Gaps between farmer and attainable yields across sunflower-growing regions of Argentina: Measurement, significance and implications.

A.J.Hall[a*], C. Feoli[b], J. Ingaramo[c], M. Balzarini[d]

[a]IFEVA, Facultad de Agronomia-Universidad de Buenos Aires/CONICET; Av. San Martín 4453, (C1417DSE) Buenos Aires, Argentina. hall@agro.uba.ar
[b] INTA-ASAGIR Agreement, Instituto Nacional de Tecnología Agropecuaria-Asociación Argentina de Girasol, Av. Corrientes 119, (C1043AAB) Buenos Aires. Argentina. carlosfeoli@asagir.org.ar
[c] ASAGIR, Asociación Argentina de Girasol, Av. Corrientes 119, (C1043AAB) Buenos Aires. Argentina. ingaramovan@arnet.com.ar
[d] Cátedra de Estadística y Biometría, Facultad de Ciencias Agropecuarias, Universidad Nacional de Córdoba/CONICET, Av. Valparaíso s/n, CC 509, (5000) Córdoba, Argentina. mbalzari@gmail.com

ABSTRACT

- Quantifying the magnitude of the gap between yields achieved by farmers and those attainable using good agronomic practice is crucial for prioritisation of research and policy efforts aimed at reducing yield gaps. In this paper we present yield gap estimates obtained from a comparison between selected reporting-district yields (best available estimate for farmer yields) in eight sunflower growing regions of Argentina and those obtained in comparative yield trials in each of those regions.
- Eight consensus rainfed sunflower growing regions within Argentina were established on the basis of perceived environmental and management characteristics. Databases, covering 5 to 9 years for each region, were constructed using official reporting-district (county) yield values and data from comparative yield trials. Harvested crop area in the sampled reporting-districts averaged 1.2 Mha y^{-1}, or *ca.* 50% of the national total for the 1999-2007 period. The attainable yield (comparative yield trial) database included 646 trials conducted over the same time period. Attainable yields were only computed for year/region combinations that included more than 5 trials per region, each including data for at least 7 commercial hybrids. Mixed linear models were used to compute Best Linear Unbiased Estimates (BLUEs) of yield for each region/year combination for both farmer and attainable yields.
- Gaps between farmer (reporting-district) yields and attainable (comparative yield trial) yields were statistically significant ($p \leq 0.05$) for all regions and ranged from 0.37 to 1.18 t ha^{-1} across regions, for a country average of 0.75 t ha^{-1} , equivalent to 41% of the mean country yield of 1.85 t ha^{-1} for the 1999-2007 period. In relative (to average farmer yields) terms, the magnitude of the farmer/attainable yield gaps ranged between 32% and 77% in the five regions that contribute 81% of the national crop. Mean yields for the top decile of comparative yield trial data ranged from 3.2 to 4.2 t ha^{-1} across regions, and the highest yields for this decile in any of the years of record ranged from 3.9 to 4.8 t ha^{-1}. A notable feature of reporting-district and comparative yield trial data was their variability. At reporting-district level within regions, contributions of spatial and temporal variability were roughly similar. The mean relative contribution of the trial effect to non-error variance of the comparative yield trials exceeded 85 % across regions, dominating the contributions of genotype and genotype-by-trial effects.
- The analysis shows that, for the main sunflower producing regions of Argentina, farmer/attainable yield gaps are substantial and exceed the floor of 25% of farmer yields which has been posited as the minimum achievable gap in other crops under prevailing economic and biophysical constraints. The magnitude of these gaps underlines the need for research into their causes and their reduction to more reasonable relative values. Our estimates of attainable yields provide a more appropriate benchmark against which to gauge farmer performance than the occasional high yields in comparative yield trials.
- To our knowledge, this is the first country-wide attempt to quantify yields gaps for sunflower and the only one, for any crop, to use comparative yield trial data combined with appropriate statistical analysis to estimate attainable yields.

Key words: reporting-district, comparative yield trials, rainfed crops, regionalisation.

INTRODUCTION

The yields obtained by farmers for several crop species and in many cropping systems around the world have almost always been shown to be lower than those attainable using locally optimised agricultural best practices and adapted, current, cultivars (e.g., Fischer et al., 2009; Lobell et al., 2009; Laborte et al., 2012).

