

# Exploring genotype, management, and environmental variables influencing grain yield of late-sown maize in central Argentina



Brenda L. Gambin<sup>a,d,\*</sup>, Tomás Coyos<sup>b</sup>, Guido Di Mauro<sup>b</sup>, Lucas Borrás<sup>a,d</sup>, Lucas A. Garibaldi<sup>c,d</sup>

<sup>a</sup> Facultad de Ciencias Agrarias, Universidad Nacional de Rosario, Campo Experimental Villarino S/N, S2125ZAA, Zavalla, Santa Fe, Argentina

<sup>b</sup> AAPRESID, Asociación Argentina de Productores en Siembra Directa, Dorrego 1639, CP 2000, Rosario, Santa Fe, Argentina

<sup>c</sup> Instituto de Investigaciones en Recursos Naturales, Agroecología y Desarrollo Rural (IRNAD), Sede Andina, Universidad Nacional de Río Negro (UNRN), Mitre 630, CP 8400 San Carlos de Bariloche, Río Negro, Argentina

<sup>d</sup> Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina

## ARTICLE INFO

### Article history:

Received 29 April 2015

Received in revised form 18 March 2016

Accepted 23 March 2016

Available online xxxx

### Keywords:

*Zea mays* L.

Grain yield

Mixed-effects models

Genotype × management interaction

## ABSTRACT

Maize is one of the most important crops worldwide. The analysis of the influences of genotype, management, and environmental variables on grain yield has important consequences for guiding farmer's decisions. Argentina is facing relevant changes in its production system, as farmers are planting later in the growing season. It is unclear, however, which management decisions are critical, and how they interact with contrasting genotypes. Using mixed-effects models we analyzed the influences of different genotypes, management, environmental predictors and relevant two-way interactions between these predictors on grain yield in late-sown maize. On-farm multi-environmental trials were conducted during two years (2013 and 2014), with a total of 9 genotypes tested at 23 different environments in the central region of Argentina. The influence of management variables like planting date, stand density, N availability, and soil P were explored. Similarly, we analyzed the influence of environmental variables like soil type, rainfall during the crop cycle, and the presence of an influencing water table.

Averaged grain yield varied from 5,555 to 12,078 kg ha<sup>-1</sup> among environments. Our best model described the spatial and temporal variation in grain yield ( $r^2 = 0.91$ ). Genotypes varied in their performance across environments and evidenced significant interaction with N availability. Management variables positively influencing yield were, in order of relevance, N availability and stand density. N availability had a positive decelerating effect, with an initial slope of 22 kg ha<sup>-1</sup> per additional kg N ha<sup>-1</sup>. Increasing the stand density had a positive linear effect of 1,001 kg ha<sup>-1</sup> per additional increment of 10,000 pl ha<sup>-1</sup> (from 54,000 to 76,000 pl ha<sup>-1</sup> explored range). Presence of an influencing water table at planting had a negative effect on yield ( $-1,361$  kg ha<sup>-1</sup>), suggesting that water availability could be in excess in later plantings. We demonstrated that, across a wide variability in soil types and rainfall, maize grain yield can be increased by choosing superior, high responsive genotypes, increasing stand density and applying optimal N rates. Results have important implications for guiding maize management and highlight that effective decisions require the combination of management options.

© 2016 Elsevier B.V. All rights reserved.

## 1. Introduction

Maize is one of the most important crops worldwide (FAO, 2014). Exploring the influence of different genotypes, management, and environmental variables has important consequences in maize production systems. The availability of information in multi-environmental data has increased exponentially in the last years, and exploiting this information is crucial for guiding farmer's decisions and testing hypothesis with regional implications. This is clearly important in the current context and challenge of substantially increasing yields while reducing

at the same time the substantial environmental impacts of agriculture (Foley et al., 2011).

Multi-environmental trials (METs) are widely applied in crop breeding and extension. In METs, a group of genotypes are grown across a number of trials within a specific region during several years to provide information covering performance of genotypes in a particular target population of environments (DeLacy et al., 1996). The term "environment" usually encompasses management variables to that particular location following "best local practice" (in terms of fertilizer, stand density, etc.) and environmental variables that could not be easily modified by farmers (soil type, water available at planting, rainfall, etc.). Given the opportunity of control or not by farmers, unraveling management from environmental variables is not trivial.

\* Corresponding author.

E-mail address: [bgambin@unr.edu.ar](mailto:bgambin@unr.edu.ar) (B.L. Gambin).

There are large possibilities to analyze data from METs (DeLacy et al., 1996; Malosetti et al., 2013). Most of them are based on limited management and environmental description. Data are traditionally analyzed by ANOVA determining the variance components associated with genotype, environment (defined as a particular trial in a particular year, including the influence of management and “truly” environmental variables), and genotype  $\times$  environment interactions (GEI; also including genotype  $\times$  management interactions; GMI), with all these effects considered fixed. More recently, the advantages afforded by linear mixed-effects models have been recognized (Smith et al., 2005). These advantages include the capacity to handle incomplete data (not all genotypes in all environments) and the ability to assume some effects to be random or fixed, based on the particular interest of the analysis (Smith et al., 2005).

Analysis of data from METs commonly shows that the environmental (trial  $\times$  year) effect is larger than the genotype or the GEI effects (DeLacy et al., 1996). If available, environmental and management covariates can be incorporated to the model to explore their influence on the response variable, usually grain yield per unit of land area. The same occurs with GEI; they are ubiquitous and often large compared to the genotype effects (Smith et al., 2005; Malosetti et al., 2013). The interaction is then described as a differential sensitivity to explicit predictors such as stand density, sowing date, or temperature. For suitable chosen covariates the associated GEI and GMI are predictable (Smith et al., 2005), having important consequences for genotype performance predictions that are specific to individual farmer conditions.

Argentina is one of the most important producers of maize, and is currently facing a relevant change in its production system. The planting date at the central region has been moved later in the growing season (September to December). A longer fallow period allows more accumulation of water and N at planting. Late-sown maize locates the critical flowering period for yield definition (Andrade et al., 1999) under conditions of higher probability of rainfall and less evaporative demand compared to earlier traditional plantings. Although yield potential at these late planting dates is lower than earlier ones (Mercau and Otegui, 2014), farmers are obtaining acceptable yields with higher yield stability. There are also commercial benefits related to lower fertilizer needs and lower stand densities. Late sowing has become a valid alternative for maize producers to reduce risk. At present, 40–60% of the total maize produced in Argentina is considered late sowing (PAS, 2015).

In the present manuscript we explore the influence of different genotypes, management, and environmental predictors on grain yield in late-sown maize, and define the model that best describes the observed data using mixed-effects models. We are interested in defining which predictors are relevant, and in quantifying the magnitude of their effects. We hypothesize that environmental variables related to water availability have limited influence on grain yield. Because some management decisions are known to depend upon the genotype (e.g., stand density; Cox, 1996; Bavec and Bavec, 2002; Sarlangue et al., 2007; Hernández et al., 2014), it was also of interest to explore potential genotype  $\times$  management interactions (GMI) as a differential response to particular management options.

