



Drought and temperature limit tropical and temperate maize hybrids differently in a subtropical region

Lucia Casali^{1,2} · Gerardo Rubio^{1,2} · Juan M. Herrera^{1,3} 

Accepted: 21 June 2018

© INRA and Springer-Verlag France SAS, part of Springer Nature 2018

Abstract

Although the semiarid and subhumid Chaco regions in northern Argentina have been traditionally considered marginal and unsuitable for cultivating grain maize for human and livestock nutrition, this crop is increasingly being adopted by local farmers. The low maize yields observed in the area suggest that climatic constraints limit productivity, while changes in genotypes and management may be useful to mitigate the effect of these constraints. We analyzed data from 792 farm paddocks with multivariate mixed models to identify and quantify the main environmental and management constraints to maize's yield. In addition, we used the crop model CERES-Maize to assess the potential of a temperate maize hybrid to overcome water constraints. Results from the mixed models identified the amount of rainfall during February as a primary determinant of maize yield and showed that tropical hybrids tended to withstand higher temperatures and heat stress better, while temperate hybrids performed better under conditions of water scarcity. CERES-Maize simulations suggested that temperate maize hybrids have the potential to increase grain yields from 18 to 21 kg ha⁻¹ (14.5% moisture content) for every millimeter of rain during February. This report is the first to identify alternative roles of temperate and tropical maize hybrids for counteracting climatic risks in the studied subtropical regions. These findings will provide plant breeders urgently needed information to breed better adapted maize genotypes for the semiarid and subhumid Chaco.

Keywords Farming systems · Subtropical semiarid areas · Big-data · Mixed models · Climate change

Abbreviations

DD Degree days

D35 Degree days accumulated above 35 °C

TJA Average maximum temperature from January to April

1 Introduction

Maize (*Zea mays* L.) in the southern cone of South America was mainly cultivated in the humid Pampas (Lavado and Taboada 2009). In the last decade, this crop has spread to other regions, previously considered

marginal and unsuitable for its cultivation. Among them, the semiarid and subhumid Chaco were epicenters of an agricultural expansion that was favored by changes in land tenure (Goldfarb and van der Haar 2016). Agricultural expansion was also promoted by the increase in annual precipitation (approximately 18%) observed in the last decade (Ricard et al. 2015) and in water use efficiency achieved by the adoption of no-till farming (Piquer-Rodríguez et al. 2015). Currently, maize is being increasingly adapted by local farmers, and its expansion is expected to continue (Giménez et al. 2015). Recent evaluations showed that these regions have the highest yield gaps for maize in Argentina (41% of water-limited yield potential) (Aramburu Merlos et al. 2015). The current situation calls for identifying main constraints to maize productivity, to minimize the climatic risk, and to improve management practices taking into account soil, climate, and genotypes (Andrade et al. 2017). It is also critical to maximize productivity of the existing agricultural lands given the high rate of deforestation in the region (Vallejos et al. 2015).

✉ Juan M. Herrera
herreraj@agro.uba.ar

¹ Instituto de Investigaciones en Biociencias Agrícolas y Ambientales (INBA-CONICET), Buenos Aires, Argentina

² Cat. Fertilidad y Fertilizantes. Facultad Agronomía, Universidad Buenos Aires, Buenos Aires, Argentina

³ Agroscope, Route de Duillier 50, 1260 Nyon, Switzerland

There is a trend by local farmers to delay sowing dates from October to December to reduce the risk of drought. By delaying sowing dates, maize encounters a less restrictive water balance during the reproductive period which is critical for yield determination. However, this delay increased the risk that the crop would be affected by heat stress during the reproductive periods (Edreira and Otegui 2012; Maddonni 2012). Furthermore, modifications in practices such as sowing dates will certainly have implications in terms of the most suitable genotype to grow. Independent of the strategy, the improvement in maize management in this area is a challenge because of scarce previous experience and research. It is thus necessary to generate valid information for the design of agronomic practices that make high productivity compatible with no or minimal environmental impacts.

Models represent valuable tools to design agronomic practices since they make the combination of a high number of variables or treatments that would be unaffordable and time consuming to study by means of field experiments feasible (Jones et al. 2003). CERES-Maize is a mechanistic model that integrates information about soil, weather, crops, and management to evaluate several combinations of production practices (Jones et al. 2003). This model was evaluated and calibrated in temperate regions of Argentina, showing low estimation errors (e.g., Brisson et al. 2001). In contrast, there are no reports of CERES-Maize performance in the studied area, where its critical evaluation is needed prior to use because the predictions of such models for subtropical areas may be biased according to recent assessments (e.g., Ray et al. 2015). Because the calibration of mechanistic models requires an extensive collection of field data (Lobell et al. 2003), the use of empirical models to mine existing data may represent an alternative to obtain insights in less time. Empirical models allow researchers to conceptualize hypotheses, generalize relationships, and develop meaningful explanations of problems (Burnham et al. 2011). Models with higher explanatory power can be derived from more simple models and be tested to accept or reject alternative explanations (Kirwan et al. 2009). Linear mixed-effect models are excellent statistical tools when databases are unbalanced and/or data have a hierarchical structure and as a result do not fulfill the assumption of independence (Smith et al. 2005).

The goals of this study were to (i) determine the main climatic and management constraints that explain variability in maize yields in the semiarid and subhumid Chaco, (ii) determine if mechanistic and empirical models can be combined to identify maize constraints, and (iii) identify genotypic characteristics with the potential to alleviate the effects of the identified climatic constraints.

2 Materials and methods

2.1 Study area

The assessment was performed in two regions of the Gran Chaco known as the subhumid and semiarid Chaco (Adamoli et al. 2011). These two regions cover a great proportion of the Northern Argentinean provinces of Santiago del Estero and Chaco (Fig. 1). In addition, the studied regions exhibit one of the highest rates of deforestation in South America caused by an expansion of agriculture, which was motivated by higher agricultural prices and changes in land tenure (Vallejos et al. 2015). In 2016, maize was planted in 661,560 ha in Santiago del Estero and 166,150 ha in Chaco.

The subhumid and semiarid Chaco have annual precipitation ranges (1901–2011) from 750 to 950 mm and from 500 to 750 mm, respectively. Annual precipitation follows a monsoonal pattern, with higher precipitation in summer (December to March) than in winter (June to September). It is in these areas that some of the highest absolute maximum temperatures of South America have been recorded (Naumann 2006), and every year, summer crops such as maize are exposed to temperatures above 30 °C during late-vegetative and reproductive stages, making them frequently affected by heat stress. Agricultural soils are mainly Haplustolls and Argiustolls.

2.2 Field data

Data on maize production were obtained from farms (85% of the observations) and genotype evaluations on fields (15% of the observations). The farm data were gathered by an association of farmers (AACREA: Asociación Argentina de Consorcios Regionales de Experimentación Agrícola) while the genotype evaluation data were from two different sources: (i) an official maize genotype evaluation network and (ii) an evaluation conducted by a seed company (Don Mario S.A., Chacabuco, Argentina).