Various approaches, each of which has particular advantages and disadvantage, have been used for estimating yield gaps. Farmer yields have been estimated using regional or national statistical reports, or by sampling farmers' fields, either directly or using remote sensing. Attainable yields have been estimated using on-farm experiments, yield contest results, research station experiments, 95 % percentile values for reporting-district yields, crop models, and breeders' trials. Research station experiments tend to have limited spatial representativity, especially under rainfed conditions. A previous attempt to use data from breeders trials to estimate attainable yields for rice, wheat and other crops (Aggarwal et al., 2008), a procedure which can provide better spatial coverage than research station experiments, did not use transparent and reproducible procedures for the collection and analysis of the data.

Here we report the results of a yield gap analysis for the sunflower growing regions of Argentina. The analysis was conducted on behalf of the Asociación Argentina de Girasol (ASAGIR), the Argentine sunflower value chain association. The objective was to quantify the farmer/attainable yield gap for this crop. ASAGIR wished to determine whether the magnitude of current yield gaps justified further research into the identification of yield constraints and guide research aimed at gap reduction, and was seeking a quantitative framework which would allow infrequent, but consistent, reports of high grain yields (4 to 5 t ha^{-1}) to be placed in the context of national yield averages in the order of 1.7-1.9 t ha^{-1}. Distinctive features of the analysis are that it applies to rainfed crops (irrigated sunflower crops in Argentina are used almost exclusively for seed production) of current commercial hybrids, it covers eight separate regions of the country, the data for the most important crop-reporting-districts within each region were used to estimate farmer yields, data from breeders' trials were used to estimate attainable yields, and the number of years considered ranged between 6 and 9 according to region.

MATERIALS AND METHODS

Regionalisation

Sunflower is grown extensively in several distinct agroecosystems in Argentina, which are distinguished by seasonal rainfall, radiation and temperature patterns; soil properties (texture, soil depth, organic matter content); the role of sunflower in the cropping system (sole within-season crop, lead crop of a seasonal sequence of two crops); and crop management (time of sowing). For the purpose of this analysis, a consensus set of eight regions was developed with input from breeders, farmers, and traders (Fig. 1). Where areas cropped to sunflower are contiguous across regions, regional boundaries follow reporting-district limits rather than biophysical boundaries. We believe that the use of eight regions provides a reasonable, although admittedly not perfect, approximation to the diversity in climate, soils and management of the current sunflower-growing areas of Argentina.

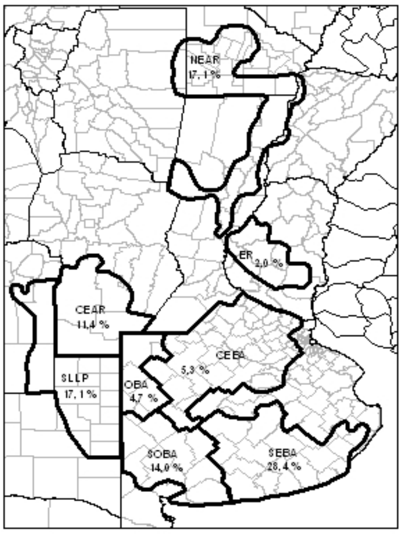

Fig. 1. Distribution of areas cropped to sunflower in Argentina. Reporting-district limits are in thin lines, provincial boundaries are in lines of intermediate thickness, and regional limits are in thick lines as mapped by G. García Accinelli (pers. comm., 2010), LART-FA Univ. Buenos Aires. Acronyms for sunflower growing regions are: South East Prov. of Buenos Aires (SEBA); South West Prov. of Buenos Aires (SOBA); West Prov. of Buenos Aires (OBA); Center Prov. of Buenos Aires (CEBA); San Luis & La Pampa Provinces (SLLP); Central Argentina (CEAR); Entre Ríos Province (ER); North East Argentina (NEAR). Values in brackets next to each acronym indicate % contribution to national harvest (mean 3.86 M t) for the 1999-2007 period.

Estimation of farmer (reporting-district) yields

Yields for reporting-districts making the greatest contribution to the total yield of each region (between 2 and 6 districts [counties] per region) over the 1999-2007 period were obtained from the Agriculture Ministry database (Ministerio de Agricultura, Pesca y Alimentación, 2011) and used as estimates of average farmer yields in each year and region. Harvested area per year and region varied between 15.000 and 500.000 ha, and total sampled harvested area for the eight regions represented about 1.200.000 ha or about 50% of total national harvested area for the 1999-2007 period. This degree of coverage was expected to provide reasonable estimates of mean farmer yields per year and region.