Multi-environmental trials involved 23 different environments (combination of trials and years), each having 9 common genotypes. We started with a model describing yield variation among environments, genotypes, GEI and relevant GMI without fixed-effects predictors, and compared it with another model incorporating fixed-effects predictors. Management predictors were planting date, stand density, N availability, and soil P. Environmental predictors were soil type, rainfall during the crop cycle, and the presence of an influencing water table at planting. Genotype  $\times$  management interaction was evaluated by exploring differential genotypic response to stand density and N availability.

## 2. Materials and methods

### 2.1. Study system

Trials were sown under different locations around the central production area in Argentina during two growing seasons (2012/2013 and 2013/2014, from now-on referred to 2013 and 2014, respectively). Environments are described in Table 1, and involved 9 locations in 2013 and 14 locations in 2014. The “location” is a loose spatial reference, as, in different seasons, the location (a summarized of the town name) may actually be different paddocks, farms, and/or soil types subject to different management practices. Since all the locations were not represented at each of the two years, the terms “environment” will be used herein to define the combination of a particular trial in a given year (Table 1). All fields used for trials were managed under no-tillage for a minimum of eight years. Fields belong to farmers grouped in AAPRESID (Argentinian Association of No-Tillage Farmers).

Nine single-cross maize hybrids from different seeds companies were evaluated in all trials (Table 2). Genotypes were selected by each seed company for being recommended under late planting. Many of these genotypes are widely used under early planting dates also. Relative maturity across genotypes ranged from 120 to 123 (Table 2). Each trial had a randomized complete block design with two replicates. One trial had three replicates (25M\_13). Plot size was 6 to 8 rows wide and 200 to 240 m length, depending on the specific trial. Inter-row spacing was 0.52 m.

All trials were managed under farmer decisions and with their available technology (planter, harvesting), making trials representative of maize production environments in Argentina. All trials were rainfed, and weeds and insects were controlled chemically. Soils are predominantly deep sandy loams (Typic Hapludoll, Entic Hapludoll, Haplustoll) and shallower clay soils (Aquic Argiudoll, Argialboll). Soil types varied across trials (Table S1), ranging from land suited to cultivation (types I, II and III) to more restrictive soils (type IV, V and VI) (Klingebiel and Montgomery, 1961). Predecessor crop in most trials was soybean. At each trial, soil samples until 60 cm depth were taken before planting to determine soil properties. Percentage of organic matter and amount of P (ppm) were determined at 0–20 cm depth, and N-NO<sub>3</sub> was determined until 60 cm depth. Organic matter was determined by semi-micro Walkley and Black technique (Walkley and Black, 1934), and P and N-NO<sub>3</sub> was determined by spectrophotometry. The amount of N and P applied was defined by farmers based on soil analysis,

**Table 1**  
List of environments tested.

Year	Site	Code	Latitude (decimal)	Longitude (decimal)
2013	Cristophersen	Cr_13	−34.2	−62.0
	Solis	So_13	−34.2	−59.2
	Laborde	La_13	−33.0	−59.4
	9 de Julio	9J_13	−35.6	−60.9
	Bustinza	Bu_13	−32.5	−61.2
	El Fortin	EF_13	−31.6	−62.2
	Rio II	RIL_13	−31.9	−63.8
	25 de Mayo	25M_13	−35.4	−60.1
	Urdinarrian	Ur_13	−32.7	−58.6
	2014	M.J. Moreno	MJM_14	−32.5
Noetinger		No_14	−32.4	−62.3
M. Juarez		MJ_14	−32.7	−62.0
Jovita		Jo_14	−34.5	−64.0
9 de Julio		9J_14	−35.4	−60.8
La Picada		LP_14	−31.7	−60.3
Colonia		Co_14	−31.8	−60.6
Rio II		RIL_14	−31.6	−63.8
Laboulaye		Lab_14	−34.0	−63.9
Godoy		Go_14	−33.3	−60.5
Bustinza		Bu_14	−32.5	−61.2
El Fortin		EF_14	−31.6	−62.1
Pergamino		Pe_14	−34.0	−60.1
Salto		S_14	−34.3	−60.4

**Table 2**

List of commercial genotypes tested. Relative maturity indicates relative differences in cycle between hybrids (for every two units of differences there will be one point of difference in moisture at harvest).

Genotype	Seed Company	Relative maturity
ACA_470	ACA	120
ADV_8112	Advanta	122
ARV_2155	Arvales	121
ARV_2194	Arvales	122
DK_7210	Monsanto	122
Dow_505	Dow Agr.	121
Dow_510	Dow Agr.	123
NK_840	Syngenta	121
NK_860	Syngenta	122

expected yields, and costs. Depth of water table was indicated when present at sowing. Rainfall during the crop cycle was recorded at each trial.

Grain yield data are presented at 14% moisture. Plot (strip) data of grain yield were determined with sensors located at the harvest machine. Most trials showed no detectable incidence of stalk break or diseases. The first winter killing frost was always latter than physiological maturity.

## 2.2. Predictor variables

One of the main interests of our study was to incorporate management and environmental variables as fixed predictors within the model. The inclusion of predictors was based on several aspects, including specific interest as how to manage the crop, data availability, and enough variation across trials. The following predictors were analyzed:

- Planting date: as days after 1 November (quantitative variable).
- Stand density at harvest ( $\text{pl m}^{-2}$ ): as quantitative variable.
- Nitrogen at planting soil ( $\text{kg ha}^{-1}$ , 0–60 cm depth) + fertilizer as quantitative variable ( $\text{kg ha}^{-1}$ ) (from now-on referred to N availability).
- Soil phosphorus (ppm, 0–20 cm depth): as quantitative variable (from now-on referred to soil P).
- Soil type: as categorical variable. Soils were grouped into three levels: I-II, III, and IV-V-VI.
- Rainfall during the crop cycle (mm): as quantitative variable.
- Water table at planting: as nominal variable (two levels: 0, absence; 1, presence at less than 2 m depth).

The first step of the analysis involved data exploration. Key concepts to consider at this stage were outliers, multicollinearity, and the type of relationships between variables (Zuur et al., 2009). Multicollinearity among quantitative variables was evaluated by matrix correlations following Pearson method and by calculating the variance inflation factor (VIF) in R software (R Core Team, 2013, version 3.0.2, fmsb package; Nakazawa, 2014). Latitude and longitude were also included in this correlation analysis as quantitative variables to explore spatial trends.

## 2.3. Statistical analysis and model selection

We used linear mixed-effects models to assess the influence of different predictor variables on grain yield (adjusted to 14% moisture) in R software (lme4 package, lmer function) (Bates et al., 2013). We applied the top-down strategy of model selection process (Zuur et al., 2009), which contains the following steps:

- We started with the “beyond optimal model”. After data exploration, which involved graphical analysis of yield response to different predictor variables for the entire data set and for individual genotypes,

we defined explanatory variables in the fixed component that were most likely to contribute to the optimal model. Fixed-effect predictors in this beyond optimal model were described previously (Section 2.2), and were classified into management (planting date, stand density, N availability, soil P) or environmental variables (soil type, rainfall and water table). Each overall partial regression coefficient ( $\beta^+$ ) was considered a fixed effect reflecting the influence of a predictor (e.g., stand density) on grain yield across all environments.