The consolidated dataset included data of 792 production paddocks and genotype evaluation plots from five growing seasons (2010/11, 2011/12, 2012/13, 2013/14, and 2014/15) across five locations in the semiarid Chaco and 14 locations in the subhumid Chaco (Fig. 1). The dataset included grain yield of maize for each production paddock and variables that characterize the paddocks in terms of previous management (e.g., years that the paddock has been cropped, years under no-till, and crop rotation) and maize management (e.g., maize hybrid, sowing and harvest dates, sowing density, crop arrangement, fertilizer application, and crop protection measures). From the 792 paddocks, 34 were fertilized with N at sowing with rates from 20 to 100 kg N ha⁻¹. Three different types of maize hybrids grown in the area were included: tropical, temperate,

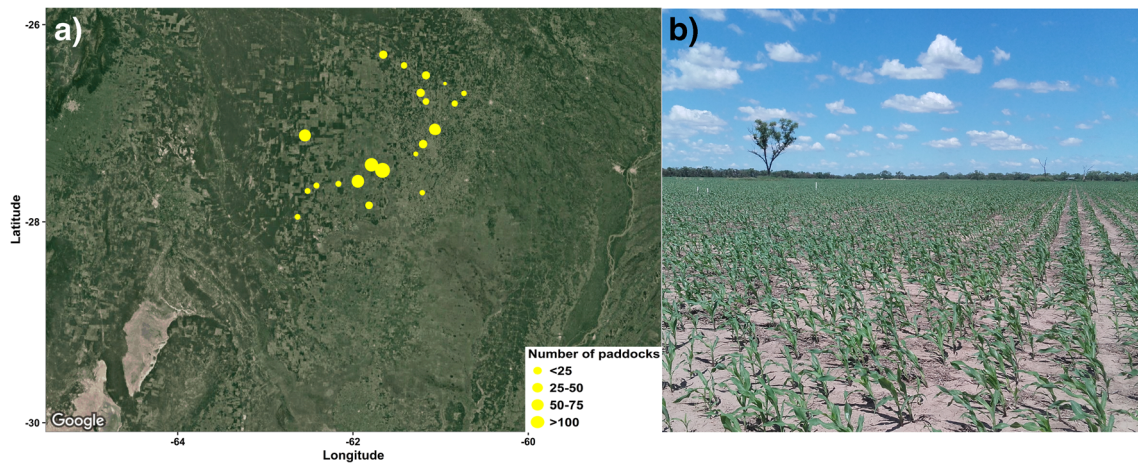


Fig. 1 **a** Location of the studied area in the subhumid and semiarid Chaco and spatial distribution of on-farm data included in the analysis. **b** Maize field at the three-leaf stage

and crosses between tropical and temperate inbred lines. Tropical hybrids compared to temperate ones showed better adaptation in environments prone to high temperature stress but tended to show undesirable agronomic traits such as tall plants, excessive foliage, long cycle, and poor harvest index. The crosses between tropical and temperate inbred lines represented a highly heterogeneous group. Most genotypes were genetically modified organisms with traits for insect resistance.

2.3 Environmental data for model inputs

Global solar radiation (total shortwave radiation onto a horizontal surface) and air temperature (1.5 m above soil surface) were obtained from six automatic weather stations (most of them from Davis Instruments, California, USA) of INTA (Instituto Nacional de Tecnología Agropecuaria; <http://siga2.inta.gov.ar/>) at Las Breñas, Añatuya, Gancedo, Los Frentones, Quimili, and Sacháyoj. Most of the precipitation data were collected at each farm on a weekly basis. For the few cases where local precipitation data were not available, data from the closest weather station were used. For some combinations of sites and years, solar radiation data was not available. In these cases, we used data from NASA (<http://power.larc.nasa.gov/>) because recent assessments showed that in flat areas such as Chaco, the correlation between measured and NASA's estimated solar radiation was reliable (Van Wart et al. 2013). We confirmed the validity of this relationship for the studied area with data from the weather stations located at Las Breñas and Añatuya ($r = 0.97$). For the few cases with missing temperature data, we made linear regressions between monthly means of minimum and maximum temperatures and between the weather stations and NASA data. Although correlations were significant and had high coefficients of determination ($R^2 > 0.8$), NASA data tended to slightly overestimate

temperatures. Due to lack of other alternatives for missing temperature data, we still used NASA data but corrected with a function obtained from the previously described correlations. From the weather database (i.e., precipitation, solar radiation, minimum and maximum temperature), we calculated effective precipitation, potential evapotranspiration according to Penman-Monteith and Hargreaves, precipitation-evapotranspiration ratios, and photothermal quotients. In addition, we calculated 12 different heat indexes based on the number of days and the degree days accumulated in which maximum temperature was above six threshold values (i.e., 20, 25, 30, 35, 40, and 45 °C). We used a threshold of 35 °C following previous studies (e.g., Lobell et al. 2013), and we considered the additional thresholds, above and below 35 °C, to determine (i) if there was an influence of the threshold values on the relationships and (ii) if there was a value different from 35 °C that should be used as a reference of heat stress in the studied regions. All the calculated variables were estimated for the entire growing season of maize, monthly periods, and for ten growth stages defined according to thermal phases. Overall, these calculations yielded 304 variables that were considered as potential predictors in empirical models to predict the grain yield of maize. Although we expected this to generate redundant variables, we still calculated them to be sure that all potential predictors with a high predictive power were considered. An additional reason to consider a high number of predictors was to address the fact that two redundant predictors may have similar explanatory power but may differ on how the information from that predictor can be translated into agronomic practices and practical interventions. Soil characterization included soil series, soil taxonomy, and soil use. The used soil data were provided by INTA. For the mechanistic model, hydraulic parameters, defined by wilting point and field capacity, were estimated by pedotransfer functions (Saxton et al. 1986), soil water saturation was estimated as a

function of soil porosity (Padarian et al. 2014), and bulk density was estimated following Rawls and Brakensiek (1989). In addition, farmers were asked to assess water availability at sowing according to a three-category scale (i.e., bad, intermediate, or good).

2.4 Empirical modeling: single and multiple constraint linear mixed models

We were interested in identifying the most relevant environmental and management variables determining grain yield variability of maize in the studied area and in quantifying the magnitude of their effects. To define the model that best describes the observed data, we considered a linear mixed-effects model since these models can handle unbalanced data and missing observations, and are particularly useful where measurements are made on related statistical units (Smith et al. 2005).

Although the final goal was to count with a multivariate model that explains variability in grain yield of maize, we began by developing single-constraint models for each group of similar variables. These groups were temperature, heat stress, water availability, management (i.e., sowing date, row distance, plant density, fertilization, type of maize hybrid), and soil characteristics (i.e., soil taxonomy, use capacity). Single-constraint models were of the type: $\text{Yield} = \mu + \text{Predictor} + (\text{Site} \times \text{Year}) + \epsilon$; with Yield: the prediction of grain yield associated with environments ($\text{Site} \times \text{Year}$), μ : mean grain yield, Predictor: variable evaluated as predictor. We defined environment as the year-site combination in order to take into consideration the fact that some observations came from nearby sites and from the same site in different years. The predictor was set as a fixed effect factor while the couple $\text{Site} \times \text{Year}$ was set as a random factor.