Estimation of attainable yields.

An estimate of attainable yield for each year and region was derived from data obtained in multi-environment yield trials of commercial hybrids. Trials of the ASAGIR-INTA testing network are conducted, in about equal proportions, on INTA (National Institute of Agricultural Technology) research stations and farmers fields. A limited number of the yield trials conducted by seed companies are located on their own research stations, but the majority are conducted in farmers' fields across the sunflower growing area. Almost all trials are conducted using plot sizes larger than 3 rows by 6-7 m, overseeded and later thinned to close to commercial crop population densities (ca. 5 pl. m^{-2}). Yield estimates are obtained harvesting the central row(s) of each plot, and are reported at 11% moisture content. Yield trials are audited (internally by seed companies, INTA -ASAGIR trials by an independent expert), and deficient (i.e., damaged by hail, poor emergence, etc.) trials are discarded. The multi-environment trial data-base used in this study, after filtering (see below), contained a total of 646 field trials including 11411 entry means, one for each hybrid in each trial. Within-trial replicates (when available) were also used in the analyses reported here. Trials were assigned to regions on the basis of location. The data base for each year and region was restricted to trials that included at least 7 commercial hybrids (including at least two entries from competing companies). To qualify for inclusion in the yield trial data-base for a given year/region combination, only years with 5 or more trials were considered. As a result of the application of these filters estimates of attainable yields were obtained for between 5 and 9 years (within the 1999-2007 window) for the various regions. Importantly, the regions that contribute most to the annual national crop (SEBA, OBA, SLLP, NEAR, SOBA, cf. Fig. 1) all had 6 to 9 years of records.

Statistical Analyses

Data for farmer (reporting-district) and multi-environment comparative yield trials were processed to obtain best linear unbiased estimates (BLUE) of annual yield derived from a mixed linear model (MLM, Litell et al., 2006) fitted to data for each crop region using SAS PROC MIXED version 9.3 (SAS Institute, 2006). All MLM variance parameters were estimated by restricted maximum likelihood (REML, Patterson and Thompson, 1971). The MLM analytical approach was chosen to analyze field yield trials because it accounted for incomplete data across years (which is frequent because of temporal changes in hybrids within trials), varying number of hybrids across trials, and different amounts of information per hybrid (Schabenberger and Pierce, 2002).

To estimate mean farmer (reporting-district) yield realised in each year and region we fitted another MLM that considered year as a fixed effect and location within year as a random effect. Resulting BLUEs of annual yields were used to estimate mean farmer yields realized in each year and region.

We fitted a second MLM to the multi-environment trial data base which considered year as a fixed effect and several random effects: trial nested within year, hybrid, hybrid-by-trial interactions, as well as an extra term for residual variability due to differences among replicates within a trial and other unknown sources of variance. The resulting non-error variance within each year and region was partitioned into the contributions (in percent of total non-error variance) of trial, hybrid and hybrid-by-trial interaction components.

Another metric extracted from the comparative yield trial data-base was an estimate, for each year and region, of the mean yield for the top decile of trial results. This metric, which we term Top10, can be regarded as an estimate of the upper yield limit for each year and region. Across all available data-years within each region, the highest observed Top10 in the series was identified. We term this value as the Upper Limit to Rainfed Yield (ULRY).

Finally, we fitted a third MLM to permit between-region contrasts for farmer, attainable, and Top10 yields. These last tests were performed using a separate mixed model (region: fixed effect, year: random effect) which allowed for the imbalance in number of years across regions.

Farmer/Attainable yield gaps were estimated for each year and region as the difference between the corresponding BLUEs of annual yields. Means of these yield gaps, as well as their 95% confidence intervals, were obtained by the bootstrap method implemented in InfoStat (Balzarini *et al*., 2008), and paired t tests were used to assess their statistical significance (α=0.05). Although all analyses were performed for all eight regions, here we report –as illustrations of the general patterns and in order to optimise the use of the space available – the results for the five regions ([SEBA, SOBA, OBA, SLLP and NEAR], see Fig. 1) that between them contributed 81% of the national sunflower harvest.

