype × location × management × year data from a wheat trial in Poland. Crop Sci. 56:2157–2164. doi:10.2135/cropsci2015.03.0152

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

Rodrigues, P.C., A. Monteiro, and V.M. Lourenço. 2016. A robust additive main effects and multiplicative interaction model for the analysis of genotype-by-environment data. Bioinformatics 32:58–66. doi:10.1093/bioinformatics/btv533

Rodrigues, P.C., D. Pereira, and J.T. Mexia. 2011. A compari-son between JRA and AMMI: The robustness with increasing amounts of missing data. Sci. Agric. (Piracicaba, Braz.) 68:679–686. doi:10.1590/S0103-90162011000600012

Romagosa, I., and P.N. Fox. 1993. Genotype × environment inter-action and adaptation. In: M.D. Hayward, et al., editors, Plant breeding, principles and prospects. Chapman & Hall, London. p. 373–390. doi:10.1007/978-94-011-1524-7_23

Samonte, S.O.P.B., L.T. Wilson, A.M. McClung, and J.C. Medley. 2005. Targeting cultivars onto rice growing environments using AMMI and SREG GGE biplot analyses. Crop Sci. 45:2414–2424. doi:10.2135/cropsci2004.0627

van Eeuwijk, F.A., D.V. Bustos-Korts, and M. Malosetti. 2016. What should students in plant breeding know about the statisti-cal aspects of genotype × environment interactions? Crop Sci. 56:2119–2140. doi:10.2135/cropsci2015.06.0375

Yan, W., and M.S. Kang. 2003: GGE biplot analysis: A graphical tool for breeders, geneticists, and agronomists. CRC Press, Boca Raton, FL.

Yang, R.C., J. Crossa, P.L. Cornelius, and J. Burgueño. 2009. Biplot analysis of genotype × environment interaction: Proceed with cau-tion. Crop Sci. 49:1564–1576. doi:10.2135/cropsci2008.11.0665