- Using the beyond optimal model, we found the optimal structure of the random component based on REML estimations. By including block nested within environment as random effects, our models estimated different intercepts for each block and environment to account for the hierarchical data structure. The same applied for the genotype by environment interaction term.

Genotype by management interaction (GMI) was analyzed as differential genotypic response to particular management variables. Variations among genotypes in the influence of individual variables on grain yield was quantified by estimating different intercepts and slope for each genotype (Gelman and Hill, 2007). Data exploration, which involved yield response to quantitative predictor variables for individual genotypes, suggested that potential GMI was present for N. Including GMI for stand density did not result in a model improvement based on AIC (Burnham and Anderson, 2002; Aho et al., 2014) and then was excluded. Genotype and GMI effects were considered random, as there is no interest in the particular response of individual genotypes. This means that the interest resides in the variation among genotypes on their response but not in comparing particular pairs of genotypes (Smith et al., 2005).

For N availability, we tested models with both linear and curvilinear response using non-transformed or log-transformed variables. Curvilinear response was explored by fitting a second-order polynomial function ( $Y_i = \alpha + \beta_1 \times X_i + \beta_2 \times X_i^2 + \epsilon_i$ ). We found model improvement (i.e., lower AIC) when considering a curvilinear relation, and therefore we present models with coefficients  $\beta_1 + \beta_2$  for this variable.

- Once the optimal random structure was found, we found the optimal fixed structure. For this, we followed the multimodel inference based on information-theoretic approach (Burnham and Anderson, 2002, 2004). This approach does not accept the notion that there is a simple “true model” in biological sciences. Selection of a best approximating model represents the inference from the data and tells us what “effects” (represented by parameters) can be supported by the data (Burnham and Anderson, 2002). We used AIC to select best-fitting models for combinations of the seven fixed-effect predictor variables (Aho et al., 2014). Based on the context and our objectives, AIC is the appropriate tool for model selection when compared to others as BIC or hypothesis testing (Aho et al., 2014; Burnham and Anderson, 2002, 2014; Burnham et al., 2011). Because models have different fixed effects (but with the same random structure), ML estimation was used and not REML.
- We present the final model using REML estimation. The final model was the following (Eq. (1)):

$$y_{ijk} = \mu + \sum_{p=1}^P \beta^+ P_j + G_i + E_j + (B_k)_j + G_i E_j + \beta_i N + \epsilon_{ijk} \quad (1)$$

where  $y_{ijk}$  is the yield of the genotype  $i$  at the environment  $j$  in block  $k$ ,  $\mu$  is the grand mean,  $\sum_{p=1}^P \beta^+$  represents the sum of fixed effects of predictor  $P$  in environment  $j$ ,  $G_i$  is the genotypic random effect,  $E_j$  is the environment random effect,  $(B_k)_j$  is the block random effect nested within environment  $j$ ,  $G_i E_j$  is the genotype by environment interaction random effect,  $\beta_i N$  represents the random genotype by N interaction effect, and  $\epsilon_{ijk}$  is the residual. Random terms were assumed normally distributed with a mean of zero and constant variance.

We checked the Gaussian and homoscedasticity assumptions (Zuur et al., 2009) for the standardized residuals of the models with graphical



analysis and these assumptions were valid in all cases. Variance heterogeneity across environments was particularly checked by fitting a model per environment and comparing the residual variance. This model included genotype and block as random effects. We found that residuals were apparently different (larger) in four particular environments, but were similar for the rest of the sites. The potential impact of this heterogeneity was determined by applying the model selection process without considering those particular four sites with larger residual variance. We found that results were similar to the analysis considering all sites, so the analysis included the entire data set.

The analysis was done in two steps. First, observations of each fixed-effect predictor variable were standardized by z-scores to allow comparison among predictors and to explore if yield was more affected by farmers decisions or by environmental variables (e.g., rainfall). Z-scores were obtained subtracting the variable mean and dividing by its standard deviation ( $\frac{x-\mu}{\sigma}$ ). Z-scores do not modify the functional relationship between the response and predictor variable. Second, analysis involved no data transformation to quantify the magnitude of the effects on crop yield (kg ha<sup>-1</sup>, adjusted to 14% moisture). Model with predictors were compared with a model without fixed-effect predictors to evaluate model improvement.

Proportional change in variance (PCV) at different grouping levels (environment, genotype and residual) was calculated as described in Merlo et al. (2005). PCV monitors changes specific to each variance component. That is, how the inclusion of additional predictor(s) has reduced (or increased) variance component at different levels. Proportional change in variance is calculated as follows (Eq. (2)):

$$PCV = \frac{V_{N-1} - V_{N-2}}{V_{N-1}} \quad (2)$$

where  $V_{N-1}$  is the variance in the null model and  $V_{N-2}$  is the variance in the final model with predictors. Positive value indicates a reduction in the variation among groups (e.g., environments) given by the incorporation of predictors.

$R^2$  of adjusted models were obtained following the methodology described in Nakagawa and Schielzeth (2013) for generalized linear mixed models. Both marginal and conditional  $R^2$  were calculated. Marginal  $R^2$  ( $R^2_m$ ) represents the variance explained by fixed factors and is given by (Eq. (3)):

$$R^2_m = \frac{\sigma_f^2}{\sigma_f^2 + \sum_{l=1}^{\mu} \sigma_l^2 + \sigma_e^2} \quad (3)$$

where  $\sigma_f^2$  is the variance calculated from the fixed effect components of the linear mixed model,  $\sigma_l^2$  is the variance component of the  $l$ th random factor, and  $\sigma_e^2$  is the residual variance. Eq. (3) can be modified to express conditional  $R^2$  ( $R^2_c$ ) (Eq. (4)):

$$R^2_c = \frac{\sigma_f^2 + \sum_{l=1}^{\mu} \sigma_l^2}{\sigma_f^2 + \sum_{l=1}^{\mu} \sigma_l^2 + \sigma_e^2} \quad (4)$$

which represents the variance explained by the entire model (fixed and random factors) (Nakagawa and Schielzeth, 2013).

### 3. Results

#### 3.1. Management and environmental variations across trials

Management and environmental variables showed ample variation across trials (Table S1). Sowing dates ranged from late November (November 20) to early January (January 6). Average stand density ranged from 5.4 to 7.6 plants m<sup>-2</sup>. The amount of N availability varied from 65 to 463 kg ha<sup>-1</sup>, and soil P ranged from 5 to 68 ppm (Table S1). Water table was detected in most environments at less

than 2 m depth. Rainfall during the crop cycle varied importantly across trials, from 296 mm to 1156 mm (Table S1). Accordingly, raw data indicated that grain yield variations across trials were also important, from 5,555 to 12,078 kg ha<sup>-1</sup>. Adjusted grain yield is depicted in Fig. 1.