RESULTS

Attainable yields were significantly greater than farmer yields in all eight regions (and at significance levels of $p< 0.01$ and $p< 0.001$ for the five regions that produced 81 % of the mean national sunflower grain harvest [Table 1]). In seven of the eight regions Top10 yields were 40% or more greater than corresponding attainable yield estimates, and ULRY values were in the 3.9 to 4.8 t/ha range across regions, and were rather similar among regions except for ER (data not shown). Some of the differences between regions in farmer, attainable, and Top10 yields were also significant (Table 1).

Table 1. Farmer (reporting-district), attainable,Top10 and Upper Limit to Rainfed Yield estimates at region and country (where appropriate) levels. Absolute and relative (as % mean farmer yields) values of yield gaps are also shown. Values in brackets in the "Region" column indicate number of years of data used to estimate BLUEs for farmer and attainable yields. Dashes indicate cells for which scaling from regions to country would be inappropriate. Values in cells are means and SEs of BLUEs for annual yields. Asterisks in the "Farmer yield" column indicate significance of differences with attainable yields (* : $p < 0.05$; ** : $p < 0.01$; ***: $p < 0.001$) for each region. Values within a single column followed by different letters are statistically different ($p < 0.05$).

Region	Farmer yield (t/ha) Mean ± SE	Attainable yield (t/ha) Mean ± SE	Absolute (t/ha) and relative (as % of farmer yield , in brackets) Farmer/ attainable yield gaps	Top10 yield (t/ha) Mean ± SE	Upper Limit to Rainfed Yield (t/ha)
SEBA (9)	1.53 ± 0.07 b ***	2.71 ± 0.12 bc	1.18a (77 %)	3.89 ± 0.17ab	4.59
SOBA (6)	1.59 ± 0.08 b ***	2.56 ± 0.12 cd	0.98ab (62 %)	4.15 ± 0.15 a	4.82
OBA (7)	2.21 ± 0.08 a ***	3.05 ± 0.12 a	0.84bc (38 %)	4.06 ± 0.24 a	4.64
SLLP (7)	1.61 ± 0.10 b ***	2.54 ± 0.18 cd	0.92ab (57 %)	3.98 ± 0.19 ab	4.74
NEAR (8)	1.69 ± 0.10 b **	2.22 ± 0.10 e	0.54cd (32 %)	3.63 ± 0.19 bc	4.52
Country	1.85 ± 0.05	2.60 ± 0.05	0.75 (41 %)	-	4.82

Farmer/attainable yield gaps across the five regions were in the 0.5 to 1.2 t ha^{-1} range (Table 1), which translated into 0.75 t ha^{-1} at a country level (encompassing the eight regions), a substantial value when contrasted with a national mean yield of 1.85 t ha^{-1} (Table 1). Importantly, the gap was close to or greater than 60% of farmer yields in three (SEBA, SOBA, SLLP) of the five regions that contribute most to the national sunflower harvest. The absolute magnitude of the farmer/attainable yield gap varied among regions (Table 1), but the significance of these inter-regional differences was limited. Of the five regions contributing most to the national harvest, the magnitude of the gap was statistically indistinguishable between SEBA, SOBA and SLLP, while NEAR exhibited a significantly smaller gap than the regions in included in this trio.

An important feature of the reporting-district and comparative yield trial data bases was yield variability in space and time within each region. Across years the minimum-maximum yield range for farmer yields in all regions exceeded 0.5 t/ha and was close to or above 1 t/ha for attainable yields (data not shown). Our estimates of the spatial and temporal components of the farmer (reporting-district) yield variances indicated that both dimensions were of roughly equal importance (data not shown).

Comparative yield trials also showed considerable between-trial variability within each combination of year and region. The mean contribution of the trial effect to total non-error variance of yields exceeded by far the contributions of the hybrid and the hybrid-by-trial interactions in all eight regions (Table 2).

Table 2. Means and SEs for relative contributions of the effects of trial, hybrid and trial by hybrid interactions to the total non-error variance in attainable yields. Values in brackets in the "Region" column indicate number of years of data used to compute estimates of variance components.