Once the most explicative single constraints within each group of similar variables were identified, we considered their potential use for the multiple constraint model. We followed an established protocol (Zuur et al. 2010) to check for (i) outliers, (ii) homogeneity of variance, (iii) normal distribution, and (iv) independence and type of relationship of the candidate variable with the response variable. Multicollinearity and relationships among quantitative variables were evaluated with correlations according to Pearson. Spatial trends, using geographic coordinates, were also explored as a potential source of patterns modifying the model's performance. To identify deviations from homoscedasticity or normality, we inspected visually residual plots for deviations and checked Gaussian and homoscedasticity assumptions for the standardized residuals of the models. The random structure of the multiconstraint model was set as the year-site combination representing environments, and it was done in the same way for the single-constraint models. Variance heterogeneity across

environments was specifically checked by fitting a model per environment and comparing the residual variance. Observations of candidate fixed effect predictors were standardized by z -scores to address the fact that predictors had very different scales; z -scores do not modify the functional relationship between the response and predictor variables. To select predictors for the final model, we followed the top-down strategy of model selection and the multimodel inference approach based on information theory (Burnham et al. 2011). The multimodel inference approach does not rely on the assumption that there is a unique "true model" but rather that model selection can identify the best approximating model that will summarize which "effects" (represented by predictors) can be supported by the data. Selection of model predictors was based on the AIC (Akaike information criterion) criteria (Burnham et al. 2011). The coefficients of the final model were estimated using REML (restricted maximum likelihood). Goodness of fit of mixed models was assessed with R^2 of adjusted models following Nakagawa and Schielzeth (2013). Marginal R^2 represents the variance explained by fixed factors, while conditional R^2 represents the variance explained by the entire model (fixed and random effects).

2.5 Mechanistic modeling: CERES-Maize

Because of the lack of reports validating or using CERES-Maize in the studied area, it was necessary to evaluate the suitability of CERES-Maize (Jones et al. 2003) for these regions. We evaluated CERES-Maize v4.5 with data from the consolidated database, except observations that had one or more of the following characteristics: (i) crop management was unusual for the area, (ii) precipitation during the growing season was unusually low (< 100 mm) or unusually high for the studied area (> 550 mm), and (iii) grain yields were unusually low (< 2000 kg ha⁻¹) or unusually high ($> 11,500$ kg ha⁻¹). The previous thresholds were chosen to use CERES-Maize with scenarios that represent average growing seasons. For example, the precipitation threshold of 550 mm was used because precipitation was greater than 550 mm in only 5% of the cases. Overall, the dataset used with CERES-Maize covered 69 paddocks (10% of the paddocks in the consolidated dataset) across four growing seasons (2010/11, 2011/12, 2012/13, and 2013/14); four locations in the semiarid Chaco: Girardet, Otumpa, Quimili, and Roversi; and nine locations in the subhumid Chaco: Campo del Cielo, Campo Largo, Charata, Gancedo, La Paloma, Las Breñas, Loro Blanco, Los Frentones, and Pampa del Infierno.

Simulations were conducted for the temperate hybrid DK 747 because it was the most grown temperate hybrid in the area during the duration of the study and data for model calibration were available. Genetic coefficients for the hybrid DK 747 were derived from previous studies (e.g., Aramburu

Merlos et al. 2015) and unpublished data from well-managed experiments.

The soil data used with the model (i.e., percentage of clay and silt, organic carbon and total nitrogen, color, drainage, and runoff potential) were obtained from the soil profiles described by the Institute of Soils of INTA. For the identification of the soil profiles in the localities of Santiago del Estero, the geographical information system of Santiago del Estero (SigSE) (Angueira et al., 2007) and the GeoINTA viewer were used (scale 1:500,000) (<http://geointa.inta.gov.ar>). For the soils of Chaco, we used the soil chart of the Argentine Republic, Province of Chaco (scale 1:50000) and the GeoINTA viewer. In each locality, a dominant series was identified, and when it presented an association of soils, we used the series with the highest percentage of representation that was suitable for agricultural use as the soil of one paddock.

Argillic horizons restrict root growth and water uptake (Dardanelli et al. 1997). Therefore, we considered the value of the soil root growth factor according to the percentage of clay: we used a value of 1.0 when clay < 32%, 0.4 when 32% < clay < 40%, and 0.2 when clay > 40% (Dardanelli et al. 1997). We considered a soil depth of 2 m based on studies that report this value as the average rooting depth of maize in similar soils (e.g., Dardanelli et al. 1997). For those sites where soil data were not available for the 2 m profile, we repeated the values of the last soil depth available until the 2 m were completed. Initial soil nitrate and soil ammonium availability were set at 70 and 15 kg N ha⁻¹, respectively, based on soil analysis conducted by local agronomists. These N contents can be taken as the usual values at these locations and sowing dates. We assumed an exponentially decreasing N distribution with depth following reports in the literature (Rimski-Korsakov et al. 2012). To parameterize water availability at sowing, we relied on the qualitative assessment that farmers did at sowing. Among the simulated paddocks, only five had been fertilized with urea (46% N) at sowing with 62 kg N ha⁻¹ on average.

The meteorological data for each simulation were obtained from the meteorological station (see Section 2.3) closest to the field (less than 60 km away). The crop management data for each simulation were those of the paddock, and one simulation was conducted for each paddock. For some paddocks, data were missing for row distance and sowing density; we used the most frequent values of 52 cm and 6 plants m⁻² for row distance and sowing density, respectively. In all cases, we used a sowing depth of 5 cm. Although there were farmers who sowed maize at other dates, the sowing dates used to run CERES-Maize were between December 23 and January 20 since these were the sowing dates used in the paddocks that were modeled. When there was no information about the preceding crop, we assumed that it was soybean as it is the

most common crop in the study area. Grain yield of preceding crops was obtained from an official repository of agricultural information (<http://www.siaa.gov.ar>). Residues (shoot and root biomass) were estimated according to Álvarez et al. (1998), and we assumed no incorporation of residues since no incorporation of residues was performed after the harvest of the preceding crop until the sowing of maize. Shoot biomass was estimated using grain yield (0% moisture content) and a shoot biomass-grain ratio of 1.33, 1.9, and 1 for soybean, wheat, and maize, respectively. Root biomass was estimated as 20% of shoot biomass. We assumed an N content of 1% in the residues. We entered the harvest date of the preceding crop as the date of simulation start, and this was generally 6 months before planting maize when the preceding crops were soybean, maize, and cotton and 2 months before when the preceding crop was wheat.

Simulations were conducted setting constraints on water and nitrogen and choosing the Priestley-Taylor method to simulate evapotranspiration. Water infiltration was simulated following the approach of the Soil Conservation Service (USDA Soil Conservation Service, 1972).

After evaluating the crop model, we explored the relationship between simulated grain yield and water availability for DK747 at sowing and during February. Available water (0–2 m) at sowing was estimated by running the crop model throughout the fallow period. The starting date of the fallow simulation was set according to the harvest date of the preceding crop.

2.6 Statistical analyses

Package lme4 (Bates et al. 2015) of the statistical software R (R Development Core Team 2007) was used to perform a mixed-effects analysis of the relationship of grain yield with environmental and management predictors. Using the final multi-constraints mixed model, we performed a variance component analysis to determine if a few single factors explained most of the variance. We compared the model having only mixed-effects predictors with other models incorporating fixed-effects predictors.

CERES-Maize performance was evaluated according to two deviation metrics: RMSE (root mean square error) and NRMSE (normalized RMSE) (Wallach et al. 2014). We quantified water productivity of the maize hybrid DK 747 by means of conventional and 95% quantile boundary regressions fitted to simulated data. The latter was done using quantile regression as implemented in the R package quantreg (Koenker 2016). Grain yield values within the 95th percentile of each precipitation class were regressed against precipitation and the fitted model was taken as the maximum bound for water productivity.