A positive correlation was found between N availability at planting and soil P (0.57;  $p < 0.05$ ; Table S2). Soil P levels were higher at lower latitudes (northern trials) (0.47;  $p < 0.05$ ; Table S2). Planting date was slightly later in lower latitudes (0.53;  $p < 0.05$ ; Table S2). There was no evidence of multicollinearity among quantitative predictors of interest (planting date, stand density, N availability, soil P and rainfall; Table S2), and thus all were included in the analysis. Because predictor variables are in different scales, the analysis was done with standardized variables by z-scores (see [Statistical analysis and model selection](#)). The final model is also presented without standardized variables to facilitate interpretation.

#### 3.2. Model without and with fixed-effects predictors

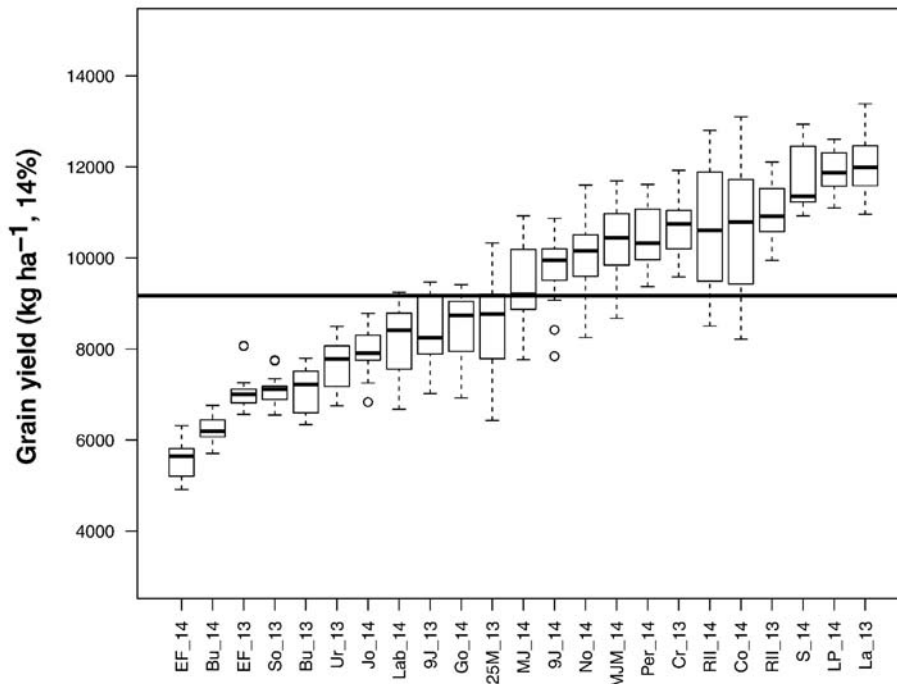
We started with a model without fixed-effects predictors. This model explores the random variation, which in this case describes how much of the total variation observed in crop yield is associated with differences across environments, genotypes, GEI and genotype  $\times$  N interaction. For our data set this model indicated that environment-to-environment variation had the greatest contribution to the total variance (68%), followed by GEI (8%), genotype-to-genotype variation (5%), and the genotypic variation in their response to N availability (1%) (Table 3). The same trend was observed for non-standardized variables (Table 3). These results are in accordance to the wide yield variation observed across environments (Fig. 1). The residual variation of this model was 10% of the total variance (Table 3).

To understand how much of the variation in crop yield among environments could be explained by management and environmental predictors (Table S1), we explored different linear mixed-effects models, and selected the best-fitting model based on AIC (Tables 3 and 4). The fixed-effects predictors of the best-fitting model explained 42.3% of the environment-to-environment variation in crop yield ( $PCV_E$ ), meaning that part of the variation in the previous model was due to the management or environmental variables we considered (Table 3). The same trend was observed in the model without standardization. While the variation (i.e.,  $\sqrt{VC}$  = standard deviation) among environments was 2072 kg ha<sup>-1</sup> in the model without predictors, this variation was reduced to 1,496 kg ha<sup>-1</sup> in the best-fitting model.

The effect of each environment on crop yield is depicted in Fig. S1. It can be observed that the ranking of environments is not the same as in Fig. 1. For example, 9J\_13 was one of the lowest yielding environments (Fig. 1) but showed a positive effect in the best-fitting model (Fig. S1). The analysis helps to separate yield variation associated with particular variables and, for the particular case of 9J\_13, indicates that fixed-effect predictors were responsible for the lower yield. The remaining variation might be explained by variables not considered here (e.g., water available at planting, other soil properties, etc.).

As done for management and environmental variables at the environment level, different genotypic attributes could be added as predictors at the genotypic level. In our case there was no predictor at this level and for this reason proportional change in variance for genotypes ( $PCV_G$ ) was low (Table 3). The same applies for GEI, where part of this interaction is due to a differential genotypic response to N availability. The best-fitting model with non-standardized variables indicated that the variation among genotypes was 606 kg ha<sup>-1</sup>.

Genotypic random effects (i.e., usually called best linear unbiased predictor for genotypes or BLUPs; Robinson, 1991) are depicted in Fig. 2. Among genotypes with positive effects or performance across environments appear DK\_7210 and ADV\_8112, and among genotypes with relative poor performance across environments appear ARV\_2194 and ARV\_2155. This indicates the importance of genotype selection for



**Fig. 1.** Boxplot of adjusted grain yield (14% moisture) for the 23 trials tested (see Table 1 for further details). The black line reflects the data mean. Across trials, a total of 9 genotypes were evaluated under on-farm conditions.

maximizing yield. The same applied for the response to N; yield response to N availability was higher or lower depending on the particular genotype (discussed below).

Proportional change in variance for residuals ( $PCV_{Residual}$ ) was close to zero, meaning that the residual of the final model with predictors was similar than the original model without predictors (Table 3). This is expected as residuals indicate the variation within environments, and our fitted models only considered predictors at the environment level.

### 3.3. Influence of predictor variables

In general, models taking into account management and environmental variables as fixed predictors improved overall model accuracy

(Table 4). This is evidenced when comparing the AIC value of the null model, with no predictor (model K), with the first ten models with at least one fixed effect (models A–J). Among the included management variables, stand density and N availability appeared increasing the accuracy in most models. Among environmental variables, the presence of an influencing water table during the crop cycle was the variable that appeared in best models. Variance explained by fixed factors ( $R^2_m$ ) across models ranged from 0.29 to 0.38. Variance explained by entire models ( $R^2_c$ ) was  $>0.90$ , indicating that they appropriately and accurately described the observed yield data.

Among models with predictors (models A–J), there were no important differences in AIC, indicating that there was no clearly best model. We found that the best model describing the data was model A. Model A included two management variables (stand density and N

**Table 3**  
Mixed-effects models of the influences of environmental, management, and genotypic predictors on grain yield.