Region	Mean relative contribution of the trial effect (%)	Mean relative contribution of the hybrid effect (%)	Mean relative contribution of the trial by hybrid effect (%)
SEBA (9)	94.0 ± 3.01	1.8 ± 0.57	4.2 ± 2.65
SOBA (7)	94.9 ± 1.44	0.4 ± 0.20	4.8 ± 1.39
OBA (7)	85.0 ± 2.05	5.1 ± 1.62	9.9 ± 2.60
SLLP (7)	87.0 ± 5.37	2.8 ± 1.16	10.2 ± 5.68
NEAR (8)	88.6 ± 0.83	5.0 ±1.05	6.4 ± 0.52

DISCUSSION

Significant (Table 1) and important, in both absolute and relative terms (Table 1), farmer (reporting-district) / attainable yield gaps were a feature of all eight sunflower producing regions of Argentina that have been distinguished in this study. In three of the four regions that contribute most to the national harvest (Table 1), these gaps are close to 1 t ha^{-1}. Expressed as percentages of mean farmer yield, in six of the eight regions, yield gaps exceed the nominal 25% floor which has been suggested as the likely lower limit to yield gaps in commercial farming (Fischer et al., 2009). Three of the four regions contributing most to the national harvest had yield gaps in the 57-77% range. At a country level, the gap is of the order of 0.75 t ha^{-1}, or 41% of the estimated national mean yield of 1.85 t ha^{-}1. Taken together, these results are a strong argument to invest research efforts in determining the management (i.e., amenable to manipulation) constraints to yield that operate in the various regions, in the expectation that identifying these constraints will contribute to devising production strategies that can reduce the gap.

Although differences between regions were not significant for all metrics computed in these analyses, the overall picture indicates that some significant inter-regional differences existed for farmer, attainable, and Top10 yields and for Farmer/Attainable yield gaps (Table 1). When taken together, these results point to the existence of regional differences that need to be studied in order to understand their origin. Their existence constitutes a justification of our use of regionalisation as an element of our approach to the study of yield gaps.

As is to be expected for rain-fed systems such as those of the sunflower producing regions of Argentina, regional attainable yields were considerably less than the mean yields obtained in the most favourable yield trials for each year and region (Top10 yields, Table 1). More surprising is the fact that the ULRY values (Table 1) tended to be rather similar across regions with the exception of the ER region. There is little data on potential (i.e., irrigated) sunflower yield in Argentina. The Top10 estimate obtained from comparative yield trials conducted under irrigation at Ascasubi (slightly South of the SOBA region) over the period 1998-2008 was 5.1 ± 0.17 t ha^{-1}, with an upper limit to irrigated yield of 5.9 t ha^{-1}. Funaro and Pérez Fernández (2005) and Funaro (pers. comm.), using a smaller number of hybrids over three years reported mean maximum yields for irrigated sunflower at Anguil (in the SLLP region) of close to 5.7 ± 0.16 t ha^{-1}. These values compare with our rainfed ULRY values of 4.7 and 4.8 t ha^{-1} for the SOBA and SLLP regions, respectively (Table 2). Although caution needs to be exercised in this context, given the limited availability of data, our ULRY must be regarded as an underestimate of potential yield (in the sense of Fischer et al., 2009) across sunflower-growing regions of Argentina. Thus while in at least some years, very favourable environmental and biotic conditions are explored by sunflower crops in some locations in most regions of Argentina, even these crops appear to have experienced some limitations to yield. Equally, the differences between attainable rainfed yields, on the one hand, and irrigated Top10 and ULRY yields, on the other, underline the inappropriateness of the last two metrics as reference points for regional and country yield gap analyses (something which does not diminish their value in other contexts).

Variability of yield estimates was an important feature of our analyses, both at the reporting-district level (data not shown) and comparative yield trial level (Table 2). At the reporting-district level, the relative importance of the spatial dimension of this variability (in contrast to the temporal dimension) may mean that our regions were too broad and that effort should be invested in searching for regional divisions characterised by a lesser spatial (in relation to temporal) variability. Equally, the temporal dimension of this variability should be explored in an effort to separate, if possible, management (i.e., potentially controllable) and climatic (seasonal water availability and other effects of weather) components. Analysis of the sources of non-error variance in the comparative yield trials (Table 2) clearly shows that trial (i.e., location within region and year) was by far the most important component, far larger than the genotype and the genotype by trial effects. This result is consistent with the findings of Chapman and de la Vega (2002) in their analysis of comparative yield trials conducted by a single seed company in the Northern and Central regions of Argentina. Anderson (2010) found similar contributions of environment to the total variance of rainfed wheat trials in Western Australia, indicating that this is not an issue restricted to sunflower. It is important to note that these results do *not* indicate that genotype is unimportant to growers when deciding the sourcing of their seed. In our analyses we compared yields obtained, in each year and region, using the most current hybrids produced by the various companies that compete in the Argentine market. De la Vega and Chapman (2010) provide examples of the yield cost of sowing superseded hybrids that lack resistance to current strains of fungal diseases, as well as that of not taking advantage to the slow but steady increase in genetic yield potential. Yield variability at these two separate levels provides further support for the argument that it makes good sense to invest further research efforts into the causes of this variability, which probably makes substantial contributions to the farmer / attainable yield gap. Equally, the importance of location in determining variability underlines the dangers of using few research-station experiments to estimate attainable yields.