3 Results and discussion

3.1 Main database results

The consolidated dataset covered the typical growing conditions of summer crops in the studied area. Most of the selected sites have been cropped more than 5 years and soybean was the previous crop in approximately two-thirds of the observations. Crop yield averaged 6156 kg ha⁻¹ but showed a great variation; with minimum and maximum values of 297 and 11,015 kg ha⁻¹ (14.5% moisture content). Most local farmers (87%) sowed maize in a relatively narrow window that spanned from December 15 to January 31. Compared to the humid Pampas, local farmers delay sowing dates to make the water balance less restrictive during critical periods as a measure to reduce the climatic risk (Giménez et al. 2015). An additional advantage of the delayed sowing date is that maize growth occur during periods of higher solar radiation. However, delayed sowing is not free of disadvantages and maize is exposed to higher temperatures and heat stress during advanced growth stages.

Farmers cropped maize mostly at an inter-row spacing of 52 cm (81%) while 14 and 5% of farmers used 76 and 70 cm, respectively. Average plant density ranged from 3.8 to 7.6 plants m⁻² and 63% of the paddocks were treated with insecticides. In the study area, there is an ongoing process of replacing tropical by temperate hybrids which was captured in our database. While in 2010, the percentage of temperate hybrids was 32%, in 2013, it was 68%. Considering all the years of the database, the most commonly grown maize hybrid was the tropical DK 390 (relative maturity 150 days) (22% of the cases), followed by the temperate hybrid DK 747 (relative maturity 125 days). On average, the length of the maize growing season was 185 days.

3.2 Single-constraint models

Single-constraint models allowed identification of the most explicative variable within groups of similar variables and reduction of the number of candidate predictors for a multiple constraint model (Kirwan et al. 2009). The variables in the consolidated dataset allowed the evaluation of 40, 73, 59, 10, and 14 single-constraint models of the temperature, heat stress, water availability, soil, and management groups, respectively. Among the evaluated variables for the relationship between temperature and grain yield, average maximum temperature from January to April (TJA) was the predictor that explained the highest amount of variance. Interestingly, the variable that was retained among the temperature variables was the most associated with heat stress. In the case of heat stress, the predictor that explained the highest amount of variance was the sum of degree days above 35 °C (D35). Sinsawat et al. (2004) indicated that 35 °C is above the

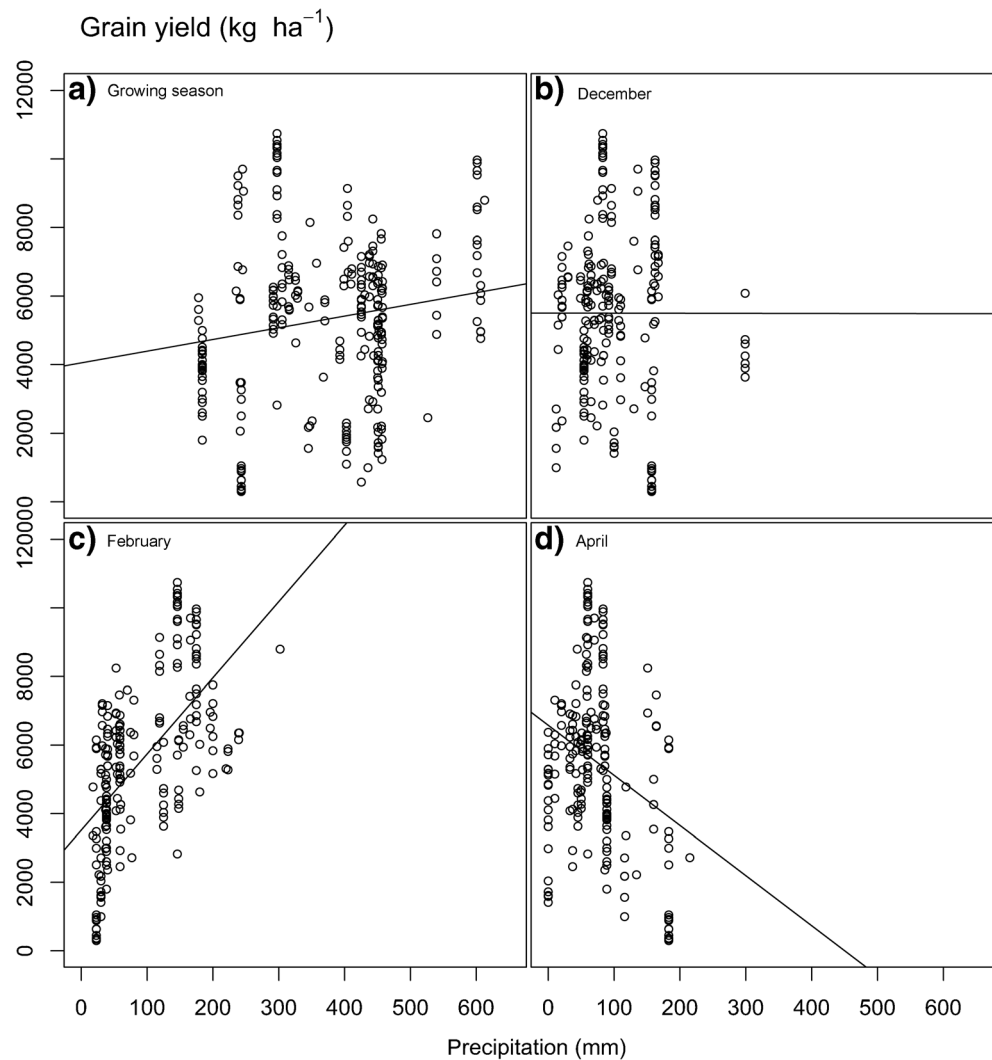
optimal temperature threshold value for maize development, growth, reproductive processes, pollen viability, and grain yield.

In terms of the variables related to water availability, the amount of rainfall during February (Fig. 2) was a better predictor of maize yield than the different indexes of evapotranspiration evaluated. Evapotranspiration may have explained less variance because by delaying sowing dates, the water balance around the critical period may be less restrictive. Only 1.8% of the farmers planted early (September–October). In the case of the predictors associated with rainfall, the relationship was stronger since the interannual variability in rainfall was greater than that of the evapotranspiration (Maddonna 2012). In addition, the precipitation data came from each farm, which could help detect more local variability than the evapotranspiration data that came from only a few meteorological stations. When we evaluated the relationship between precipitation and grain yield, the goodness of fit, indicated by R^2 , was 0.02 (precipitation during the entire growing season, Fig. 2a), 0.005 (December, Fig. 2b), 0.35 (February, Fig. 2c), and 0.09 (April, Fig. 2d). The environmental variables in monthly periods explained higher variance in grain yield than the values of the same variables during the entire growing season because there are periods in which the occurrence of abiotic stress has higher influence on grain yield than during other growth stages (Maddonna 2012). In the case of maize, this period occurs around flowering, when the incidence of water stress negatively affects grain yield more than in any other period (Edreira and Otegui 2012).

A practice that could reduce water limitation is more efficient weed control. However, the data collection scheme did not consider quantifying weed control efficiency. Ricard et al. (2015) suggested irrigation as an effective measure to counteract precipitation variability in the studied area. Double cropping, an efficient water use practice widespread in other parts of the continent, is considered unsuitable in this region because of the high water consumption. However, a reconsideration of double cropping supported by further research is recommended since highly positive water balances are also to be avoided; unutilized water excess represents an environmental risk from erosion, flooding, and salinization (Giménez et al. 2015). Information for Chaco about the effect of weed control on the water balance is also missing and should be collected in future projects. Although, irrigation is practically null in the area, collected data shows that is a promising alternative to reduce yield variability.