Model name		Model without fixed effects (z-scores)	Final model (z-scores)	Model without fixed effects	Final model
Random effects		VC <sup>b</sup>	VC	VC	VC
Environment (E)		0.7315	0.4224	4,293,000	2,237,000
Genotype (G)		0.0574	0.0571	333,400	367,800
GEI <sup>a</sup>		0.0812	0.0811	142,600	405,400
N	$\beta_1^c$	0.0088	0.0088	831,300	83,840
	$\beta_2$	0.0045	0.0043	2	0.3295
Residual		0.1088	0.1088	586,800	438,800
PCV <sub>E</sub> <sup>d</sup>		–	42.3%	–	–
PCV <sub>G</sub>		–	0.5%	–	–
PCV <sub>Residual</sub>		–	0.02%	–	–
Fixed effects		Estimate ± SE	Estimate ± SE	Estimate ± SE	Estimate ± SE
Intercept		–0.02 ± 0.19	0.51 ± 0.25	4,024 ± 2,213	397 ± 4,732
N	$\beta_1$	–	0.67 ± 0.24	–	22 ± 98
	$\beta_2$	–	–0.16 ± 0.17	–	–0.02 ± 0.19
Plant density		–	0.22 ± 0.15	–	1,001 ± 746
Water table		–	–0.63 ± 0.34	–	–1,361 ± 785
VIFs <sup>e</sup>		–	<1.2	–	<1.2

<sup>a</sup> GEI, genotype by environment interaction.

<sup>b</sup> VC, variance components.

<sup>c</sup>  $\beta_1$  and  $\beta_2$ , parameters of a second-order polynomial function fitted to yield–nitrogen relationship.

<sup>d</sup> PCV, proportional change in variance.

<sup>e</sup> VIFs, variance inflation factor.

**Table 4**  
Akaike's Information Criterion (AIC) for mixed effects models of the potential effect of management and environmental variables on grain yield in late-sown maize. The best 10 models are shown (from a total of 128 possible models) plus the model without fixed effects (model K). Each column represents a different predictor variable (management or environment). Uncross cells indicate variables that were not included in a particular model.  $R^2_m$  represents the variance explained by fixed factors, while  $R^2_c$  represents the variance explained by the entire model. AIC measures the relative goodness of fit of a given model; the lower its value, the more likely it is that this model is correct. The  $\Delta$  column indicates the difference between a model's AIC and that of the best-fitting model. Models were compared using the maximum likelihood (ML) method, with the exception of  $R^2_m$  and  $R^2_c$  which were calculated from REML estimations. Models have the same random effects, to account for the hierarchical data structure (blocks nested within environment), differences among environments, genotypes, genotype  $\times$  environment and genotype  $\times$  N interaction. Environment refers to a particular location in a given year. Data were standardized by z-scores prior to analysis. See [Materials and methods](#) for further details.

Model	Management variables				Environment variables				$R^2_m$	$R^2_c$	AIC	$\Delta$ AIC
	Stand density	Planting date	Soil P	N availability at planting	Soil class	Rainfall	Water table					
A	X			X			X	0.38	0.91	622	0	
B				X			X	0.34	0.91	623	1	
C	X	X		X			X	0.37	0.91	624	2	
D				X		X	X	0.34	0.91	625	3	
E			X	X			X	0.34	0.91	625	3	
F		X		X			X	0.34	0.91	625	3	
G				X	X		X	0.36	0.91	625	3	
H	X		X	X	X		X	0.36	0.91	625	3	
I				X		X	X	0.29	0.91	626	3	
J	X		X	X		X	X	0.37	0.91	626	3	
K								–	–	632	10	

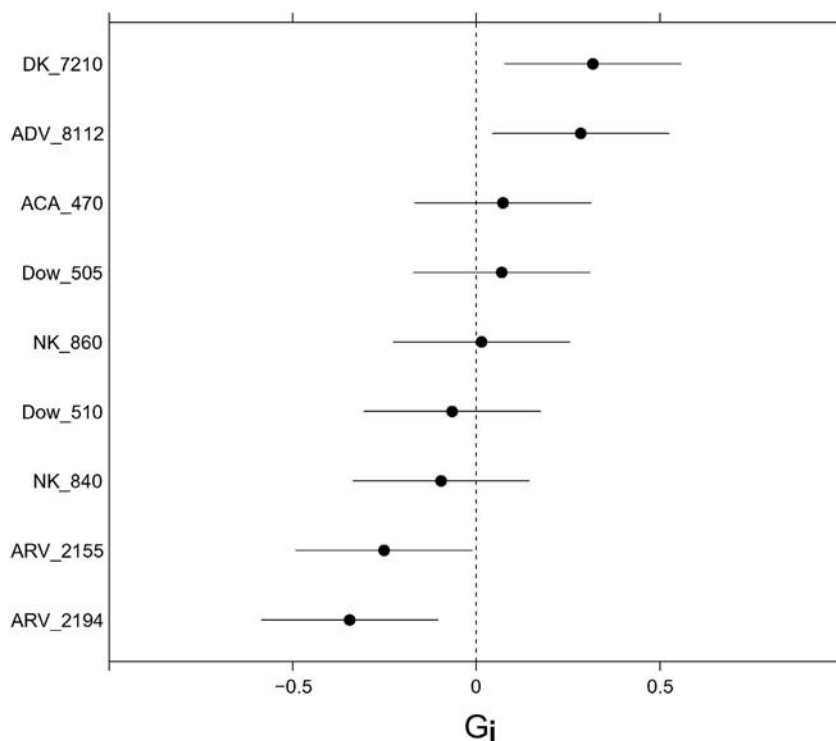
availability) and one environmental variable (the presence of water table at planting) (Table 4). Despite this, it is interesting to note that N availability at planting, a management variable that can be easily managed by farmers, appeared in best models (Table 4).

We further examined the estimates of regression coefficients ( $\beta$ ) for the best model (model A, Table 4). This allows quantifying the particular influence of each predictor variable on grain yield. The analysis with standardized variables allows comparing the influence of different predictors (Table 3). N availability was the most important management variable. Presence of water table followed in importance. Less important was stand density (Table 3).

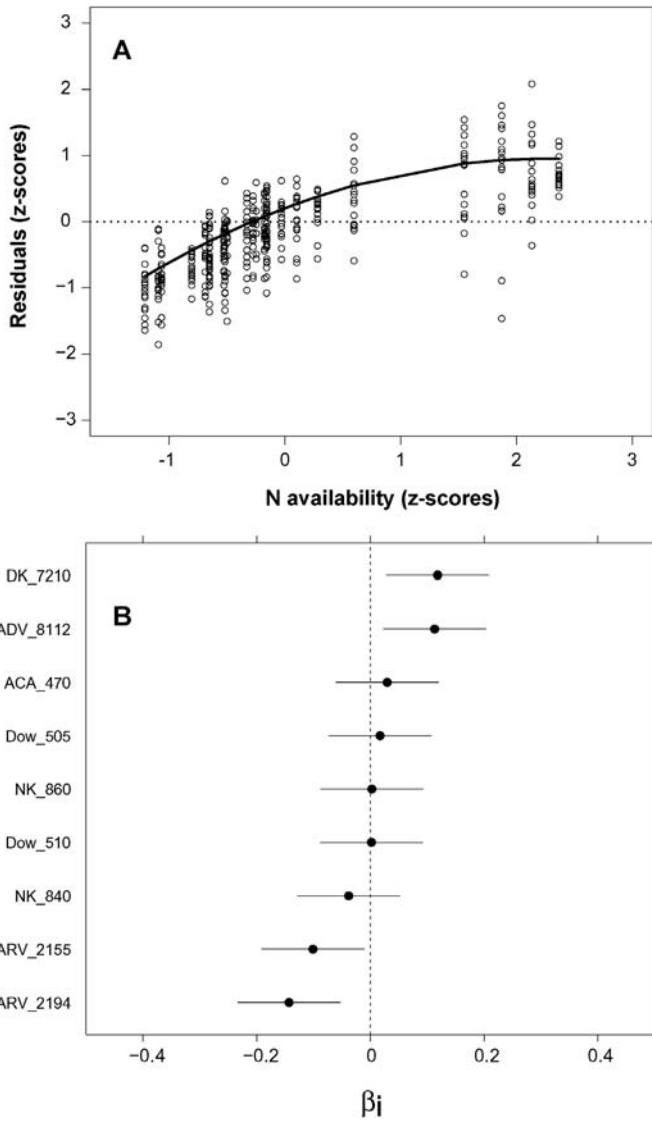
Nitrogen availability showed a positive decelerating effect on grain yield (Fig. 3A). The initial slope was 22 kg ha<sup>-1</sup> per additional kg N ha<sup>-1</sup> available at planting. Importantly, this initial yield response