ACKNOWLEDGEMENTS

Data for the comparative yield trial database was provided by the INTA-ASAGIR testing network and by private seed companies (Asociación de Cooperativas Argentinas ACA S.A.; Advanta S.A.I.C.; Asociados Don Mario S.A. ; Dow AgroSciences Argentina S.A.; Monsanto Argentina S.A.I.C; Nidera S.A.; Pannar Semillas SRL; Profertil S.A.; SPS Argentina). We gratefully acknowledge these contributions. We thank Cecilia Bruno, Laura González and Margot Tablada for data analysis; Valeria Zivec and Constanza Palacios for chasing up data sources; Gonzálo García Accinelli for assistance in drafting Fig. 1 and Abelardo de la Vega and Guillermo Pozzi-Jauregui for critical reading of an earlier version of the manuscript. This project was financed by a grant from ASAGIR. AJH and MB are research scientists with CONICET (National Council for Scientific and Technological Research).

REFERENCES

Aggarwal, P.K., Hebbar, K.B., Venugopalam, M.V., Rani, S., Bala, A., Biswal, A., Wani, S.P., 2008. Quantification of yield gaps in rain-fed rice, wheat, cotton and mustard in India. Global theme on agroecosystems Rept. Nº 43. ICRISAT, Patancheru, India. 36 pp.

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

Balzarini M., Gonzalez L., Tablada M., Casanoves F., Di Rienzo, J., Robledo, C., 2008. InfoStat software estadístico. Manual del usuario. Ed. Brujas, Córdoba, Argentina.

Chapman, S.C., de la Vega, A.J., 2002. Spatial and seasonal effects confounding interpretation of sunflower yields in Argentina. Field Crops Res. 73:107-120.

de la Vega, A.J, Chapman, S.C., 2010. Mega-environment differences affecting genetic progress for yield and relative value of component traits. Crop Sci. 50: 574-583.

Fischer, R.A., Byerlee, D. Edmeades, G.O., 2009. Can technology deliver on the yield challenge to 2050? FAO expert meeting on how to feed the world in 2050. 24-26 June 2009, FAO, Rome. ftp://ftp.fao.org/docrep/fao/012/ak977e/ak977e00.pdf (last accessed, November 12, 2011).

Funaro, D., Pérez Fernández, J., 2005. Rendimiento potencial y fechas de siembra para maíz y girasol, in Cultivos de Cosecha Gruesa. Actualización 2005, Public. Técnica N° 61, EEA Anguil, INTA, pp. 37-39.

Laborte, A.G., de Bie, C.A.J.M., Smaling, E.M.A., Moya, P.F., Boling, A.A., van Ittersum, M.K., 2012 Rice yields and yield gaps in South East Asia: Past trends and future outlook. Eur. J. Agron. 36: 9-20.

Littell R, Milliken G, Stroup W, Wolfinger R, Schabenberger O., 2006. SAS® for Mixed Models, second ed. SAS Institute Inc., Cary, NC.

Lobell, D.B., Cassman, K.G., Field, C.B., 2009. Crop yield gaps: Their importance, magnitudes and causes. Annu. Rev. Environ. Resour. 34; 179-204.

Ministerio de Agricultura, Pesca y Alimentación de la República Argentina, 2011 (http://www.siia.gov.ar/index.php/series-por-tema/agricultura, last accessed November 12, 2011).

Patterson H., Thompson R., 1971. Recovery of inter-block information when block sizes are unequal. Biometrika 58: 545–554.

SAS Institute Inc., 2006. SAS 9.3. SAS Institute Inc., Cary, NC.

Schabenberger, O and Pierce, F., 2002. Contemporary statistical models for the plant and soil sciences. Taylor and Francis-CRC Press.