Among the management variables, the type of maize hybrid explained a significant amount of variance (Table 1). This result suggests that farmers can influence the attainable maize yield by not only choosing a specific hybrid but also by selecting between tropical or temperate hybrids. Indeed, genotype choice is one of their primary options for adapting crop production to climatic conditions (Maddonna 2012). Although

Fig. 2 Relationships between grain yield (14.5% moisture content) and precipitation during the whole growing season (a), December (b), February (c), and April (d) at selected analyzed periods. Data were obtained from 792 maize production paddocks in the subhumid and semiarid Chaco. Lines are linear regressions between grain yield and precipitation



the amount of variance explained was lower than for the type of maize hybrid, the preceding crop and number of years the site has been cropped also explained significant amounts of variance on the grain yield of maize. In contrast, all the evaluated variables associated with soil were poor predictors of grain yield. A qualitative variable that classified sites

according to whether they were located in the semiarid or subhumid Chaco was also not retained by the models, suggesting that this classification may have little operational application for modeling purposes. Although other studies reported significant associations between plant density and maize grain yield (e.g., Andrade et al. 2017), in our case, plant

Table 1 General structure of multiple constraint models for predicting maize yields in the subhumid and semiarid Chaco shown with selected examples. Models differed in the predictors included and in the additive

or interactive relationship among predictors. We arrived at a final model (# 2) by a step-wise selection process according to Akaike's information criterion (AIC)

Models	Predictors	AIC	Marg. R^2	Cond. R^2
1. Full model	$Yield_i = D35 \times TJA \times RF \times HT \times (Site \times Year)_i + \epsilon_i$	3816	0.55	0.66
2. Heat \times temperature	$Yield_i = D35 + TJA + RF + HT + D35 \times TJA \times (Site \times Year)_i + \epsilon_i$	3722	0.40	0.60
3. No heat and interactive effects	$Yield_i = TJA \times RF \times HT + (Site \times Year)_i + \epsilon_i$	8920	0.32	0.65
4. No heat and additive effects	$Yield_i = TJA + RF + HT + (Site \times Year)_i + \epsilon_i$	8990	0.32	0.65
5. No heat and no genotype	$Yield_i = TJA \times RF + (Site \times Year)_i + \epsilon_i$	9247	0.31	0.56

Marg. R^2 represents the variance explained by fixed factors; Cond. R^2 represents the variance explained by the entire model; D35, degree days accumulated above 35 °C; TJA, average maximum temperature from January to April; RF, rainfall during the month of February; HT, type of maize hybrid; ϵ , error

density did not explain variability in grain yield. This result may be explained by the fact that farmers in the area used a very narrow range of plant densities.

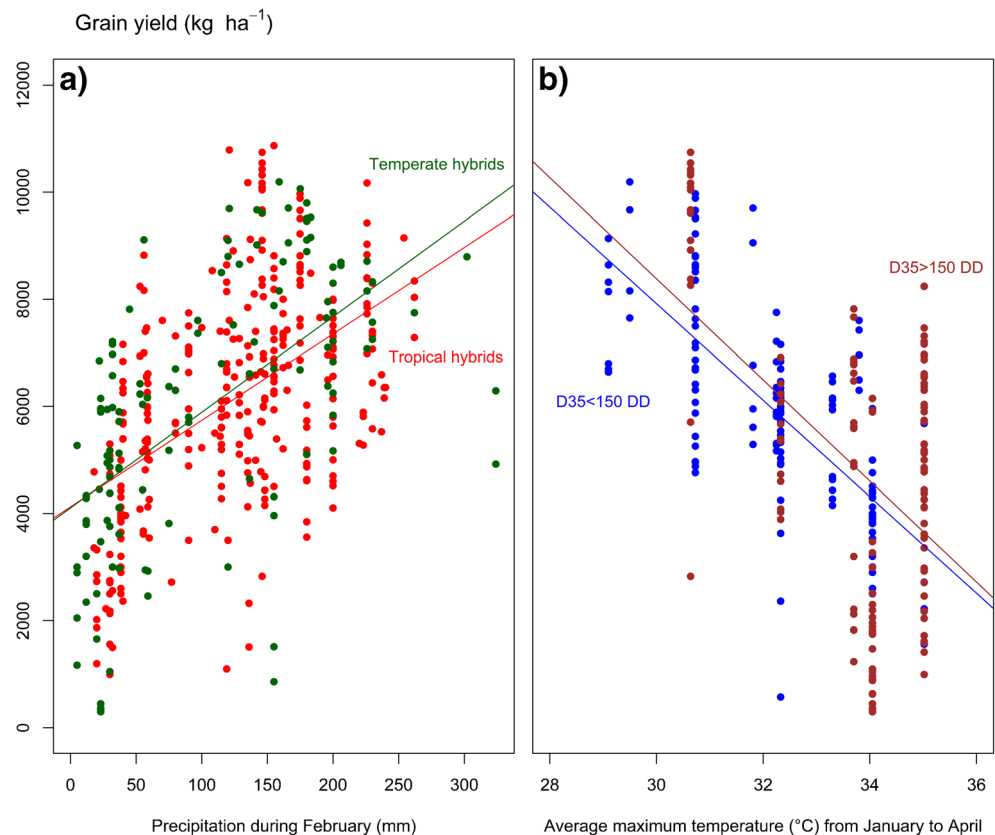
3.3 Multiple constraint models

After identifying predictor candidates with single-constraint models, we developed a multiple constraint model with the goal of summarizing the main environmental and management factors that influenced maize grain yield in the studied area and of understanding whether the relationships among variables were of an additive or interactive type. From a longer list of candidate models, Table 1 shows five examples of candidate models to document our model selection criteria and procedure. Among the considered models, model 2 showed the best overall fit. Model 2 showed the lowest AIC, which is the most informative criterion for multivariate model selection (Burnham et al. 2011). The higher conditional R^2 of other models could show a tendency of these models to be overfitted, while the marginal R^2 , which indicates the variance explained by the fixed factors, was higher for model 2 than for all the other simplified models. In general, models that ignored heat stress had worse approximations of the data than those including it. To illustrate this, Table 1 shows models where D35 was included (models 1 and 2) and not included (models 3 to 5) as a predictor. A full model including interactions among all the identified single predictors (model 1 in

Table 1) had a better fit than other combinations of interactive and additive predictors but a worse fit than the selected model 2 that included all parameters as additive except the interaction between an index of heat stress (i.e., D35) and TJA (average maximum temperature from January to April).