to N varied depending on the genotype (Fig. 3B). For example, this response was 6 kg ha<sup>-1</sup> higher for DK\_7210, meaning that the response of this particular genotype raised to 28 kg ha<sup>-1</sup> per additional kg N ha<sup>-1</sup>. Contrarily, this response was 6 kg ha<sup>-1</sup> lower for ARV\_2194 (Fig. 3B), which means that the response of this particular genotype was 16 kg ha<sup>-1</sup> per additional kg N ha<sup>-1</sup>. Response to N for particular genotypes is depicted in Fig. S2.

Stand density showed a positive effect on grain yield (Fig. 4A). Increasing the stand density by 10,000 plants ha<sup>-1</sup> promoted an overall increase in grain yield of 1,001 kg ha<sup>-1</sup>. While this was the overall effect, there was no clear indication of a differential yield response to stand density among genotypes (data not shown). There was no model improvement (lower AIC) when considering stand density  $\times$  genotype interaction.



**Fig. 2.** Intercepts ( $G_i$ ) represent the effect of genotype, and reflect grain yield variations among genotypes not explained by predictors considered in the final model. Data were standardized by z-scores prior to analysis.

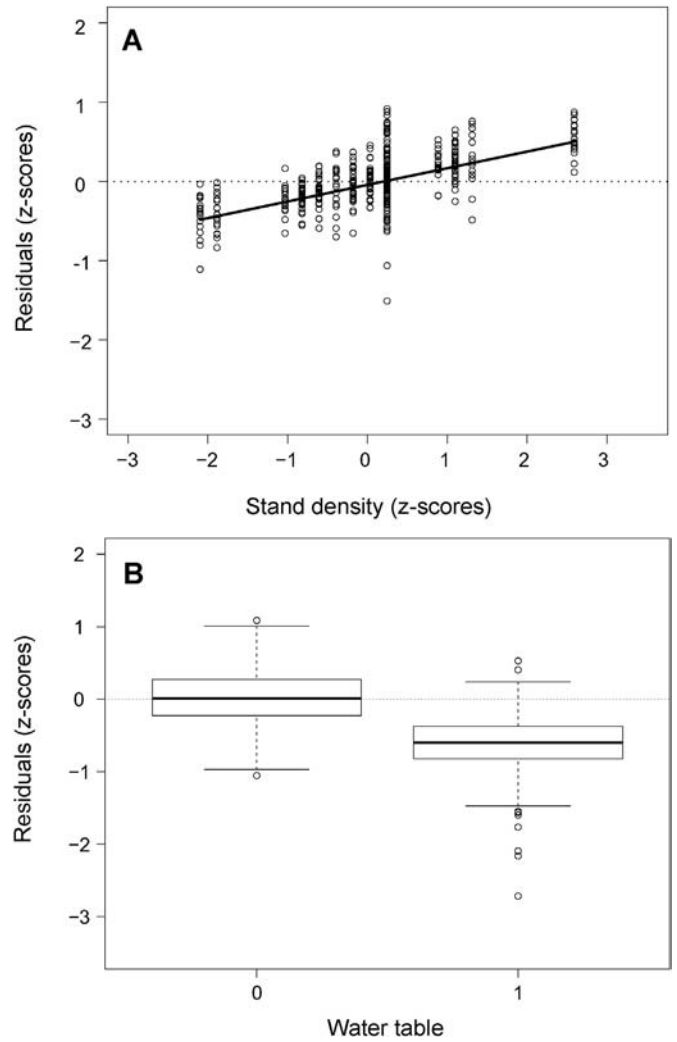


**Fig. 3.** (A) Residuals of grain yield versus N availability at planting for the entire data set. Residuals were obtained after subtracting to each observed yield the estimated yield from the final model without considering the effect of N availability. The black line reflects the  $\beta +$  for N availability. (B) Slopes ( $\beta_i$ ) represent the effect of N for individual genotypes. Data were standardized by z-scores prior to analysis.

Presence of a water table at planting showed a negative effect on grain yield (Fig. 4B). The effect was  $-1,361 \text{ kg ha}^{-1}$ . Water table presence was not related to poor soils, as shown in Fig. S3. This spine plot showed that presence of water table was mostly found in best soils, roughly 70% of the observations of soil type I-II.

**4. Discussion**

Mixed-effects models are powerful statistical tools to analyze hierarchical data sets (Zuur et al., 2009). They are currently applied in many disciplines where simple and classical analyses (like ANOVA) fail (Smith et al., 2001; Bolker et al., 2009; Merlo et al., 2005). In agronomy sciences, mixed-effects models are widely applied for breeding decisions (DeLacy et al., 1996; Smith et al., 2001, 2005). The advantages of these models include the capacity to handle incomplete data, nested (hierarchical) data structure, the capacity to use more realistic within-trial models for error variation (e.g., spatial correlation), and the ability to assume specific effects to be random or fixed (Smith et al., 2005; Zuur et al., 2009). Their potential application for explaining genotype  $\times$  environmental interactions and assist selection in a breeding program



**Fig. 4.** Residuals of grain yield versus stand density (A) and water table (B) for the entire data set. Residuals were obtained after subtracting to each observed yield the estimated yield from the final model without considering the effect of stand density (A) or water table (B). The black line in (A) reflects the  $\beta +$  for stand density. In (B), absence of water table is indicated with 0, and presence with 1. Data were standardized by z-scores prior to analysis.

has been explored (Malosetti et al., 2013). Here, we extended their use for dissecting the effect of crop management decisions and relevant GMI. Our model satisfactorily described the spatial and temporal variation in maize grain yield ( $r^2 = 0.91$ ), which ranged from 5,555 to 12,078  $\text{kg ha}^{-1}$ .