Figure 3a shows the relationship between grain yield and precipitation during February. Although the interaction between type of hybrid and precipitation did not have a significant effect on grain yield, there was an additive effect of the type of hybrid on grain yield. With low precipitation, temperate hybrids tended to perform better than tropical ones, which supports the ongoing adoption of temperate hybrids by local farmers. The adoption of tropical hybrids was encouraged in the past because the expectation that they would be better adapted to the local temperature and photoperiod requirements. Figure 3b shows the relationship between grain yield and TJA and the interactive effect of a heat stress index (D35) on this relationship. As expected, grain yield of maize diminished as TJA increased over the already high average of 29 °C that was measured in the studied area. As could also be expected, the heat stress index was correlated with TJA ($r = 0.58$ according to Pearson's product-moment). However, the correlation between these two variables was not tight enough to prevent different intensities of heat stress for a similar TJA. These combinations of maximum temperatures and heat stress have implications for maize productivity, and they were captured by the multiple constraint model. The significant

Fig. 3 **a** Relationship between grain yield (14.5% moisture content) of tropical and temperate maize hybrids and precipitation during February. **b** Relationship between grain yield and average maximum temperature from January to April (TJA) as influenced by a heat stress index (D35). Data were obtained from 792 maize production paddocks in the subhumid and semiarid Chaco. Lines are linear regressions fitted to the data



interaction between temperature and heat stress is visible by the fact that at the lowest extreme of TJA (i.e., < 30 °C), there were no observations with D35 > 150 DD (degree days), while at the highest extreme of TJA (i.e., > 34 °C), there were no observations with D35 < 150 DD. At TJA of approximately 31 °C, grain yield tended to be higher with D35 > 150 DD, while at approximately 34 °C, grain yield tended to be higher with D35 < 150 DD. Thus, more heat events resulted in lower grain yield with higher temperatures, but not with lower temperatures. Although this effect may look subtle, it indeed shows how complex the dilemma is for local farmers to choose a type of hybrid, sowing dates, and other management decisions that mediate the effect on productivity of two climatic effects that in a narrow range can result in very different outcomes. This result, in turn, suggests that besides improvements from breeding or better crop management, strategies towards diversification (e.g., rotations or genotypes) are advisable to minimize the risk of climatic effects that are difficult to anticipate.

With the selected multiple constraints model (model 2), we conducted a variance component analysis to identify the impact of these sources of variability on grain yield. Eight percent of the variance was ascribed to the type of hybrid, a factor that can be controlled by farmers. In contrast, most of the effects on grain yield were explained by environmental variables for which farmers have no or indirect control; D35 (11%), average maximum temperature from January to April (26%), rainfall during February (13%), the interaction D35 × TJA (13%), and the interaction Site × Year (11%). However, the choice of hybrid had implications on the effects that the environmental variables exert on grain yield. When we circumscribed the variance component analysis to a subset with only tropical genotypes, precipitation in February explained 24% of variance in grain yield, the environment (i.e., Site × Year) 17% and TJA 29%, whereas heat stress alone or interacting with temperature explained almost no variance. The partition of variance was rather different for temperate hybrids; i.e., 20, 22, 34, and 8% were explained by the interaction heat × TJA, heat alone, temperature alone, and precipitation, respectively. The respective values for the tropical × temperate were 6, 18, 36, and 10%. Similarly, in field experiments carried out in the humid Pampas of Argentina, the effects of heat stress on grain yield were larger for temperate than for tropical hybrids (Edreira and Otegui 2012). An analysis covering the entire Chaco region highlights the role of water availability as another factor determining maize yields (Adamoli et al. 2011). Although our results show that water availability is one important constraint on maize productivity, depending on the type of hybrid it may not be the one explaining the higher amount of variance. Therefore, our results support the conclusions of other studies (e.g., Baldi et al. 2015) that multiple factors should be taken into consideration in the

design of improved agronomic management and farming systems in semiarid and subhumid subtropical areas.

Although the type of hybrid explains less variability than other variables, it would be a strategic factor to increase maize's productivity in the area because through the choice of genotypes farmers can influence the effects that climate constraints have on maize productivity more easily or affordably than through other interventions. Farmers have no means to directly influence temperature while influencing soil moisture with irrigation needs major investments that most local farmers cannot afford. Tropical hybrids were less affected by heat stress and higher temperatures than temperate ones. However, tropical hybrids were more affected by precipitation than temperate ones, presumably due to their higher biomass and the associated higher consumption of water (Edreira and Otegui 2012). The rationale of combining two types of hybrids on one farm is to cope with uncertainties that arise from climate variability. The crosses between tropical and temperate inbred lines represented a highly heterogeneous group that as a category did not show clear cut differences compared to tropical or temperate hybrids. Although this scenario could have been an advantage in terms of adaptation and stress tolerance, we did not observe this. This result may be explained because the crosses included in this study were the first attempts to generate this type of genotype in the region and the physiological traits from tropical lines that confer advantages for heat tolerance were probably not targeted.

An unexpected outcome of the model selection was that N fertilization was not retained in the final multiple constraint model and it was not identified as a predictor with acceptable explanatory capacity by a single-constraint model. Three hypotheses may explain this outcome. First, the observed variability in N fertilization rates was low for adequately testing N effects and the real soil N content was uncertain. Second, the soil N availability could be high enough for summer crops due to the local soils that have been only recently cropped and still hold the fertility of virgin soils, and the high summer temperatures exacerbate N mineralization, generating N pulses during the periods of high demand. Third, limitations imposed by N availability may have been obscured by other constraints, e.g., drought and heat stress. In general, local farmers consider that the application of fertilizers does not assure economic benefits, local soils are still fertile enough to achieve satisfactory maize yields, and/or climatic variability makes the response to N fertilization highly uncertain.

3.4 Mechanistic modeling

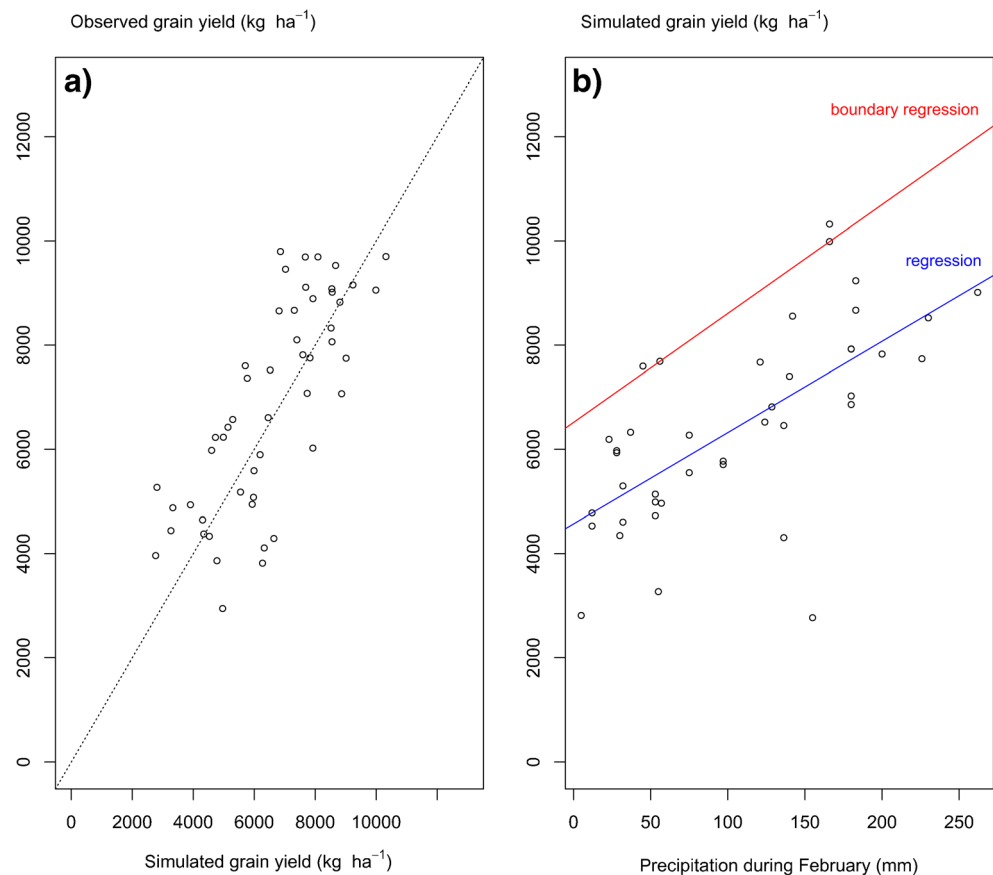
Mechanistic models became powerful tools to assess the effects of climate change and climate variability (Ray et al. 2015). In this study, the goal of using CERES-Maize was to extend the conclusions suggested by the field data, to increase statistical power, to solve uncertainties, and to improve effect

size estimation. An initial evaluation of CERES-Maize for the most used temperate hybrid in the area (DK747) showed that RMSE, NRMSE, and R^2 of the regression between simulated and observed grain yields were 1680 kg ha^{-1} , 25%, and 0.38, respectively. A question in this study was to better understand why CERES-Maize performs generally poorly in subtropical areas (e.g., Ray et al. 2015). The opportunity to overcome this lack of accuracy has implications for our understanding of the potential effects of global climate change and for aligning sustainable development efforts taking into consideration these effects. Therefore, we searched for sources of deviations between simulated and observed values among variables associated with weather, crop characteristics, management, and environment. We could not identify a consistent source, such as heat stress or intense drought, for reducing the accuracy of CERES-Maize. We identified three sources of deviations: (i) biases at specific locations (Charata, Loro Blanco, and Girardet that we attributed to inaccurate soil data), (ii) biases at specific fields (two fields in which the soil was probably degraded to some extent), and (iii) biases from specific sowing dates (three observations from 2014 where there were problems in the emergence of maize). When observations had one of the above-mentioned characteristics were not included in the subdataset, RMSE, NRMSE, and R^2 of the regression between

simulated and observed values were 1246 kg ha^{-1} , 16%, and 0.61, respectively. Therefore, CERES-Maize performance was satisfactory (Fig. 4a), and it included a wide range of combinations of factors to support decisions in a relatively new cropping area where locally available information is scarce.

Mechanistic models allow researchers to extend the scope of observational results (Jones et al. 2003). This is an important opportunity to overcome a potential problem associated to mining observational data, which is that for certain response variables the range of values is incompletely covered by field data (Rosenbaum 2002), and data may be characterized by forming discrete groups of data within the observed ranges. This process can be a source of inaccuracy since regressions can be the result of a few groups having high leverage effects. After evaluating CERES-Maize for the studied area, we conducted simulations to quantify the water productivity of this hybrid. Figure 4b shows conventional and 95% quantile boundary regression analyses applied to the simulated data. Our analysis with conventional regression indicated that the hybrid DK 747 has a potential to increase grain yield by 18.5 kg ha^{-1} (14.5% moisture content) for every millimeter of rainfall during February. We additionally used a boundary regression fitted to the 95% quantile of the simulated data to assess the maximum water harvest potential of the temperate

Fig. 4 **a** Relationship between the grain yields (14.5% moisture content) of the temperate maize hybrid DK-747 simulated by CERES-Maize and grain yields observed on farms located in the subhumid and semiarid Chaco. Dotted line represents 1:1 relationship. **b** Relationship between simulated grain yields of the temperate maize hybrid DK-747 by CERES-Maize and precipitation during February (mm). Solid and dashed lines represent regression and boundary regression of the 95% quantile, respectively



hybrid DK 747. The results indicated a maximum possible increase in grain yield by 21.3 kg ha⁻¹ (14.5% moisture content) for every millimeter of rain fell during February. Even though the mechanistic model left unanswered questions, it significantly reduced the number of queries that were posed at the inception of this project and will allow a more efficient investment of resources in subsequent projects by focusing only on those questions that remain open.

4 Conclusion

The combination of empirical and mechanistic modeling of farm data allowed the identification of constraints to maize production in an area where maize cropping is in its early stages. The empirical model identified the amount of rainfall during February as a primary determinant of maize yields. Based on these observations, CERES-Maize simulations indicated that suitable temperate hybrids have the potential to increase grain yield from 18 to 21 kg ha⁻¹ (14.5% moisture content) for every millimeter of rainfall during February.

An additional key contribution of the multivariate mixed model was to elucidate the role of genotypes since most mechanistic models are still not advanced enough to capture differences at the genotypic level (Jeuffroy et al. 2014). Our empirical model showed that temperate hybrids tended to perform better under conditions of water scarcity while tropical hybrids tended to withstand better conditions that arose from higher temperatures and heat stress (temperature > 35 °C). This situation suggests that farmers face a difficult dilemma when choosing between temperate and tropical hybrids to reduce vulnerability to drought and heat stress, which are two stresses that tend to occur simultaneously. Farmers are thus exposed to a deadlock where the main option in terms of genotype choice to minimize climatic risks may be to diversify and mix them in different paddocks within a farm as a measure to minimize the overall climatic vulnerability. The findings of this study provide plant breeders urgently needed information to breed better adapted maize genotypes for these regions, which in turn can increase local farmers' options.