We demonstrate that genotype selection and management decisions are relevant for optimizing maize yield under late sowing. These results have important implications; they indicate that the success of late-sown maize is highly dependent on farmer decisions. Important management variables were N availability and stand density. The lack of soil P effect is in accordance to recent results in this region, showing lower P response and lower saturation thresholds due to higher temperatures compared to early plantings (Ferraris and Couretot, 2014). We confirmed that weather variables like rainfall during the crop cycle have negligible influence on grain yield. This is in accordance with the lower evaporative demand under these environments compared to traditional early sowing (Maddoni, 2012). We even found a considerable negative effect of the presence of water table at sowing on grain yield, suggesting that water availability could be on excess under late plantings. Our results are not contemplating a maize crop sown after a winter crop (wheat, barley, pea), but a maize where the fallow is extended for several months. In the case of maize after a winter crop, water availability



at sowing should have an important influence on yield (Mercau and Otegui, 2014). The implications of extended fallows in terms of soil conservation and sustainability are still to be studied, especially when considering late sown maize is commonly planted after soybean.

By including predictor variables explicitly within the model we removed part of the random source of variation at the environment grouping level. In our particular case, 42.3% of yield variation across environments was explained by management and environmental variables. This result is very useful considering that predictors varied independently across environments. The analysis helps to reduce the uncertainty and define factors affecting yield. As mentioned above, these factors are mostly related to management decisions. The remaining variation, on the other hand, means that there is variation across environments not explained by the considered predictors. This variation is relevant, and denotes that there are other important variables influencing yield.

Genotypes represent a random sample of commercial germplasm currently available. The same applied for the combinations for studying possible GMI. This means that there is no particular interest in estimating neither an effect for a particular genotype, nor the effect of a particular genotype when a particular management variable is changed. Our interest is to explore the variation among genotypes and in their response to management decisions (for example, stand density). This might sound trivial but has important implications when extrapolating results to a wider population of genotypes or environments (Zuur et al., 2009; Smith et al., 2005). We found that there was GMI for N availability. This was not the case for stand density. Management recommendations depending on genotype is an increasing practice offered by seed companies, especially for stand density, and this information could help to reduce the costs associated with required experiments or increased recommendation efficiencies.

Although recommended N rates vary depending on the environmental quality (Alvarez, 2008; Salvagiotti et al., 2011), we found an overall positive response across an important range of yield variation. The overall regression describes a common saturation response curve (Fig. 3A; de Wit, 1953). Interestingly, we showed that part of the variation depicted in Fig. 3A is due to genotype differences in the response to this nutrient (Figs. 3B; S2). Our results describe that farmer N fertilizer investment need to be coupled with genotype selection. While a differential yield response to N in older versus newer maize genotypes has been documented (Ma and Dwyer, 1998), variations across current commercial genotypes is more scarce.

The response of maize yield to stand density for an individual genotype growing at a particular environment and under particular management conditions is typically parabolic (Williams et al., 1968; Giebrech, 1969; Westgate et al., 1997; Hashemi et al., 2005). We found an overall positive response to stand density across environments, suggesting that we were exploring the linear part of the yield response to stand density. This positive response, however, is not in accordance to the recommended reduction in stand density for more limited environments (Bavec and Bavec, 2002; Al-Kaisi and Yin, 2003), and indicates that farmers are underestimating the optimum stand density. More information in this sense will allow exploring if recommended stand density vary depending on the genotype or the quality of the environment. This is important as optimum stand density depends on both genotype and environmental quality (Duncan, 1954; Rutger and Crowder, 1967; Brown et al., 1970; Carlone and Russell, 1987; Bavec and Bavec, 2002; Hernández et al., 2014).

A global-scale assessment of future agricultural prospects has shown that yield gaps (differences between observed yields and those attainable in a given region) can be importantly reduced by management practices, particularly nutrient and water management (Licker et al., 2010; Mueller et al., 2012). For this particular region, attainable yields were close to 12 t ha<sup>-1</sup> considering the highest-yielding location in this study, which is similar to simulated potential yield under no water and nutrient limitations (Mercau and Otegui, 2014). Considering

the limitations of the sample we analyzed we demonstrated that yield increments through N rates and stand density management represented 40% of attainable yields based on our analysis. We provide information of high-yielding cropping practices that could be extrapolated to other regions of similar climate towards reducing yield gap. Importantly, this intensification will need to be sustainable to increase yield while simultaneously reducing the agricultural global environmental footprint (Foley et al., 2011).

## 5. Conclusion

Using mixed-effects models we explored the influence of different genotypes, managements, environments and relevant genotype × management interactions on grain yield in late-sown maize. Our model satisfactorily described the spatial and temporal variation in grain yield ( $r^2 = 0.91$ ), which ranged from 5,555 to 12,078 kg ha<sup>-1</sup>.

However the limitations of our sample (environments, genotypes, management combinations) we demonstrated that farmers' decisions related to genotype selection, and management variables like N availability and stand density are relevant in terms of the crop yield. Rainfall and soil type showed limited influence. Presence of water table at sowing showed an overall negative effect, suggesting that water availability could be in excess in these environments. N availability and genotype selection need to be managed in combination, as results showed that commercial genotypes differed in their response to N availability.

We provide information of high-yielding cropping practices that could be extrapolated to other regions of similar climate towards reducing maize yield gap and increasing cropping efficiency.

## Acknowledgments

Authors wish to thank AAPRESID agronomists and farmers for field technical assistance and for providing fields for experimentation, and maize seed companies for providing seed and financial support.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.agry.2016.03.011>.

## References

- Aho, K., Derryberry, D., Peterson, T., 2014. Model selection for ecologists: the worldviews of AIC and BIC. *Ecology* 95, 631–636.
- Al-Kaisi, M.M., Yin, X., 2003. Effects of nitrogen rate, irrigation rate, and plant population on corn yield and water use efficiency. *Agron. J.* 95, 1475–1482.
- Alvarez, R., 2008. Analysis of yield response variability to nitrogen fertilization in experiments performed in the Argentine Pampas. *Commun. Soil Sci. Plant Anal.* 39, 1235–1244.
- Andrade, F.H., Vega, C., Uhart, S., Cirilo, A., Cantarero, M., Valentinuz, O., 1999. Kernel number determination in maize. *Crop Sci.* 39, 453–459.
- Bates, D., Maechler, M., Bolker, B., Walker, S., 2013. lme4: linear mixed-effects models using Eigen and S4. R package version 1.0-5. <http://CRAN.R-project.org/package=lme4>.
- Bavec, F., Bavec, M., 2002. Effects of plant population on leaf area index, cob characteristics and grain yield of early maturing maize cultivars (FAO 100–400). *Eur. J. Agron.* 16, 151–159.
- Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H.H., White, J.S.S., 2009. Generalized linear mixed models: a practical guide for ecology and evolution. *Trends Ecol. Evol.* 24, 127–135.
- Brown, R.H., Beatty, E.R., Ethredge, W.J., Hayes, D.D., 1970. Influence of row width and plant population on yield of two varieties of corn (*Zea mays* L.). *Agron. J.* 62, 767–770.
- Burnham, K.P., Anderson, D.R., 2002. Model selection and multimodel inference: a practical information-theoretic approach. Springer Science and Business Media.
- Burnham, K.P., Anderson, D.R., 2004. Multimodel inference understanding AIC and BIC in model selection. *Sociol. Methods Res.* 33, 261–304.
- Burnham, K.P., Anderson, D.R., 2014. P values are only an index to evidence: 20th- vs. 21st-century statistical science. *Ecology* 95, 627–630.
- Burnham, K.P., Anderson, D.R., Huyvaert, K.P., 2011. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behav. Ecol. Sociobiol.* 65, 23–35.
- Cox, W.J., 1996. Whole-plant physiological and yield responses of maize to plant density. *Agron. J.* 88, 489–496.