Acknowledgements Financial support was provided by CONICET, ANPCYT and UBA. We are grateful to Guillermo Peralta, Jose Micheloud, Luciana Bolañez, Salvador Prieto, Hernan Ingrassia, Nestor Gomez, Tomas Nolting and to the members of AAPRESID and INTA Quimili for their valuable help.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Adamoli JM, Guinzburg R, Torrella S (2011) Escenarios productivos y ambientales del Chaco Argentino: 1977–2010. In: Fundación Producir Conservando (ed). Buenos Aires, p 101
- Alvarez R, Alvarez CR, Daniel PE, Richter V, Blotta L (1998) Nitrogen distribution in soil density fractions and its relation to nitrogen mineralisation under different tillage systems. *Aust J Soil Res* 36: 247–256
- Andrade JF, Poggio SL, Ermácora M, Satorre EH (2017) Land use intensification in the Rolling Pampa, Argentina: diversifying crop sequences to increase yields and resource use. *Eur J Agron* 82:1–10. <https://doi.org/10.1016/j.eja.2016.09.013>
- Angueira MC, Prieto DR, López J, Barraza G. 2007. SigSE 2.0. Sistema de Información Geográfica de Santiago del Estero. Ediciones INTA. Buenos Aires, Argentina
- Aramburu Merlos F, Monzon JP, Mercu JL, Taboada M, Andrade FH, Hall AJ, Jobbagy E, Cassman KG, Grassini P (2015) Potential for crop production increase in Argentina through closure of existing yield gaps. *Field Crop Res* 184:145–154. <https://doi.org/10.1016/j.fcr.2015.10.001>
- Baldi G, Houspanossian J, Murray F, Rosales AA, Rueda CV, Jobbágy EG (2015) Cultivating the dry forests of South America: diversity of land users and imprints on ecosystem functioning. *J Arid Environ* 123:47–59. <https://doi.org/10.1016/j.jaridenv.2014.05.027>
- Bates D, Mächler M, Bolker BM, Walker SC (2015) Fitting linear mixed-effects models using lme4. *J Stat Softw* 67:1–48. <https://doi.org/10.18637/jss.v067.i01>
- Brisson N, Guevara E, Meira S, Maturano M, Coca G (2001) Response of five wheat cultivars to early drought in the Pampas. *Agronomie* 21: 483–495
- Burnham KP, Anderson DR, Huyvaert KP (2011) AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behav Ecol Sociobiol* 65:23–35. <https://doi.org/10.1007/s00265-010-1029-6>
- Dardanelli JL, Bachmeier OA, Sereno R, Gil R (1997) Rooting depth and soil water extraction patterns of different crops in a silty loam haplustoll. *Field Crop Res* 54:29–38. [https://doi.org/10.1016/S0378-4290\(97\)00017-8](https://doi.org/10.1016/S0378-4290(97)00017-8)
- Edreira JIR, Otegui ME (2012) Heat stress in temperate and tropical maize hybrids: differences in crop growth, biomass partitioning and reserves use. *Field Crop Res* 130:87–98. <https://doi.org/10.1016/j.fcr.2012.02.009>
- Giménez R, Mercu JL, Houspanossian J, Jobbágy EG (2015) Balancing agricultural and hydrologic risk in farming systems of the Chaco plains. *J Arid Environ* 123:81–92. <https://doi.org/10.1016/j.jaridenv.2014.09.004>
- Goldfarb L, van der Haar G (2016) The moving frontiers of genetically modified soy production: shifts in land control in the Argentinian Chaco. *J Peasant Stud* 43:562–582. <https://doi.org/10.1080/03066150.2015.1041107>
- Jeuffroy MH, Casadebaig P, Debaeque P, Loyce C, Meynard JM (2014) Agronomic model uses to predict cultivar performance in various environments and cropping systems. A review. *Agron Sustain Dev* 34:121–137. <https://doi.org/10.1007/s13593-013-0170-9>
- Jones JW, Hoogenboom G, Porter CH, Boote KJ, Batchelor WD, Hunt LA, Wilkens PW, Singh U, Gijsman AJ, Ritchie JT (2003) The DSSAT cropping system model. *Eur J Agron* 18:235–265. [https://doi.org/10.1016/S1161-0301\(02\)00107-7](https://doi.org/10.1016/S1161-0301(02)00107-7)
- Kirwan L, Connolly J, Finn JA, Brophy C, Lüscher A, Nyfeler D, Sebastiá MT (2009) Diversity-interaction modeling: estimating contributions of species identities and interactions to ecosystem function. *Ecology* 90:2032–2038. <https://doi.org/10.1890/08-1684.1>
- Koenker R (2016) quantreg: Quantile regression. R package version 5.21. <http://CRAN.R-project.org/package=quantreg>

- Lavado RS, Taboada MA (2009) The Argentinean Pampas: a key region with a negative nutrient balance and soil degradation needs better nutrient management and conservation programs to sustain its future viability as a world agresource. *J Soil Water Conserv* 64:150A–153A. <https://doi.org/10.2489/jswc.64.5.150A>
- Lobell DB, Asner GP, Ortiz-Monasterio JI, Benning TL (2003) Remote sensing of regional crop production in the Yaqui Valley, Mexico: estimates and uncertainties. *Agric Ecosyst Environ* 94:205–220. [https://doi.org/10.1016/S0167-8809\(02\)00021-X](https://doi.org/10.1016/S0167-8809(02)00021-X)
- Lobell DB, Hammer GL, McLean G, Messina C, Roberts MJ, Schlenker W (2013) The critical role of extreme heat for maize production in the United States. *Nat Clim Chang* 3:497–501. <https://doi.org/10.1038/nclimate1832>
- Maddoni GA (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. <https://doi.org/10.1007/s00704-011-0478-9>
- Nakagawa S, Schielzeth H (2013) A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods Ecol Evol* 4:133–142. <https://doi.org/10.1111/j.2041-210x.2012.00261.x>
- Naumann M (2006) Atlas del Gran Chaco Sudamericano. Deutsche Gesellschaft für Technische Zusammenarbeit. ErreGé & Asoc. Buenos Aires, Argentina
- Padarian J, Minasny B, McBratney AB, Dalgliesh N (2014) Mapping the available water capacity of Australian soils. In: *GlobalSoilMap: basis of the global spatial soil information system—Proceedings of the 1st GlobalSoilMap Conference*. pp 173–179
- Piquer-Rodríguez M, Torella S, Gavier-Pizarro G, Volante J, Somma D, Ginzburg R, Kuemmerle T (2015) Effects of past and future land conversions on forest connectivity in the Argentine Chaco. *Landsc Ecol* 30:817–833. <https://doi.org/10.1007/s10980-014-0147-3>
- R Development Core Team (2007) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
- Rawls WJ, Brakensiek DL (1989) Estimation of soil water retention and hydraulic properties. In: Morel-Seytoux HJ (ed) *Unsaturated flow in hydrologic modeling: theory and practice*. Springer, Netherlands, pp 275–300. https://doi.org/10.1007/978-94-009-2352-2_10
- Ray DK, Gerber JS, Macdonald GK, West PC (2015) Climate variation explains a third of global crop yield variability. *Nat Commun* 6:5989. <https://doi.org/10.1038/ncomms6989>
- Ricard MF, Viglizzo EF, Podestá G (2015) Comparison of adaptive strategies to climate variability in rural areas of Argentine Chaco and US Southern Plains during the last century. *J Arid Environ* 123:68–80. <https://doi.org/10.1016/j.jaridenv.2014.10.009>
- Rimski-Korsakov H, Rubio G, Lavado RS (2012) Fate of the nitrogen from fertilizers in field-grown maize. *Nutr Cycl Agroecosyst* 93:253–263. <https://doi.org/10.1007/s10705-012-9513-1>
- Rosenbaum PR (2002) *Observational studies*, 2nd edn. Springer-Verlag, New York
- Saxton KE, Rawls WJ, Romberger JS, Papendick RI (1986) Estimating generalized soil-water characteristics from texture. *Soil Sci Soc Am J* 50:1031–1036
- Sinsawat V, Leipner J, Stamp P, Fracheboud Y (2004) Effect of heat stress on the photosynthetic apparatus in maize (*Zea mays* L.) grown at control or high temperature. *Environ Exp Bot* 52:123–129. <https://doi.org/10.1016/j.envexpbot.2004.01.010>
- Smith AB, Cullis BR, 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. <https://doi.org/10.1017/S0021859605005587>
- Vallejos M, Volante JN, Mosciaro MJ, Vale LM, Bustamante ML, Paruelo JM (2015) Transformation dynamics of the natural cover in the Dry Chaco ecoregion: a plot level geo-database from 1976 to 2012. *J Arid Environ* 123:3–11. <https://doi.org/10.1016/j.jaridenv.2014.11.009>
- Van Wart J, Grassini P, Cassman KG (2013) Impact of derived global weather data on simulated crop yields. *Glob Chang Biol* 19:3822–3834. <https://doi.org/10.1111/gcb.12302>
- Wallach D, Makowski D, Jones JW, Brun F (2014) Chapter 9—model evaluation. In: *Working with dynamic crop models*, 2nd edn. Academic Press, San Diego, pp 345–406. <https://doi.org/10.1016/B978-0-12-397008-4.00009-5>
- Zuur AF, Ieno EN, Elphick CS (2010) A protocol for data exploration to avoid common statistical problems. *Methods Ecol Evol* 1:3–14. <https://doi.org/10.1111/j.2041-210X.2009.00001.x>