- Carlone, M.R., Russell, W.A., 1987. Response to plant densities and nitrogen levels for four maize cultivars from different eras of breeding. *Crop Sci.* 27, 465–470.
- de Wit, C.T., 1953. A physical theory on placement of fertilizers. *Vers. Landbouwk. Onderz. (Agric. Res. Rep.)* Vol. 59.4. Staatsdrukkerij, 's-Gravenhage
- DeLacy, I.H., Basford, K.E., Cooper, M., Bull, J.K., McLaren, C.G., 1996. Analysis of multi-environment trials – an historical perspective. In: Cooper, M., Hammer, G.L. (Eds.), *Plant Adaptation and Crop Improvement*. CAB International, Wallingford, UK, pp. 39–124.
- Duncan, E.R., 1954. Influences of varying plant population, soil fertility and hybrid on corn yields. *Soil Sci. Soc. Proc.* 18, 437–440.
- FAO, Food and Agricultural Organization of the United Nations, 2014o. Statistics of farming production. Maize. Available in <http://faostat3.fao.org/home/E>.
- Ferraris, G.N., Couretot, L.A., 2014. Elección de ambientes, rendimiento y fertilización de maíz según fecha de siembra. *Revista Técnica Siembra Directa Maíz*, pp. 58–63.
- Foley, J.A., Ramankutty, N., Brauman, K.A., Cassidy, E.S., Gerber, J.S., Johnston, M., Mueller, N.D., O'Connell, C., Ray, D.K., West, P.C., Balzer, C., Bennett, E.M., Carpenter, S.R., Hill, J., Monfreda, C., Polasky, P., Rockström, J., Sheehan, J., Siebert, S., Tilman, D., Zaks, D.P.M., 2011. Solutions for a cultivated planet. *Nature* 478, 337–342.
- Gelman, A., Hill, J., 2007. Data analysis using regression and multilevel/hierarchical models. Cambridge Univ. Press, Cambridge.
- Giebrech, J., 1969. Effect of population and row spacing on the performance of four corn (*Zea mays* 1L.) hybrids. *Agron. J.* 61, 439–441.
- Hashemi, A.M., Helbert, S.J., Putnam, D.H., 2005. Yield response of corn to crowding stress. *Agron. J.* 97, 839–846.
- Hernández, F., Amelong, A., Borrás, L., 2014. Genotypic differences among Argentinean maize hybrids in yield response to stand density. *Agron. J.* 106, 2316–2324.
- Klingebiel, A.A., Montgomery, P.H., 1961. Land-capability classification. Soil conservation service, USDA. *Agriculture Handbook* No. 210.
- Licker, R., Johnston, M., Foley, J.A., Barford, C., Kucharik, C.J., Monfreda, C., Ramankutty, N., 2010. *Glob. Ecol. Biogeogr.* 19, 769–782.
- Ma, B.L., Dwyer, L.M., 1998. Nitrogen uptake and use of contrasting maize hybrids differing in leaf senescence. *Plant Soil* 199, 283–291.
- Maddonni, G.A., 2012. Analysis of the climatic constraints to maize production in the current agricultural region of Argentina—a probabilistic approach. *Theor. Appl. Climatol.* 107, 325–345.
- Malosetti, M., Ribaut, J., van Eeuwijk, F.A., 2013. The statistical analysis of multi-environment data: modeling genotype by environment interaction and its genetic basis. *Front. Physiol.* 4, 44.
- Mercau, J.L., Otegui, M.E., 2014. Advances in agricultural systems modeling. In: Ahuja, L.R., Ma, L., Lascano, R.J. (Eds.), *Practical Applications of Agricultural System Models to Optimize the Use of Limited Water*. ASA, CSSA, SSSA, Madison, USA, pp. 301–323.
- Merlo, J., Yang, M., Chaix, B., Lynch, J., Råstam, L., 2005. A brief conceptual tutorial on multilevel analysis in social epidemiology: investigating contextual phenomena in different groups of people. *J. Epidemiol. Community Health* 59, 729–736.
- Mueller, N.D., Gerber, J.S., Johnston, M., Ray, D.K., Ramankutty, N., Foley, J.A., 2012. Closing yield gaps through nutrient and water management. *Nature* 490, 254–257.
- Nakagawa, S., Schielzeth, H., 2013. A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods Ecol. Evol.* 4, 133–142.
- Nakazawa, M., 2014. fmsb: functions for medical statistics book with some demographic data. R package version 0.5.1. <http://CRAN.R-project.org/package=fmsb>.
- PAS. Panorama Agrícola Semanal, 2015. Bolsa de Cereales, Argentina. URL <http://www.bolsadecereales.org/>.
- R Core Team, 2013. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (URL <http://www.R-project.org/>).
- Robinson, G.K., 1991. That BLUP is a good thing: the estimation of random effects. *Stat. Sci.* 6, 15–32.
- Rutger, J.N., Crowder, L.V., 1967. Effect of high plant density on silage and grain yields of six com hybrids. *Crop Sci.* 7, 182–184.
- Salvagiotti, F., Catellarín, J., Ferraguti, F., Pedrol, H., 2011. Economic optimal nitrogen rate as affected by yield potential and nitrogen supply in the northern pampas. *Cienc. Suelo* 29, 199–212.
- Sarlangue, T., Andrade, F.H., Calviño, P.A., Purcell, L.C., 2007. Why do maize hybrids respond differently to variations in plant density? *Agron. J.* 99, 984–991.
- Smith, A.B., Cullis, B.R., Thompson, R., 2001. Analyzing variety by environmental data using multiplicative mixed models and adjustments for spatial field trend. *Biometrics* 57, 1138–1147.
- Smith, A.B., Cullis, B.R., Thompson, R., 2005. The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. *J. Agric. Sci.* 143, 449–462.
- Walkley, A., Black, A., 1934. An examination of the Degtjareff method for determining soil organic matter, and proposed modification of the chromic acid titration method. *Soil Sci.* 37, 29–38.
- Westgate, M.E., Forcella, F., Reicosky, D.C., Somsen, J., 1997. Rapid canopy closure for maize production in the northern US corn belt: radiation-use efficiency and grain yield. *Field Crop Res.* 49, 249–258.
- Williams, W.A., Loomis, R.S., Duncan, W.G., Dovrat, A., Nunez, A.F., 1968. Canopy architecture at various population densities and the growth and grain yield of corn. *Crop Sci.* 8, 303–308.
- Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A., Smith, G.M., 2009. *Mixed effects models and extensions in ecology with R*. Springer, New York.